FEDERAL UNIVERSITY OF PERNAMBUCO
Centre for Natural and Exact Sciences
Graduate Program in Statistics

ESSAYS ON DISTRIBUTION THEORY AND A MODIFIED MOMENT ESTIMATOR FOR THE PRECISION PARAMETER IN A CLASS OF REGRESSION MODELS

RODRIGO BERNARDO DA SILVA

Doctoral thesis

Recife, December 2013.

Federal University of Pernambuco
Centre for Natural and Exact Sciences

Rodrigo Bernardo da Silva

ESSAYS ON DISTRIBUTION THEORY AND A MODIFIED MOMENT ESTIMATOR FOR THE PRECISION PARAMETER IN A CLASS OF REGRESSION MODELS

Doctoral thesis submitted to the Graduate Program in Statistics, Department of Statistics, Federal University of Pernambuco as a partial requirement for obtaining a Ph.D. in Statistics.

Advisor: Professor Dr. Gauss Moutinho Cordeiro
Co-advisor: Professor Dr. Alexandre de Bustamante Simas

Recife, December 2013.

Catalogação na fonte
Bibliotecária Jane Souto Maior, CRB4-571

Silva, Rodrigo Bernardo da
Essays on distribution theory and a modified moment estimator for the precision parameter in a class of regression models / Rodrigo Bernardo da Silva. - Recife: O Autor, 2013.
111 f., il., fig., tab.

Orientador: Gauss Moutinho Cordeiro.
Tese (doutorado) - Universidade Federal de Pernambuco. CCEN, Estatística, 2013.

Inclui referências e apêndice.

1. Estatística Matemática. 2. Probabilidade. 3. Distribuições de probabilidade. I. Cordeiro, Gauss Moutinho (orientador). II. Título.

519.9 CDD (23. ed.) MEI2013 – 169

Universidade Federal de Pernambuco
Pós-Graduação em Estatística

18 de novembro de 2013

Nós recomendamos que a tese de doutorado de autoria de

Rodrigo Bernardo da Silva

intitulada

"*Essays on distribution theory and a modified moment estimator for the precision parameter in a class of regression models*"

seja aceita como cumprimento parcial dos requerimentos para o grau de Doutor em Estatística.

____
Coordenador da Pós-Graduação em Estatística

Banca Examinadora:

____
Gauss Moutinho Cordeiro orientador

____
Francisco Cribari Neto

____
Leandro Chaves Rêgo

____
Artur José Lemonte

____
Giovana Oliveira Silva (UFBA)

Este documento será anexado à versão final da tese.

Agradecimentos

Uma tese de doutorado é, pela sua finalidade acadêmica, um trabalho individual. Entretanto, há contribuições de diversas formas que não podem e nem devem deixar de ser realçadas. Por isso, desejo expressar os meus mais sinceros agradecimentos:

À Deus por iluminar meu caminho todos esses anos e me dar forças para seguir sempre em frente. Gostaria de agradecer à minha mãe Manuela Rebelo e meu pai Paulo Bernardo por todo o amor e suporte, das mais variadas formas, nesses dez anos de universidade. À eles, minha eterna gratidão.

Quero agradecer também minha noiva Adélia Pereira por todo o amor e compreensão com minha falta de tempo para tudo que não fosse acadêmico.

Agradeço de forma especial ao meu orientador, Gauss Moutinho Cordeiro, por toda a ajuda, por ser um brilhante profissional e uma excelente pessoa. Gostaria que ele soubesse que o considero mais que meu orientador, o considero um amigo.

Aos meus amigos de longa data Wagner Barreto, Alexandre Simas, Andréa Rocha, Alessandro Henrique e Saulo Leão. Apesar de não termos mais uma convivência diária os considero como irmãos e espero que nossa amizade seja eterna.

Agradeço aos grandes amigos que fiz durante o doutorado Marcelo Bourguignon, Luz Milena, Jymmy Barreto, Manoel Santos, Cícero Rafael, Fernanda de Bastiani e Waldemar Santa Cruz. À Abraão Nascimento, que conheço desde a minha época de graduação, mas só tive mais convivência durante o doutorado.

À Valéria Bittencourt pela amizade e incrível competência. À Lódino Serbim, seu Cícero e todos os funcionários do departamento de Estatística da UFPE.

À Professora Cristina Raposo pelo apoio e amizade durante toda minha jornada nesses dez anos de universidade.

Aos Professores membros da banca Francisco Cribari, Leandro Rêgo, Giovana Silva e Artur Lemonte por terem aceitado o convite de participar da minha defesa e pelos valiosos comentários.

Agradeço à CAPES pelos auxílios financeiros.

Resumo

Este trabalho está dividido em quatro capítulos independentes. No primeiro, introduzimos um método geral para obter distribuições de probabilidade mais flexíveis. Esse método consiste em compor duas classes de distribuições: as classes Weibull estendida e de séries de potências. O procedimento de composição segue a mesma ideia implementada por Adamidis e Loukas (1998) e, mais recentemente, por Morais e Barreto-Souza (2011). Algumas propriedades matemáticas da nova classe são estudadas, incluindo momentos e função geradora. O método de máxima verossimilhança é utilizado para obter estimativas dos parâmetros. A utilidade da nova classe é exemplicada através de dois exemplos com conjuntos de dados reais. No segundo capítulo, introduzimos e estudamos as propriedades matemáticas de um novo gerador de distribuições de probabilidade contínuas, que adiciona três parâmetros extras. A nova densidade pode ser expressa como combinação linear de densidades exponencializadas da distribuição de origem. Obtemos expressões explícitas para os momentos ordinários e incompletos, funções quantílica e geradora, distribuição assintótica de valores extremos, entropias de Shannon e Rényi e estatísticas de ordem, que valem para qualquer distribuição de origem. O método de máxima verossimilhança é utilizado para obter estimativas dos parâmetros. A potencialidade da nova classe é exemplicada através de dois exemplos com conjuntos de dados reais. No terceiro, propomos um teste da razão de verossimilhanças baseado na estatística de Cox (1961) para discriminar as distribuições exponencial-Poisson e gama. Para isso, consideramos as duas hipóteses nulas: os dados seguem distribuição exponencial-Poisson/gama. A distribuição do logaritmo da razão de verossimilhanças sob a hipótese nula é obtida para os dois casos. Além disso, determinamos o tamanho mínimo de amostra para discriminar as duas distribuições quando a probabilidade de seleção correta é previamente estabelecida. Estudos de simulação e aplicação a dois conjuntos de dados reais exemplificam o comportamento e a utilidade da metodologia proposta, respectivamente. Por fim, propomos um estimador modificado para o parâmetro de precisão para uma extensa classe de modelos de regressão. Por exemplo, os resultados propostos valem para os modelos lineares generalizados (McCullagh e Nelder, 1989), modelos de quase-verossimilhança (Wedderburn, 1974), entre outros. Sabemos que estimador de Pearson para o parâmetro de dispersão funciona bem para ambos os modelos de regressão citados, mas sua versão para o parâmetro de precisão é significativamente viesada para pequenos e médios tamanhos de amostra. Propomos, então, um método simples para a redução do viés do parâmetro de precisão. Estudos de simulação são usados para comparar o desempenho do estimador proposto com outros estimadores.

Palavras-chave: Classe Kumaraswamy generalizada; Distribuição Weibull estendida; Distribuição exponencial-Poisson; Distribuição gama; Parâmetro de dispersão; Parâmetro de precisão; Séries de potência.

Abstract

This work is divided in four independent papers. In the first one, we introduce a general method for obtaining more flexible new distributions by compounding the extended Weibull and power series distributions. The compounding procedure follows the same set-up carried out by Adamidis and Loukas (1998) and, more recently, by Morais and Barreto-Souza (2011). Some mathematical properties of the new class are studied including moments and generating function. The method of maximum likelihood is used for estimating the model parameters. We illustrate the usefulness of the new distributions by means of two applications to real data sets. In the second chapter, we introduce and study general mathematical properties of a new generator of continuous distributions with three extra parameters. The new density function can be expressed as a linear combination of exponentiated densities based on the same baseline distribution. Explicit expressions for the ordinary and incomplete moments, quantile and generating functions, asymptotic distribution of the extreme values, Shannon and Rényi entropies and order statistics, which hold for any baseline model, are determined. We discuss the estimation of the model parameters by maximum likelihood and illustrate the potentiality of the family by means of two applications to real data. In the third chapter, we propose a likelihood ratio test based on Cox's statistic Cox (1961) to discriminate between the exponential-Poisson and gamma distributions. We consider two null hypothesis: the data come from exponential-Poisson/gamma distribution. The asymptotic distribution of the logarithm of the ratio of the maximized likelihoods under the null hypothesis is provided for both cases. We also determine the minimum sample size required for discriminating the two distributions when the probability of correct selection is previously stated. A simulation study and applications to real data sets are presented in order to show the behavior and usefulness of the proposed methodology. Finally, we propose a modified estimator for an extensive class of regression models. For instance, our results hold for generalized linear models (see McCullagh and Nelder, 1989), quasi-likelihood models (see Wedderburn, 1974), among others. The Pearson-based dispersion estimator is known to work well for both generalized linear models and quasi-likelihood models, but the precision parameter version of this estimator is significatively biased in small and medium sample sizes. We thus propose a simple bias-reduction method to reduce the bias of this precision parameter estimator. Monte Carlo simulation is used to compare the performance of the proposed estimator against others.

Keywords: Dispersion parameter; Exponential-Poisson distribution; Extended Weibull distribution; Gamma distribution; Generalized Kumaraswamy class; Power series distribution; Precision parameter.

List of Figures

List of Tables

Contents

CHAPTER 1

Introduction

This thesis is divided in three parts, composed by four independent papers. Two of them introduce new families of distributions; the other two deal with a method to discriminate the exponential-Poisson and gamma distributions, and a modified moment estimator for a general class of regression models, respectively. So, we decide that, for this thesis, each of the papers fills a distinct chapter. Therefore, each chapter can be read independently to each other, since each is self-contained. Additionally, we emphasize that each chapter contains a thorough introduction to the presented matter, so this general introduction only shows, quite briefly, the context of each chapter.

In Chapters 1 and 2, we are interested in the study of probability distributions defined on the positive real line. Roughly speaking, any probability distribution defined on the positive real line can be considered as a lifetime distribution. Obviously, not all such distributions are meaningful for describing an ageing (lifetime) phenomenon. The analysis of lifetime data is an important topic in statistical literature, since its applications range from industrial applications to biological studies. Typically, "lifetime" refers to the time until a specified event, not necessarily the end of a life, such as the life span of a device before it fails, the survival time of a patient with serious disease from the date of diagnosis or major treatment, the time until retirement, the time until marriage/divorce/remarriage amongst others. Because different shapes of lifetime distributions are required for fitting various types of lifetime data, numerous lifetime models were proposed and tested. In fact, Chapters 1 and 2 refer to methods of construction of lifetime distributions that allow to expand the range of shapes of the hazard rate function.

More specifically, Chapter 1 introduces a class of univariate distributions obtained by compounding the extended Weibull and power series distributions. The compounding procedure follows the same one carried out by Adamidis and Loukas (1998) or, more generally, by Chahkandi and Ganjali (2009) and Morais and Barreto-Souza (2011). The hazard function of the proposed class can be decreasing, increasing, bathtub and upside down bathtub. This extension can be derived as follows. Given N, let $X_1, \ldots, X_N$ be independent and identically distributed (iid) continuous random variables following the extended Weibull distribution. Let $X_{(1)} = \min\{X_i\}_{i=1}^{N}$. If we assume that the X_i's are independent of N, which follows a power series distribution (truncated at zero), then $X_{(1)}$ has the extended Weibull power series distribution. To attach a general interpretation to this extension, we may think of a situation where failure, which follows the extended Weibull distribution, occurs due to the presence of an unknown number, say N (which follows the power series distributions), of initial defects of the same kind (a number of semiconductors from a defective lot, for example). The X_i's represent their lifetimes and each defect can be detected only after causing failure, in which case it is repaired perfectly. Then, the distributional assumptions given earlier lead to the extended Weibull power series distribution for modeling the time to the first failure.

Chapter 2 presents a new family of probability distributions, which is based on the alternative gamma-generator defined by Ristíc and Balakrishnan (2011) and extends the generalized Kumaraswamy class (Cordeiro and de Castro, 2011) of distributions, the proportional and reversed hazard rate models, Marshal-Olkin family of distributions and other sub-families. This class provides greater flexibility of its tails and can be widely applied in many areas of engineering and biology. The hazard function of the new family can be decreasing, increasing, bathtub and upside down bathtub. For given parameters values, a physical interpretation of the new distribution can be given as follows. Consider a system formed by α independent components and that each component is made up of λ independent subcomponents. Suppose that the system fails if any of the α components fails and that each component fails if all of the λ subcomponents fail. Let $X_{j1}, \ldots, X_{j\lambda}$ denote the lifetimes of the subcomponents within the jth component, $j = 1, \ldots, \alpha$, having a common cdf G. Let X_j denote de lifetime of the jth component, for $j = 1, \ldots, \alpha$, and let X denote the lifetime of the entire system. Thus, the family of distributions models precisely the time to failure of the entire system.

Chapter 3 deals with the discrimination of two popular lifetime models: the exponential-Poisson and the gamma families. These models have many similarities and in a practical situation, an issue of interest is the selection of the most adequate model (between exponential-Poisson and gamma distributions) to fit a certain continuous lifetime dataset. Therefore, we propose a selection criterion between exponential-Poisson and gamma distributions. To do this, we obtain the asymptotic distribution of the likelihood ratio statistic proposed by Cox (1961, 1962) and with this, we propose a modified test statistic.

Finally, in Chapter 4, we study and reduce the bias of the Pearson-based precision parameter estimation for a large class of regression models. We are considering a large class of regression models, namely, the class of regression models in which the response variables are modelled through the mean, say μ, and which the variance has the form $\sigma^2 V(\mu)$, where σ^2 is a dispersion parameter, $V(\mu)$ is a function of the mean, and $\phi = \sigma^{-2}$ is called the precision parameter. This class of regression models covers many important regression models, such as the generalized linear models (see McCullagh and Nelder, 1989), the exponential family nonlinear models (see Cordeiro and Paula, 1989), the quasi-likelihood models (see Wedderburn, 1974), the extended quasi-likelihood models (see Nelder and Pregibon, 1987), the beta regression models (see Ferrari and Cribari-Neto, 2004), the symmetrical models with mean $\mu \in (0, \infty)$ (see Lange et al., 1989), the dispersion models which are modelled through the mean (see Jørgensen, 1987), among others.

References

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

[2] Chahkandi, M., Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis*, **53**, 4433–4440.

[3] Cordeiro G. M., de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, **81**, 883–898.

[4] Cordeiro, G.M., Paula, G.A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika*, **76**, 93–100.

[5] Cox, D.R. (1961). Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability, Berkeley, University of California Press. p. 105–123.

[6] Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society: Series B*, **24**, 406–424.

[7] Ferrari S.L.P, Cribari-Neto, F. (2004) Beta regression for modeling rates and proportions. *Journal Applied Statistics*, **31**, 799–815.

[8] Lange, K.L., Little, R.J.A., Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, **84**, 881–896.

[9] McCullagh, P., Nelder, J. (1989). Generalized Linear Models, second ed. Chapman & Hall, London.

[10] Morais, A.L., Barreto-Souza, W. (2011). A Compound Class of Weibull and Power Series Distributions. *Computational Statistics and Data Analysis*, **55**, 1410–1425.

[11] Nelder, J.A., Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, **74**, 221–32.

[12] Ristíc, M.M., Balakrishnan, N. (2011). The gamma-exponentiated exponential distribution. *Journal Statistical Computation and Simulation*, **82**, 1191–1206.

[13] Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, **61**, 439–447.

CHAPTER 2

The compound class of extended Weibull power series distributions

Artigo publicado no periódico *Computational Statistics and Data Analysis*, 58, p. 352-367, 2013.

Resumo

Introduzimos um método geral para obter novas distribuições de probabilidade mais flexíveis através da composição das distribuições Weibull estendida e de séries de potências. O processo de composição é o mesmo estabelecido por Marshall e Olkin (1997) e define 68 novos submodelos. A nova classe estende algumas distribuições compostas bem conhecidas como Weibull séries de potências (Morais e Barreto-Souza, 2011) e exponencial séries de potências (Chahkandi e Ganjali, 2009). Algumas propriedades matemáticas da nova classe são estudadas, incluindo momentos e função geradora. Obtemos a densidade das estatísticas de ordem e seus momentos. O método de máxima verossimilhança é usado para a estimação dos parâmetros. Distribuições especiais são investigadas. Ilustramos a utilidade da nova classe de distribuições através de duas aplicações a conjuntos de dados reais.

Palavras-chave: Distribuição de séries de potências; Distribuição Weibull estendida; Distribuição Weibull estendida séries de potências; Estatísticas de ordem.

Abstract

We introduce a general method for obtaining more flexible new distributions by compounding the extended Weibull and power series distributions. The compounding procedure follows the same set-up carried out by Marshall e Olkin (1997) and defines 68 new sub-models. The new class of generated distributions includes some well-known compound distributions, such as the Weibull power series (Morais and Barreto-Souza, 2011) and exponential power series

(Chahkandi and Ganjali, 2009) distributions. Some mathematical properties of the new class are studied including moments and generating function. We provide the density function of the order statistics and their moments. The method of maximum likelihood is used for estimating the model parameters. Special distributions are investigated. We illustrate the usefulness of the new distributions by means of two applications to real data sets.

Keywords: Extended Weibull distribution; Extended Weibull power series distribution; Order statistic; Power series distribution.

2.1 Introduction

The modeling and analysis of lifetimes is an important issue of statistical work in a wide variety of scientific and technological fields. Several distributions have been proposed in the literature to model lifetime data by compounding some useful lifetime distributions. Adamidis and Loukas (1998) introduced a two-parameter exponential geometric (EG) distribution by compounding the exponential and geometric distributions. In a similar manner, the exponential Poisson (EP) and exponential logarithmic (EL) distributions were introduced and studied by Kus (2007) and Tahmasbi and Rezaei (2008), respectively. Recently, Chahkandi and Ganjali (2009) proposed the exponential power series (EPS) family of distributions, which contains as special cases these distributions. Barreto-Souza et al. (2010) and Lu and Shi (2011) introduced the Weibull geometric (WG) and Weibull Poisson (WP) distributions which naturally extend the EG and EP distributions, respectively. In a recent paper, Morais and Barreto-Souza (2011) defined the Weibull power series (WPS) class of distributions which includes as submodels the EPS distributions. The WPS distributions can have increasing, decreasing and upside down bathtub failure rate functions. The generalized exponential power series (GEPS) distributions were proposed by Mahmoudi and Jafari (2012) following the same approach developed by Morais and Barreto-Souza (2011). Another recent compounded distribution can be found in Cancho et al. (2011, 2012) who introduced the Poisson exponential (PE) and geometric Birnbaum-Saunders (GBS) distributions, and Barreto-Souza and Bakouch (2012) who defined the Poisson Lindley (PL) distribution. Further, Louzada et al. (2011) and Cordeiro et al. (2012) proposed the complementary exponential geometric (CEG) and the exponential COM Poisson (ECOMP) distributions, respectively.

The Weibull distribution was one of the earliest and most popular model for failure times. In recent years, many authors have proposed generalizations of the Weibull model based on extended types of failure of a system. In the context of modeling random strength of brittle materials and failure times, Gurvich et al (1997) proposed an extended Weibull (EW) family of distributions. Nadarajah and Kotz (2005) and Pham and Lai (2007) presented much more than twenty useful distributions in their family. The EW cumulative distribution function (cdf) is given by

$$G(x;\alpha,\boldsymbol{\xi}) = 1 - \mathrm{e}^{-\alpha\, H(x;\boldsymbol{\xi})}, \quad x > 0,\ \ \alpha > 0, \tag{2.1}$$

where $H(x;\boldsymbol{\xi})$ is a non-negative monotonically increasing function which depends on a parameter vector $\boldsymbol{\xi}$. The corresponding probability density function (pdf) becomes

$$g(x;\, \alpha, \boldsymbol{\xi}) = \alpha\, h(x;\, \boldsymbol{\xi})\, \mathrm{e}^{-\alpha\, H(x;\boldsymbol{\xi})}, \quad x > 0,\ \ \alpha > 0, \tag{2.2}$$

where $h(x;\, \boldsymbol{\xi})$ is the first derivative of $H(x;\, \boldsymbol{\xi})$. Many well-known models are special cases of equation (2.1) such as:

(*i*) $H(x; \boldsymbol{\xi}) = x$ gives the exponential distribution;
(*ii*) $H(x; \boldsymbol{\xi}) = x^2$ yields the Rayleigh distribution (Burr type-X distribution);
(*iii*) $H(x; \boldsymbol{\xi}) = \log(x/k)$ leads to the Pareto distribution;
(*iv*) $H(x; \boldsymbol{\xi}) = \beta^{-1}[\exp(\beta x) - 1]$ gives the Gompertz distribution.

We emphasize that several other distributions could be re-written in form (2.1) (see some examples in Nadarajah and Kotz, 2005; and Pham and Lai, 2007). In this chapter, we define the extended Weibull power series (EWPS) class of univariate distributions obtained by compounding the extended Weibull and power series distributions. The compounding procedure follows the key idea of Adamidis and Loukas (1998) or, more generally, by Chahkandi and Ganjali (2009) and Morais and Barreto-Souza (2011). The new class of distributions includes as special models the WPS distributions, which in turn extends the EPS distributions and defines 68 (17×4) new sub-models as special cases. The hazard function of the proposed class can be decreasing, increasing, bathtub and upside down bathtub. We are motivated to introduce the EWPS distributions because of the wide usage of (2.1) and the fact that the current generalization provides means of its continuous extension to still more complex situations.

This chapter is organized as follows. In Section 2.2, we define the EWPS class of distributions and demonstrate that there are many existing models which can be deduced as special cases of the proposed unified model. In Section 2.3, we provide general properties of the EWPS class including the density, survival and hazard rate functions, some useful expansions, quantiles, ordinary and incomplete moments, generating function, order statistics and their moments, reliability and average lifetime. Estimation of the parameters by maximum likelihood is investigated in Section 2.4. In Section 2.5, we present suitable constraints leading to the maximum entropy characterization of the new class. Three special models of the proposed class are studied in Section 2.6. Applications to two real data sets are presented in Section 2.7. Some concluding remarks are addressed in Section 2.8.

2.2 The new class

Let N be a discrete random variable having a power series distribution (truncated at zero) with probability mass function

$$p_n = P(N = n) = \frac{a_n\, \theta^n}{C(\theta)},\ n = 1, 2, \ldots, \tag{2.3}$$

where a_n depends only on n, $C(\theta) = \sum_{n=1}^{\infty} a_n\, \theta^n$ and $\theta > 0$ is such that $C(\theta)$ is finite. Table 2.1 summarizes some power series distributions (truncated at zero) defined according to (2.3) such as the Poisson, logarithmic, geometric and binomial distributions. The proposed class of distributions can be derived as follows. Given N, let $X_1, \ldots, X_N$ be independent and identically distributed (iid) random variables following (2.1). Let $X_{(1)} = \min\{X_i\}_{i=1}^{N}$. The conditional cumulative distribution of $X_{(1)}|N = n$ is given by

$$G_{X_{(1)}|N=n}(x) = 1 - \mathrm{e}^{-n\alpha H(x;\boldsymbol{\xi})},$$

i.e., $X_{(1)}|N = n$ has the general class of distributions (2.1) with parameters $n\alpha$ and $\boldsymbol{\xi}$ based on the same $H(x;\boldsymbol{\xi})$ function. Hence, we obtain

$$P(X_{(1)} \leq x, N = n) = \frac{a_n\,\theta^n}{C(\theta)}\left[1 - \mathrm{e}^{-n\alpha H(x;\boldsymbol{\xi})}\right], \quad x > 0, \quad n \geq 1.$$

The EWPS class of distributions is defined by the marginal cdf of $X_{(1)}$:

$$F(x;\theta,\alpha,\boldsymbol{\xi}) = 1 - \frac{C(\theta\,\mathrm{e}^{-\alpha H(x;\boldsymbol{\xi})})}{C(\theta)}, \quad x > 0. \tag{2.4}$$

We provide at least four motivations for the EWPS class of distributions, which can be applied in some interesting situations as follows:

1. Time to the first failure. Suppose that the failure of a device occurs due to the presence of an unknown number N of initial defects of same kind, which can be identifiable only after causing failure and are repaired perfectly. Define by X_i the time to the failure of the device due to the ith defect, for $i \geq 1$. If we assume that the X_i's are iid EW random variables independent of N, which follows a power series distribution (truncated at zero), then the time to the first failure is appropriately modeled by the EWPS distribution.

2. Reliability. From the stochastic representations $X = \min\{X_i\}_{i=1}^{N}$ or $X = \max\{X_i\}_{i=1}^{N}$, we note that the EWPS model arises in series (for the minimum of EW distributions) or parallel systems (for the maximum of the EW distributions) with identical components, which appear in many industrial applications and biological organisms.

3. Time to relapse of cancer under the first-activation scheme. Here N is the number of carcinogenic cells for an individual left active after the initial treatment and X_i is the time spent for the ith carcinogenic cell to produce a detectable cancer mass, for $i \geq 1$. Assuming that $\{X_i\}_{i\geq 1}$ is a sequence of iid EW random variables independent of N, which follows a power series distribution (truncated at zero), we have that the time to relapse of cancer of a susceptible individual can be modeled by the EWPS class of distributions.

4. Last-activation scheme. As discussed by Cooner et al. (2007), the first activation scheme may be questioned by certain diseases. Let N be the number of latent factors that must all be active by failure and X_i be the time of resistance to a disease manifestation due to the

ith latent factor. In the last-activation scheme (for the maximum of the EW distributions), it is assumed that failure occurs after all N factors have been active. So, if the X_i's are iid EW random variables independent of N, where N follows a zero-truncated power series distribution, the EWPS class can be able for modeling the time to the failure under the last-activation scheme.

Table 2.1: Useful quantities for some power series distributions.

Distribution	a_n	$C(\theta)$	$C'(\theta)$	$C''(\theta)$	$C(\theta)^{-1}$	Θ
Poisson	$n!^{-1}$	$e^{\theta}-1$	e^{θ}	e^{θ}	$\log(\theta+1)$	$\theta \in (0,\infty)$
Logarithmic	n^{-1}	$-\log(1-\theta)$	$(1-\theta)^{-1}$	$(1-\theta)^{-2}$	$1-e^{-\theta}$	$\theta \in (0,1)$
Geometric	1	$\theta(1-\theta)^{-1}$	$(1-\theta)^{-2}$	$2(1-\theta)^{-3}$	$\theta(\theta+1)^{-1}$	$\theta \in (0,1)$
Binomial	$\binom{m}{n}$	$(\theta+1)^m-1$	$m(\theta+1)^{m-1}$	$\dfrac{m(m-1)}{(\theta+1)^{2-m}}$	$(\theta-1)^{1/m}-1$	$\theta \in (0,1)$

Hereafter, the random variable X following (2.4) with parameters θ and α and vector of parameters $\boldsymbol{\xi}$ is denoted by $X \sim \text{EWPS}(\theta,\alpha,\boldsymbol{\xi})$. Equation (2.4) extends several distributions which have been studied in the literature. The EG distribution (Adamidis and Loukas, 1998) is obtained by taking $H(x;\boldsymbol{\xi}) = x$ and $C(\theta) = \theta(1-\theta)^{-1}$ with $\theta \in (0,1)$. Further, for $H(x;\boldsymbol{\xi}) = x$, we obtain the EP (Kus, 2007) and EL (Tahmasbi and Rezaei, 2008) distributions by taking $C(\theta) = e^{\theta}-1, \theta > 0$, and $C(\theta) = -\log(1-\theta), \theta \in (0,1)$, respectively. In the same way, for $H(x;\boldsymbol{\xi}) = x^{\gamma}$, we obtain the WG (Barreto-Souza et al., 2009) and WP (Lu and Shi, 2011) distributions. The EPS distributions come from (2.4) by combining $H(x;\boldsymbol{\xi}) = x$ with any $C(\theta)$ listed in Table 2.1 (see Chahkandi and Ganjali, 2009). Finally, we obtain the WPS distributions from (2.4) by compounding $H(x;\boldsymbol{\xi}) = x^{\gamma}$ with any $C(\theta)$ in Table 2.1 (see Morais and Barreto-Souza, 2011). Table 2.2 displays some useful quantities and corresponding parameter vectors for special distributions.

2.3 General properties

2.3.1 Density, survival and hazard functions

The density function associated to (2.4) is given by

$$f(x;\theta,\alpha,\boldsymbol{\xi}) = \theta\,\alpha\,h(x;\boldsymbol{\xi})\,e^{-\alpha H(x;\boldsymbol{\xi})}\,\frac{C'(\theta\,e^{-\alpha H(x;\boldsymbol{\xi})})}{C(\theta)}, \quad x > 0. \tag{2.5}$$

Proposition 1. *The EW class of distributions with parameters $c\alpha$ and $\boldsymbol{\xi}$ is a limiting special case of the EWPS class of distributions when $\theta \to 0^+$, where $c = \min\{n \in \mathbb{N} : a_n > 0\}$.*

Table 2.2: Special distributions and corresponding $H(x;\,\boldsymbol{\xi})$ and $h(x;\,\boldsymbol{\xi})$ functions.

Distribution	$H(x;\boldsymbol{\xi})$	$h(x;\boldsymbol{\xi})$	α	$\boldsymbol{\xi}$	References
Exponential ($x \geq 0$)	x	1	α	$\emptyset$	Johnson et al. (1994)
Pareto ($x \geq k$)	$\log(x/k)$	$1/x$	α	k	Johnson et al. (1994)
Rayleigh ($x \geq 0$)	x^2	$2x$	α	$\emptyset$	Rayleigh (1880)
Weibull ($x \geq 0$)	x^γ	$\gamma x^{\gamma-1}$	α	γ	Johnson et al. (1994)
Modified Weibull ($x \geq 0$)	$x^\gamma \exp(\lambda x)$	$x^{\gamma-1}\exp(\lambda x)(\gamma+\lambda x)$	α	$[\gamma, \lambda]$	Lai et al. (2003)
Weibull extension ($x \geq 0$)	$\lambda[\exp(x/\lambda)^\beta - 1]$	$\beta\exp(x/\lambda)^\beta(x/\lambda)^{\beta-1}$	α	$[\gamma, \lambda, \beta]$	Xie et al. (2002)
Log-Weibull ($-\infty < x < \infty$)	$\exp[(x-\mu)/\sigma]$	$(1/\sigma)\exp[(x-\mu)/\sigma]$	1	$[\mu, \sigma]$	White (1969)
Phani ($0 < \mu < x < \sigma < \infty$)	$[(x-\mu)/(\sigma-x)]^\beta$	$\beta[(x-\mu)/(\sigma-x)]^{\beta-1}[(\sigma-\mu)/(\sigma-t)^2]$	α	$[\mu, \sigma, \beta]$	Phani (1987)
Weibull Kies ($0 < \mu < x < \sigma < \infty$)	$(x-\mu)^{\beta_1}/(\sigma-x)^{\beta_2}$	$(x-\mu)^{\beta_1-1}(\sigma-x)^{-\beta_2-1}[\beta_1(\sigma-x)+\beta_2(x-\mu)]$	α	$[\mu, \sigma, \beta_1, \beta_2]$	Kies (1958)
Additive Weibull ($x \geq 0$)	$(x/\beta_1)^{\alpha_1} + (x/\beta_2)^{\alpha_2}$	$(\alpha_1/\beta_1)(x/\beta_1)^{\alpha_1-1} + (\alpha_2/\beta_2)(x/\beta_2)^{\alpha_2-1}$	1	$[\alpha_1, \alpha_2, \beta_1, \beta_2]$	Xie and Lai (1995)
Traditional Weibull ($x \geq 0$)	$x^b[\exp(cx^d - 1)]$	$bx^{b-1}[\exp(cx^d) - 1] + cdx^{b+d-1}\exp(cx^d)$	α	$[b, c, d]$	Nadarajah and Kotz (2005)
Gen. power Weibull ($x \geq 0$)	$[1+(x/\beta)^{\alpha_1}]^\theta - 1$	$(\theta\alpha_1/\beta)[1+(x/\beta)^{\alpha_1}]^{\theta-1}(x/\beta)^{\alpha_1}$	1	$[\alpha_1, \beta, \theta]$	Nikulin and Haghighi (2006)
Flexible Weibull extension($x \geq 0$)	$\exp(\alpha_1 x - \beta/x)$	$\exp(\alpha_1 x - \beta/x)(\alpha_1 + \beta/x^2)$	1	$[\alpha_1, \beta]$	Bebbington et al. (2007)
Gompertz ($x \geq 0$)	$\beta^{-1}[\exp(\beta x) - 1]$	$\exp(\beta x)$	α	β	Gompertz (1825)
Exponential power ($x \geq 0$)	$\exp[(\lambda x)^\beta] - 1$	$\beta\lambda\exp[(\lambda x)^\beta](\lambda x)^{\beta-1}$	1	$[\lambda, \beta]$	Smith and Bain (1975)
Chen ($x \geq 0$)	$\exp(x^b) - 1$	$bx^{b-1}\exp(x^b)$	α	b	Chen (2000)
Pham ($x \geq 0$)	$(a^x)^\beta - 1$	$\beta(a^x)^\beta\log(a)$	1	$[a, \beta]$	Pham (2002)

Proof. This proof uses a similar argument given by Morais and Barreto-Souza (2011). Define $c = \min\{n \in \mathbb{N} : a_n > 0\}$. For $x > 0$, we have

$$\lim_{\theta\to 0^+} F(x) = 1 - \lim_{\theta\to 0^+} \frac{\sum_{n=c}^{\infty} a_n \left(\theta\,\mathrm{e}^{-\alpha H(x;\boldsymbol{\xi})}\right)^n}{\sum_{n=c}^{\infty} a_n\,\theta^n}$$

$$= 1 - \lim_{\theta\to 0^+} \frac{\mathrm{e}^{-c\alpha H(x;\boldsymbol{\xi})} + a_c^{-1} \sum_{n=c+1}^{\infty} a_n\,\theta^{n-c}\mathrm{e}^{-n\alpha H(x;\boldsymbol{\xi})}}{1 + a_c^{-1} \sum_{n=c+1}^{\infty} a_n\,\theta^{n-c}}$$

$$= 1 - \mathrm{e}^{-c\alpha H(x;\boldsymbol{\xi})}.$$

□

We now provide an interesting expansion for (2.5). We have $C'(\theta) = \sum_{n=1}^{\infty} n\,a_n\,\theta^{n-1}$. By using this result in (2.5), we obtain

$$f(x;\theta,\alpha,\boldsymbol{\xi}) = \sum_{n=1}^{\infty} p_n\, g(x;\, n\alpha,\boldsymbol{\xi}), \tag{2.6}$$

where $g(x;\, n\alpha,\boldsymbol{\xi})$ is given by (2.2). Based on equation (2.6), we obtain

$$F(x;\theta,\alpha,\boldsymbol{\xi}) = 1 - \sum_{n=1}^{\infty} p_n\,\mathrm{e}^{-n\alpha H(x;\boldsymbol{\xi})}.$$

Hence, the EWPS density function is an infinite mixture of EW densities. So, some mathematical quantities (such as ordinary and incomplete moments, generating function and mean deviations) of the EWPS distributions can be obtained by knowing those quantities for the baseline density function $g(x;\, n\alpha,\boldsymbol{\xi})$. The EWPS survival function becomes

$$S(x;\theta,\alpha,\boldsymbol{\xi}) = \frac{C(\theta\,\mathrm{e}^{-\alpha H(x;\boldsymbol{\xi})})}{C(\theta)}, \tag{2.7}$$

and the corresponding hazard rate function reduces to

$$\tau(x;\theta,\alpha,\boldsymbol{\xi}) = \theta\alpha\, h(x;\,\boldsymbol{\xi})\,\mathrm{e}^{-\alpha H(x;\boldsymbol{\xi})}\,\frac{C'(\theta\,\mathrm{e}^{-\alpha H(x;\boldsymbol{\xi})})}{C(\theta\,\mathrm{e}^{-\alpha H(x;\boldsymbol{\xi})})}.$$

2.3.2 Quantiles, moments and order statistics

The EWPS distribution is simulated from (2.4) as follows: if $U \sim (0,1)$, the solution of the nonlinear equation

$$X = H^{-1}\left\{-\frac{1}{\alpha}\log\left[\frac{C^{-1}(C(\theta)(1-U))}{\theta}\right]\right\}$$

has the EWPS$(\theta,\alpha,\boldsymbol{\xi})$ distribution, where $H^{-1}(\cdot)$ and $C^{-1}(\cdot)$ are the inverse functions of $H(\cdot)$ and $C(\cdot)$, respectively. To simulate data from this nonlinear equation, we can use the matrix programming language `Ox` through `SolveNLE` subroutine (see Doornik, 2007).

Many of the important characteristics and features of a distribution are obtained through the moment generating function (mgf) and moments. The rth raw moment of X can be determined from (2.6) and the monotone convergence theorem. So, for $r \in \mathbb{N}$, we obtain

$$\mathrm{E}(X^r) = \sum_{n=1}^{\infty} p_n \; \mathrm{E}(Z^r).$$

Hereafter, Z denotes a random variable with density function $g(z; n\alpha, \boldsymbol{\xi})$.

The incomplete moments (I_X) and mgf (M_X) of X can be determined from (2.6) using the monotone convergence theorem:

$$I_X(y) = \int_0^y x^r f(x) dx = \sum_{n=1}^{\infty} p_n \, I_Z(y)$$

and

$$M_X(t) = \sum_{n=1}^{\infty} p_n \; \mathrm{E}\left(\mathrm{e}^{tZ}\right).$$

Order statistics are among the most fundamental tools in non-parametric statistics and inference. They enter in problems of estimation and hypothesis tests in a variety of ways. Then, we now discuss some properties of the order statistics for the proposed class of distributions. The pdf $f_{i:m}(x)$ of the ith order statistic from a random sample $X_1, \ldots, X_m$ having density function (2.5) is given by

$$f_{i:m}(x) = \frac{m!}{(i-1)!(m-i)!} f(x; \theta, \alpha, \boldsymbol{\xi}) \left[1 - \frac{C(\theta \, \mathrm{e}^{-\alpha H(x;\boldsymbol{\xi})})}{C(\theta)}\right]^{i-1} \left[\frac{C(\theta \mathrm{e}^{-\alpha H(x;\boldsymbol{\xi})})}{C(\theta)}\right]^{m-i}, \quad x > 0. \tag{2.8}$$

By using the binomial expansion, we can write (2.8) as

$$f_{i:m}(x) = \frac{m!}{(i-1)!(m-i)!} f(x; \theta, \alpha, \boldsymbol{\xi}) \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} S(x; \theta, \alpha, \boldsymbol{\xi})^{m+j-i},$$

where $S(x; \theta, \alpha, \boldsymbol{\xi})$ is given by (2.7). The corresponding cumulative function becomes

$$F_{i:m}(x) = \sum_{j=0}^{k} \sum_{k=i}^{m} (-1)^j \binom{k}{j} \binom{m}{k} S(x; \theta, \alpha, \boldsymbol{\xi})^{m+j-k}.$$

An alternative form for (2.8) can be obtained from (2.6) as

$$f_{i:m}(x) = \frac{m!}{(i-1)!(m-i)!} \sum_{n=1}^{\infty} \sum_{j=0}^{i-1} \omega_j \, p_n \, g(x; n\alpha, \boldsymbol{\xi}) S(x; \theta, \alpha, \boldsymbol{\xi})^{m+j-1}, \tag{2.9}$$

where $\omega_j = (-1)^j \binom{i-1}{j}$. So, the sth raw moment of $X_{i:m}$ comes immediately from the above equation

$$\mathrm{E}\left(X_{i:m}^s\right) = \frac{m!}{(i-1)!(m-i)!} \sum_{n=1}^{\infty} \sum_{j=0}^{i-1} \omega_j \, p_n \; \mathrm{E}\left[Z^s \, S(Z; \theta, \alpha, \boldsymbol{\xi})^{m+j-i}\right]. \tag{2.10}$$

2.3.3 Average lifetime

The average lifetime is given by

$$t_m = \sum_{n=1}^{\infty} p_n \int_0^{\infty} \mathrm{e}^{-n\alpha H(x;\boldsymbol{\xi})}\, dx.$$

In fields such as actuarial sciences, survival studies and reliability theory, the concept of mean residual life has been of much interest; see a survey by Guess and Proschan (1985). Given that there was no failure prior to x_0, the residual life is the period from time x_0 until the time of failure. The mean residual lifetime can be expressed as

$$\begin{aligned} m(x_0;\theta,\alpha,\boldsymbol{\xi}) &= [\Pr(X > x_0)]^{-1} \int_0^{\infty} y\, f(x_0+y;\theta,\alpha,\boldsymbol{\xi})dy \\ &= [S(x_0)]^{-1} \sum_{n=1}^{\infty} p_n \int_0^{\infty} y\, g(x_0+y;n\alpha,\boldsymbol{\xi})dy. \end{aligned}$$

The latter integral can be computed from the baseline EW distribution. Further, we have that $m(x_0;\theta,\alpha,\boldsymbol{\xi}) \to \mathrm{E}(X)$ as $x_0 \to 0$. Some results of this section can be obtained numerically in any symbolic software such as `MAPLE` (Garvan, 2002), `MATLAB` (Sigmon and Davis, 2002), `MATHEMATICA` (Wolfram, 2003), `Ox` (Doornik, 2007) and `R` (`R` Development Core Team, 2009). The `Ox` (for academic purposes) and `R` are freely distributed and available at `http://www.doornik.com` and `http://www.r-project.org`, respectively. The infinity limit in these sums can be substituted by a large positive integer such as 20 or 30 for most practical purposes.

2.4 Maximum likelihood estimation

Here, we determine the maximum likelihood estimates (MLEs) of the parameters of the EWPS class of distributions from complete samples only. Let $x_1,\ldots,x_n$ be observed values from the EWPS distribution with parameters θ,α and $\boldsymbol{\xi}$. Let $\Theta = (\theta,\alpha,\boldsymbol{\xi})^{\top}$ be the $p\times 1$ parameter vector. The total log-likelihood function for Θ is given by

$$\begin{aligned} \ell_n &= \ell_n(\Theta) = n\,[\log(\theta) + \log(\alpha) - \log(C(\theta))] - \alpha \sum_{i=1}^{n} H(x_i;\,\boldsymbol{\xi}) + \sum_{i=1}^{n} \log[h(x_i;\,\boldsymbol{\xi})] \\ &+ \sum_{i=1}^{n} \log[C'(\theta\, \mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})})]. \end{aligned} \qquad (2.11)$$

The log-likelihood function can be maximized either directly by using the `SAS` (`PROC NLMIXED`) or the `Ox` program (sub-routine `MaxBFGS`) (see Doornik, 2007) or by solving the nonlinear likelihood equations obtained by differentiating (2.11). The components of the score function

$U_n(\Theta) = (\partial\ell_n/\partial\theta, \partial\ell_n/\partial\alpha, \partial\ell_n/\partial\boldsymbol{\xi})^\top$ are

$$\frac{\partial\ell_n}{\partial\alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} H(x_i;\,\boldsymbol{\xi}) - \theta\sum_{i=1}^{n} H(x_i;\,\boldsymbol{\xi})\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})}\,\frac{C''(\theta\,\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})})}{C'(\theta\,\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})})},$$

$$\frac{\partial\ell_n}{\partial\theta} = \frac{n}{\theta} - n\frac{C'(\theta)}{C(\theta)} + \sum_{i=1}^{n}\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})}\,\frac{C''(\theta\,\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})})}{C'(\theta\,\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})})},$$

$$\frac{\partial\ell_n}{\partial\boldsymbol{\xi}_k} = \sum_{i=1}^{n}\frac{\partial\log h(x_i;\,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_k} - \alpha\sum_{i=1}^{n}\frac{\partial H(x_i;\,\xi)}{\partial\boldsymbol{\xi}_k}\left[1+\theta\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})}\,\frac{C''(\theta\,\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})})}{C'(\theta\,\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})})}\right].$$

For interval estimation on the model parameters, we require the observed information matrix

$$J_n(\Theta) = -\begin{pmatrix} J_{\theta\theta} & J_{\theta\alpha} & | & J_{\theta\boldsymbol{\xi}}^\top \\ J_{\alpha\theta} & J_{\alpha\alpha} & | & J_{\alpha\boldsymbol{\xi}}^\top \\ -- & -- & -- & -- \\ J_{\theta\boldsymbol{\xi}} & J_{\alpha\boldsymbol{\xi}} & | & J_{\boldsymbol{\xi}\boldsymbol{\xi}} \end{pmatrix},$$

whose elements are listed in Appendix A. Let $\widehat{\Theta}$ be the MLE of Θ. Under standard regular conditions (Cox and Hinkley, 1974) that are fulfilled for the proposed model whenever the parameters are in the interior of the parameter space, we can approximate the distribution of $\sqrt{n}(\widehat{\Theta}-\Theta)$ by the multivariate normal $N_p(0, K(\Theta)^{-1})$, where $K(\Theta) = \lim_{n\to\infty}\frac{1}{n}J_n(\Theta)$ is the unit information matrix and p is the number of parameters of the compounded distribution.

Often with lifetime data and reliability studies, one encounters censoring. A very simple random censoring mechanism very often realistic is one in which each individual i is assumed to have a lifetime X_i and a censoring time C_i, where X_i and C_i are independent random variables. Suppose that the data consist of n independent observations $x_i = \min(X_i, C_i)$ and $\delta_i = I(X_i \le C_i)$ is such that $\delta_i = 1$ if X_i is a time to event and $\delta_i = 0$ if it is right censored for $i = 1, \ldots, n$. The censored likelihood $L(\Theta)$ for the model parameters is

$$L(\Theta) \propto \prod_{i=1}^{n}[f(x_i;\theta,\alpha,\boldsymbol{\xi})]^{\delta_i}\,[S(x_i;\theta,\alpha,\boldsymbol{\xi})]^{1-\delta_i},$$

where $f(x;\theta,\alpha,\boldsymbol{\xi})$ and $S(x;\theta,\alpha,\boldsymbol{\xi})$ are given in (2.5) and (2.7), respectively.

2.5 Maximum entropy identification

The concept of Shannon entropy is the central role of information theory sometimes referred as measure of uncertainty. The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty. Shannon (1948) introduced the probabilistic definition of entropy which is closely connected with the definition of entropy in statistical mechanics. Let X be a random variable of a continuous distribution with density f. The Shannon entropy of X is defined by

$$\mathbb{H}_{Sh}(f) = -\int_{\mathbb{R}} f(x;\theta,\alpha,\boldsymbol{\xi})\log\left[f(x;\theta,\alpha,\boldsymbol{\xi})\right]dx. \tag{2.12}$$

Jaynes (1957) introduced one of the most powerful techniques employed in the field of probability and statistics called the "*maximum entropy method*". This method is closely related to the Shannon entropy and considers a class of density functions

$$\mathbb{F} = \left\{ f(x;\theta,\alpha,\boldsymbol{\xi}) : \mathrm{E}_f(T_i(X)) = \alpha_i,\, i = 0,\ldots,m \right\},$$

where $T_1(X),\ldots,T_m(X)$ are absolutely integrable functions with respect to f, and $T_0(X) = a_0 = 1$. In the continuous case, the maximum entropy principle suggests deriving the unknown density function of the random variable X by the model that maximizes the Shannon entropy in (2.12), subject to the information constraints defined in the class $\mathbb{F}$. Shore and Johnson (1980) treated the maximum entropy method axiomatically. This method has been successfully applied in a wide variety of fields and has also been used for the characterization of several standard probability distributions; see, for example, Kapur (1989), Soofi (2000) and Zografos and Balakrishnan (2009).

The maximum entropy distribution is the density of the class $\mathbb{F}$, denoted by f^{ME}, determined as the solution of the optimization problem

$$f^{ME}(x;\theta,\alpha,\boldsymbol{\xi}) = \arg\max_{f\in\mathbb{F}} \mathbb{H}_{Sh}.$$

Jaynes (1957, p. 623) states that the maximum entropy distribution f^{ME} obtained by the constrained maximization problem described above, "*is the only unbiased assignment we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have*". It is the distribution which should not incorporate additional exterior information other than which is specified by the constraints. We now derive suitable constraints in order to provide a maximum entropy characterization for the class (2.4). For this purpose, the next result plays an important role.

Proposition 2. *Let X be a random variable with pdf given by* (2.5)*. Then,*

C1. $\mathrm{E}\left\{\log[C'(\theta\, \mathrm{e}^{-\alpha H(X;\boldsymbol{\xi})})]\right\} = \dfrac{\theta}{C(\theta)}\, \mathrm{E}\left\{C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})}) \log[C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})})]\right\};$

C2. $\mathrm{E}\left\{\log[h(X;\,\boldsymbol{\xi})]\right\} = \dfrac{\theta}{C(\theta)}\, \mathrm{E}\left\{C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})}) \log[h(Y;\,\boldsymbol{\xi})]\right\};$

C3. $\mathrm{E}\left[H(X;\,\boldsymbol{\xi})\right] = \dfrac{\theta}{C(\theta)}\, \mathrm{E}\left[C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})}) H(Y;\,\boldsymbol{\xi})\right],$

where Y follows the EW distribution with density function (2.2)*.*

Proof. The constraints C1, C2 and C3 are easily demonstrated and then the proofs are omitted. □

The next proposition reveals that the EWPS distribution has maximum entropy in the class of all probability distributions specified by the constraints stated in the previous proposition.

Proposition 3. *The pdf f of a random variable X, given by* (2.5)*, is the unique solution of the optimization problem*

$$f(x;\theta,\alpha,\boldsymbol{\xi}) = \arg\max_{\tau} \mathbb{H}_{Sh},$$

under the constraints C1*,* C2 *and* C3 *presented in the Proposition 2.*

Proof. Let τ be a pdf representing the distribution of X that satisfies the constraints C1, C2 and C3. The Kullback-Leibler divergence between τ and f is

$$D(\tau, f) = \int_{\mathbb{R}} \tau(x;\theta,\alpha,\boldsymbol{\xi}) \log\left[\frac{\tau(x;\theta,\alpha,\boldsymbol{\xi})}{f(x;\theta,\alpha,\boldsymbol{\xi})}\right] dx.$$

Following Cover and Thomas (1991), we obtain

$$\begin{aligned} 0 \le D(\tau, f) &= \int_{\mathbb{R}} \tau(x;\theta,\alpha,\boldsymbol{\xi}) \log\left[\tau(x;\theta,\alpha,\boldsymbol{\xi})\right] dx - \int_{\mathbb{R}} \tau(x;\theta,\alpha,\boldsymbol{\xi}) \log\left[f(x;\theta,\alpha,\boldsymbol{\xi})\right] dx \\ &= -\mathbb{H}_{Sh}(\tau;\theta,\alpha,\boldsymbol{\xi}) - \int_{\mathbb{R}} \tau(x;\theta,\alpha,\boldsymbol{\xi}) \log\left[f(x;\theta,\alpha,\boldsymbol{\xi})\right] dx. \end{aligned}$$

From the definition of f and based on the constraints C1, C2 and C3, we have

$$\begin{aligned} \int_{\mathbb{R}} \tau(x) \log\left[f(x)\right] dx &= \log(\theta\alpha) + \frac{\theta}{C(\theta)} \,\mathrm{E}\left\{C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})}) \log\left[h(Y;\,\boldsymbol{\xi})\right]\right\} - \log\left[C(\theta)\right] \\ &- \alpha\frac{\theta}{C(\theta)} \,\mathrm{E}\left[C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})}) H(Y;\,\boldsymbol{\xi})\right] \\ &+ \frac{\theta}{C(\theta)} \,\mathrm{E}\left\{\log\left[C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})})\right] C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})})\right\} \\ &= \int_{\mathbb{R}} f(x;\theta,\alpha,\boldsymbol{\xi}) \log\left[f(x;\theta,\alpha,\boldsymbol{\xi})\right] dx = -\mathbb{H}_{Sh}(f), \end{aligned}$$

where Y is defined as before. So, we obtain $\mathbb{H}_{Sh}(\tau) \le \mathbb{H}_{Sh}(f)$ with equality if and only if $\tau(x;\theta,\alpha,\boldsymbol{\xi}) = f(x;\theta,\alpha,\boldsymbol{\xi})$ for all x, except for a null measure set, thus proving the uniqueness. □

The intermediate steps in the above proof in fact provide the following explicit expression for the Shannon entropy of X

$$\begin{aligned} &\mathbb{H}_{Sh}(f) = -\log(\theta\alpha) - \frac{\theta}{C(\theta)} \,\mathrm{E}\left\{C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})}) \log\left[h(Y;\,\boldsymbol{\xi})\right]\right\} + \log\left[C(\theta)\right] \\ &+\alpha\frac{\theta}{C(\theta)} \,\mathrm{E}\left[C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})}) H(Y;\,\boldsymbol{\xi})\right] - \frac{\theta}{C(\theta)} \,\mathrm{E}\left\{C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})}) \log\left[C'(\theta\, \mathrm{e}^{-\alpha H(Y;\boldsymbol{\xi})})\right]\right\}. \end{aligned}$$

For some EWPS distributions, the above results can only be obtained numerically.

2.6 Special models

In this section, we study three special models of the EWPS class of distributions. We provide plots of the density and hazard rate functions for selected parameter values to illustrate the flexibility of these distributions. We offer some explicit expressions for the moments and moments of the order statistics.

2.6.1 Modified Weibull geometric distribution

The modified Weibull geometric (MWG) distribution is defined by the cdf (2.4) with $H(x;\boldsymbol{\xi}) = x^{\gamma}\exp(\lambda x)$ and $C(\theta) = \theta(1-\theta)^{-1}$ leading to

$$F(x;\theta,\alpha,\gamma,\lambda) = 1 - \frac{(1-\theta)\exp\left(-\alpha x^{\gamma} e^{\lambda x}\right)}{1-\theta\exp\left(-\alpha x^{\gamma} e^{\lambda x}\right)}, \quad x > 0,$$

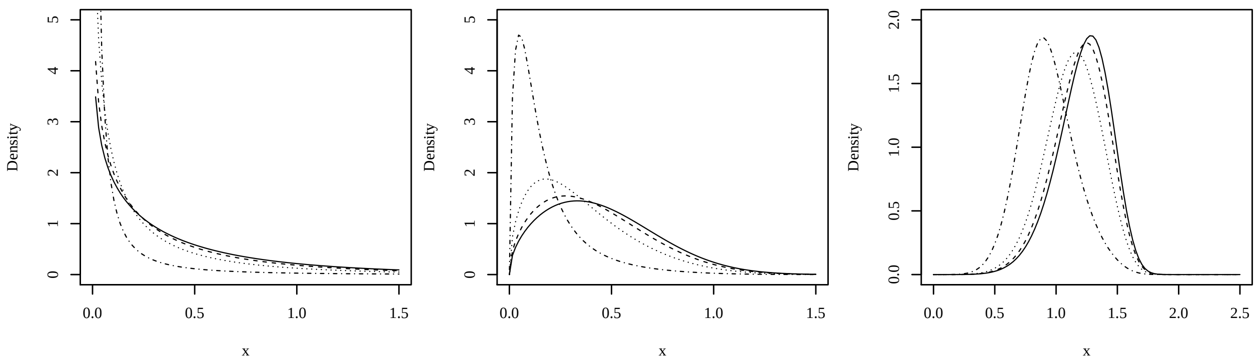
where $\theta \in (0,1)$. The associated pdf and hazard rate function are

$$f(x;\theta,\alpha,\gamma,\lambda) = \alpha(1-\theta)(\gamma+\lambda x)\,x^{\gamma-1}\frac{\exp\left(\lambda x - \alpha x^{\gamma} e^{\lambda x}\right)}{\left[1-\theta\exp\left(-\alpha x^{\gamma} e^{\lambda x}\right)\right]^2}$$

and

$$\tau(x;\theta,\alpha,\gamma,\lambda) = \alpha(\gamma+\lambda x)\,x^{\gamma-1}\frac{\exp(\lambda x)}{1-\theta\exp\left(-\alpha x^{\gamma} e^{\lambda x}\right)},$$

respectively. The MWG distribution includes the WG distribution (Barreto-Souza *et al.*, 2010) when $\lambda = 0$. Further, for $\lambda = 0$ and $\gamma = 1$, we obtain the EG distribution (Adamidis and Loukas, 1998). Figures 2.1 and 2.2 display the density and hazard functions of the MWG distribution for selected parameter values.



(a) $\alpha = 2$, $\gamma = 1.5$ and $\lambda = 0.5$ (b) $\alpha = 2$, $\gamma = 0.8$ and $\lambda = 0.01$ (c) $\alpha = 0.1$, $\gamma = 6$ and $\lambda = 0.5$

Figure 2.1: Plots of the MWG density function for $\theta = 0.01$ (solid line), $\theta = 0.2$ (dashed line), $\theta = 0.5$ (dotted line) and $\theta = 0.9$ (dotdash line).

The rth raw moment of the random variable X having the MWG distribution is determined in closed-form from (2.6) as

$$\mathrm{E}(X^r) = \sum_{n=1}^{\infty} p_n\,\mu_r(n), \tag{2.13}$$

where $\mu_r(n) = \int_0^{\infty} x^r\,g(x;\,n\alpha,\gamma,\lambda)dx$ denotes the rth raw moment of the MW distribution with parameters $n\alpha, \gamma$ and λ. Here, p_n corresponds to the geometric probability function. Carrasco *et al.* (2008) obtained an infinite representation for the rth raw moment of the MW distribution with these parameters given by

$$\mu_r(n) = \sum_{i_1,\ldots,i_r=1}^{\infty} \frac{A_{i_1,\ldots,i_r}\,\Gamma(s_r/\gamma+1)}{(n\alpha)^{s_r/\gamma}}, \tag{2.14}$$

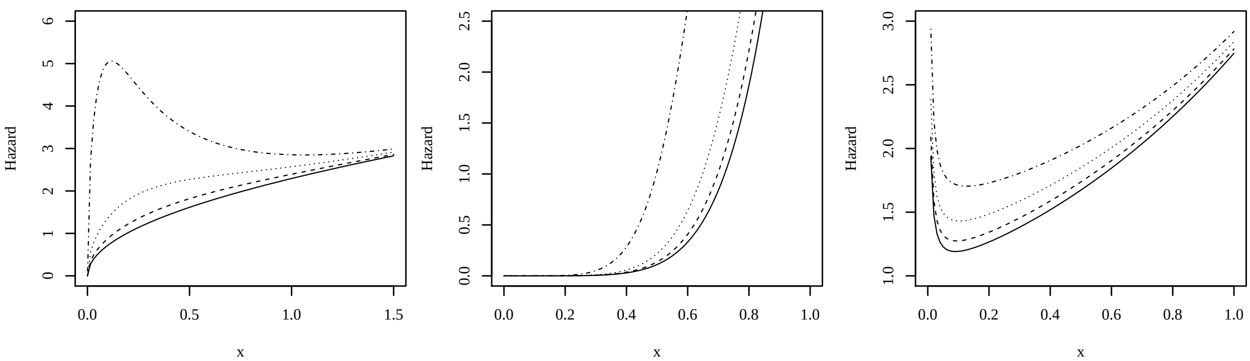
where

$$A_{i_1,\ldots,i_r} = a_{i_1} \times \ldots \times a_{i_r} \quad \text{and} \quad s_r = i_1 + \ldots + i_r,$$

and

$$a_i = \frac{(-1)^{i+1} i^{i-2}}{(i-1)!}\left(\frac{\lambda}{\gamma}\right)^{i-1}.$$

Hence, the ordinary moments of X can be obtained directly from equations (2.13) and (2.14).



(a) $\alpha = 2$, $\gamma = 1.5$ and $\lambda = 0.5$ (b) $\alpha = 2$, $\gamma = 0.8$ and $\lambda = 0.01$ (c) $\alpha = 0.1$, $\gamma = 6$ and $\lambda = 0.5$

Figure 2.2: Plots of the MWG hazard rate function for $\theta = 0.01$ (solid line), $\theta = 0.2$ (dashed line), $\theta = 0.5$ (dotted line) and $\theta = 0.9$ (dotdash line).

The density of the ith order statistic $X_{i:m}$ in a random sample of size m from the MWG distribution is given by (for $i = 1, \ldots, m$)

$$f_{i:m}(x) = \frac{m!}{(i-1)!(m-i)!} \sum_{n=1}^{\infty} \sum_{j=0}^{i-1} \omega_j\, p_n \left[\frac{(1-\theta)\exp\left(-\alpha x^{\gamma} e^{\lambda x}\right)}{1-\theta \exp\left(-\alpha x^{\gamma} e^{\lambda x}\right)}\right]^{m+j-i} g(x; n\alpha, \gamma, \lambda),$$

where $g(x; n\alpha, \gamma, \lambda)$ denotes the MW density function with parameters $n\alpha$, γ and λ. From (2.10), we obtain

$$\mathrm{E}\left(X_{i:m}^{s}\right) = \frac{m!}{(i-1)!(m-i)!} \sum_{n=1}^{\infty} \sum_{j=0}^{i-1} \omega_j\, p_n\, \mathrm{E}\left\{ X^{s} \left[\frac{(1-\theta)\exp\left(-\alpha X^{\gamma} e^{\lambda X}\right)}{1-\theta \exp\left(-\alpha X^{\gamma} e^{\lambda X}\right)}\right]^{m+j-i} \right\}.$$

2.6.2 Pareto Poisson distribution

The Pareto Poisson (PP) distribution is defined by taking $H(x;\boldsymbol{\xi}) = \log(x/k)$, for $x \geq k$, and $C(\theta) = \mathrm{e}^{\theta} - 1$ in (2.4) leading to

$$F(x;\theta,\alpha,k) = 1 - \frac{\exp\left[\theta\,(k/x)^{\alpha}\right] - 1}{\mathrm{e}^{\theta} - 1}, \quad x \geq k.$$

The corresponding pdf and hazard rate function are

$$f(x;\theta,\alpha,k) = \frac{\theta\,\alpha\,k^{\alpha}\exp\left[\theta\,(k/x)^{\alpha}\right]}{(\mathrm{e}^{\theta} - 1)\,x^{\alpha+1}}$$

and

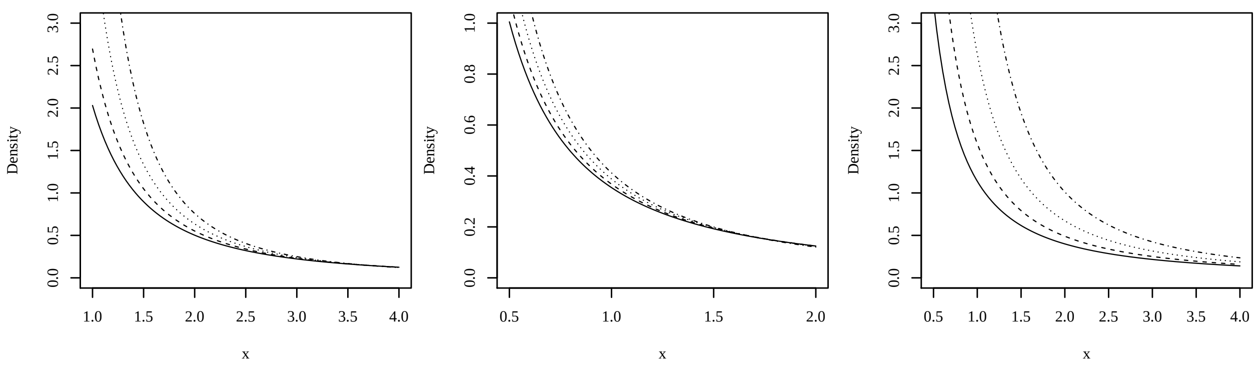
$$\tau(x;\theta,\alpha,k) = \frac{\theta\,\alpha\,k^{\alpha}\exp\left[\theta\,(k/x)^{\alpha}\right]}{x^{\alpha+1}\left\{\exp\left[\theta\,(k/x)^{\alpha}\right] - 1\right\}},$$

respectively. We obtain the Pareto distribution as a sub-model when $\theta \to 0$. The rth moment of the random variable X having the PP distribution becomes

$$\mathrm{E}(X^{r}) = \frac{\alpha k^{r}}{(\mathrm{e}^{\theta} - 1)}\sum_{n=1}^{\infty}\frac{\theta^{n}}{(n-1)!\,(n\alpha - r)}, \quad n\alpha > r. \tag{2.15}$$

In particular, setting $r = 1$ in (2.15), the mean of X reduces to

$$\mu = \frac{\alpha k}{\mathrm{e}^{\theta} - 1}\sum_{n=1}^{\infty}\frac{\theta^{n}}{(n-1)!\,(n\alpha - 1)}, \quad n\alpha > 1.$$

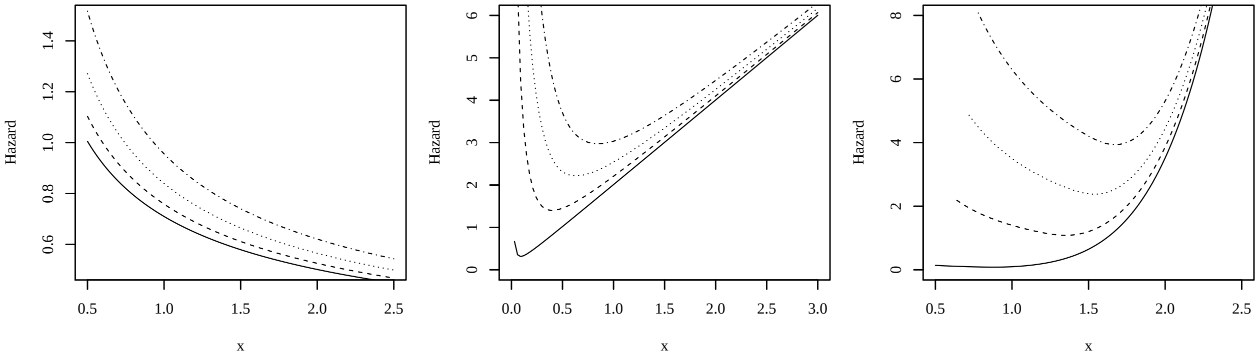


(a) $\alpha = 2$, $\gamma = 1.5$ and $\lambda = 0.5$ (b) $\alpha = 2$, $\gamma = 0.8$ and $\lambda = 0.01$ (c) $\alpha = 0.1$, $\gamma = 6$ and $\lambda = 0.5$

Figure 2.3: Plots of the PP density function for $\theta = 0.01$ (solid line), $\theta = 0.2$ (dashed line), $\theta = 0.5$ (dotted line) and $\theta = 0.9$ (dotdash line).

From equation (2.10), the sth moment of the ith order statistic (for $i = 1, \ldots, m$) is given by

$$\mathrm{E}\left(X_{i:m}^{s}\right) = \frac{m!}{(i-1)!(m-i)!}\sum_{n=1}^{\infty}\sum_{j=0}^{i-1}\omega_j\,p_n\,\mathrm{E}\left[X^{s}\left(\frac{\exp(\theta\,(k/X)^{\alpha}) - 1}{\mathrm{e}^{\theta} - 1}\right)^{m+j-i}\right],$$



(a) $\alpha = k = 0.5$ (b) $\alpha = 2$ and $k = 1$ (c) $\alpha = 7$ and $k = 2$

Figure 2.4: Plots of the PP hazard rate function for $\theta = 0.01$ (solid line), $\theta = 0.2$ (dashed line), $\theta = 0.5$ (dotted line) and $\theta = 0.9$ (dotdash line).

where p_n denotes the Poisson probability function. In addition, after some algebra, the Shannon entropy for the PP distribution reduces to

$$\mathbb{H}_{Sh}(f) = \log\left(\frac{e^{\theta}-1}{\theta\alpha}\right) - \frac{\theta}{e^{\theta}-1}\left(\mu_1 - \alpha\mu_2 + \mu_3\right),$$

where

$$\mu_1 = \mathrm{E}\left[\exp\left\{\theta\left(\frac{k}{X}\right)^{\alpha}\right\}\log\left(\frac{1}{X}\right)\right] = \frac{1}{2(e^{\theta}-1)}\left\{\frac{\mathrm{Chi}(2\theta) - \log(2\theta) + \mathrm{Shi}(2\theta) - \gamma}{\alpha} - (e^{2\theta}-1)\log k\right\},$$

$$\mu_2 = \mathrm{E}\left[\exp\left\{\theta\left(\frac{k}{X}\right)^{\alpha}\right\}\log\left(\frac{X}{k}\right)\right] = \frac{\mathrm{Chi}(2\theta) - \log(2\theta) + \mathrm{Shi}(2\theta) - \gamma}{2\alpha(e^{\theta}-1)}$$

and

$$\mu_3 = \mathrm{E}\left[\theta\exp\left\{\theta\left(\frac{k}{X}\right)^{\alpha}\right\}\left(\frac{k}{X}\right)^{\alpha}\right] = \frac{\alpha\,\theta\,k^{2\alpha}}{4(e^{\theta}-1)}\left\{1 - (2\theta+1)e^{2\theta}\right\},$$

where

$$\mathrm{Chi}(z) = \gamma + \log z + \int_0^z \frac{\cosh(t)-1}{t}dt$$

is the hyperbolic cosine integral,

$$\mathrm{Shi}(z) = \int_0^z \frac{\sinh(t)-1}{t}dt$$

is the hyperbolic sine integral and $\gamma \approx 0.577216$ is the Euler-Mascheroni constant.

2.6.3 Chen logarithmic distribution

The Chen logarithmic (CL) distribution is defined by the cdf (2.4) with $H(x;\boldsymbol{\xi}) = \exp(x^{\beta}) - 1$ and $C(\theta) = -\log(1-\theta)$ leading to

$$F(x) = 1 - \frac{\log\left\{1 - \theta\exp\left[-\alpha(\exp(x^{\beta})-1)\right]\right\}}{\log(1-\theta)}, \quad x > 0,$$

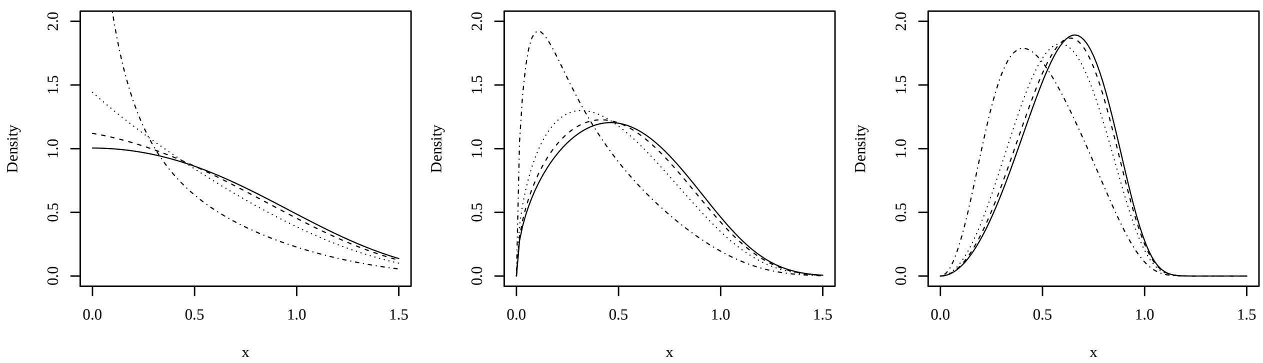
where $\theta \in (0,1)$. The associated pdf and hazard rate function are

$$f(x) = \frac{\theta\,\alpha\,\beta\,x^{\beta-1}\exp\left\{x^{\beta} - \alpha\left[\exp(x^{\beta}) - 1\right]\right\}}{\log(1-\theta)\left\{\theta\exp\left[-\alpha(\exp(x^{\beta}) - 1)\right] - 1\right\}}$$

and

$$\tau(x) = \frac{\theta\,\alpha\,\beta\,x^{\beta-1}\exp\left[x^{\beta} - \alpha(\exp(x^{\beta}) - 1)\right]}{\left\{\theta\exp\left[-\alpha(\exp(x^{\beta}) - 1)\right] - 1\right\}\log\left\{1 - \theta\exp\left[-\alpha(\exp(x^{\beta}) - 1)\right]\right\}},$$

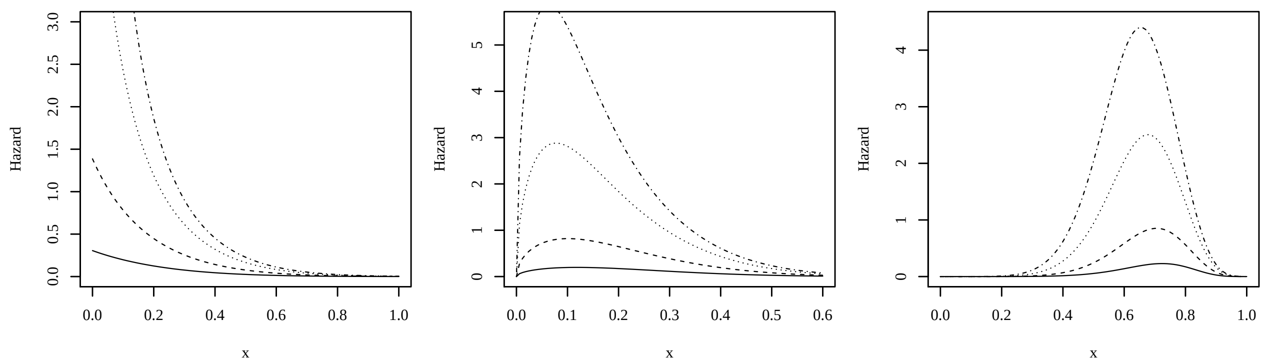
respectively.



(a) $\alpha = b = 1$ (b) $\alpha = b = 1.5$ (c) $\alpha = 2.5$ and $b = 3$

Figure 2.5: Plots of the CL density function for $\theta = 0.01$ (solid line), $\theta = 0.2$ (dashed line), $\theta = 0.5$ (dotted line) and $\theta = 0.9$ (dotdash line).

Proposition 1 implies that the Chen distribution is a limiting special case when $\theta \to 0^{+}$.



(a) $\alpha = 2$ and $b = 1$ (b) $\alpha = 3.5$ and $b = 7$ (c) $\alpha = 3.5$ and $b = 1.5$

Figure 2.6: Plots of the CL hazard rate function for $\theta = 0.01$ (solid line), $\theta = 0.2$ (dashed line), $\theta = 0.5$ (dotted line) and $\theta = 0.9$ (dotdash line).

The density of the ith order statistic $X_{i:m}$ in a random sample of size m from the CL distribution is given by (for $i = 1, \ldots, m$)

$$f_{i:m}(x) = \frac{m!}{(i-1)!(m-i)!} \sum_{n=1}^{\infty} \sum_{j=0}^{i-1} \omega_j^* \, p_n \, g(x; n\alpha, b) \left\{ \log \left[1 - \theta \exp(\alpha - \alpha \, \mathrm{e}^{x^b}) \right] \right\}^{m+j-1},$$

where $g(x; n\alpha, b)$ denotes the Chen density function with parameters $n\alpha$ and b and p_n denotes the logarithmic probability function and

$$\omega_j^* = (-1)^j \binom{i-1}{j} \left[\frac{1}{\log(1-\theta)} \right]^{m+j-1}.$$

In a similar manner, the sth raw moment of $X_{i:m}$ is obtained directly from

$$\mathrm{E}\,(X_{i:m}^s) = \frac{m!}{(i-1)!(m-i)!} \sum_{n=1}^{\infty} \sum_{j=0}^{i-1} \omega_j \, p_n \; \mathrm{E}\left\{ Z^s \exp\left[n\alpha(m+j-1)(1-\exp(Z^b)) \right] \right\},$$

where $Z \sim \text{Chen}(n\alpha, b)$.

2.7 Applications

In this section, we compare the results of the fitted special models of the EWPS class by means of two real data sets for illustrative purposes. In order to estimate the parameters of these special models, we adopt the maximum likelihood method (as discussed in Section 4) and all the computations were done using the subroutine `NLMixed` of the `SAS` software. A good alternative is to use the software `R` for which Nadarajah et al. (2012) introduced the package `Compounding` for dealing with continuous distributions obtained by compounding continuous distributions with discrete distributions. They demonstrated its use by computing values of the cumulative and density functions, quantile and hazard rate functions, generating random samples from a population with compounding distribution, and computing mean and variance of a random variable with a compounding distribution.

First, we consider a data set from Fonseca and França (2007), who studied the soil fertility influence and the characterization of the biologic fixation of N_2 for the *Dimorphandra wilsonii rizz growth*. For 128 plants, they made measures of the phosphorus concentration in the leaves. The data are listed in Table 2.3. We fit the Gompertz Poisson (GP), Chen Poisson (CP) and CL models to these data. We also fit the three-parameter sub-model WG (Barreto-Souza *et al.*, 2010).

Tables 2.4 and 2.5 display some descriptive statistics and the MLEs (with corresponding standard errors in parentheses), the maximized log-likelihood and the Kolmogorov-Smirnov statistic for the fitted models. Since the values of the Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC) are smaller for the CL distribution compared with those values of the other models, this new distribution seems to be a very competitive model for these data.

Plots of the pdf and cdf of the fitted WG, GP, CP and CL models to these data are displayed in Figure 2.7. They indicate that the CL distribution is better than the other distributions in

0.22	0.17	0.11	0.10	0.15	0.06	0.05	0.07	0.12	0.09	0.23	0.25	0.23
0.24	0.20	0.08	0.11	0.12	0.10	0.06	0.20	0.17	0.20	0.11	0.16	0.09
0.10	0.12	0.12	0.10	0.09	0.17	0.19	0.21	0.18	0.26	0.19	0.17	0.18
0.20	0.24	0.19	0.21	0.22	0.17	0.08	0.08	0.06	0.09	0.22	0.23	0.22
0.19	0.27	0.16	0.28	0.11	0.10	0.20	0.12	0.15	0.08	0.12	0.09	0.14
0.07	0.09	0.05	0.06	0.11	0.16	0.20	0.25	0.16	0.13	0.11	0.11	0.11
0.08	0.22	0.11	0.13	0.12	0.15	0.12	0.11	0.11	0.15	0.10	0.15	0.17
0.14	0.12	0.18	0.14	0.18	0.13	0.12	0.14	0.09	0.10	0.13	0.09	0.11
0.11	0.14	0.07	0.07	0.19	0.17	0.18	0.16	0.19	0.15	0.07	0.09	0.17
0.10	0.08	0.15	0.21	0.16	0.08	0.10	0.06	0.08	0.12	0.13		

Table 2.3: Phosphorus concentration in leaves data.

Min.	Q_1	Q_2	Mean	Q_3	Max.	Var.
0.0500	0.1000	0.1300	0.1408	0.1800	0.2800	0.0030

Table 2.4: Descriptive statistics.

Model	$\widehat{\theta}$	$\widehat{\alpha}$	$\widehat{\gamma}$	AIC	BIC	CAIC	K–S	$-2\ell(\widehat{\Theta})$
WG	0.9995	2.4471	4.2041	−378.5	−370.0	−378.3	0.0873	−384.5
	(0.0017)	(8.7059)	(0.3022)					
	$\widehat{\theta}$	$\widehat{\alpha}$	$\widehat{\beta}$					
GP	2.9478	0.3169	19.7047	−368.7	−360.2	−368.5	0.1201	−374.7
	(1.2627)	(0.1473)	(1.6135)					
	$\widehat{\theta}$	$\widehat{\alpha}$	$\widehat{b}$					
CP	15.4386	14.7817	2.9212	−383.7	−375.2	−383.5	0.1159	−389.7
	(22.8318)	(28.1576)	(0.2634)					
CL	0.9999	52232	7.5882	−395.8	−387.2	−395.6	0.0678	−401.8
	(0.0001)	(0.0000)	(0.2039)					

Table 2.5: MLEs of the parameters with corresponding SE's (given in parentheses) and maximized log-likelihoods of the WG, GP, CP and CL models for the first data set. The statistics AIC, BIC and CAIC are also displayed.

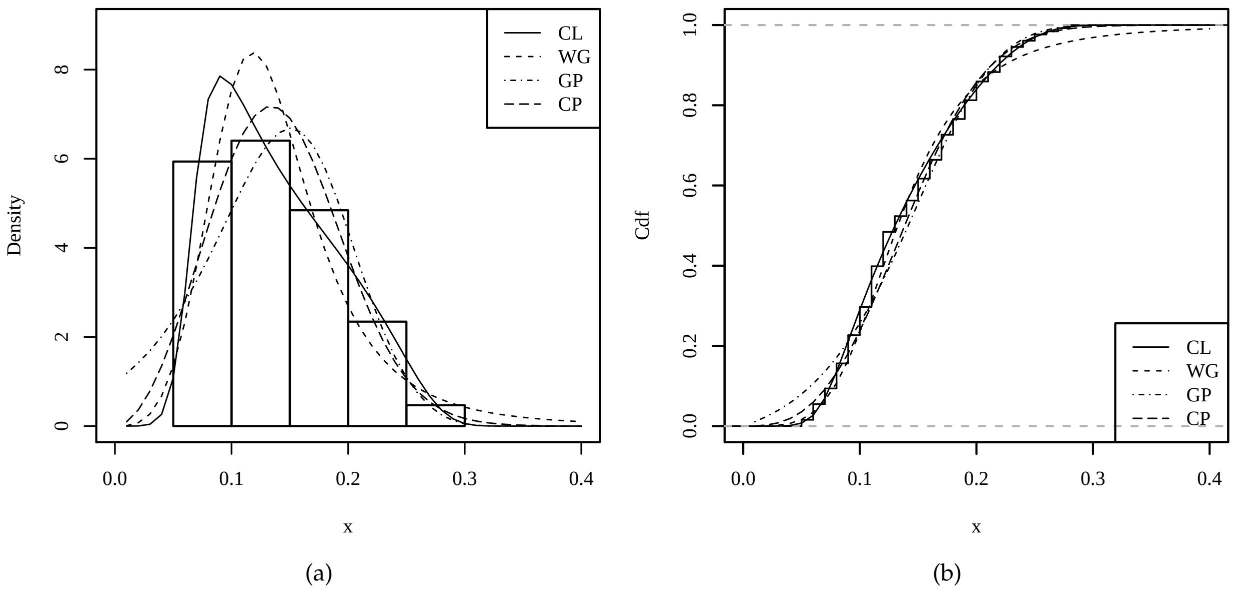


Figure 2.7: Estimated (a) pdf and (b) cdf for the CL, CP, WG and GP models to the percentage of Phosphorus concentration in leaves data.

terms of model fitting. Based on these plots, we conclude that the CL distribution provides a better fit to these data than the WG, GP and CP models.

As a second application, we consider the data consisting of the failure times of 20 mechanical components given in Murthy *et al.* (2004) and listed in Table 2.6. Obviously, due to the genesis of the EW family, the failure times are ideally modeled by this distribution. Thus, the use of the EWPS class for fitting these data is justified.

0.067	0.068	0.076	0.081	0.084	0.085	0.085	0.086	0.089	0.098
0.098	0.114	0.114	0.115	0.121	0.125	0.131	0.149	0.160	0.485

Table 2.6: The failure times of 20 mechanical components.

Tables 2.7 display some descriptive statistics. The MLEs of the parameters (standard errors between parentheses), the Kolmogorov-Smirnov statistic, $-2\ell(\widehat{\Theta})$ and the values of the AIC, BIC and CAIC statistics are listed in Table 2.8. The values of these statistics indicate that the WG model yields a better fit to these data than the GP and CP models.

Min.	Q_1	Q_2	Mean	Q_3	Max.	Var.
0.0670	0.0848	0.0980	0.1216	0.1220	0.4850	0.0080

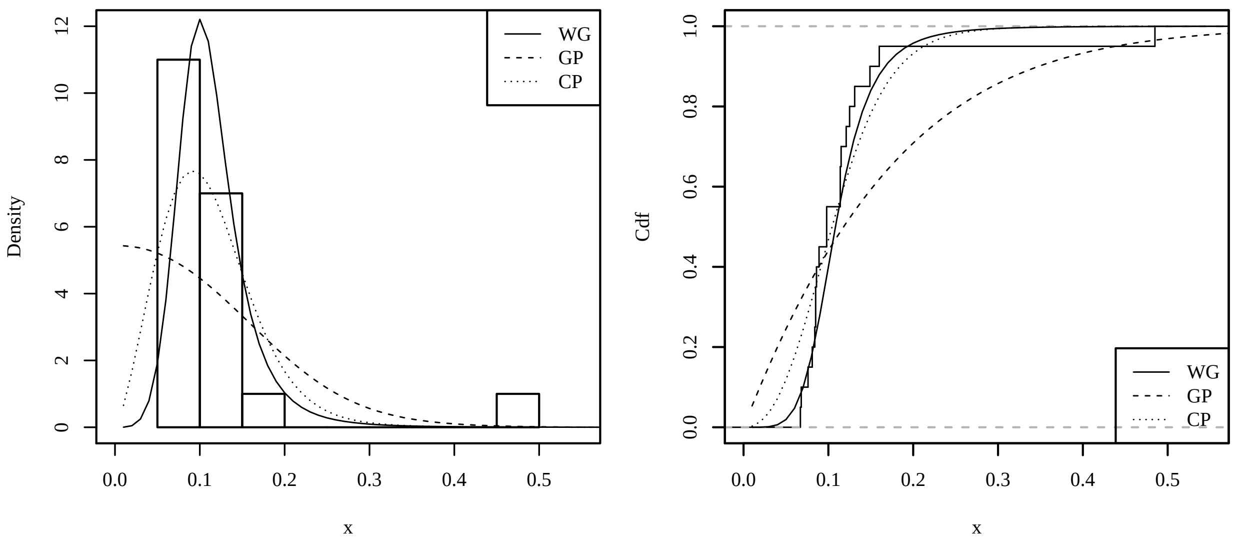
Table 2.7: Descriptive statistics.

Plots of the estimated pdf and cdf of the fitted WG, GP and CP models to these data are displayed in Figure 2.8. They indicate that the WG distribution is better then the other distribu-

Model	$\widehat{\theta}$	$\widehat{\alpha}$	$\widehat{\gamma}$	AIC	BIC	CAIC	K–S	$-2\ell(\widehat{\Theta})$
WG	0.9999	8.1443	5.0876	−66.4	−63.4	−64.9	0.1810	−72.4
	(0.0001)	(0.0137)	(0.8002)					
	$\widehat{\theta}$	$\widehat{\alpha}$	$\widehat{\beta}$					
GP	5.4566	0.9909	6.5683	−41.9	−38.9	−40.4	0.3312	−47.9
	(2.4140)	(0.5504)	(2.2144)					
	$\widehat{\theta}$	$\widehat{\alpha}$	$\widehat{b}$					
CP	6.2426	25.3554	2.3796	−54.7	−51.7	−53.2	0.2214	−60.7
	(2.2755)	(15.8907)	(0.3380)					

Table 2.8: MLEs of the parameters with corresponding SE's (given in parentheses) and maximized log-likelihoods of the WG, GP and CP models for the second data set. The statistics AIC, BIC and CAIC are also displayed.

tions in terms of model fitting. From these figures, we conclude that this distribution provides a better fit to these data than the GP and CP models.



(a) (b)

Figure 2.8: Estimated (a) pdf and (b) cdf for the WG, GP and CP models to the failure times.

2.8 Concluding remarks

We define a new class of lifetime distributions called the extended Weibull power series (EWPS) class, which generalizes the Weibull power series class of distributions (Morais and Barreto-Souza, 2011). Further, the new class extends the exponential power series distributions (Chahkandi and Ganjali, 2009). We provide a mathematical treatment of the new class including expansions for the density function, moments, generating function, incomplete moments and reliability. Further, explicit expressions for the order statistics and Shannon entropy are derived. The EWPS density function can be expressed as a mixture of extended Weibull (EW) density functions. This mixture representation is important to derive several properties of the new class. Maximum likelihood inference is implemented straightforwardly for estimating the model parameters. We obtain the observed information matrix. Maximum entropy identification was discussed and some special models are explored. We fit some EWPS distributions to two real data sets to show the usefulness of the proposed class. In conclusion: we define a general approach for generating new lifetime distributions, at least 68 distributions, some of them known and the great majority new ones. Further, we motivate the use of the new class in four different ways. We think these two facts combined may attract more complex applications in the literature of lifetime distributions. Finally, the formulae derived are manageable by using modern computer resources with analytic and numerical capabilities.

Appendix A

The elements of the $p \times p$ information matrix $J_n(\Theta)$ are

$$J_{\theta\theta} = -\frac{n}{\theta^2} - n\left[\frac{C''(\theta)}{C(\theta)} - \left(\frac{C'(\theta)}{C(\theta)}\right)^2\right] + \theta\sum_{i=1}^{n}\left(\frac{z_{2i}}{z_{1i}}\right)^2 H(x_i;\,\xi)\mathrm{e}^{-2\alpha H(x_i;\xi)}$$
$$-\,\theta\sum_{i=1}^{n}\frac{z_{3i}}{z_{1i}}H(x_i;\,\boldsymbol{\xi})\mathrm{e}^{-2\alpha H(x_i;\boldsymbol{\xi})}$$

$$J_{\alpha\alpha} = -\frac{n}{\alpha^2} + \theta\sum_{i=1}^{n}\frac{z_{2i}}{z_{1i}}H^2(x_i;\boldsymbol{\xi})\mathrm{e}^{-\alpha H(x_i;\xi)} + \theta^2\sum_{i=1}^{n}\frac{(z_{3i}-z_{2i}^2)}{z_{1i}}H^2(x_i;\xi)\mathrm{e}^{-2\alpha H(x_i;\xi)}$$

$$J_{\alpha\theta} = \theta\sum_{i=1}^{n}\left[\left(\frac{z_{2i}}{z_{1i}}\right)^2 - \frac{z_{3i}}{z_{1i}}\right]H^2(x_i;\xi)\mathrm{e}^{-2\alpha H(x_i;\xi)} - \sum_{i=1}^{n}\frac{z_{2i}}{z_{1i}}H^2(x_i;\xi)\mathrm{e}^{-\alpha H(x_i;\xi)}$$

$$J_{\alpha\boldsymbol{\xi}_k} = -\sum_{i=1}^{n}\frac{\partial H(x_i;\,\boldsymbol{\xi})}{\partial\xi_k} - \theta\sum_{i=1}^{n}\frac{z_{2i}}{z_{1i}}\frac{\partial H(x_i;\,\xi)}{\partial\xi_k}\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})}\left[1-\alpha H(x_i;\,\boldsymbol{\xi})\right]$$
$$+\,\alpha\theta^2\sum_{i=1}^{n}\left[\frac{z_{3i}}{z_{1i}} - \left(\frac{z_{2i}}{z_{1i}}\right)^2\right]\frac{\partial H(x_i;\,\xi)}{\partial\boldsymbol{\xi}_k}H(x_i;\,\xi)\mathrm{e}^{-2\alpha H(x_i;\boldsymbol{\xi})}$$

$$J_{\theta\boldsymbol{\xi}_k} = \theta\alpha\sum_{i=1}^{n}\left[\left(\frac{z_{2i}}{z_{1i}}\right)^2 - \frac{z_{3i}}{z_{1i}}\right]\frac{\partial H(x_i;\,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_k}\mathrm{e}^{-2\alpha H(x_i;\xi)} - \alpha\sum_{i=1}^{n}\frac{z_{2i}}{z_{1i}}\frac{\partial H(x_i;\,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_k}\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})}$$

$$J_{\xi_k\boldsymbol{\xi}_l} = -\alpha\sum_{i=1}^{n}\frac{\partial^2 H(x_i;\,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_k\partial\boldsymbol{\xi}_l} - \sum_{i=1}^{n}\frac{1}{H(x_i;\,\boldsymbol{\xi})^2}\frac{\partial H(x_i;\,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_k}\frac{\partial H(x_i;\,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_l} + \sum_{i=1}^{n}\frac{1}{H(x_i;\,\boldsymbol{\xi})}\frac{\partial^2 H(x_i;\,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_k\partial\boldsymbol{\xi}_l}$$
$$+\,(\alpha\theta)^2\sum_{i=1}^{n}\left[\left(\frac{z_{2i}}{z_{1i}}\right)^2 + \frac{z_{3i}}{z_{1i}}\right]\frac{\partial H(x_i;\,\boldsymbol{\xi})}{\partial\xi_k}\frac{\partial H(x_i;\,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_l}\mathrm{e}^{-2\alpha H(x_i;\boldsymbol{\xi})}$$
$$-\,\alpha\theta\sum_{i=1}^{n}\frac{z_{2i}}{z_{1i}}\frac{\partial^2 H(x_i;\,\xi)}{\partial\boldsymbol{\xi}_k\partial\boldsymbol{\xi}_l}\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})} + \alpha^2\theta\sum_{i=1}^{n}\frac{z_{2i}}{z_{1i}}\frac{\partial H(x_i;\xi)}{\partial\boldsymbol{\xi}_k}\frac{\partial H(x_i;\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_l}\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})}$$

where $z_{1i} = C'(\theta e^{-\alpha H(x_i;\boldsymbol{\xi})})$, $z_{2i} = C''(\theta\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})})$ and $z_{3i} = C'''(\theta\mathrm{e}^{-\alpha H(x_i;\boldsymbol{\xi})})$, for $i = 1,\ldots,n$.

References

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

[2] Barreto-Souza, W., Cribari-Neto, F. (2009). A generalization of the exponential-Poisson distribution. *Statistics and Probability Letters*, **79**, 2493–2500.

[3] Barreto-Souza, W., Morais, A.L., Cordeiro, G.M. (2010). The Weibull-geometric distribution. *Journal of Statistical Computation and Simulation*, **81**, 645–657.

[4] Barreto-Souza, W., Bakouch, H.S. (2012). A new lifetime model with decreasing failure rate. *Statistics*. doi:`10.1080/02331888.2011.595489`.

[5] Bebbington, M., Lai, C.D. and Zitikis, R. (2007). A flexible Weibull extension. *Reliability Engineering and System Safety*, **92**, 719–726.

[6] Cancho, V.G., Louzada, F., Barriga. G.D.C. (2011). The Poisson-exponential lifetime distribution. *Computational Statistics and Data Analysis*, **55**, 677–686.

[7] Cancho, V.G., Louzada, F., Barriga. G.D.C. (2012). The Geometric Birnbaum-Saunders regression model with cure rate. *Journal of Statistical Planning and Inference*, **142**, 993–1000.

[8] Carrasco J.M.F., Ortega, E.M.M., Cordeiro, G.M. (2008). A generalized modified Weibull distribution for lifetime modeling. *Computational Statistics and Data Analysis*, **53**, 450–462.

[9] Chahkandi, M., Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis*, **53**, 4433–4440.

[10] Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters*, **49**, 155–161.

[11] Cooner, F., Banerjee, S., Carlin, B.P., Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*. **102**, 560–572.

[12] Cordeiro, G.M., Rodrigues, J., Castro, M. (2012). The exponential COM-Poisson distribution. *Statistical Papers*. doi:10.1007/s00362-011-0370-9.

[13] Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory*. John Wiley and Sons, New York.

[14] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

[15] Doornik, J. (2007) Ox 5: object-oriented matrix programming language, 5th ed. Timberlake Consultants, London.

[16] Fonseca, M.B. and França, M.G.C. (2007). A influência da fertilidade do solo e caracterização da fixação biológica de N_2 para o crescimento de *Dimorphandra wilsonii rizz. Master's thesis, Universidade Federal de Minas Gerais*.

[17] Garvan, F. (2002). The Maple Book. Chapman and Hall/CRC, London.

[18] Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality and on the new model of determining the value of life contingencies. Philosophical Trans. *Royal Society of London*, **115**, 513–585.

[19] Guess F. and Proschan F. (1985). Mean residual life: theory and applications. In P.R. Krishnaiah and C.R. Rao (eds.) *Handbook of Statistics*. Reliability and Qualiy Control, 7, 215–224.

[20] Gurvich, M., DiBenedetto, A., Ranade, S. (1997). A new statistical distribution for characterizing the random strength of brittle materials. *Journal of Materials Science*, **32**, 2559–2564.

[21] Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Reviews*, **106**, 620–630.

[22] Johnson, N.L, Kotz, S. and Balakrishnan, N. (1994). Continuous Univariate Distributions volume 1. John Wiley and Sons, New York.

[23] Kapur, J.N. (1989). Maximum Entropy Models in Science and Engineering. John Wiley and Sons, New York.

[24] Kies, J.A. (1958). *The strength of glass*. Washington D.C. Naval Research Lab, 5093.

[25] Kus, C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis*, **51**, 4497–4509.

[26] Lai, C.D., Xie, M., Murthy, D.N.P. (2003). A modified weibull distribution. *Transactions on Reliability*, **52**, 33–37.

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

[28] Lu, W., Shi, D. (2011). A new compounding life distribution: the Weibull-Poisson distribution. *Journal of Applied Statistics*, DOI:10.1080/02664763.2011.575126.

[29] Mahmoudi, E., Jafari, A.A. (2012). Generalized exponential?power series distributions. *Computational Statistics and Data Analysis*, **56**, 4047–4066.

[30] Morais, A.L., Barreto-Souza, W. (2011). A Compound Class of Weibull and Power Series Distributions. *Computational Statistics and Data Analysis*, **55**, 1410–1425.

[31] Murthy, D.N.P., Xie, M., Jiang, R. (2004). *Weibull models*. Vol. 1.

[32] Nadarajah, S. and Kotz, S. (2005). On some recent modifications of Weibull distribution. *IEEE Trans. Reliability*, **54**, 561–562.

[33] Nadarajah, S., Popović, B.V. Ristić, M.M. (2012). Compounding: an R package for computing continuous distributions obtained by compounding a continuous and a discrete distribution. *Computational Statistics*. doi:10.1007/s00180-012-0336-y.

[34] Nikulin, M., Haghighi, F. (2006). A Chi-squared test for the generalized power Weibull family for the head-and-neck cancer censored data. *Journal of Mathematical Sciences*, **133**, 1333–1341.

[35] Pham, H. (2002). A vtub-shaped hazard rate function with applications to sys- tem safety. *International Journal of Reliability and Applications*, **3**, 1–16.

[36] Pham, H., Lai, C.D. (2007). On recent generalizations of the Weibull Distribution. *IEEE Transactions on Reliabilty*, **56**, 454–458.

[37] Phani, K.K. (1987). A new modified Weibull distribution function. *Communications of the American Ceramic Society*, **70**, 182–184.

[38] R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. Vienna, Austria.

[39] Rayleigh, J.W.S. (1880). On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philosophical Magazine*, **10**, 73–78.

[40] Shannon, C.E., (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–432.

[41] Sigmon, K., Davis, T.A., 2002. MATLAB Primer, 6th ed. Chapman and Hall/CRC, London.

[42] Shore, J.E., Johnson, R.W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, **26**, 26–37.

[43] Smith, R.M. and Bain, L.J. (1975). An exponential power life-testing distribution. *Communications Statistics*, **4**, 469–481.

[44] Soofi, E.S. (2000). Principal information theoretic approaches. *Journal of the American Statistical Association*, **95**, 1349–1353.

[45] Tahmasbi, R., Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics and Data Analysis*, **52**, 3889–3901.

[46] Xie, M., Lai, C.D. (1995). Reliability analysis using additive Weibull model with bathtub-shaped failure rate function. *Reliability Engineering and System Safety*, **52**, 87–93.

[47] Xie, M., Tang, Y. and Goh, T.N. (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering and System Safety*, **76**, 279–285.

[48] Zografos, K., Balakrishnan, N. (2009). On families of beta-and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, **6**, 344–362.

[49] White, J.S. (1969). The moments of log-Weibull order statistics. *Technometrics*, **11**, 373–386.

[50] Wolfram, S. (2003). The Mathematica Book, 5th ed. Cambridge University Press.

CHAPTER 3

A New Wider Family of Continuous Models: The Extended Cordeiro and de Castro Family

Artigo atualmente submetido para publicação.

Resumo

Introduzimos e estudamos propriedades matemáticas gerais de um novo gerador de distribuições contínuas com três parâmetros extras chamado família Cordeiro e Castro estendida de distribuições. Investigamos as assíntotas e formas das funções de densidade e de risco. A nova função densidade pode ser expressa como uma combinação linear das densidades de origem exponencializadas. Obtemos uma série de potências para a função quantílica. Expressões explícitas para os momentos ordinários e incompletos, funções quantílica e geradora, a distribuição assintótica dos valores extremos, entropias de Shannon e Rényi e estatísticas de ordem, que valem para qualquer modelo de origem, são determinados. Discutimos a estimativa dos parâmetros do modelo por máxima verossimilhança e ilustramos a potencialidade da nova família por meio de duas aplicações a dados reais.

Palavras-chave: Entropia de Rényi; Estatística de ordem; Família generalizada; Função quantílica; distribuição exponencial geométrica generalizada; Máxima verossimilhança ; Momentos.

Abstract

We introduce and study general mathematical properties of a new generator of continuous distributions with three extra parameters called the *extended Cordeiro and de Castro* family. We investigate the asymptotes and shapes of the density and hazard rate functions. The new density function can be expressed as a linear combination of exponentiated densities based on the same baseline distribution. We derive a power series for the quantile function of this family. Explicit expressions for the ordinary and incomplete moments, quantile and generating functions, asymptotic distribution of the extreme values, Shannon and Rényi entropies and order statistics, which hold for any baseline model, are determined. We discuss the estimation of the model parameters by maximum likelihood and illustrate the potentiality of the family by means of two applications to real data.

Keywords: Generalized exponential geometric distribution; Generated family; Maximum likelihood; Moment; Order statistic; Quantile function; Rényi entropy.

3.1 Introduction

In the past few years, several ways of generating new distributions from classic ones were developed and discussed. Jones (2004) studied a family of distributions that arises naturally from the distribution of the order statistics. The beta-generated family proposed by Eugene *et al*. (2002) was disscussed in Zografos and Balakrishnan (2009), who introduced the gamma-generated family of distributions. Based on a baseline continuous distribution $G(x)$ with survival function $\overline{G}(x)$ and probability density function (pdf) $g(x)$, Zografos and Balakrishnan (2009) defined the cumulative distribution function (cdf) and the pdf of the gamma-generator (for $x \in \mathbb{R}$) by

$$F(x) = \frac{1}{\Gamma(\delta)} \int_0^{-\log\left[\overline{G}(x;\boldsymbol{\xi})\right]} t^{\delta-1}\, \mathrm{e}^{-t} dt \tag{3.1}$$

and

$$f(x) = \frac{1}{\Gamma(\delta)} \left\{-\log\left[\overline{G}(x;\boldsymbol{\xi})\right]\right\}^{\delta-1} g(x;\boldsymbol{\xi}), \tag{3.2}$$

respectively, where $\Gamma(\cdot)$ is the gamma function. Ristíc and Balakrishnan (2011) proposed an alternative gamma-generator defined by the cdf and pdf (for $x \in \mathbb{R}$) given by

$$F(x) = 1 - \frac{1}{\Gamma(\delta)} \int_0^{-\log[G(x;\boldsymbol{\xi})]} t^{\delta-1}\mathrm{e}^{-t} dt, \quad \delta > 0 \tag{3.3}$$

and

$$f(x) = \frac{1}{\Gamma(\delta)} \left\{-\log\left[G(x;\boldsymbol{\xi})\right]\right\}^{\delta-1} g(x;\boldsymbol{\xi}), \tag{3.4}$$

respectively.

Based on this paper, by replacing the gamma distribution by the more flexible three parameter generalized exponential-geometric (GEG) distribution (Silva *et al*., 2010), we propose a new wider family of distributions given by

$$F(x) = 1 - \int_0^{-\log[G(x;\boldsymbol{\xi})]} \frac{\alpha\lambda(1-p)\mathrm{e}^{-\lambda t}[1-\mathrm{e}^{-\lambda t}]^{\alpha-1}}{[1-p\,\mathrm{e}^{-\lambda t}]^{\alpha+1}} \mathrm{d}t = 1 - \left[\frac{1-G(x;\boldsymbol{\xi})^{\lambda}}{1-p\,G(x;\boldsymbol{\xi})^{\lambda}}\right]^{\alpha}, \tag{3.5}$$

where $G(x;\boldsymbol{\xi})$ is the baseline cdf depending on a parameter vector $\boldsymbol{\xi}$ and $\alpha > 0$, $\lambda > 0$ and $p \in (0,1)$ are three additional parameters. For each baseline G, the *extended Cordeiro and de Castro-G* ("ECC-G" for short) family of distributions is defined by the cdf (3.5). Equation (3.5) is a wider family of continuous distributions. It includes the *generalized Kumaraswamy class* (Cordeiro and de Castro, 2011) of distributions, the proportional and reversed hazard rate models, Marshal-Olkin family of distributions and other sub-families. Some special models are given in Table 3.1.

Table 3.1: Some special models.

λ	α	p	$G(x)$	Reduced distribution
-	-	0	-	Generalized Kumaraswamy distribution (Cordeiro and de Castro, 2011)
1	1	0	-	$G(x)$
-	1	0	-	Reversed hazard rate model (Gupta and Gupta, 2007)
1	-	0	-	Proportional hazard rate model (Gupta and Gupta, 2007)
-	1	p	-	Marshall-Olkin family of distributions (Marshall-Olkin, 1997)
-	-	0	Generalized Rayleigh	Kumaraswamy generalized Rayleigh distribution (Gomes *et al.*, 2012)
-	-	0	Burr XII distribution	Kumaraswamy Burr XII distribution (Paranaíba *et al.*, 2012)
-	-	0	Modified Weibull distribution	Kumaraswamy modified Weibull distribution (Cordeiro *et al.*, 2012)
-	-	0	Pareto distribution	Kumaraswamy Pareto distribution (Bourguignon *et al.*, 2012)

This chapter is organized as follows. In Section 2, we provide a physical interpretation of the ECC-G family. Four special cases of this family are defined in Section 3. Some useful expansions are derived in Section 4. In Section 5, we propose explicit expressions for the moments and generating function using a power series expansion for the quantile function. Further, we present general expressions for the Rényi and Shannon entropies and mean deviations are addressed. Estimation of the model parameters by maximum likelihood is performed in Section 6. Applications to two real data sets illustrate the performance of the new family in Section 7. The chapter is concluded in Section 8.

3.2 The new family

The corresponding density function to (3.5) is given by

$$f(x;\alpha,\lambda,p,\boldsymbol{\xi}) = \alpha\,\lambda\,(1-p)\,g(x;\boldsymbol{\xi})\,G(x;\boldsymbol{\xi})^{\lambda-1}\,\frac{[1-G(x;\boldsymbol{\xi})^{\lambda}]^{\alpha-1}}{[1-p\,G(x;\boldsymbol{\xi})^{\lambda}]^{\alpha+1}}, \tag{3.6}$$

where $g(x;\boldsymbol{\xi})$ is the baseline pdf. Equation (3.6) will be most tractable when the cdf $G(x)$ and the pdf $g(x)$ have simple analytic expressions. Hereafter, a random variable X with density function (3.6) is denoted by $X \sim$ ECC-G$(p,\alpha,\lambda,\boldsymbol{\xi})$. Further, we can omit sometimes the dependence on the vector $\boldsymbol{\xi}$ of the parameters and write simply $G(x) = G(x;\boldsymbol{\xi})$.

When $p > 0$, we consider a system formed by α independent components following the

Marsha-Olkin cdf (see Table 3.1) given by

$$H(x) = \frac{(1-p)\,G(x)^{\lambda}}{1-p\,G(x)^{\lambda}}.$$

Suppose the system fails if any of the α components fails and let X denote the lifetime of the entire system. Then, the cdf of X is

$$F(x) = 1 - [1-H(x)]^{\alpha} = 1 - \left[\frac{1-G(x)^{\lambda}}{1-p\,G(x)^{\lambda}}\right]^{\alpha},$$

which is the proposed generator.

When $p = 0$, a physical interpretation of the ECC-G distribution can be given as follows. Consider a system formed by α independent components and that each component is made up of λ independent subcomponents. Suppose that the system fails if any of the α components fails and that each component fails if all of the λ subcomponents fail. Let $X_{j1}, \ldots, X_{j\lambda}$ denote the lifetimes of the subcomponents within the jth component, $j = 1, \ldots, \alpha$, having a common cdf G. Let X_j denote de lifetime of the jth component, for $j = 1, \ldots, \alpha$, and let X denote the lifetime of the entire system. Then, the cdf of X is

$$\begin{aligned} P(X \le x) &= 1 - P(X_1 > x, \ldots, X_{\alpha} > x) = 1 - P(X_1 > x)^{\alpha} \\ &= 1 - [1 - P(X_1 \le x)]^{\alpha} = 1 - [1 - P(X_{11} \le x, \ldots, X_{1\lambda})]^{\alpha} \\ &= 1 - \left[1 - P(X_{11} \le x)^{\lambda}\right]^{\alpha} = 1 - \left[1 - G(x)^{\lambda}\right]^{\alpha}. \end{aligned}$$

Thus, the family of distributions (3.6) with $p = 0$ is precisely the time to failure of the entire system.

The hazard rate function (hrf) of X becomes

$$h(x;\alpha,\lambda,p,\boldsymbol{\xi}) = \alpha\,\lambda\,(1-p)\,g(x;\boldsymbol{\xi})\,G(x;\boldsymbol{\xi})^{\lambda-1}\,\left[\frac{1-p\,G(x;\boldsymbol{\xi})^{\lambda}}{1-G(x;\boldsymbol{\xi})^{\lambda}}\right]. \tag{3.7}$$

The ECC-G family of distributions is easily simulated by inverting (3.5) as follows: if u has a uniform $U(0,1)$ distribution, the solution of the nonlinear equation

$$x_q = G^{-1}\left[\frac{1-(1-u)^{1/\alpha}}{1-p\,(1-u)^{1/\alpha}}\right]^{1/\lambda}, \quad q \in (0,1), \tag{3.8}$$

has the density function (3.6).

3.3 Special ECC-G distributions

For $p = 0$, we obtain, as an important special case of (3.6), the *Cordeiro and de Castro*'s (CC) (2011) class of density functions. This class provides greater flexibility of its tails and can be widely applied in many areas of engineering and biology. Here, we present some special cases of the ECC-G family since it extends several useful distributions in the literature. For all cases listed below, $p \in (0,1)$, $\alpha > 0$ and $\lambda > 0$.

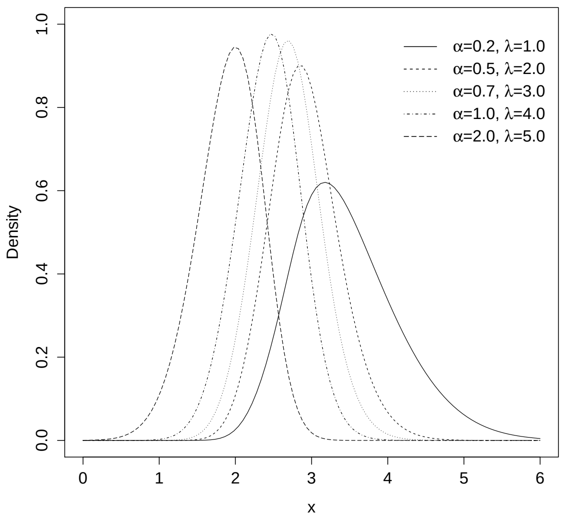
3.3.1 The ECC-normal (ECCN) distribution

The ECCN distribution is defined from (3.6) by taking $G(x)$ and $g(x)$ to be the cdf and pdf of the normal $N(\mu, \sigma^2)$ distribution. Its density function is given by

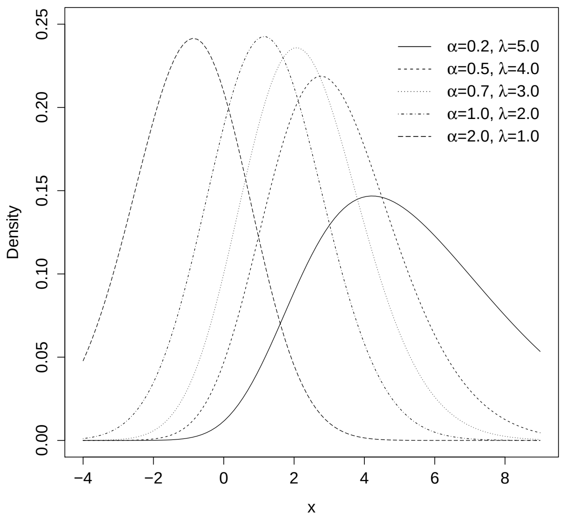
$$f(x) = \frac{\alpha\,\lambda\,(1-p)}{\sigma}\,\phi\left(\frac{x-\mu}{\sigma}\right)\,\left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\lambda-1}\,\frac{[1-\Phi(\frac{x-\mu}{\sigma})^{\lambda}]^{\alpha-1}}{[1-p\,\Phi(\frac{x-\mu}{\sigma})^{\lambda}]^{\alpha+1}}, \tag{3.9}$$

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. A random variable with density function (3.9) is denoted by $X \sim \text{ECCN}(p, \alpha, \lambda, \mu, \sigma^2)$. For $\mu = 0$, $\sigma = 1$ and $p \to 0$, we obtain the standard Kumaraswamy-normal (KwN) distribution. Furthermore, the KwN distribution with $\lambda = 1$ and $\alpha = 1$ reduces to the normal distribution.

Plots of the ECCN density function for selected parameter values are displayed in Figure 3.1. Based on these plots, we note that the parameter σ has the same dispersion property such as in the normal density.



(a) $p = 0.5, \mu = 2.0$ and $\sigma = 0.5$



(b) $p = 0.9, \mu = -2.0$ and $\sigma = 2.0$

Figure 3.1: Plots of the ECCN density function for some parameter values.

3.3.2 The ECC-Weibull (ECCW) distribution

Taking $G(x)$ as the Weibull cumulative distribution with scale parameter $\beta > 0$ and shape parameter $c > 0$, say $G(x) = 1 - e^{-(\beta x)^c}$, it follows from equation (3.6) the ECCW density function (for $x > 0$)

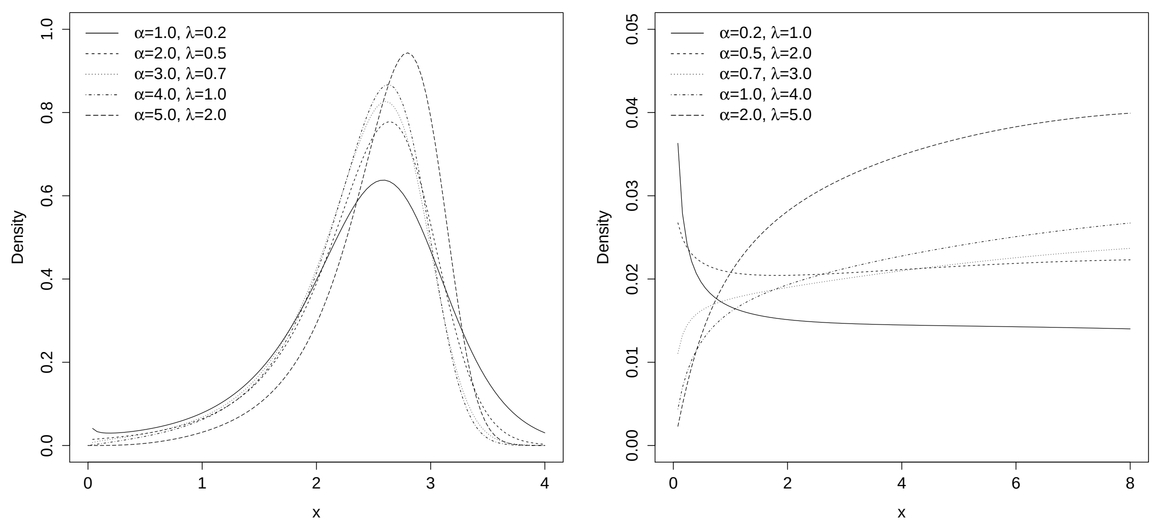
$$f(x) = \alpha\,\lambda\,(1-p)\,c\,\beta^c\,x^{c-1}\,e^{-(\beta x)^c}\,\left[1-e^{-(\beta x)^c}\right]^{\lambda-1}\,\frac{\left[1-\left(1-e^{-(\beta x)^c}\right)^{\lambda}\right]^{\alpha-1}}{\left[1-p\left(1-e^{-(\beta x)^c}\right)^{\lambda}\right]^{\alpha+1}}. \tag{3.10}$$

For $p = 0$ and $\alpha = \lambda = 1$, the ECCW distribution reduces to the classical Weibull distribution. A random variable with density function (3.10) is denoted by $X \sim \text{ECCW}(p, \alpha, \lambda, \beta, c)$. For $c = 1$, the ECCW model reduces to the Kumaraswamy-exponential-geometric (KwEG) distribution. The Kumaraswamy-Weibull (KwW) distribution follows as a special case when $p \to 0$.

The hrf corresponding to (3.10) is given by

$$h(x) = \alpha\,\lambda\,(1-p)\,c\,\beta^c\,x^{c-1}\,\mathrm{e}^{-(\beta x)^c}\,\left[1-\mathrm{e}^{-(\beta x)^c}\right]^{\lambda-1}\left[\frac{1-p\left(1-\mathrm{e}^{-(\beta x)^c}\right)^{\lambda}}{1-\left(1-\mathrm{e}^{-(\beta x)^c}\right)^{\lambda}}\right]. \tag{3.11}$$

Plots of the ECCW density and hrf for selected parameter values are displayed in Figure 3.2 and 3.3, respectively.



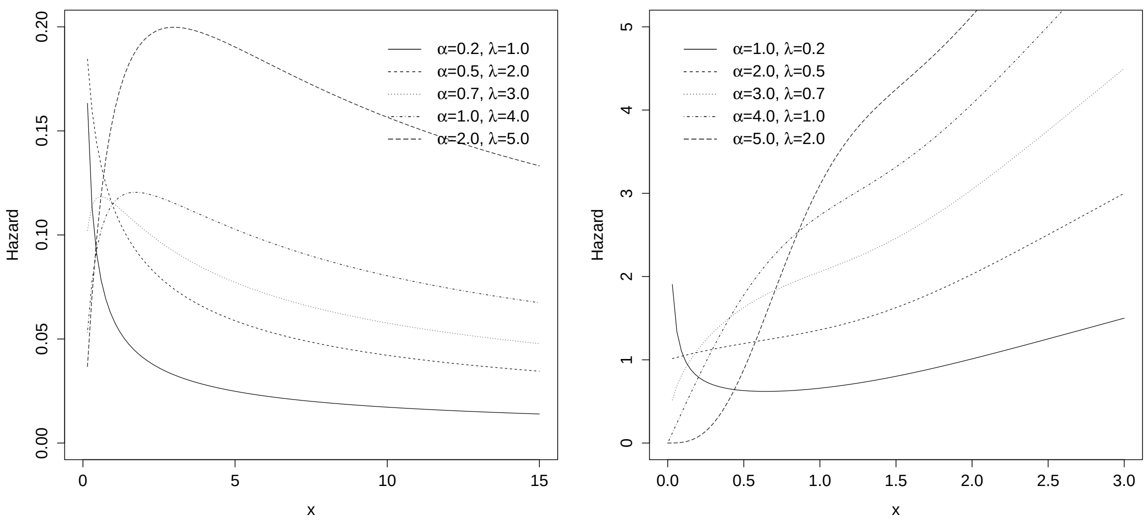
(a) $p = 0.9, c = 2.0$ and $\beta = 0.7$ (b) $p = 0.9, c = 0.5$ and $\beta = 0.7$

Figure 3.2: Plots of the ECCW density function for some parameter values.

3.3.3 The ECC-gamma (ECCG) distribution

Consider the gamma distribution with shape parameter $a > 0$ and scale parameter $b > 0$, where the pdf and cdf (for $x > 0$) are given by

$$g(x) = \frac{b^a}{\Gamma(a)}\,x^{a-1}\,\mathrm{e}^{-bx} \quad \text{and} \quad G(x) = \frac{\gamma(a, bx)}{\Gamma(a)},$$



(a) $p = 0.2, c = 0.5$ and $\beta = 0.7$ (b) $p = 0.2, c = 0.5$ and $\beta = 0.7$

Figure 3.3: Plots of the ECCW hrf for some parameter values.

where $\gamma(a, bx)$ is the incomplete gamma function. Inserting these expressions in (3.6) gives the ECCG density function

$$f(x) = \frac{\alpha\,\lambda\,(1-p)\,b^a}{\Gamma(a)^\lambda}\, x^{a-1}\,\mathrm{e}^{-bx}\,\gamma(a,bx)^{\lambda-1}\,\frac{\left[1-\left(\frac{\gamma(a,bx)}{\Gamma(a)}\right)^\lambda\right]^{\alpha-1}}{\left[1-p\left(\frac{\gamma(a,bx)}{\Gamma(a)}\right)^\lambda\right]^{\alpha+1}}.$$

The Kumaraswamy-gamma (KwG) distribution follows from this model when $p \to 0$. Plots of the ECCG density and its hrf for selected parameter values are displayed in Figures 3.4 and 3.5, respectively.

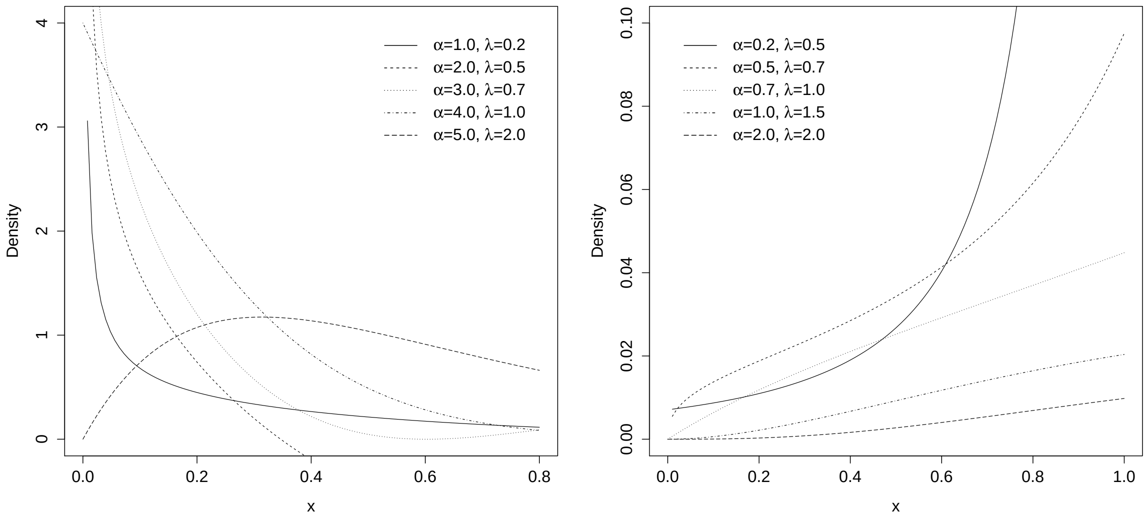
3.3.4 The ECC-beta (ECCB) distribution

Consider the beta distribution with positive shape parameters a and b and pdf and cdf (for $0 < x < 1$) given by

$$g(x) = \frac{1}{B(a,b)}\, x^{a-1}\,(1-x)^{b-1} \quad \text{and} \quad G(x) = \frac{I_x(a,b)}{B(a,b)},$$

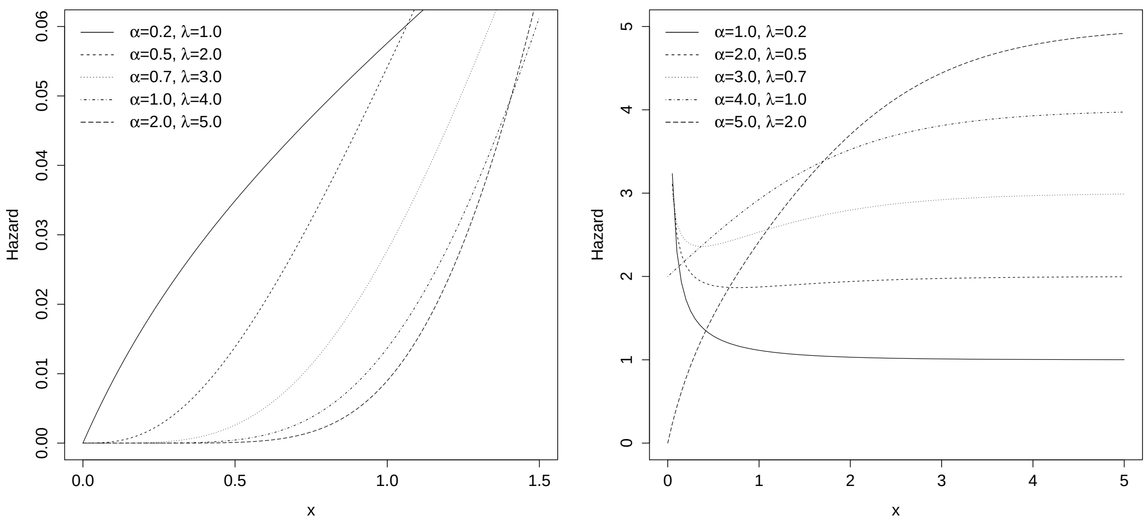
where $I_x(a,b) = \int_0^x w^{a-1}(1-w)^{b-1}\mathrm{d}w$ is the incomplete beta function and $B(a,b) = \int_0^1 w^{a-1}(1-w)^{b-1}\mathrm{d}w = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function. Inserting these expressions in (3.6) yields the ECCB density function (for $0 < x < 1$)

$$f(x) = \frac{\alpha\,\lambda\,(1-p)}{B(a,b)^\lambda}\, x^{a-1}\,(1-x)^{b-1}\, I_x(a,b)^{\lambda-1}\,\frac{\left[1-\left(\frac{I_x(a,b)}{B(a,b)}\right)^\lambda\right]^{\alpha-1}}{\left[1-p\,\left(\frac{I_x(a,b)}{B(a,b)}\right)^\lambda\right]^{\alpha+1}}.$$



(a) $a = 1, b = 2$ and $p = 0.5$ (b) $a = 2, b = 1$ and $p = 0.9$

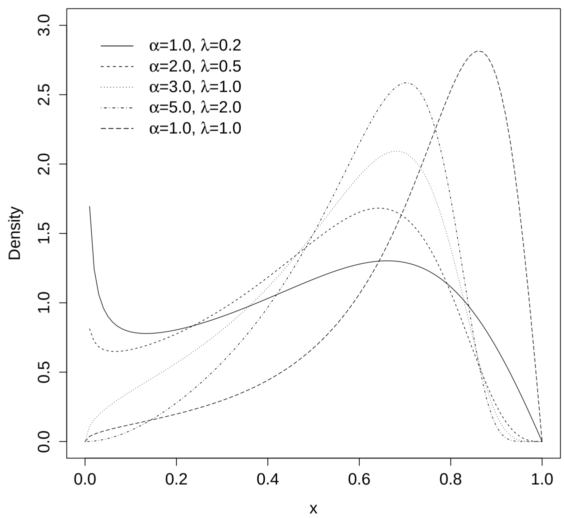
Figure 3.4: Plots of the ECCG density function for some parameter values.



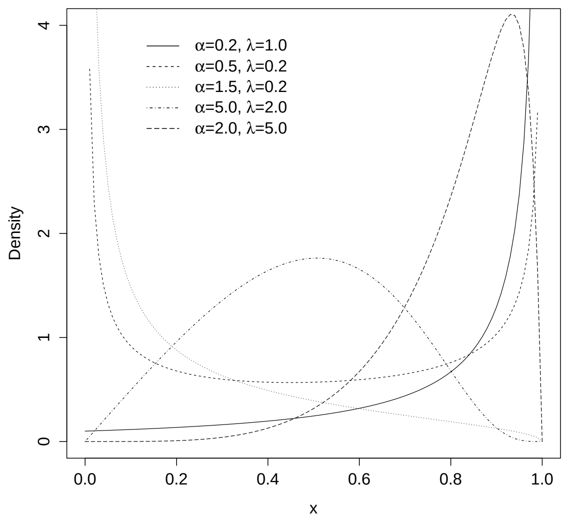
(a) $a = 2, b = 1$ and $p = 0.9$ (b) $a = 2, b = 1$ and $p = 0.5$

Figure 3.5: Plots of the ECCG hrf for some parameter values.

The Kumaraswamy beta (KwB) distribution arises as a special case when $p \to 0$. The beta distribution corresponds to the limiting case: $p \to 0$ and $\alpha = \lambda = 1$. Plots of the ECCB density function for selected parameter values are displayed in Figure 3.6.



(a) $a = 1.5, b = 2$ and $p = 0.9$



(b) $a = b = 1$ and $p = 0.5$

Figure 3.6: Plots of the ECCB density function for some parameter values.

3.4 Useful expansions

We can demonstrate that the cdf (3.5) of X has the expansion

$$F(x) = 1 - \sum_{j,k=0}^{\infty} w_{j,k} H_{(j+k)\lambda}(x), \tag{3.12}$$

where

$$w_{j,k} = (-1)^{j+k} \binom{-\alpha}{j} \binom{\alpha}{k}$$

and $H_a(x) = G(x)^a$ denotes the exponentiated-G ("exp-G" for short) cumulative distribution. Some structural properties of the exp-G distributions are studied by Mudholkar *et al.* (1996), Gupta and Kundu (2001) and Nadarajah and Kotz (2006), among others.

The density function of X can be expressed as an infinite linear combination of exp-G density functions

$$f(x; \alpha, \lambda, p, \boldsymbol{\xi}) = \sum_{j,k=0}^{\infty} \omega_{j,k}\, h_{\lambda(j+k+1)}(x), \tag{3.13}$$

where

$$\omega_{j,k} = \frac{\alpha\,\lambda\,(1-p)\,p^k}{\lambda\,(j+k+1)} (-1)^{j+k} \binom{\alpha-1}{j} \binom{-\alpha-1}{k}$$

and $h_{\lambda(j+k+1)}(x;\boldsymbol{\xi}) = \lambda(j+k+1)\, g(x;\boldsymbol{\xi})\, G(x;\boldsymbol{\xi})^{\lambda(j+k+1)-1}$ denotes the exp-G density function with power parameter $\lambda(j+k+1)$. Hereafter, a random variable having this density function is denoted by $Y_{j,k} \sim$ exp-G$(\lambda(j+k+1))$. Equation (3.13) reveals that the ECC-G density function is a linear combination of exp-G density functions. Thus, some mathematical properties of the new model can be derived from those properties of the exp-G distribution. For example, the ordinary and incomplete moments and moment generating function (mgf) of X can be obtained from those quantities of the exp-G distribution.

The formulae derived throughout the paper can be easily handled in most symbolic computation software plataforms such as Maple, Mathematica. These plataforms have currently the ability to deal with analytic expressions of formidable size and complexity. Established explicit expressions to calculate statistical measures can be more efficient than computing them directly by numerical integration. The infinity limit in these sums can be substituted by a large positive integer such as 20 or 30 for most practical purposes.

3.5 General properties

3.5.1 Characterization

The shapes of the density and hazard rate functions can be described analytically. The critical points of the ECC-G density function are the roots of the equation:

$$(\lambda-1)\frac{g(x)}{G(x)}+\frac{g'(x)}{g(x)}=\lambda\, g(x)\, G(x)^{\lambda-1}\left[\frac{\alpha-1}{1-G(x)^{\lambda}}+\frac{p\,(\alpha+1)}{1-p\,G(x)^{\lambda}}\right]. \tag{3.14}$$

There may be more than one root to (3.14). Let $\lambda(x)=\partial^2\log[f(x)]/\partial x^2$. We have

$$\begin{aligned}\lambda(x) &= (\lambda-1)\frac{g'(x)G(x)-g(x)^2}{G(x)^2}+\frac{g''(x)g(x)-g'(x)^2}{g(x)^2}\\ &- \lambda g'(x)G(x)^{\lambda-1}\left[\frac{\alpha-1}{1-G(x)^{\lambda}}+\frac{p\,(\alpha+1)}{1-p\,G(x)^{\lambda}}\right]\\ &- \lambda(\lambda-1)g(x)^2G(x)^{\lambda-2}\left[\frac{\alpha-1}{1-G(x)^{\lambda}}+\frac{p\,(\alpha+1)}{1-p\,G(x)^{\lambda}}\right]\\ &- \lambda^2 g(x)^2G(x)^{2\lambda-2}\left[\frac{\alpha-1}{(1-G(x)^{\lambda})^2}+\frac{p\,(\alpha+1)}{(1-p\,G(x)^{\lambda})^2}\right].\end{aligned}$$

If $x=x_0$ is a root of (3.14) then it corresponds to a local maximum if $\lambda(x_0)>0$ for all $x<x_0$ and $\lambda(x_0)<0$ for all $x>x_0$. It corresponds to a local minimum if $\lambda(x_0)<0$ for all $x<x_0$ and $\lambda(x_0)>0$ for all $x>x_0$. It refers to a point of inflexion if either $\lambda(x)>0$ for all $x\neq x_0$ in the neighbourhood of x_0 or $\lambda(x)<0$ for all $x\neq x_0$ in the neighbourhood of x_0.

The critical point of $h(x)$ is obtained from the equation

$$\frac{g'(x)}{g(x)}+(\lambda-1)\frac{g(x)}{G(x)}-p\,\lambda\, g(x)\frac{G(x)^{\lambda-1}}{1-pG(x)^{\lambda-1}}=\lambda\, g(x)\frac{G(x)^{\lambda-1}}{1-G(x)^{\lambda-1}}. \tag{3.15}$$

There may be more than one root to (3.15). Let $\tau(x) = d^2 \log[h(x)]/dx^2$. We have

$$\begin{aligned}\tau(x) = & \frac{g''(x)g(x) - [g'(x)]^2}{g(x)^2} + (\lambda - 1)\,\frac{g'(x)G(x) - [g(x)]^2}{G(x)^2} \\ & - p\,\lambda\, g'(x)\frac{G(x)^{\lambda-1}}{1 - p\,G(x)^\lambda} - p\,\lambda\,(\lambda-1)\,g(x)^2\frac{G(x)^{\lambda-2}}{1 - p\,G(x)^\lambda} - p^2\lambda^2 g(x)^2\frac{G(x)^{2\lambda-2}}{[1 - p\,G(x)^\lambda]^2} \\ & - \lambda\, g'(x)\frac{G(x)^{\lambda-1}}{1 - G(x)^\lambda} - \lambda\,(\lambda-1)\,g(x)^2\frac{G(x)^{\lambda-2}}{1 - G(x)^\lambda} - \lambda^2 g(x)^2\frac{G(x)^{2(\lambda-1)}}{[1 - G(x)^\lambda]^2} = 0.\end{aligned}$$

If $x = x_0$ is a root of (3.15) then it refers to a local maximum if $\tau(x_0) > 0$ for all $x < x_0$ and $\tau(x_0) < 0$ for all $x > x_0$. It corresponds to a local minimum if $\tau(x_0) < 0$ for all $x < x_0$ and $\tau(x_0) > 0$ for all $x > x_0$. It gives an inflexion point if either $\tau(x) > 0$ for all $x \neq x_0$ or $\tau(x) < 0$ for all $x \neq x_0$.

3.5.2 Quantile power series

Power series methods are at the heart of many aspects of applied mathematics and statistics. Quantile functions are in widespread use in probability distributions and general statistics and often find representations in terms of power series. The quantile function for a distribution has many uses in both the theory and statistical applications. It may be used to generate values of a random variable having $F(x)$ as its distribution function. This fact serves as the basis of a method for simulating a sample from an arbitrary distribution with the aid of a random number generator.

We derive explicit expressions for the moments and generating function of the ECC family of distributions using a power series for the quantile function $x = Q(u) = F^{-1}(u)$ of X obtained by expanding (3.8), which is easily computed using a linear recurrent equation for its coefficients. If the G quantile function, say $Q_G(u)$, does not have a closed-form expression, it can usually be expressed in terms of a power series

$$Q_G(u) = \sum_{i=0}^{\infty} a_i\, u^i, \tag{3.16}$$

where the coefficients a_i's are suitably chosen real numbers which depend on the parameters of the G distribution. For several important distributions, such as the normal, Student t, gamma and beta distributions, $Q_G(u)$ does not have explicit expressions but it can be expanded as in equation (3.16). As a simple example, for the normal $N(0,1)$ distribution, $a_i = 0$ for $i = 0, 2, 4, \ldots$ and $a_i = b_{(i-1)/2}$ for $i = 1, 3, 5, \ldots$, where the quantities $b_{(i-1)/2}$ can be determinated recursively from

$$b_{k+1} = \frac{1}{2(2k+3)} \sum_{r=0}^{k} \frac{(2r+1)\,(2k-2r+1)\,b_r\,b_{k-r}}{(r+1)\,(2r+1)}.$$

We have $a_1 = 1$, $a_3 = 1/6$, $a_5 = 7/120$ and $a_7 = 127/7560, \ldots$.

From now on, we use a result by Gradshteyn and Ryzhik (2000, Section 0.314) for a power series raised to a positive integer n (for $n \geq 1$)

$$Q_G(u)^n = \left(\sum_{i=0}^{\infty} a_i\, u^i\right)^n = \sum_{i=0}^{\infty} c_{n,i}\, u^i, \tag{3.17}$$

where the coefficients $c_{n,i}$ (for $i = 1, 2, \ldots$) are easily obtained from the recurrence equation (with $c_{n,0} = a_0^n$)

$$c_{n,i} = (i\, a_0)^{-1} \sum_{m=1}^{i} [m(n+1) - i]\, a_m\, c_{n,i-m}. \tag{3.18}$$

Clearly, the quantity $c_{n,i}$ can be determined from $c_{n,0}, \ldots, c_{n,i-1}$ and then from the quantities $a_0, \ldots, a_i$. The coefficient $c_{n,i}$ can be given explicitly in terms of the coefficients a_i's, although it is not necessary for programming numerically our expansions in any algebraic or numerical software. For the normal $N(0,1)$ distribution, the coefficients $c_{n,i}$ can be obtained from (3.17) using the a_i's given above.

Next, we derive an expansion for the argument of $Q_G(\cdot)$ in (3.8)

$$A = \frac{[1 - (1-u)^{1/\alpha}]^{1/\lambda}}{[1 - p\,(1-u)^{1/\alpha}]^{1/\lambda}}.$$

Using the generalized binomial expansion three times since $u \in (0,1)$, we can write

$$A = \sum_{r,s,t=0}^{\infty} (-1)^{r+s+t}\, p^r \binom{-\lambda^{-1}}{r} \binom{\lambda^{-1}}{s} \binom{(r+s)\alpha^{-1}}{t} u^t.$$

Then, the quantile function of X can be expressed from (3.8) as

$$Q(u) = Q_G\left(\sum_{t=0}^{\infty} \delta_t\, u^t\right), \tag{3.19}$$

where

$$\delta_t = \sum_{r,s=0}^{\infty} (-1)^{r+s+t}\, p^r \binom{\lambda^{-1}}{r} \binom{\lambda^{-1}}{s} \binom{(r+s)\alpha^{-1}}{t}.$$

For any baseline G distribution, we combine (3.16) and (3.19) to obtain

$$Q(u) = Q_G\left(\sum_{t=0}^{\infty} \delta_t\, u^t\right) = \sum_{i=0}^{\infty} a_i \left(\sum_{t=0}^{\infty} \delta_t\, u^t\right)^i,$$

and then using (3.17) and (3.18), we have

$$Q(u) = \sum_{t=0}^{\infty} e_t\, u^t, \tag{3.20}$$

where $e_t = \sum_{i=0}^{\infty} a_i\, d_{i,t}$, $d_{i,0} = \delta_0^i$ and (for $t > 1$)

$$d_{i,t} = (t\, \delta_0)^{-1} \sum_{m=1}^{t} [m(i+1) - t]\, \delta_m\, d_{i,t-m}.$$

Equation (3.20) is the main result of this section. It allows to obtain various mathematical quantities for the ECC-G family as demonstrated in the next sections.

3.5.3 Generating function

Here, we provide two general formulae for the mgf $M(t) = E(e^{tX})$ of X. A first formula for $M(t)$ follows from (3.13) as

$$M(t) = \sum_{j,k=0}^{\infty} \omega_{j,k} \, M_{j,k}(t), \tag{3.21}$$

where $M_{j,k}(t)$ is the mgf of $Y_{j,k}$. Hence, $M(t)$ can be immediately determined from the generating function of the exp-G distribution. We now provide three applications of equation (3.21). For example, the generating functions of the ECC-exponential (with parameter β) (for $t < 1/\beta$), ECCPa (with parameter $\nu > 0$ real non integer) and ECCSL (for $t < 1$) distributions follow from equation (3.21) as

$$M(t) = \sum_{j,k=0}^{\infty} [\lambda(j+k+1)] \, B(\lambda(j+k+1), 1-\beta t) \, \omega_{j,k},$$

$$M(t) = e^{-t} \sum_{j,k,m=0}^{\infty} [\lambda(j+k+1)] \; B\left(\lambda(j+k+1), 1 - m\nu^{-1}\right) \, \omega_{j,k} \, \frac{t^m}{m!}$$

and

$$M(t) = \sum_{j,k=0}^{\infty} [\lambda(j+k+1)] \, B(t + \lambda(j+k+1), 1-t) \, \omega_{j,k},$$

respectively.

We now provide a fourth application of (3.21) by taking again as the baseline the Weibull distribution with scale parameter β and shape parameter c (see Section 3.2). The generating function of the exp-Weibull distribution with power parameter $\lambda(j+k+1)$ is given by

$$M_{j,k}(t) = \sum_{r=0}^{\infty} v_{j,k}^{(r)} \, I_r(t), \tag{3.22}$$

where

$$v_{j,k}^{(r)} = \beta \, c^{\beta} \, [\lambda(j+k+1)] \sum_{i=0}^{\infty} (-1)^{i+r} \binom{[\lambda(j+k+1)](i+1) - 1}{r},$$

$\delta_r = \beta \, (r+1)^{1/c}$ and

$$I_r(t) = \int_0^{\infty} x^{c-1} \, \exp\{t\,x - (\delta_r \, x)^c\} dx.$$

Pascoa *et al.* (2011) derived two different formulae for $I_r(t)$ which hold for: (i) $c > 1$ or (ii) for $c = p/q$, where $p \geq 1$ and $q \geq 1$ are co-prime integers. The first representation for $I_r(t)$ is given in terms of the Wright generalized hypergeometric function (Wright, 1935) defined by

$${}_p\Psi_q \left[\begin{array}{c} (\alpha_1, A_1), \cdots, (\alpha_p, A_p) \\ (\beta_1, B_1), \cdots, (\beta_q, B_q) \end{array} ; x \right] = \sum_{n=0}^{\infty} \frac{\prod_{j=1}^{p} \Gamma(\alpha_j + A_j \, n)}{\prod_{j=1}^{q} \Gamma(\beta_j + B_j \, n)} \, \frac{x^n}{n!}.$$

We can write

$$\begin{aligned} I_r(t) &= \sum_{m=0}^{\infty} \frac{t^m}{m!} \int_0^{\infty} x^{m+\beta-1} \exp\{-(\delta_r\, x)^c\} dx = \frac{1}{\beta\, \delta_r^{\beta}} \sum_{m=0}^{\infty} \frac{t^m}{\delta_r^m\, m!} \Gamma(mc^{-1}+1) \\ &= \frac{1}{\beta\, \delta_r^{\beta}}\, {}_1\Psi_0 \left[\begin{array}{c} (1, \beta^{-1}) \\ - \end{array} ; \frac{t}{\delta_r} \right]. \end{aligned} \tag{3.23}$$

The function $I_r(t)$ exists if $1 + \sum_{j=1}^{q} B_j - \sum_{j=1}^{p} A_j > 0$.

Using equations (3.21), (3.22) and (3.23), we obtain (assuming that $\lambda > 1$)

$$M(t) = c^{-1} \sum_{j,k,r=0}^{\infty} \frac{\omega_{j,k}\, v_{j,k}^{(r)}}{\delta_r}\, {}_1\Psi_0 \left[\begin{array}{c} (1, c^{-1}) \\ - \end{array} ; \frac{t}{\delta_r} \right]. \tag{3.24}$$

A second representation for $I_r(t)$ is based on the Meijer G-function defined by

$$G_{p,q}^{m,n}\left(x \left| \begin{array}{c} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{array} \right. \right) = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^{m} \Gamma\left(b_j + t\right) \prod_{j=1}^{n} \Gamma\left(1 - a_j - t\right)}{\prod_{j=n+1}^{p} \Gamma\left(a_j + t\right) \prod_{j=m+1}^{p} \Gamma\left(1 - b_j - t\right)}\, x^{-t} dt,$$

where $\mathrm{i} = \sqrt{-1}$ is the complex unit and L denotes an integration path; see Section 9.3 in Gradshteyn and Ryzhik (2000) for a description of this path. The Meijer G-function contains many integrals with elementary and special functions (Prudnikov *et al.*, 1986). From the result $\exp\{-g(x)\} = G_{0,1}^{1,0}\left(g(x) \mid \begin{array}{c} - \\ 0 \end{array} \right)$ for an arbitrary $g(\cdot)$ function, $I_r(t)$ becomes

$$I_r(t) = \int_0^{\infty} x^{c-1} \exp\{sx - (\delta_r\, x)^c\} dx = \int_0^{\infty} x^{v-1}\, \mathrm{e}^{sx}\, G_{0,1}^{1,0}\left(\delta_r^c\, x^c \middle| \begin{array}{c} - \\ 0 \end{array} \right) dx.$$

We now assume that $c = p/q$, where $p \geq 1$ and $q \geq 1$ are co-prime integers. Note that this condition for calculating the integral $I_r(t)$ is not restrictive since every real number can be approximated by a rational number. Using equation (2.24.1.1) in Prudnikov *et al.* (1986, volume 3), we have

$$I_r(t) = \frac{p^{p/q-1/2}(-t)^{-p/q}}{(2\pi)^{(p+q)/2-1}}\, G_{q,p}^{p,q}\left(\frac{\delta_r^q\, p^{p+q}}{(-t)^p q^{2q}} \middle| \begin{array}{c} \frac{q-p}{pq}, \frac{2q-p}{pq}, \ldots, \frac{pq-p}{pq} \\ 0, \frac{1}{q}, \ldots, \frac{q-1}{q} \end{array} \right). \tag{3.25}$$

Using (3.21), (3.22) and (3.25), we can obtain $M(t)$ for the ECCW distribution.

A second general formula for $M(t)$ can be derived from (3.13) as

$$M(t) = \sum_{j,k=0}^{\infty} [\lambda(j+k+1)]\, \omega_{j,k}\, \rho(t, \lambda(j+k+1)-1), \tag{3.26}$$

where $\rho(t,a)$ can be determined from the baseline quantile function $Q_G(x)$ by

$$\rho(t,a) = \int_{-\infty}^{\infty} \mathrm{e}^{t\,x}\, G(x)^a\, g(x) dx = \int_0^1 \exp\left\{t\, Q_G(u)\right\}\, u^a du. \tag{3.27}$$

An alternative expression for $\rho(t,a)$ in terms of the coefficients of the G quantile function follows using the power series for the exponential function and (3.17) and then integrating the result. We obtain

$$\rho(t,a) = \sum_{n,i=0}^{\infty} \frac{c_{n,i}\ t^n}{(a+i+1)\, n!}. \tag{3.28}$$

We can derive the mgf's of several ECC distributions from equations (3.21) and (3.26), the last one combining with (3.27) or (3.28). Equations (3.21) and (3.26) are the main results of this section.

3.5.4 Moments

Here, we provide two general formulae for the nth moment of X. The first one is obtained from (3.13) as

$$\mu'_n = E(X^n) = \sum_{j,k=0}^{\infty} \omega_{j,k}\, E(Y_{j,k}^n) = \sum_{j,k=0}^{\infty} \omega_{j,k} \int_{-\infty}^{\infty} x^n\, h_{\lambda(j+k+1)}(x;\boldsymbol{\xi}). \tag{3.29}$$

Expressions for moments of some exponentiated distributions are given by Nadarajah and Kotz (2006). They can be used to obtain μ'_n. We now provide an application of (3.29) for the ECCW distribution discussed in Section 3.2, where $G(x) = 1 - e^{-(\beta\, x)^c}$, $c > 0$ is a shape parameter and $\beta > 0$ a scale parameter. The corresponding exp-Weibull (exp-W) density function with power parameter $\lambda(j+k+1)$ is given by

$$h_{\lambda(j+k+1)}(x;\beta,c) = \lambda(j+k+1)\, c\, \beta^c\, x^{c-1}\, \mathrm{e}^{-(\beta x)^c}\, [1-\mathrm{e}^{-(\beta x)^c}]^{\lambda(j+k+1)-1}. \tag{3.30}$$

The nth moment of (3.30), say $\rho_{j,k}^{(n)}$, can be obtained from Cordeiro *et al.* (2011) as

$$\rho_{j,k}^{(n)} = \frac{\Gamma(n/c+1)}{\beta^n} \sum_{r=0}^{\infty} \frac{w_{j,k}^{(r)}}{(r+1)^{n/c}}, \tag{3.31}$$

where

$$w_{j,k}^{(r)} = \frac{[\lambda(j+k+1)]}{(r+1)} \sum_{i=0}^{\infty} (-1)^{i+r} \binom{[\lambda(j+k+1)](i+1)-1}{r}.$$

Combining equations (3.29) and (3.31), we can write μ'_n as

$$\mu'_n = \frac{\Gamma(n/c+1)}{\beta^n} \sum_{k,j,r,i=0}^{\infty} \frac{(-1)^{i+r}\, [\lambda(j+k+1)]\, \omega_{j,k}}{(r+1)^{n/c+1}} \binom{[\lambda(j+k+1)](i+1)-1}{r}.$$

Next, we provide two more examples from (3.29). First, for the ECC-Pareto (ECCPa) distribution, where the baseline cdf is $G(x) = 1-(1+x)^{-\nu}$ and $\nu > 0$, we obtain (for ν real non integer)

$$\mu'_n = \sum_{k,j,m=0}^{\infty} (-1)^{n+m}\, [\lambda(j+k+1)]\, B(\lambda(j+k+1)-1, 1-m\nu^{-1})\, \omega_{j,k} \binom{n}{m}.$$

Second, for the ECC-standard logistic (ECCSL) distribution, where $G(x) = (1 + e^{-x})^{-1}$, we can write using a result by Prudnikov *et al.* (1986, Section 2.6.13, equation 4) (for $t < 1$)

$$\mu'_n = \sum_{k,j=0}^{\infty} [\lambda(j+k+1)]\, \omega_{j,k} \left(\frac{\partial}{\partial t}\right)^n B(t + \lambda(j+k+1), 1-t)\bigg|_{t=0}.$$

A second general formula for μ'_n follows from (3.13) and the baseline quantile function $Q_G(u)$. We can write

$$\mu'_n = \sum_{k,j=0}^{\infty} (\alpha + k + j)\, \omega_{j,k}\, \tau(n, \lambda(j+k+1) - 1), \tag{3.32}$$

where $\tau(n, a)$ is given by

$$\tau(n, a) = \int_{-\infty}^{\infty} x^n\, G(x)^a\, g(x) dx = \int_0^1 Q_G(u)^n\, u^a du.$$

Inserting (3.17) in the last equation and integrating, we obtain

$$\tau(n, a) = \sum_{i=0}^{\infty} \frac{c_{n,i}}{(a+1)}, \tag{3.33}$$

where the quantities $c_{n,i}$ can be determined from (3.18).

The central moments (μ_n) and cumulants (κ_n) of X can be determined from (3.29) or (3.32) as

$$\mu_n = \sum_{k=0}^{r} (-1)^k \binom{n}{k} \mu_1'^{n}\, \mu'_{n-k} \qquad \text{and} \qquad \kappa_n = \mu'_n - \sum_{k=1}^{n-1} \binom{n-1}{k-1} \kappa_k\, \mu'_{n-k},$$

respectively, where $\kappa_1 = \mu'_1$. Thus, $\kappa_2 = \mu'_2 - \mu_1'^2$, $\kappa_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu_1'^3$, $\kappa_4 = \mu'_4 - 4\mu'_3\mu'_1 - 3\mu_2'^2 + 12\mu'_2\mu_1'^2 - 6\mu_1'^4$, etc. The skewness $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and kurtosis $\gamma_2 = \kappa_4/\kappa_2^2$ can be calculated from the third and fourth standardized cumulants.

3.5.5 Incomplete moments

The answers to many important questions in economics require more than just knowing the mean of a distribution, but its shape as well. This is obvious not only in the study of econometrics (for example, asymmetric error terms cannot be generated by the commonly assumed spherical distributions) and income distribution, but in other areas as well. Incomplete moments of the income distribution form natural building blocks for measuring inequality: for example, the Lorenz and Bonferroni curves and Pietra and Gini measures of inequality all depend upon the incomplete moments of the income distribution.

The nth incomplete moment of X is defined as $m_n(y) = E(X^n | X < y) = \int_{-\infty}^{y} x^r f(x) dx$. Here, we propose two methods to determine the incomplete moments of the new family. First, the nth incomplete moment of X can be expressed as

$$m_n(y) = \sum_{j,k=0}^{\infty} [\lambda(j+k+1)]\, \omega_{j,k} \int_0^{G(y;\boldsymbol{\xi})} Q_G(u)^n\, u^{\lambda(j+k+1)}\, du. \tag{3.34}$$

The integral in (3.34) can be computed at least numerically for most baseline distributions. A second method to obtain the incomplete moments of X follows from (3.34) using equations (3.17) and (3.18). We obtain

$$m_n(y) = \sum_{j,k,i=0}^{\infty} \frac{[\lambda(j+k+1)]\,\omega_{j,k}\,c_{n,i}}{[\lambda(j+k+1)+i]}\,G(y;\boldsymbol{\xi})^{\lambda(j+k+1)+i}. \tag{3.35}$$

3.5.6 Mean deviations

Let $X \sim$ ECC-G$(p,\alpha,\lambda,\boldsymbol{\xi})$. The mean deviations about the mean $(\delta_1(X))$ and about the median $(\delta_2(X))$ can be expressed as

$$\delta_1(X) = E(|X-\mu_1'|) = 2\mu_1' F(\mu_1') - 2T(\mu_1') \quad \text{and} \quad \delta_2(X) = E(|X-M|) = \mu_1' - 2T(M),$$

respectively, where $\mu_1' = E(X)$, $M = Median(X)$ denotes the median, $F(\mu_1')$ comes from equation (3.5) and $T(z) = \int_{-\infty}^{z} x f(x)\mathrm{d}x$. The median M follows from equation (3.5) as

$$M = G^{-1}\left[\left(\frac{1-2^{-1/\alpha}}{1-p\,2^{-1/\alpha}}\right)^{1/\lambda}\right].$$

Then, using ordinary and incomplete moments, we can easily obtain $\delta_1(X)$ and $\delta_2(X)$.

3.5.7 Quantile measure

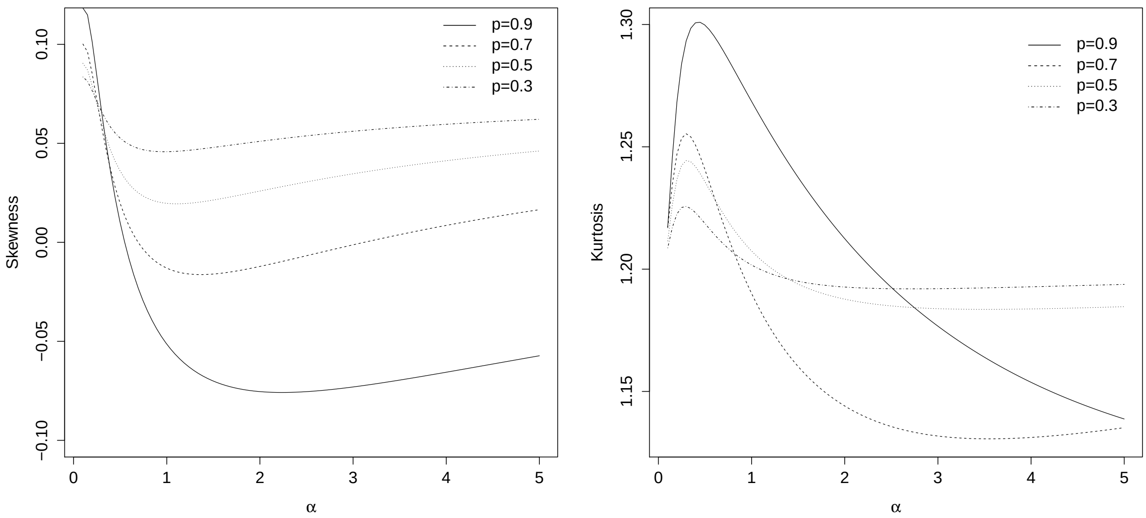
The effects of the shape parameters a and b on the skewness and kurtosis can be based on quantile measures. The shortcomings of the classical kurtosis measure are well-known. The Bowley skewness (Kenney and Keeping, 1962) is one of the earliest skewness measures defined by the average of the quartiles minus the median, divided by half the interquartile range, namely

$$B = \frac{Q\left(\frac{3}{4}\right) + Q\left(\frac{1}{4}\right) - 2Q\left(\frac{1}{2}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}.$$

Since only the middle two quartiles are considered and the other two quartiles are ignored, this adds robustness to the measure. The Moors kurtosis (Moors, 1998) is based on octiles

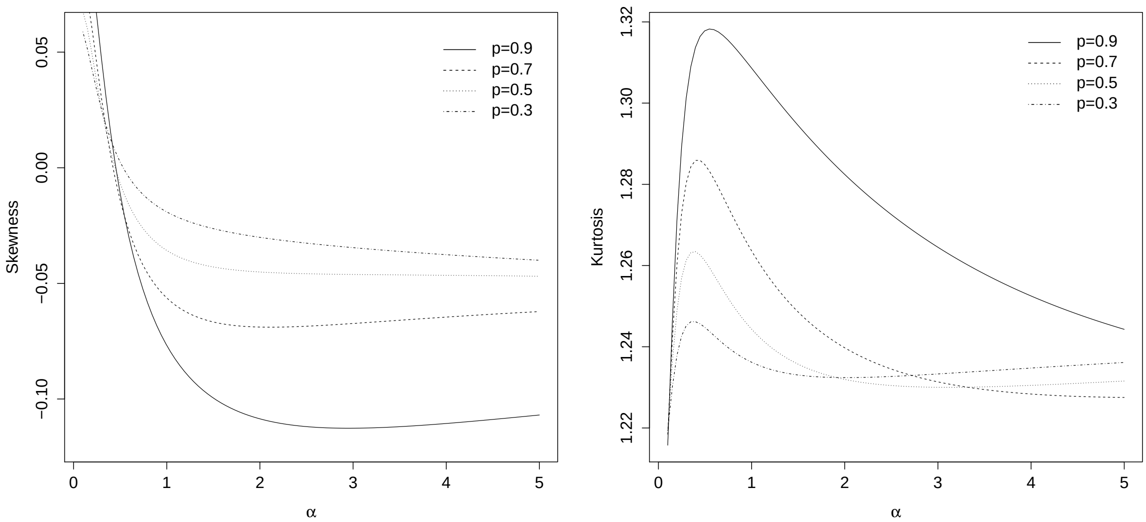
$$M = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)}.$$

These measures are less sensitive to outliers and they exist even for distributions without moments. Plots of the skewness and kurtosis for the distributions ECCW and ECCN (discussed in Section 3) and selected parameter values are displayed in Figures 3.7 and 3.8, respectively. These plots indicate how both measures B and M vary depending on the values of the shape parameters.



(a) $\lambda = 1$ and $\beta = c = 2$ (b) $\lambda = 1$ and $\beta = c = 2$

Figure 3.7: Skewness (a) and Kurtosis (b) of the ECCW distribution.



(a) $\lambda = 1$, $\mu = -2$ and $\sigma = 0.5$ (b) $\lambda = 1$, $\mu = -2$ and $\sigma = 0.5$

Figure 3.8: Skewness (a) and Kurtosis (b) of the ECCN distribution.

3.5.8 Entropies

An entropy is a measure of variation or uncertainty of a random variable X. Two popular entropy measures are the Rényi and Shannon entropies (Shannon, 1948; Rényi, 1961). The Rényi entropy of a random variable with pdf $f(x)$ is defined as

$$I_R(c) = \frac{1}{1-c}\log\left(\int_0^\infty f^c(x)dx\right),$$

for $c > 0$ and $c \neq 1$. After some algebraic developments, we obtain an alternative expression for $I_R(c)$

$$I_R(c) = \frac{1}{1-c}\log\left[\frac{(\alpha\lambda(1-p))^c}{\lambda}\right]\sum_{i=0}^{\infty}(-p)^i B(i+1,(\alpha-1)c+1)\,\mathrm{E}_{Y_i}\left\{g^{c-1}\left[G^{-1}\left(Y^{1/\lambda}\right)\right]\right\},$$

where $Y_i \sim B(i+1,(\alpha-1)c+1)$. The Shannon entropy of a random variable X is defined by $E\{-\log[f(X)]\}$. It is the special case of the Rényi entropy when $c \uparrow 1$. Direct calculation yields

$$\begin{aligned}\mathrm{E}\{-\log[f(X)]\} &= -\log[\alpha\lambda(1-p)] - \mathrm{E}\{\log[g(X;\boldsymbol{\xi})]\} - (\lambda-1)\,\mathrm{E}\{\log[G(x;\boldsymbol{\xi})]\}\\ &- (\alpha-1)\,\mathrm{E}\left\{\log\left[1-G(x;\boldsymbol{\xi})^\lambda\right]\right\} + (\alpha+1)\,\mathrm{E}\left\{\log\left[1-p\,G(x;\boldsymbol{\xi})^\lambda\right]\right\}.\end{aligned}$$

After some algebraic manipulations we obtain:

Proposition 4. *Let X be a random variable with pdf* (3.6). *Then,*

$$\mathrm{E}\{\log[G(X)]\} = \frac{(1-\alpha)}{\lambda}\sum_{i=0}^{\infty}(-p)^i\binom{-\alpha-1}{i}B(i+1,\alpha)\,[\psi(i+1)-\psi(i+\alpha+1)],$$

$$\mathrm{E}\left\{\log\left[1-G(X)^\lambda\right]\right\} = \alpha(1-p)\sum_{i=0}^{\infty}(-p)^i\binom{-\alpha-1}{i}B(i+1,\alpha)\,[\psi(\alpha)-\psi(i+\alpha+1)],$$

$$\mathrm{E}\left\{\log\left[1-p\,G(X)^\lambda\right]\right\} = \alpha(1-p)\sum_{i=0}^{\infty}(-p)^i\,B(\alpha,i+1)\left[\frac{d}{dt}\binom{t-\alpha-1}{i}\bigg|_{t=0}\right],$$

where $\psi(\cdot)$ is the digamma function.

The simplest formula for the entropy of X is given by

$$\begin{aligned}\mathrm{E}\{-\log[f(X)]\} &= -\log[\alpha\lambda(1-p)] - \mathrm{E}\{\log[g(X;\boldsymbol{\xi})]\}\\ &+ \frac{(\alpha-1)(\lambda-1)}{\lambda}\sum_{i=0}^{\infty}(-p)^i\binom{-\alpha-1}{i}B(i+1,\alpha)\,[\psi(i+1)-\psi(i+\alpha+1)]\\ &+ \alpha\,(1-\alpha)(1-p)\sum_{i=0}^{\infty}(-p)^i\binom{-\alpha-1}{i}B(i+1,\alpha)\,[\psi(\alpha)-\psi(i+\alpha+1)]\\ &+ \alpha\,(\alpha+1)(1-p)\sum_{i=0}^{\infty}(-p)^i\,B(\alpha,i+1)\left[\frac{d}{dt}\binom{t-\alpha-1}{i}\bigg|_{t=0}\right].\end{aligned}$$

3.5.9 Extreme values

If $\overline{X} = (X_1 + \cdots + X_n)/n$ denotes the mean of a random sample from (3.5), then by the usual central limit theorem $\sqrt{n}(\overline{X} - E(X))/\sqrt{Var(X)}$ approaches the standard normal distribution as $n \to \infty$ under suitable conditions. Sometimes one would be interested in the asymptotics of the extreme values $M_n = \max(X_1, \ldots, X_n)$ and $m_n = \min(X_1, \ldots, X_n)$.

First, suppose that G belongs to the max domain of attraction of Gumbel extreme value distribution. Then by Leadbetter *et al.* (1987, Chapter 1), there must exist a strictly positive function, say $h(t)$, such that

$$\lim_{t\to\infty} \frac{1 - G(t + xh(t))}{1 - G(t)} = e^{-x},$$

for every $x \in (-\infty, \infty)$. But

$$\begin{aligned} \lim_{t\to\infty} \frac{1 - F(t + xh(t))}{1 - F(t)} &= \lim_{t\to\infty} \left[\frac{1 - G(t + xh(t))^{\lambda}}{1 - G(t)^{\lambda}} \right]^{\alpha} \\ &= \lim_{t\to\infty} \left[\frac{1 - G(t + xh(t))}{1 - G(t)} \right]^{\alpha} = e^{-\alpha x}, \end{aligned}$$

for every $x \in (-\infty, \infty)$. So, it follows by Leadbetter *et al.* (1987, Chapter 1) that F belongs to the max domain of attraction of the Gumbel extreme value distribution with

$$\lim_{n\to\infty} P[a_n(M_n - b_n \le x)] = \exp\left[-\exp(-\lambda x)\right]$$

for some suitable norming constants $a_n > 0$ and b_n. Second, suppose that G belongs to the max domain of attraction of the Fréchet extreme value distribution. Then by Leadbetter *et al.* (1987, chapter 1), there must exist a $\beta > 0$ such that

$$\lim_{t\to\infty} \frac{1 - G(t + xh(t))}{1 - G(t)} = x^{\beta}$$

for every $x \in (-\infty, \infty)$. But

$$\begin{aligned} \lim_{t\to\infty} \frac{1 - F(t + xh(t))}{1 - F(t)} &= \lim_{t\to\infty} \left[\frac{1 - G(t + xh(t))^{\lambda}}{1 - G(t)^{\lambda}} \right]^{\alpha} \\ &= \lim_{t\to\infty} \left[\frac{1 - G(t + xh(t))}{1 - G(t)} \right]^{\alpha} = x^{\alpha\beta}, \end{aligned}$$

for every $x > 0$. So, it follows by Leadbetter *et al.* (1987, chapter 1) that F belongs to the max domain of attraction of the Gumbel extreme value distribution with

$$\lim_{n\to\infty} P[a_n(M_n - b_n \le x)] = \exp(-x^{\alpha\beta})$$

for some suitable norming constants $a_n > 0$ and b_n. Third, suppose that G belongs to the max domain of attraction of the Weibull extreme value distribution. Then, by Leadbetter *et al.* (1987, chapter 1), there must exist a $\gamma > 0$ such that

$$\lim_{t\to 0} \frac{G(tx)}{G(t)} = x^{\gamma}$$

for every $x < 0$. But
$$\lim_{t\to 0}\frac{F(tx)}{F(t)} = \lim_{t\to 0}\left[\frac{G(tx)}{G(t)}\right]^{\alpha} = x^{\alpha\gamma}$$
for every $x < 0$. So, it follows by Leadbetter *et al.* (1987, chapter 1) that F belongs to the max domain of attraction of the Weibull extreme value distribution with

$$\lim_{n\to\infty} P[a_n(M_n - b_n \leq x)] = \exp\left[-(-x)^{\alpha\gamma}\right]$$

for some suitable norming constants $a_n > 0$ and b_n. We conclude that F belongs to the same min domain of attraction as that of G. If $X \sim \text{ECCG}(\boldsymbol{\xi}, \lambda, p, \alpha)$, then $M_n \sim \text{ECCG}(\boldsymbol{\xi}, \lambda, p, n\alpha)$.

3.6 Estimation

Here, we determine the maximum likelihood estimates (MLEs) of the model parameters of the new family from complete samples only. Let $x_1, \ldots, x_n$ be observed values from the ECC-G distribution with parameters p, α, λ and $\boldsymbol{\xi}$. Let $\Theta = (p, \alpha, \lambda, \boldsymbol{\xi})^\top$ be the $r \times 1$ parameter vector. The total log-likelihood function for Θ is given by

$$\begin{aligned}\ell_n &= \ell_n(\Theta) = n\,\log\alpha + n\,\log\lambda + n\,\log(1-p) + \sum_{i=1}^{n}\log\left[g(x;\boldsymbol{\xi})\right] + (\lambda-1)\sum_{i=1}^{n}\log\left[G(x;\boldsymbol{\xi})\right] \\ &+ (\alpha-1)\sum_{i=1}^{n}\log\left[1 - G(x;\boldsymbol{\xi})^{\lambda}\right] - (\alpha+1)\sum_{i=1}^{n}\log\left[1 - p\,G(x;\boldsymbol{\xi})^{\lambda}\right]. \qquad (3.36)\end{aligned}$$

The log-likelihood function can be maximized either directly by using the SAS (PROC NLMIXED) or the Ox program (sub-routine MaxBFGS) (see Doornik, 2007) or by solving the nonlinear likelihood equations obtained by differentiating (3.36). The components of the score function $U_n(\Theta) = (\partial\ell_n/\partial p, \partial\ell_n/\partial\alpha, \partial\ell_n/\partial\lambda, \partial\ell_n/\partial\boldsymbol{\xi})^\top$ are

$$\frac{\partial\ell_n}{\partial p} = (\alpha+1)\sum_{i=1}^{n}\frac{G(x;\boldsymbol{\xi})^{\lambda}}{1 - p\,G(x;\boldsymbol{\xi})^{\lambda}} - \frac{n}{1-p},$$
$$\frac{\partial\ell_n}{\partial\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\log\left[1 - G(x;\boldsymbol{\xi})^{\lambda}\right] - \sum_{i=1}^{n}\log\left[1 - p\,G(x;\boldsymbol{\xi})^{\lambda}\right],$$
$$\frac{\partial\ell_n}{\partial\lambda} = \frac{n}{\lambda} + \sum_{i=1}^{n}\log\left[G(x;\boldsymbol{\xi})\right] - (\alpha-1)\sum_{i=1}^{n}\frac{G(x;\boldsymbol{\xi})^{\lambda}\log\left[G(x;\boldsymbol{\xi})\right]}{1 - G(x;\boldsymbol{\xi})^{\lambda}} + p\,(\alpha+1)\sum_{i=1}^{n}\frac{G(x;\boldsymbol{\xi})^{\lambda}\log\left[G(x;\boldsymbol{\xi})\right]}{1 - p\,G(x;\boldsymbol{\xi})^{\lambda}}$$

and

$$\begin{aligned}\frac{\partial\ell_n}{\partial\boldsymbol{\xi}} &= p\,\lambda\,(\alpha+1)\sum_{i=1}^{n}\frac{G(x;\boldsymbol{\xi})^{\lambda-1}}{\left[1 - p\,G(x;\boldsymbol{\xi})^{\lambda}\right]}G^{(\xi)}(x;\boldsymbol{\xi}) - \lambda(\alpha-1)\sum_{i=1}^{n}\frac{G(x;\boldsymbol{\xi})^{\lambda-1}}{\left[1 - G(x;\boldsymbol{\xi})^{\lambda}\right]}G^{(\xi)}(x;\boldsymbol{\xi}) \\ &+ \sum_{i=1}^{n}\frac{g^{(\xi)}(x;\boldsymbol{\xi})}{g(x;\boldsymbol{\xi})} + (\lambda-1)\sum_{i=1}^{n}\frac{G^{(\xi)}(x;\boldsymbol{\xi})}{G(x;\boldsymbol{\xi})},\end{aligned}$$

where $h^{(\xi)}(\cdot)$ means the derivative of the function h with respect to $\boldsymbol{\xi}$. For interval estimation on the model parameters, we require the observed information matrix

$$J_n(\Theta) = -\begin{pmatrix} U_{pp} & U_{p\alpha} & U_{p\lambda} & | & U_{p\boldsymbol{\xi}}^{\top} \\ U_{\alpha p} & U_{\alpha\alpha} & U_{\alpha\lambda} & | & U_{\alpha\boldsymbol{\xi}}^{\top} \\ U_{\lambda p} & U_{\lambda\alpha} & U_{\lambda\lambda} & | & U_{\lambda\boldsymbol{\xi}}^{\top} \\ -- & -- & -- & -- & -- \\ U_{\boldsymbol{\xi}p} & U_{\boldsymbol{\xi}\alpha} & U_{\boldsymbol{\xi}\lambda} & | & U_{\boldsymbol{\xi}\boldsymbol{\xi}} \end{pmatrix},$$

whose elements are listed in Appendix A. Let $\widehat{\Theta}$ be the MLE of Θ. Under standard regularity conditions (Cox and Hinkley, 1974) that are fulfilled for the proposed model whenever the parameters are in the interior of the parameter space, we can approximate the distribution of $\sqrt{n}(\widehat{\Theta}-\Theta)$ by the multivariate normal $N_r(0, K(\Theta)^{-1})$, where $K(\Theta) = \lim_{n\to\infty} \frac{1}{n} J_n(\Theta)$ is the unit information matrix and r is the number of parameters of the new distribution.

3.7 Empirical illustrations

In this section, we compare the fits of some special models of the ECC-G family by means of two real data sets to show the potentiality of the new family. In order to estimate the parameters of these special models, we adopt the maximum likelihood method (as discussed in Section 14). All the computations were done using the subroutine `NLMixed` of the `SAS` software.

The first data set consists of fracture toughness from the silicon nitride. The data taken from the web-site `http://www.ceramics.nist.gov/srd/summary/ftmain.htm` was already studied by Nadarajah and Kotz (2007). The ECC-G model used in the first application is defined by equation (3.10) with $\boldsymbol{\theta_1} = (\alpha, \beta, \lambda, c, p)$. Further, the *extended Cordeiro and de Castro-exponential* (ECCE) density function is given by

$$f_2(x;\boldsymbol{\theta_2}) = \alpha\,\beta\,\lambda\,(1-p)\exp\left(-\beta x\right)\left[1-\exp\left(-\beta x\right)\right]^{\lambda-1}\frac{\left\{1-\left[1-\exp(-\beta x)\right]^{\lambda}\right\}^{\alpha-1}}{\left\{1-p\,\left[1-\exp\left(-\beta x\right)\right]^{\lambda}\right\}^{\alpha+1}}, \quad x>0,$$

where $\boldsymbol{\theta_2} = (\alpha, \beta, \lambda, p)$. These ECC-G models are compared with the Kumaraswamy Weibull (KwW) and beta Weibull (BW) models with corresponding densities

$$f_3(x;\boldsymbol{\theta_3}) = \alpha\,\lambda\,c\,\beta^c\,x^{c-1}\,\mathrm{e}^{-(\beta x)^c}\left[1-\mathrm{e}^{-(\beta x)^c}\right]^{\lambda-1}\left\{1-\left[1-\mathrm{e}^{-(\beta x)^c}\right]^{\lambda}\right\}^{\alpha-1}, \quad x>0$$

and

$$f_4(x;\boldsymbol{\theta_4}) = \frac{c\,\lambda^c}{B(a,b)}\,x^{c-1}\exp\left[-b(\lambda x)^c\right]\left[1-\mathrm{e}^{-(\lambda x)^c}\right]^{a-1}, \quad x>0,$$

where $\boldsymbol{\theta_3} = (\alpha, \beta, \lambda, c)$ and $\boldsymbol{\theta_4} = (a, b, c, \lambda)$.

As a second application, we consider a real data set on the strengths of 1.5 cm glass fibres, measured at the National Physical Laboratory, England, see Smith and Naylor (1987). We fit the ECCW and ECCE models to these data. These models are compared with the BW model and beta Birnbaum-Saunders (BBS) model defined by

$$f_5(x;\boldsymbol{\theta_5}) = \frac{\kappa(\alpha,\beta)}{B(a,b)}\, x^{-3/2}(x+\beta)\exp\left[-\tau(x/\beta)/(2\alpha^2)\right]\Phi(\nu)^2\left[1-\Phi(\nu)\right]^{b-1}, \quad x>0,$$

where $\boldsymbol{\theta_5} = (\alpha,\beta,a,b)$, $\nu = \alpha^{-1}\rho(x/\beta)$, $\rho(z) = z^{1/2} + z^{-1/2}$, $\kappa(\alpha,\beta) = \exp(\alpha^{-2})/(2\alpha\sqrt{2\pi\beta})$, $\tau(z) = z + z^{-1}$ and $\Phi(\cdot)$ is the standard normal cumulative function.

Table 3.2: Estimates ([a] denotes standard errors) and K-S statistics.

Data set	Distribution	Estimates	K-S
1 ($n = 119$)	ECCW	$\widehat{\boldsymbol{\theta}}_1 = (1.9217, 0.5422, 4.3091, 1.5526, 0.9589)$	0.0537
		$(3.7100, 1.3219, 17.9943, 2.6544, 0.1602)^a$	
	ECCE	$\widehat{\boldsymbol{\theta}}_2 = (3.0949, 1.2853, 16.8832, 0.9848)$	0.0583
		$(2.2243, 0.2272, 10.6914, 0.0137)^a$	
	KwW	$\widehat{\boldsymbol{\theta}}_3 = (7.0242, 0.1450, 0.8329, 5.8042)$	0.0763
		$(2.4542, 0.1655, 0.5797, 3.4765)^a$	
	BW	$\widehat{\boldsymbol{\theta}}_4 = (5.6663, 0.1634, 0.8054, 3.4077)$	0.0697
		$(0.5568, 0.3708, 0.0067, 3.5823)^a$	
2 ($n = 51$)	ECCW	$\widehat{\boldsymbol{\theta}}_1 = (0.9367, 0.8276, 0.6341, 3.6621, 0.9420)$	0.2339
		$(0.7763, 0.2411, 0.7032, 2.2503, 0.1053)^a$	
	ECCE	$\widehat{\boldsymbol{\theta}}_2 = (3.6519, 4.5125, 11.1218, 0.9978)$	0.2390
		$(2.7515, 0.7429, 22.1433, 0.0049)^a$	
	BW	$\widehat{\boldsymbol{\theta}}_4 = (7.0127, 0.9199, 0.4493, 0.0496)$	0.2508
		$(0.1867, 0.0484, 0.8872, 0.1522)^a$	
	BBS	$\widehat{\boldsymbol{\theta}}_5 = (0.3638, 7857.5658, 1.0505, 30.4783)$	0.2432
		$(0.1517, 2558.5670, 0.2506, 18.1233)^a$	

The MLEs of the parameters, their standards errors, and Kolmogorov-Smirnov (K-S) statistics are given in Table 3.2. From the values of K-S statistics, we conclude that the ECCW model provides a better fit to the first data set, since it yields the lowest value.

We can also perform formal goodness-of-fit tests in order to verify which distribution fits better to these data. We apply the Cramér-von Mises (W^*) and Anderson-Darling (A^*) tests. The W^* and A^* test statistics are described in details in Chen and Balakrishnan (1995). In general, the smaller the values of W^* and A^*, the better the fit to the data. Table 3.3 gives the values of the W^* and A^* statistics for the first and second data sets. According to these statistics, the ECCW model fits the first data set better than the others competing models.

The figures in Table 3.2 for the second data set indicate that the ECCW model is a very competitive model to the other fitted models to these data, although it does not give the smallest AIC. However, the smallest values of the W^* and A^* statistics in Table 3.3 indicate that the ECCW model provides a more adequate fit to these data than the other distributions. Overall, these results illustrate the potentiality of the ECCW model for lifetime data and the importance of its additional parameters.

Table 3.3: Goodness-of-fit tests statistics.

Data set	Model	Statistics	
		W^*	A^*
1	ECCW	0.0469	0.3116
	ECCE	0.0577	0.5629
	KwW	0.0784	0.6064
	BW	0.1967	1.3748
2	ECCW	0.1824	1.1636
	ECCE	0.2350	1.3730
	BW	0.2390	1.3750
	BBS	0.3651	1.9727

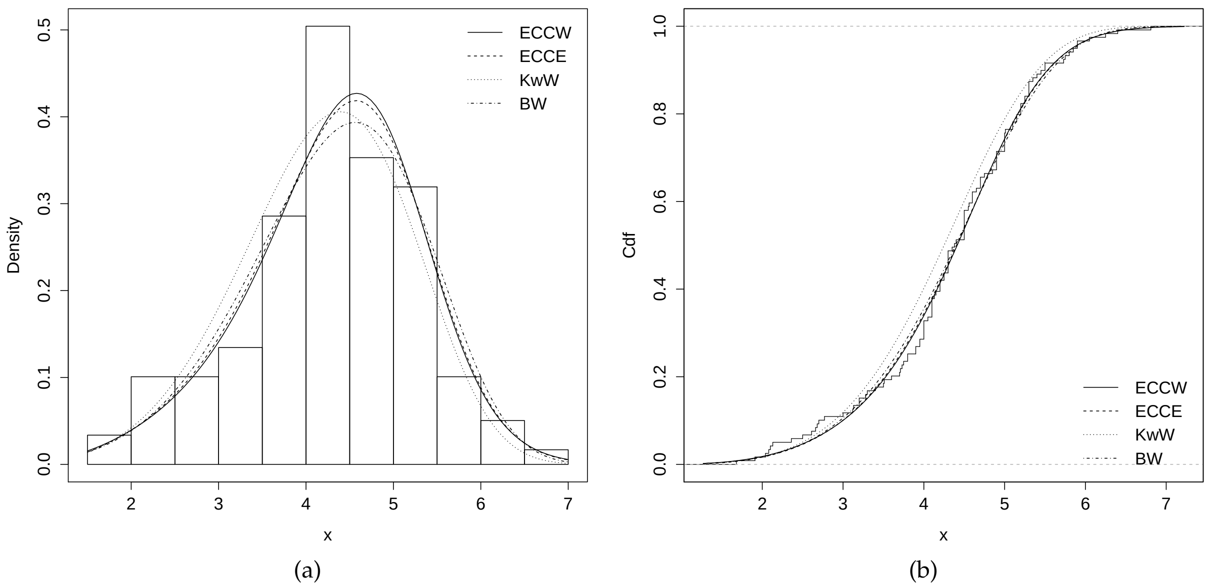
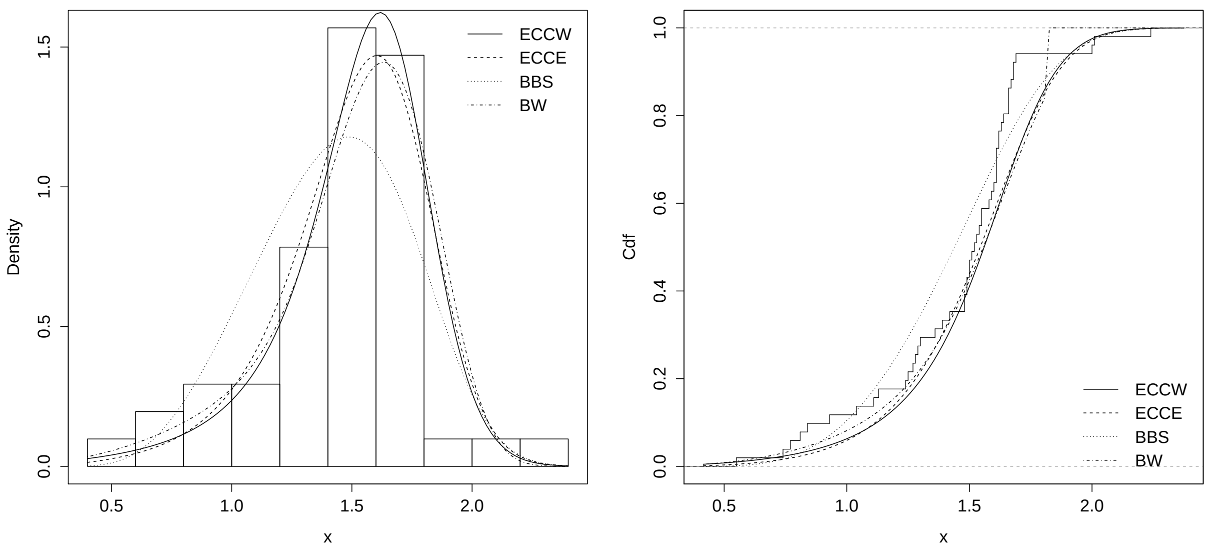


Figure 3.9: Estimated (a) pdf and (b) cdf for the ECCW, ECCE, KwW and BW models for the first data set.



(a) (b)

Figure 3.10: Estimated (a) pdf and (b) cdf for the ECCW, ECCE, BBS and BW models for the second data set.

3.8 Concluding remarks

We define a new family of distributions, called the *extended Cordeiro and de Castro* (ECC-G) family of distributions, which generalizes several well known distributions in the statistical literature such as the normal, Weibull and beta distributions by adding three shape parameters. We provide a mathematical treatment of the new family including expansions for the density function, moments, generating function and incomplete moments. The ECC-G density function can be expressed as a mixture of exponentiated density functions. This property is important to obtain several other results. We derive a power series for the quantile function of this family. Our formulas related with the ECC-G model are manageable, and with the use of modern computer resources with analytic and numerical capabilities, they may turn into adequate tools comprising the arsenal of applied statisticians. Some special models are studied in some detail. The estimation of the model parameters is approached by the method of maximum likelihood. The observed information matrix is derived. Finally, we fit the ECC-G models to two real data sets to demonstrate the potentiality of the proposed family.

Appendix A

Observed information matrix

The elements of the $r \times r$ observed information matrix $J_n(\Theta)$ are

$$U_{pp} = (\alpha+1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{2\lambda}}{[1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}]^2} - \frac{n}{(1-p)^2}, \quad U_{p\alpha} = \sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda}}{1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}},$$

$$U_{p\lambda} = (\alpha+1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda}\log G(x_i;\boldsymbol{\xi})}{[1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}]^2}, \quad U_{p\xi} = \lambda(\alpha+1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-1}\,G^{(\xi)}(x;\boldsymbol{\xi})}{[1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}]^2},$$

$$U_{\alpha\alpha} = -\frac{n}{\alpha^2}, \quad U_{\alpha\lambda} = p\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda}\log G(x_i;\boldsymbol{\xi})}{1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}} - p\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda}\log G(x_i;\boldsymbol{\xi})}{1-G(x_i;\boldsymbol{\xi})^{\lambda}},$$

$$U_{\alpha\boldsymbol{\xi}} = \lambda\,p\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-1}\,G^{(\xi)}(x;\boldsymbol{\xi})}{1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}} - \lambda\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-1}\,G^{(\xi)}(x;\boldsymbol{\xi})}{1-G(x_i;\boldsymbol{\xi})^{\lambda}},$$

$$\begin{aligned}U_{\lambda\lambda} &= -\frac{n}{\lambda^2} - \lambda(\alpha-1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-1}\,G^{(\xi)}(x;\boldsymbol{\xi})}{[1-G(x_i;\boldsymbol{\xi})^{\lambda}]^2}\log G(x_i;\boldsymbol{\xi}) \\ &\quad + \lambda\,p\,(\alpha+1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-1}\,G^{(\xi)}(x;\boldsymbol{\xi})}{[1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}]^2}\log G(x_i;\boldsymbol{\xi}),\end{aligned}$$

$$\begin{aligned}U_{\lambda\boldsymbol{\xi}} &= \sum_{i=1}^{n}\frac{G^{(\xi)}(x;\boldsymbol{\xi})}{G(x_i;\boldsymbol{\xi})} - (\alpha-1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-1}\,G^{(\xi)}(x;\boldsymbol{\xi})\,\left[1-G(x_i;\boldsymbol{\xi})^{\lambda}+\log G(x_i;\boldsymbol{\xi})^{\lambda}\right]}{[1-G(x_i;\boldsymbol{\xi})^{\lambda}]^2} \\ &\quad + p\,(\alpha+1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-1}\,G^{(\xi)}(x;\boldsymbol{\xi})\,\left[1-G(x_i;\boldsymbol{\xi})^{\lambda}+\log G(x_i;\boldsymbol{\xi})^{\lambda}\right]}{[1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}]^2}\end{aligned}$$

and

$$\begin{aligned}U_{\boldsymbol{\xi}\boldsymbol{\xi}} &= (\lambda-1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})\,G^{(2\xi)}(x;\boldsymbol{\xi})}{G^{(\xi)}(x_i;\boldsymbol{\xi})^2} + \sum_{i=1}^{n}\frac{g(x_i;\boldsymbol{\xi})\,g^{(2\xi)}(x;\boldsymbol{\xi})}{g^{(\xi)}(x_i;\boldsymbol{\xi})^2} - n\lambda \\ &\quad + \lambda\,(\alpha-1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-2}\,G^{(\xi)}(x;\boldsymbol{\xi})^2}{[1-G(x_i;\boldsymbol{\xi})^{\lambda}]^2} - \sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-2}\,G^{(\xi)}(x;\boldsymbol{\xi})^2}{1-G(x_i;\boldsymbol{\xi})^{\lambda}}\left(\frac{G(x_i;\boldsymbol{\xi})}{G^{(\xi)}(x;\boldsymbol{\xi})}\right)^{(\xi)} \\ &\quad - \lambda\,p\,(\alpha+1)\sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-2}\,G^{(\xi)}(x;\boldsymbol{\xi})^2}{[1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}]^2} - \sum_{i=1}^{n}\frac{G(x_i;\boldsymbol{\xi})^{\lambda-2}\,G^{(\xi)}(x;\boldsymbol{\xi})^2}{1-p\,G(x_i;\boldsymbol{\xi})^{\lambda}}\left(\frac{G(x_i;\boldsymbol{\xi})}{G^{(\xi)}(x;\boldsymbol{\xi})}\right)^{(\xi)},\end{aligned}$$

where $h^{(2\xi)}(\cdot)$ denotes the second derivative of the function h with respect to ξ.

References

[1] Barreto-Souza, W., Lemonte, A.J. (2012). Bivariate Kumaraswamy distribution: properties and a new method to generate bivariate classes. *Statistics: A Journal of Theoretical and Applied Statistics*. http://www.tandfonline.com/doi/abs/10.1080/02331888.2012.694446.

[2] Bourguignon, M., Silva, R.B., Zea, L.M., Cordeiro, G.M. (2013). The Kumaraswamy Pareto distribution. *Journal of Statistical Theory and Applications*, **12**, 1–21.

[3] Cordeiro G. M., de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, **81**, 883–898.

[4] Cox, D.R., Hinkley, D.V. (1974). Theoretical Statistics. Chapman and Hall, London.

[5] Doornik, J. (2007). Ox 5: object-oriented matrix programming language, fifth ed. Timberlake Consultants, London.

[6] Eugene, N., Lee, C., Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics, Theory and Methods*, **31**, 497–512.

[7] Gomes, A.E., da-Silva, C.Q., Cordeiro, G.M., Ortega, E.M.M. (2012). A new lifetime model: the Kumaraswamy generalized Rayleigh distribution. *Journal of Statistical Computation and Simulation*. To appear.

[8] Gradshteyn, I.S., Ryzhik, I.M. (2000). Table of integrals, series, and products. Academic Press, San Diego.

[9] Gupta, R.C., Gupta, R.D. (2007). Proportional reversed hazard rate model and its applications. *Journal of Statistical Planning and Inference*, **137**, 3525–3536.

[10] Gupta, R.D., Kundu, D. (2001). Exponentiated exponential family: an alternative to gamma and Weibull. *Biometrical Journal*, **43**, 117–130.

[11] Jones, M.C. (2004). Families of distributions arising from distributions of order statistics. *TEST: A Journal of the Spanish Statistical Society*, **13**, 1–43.

[12] Leadbetter, M.R., Lindgren, G., Rootzén, H. (1987). Extremes and related properties of random sequences and processes. Springer Verlag, New York.

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

[14] Mudholkar, G.S., Srivastava, D.K., Kollia, G.D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, **91**, 1575–1583.

[15] Nadarajah, S., Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae*, **92**, 97–111.

[16] Nadarajah, S., Kotz, S. (2007). On the alternative to the Weibull function. *Engeneering Fracture Mechanics*, **74** , 451–456.

[17] Paranaíba, P.F., Ortega, E.M.M., Cordeiro, G.M., Pascoa, M.A.R. (2012). The Kumaraswamy Burr XII distribution: theory and practice. *Journal of Statistical Computation and Simulation*, **82**, 1–27.

[18] Pascoa, M.A.R.,; Ortega, E.M.M., Cordeiro, G.M. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical Methodology*, **8**, 411–433.

[19] Prudnikov, A.P., Brychkov, Y.A., Marichev, O.I. (1986). Integrals and series. Gordon and Breach Science Publishers, Amsterdam.

[20] Rényi, A. (1961). On measures of information and entropy. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, 547–561.

[21] Ristíc, M.M., Balakrishnan, N. (2011). The gamma-exponentiated exponential distribution. *Journal Statistical Computation and Simulation*, **82**, 1191–1206.

[22] Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–432.

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

[24] Smith, R. L., Naylor, J. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics*, **36**, 358–369.

[25] Wright, E.M. (1935). The asymptotic expansion of the generalized hypergeometric function. *Proceedings of the London Mathematical Society*, **10**, 286–293.

[26] Zografos, K., Balakrishnan, N. (2009). On families of beta and generalized gamma generated distributions and associated inference. *Statistical Methodology*, **6**, 344–362.

CHAPTER 4

A likelihood ratio test to discriminate between the exponential-Poisson and gamma distributions

Artigo aceito para publicação no periódico *Journal of Statistical Computation and Simulation*.

Resumo

A distribuição exponencial-Poisson (EP), com parâmetros de escala e de forma $\beta > 0$ e $\lambda \in \mathbb{R}$, respectivamente, é uma distribuição de tempo de vida obtida através da composição dos modelos exponencial e Poisson truncado (em zero). A distribuição EP tem sido uma boa alternativa para a distribuição gama para a modelagem de tempo de vida, confiabilidade e intervalos de tempo entre sucessivos desastres naturais. Ambas as distribuições EP e gama apresentam algumas similaridades e propriedades em comum. Por exemplo, suas densidades podem ser estritamente decrescentes ou unimodais e suas funções de risco podem ser decrescentes, crescentes ou constantes, dependendo dos valores de seus parâmetros de forma. Por outro lado, a distribuição EP apresenta diversas aplicações interessantes baseadas em representações estocásticas envolvendo o máximo e o mínimo de variáveis aleatórias iid (com tamanho de amostra aleatório), que a tornam de importância científica distinguível da distribuição gama. Dadas as similaridades e importância científica diferentes, uma questão de interesse é como discriminá-las. Com isto em mente, neste capítulo, propomos um teste da razão de verossimilhanças baseado na estatística de Cox para discriminar as distribuições EP e gama. A distribuição assintótica do logaritmo normalizado da razão das verossimilhanças maximizadas sob as duas hipóteses – os dados vêm de uma distribuição EP | gama – são obtidos. Além disso, determinamos o tamanho mínimo de amostra necessário para discriminar os dois modelos quando a probabilidade de seleção correta e um dado nível de tolerância são previamente estabelecidos. Apresentamos um estudo de simulação para avaliar a precisão das probabilidades assintóticas de seleção correta. O trabalho é motivado por duas aplicações para conjuntos de dados reais.

Palavras-chave: Distribuição assintótica; Distribuição de tempo de vida; Estatística de Cox; Probabilidade de seleção correta.

Abstract

The exponential-Poisson (EP) distribution, with scale and shape parameters $\beta > 0$ and $\lambda \in \mathbb{R}$, respectively, is a lifetime distribution which is obtained by compounding exponential and zero truncated Poisson models. The EP distribution has been a good alternative to the gamma distribution for modeling lifetime, reliability and time intervals of successive natural disasters. Both EP and gamma distributions present some similarities and properties in common, for example, their densities may be strictly decreasing or unimodal, and their hazard functions may be decreasing, increasing or constant depending on their shape parameters. On the other hand, the EP distribution has several interesting applications based on stochastic representations involving maximum and minimum of iid exponential variables (with random sample size) which make it of distinguishable scientific importance from the gamma distribution. Given the similarities and different scientific relevance between these models, one question of interest is how to discriminate them. With this in mind, in this paper we propose a likelihood ratio test based on Cox's statistic to discriminate the exponential-Poisson and gamma distributions. The asymptotic distributions of the normalized logarithm of the ratio of the maximized likelihoods under two null hypotheses – data come from exponential-Poisson or gamma distributions – are provided. We also determinate the minimum sample size required to discriminate the EP and gamma distributions when the probability of correct selection and a given tolerance level based on some distance are previously stated. A simulation study to evaluate the accuracy of the asymptotic probabilities of correct selection is also presented. The paper is motivated by two applications to real data sets.

Keywords: Asymptotic distribution, Cox's statistic, Lifetime distribution, Probability of correct selection.

4.1 Introduction

Recently, Kus (2007) introduced and studied the two-parameter exponential-Poisson (EP) distribution, which is obtained by compounding the exponential and zero truncated Poisson distributions, where the compounding procedure follows the same one carried out by Marshall and Olkin (1997). More specifically, if $\{X_i\}_{i=1}^{\infty}$ is a sequence of independent and identically distributed (iid) exponential random variables (with mean $1/\theta$) independent of a discrete random variable N following a zero truncated Poisson distribution with the probability mass function

$$P(N = n) = \{\exp(\lambda) - 1\}^{-1} \frac{\lambda^n}{n!}, \quad n = 1, 2, \ldots,$$

where $\lambda > 0$, then

$$X = \min(X_1, \ldots, X_N) \tag{4.1}$$

has exponential-Poisson distribution and its probability density function (pdf) is

$$f_{\text{EP}}(x; \theta, \lambda) = \frac{\theta\lambda}{\text{e}^{\lambda} - 1} \exp\{-\theta x + \lambda \text{e}^{-\theta x}\}, \quad x > 0, \tag{4.2}$$

with corresponding hazard rate function

$$\tau_{\text{EP}}(x; \theta, \lambda) = \frac{\theta\lambda \left(1 - \text{e}^{\lambda}\right) \exp(-\lambda - \theta x + \lambda \text{e}^{-\theta x})}{\left(1 - \text{e}^{-\lambda}\right) \left(1 - \exp(\lambda \text{e}^{-\theta x})\right)}, \quad x > 0.$$

Barreto-Souza and Simas (2013) introduced and studied the exp-G distribution, which has its pdf given by

$$p(x) = \frac{\gamma}{1 - \text{e}^{-\gamma}} g(x) \exp\{-\gamma G(x)\}, \quad \gamma \neq 0, \quad x \in \mathbb{R}, \tag{4.3}$$

and $p(x) = g(x)$ when γ approaches 0, where $G(\cdot)$ is the cumulative distribution function (cdf) of a continuous random variable and $g(\cdot)$ is the derivative of $G(\cdot)$. The EP distribution may be obtained through the exp-G class by taking $G(x) = 1 - \exp(-\theta x)$, for $x > 0$, in (4.3). So, the EP distribution can be generalized by extending the parameter space with respect to λ, that is, by taking $\lambda \in \mathbb{R}$. The exp-G model may be seen as a particular proper distribution in the cure rate models by Yakovlev and Tsodikov (1996) and Cooner et al. (2007). Therefore it has an interesting representation based on maximum and minimum of iid random variables with random sample size following the zero-truncated Poisson distribution. In particular, if $\lambda < 0$, N is a discrete random variable with zero truncated Poisson distribution with parameter $-\lambda$ and the sequence $\{X_i\}_{i=1}^{\infty}$ is defined as before, we have that

$$X = \max(X_1, \ldots, X_N) \tag{4.4}$$

has pdf given in (4.2). From now on we consider the EP distribution with density (4.2) and $\lambda \in \mathbb{R}$. The EP distribution is also a particular case of the exponential power series (Chahkandi and Ganjali, 2009), Weibull power series (Morais and Barreto-Souza, 2011), generalized exponential-Poisson (Barreto-Souza and Cribari, 2009) and exponentiated exponential-Poisson (Ristic and Nadarajah, 2012) distributions.

Karlis (2009) introduced a nested EM algorithm to find the maximum likelihood estimates of the parameters of the EP distribution. Mutairi et al. (2011) proposed the EP model for estimating reliability in a series system with random sample size. Kus (2007) showed the usefulness of the EP distribution for modeling the time of successive failures for air conditioning system of Boeing 720 jet airplanes, the period between successive coal-mining disasters and the time intervals of successive earthquakes. Furthermore, the EP distribution provides a better fit than well-known lifetime models, including gamma distribution, based on Kolmogorov-Smirnov test.

The EP and gamma distributions are both generalizations of the exponential distribution but in different ways. Both models are used quite effectively when data are coming from a right tailed distribution. Further, these models present some similarities and properties in common. For example, their densities may be strictly decreasing or unimodal and their hazard functions may be decreasing, increasing or constant depending on their shape parameters.

On the other hand, the EP distribution has several interesting applications based on the stochastic representations (4.1) and (4.4) which make it of distinguishable scientific importance from the gamma distribution. Below we list some of these interesting applications.

- *Time to the first failure (Adamidis and Loukas (1998) and Kus (2007)).* Suppose the failure of a device occurs due to the presence of an unknown number N of initial defects of same kind, which can be identifiable only after causing failure and are repaired perfectly. Denote by X_i the time to the failure of the device due to the ith defect, for $i \geq 1$. Under the assumptions that the X_i's are iid exponential variables independent of N, which follows a truncated-zero Poisson distribution, we obtain that the EP distribution is appropriate for modeling the time to the first failure.

- *Reliability.* From stochastic representations (4.1) and (4.4), we have that the EP model can emerges in series system with identical components (for $\lambda > 0$) and parallel system with identical components (for $\lambda < 0$), which appears in many industrial applications and biological organisms.

- *Time to relapse of cancer under the first-activation scheme (Chen et al., 1999).* Suppose that an individual in the population is susceptible to a certain type of cancer. Let N be the number of carcinogenic cells for that individual left active after the initial treatment and denote by X_i the time spent for the ith carcinogenic cell to produce a detectable cancer mass, for $i \geq 1$. Under the assumptions that $\{X_i\}_{i\geq 1}$ is a sequence of iid exponential variables independent of N, which follows a Poisson distribution (truncated at zero), we have that the time to relapse of cancer of a susceptible individual, given by the random variable $\min\{X_1, \ldots, X_N\}$, follows the EP distribution.

- *Last-activation scheme (Cooner et al., 2007).* As discussed by Cooner et al. (2007), the first-activation scheme may be questioned by certain diseases. Let N be the number of latent factors that must all be actived by failure and X_i be the time of resistance to a disease manifestation due to the ith latent factor. In the last-activation scheme it is assumed that failure occurs after all N factors have been actived. So, if the X_i's are iid exponential variables independent of N, and N follows a zero-truncated Poisson distribution, the EP distribution is able for modeling the time to the failure under last-activation scheme.

Given the similarities between these models and the fact that EP model has important characteristics which make it of perceptible scientific importance, one question of interest is how to discriminate them.

With this mind, our aim in this paper is to propose a likelihood ratio test based on Cox's statistic to discriminate the EP and gamma distributions. The pdf associated to the gamma distribution we consider here is

$$f_{\mathrm{GA}}(x;\beta,\alpha) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1}\exp(-\beta x), \quad x > 0, \tag{4.5}$$

where $\beta > 0$ and $\alpha > 0$. Note that for certain ranges of the parameters, shapes of the EP and GA density functions are quite similar. See, for example, in Figure 4.1(a) the densities of $\mathrm{EP}(0.75, -5.2)$ and $\mathrm{GA}(1, 3)$ and in Figure 4.1(b) the densities of $\mathrm{EP}(1, 2)$ and $\mathrm{GA}(1.42, 0.82)$, where they are almost indistinguishable from a practical point of view.

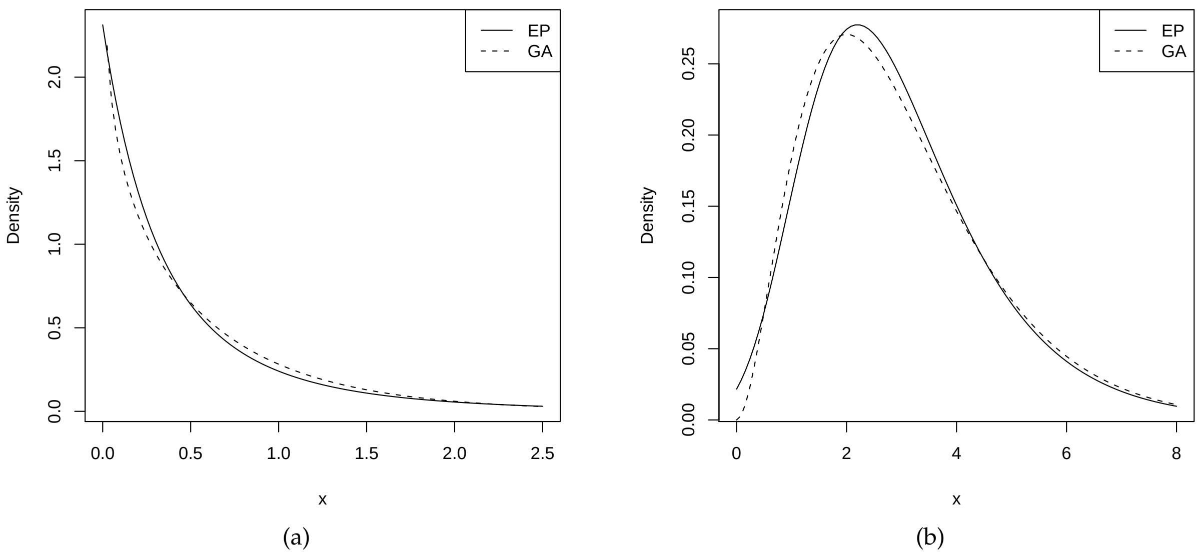


Figure 4.1: The density functions of (a) $\mathrm{EP}(0.75, -5.2)$ and $\mathrm{GA}(1, 3)$ and (b) $\mathrm{EP}(1, 2)$ and $\mathrm{GA}(1.42, 0.82)$.

For lifetime and reliability studies the tail of the assumed distribution plays an important role, therefore to choose the more adequate model is crucial. The idea of discriminating two separate distributions was originally proposed in pioneering works of Cox (1961) and Cox (1962). The interest in discriminating two distributions is an old and well known problem in the statistical literature. A special attention is given in discrimination between lifetime distributions due to increasing applications of such distributions. For example, see the works of Dumonceaux and Antle (1973), Dumonceaux et al. (1973), Quesenberry and Kent (1982) and Balasooriya and Abeysinghe (1994). Dumonceaux and Antle (1973) proposed a procedure to discriminate the log-normal and Weibull distributions, and the discrimination between the log-normal and gamma distributions was studied by Wiens (1999). The discrimination between the gamma and Weibull distributions was studied by Bain and Engelhardt (1980) and Fearn and Nebenzahl (1991). Some papers proposed discrimination tests involving the generalized exponential distribution and other well known lifetime distributions; for instance, see

Gupta and Kundu (2003), Gupta and Kundu (2004) and Kundu et al. (2005). Selection between the Weibull and log-normal distributions under Type-II censoring was considered by Dey and Kundu (2012). References about discrimination of separate families of hypotheses are widespread and we recommend the reader to see references contained in the above papers.

General regularity conditions and a rigorous proof of the asymptotic normality of the Cox's statistic for testing separate families of hypotheses was firstly given by White (1982a) and White (1982b). We here obtain the asymptotic distribution of the normalized logarithm of the ratio of the maximized likelihoods based on these papers. We will use these asymptotic results to compute the probability of correct selection.

The present chapter is organized in the following way. In Section 4.2, we define the likelihood ratio test to discriminate the exponential-Poisson and gamma models and obtain asymptotic distributions of the normalized logarithms of the ratio of the maximized likelihoods under the two null hypotheses: data come from exponential-Poisson or gamma distributions. The minimum sample size required to discriminate the EP and gamma distributions when the probability of correct selection and a tolerance level (based on the Kolmogorov-Smirnov or Hellinger distances) are previously handled is provided in Section 4.3. Further, we obtain an expression for the Hellinger distance between the EP and gamma distributions. A simulation study about the accuracy of the asymptotic probability of correct selection is presented in Section 4.4. Two applications to real data sets are presented in Section 4.5. Proof of the results stated in the chapter are presented in Section 4.6. Concluding remarks and future research are addressed in Section 4.7.

4.2 Test statistic and asymptotic distributions

Let $\{X_i\}_{i=1}^n$ be a sequence of iid random variables with observed values $x_1, \ldots, x_n$, where the data come from $\text{EP}(\theta, \lambda)$ distribution or $\text{GA}(\beta, \alpha)$ distribution with densities given by (4.2) and (4.5), respectively. We are interested in discriminating the EP and GA models and obtain the asymptotic distributions of the normalized test statistic under the hypotheses H_{EP} : $\{X_i\}_{i=1}^n \sim \text{EP}(\theta, \lambda)$ and $H_{\text{GA}} : \{X_i\}_{i=1}^n \sim \text{GA}(\beta, \alpha)$.

The log-likelihood function associated to the EP distribution is given by

$$\ell_{\text{EP}}^{(n)}(\theta, \lambda) = n \log \theta + n \log \left(\frac{\lambda}{\mathrm{e}^{\lambda} - 1} \right) - \theta n \bar{x}_n + \lambda \sum_{i=1}^{n} \exp(-\theta x_i),$$

where $\bar{x}_n = \sum_{i=1}^n x_i / n$. The maximum likelihood estimates (MLEs) $\widehat{\theta}_n$ and $\widehat{\lambda}_n$ of θ and λ, respectively, are obtained by solving the nonlinear system of equations

$$\widehat{\lambda}_n = \frac{n(1/\widehat{\theta}_n - \bar{x}_n)}{\sum_{i=1}^n x_i \exp(-\widehat{\theta}_n x_i)} \quad \text{and} \quad n \left(\frac{1}{\widehat{\lambda}_n} - \frac{1}{1 - e^{-\widehat{\lambda}_n}} \right) + \sum_{i=1}^{n} \exp(-\widehat{\theta}_n x_i) = 0. \tag{4.6}$$

On the other hand, the log-likelihood corresponding to the gamma distribution is

$$\ell_{\text{GA}}^{(n)}(\beta, \alpha) = n\{\alpha \log \beta - \log \Gamma(\alpha)\} - n\beta \bar{x}_n + (\alpha - 1) \sum_{i=1}^{n} \log x_i.$$

The MLEs $\widehat{\beta}_n$ and $\widehat{\alpha}_n$ of β and α, respectively, are given as solution of the nonlinear system of equations

$$\widehat{\beta} = \frac{\widehat{\alpha}_n}{\bar{x}_n} \quad \text{and} \quad n \log \widehat{\beta}_n - n\Psi(\widehat{\alpha}_n) + \sum_{i=1}^{n} \log x_i = 0, \tag{4.7}$$

where $\Psi(z) = d \log \Gamma(z)/dz = \Gamma'(z)/\Gamma(z)$. The logarithm of the ratio of the maximized likelihoods is given by

$$T_n = \log \left(\frac{\prod_{i=1}^{n} f_{\mathrm{EP}}(x_i; \widehat{\theta}_n, \widehat{\lambda}_n)}{\prod_{i=1}^{n} f_{\mathrm{GA}}(x_i; \widehat{\alpha}_n, \widehat{\beta}_n)} \right) = \ell_{\mathrm{EP}}^{(n)}(\widehat{\theta}_n, \widehat{\lambda}_n) - \ell_{\mathrm{GA}}^{(n)}(\widehat{\beta}_n, \widehat{\alpha}_n),$$

where $(\widehat{\theta}_n, \widehat{\lambda}_n)$ and $(\widehat{\beta}_n, \widehat{\alpha}_n)$ are given in (4.6) and (4.7), respectively. More explicitly, the statistic T_n can be expressed as

$$T_n = n \left\{ \log \left(\frac{\Gamma(\widehat{\alpha}_n)}{\widehat{\beta}_n^{\widehat{\alpha}_n}} \right) + \log \left(\frac{\widehat{\theta}_n \, \widehat{\lambda}_n}{e^{\widehat{\lambda}_n} - 1} \right) + (\widehat{\beta}_n - \widehat{\theta}_n)\bar{x}_n \right\} + \sum_{i=1}^{n} \left\{ \widehat{\lambda}_n \exp(-\widehat{\theta}_n x_i) - (\widehat{\alpha}_n - 1) \log x_i \right\}.$$

With this, the following decision rule could be used: choose the exponential-Poisson distribution if $T_n > 0$, otherwise choose the gamma distribution as the preferred model. Since that both models have the same number of parameters, this rule decision is equivalent to Akaike criteria (Akaike, 1974).

In this chapter we adopt the following rule: we choose the model that maximizes the probability of correct selection. The probability of correct selection is presented in Section 4.3. This procedure is illustrated in Section 4.5, where two applications to real data sets are presented.

We now obtain the asymptotic distributions of T_n under the hypotheses H_{EP} and H_{GA}. From now on, a.s. denotes almost sure convergence.

4.2.1 H_0 : EP distribution $\times$ H_1 : Gamma distribution

Suppose that $X_1, \ldots, X_n$ are iid random variables from $\mathrm{EP}(\theta, \lambda)$ distribution. For our purposes in this Section, we now introduce some notations. Let $h(\cdot)$ and $g(\cdot)$ be two real mensurable functions and U be a random variable following $\mathrm{EP}(\theta, \lambda)$ distribution. We denote by $\mathrm{E}_{\mathrm{EP}}(h(U))$ and $\mathrm{Var}_{\mathrm{EP}}(h(U))$ the mean and variance of $h(U)$. Further, we denote the covariance between $h(U)$ and $g(U)$ by

$$\mathrm{Cov}_{\mathrm{EP}}(h(U), g(U)) = \mathrm{E}_{\mathrm{EP}}(h(U)g(U)) - \mathrm{E}_{\mathrm{EP}}(h(U))\mathrm{E}_{\mathrm{EP}}(g(U)),$$

with U defined as before.

Lemma 5. *Under the hypothesis H_{EP}, as $n \to \infty$ we have*

(i) $\widehat{\alpha}_n \to \widetilde{\alpha}$ a.s., $\widehat{\beta}_n \to \widetilde{\beta}$ a.s., where

$$\mathrm{E}_{\mathrm{EP}} \left[\log f_{\mathrm{GA}}(X; \widetilde{\alpha}, \widetilde{\beta}) \right] = \max_{\alpha, \beta} \mathrm{E}_{\mathrm{EP}} \left[\log f_{\mathrm{GA}}(X; \alpha, \beta) \right].$$

(ii) $\widehat{\theta}_n \to \theta$ *a.s.,* $\widehat{\lambda}_n \to \lambda$ *a.s., and* θ *and* λ *satisfy*

$$\mathrm{E}_{\mathrm{EP}}\left[\log f_{\mathrm{EP}}(X;\theta,\lambda)\right] = \max_{\bar{\theta},\bar{\lambda}} \mathrm{E}_{\mathrm{EP}}\left[\log f_{\mathrm{EP}}(X;\bar{\theta},\bar{\lambda})\right].$$

Remark: We call attention of the reader that $\widetilde{\alpha}$ and $\widetilde{\beta}$ are functions of θ and λ, which is not explicit in order to simplify the notation.

We now discuss how to obtain $\widetilde{\beta}$ and $\widetilde{\alpha}$. For this, let $F_{p,q}(\mathbf{b},\mathbf{d},z)$ be the known Barnes extended hypergeometric function

$$F_{p,q}(\mathbf{b},\mathbf{d},z) = \sum_{k=0}^{\infty} \frac{z^k \prod_{i=1}^{p} \Gamma(b_i+k)\Gamma(b_i)^{-1}}{k! \prod_{i=1}^{q} \Gamma(d_i+k)\Gamma(d_i)^{-1}}, \quad z \in \mathbb{R},$$

where $\mathbf{b} = [b_1,\ldots,b_p]$ and $\mathbf{d} = [d_1,\ldots,d_q]$. Further, for $z \in \mathbb{R}$ and $i \in \{0,1\}$, define

$$\Phi(z;i) = \sum_{k=2}^{\infty} \frac{z^k (k-1)^i \log k}{k!}.$$

Hence, we have that $\omega_{\mathrm{EP}}(\beta,\alpha) \equiv \mathrm{E}_{\mathrm{EP}}\left[\log f_{\mathrm{GA}}(X;\beta,\alpha)\right]$ equals

$$\omega_{\mathrm{EP}}(\beta,\alpha) = \log\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right) - \frac{\beta\lambda}{\theta(\mathrm{e}^{\lambda}-1)} F_{2,2}([1,1],[2,2],\lambda) + (\alpha-1)\left(\Psi(1) - \log\theta - \frac{\Phi(\lambda;0)}{\mathrm{e}^{\lambda}-1}\right).$$

With this, we have that $\widetilde{\beta}$ and $\widetilde{\alpha}$ are the solutions of the nonlinear system of equations $(\partial\omega_{\mathrm{EP}}/\partial\beta, \partial\omega_{\mathrm{EP}}/\partial\alpha)^{\top} = 0$. These equations are given by

$$\widetilde{\beta} = \theta \exp\left\{\Psi(\widetilde{\alpha}) - \Psi(1) + \frac{\Phi(\lambda;0)}{\mathrm{e}^{\lambda}-1}\right\} \text{ and } \widetilde{\alpha} = \frac{\widetilde{\beta}\lambda}{\theta(\mathrm{e}^{\lambda}-1)} F_{2,2}([1,1],[2,2],\lambda). \tag{4.8}$$

Now, in order to state a theorem which gives us the asymptotic distribution of the normalized test statistic under H_{EP}, we compute the mean and variance of the random variable $\log f_{\mathrm{EP}}(X;\theta,\lambda) - \log f_{\mathrm{GA}}(X;\widetilde{\beta},\widetilde{\alpha})$ (with $X \sim \mathrm{EP}(\theta,\lambda)$), which we will be denoted by $\mathrm{AM}_{\mathrm{EP}}$ and $\mathrm{AV}_{\mathrm{EP}}$, respectively.

We observe that $\widetilde{\beta}/\theta$ is a function of $\widetilde{\alpha}$ and λ. We also have that $\widetilde{\alpha}$ only depends on λ. Hence, it can be showed that $\mathrm{AM}_{\mathrm{EP}}$ and $\mathrm{AV}_{\mathrm{EP}}$ only depend on λ and are given by

$$\begin{aligned} \mathrm{AM}_{\mathrm{EP}}(\lambda) &= \widetilde{\alpha}\log\widetilde{\beta} - \log\Gamma(\widetilde{\alpha}) - \log\theta - \log\left(\frac{\lambda}{e^{\lambda}-1}\right) + 1 - \frac{\lambda}{1-e^{-\lambda}} + \\ &\quad \lambda\frac{\widetilde{\beta}/\theta - 1}{\mathrm{e}^{\lambda}-1} F_{2,2}([1,1],[2,2],\lambda) + (\widetilde{\alpha}-1)\left(\Psi(1) - \log\theta - \frac{\Phi(\lambda;0)}{\mathrm{e}^{\lambda}-1}\right) \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} \mathrm{AV}_{\mathrm{EP}}(\lambda) &= (\widetilde{\beta}-\theta)^2\mathrm{Var}_{\mathrm{EP}}(X) + (\widetilde{\alpha}-1)^2\mathrm{Var}_{\mathrm{EP}}(\log X) + \lambda^2\mathrm{Var}_{\mathrm{EP}}(e^{-\theta X}) - \\ &\quad 2(\widetilde{\alpha}-1)(\widetilde{\beta}-\theta)\mathrm{Cov}_{\mathrm{EP}}(X,\log X) - 2\lambda(\widetilde{\alpha}-1)\mathrm{Cov}_{\mathrm{EP}}(\log X, \mathrm{e}^{-\theta X}) + \\ &\quad 2\lambda(\widetilde{\beta}-\theta)\mathrm{Cov}_{\mathrm{EP}}(X, \mathrm{e}^{-\theta X}), \end{aligned} \tag{4.10}$$

with

$$\mathrm{Var}_{\mathrm{EP}}(X) = \frac{\lambda}{\theta^2(\mathrm{e}^\lambda - 1)} \left\{ 2F_{3,3}([1,1,1],[2,2,2],\lambda) - \frac{\lambda}{\mathrm{e}^\lambda - 1} F_{2,2}^2([1,1],[2,2],\lambda) \right\},$$

$$\begin{aligned} \mathrm{Var}_{\mathrm{EP}}(\log X) &= \Psi'(1) + \frac{1}{\mathrm{e}^\lambda - 1} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \{\Psi(1) - \log(\theta k)\}^2 - \\ &\quad \frac{1}{(\mathrm{e}^\lambda - 1)^2} \left\{ (\mathrm{e}^\lambda - 1)(\Psi(1) - \log\theta) - \Phi(\lambda;0) \right\}^2, \end{aligned}$$

$$\mathrm{Var}_{\mathrm{EP}}(\mathrm{e}^{-\theta X}) = \frac{1 - e^\lambda(\lambda^2 - e^\lambda + 2)}{\lambda^2(e^\lambda - 1)^2},$$

$$\begin{aligned} \mathrm{Cov}_{\mathrm{EP}}(X, \log X) &= \frac{1}{\theta(\mathrm{e}^\lambda - 1)} \left\{ \sum_{k=1}^{\infty} \frac{\lambda^k}{k!k} (\Psi(2) - \log(\theta k)) - \right. \\ &\quad \left. \frac{\lambda}{\mathrm{e}^\lambda - 1} F_{2,2}([1,1],[2,2],\lambda) \left[(\mathrm{e}^\lambda - 1)(\Psi(1) - \log\theta) - \Phi(\lambda;0) \right] \right\}, \end{aligned}$$

$$\mathrm{Cov}_{\mathrm{EP}}(X, \mathrm{e}^{-\theta X}) = \frac{\lambda \mathrm{e}^{-\lambda}}{4\theta(1 - \mathrm{e}^{-\lambda})} F_{2,2}([2,2],[3,3],\lambda) - \left(\frac{\lambda}{1 - \mathrm{e}^{-\lambda}} - 1 \right) \frac{F_{2,2}([1,1],[2,2],\lambda)}{\theta(\mathrm{e}^\lambda - 1)},$$

and

$$\mathrm{Cov}_{\mathrm{EP}}(\log X, \mathrm{e}^{-\theta X}) = \left(\frac{1}{1 - \mathrm{e}^{-\lambda}} - \frac{1}{\lambda} \right) \frac{\Phi(\lambda;0)}{\mathrm{e}^\lambda - 1} - \frac{\Phi(\lambda;1)}{\lambda(\mathrm{e}^\lambda - 1)}.$$

In Table 4.1, we list values of $\mathrm{AM}_{\mathrm{EP}}(\lambda)$, $\mathrm{AV}_{\mathrm{EP}}(\lambda)$, $\widetilde{\alpha}$ and $\widetilde{\beta}$ for $\theta = 1$ and some values of the parameter λ. Note that as $\lambda \to 0$, $\mathrm{AM}_{\mathrm{EP}}(\lambda)$ and $\mathrm{AV}_{\mathrm{EP}}(\lambda)$ approach 0 and $\widetilde{\beta}$ and $\widetilde{\alpha}$ approach 1, what was expected since $\mathrm{E}P(1, \lambda)$ distribution approaches the exponential (with mean 1) distribution when $\lambda \to 0$.

λ	$\mathrm{AM}_{\mathrm{EP}}(\lambda)$	$\mathrm{AV}_{\mathrm{EP}}(\lambda)$	$\widetilde{\beta}$	$\widetilde{\alpha}$
−2.0	0.008514	0.019892	0.9600	1.4648
−1.0	0.002798	0.005860	0.9432	1.1887
−0.5	0.000760	0.001544	0.9612	1.0843
0.5	0.000804	0.001609	1.0625	0.9338
1.0	0.003134	0.006366	1.1518	0.8834
2.0	0.010723	0.023666	1.4246	0.8214

Table 4.1: Values of $\mathrm{AM}_{\mathrm{EP}}(\lambda)$, $\mathrm{AV}_{\mathrm{EP}}(\lambda)$, $\widetilde{\beta}$ and $\widetilde{\alpha}$ for $\theta = 1$ and some values of λ.

We now are ready to state one of our main results about asymptotic distribution of the statistic T_n. For this, define $\widetilde{T}_n^{\mathrm{EP}} = \ell_{\mathrm{EP}}^{(n)}(\theta, \lambda) - \ell_{\mathrm{GA}}^{(n)}(\widetilde{\beta}, \widetilde{\alpha})$.

Theorem 6. *Under* $H_0 : \{X_i\}_{i=1}^n \sim \text{EP}(\theta, \lambda)$*,*

$$n^{-1/2}\{T_n - \text{E}_{\text{EP}}(T_n)\} \sim n^{-1/2}\{\widetilde{T}_n^{\text{EP}} - n\text{AM}_{\text{EP}}(\lambda)\} \xrightarrow{d} N(0, \text{AV}_{\text{EP}}(\lambda)),$$

as $n \to \infty$*, where* $\text{AM}_{\text{EP}}(\lambda)$ *and* $\text{AV}_{\text{EP}}(\lambda)$ *are given by (4.9) and (4.10), respectively, and "*$\sim$*" denotes asymptotically equivalent.*

Remark: From classical Central Limit Theorem for iid random variables with finite second moment, it follows that $n^{-1/2}\{\widetilde{T}_n^{\text{EP}} - n\text{AM}_{\text{EP}}(\lambda)\} \xrightarrow{d} \text{N}(0, \text{AV}_{\text{EP}}(\lambda))$ as $n \to \infty$.

4.2.2 H_0 : Gamma distribution $\times$ H_1 : EP distribution

Suppose now $X_1, \ldots, X_n$ are iid random variables from $\text{GA}(\beta, \alpha)$ distribution. Let $h(\cdot)$ and $g(\cdot)$ be two real mensurable functions and U be a random variable with $\text{GA}(\beta, \alpha)$ distribution. Similarly as before, we introduce the following notation. We will denote $\text{E}_{\text{GA}}(h(U))$ and $\text{Var}_{\text{GA}}(h(U))$ as the mean and variance of $h(U)$. Further, the covariance between the random variables $h(U)$ and $g(U)$ is denoted by

$$\text{Cov}_{\text{GA}}(h(U), g(U)) = \text{E}_{\text{GA}}(h(U)g(U)) - \text{E}_{\text{GA}}(h(U))\text{E}_{\text{GA}}(g(U)).$$

Lemma 7. *Under the hypothesis* H_{GA}*, when* $n \to \infty$ *we have*

(i) $\widehat{\theta}_n \to \widetilde{\theta}$ *a.s.,* $\widehat{\lambda}_n \to \widetilde{\lambda}$ *a.s., where*

$$\text{E}_{\text{GA}}\left[\log f_{\text{EP}}(X; \widetilde{\theta}, \widetilde{\lambda})\right] = \max_{\theta,\lambda} \text{E}_{\text{GA}}\left[\log f_{\text{EP}}(X; \theta, \lambda)\right].$$

(ii) $\widehat{\beta}_n \to \beta$ *a.s.,* $\widehat{\alpha}_n \to \alpha$ *a.s., and* β *and* α *satisfy*

$$\text{E}_{\text{GA}}\left[\log f_{\text{GA}}(X; \beta, \alpha)\right] = \max_{\bar{\beta},\bar{\alpha}} \text{E}_{\text{GA}}\left[\log f_{\text{GA}}(X; \bar{\beta}, \bar{\alpha})\right].$$

Remark: As before, we call attention of the reader that $\widetilde{\theta}$ and $\widetilde{\lambda}$ are functions of β and α, which is not explicit for simplicity.

We now discuss how to obtain $\widetilde{\theta}$ and $\widetilde{\lambda}$. For this, we first need to compute the expectation $\omega_{\text{GA}}(\theta, \lambda) \equiv \text{E}_{\text{GA}}\left[\log f_{\text{EP}}(X; \lambda, \theta)\right]$ which, after some algebra, results

$$\omega_{\text{GA}}(\theta, \lambda) = \log\theta + \log\left(\frac{\lambda}{1 - e^{-\lambda}}\right) - \theta\frac{\alpha}{\beta} - \lambda\left[1 - \left(\frac{\beta}{\beta + \theta}\right)^{\alpha}\right].$$

Hence, we find $\widetilde{\theta}$ and $\widetilde{\lambda}$ as solutions of the nonlinear system of equations

$$(\partial\omega_{\text{GA}}/\partial\theta, \partial\omega_{\text{GA}}/\partial\lambda)^{\top} = 0,$$

which leads to

$$\widetilde{\lambda} = \frac{(\beta + \widetilde{\theta})^{\alpha+1}}{\alpha\beta^{\alpha}}\left(\frac{1}{\widetilde{\theta}} - \frac{\alpha}{\beta}\right) \quad \text{and} \quad \frac{1}{\widetilde{\lambda}} - \frac{1}{1 - \text{e}^{-\widetilde{\lambda}}} + \left(\frac{\beta}{\beta + \widetilde{\theta}}\right)^{\alpha} = 0. \tag{4.11}$$

We now compute the mean and variance of the random variable

$$\log f_{\mathrm{EP}}(X;\widetilde{\theta},\widetilde{\lambda}) - \log f_{\mathrm{GA}}(X;\beta,\alpha)$$

(with $X \sim \mathrm{GA}(\beta,\alpha)$), which we will denote by $\mathrm{AM}_{\mathrm{GA}}$ and $\mathrm{AV}_{\mathrm{GA}}$. These results will be important to state our theorem about asymptotic distribution of the normalized test statistic under H_{GA}. It may be checked that $\beta/\widetilde{\theta}$ and $\widetilde{\lambda}$ depend only on α. With this, it is easy to see that $\mathrm{AM}_{\mathrm{GA}}$ and $\mathrm{AV}_{\mathrm{GA}}$ depend only on α. Further, after some algebra, these quantities can be expressed by

$$\begin{aligned}\mathrm{AM}_{\mathrm{GA}}(\alpha) &= \log\widetilde{\theta} + \log\Gamma(\alpha) + \log\left(\frac{\widetilde{\lambda}}{1-e^{-\widetilde{\lambda}}}\right) - \widetilde{\lambda}\left[1-\left(\frac{\beta}{\beta+\widetilde{\theta}}\right)^{\alpha}\right] - \\ &\quad \alpha\left(\frac{\widetilde{\theta}}{\beta} + \log\beta - 1\right) - (\alpha-1)\left(\Psi(\alpha) - \log\beta\right) \qquad (4.12)\end{aligned}$$

and

$$\begin{aligned}\mathrm{AV}_{\mathrm{GA}}(\alpha) &= (\widetilde{\theta}-\beta)^2\mathrm{Var}_{\mathrm{GA}}(X) + \widetilde{\lambda}\mathrm{Var}_{\mathrm{GA}}(\mathrm{e}^{-\widetilde{\theta}X}) + (\alpha-1)^2\mathrm{Var}_{\mathrm{GA}}(\log X) - \\ &\quad 2\widetilde{\lambda}(\widetilde{\theta}-\beta)\mathrm{Cov}_{\mathrm{GA}}(X,\mathrm{e}^{-\widetilde{\theta}X}) + 2(\alpha-1)(\widetilde{\theta}-\beta)\mathrm{Cov}_{\mathrm{GA}}(X,\log X) - \\ &\quad 2\widetilde{\lambda}(\alpha-1)\mathrm{Cov}_{\mathrm{GA}}(\log X,\mathrm{e}^{-\widetilde{\theta}X}), \qquad (4.13)\end{aligned}$$

where

$$\mathrm{Var}_{\mathrm{GA}}(X) = \frac{\alpha}{\beta^2},\quad \mathrm{Var}_{\mathrm{GA}}(\mathrm{e}^{-\widetilde{\theta}X}) = \left(\frac{\beta}{\beta+2\widetilde{\theta}}\right)^{\alpha} - \left(\frac{\beta}{\beta+\widetilde{\theta}}\right)^{2\alpha},\quad \mathrm{Var}_{\mathrm{GA}}(\log X) = \Psi'(\alpha),$$

$$\mathrm{Cov}_{\mathrm{GA}}(X,\mathrm{e}^{-\widetilde{\theta}X}) = -\frac{\alpha\widetilde{\theta}\beta^{\alpha-1}}{(\beta+\widetilde{\theta})^{\alpha+1}},\quad \mathrm{Cov}_{\mathrm{GA}}(X,\log X) = \frac{1}{\beta},$$

and

$$\mathrm{Cov}_{\mathrm{GA}}(\log X,\mathrm{e}^{-\widetilde{\theta}X}) = \left(\frac{\beta}{\beta+\widetilde{\theta}}\right)^{\alpha}\log\left(\frac{\beta}{\beta+\widetilde{\theta}}\right).$$

In Table 4.2, we present values of $\mathrm{AM}_{\mathrm{GA}}(\alpha)$, $\mathrm{AV}_{\mathrm{GA}}(\alpha)$, $\widetilde{\lambda}$ and $\widetilde{\theta}$ for $\beta = 1$ and some values of the parameter α. We observe that $\mathrm{AM}_{\mathrm{GA}}(\alpha)$ and $\mathrm{AV}_{\mathrm{GA}}(\alpha)$ approach 0 as $\alpha \to 1$. Further, for $\alpha < 1$ we obtain $\widetilde{\lambda} > 0$ and for $\alpha > 1$ we have $\widetilde{\lambda} < 0$, which makes sense since $\alpha < 1$ and $\lambda > 0$ corresponds to decreasing gamma and EP densities (respectively) and $\alpha > 1$ and $\lambda < 0$ corresponds to increasing gamma and EP densities (respectively).

We now introduce the quantity $\widetilde{T}_n^{\mathrm{GA}} = \ell_{\mathrm{EP}}^{(n)}(\widetilde{\theta},\widetilde{\lambda}) - \ell_{\mathrm{GA}}^{(n)}(\beta,\alpha)$ in order to state one of our main results.

Theorem 8. *Under $H_0 : \{X_i\}_{i=1}^n \sim \mathrm{GA}(\beta,\alpha)$,*

$$n^{-1/2}\{T_n - \mathrm{E}_{\mathrm{GA}}(T_n)\} \sim n^{-1/2}\{\widetilde{T}_n^{\mathrm{GA}} - n\mathrm{AM}_{\mathrm{GA}}(\alpha)\} \xrightarrow{d} N(0,\mathrm{AV}_{\mathrm{GA}}(\alpha)),$$

as $n \to \infty$, where $\mathrm{AM}_{\mathrm{GA}}(\alpha)$ and $\mathrm{AV}_{\mathrm{GA}}(\alpha)$ are given by (4.12) and (4.13), respectively, and "$\sim$" denotes asymptotically equivalent.

Remark: From classical Central Limit Theorem for iid random variables with finite second moment, we have that $n^{-1/2}\{\widetilde{T}_n^{\mathrm{GA}} - n\mathrm{AM}_{\mathrm{GA}}(\lambda)\} \xrightarrow{d} \mathrm{N}(0,\mathrm{AV}_{\mathrm{GA}}(\lambda))$ as $n \to \infty$.

α	$\text{AM}_{\text{GA}}(\alpha)$	$\text{AV}_{\text{GA}}(\alpha)$	$\widetilde{\lambda}$	$\widetilde{\theta}$
0.5	−0.136915	0.466563	2.6008	1.0663
0.6	−0.059737	0.176354	2.2015	0.9484
0.7	−0.023332	0.060862	1.6656	0.9228
1.2	−0.002271	0.004070	−0.8259	1.0128
1.5	−0.006794	0.010915	−1.8080	0.9873
2.0	−0.009330	0.013492	−3.1021	0.9059

Table 4.2: Values of $\text{AM}_{\text{GA}}(\alpha)$, $\text{AV}_{\text{GA}}(\alpha)$, $\widetilde{\lambda}$ and $\widetilde{\theta}$ for $\beta = 1$ and some values of α.

4.3 Distances and minimum sample size

We now propose a method to determine the minimum sample size required in order to discriminate exponential-Poisson and gamma distributions for a specified probability of correct selection (PCS) and a given tolerance level, which is defined in terms of some distance to measure the closeness between the EP and gamma distributions.

There are several ways to measure how close are two probability distributions. The most common measures are the Kolmogorov-Smirnov ($\mathcal{KS}$) distance and the Hellinger ($\mathcal{H}$) distance and we will use both in this paper.

Let f and g (with same support Ω) be two absolutely continuous density functions with distribution functions $F(x)$ and $G(x)$, respectively. The Kolmogorov-Smirnov distance between F and G is given by

$$\mathcal{KS}(F,G) = \sup_{x\in\Omega} |F(x) - G(x)|.$$

The Hellinger distance between f and g is defined by

$$\mathcal{H}(f,g) = \frac{1}{2}\int_{\Omega}(\sqrt{f(x)} - \sqrt{g(x)})^2 dx = 1 - \int_{\Omega}\sqrt{f(x)g(x)}dx.$$

In the following proposition we give an explicit expression for the Hellinger distance between EP and gamma distributions.

Proposition 9. *The Hellinger distance between the EP and gamma distributions is given by*

$$\mathcal{H}(f_{\text{EP}}, f_{\text{GA}}) = 1 - \left(\frac{\theta\lambda\beta^{\alpha}}{e^{\lambda}-1}\right)^{1/2} \frac{\Gamma((\alpha+1)/2)}{\Gamma(\alpha)^{1/2}} \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k![\theta(k+1/2)+\beta/2]^{(\alpha+1)/2}}.$$

Proof. By expanding the term $\exp\{(\lambda/2)e^{-\theta x}\}$ of $\sqrt{f_{\text{EP}}(x)f_{\text{GA}}(x)}$ in Taylor series and using Dominate Convergence Theorem, the result follows. □

If the distance between two probability distributions is small, it is expected that the minimum sample size required to discriminate them be large. Otherwise, a small or moderate sample size are sufficient to discriminate the models.

We assume that the user will specify before hand the PCS and the tolerance level in terms of the distance function between the EP and gamma distributions. When a tolerance level is specified (in term of some distance), this means that two distribution functions are not considered to be significantly different if their distance is less than the tolerance level. PCS and tolerance level play a similar role that the power and Type-I error in the corresponding hypothesis testing problem.

Based on PCS and tolerance level we can to determinate the minimum sample size required to discriminate the EP and gamma distributions. Here, the tolerance level is defined for the Kolmogorov-Smirnov and Hellinger distances. We observed in Section 4.2 that the normalized logarithm of the ratio of the maximized likelihoods is asymptotically normally distributed. Now, this and Hellinger (or Kolmogorov-Smirnov) distance will be used to determine the required sample size n such that the probability of correct selection achieves a certain protection level p for a given tolerance level D. We explain the procedure under the null hypothesis H_{EP} (the PCS under this hypothesis will be denoted by $\mathrm{PCS}_{\mathrm{EP}}$). The procedure under H_{GA} (the PCS under this hypothesis will be denoted by $\mathrm{PCS}_{\mathrm{GA}}$) follows in a similar way and therefore is omitted.

From Theorem 6, under H_{EP}, we have that $\mathrm{PCS}_{\mathrm{EP}}(\lambda) = P(T_n > 0)$ may be approximate by

$$\mathrm{PCS}_{\mathrm{EP}}(\lambda) \approx \Phi\left(-\frac{\sqrt{n}\mathrm{AM}_{\mathrm{EP}}(\lambda)}{\sqrt{\mathrm{AV}_{\mathrm{EP}}(\lambda)}}\right),$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution and $\mathrm{AM}_{\mathrm{EP}}(\lambda)$ and $\mathrm{AV}_{\mathrm{EP}}(\lambda)$ are given in (4.9) and (4.10), respectively. Now to determine the sample size needed to achieve at least a protection level p, we equate

$$\Phi\left(-\frac{\sqrt{n}\mathrm{AM}_{\mathrm{EP}}(\lambda)}{\sqrt{\mathrm{AV}_{\mathrm{EP}}(\lambda)}}\right) = p$$

and by solving for n we obtain

$$n = \left[\frac{z_p^2 \mathrm{AV}_{\mathrm{EP}}(\lambda)}{\mathrm{AM}_{\mathrm{EP}}^2(\lambda)}\right], \tag{4.14}$$

where z_p is the $100p$ percentile point of a standard normal distribution and $[a]$ denotes the smallest integer b such that $b > a$, for $a \in \mathbb{R}$. In analogous way, under the null hypothesis H_{GA} and using Theorem 6 we need

$$n = \left[\frac{z_p^2 \mathrm{AV}_{\mathrm{GA}}(\alpha)}{\mathrm{AM}_{\mathrm{GA}}^2(\alpha)}\right], \tag{4.15}$$

to choose the gamma distribution with PCS equal to p. Values of (4.14) for some values of λ, setting $\theta = 1$ and $p = 0.7$, are given in Table 4.3. In the same way, Table 4.4 lists values of (4.15) for some values of α, setting $\beta = 1$ and $p = 0.7$. In these Tables values of $\mathcal{KS}$ and $\mathcal{H}$ distances are also presented.

$\lambda \rightarrow$	-2	-1	-0.5	0.5	1	2
n	75	206	734	683	178	57
$\mathcal{KS}$	0.0265	0.0161	0.0086	0.0090	0.0176	0.0317
$\mathcal{H}$	0.0020	0.0007	0.0002	0.0002	0.0008	0.0026

Table 4.3: Values of n and the $\mathcal{KS}$ and $\mathcal{H}$ distances between $\text{EP}(1,\lambda)$ and $\text{GA}(\widetilde{\beta},\widetilde{\alpha})$ distributions for $\theta = 1$ and some values of λ.

$\alpha \rightarrow$	0.5	0.6	0.7	1.2	1.5	2.0
n	7	14	30	217	65	43
$\mathcal{KS}$	0.0433	0.0305	0.0204	0.0079	0.0145	0.0177
$\mathcal{H}$	0.0255	0.0121	0.0051	0.0006	0.0019	0.0027

Table 4.4: Values of n and the $\mathcal{KS}$ and $\mathcal{H}$ distances between $\text{GA}(1,\alpha)$ and $\text{EP}(\widetilde{\theta},\widetilde{\lambda})$ distributions for $\beta = 1$ and some values of α.

We observe that under H_{EP}, the distances are close to 0 when λ approaches 0. Under H_{GA}, when α is close to 1 the distances approach 0. This was expected since in this cases the distribution under null hypothesis approximate the exponential distribution, which is a particular case of both distributions considered in this chapter.

On the other hand, when λ moves away from 0 (under H_{EP}), the distances increase. The same is observed under H_{GA} when α moves away from 1. In Figures 4.2 and 4.3 plots of the distances as function of λ and α, respectively, are displayed. From these Figures, the discussion above may be clearly seen.

We now briefly discuss how to use the PCS and the tolerance level in a practical situation. Suppose one is interested in discriminating EP and gamma models, where the null hypothesis is H_{EP}. Further, suppose that the tolerance level is based on Hellinger distance and fixed at 0.0002. Therefore, from Table 4.3 one needs to take the sample size at least equal to $n > \max\{683, 734\} = 734$ to discriminate EP from gamma distribution. For a more accurate result, under the hypothesis H_{EP} (H_{GA}), a greater range of λ (and α) is necessary, as it was made in Figures 4.2 and 4.3.

4.4 Simulation

In this Section we perform some numerical experiments to observe how our asymptotic results derived in Section 4.3 work for different sample sizes. We here are interested in comparing the asymptotic probabilities of correct selection (PCS) under the hypothesis H_{EP} and H_{GA} with respect to the simulated probabilities based on Monte Carlo simulations.

Let us now to describe how the simulated results are obtained. First of all, suppose that

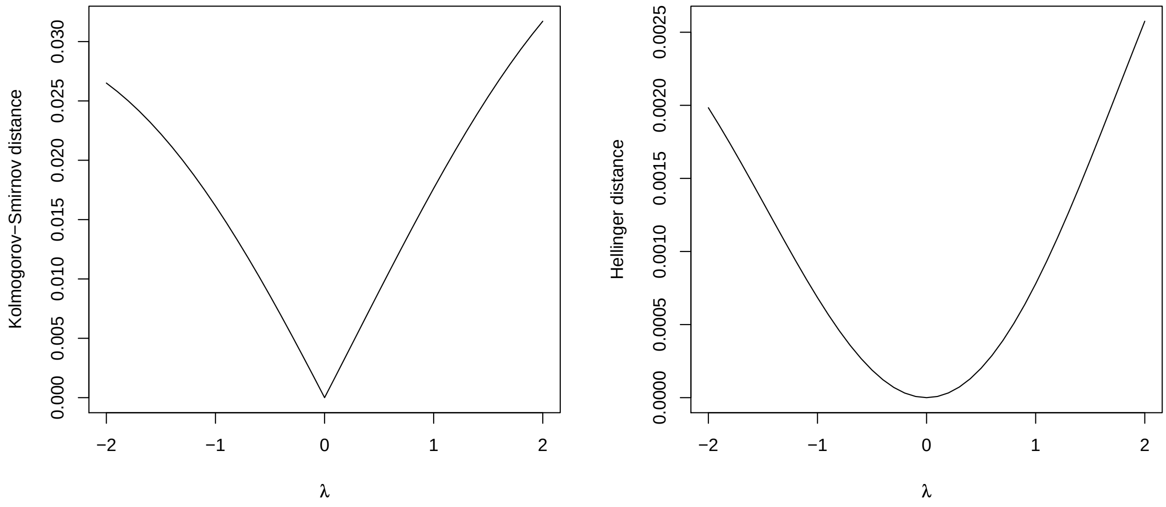


Figure 4.2: Kolmogorov-Smirnov (picture to the left) and Hellinger (picture to the right) distances between $\mathrm{EP}(1, \lambda)$ and $\mathrm{GA}(\widetilde{\beta}, \widetilde{\alpha})$ distributions as function of λ.

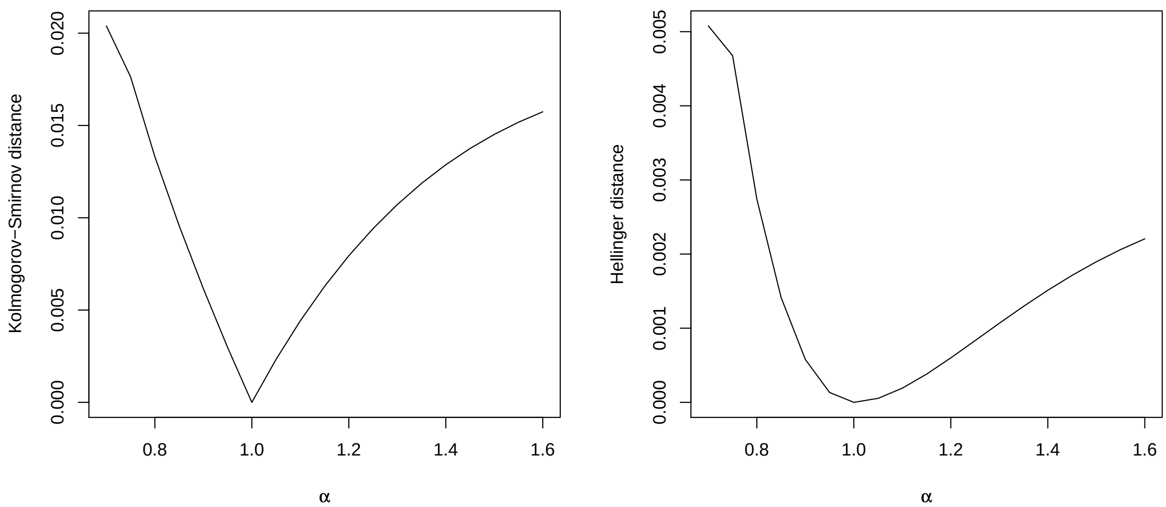


Figure 4.3: Kolmogorov-Smirnov (picture to the left) and Hellinger (picture to the right) distances between $\mathrm{GA}(1, \alpha)$ and $\mathrm{EP}(\widetilde{\theta}, \widetilde{\lambda})$ distributions as function of α.

the null hypothesis is that data comes from $\mathrm{EP}(\theta, \lambda)$ distribution, and the sample size equals n. The procedure in what follows holds in a similar way for the null hypothesis that the data comes from the $\mathrm{GA}(\beta, \alpha)$ distribution and therefore is omitted. Let N be the number of replicas of the Monte Carlo simulation and $I = (I_1, \ldots, I_N)^\top$ be a vector with length N. The steps, for each replica j, are the following:

1. Generate a random sample from the $\mathrm{EP}(\theta, \lambda)$ distribution with size n. This can be done by using the stochastic representations (1) or (4) (depending of the value of λ) or by using the inversion method;

2. Find the MLEs of (θ,λ) and (β,α), that are based on the EP and gamma distributions, respectively;

3. Compute the statistic $T_n = \ell_{\mathrm{EP}}^{(n)}(\widehat{\theta}_n, \widehat{\lambda}_n) - \ell_{\mathrm{GA}}^{(n)}(\widehat{\beta}_n, \widehat{\alpha}_n)$;

4. If $T_n > 0$, take $I_j = 1$, otherwise $I_j = 0$.

After running the above Monte Carlo simulation, the simulated probability of correct selection (PCS) is given by $\sum_{j=1}^{N} I_j / N$.

We compute the PCS based on simulations and we also compute it based on the asymptotic results derived in Section 3. Since the distribution of T_n does not depend of the scale parameters, we varied the shape parameter and set the scale parameter to be one in all cases.

First we consider that the null hypothesis is H_{EP} and compute the probabilities of correct selection for $\lambda = -1.2, -0.8, -0.4, 0.4, 0.8, 1.2$ and $n = 60, 80, 100, 200, 300, 400, 500$. These results are given in the Table 4.5. We see a good agreement between the asymptotic and empirical probabilities, mainly for moderate and large values of n. Further, when λ approaches to zero, the probabilities of correct selection approaches 0.5. This was expected since when λ goes to 0 both exponential-Poisson and gamma distributions converges to the same law, in this case, the exponential distribution. Another expected result is that when n increases the PCSs approaches 1.

In the Table 4.6 we present the asymptotic and simulated probabilities of correct selection under the null hypothesis H_{GA} for $\alpha = 0.75, 0.8, 0.9, 1.2, 1.5, 1.75$ and $n = 60, 80, 100, 200, 300, 400, 500$. We also observe a good agreement between the PCSs in this case. When α is close to 1, the probabilities are close to 0.5 and as n increases the probabilities goes to 1, as expected and discussed in the previous case.

	Asymptotic probability under H_{EP}						
$\lambda \downarrow n \rightarrow$	60	80	100	200	300	400	500
−1.2	0.6291	0.6483	0.6648	0.7264	0.7695	0.8027	0.8293
−0.8	0.5919	0.6058	0.6179	0.6644	0.6984	0.7258	0.7489
−0.4	0.5481	0.5555	0.5620	0.5873	0.6065	0.6225	0.6364
0.4	0.5495	0.5571	0.5638	0.5898	0.6096	0.6260	0.6403
0.8	0.5972	0.6119	0.6247	0.6735	0.7090	0.7375	0.7614
1.2	0.6403	0.6609	0.6787	0.7441	0.7892	0.8233	0.8502
	Empirical probability under H_{EP}						
$\lambda \downarrow n \rightarrow$	60	80	100	200	300	400	500
−1.2	0.5791	0.6137	0.6370	0.7177	0.7670	0.8030	0.8260
−0.8	0.5316	0.5502	0.5785	0.6395	0.6886	0.7180	0.7488
−0.4	0.5056	0.5192	0.5227	0.5410	0.5745	0.5931	0.6139
0.4	0.4989	0.5117	0.5296	0.5523	0.5746	0.5869	0.6169
0.8	0.5476	0.5606	0.5776	0.6407	0.6984	0.7351	0.7614
1.2	0.5897	0.6175	0.6417	0.7369	0.7889	0.8399	0.8641

Table 4.5: The PCS based on the Monte Carlo simulation and based on the asymptotic result under H_{EP} for some values of λ and for $n = 60, 80, 100, 200, 300, 400, 500$.

	Asymptotic probability under H_{GA}						
$\alpha \downarrow n \rightarrow$	60	80	100	200	300	400	500
0.75	0.7174	0.7467	0.7711	0.8531	0.9008	0.9312	0.9515
0.8	0.6677	0.6917	0.7122	0.7856	0.8339	0.8685	0.8946
0.9	0.5764	0.5881	0.5983	0.6376	0.6668	0.6907	0.7112
1.2	0.6086	0.6249	0.6390	0.6927	0.7314	0.7616	0.7869
1.5	0.6928	0.7197	0.7423	0.8211	0.8700	0.9033	0.9271
1.75	0.7227	0.7525	0.7771	0.8596	0.9068	0.9364	0.9560
	Empirical probability under H_{GA}						
$\alpha \downarrow n \rightarrow$	60	80	100	200	300	400	500
0.75	0.7392	0.7756	0.8087	0.8951	0.9381	0.9582	0.9723
0.8	0.6635	0.7006	0.7186	0.8157	0.8683	0.9003	0.9230
0.9	0.5587	0.5577	0.5694	0.6175	0.6504	0.6871	0.7120
1.2	0.6044	0.6054	0.6292	0.6958	0.7360	0.7640	0.7835
1.5	0.7150	0.7399	0.7530	0.8274	0.8670	0.8982	0.9170
1.75	0.7543	0.7765	0.7982	0.8645	0.8993	0.9308	0.9508

Table 4.6: The PCS based on the Monte Carlo simulation and based on the asymptotic result under H_{GA} for some values of α and for $n = 60, 80, 100, 100, 200, 300, 400, 500$.

4.5 Empirical illustrations

In this Section we applied our results in two real data set. In the first set, the data are 213 observations of times between failures of the air-conditioning equipment in 13 Boeing 720 jet airplanes reported from Proshan (1963). This data set has been used by Kus (2007). There we have that, based on the Kolmogorov-Smirnov test, the EP distribution presents a better fit than the gamma distribution.

The MLEs of the parameters of the EP and gamma distributions are $(\widehat{\theta}, \widehat{\lambda}) = (0.0075, 1.3130)$ and $(\widehat{\beta}, \widehat{\alpha}) = (0.0099, 0.9215)$, respectively. Figure 4.4 shows the histogram and the plots of the fitted densities of the EP and gamma distributions for the first data set. Empirical and fitted survival functions are also presented in this Figure.

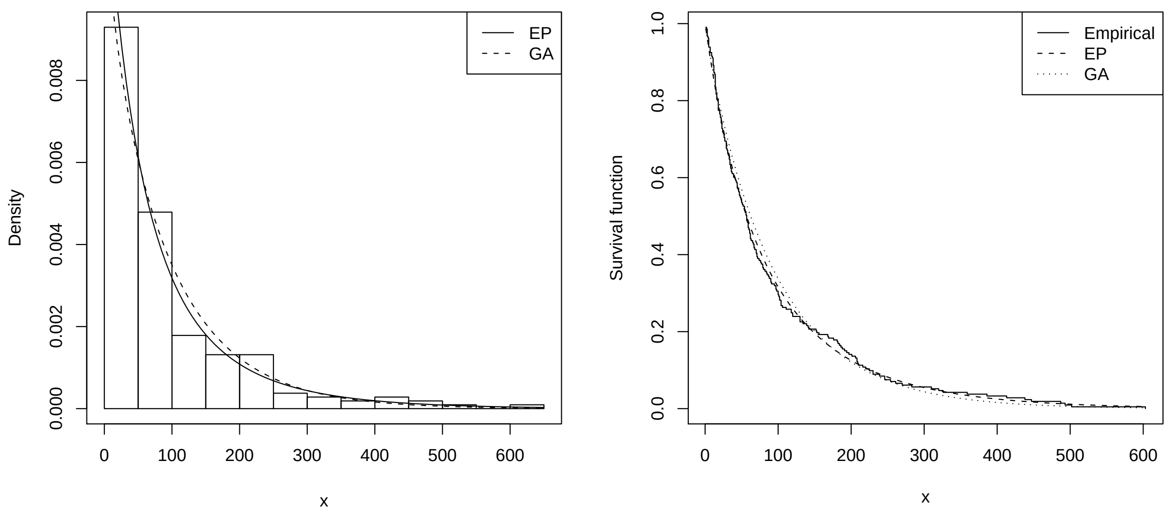


Figure 4.4: Histogram and plots of the fitted densities (picture to the left) of the EP and gamma distributions for the first data set. Empirical and fitted survival functions are presented in the picture to the right.

The logarithm of the likelihood ratio statistic equals 2.4975, so indicating that the EP distribution is more adequate than the gamma distribution to model the considered data set. Under the hypothesis that the data come from an EP distribution, we obtain the estimated quantities $\widetilde{\beta} = 0.0092$, $\widetilde{\alpha} = 0.8589$, $\mathrm{AM_{EP}}(\widehat{\lambda}) = 0.0052$ and $\mathrm{AV_{EP}}(\widehat{\lambda}) = 0.0108$, hence, it follows that $\mathrm{PCS_{EP}}(\widehat{\lambda}) = 0.7682$. Under the hypothesis that the data come from a gamma distribution, we have $\widetilde{\theta} = 0.0097$, $\widetilde{\lambda} = 0.3929$, $\mathrm{AM_{GA}}(\widehat{\alpha}) = -0.0008$ and $\mathrm{AV_{GA}}(\widehat{\alpha}) = 0.0016$. With this, we obtain $\mathrm{PCS_{GA}}(\widehat{\alpha}) = 0.6097$. We have that the probability of correct selection is at least equal to $\min\{0.7682, 0.6097\} = 0.6097$. Since the PCS is maxima under the hypothesis H_{EP}, we choose the EP distribution.

For our second application we use a real data set taken from the work of Bryson and Siddiqui(1969). These data are survival times of 43 patients suffering from chronic granulocytic leukemia. The MLEs of the parameters of the EP and gamma distributions are given by $(\widehat{\theta}, \widehat{\lambda}) = (0.0016, -1.7324)$ and $(\widehat{\beta}, \widehat{\alpha}) = (0.0014, 1.3074)$, respectively. Figure 4.4 shows the histogram and the plots of the fitted densities of the EP and gamma distributions for the second data set. Empirical and fitted survival function are also displayed.

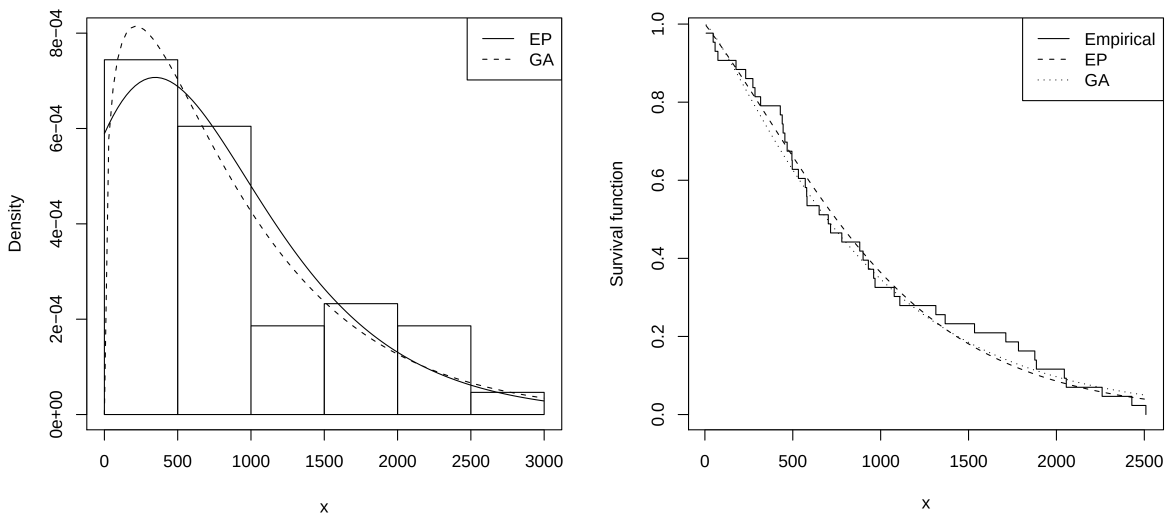


Figure 4.5: Histogram and plots of the fitted densities (picture to the left) of the EP and gamma distributions for the second data set. Empirical and fitted survival functions are presented in the picture to the right.

The test statistic equals 0.7421, which indicates that the EP distribution should be chosen. Under the hypothesis that the data come from an EP distribution, we obtain the estimated quantities $\widetilde{\beta} = 0.0015$, $\widetilde{\alpha} = 1.3817$, $\mathrm{AM}_{\mathrm{EP}}(\widehat{\lambda}) = 0.0070$ and $\mathrm{AV}_{\mathrm{EP}}(\widehat{\lambda}) = 0.0157$, hence, we have that $\mathrm{PCS}_{\mathrm{EP}}(\widehat{\lambda}) = 0.6422$. Under the hypothesis H_{GA}, we have that $\widetilde{\theta} = 0.0014$, $\widetilde{\lambda} = -1.2045$, $\mathrm{AM}_{\mathrm{GA}}(\widehat{\alpha}) = -0.0041$ and $\mathrm{AV}_{\mathrm{GA}}(\widehat{\alpha}) = 0.0070$. With this, we obtain that $\mathrm{PCS}_{\mathrm{GA}}(\widehat{\alpha}) = 0.6255$. Here the PCS is at least equal to $\min\{0.6422, 0.6255\} = 0.6255$. Since the PCS is maxima under the hypothesis H_{EP}, we choose the exponential-Poisson distribution.

4.6 Proof of the results

Proof of the Lemma 5

We now prove the part (i) of the Lemma 5. Since part (ii) follows in a similar way, we omitted its proof. We first check that $(\widetilde{\beta}, \widetilde{\alpha})$ is the unique solution of the nonlinear system of equations (4.8) and is the global maximum point of the function $\omega_{\mathrm{EP}}(\beta, \alpha)$.

From equations (4.8), we have that

$$\Psi(\widetilde{\alpha}) - \log \widetilde{\alpha} = \log \mathrm{E}_{\mathrm{EP}}(X) - \mathrm{E}_{\mathrm{EP}}(\log X). \tag{4.16}$$

Since the values of the expectations $\mathrm{E}_{\mathrm{EP}}(X)$ and $\mathrm{E}_{\mathrm{EP}}(\log X)$ are not important to prove the Lemma, we do not explicit them. We now show that the function $w(x) = \Psi(x) - \log x$ maps $[0, \infty)$ in $(-\infty, 0)$ and it is continuous and strictly increasing. By using the relation $\Psi(x) + 1/x = \Psi(x+1)$, we find that $\lim_{x\to 0^+} w(x) = \lim_{x\to 0^+}\{\Psi(x+1) - 1/x - \log x\} = \Psi(1) - \lim_{x\to 0^+} x^{-1}(1 + x\log x) = -\infty$. From Abramowitz and Stegun (1965) we have $\Psi(x) = \log x + O(x^{-1})$ for $x \to \infty$. Hence, it follows that $\lim_{x\to\infty} w(x) = 0$.

In order to study the behaviour of the function $w(\cdot)$, we take its derivative, that is $w'(x) = \Psi'(x) - 1/x$, where $\Psi'(x) = d\Psi(x)/dx$. By using the integral representation

$$\Psi'(x) = \int_0^\infty te^{-xt}/(1 - e^{-t})dt$$

and $1/x = \int_0^\infty e^{-xt}dt$, we obtain $w'(x) = \int_0^\infty (t + e^{-t} - 1)e^{-xt}/(1 - e^{-t})dt$. It may easily check that $t + e^{-t} - 1 \geq 0$ for all $t \geq 0$, which implies $w'(x) > 0$ for all $x > 0$. So, we obtain that the function $w(\cdot)$ is strictly increasing. Further, it is easy to see that $w(\cdot)$ is continuous. Since $\mathrm{E}_{\mathrm{EP}}(X) < \log \mathrm{E}_{\mathrm{EP}}(X)$ (by using Jensen's inequality for concave functions and using the fact $\mathrm{E}_{\mathrm{EP}}(X) \neq \log \mathrm{E}_{\mathrm{EP}}(X)$), we have that the solution $\widetilde{\alpha}$ in (4.16) exists and is unique. With this, we obtain from (4.8) that $\widetilde{\beta}$ exists and is also unique. It remains to show that $(\widetilde{\beta}, \widetilde{\alpha})$ is the global maximum point of the function $\omega_{\mathrm{EP}}(\beta, \alpha)$. We have that $\partial\omega_{\mathrm{EP}}^2/\partial\beta^2 = -\alpha/\beta^2 < 0$, $\partial\omega_{\mathrm{EP}}^2/\partial\beta^2 = -\Psi'(\alpha) < 0$ and

$$\det\left|\frac{\partial\omega_{\mathrm{EP}}^2(\beta,\alpha)}{\partial(\beta,\alpha)/\partial(\beta,\alpha)^\top}\right| = \frac{\alpha}{\beta^2}\{\Psi'(\alpha) - 1/\alpha\}.$$

As discussed before, we have that $\Psi'(\alpha) - 1/\alpha > 0$ for all $\alpha > 0$ and hence the above determinant is positive. Therefore, $(\widetilde{\beta}, \widetilde{\alpha})$ is the global maximum point of $\omega_{\mathrm{EP}}(\cdot, \cdot)$.

In order to show a.s. convergence stated in the Lemma, we need to verify that the three first regularity conditions given by White (1982a) holds. Assumptions 2.1 and 2.2 are clearly satisfied and Assumption 2.3 (a) may be easily checked. The result above on existence and uniqueness of the global maximum point of $\omega_{\mathrm{EP}}(\cdot, \cdot)$ proves that the Assumption 2.3 (b) is satisfied. Therefore, from Theorem 2.2 of White (1982b) we obtain that $\widehat{\beta}_n \overset{a.s.}{\to} \widetilde{\beta}$ and $\widehat{\alpha}_n \overset{a.s.}{\to} \widetilde{\alpha}$ as $n \to \infty$. □

Proof of the Theorem 6

It may be checked that Assumptions 2.4, 2.5 and 2.6 from White (1982a) holds with the EP and gamma distributions being the null and the alternative hypotheses, respectively.

Under H_{EP}, we have that $n^{-1/2}\{T_n - \mathrm{E}_{\mathrm{EP}}(T_n)\}$ equals

$$n^{-1/2}\left\{\ell_{\mathrm{EP}}^{(n)}(\widehat{\theta}_n, \widehat{\lambda}_n) - \ell_{\mathrm{GA}}^{(n)}(\widehat{\beta}_n, \widehat{\alpha}_n) - \int_0^\infty (\ell_{\mathrm{EP}}^{(n)}(\widehat{\theta}_n, \widehat{\lambda}_n) - \ell_{\mathrm{GA}}^{(n)}(\widehat{\beta}_n, \widehat{\alpha}_n))\tilde{f}_{\mathrm{EP}}(\tilde{x}; \theta, \lambda)d\tilde{x}\right\}, \quad (4.17)$$

where $\tilde{x} = (x_1, \ldots, x_n)$ and $\tilde{f}_{\mathrm{EP}}(\tilde{x}; \theta, \lambda) = \prod_{i=1}^n f_{\mathrm{EP}}(x_i; \theta, \lambda)$. In order to prove our Theorem we follow the ideas of the Theorem 1 of White (1982a). Then, following that proof and by using Lemma 5 we have that there exist sequences $(\theta_n^*, \lambda_n^*)$ and (β_n^*, α_n^*) which are tail equivalent to $(\widehat{\theta}_n, \widehat{\lambda}_n)$ and $(\widehat{\beta}_n, \widehat{\alpha}_n)$, respectively, and they assume their values in a compact convex neighborhood of (θ,λ) and $(\widetilde{\beta},\widetilde{\alpha})$, respectively. Since Assumption 2.4 from White (1982a) is satisfied, there exist measurable functions $(\bar{\theta}_n,\bar{\alpha}_n)$ and $(\bar{\beta}_n,\bar{\alpha}_n)$ which lies on the segments joining (θ_n^*,λ_n^*) to (θ,λ) and (β_n^*,α_n^*) to $(\widetilde{\beta},\widetilde{\alpha})$, respectively, such that the following first-order Taylor's expansion of the first term of (4.17) around (θ,λ) and $(\widetilde{\beta},\widetilde{\alpha})$ holds:

$$\ell_{\mathrm{EP}}^{(n)}(\widehat{\theta}_n,\widehat{\lambda}_n)-\ell_{\mathrm{GA}}^{(n)}(\widehat{\beta}_n,\widehat{\alpha}_n)=\ell_{\mathrm{EP}}^{(n)}(\theta,\lambda)-\ell_{\mathrm{GA}}^{(n)}(\widetilde{\beta},\widetilde{\alpha})+\frac{\partial\ell_{\mathrm{EP}}^{(n)}(\bar{\theta}_n,\bar{\lambda}_n)}{\partial\bar{\theta}_n}(\theta_n^*-\theta)+$$
$$\frac{\partial\ell_{\mathrm{EP}}^{(n)}(\bar{\theta}_n,\bar{\lambda}_n)}{\partial\bar{\lambda}_n}(\lambda_n^*-\lambda)-\frac{\partial\ell_{\mathrm{GA}}^{(n)}(\bar{\beta}_n,\bar{\alpha}_n)}{\partial\bar{\beta}_n}(\beta_n^*-\widetilde{\beta})-\frac{\partial\ell_{\mathrm{GA}}^{(n)}(\bar{\theta}_n,\bar{\alpha}_n)}{\partial\bar{\alpha}_n}(\alpha_n^*-\widetilde{\alpha}). \tag{4.18}$$

On the other hand, the Mean Value Theorem of the calculus gives us the existence of $(\ddot{\theta}_n,\ddot{\lambda}_n)$ and $(\ddot{\beta}_n,\ddot{\alpha}_n)$ (belonging on the segments joining (θ_n^*,λ_n^*) to (θ,λ) and (β_n^*,α_n^*) to $(\widetilde{\beta},\widetilde{\alpha})$, respectively) such that the integral in (4.17) admits the following Taylor's expansion around (θ,λ) and $(\widetilde{\beta},\widetilde{\alpha})$:

$$\begin{aligned}\int_0^\infty(\ell_{\mathrm{EP}}^{(n)}(\widehat{\theta},\widehat{\lambda})-\ell_{\mathrm{GA}}^{(n)}(\widehat{\beta},\widehat{\alpha}))\tilde{f}_{\mathrm{EP}}(\tilde{x};\theta,\lambda)d\tilde{x} \;=\;& \int_0^\infty(\ell_{\mathrm{EP}}^{(n)}(\theta,\lambda)-\ell_{\mathrm{GA}}(\widetilde{\beta},\widetilde{\alpha}))\tilde{f}_{\mathrm{EP}}(\tilde{x};\theta,\lambda)d\tilde{x}+\\ &(\theta_n^*-\theta)\frac{\partial}{\partial\ddot{\theta}_n}\int_0^\infty\ell_{\mathrm{EP}}^{(n)}(\ddot{\theta}_n,\ddot{\lambda}_n)\tilde{f}_{\mathrm{EP}}(\tilde{x};\theta,\lambda)d\tilde{x}+\\ &(\lambda_n^*-\lambda)\frac{\partial}{\partial\ddot{\lambda}_n}\int_0^\infty\ell_{\mathrm{EP}}^{(n)}(\ddot{\theta}_n,\ddot{\lambda}_n)\tilde{f}_{\mathrm{EP}}(\tilde{x};\theta,\lambda)d\tilde{x}-\\ &(\beta_n^*-\widetilde{\beta})\frac{\partial}{\partial\ddot{\beta}_n}\int_0^\infty\ell_{\mathrm{GA}}^{(n)}(\ddot{\beta}_n,\ddot{\alpha}_n)\tilde{f}_{\mathrm{EP}}(\tilde{x};\theta,\lambda)d\tilde{x}-\\ &(\alpha_n^*-\widetilde{\alpha})\frac{\partial}{\partial\ddot{\alpha}_n}\int_0^\infty\ell_{\mathrm{GA}}^{(n)}(\ddot{\beta}_n,\ddot{\alpha}_n)\tilde{f}_{\mathrm{EP}}(\tilde{x};\theta,\lambda)d\tilde{x}.\end{aligned} \tag{4.19}$$

It can be showed that arguing with the Dominate Convergence Theorem the differential and integral signs may be permuted in (4.19). By using (4.18) and (4.19) in (4.17), we obtain

$$\begin{aligned}\frac{T_n-\mathrm{E}_{\mathrm{EP}}(T_n)}{\sqrt{n}} \;=\;& n^{-1/2}\left\{\ell_{\mathrm{EP}}^{(n)}(\theta,\lambda)-\ell_{\mathrm{GA}}^{(n)}(\widetilde{\beta},\widetilde{\alpha})-n\int_0^\infty(\ell_{\mathrm{EP}}^{(n)}(\theta,\lambda)-\ell_{\mathrm{GA}}^{(n)}(\widetilde{\beta},\widetilde{\alpha}))f_{\mathrm{EP}}(x;\theta,\lambda)dx\right\}+\\ &\sqrt{n}(\theta_n^*-\theta)\left\{n^{-1}\frac{\partial\ell_{\mathrm{EP}}^{(n)}(\bar{\theta}_n,\bar{\lambda}_n)}{\partial\bar{\theta}_n}-\int_0^\infty\frac{\partial\log f_{\mathrm{EP}}(x;\ddot{\theta}_n,\ddot{\lambda}_n)}{\partial\ddot{\theta}_n}f_{\mathrm{EP}}(x;\theta,\lambda)dx\right\}+\\ &\sqrt{n}(\lambda_n^*-\lambda)\left\{n^{-1}\frac{\partial\ell_{\mathrm{EP}}^{(n)}(\bar{\theta}_n,\bar{\lambda}_n)}{\partial\bar{\lambda}_n}-\int_0^\infty\frac{\partial\log f_{\mathrm{EP}}(x;\ddot{\theta}_n,\ddot{\lambda}_n)}{\partial\ddot{\lambda}_n}f_{\mathrm{EP}}(x;\theta,\lambda)dx\right\}-\\ &\sqrt{n}(\beta_n^*-\widetilde{\beta})\left\{n^{-1}\frac{\partial\ell_{\mathrm{GA}}^{(n)}(\bar{\beta}_n,\bar{\alpha}_n)}{\partial\bar{\beta}_n}-\int_0^\infty\frac{\partial\log f_{\mathrm{GA}}(x;\ddot{\beta}_n,\ddot{\alpha}_n)}{\partial\ddot{\beta}_n}f_{\mathrm{GA}}(x;\beta,\alpha)dx\right\}-\\ &\sqrt{n}(\alpha_n^*-\widetilde{\alpha})\left\{n^{-1}\frac{\partial\ell_{\mathrm{GA}}^{(n)}(\bar{\beta}_n,\bar{\alpha}_n)}{\partial\bar{\alpha}_n}-\int_0^\infty\frac{\partial\log f_{\mathrm{GA}}(x;\ddot{\beta}_n,\ddot{\alpha}_n)}{\partial\ddot{\alpha}_n}f_{\mathrm{GA}}(x;\beta,\alpha)dx\right\}.\end{aligned} \tag{4.20}$$

Lemma 5 gives us that $(\bar{\theta}_n,\bar{\lambda}_n)\overset{a.s.}{\to}(\theta,\lambda)$, $(\bar{\beta}_n,\bar{\alpha}_n)\overset{a.s.}{\to}(\widetilde{\beta},\widetilde{\alpha})$, $(\ddot{\theta}_n,\ddot{\lambda}_n)\overset{a.s.}{\to}(\theta,\lambda)$ and $(\ddot{\beta}_n,\ddot{\alpha}_n)\overset{a.s.}{\to}(\widetilde{\beta},\widetilde{\alpha})$ as $n\to\infty$. Hence, we obtain that

$$\int_0^\infty\frac{\partial\log f_{\mathrm{EP}}(x;\ddot{\theta}_n,\ddot{\lambda}_n)}{\partial\ddot{\theta}_n}f_{\mathrm{EP}}(x;\theta,\lambda)dx\overset{a.s.}{\longrightarrow}\int_0^\infty\frac{\partial\log f_{\mathrm{EP}}(x;\theta,\lambda)}{\partial\theta}f_{\mathrm{EP}}(x;\theta,\lambda)dx=0,$$

$$\int_0^\infty\frac{\partial\log f_{\mathrm{EP}}(x;\ddot{\theta}_n,\ddot{\lambda}_n)}{\partial\ddot{\lambda}_n}f_{\mathrm{EP}}(x;\theta,\lambda)dx\overset{a.s.}{\longrightarrow}\int_0^\infty\frac{\partial\log f_{\mathrm{EP}}(x;\theta,\lambda)}{\partial\lambda}f_{\mathrm{EP}}(x;\theta,\lambda)dx=0,$$

$$\int_0^\infty \frac{\partial \log f_{\mathrm{GA}}(x;\ddot{\beta}_n,\ddot{\alpha}_n)}{\partial \ddot{\beta}_n} f_{\mathrm{GA}}(x;\beta,\alpha)dx \xrightarrow{a.s.} \int_0^\infty \frac{\partial \log f_{\mathrm{GA}}(x;\beta,\alpha)}{\partial \beta} f_{\mathrm{GA}}(x;\beta,\alpha)dx = 0$$

and

$$\int_0^\infty \frac{\partial \log f_{\mathrm{GA}}(x;\ddot{\beta}_n,\ddot{\alpha}_n)}{\partial \ddot{\alpha}_n} f_{\mathrm{GA}}(x;\beta,\alpha)dx \xrightarrow{a.s.} \int_0^\infty \frac{\partial \log f_{\mathrm{GA}}(x;\beta,\alpha)}{\partial \alpha} f_{\mathrm{GA}}(x;\beta,\alpha)dx = 0$$

Furthermore, from Lemma 3.1 of White (1981) it follows that

$$n^{-1}\frac{\partial \ell_{\mathrm{EP}}^{(n)}(\bar{\theta}_n,\bar{\lambda}_n)}{\partial \bar{\theta}_n} \xrightarrow{a.s.} \int_0^\infty \frac{\partial \log f_{\mathrm{EP}}(x;\theta,\lambda)}{\partial \theta} f_{\mathrm{EP}}(x;\theta,\lambda)dx = 0,$$

$$n^{-1}\frac{\partial \ell_{\mathrm{EP}}^{(n)}(\bar{\theta}_n,\bar{\lambda}_n)}{\partial \bar{\lambda}_n} \xrightarrow{a.s.} \int_0^\infty \frac{\partial \log f_{\mathrm{EP}}(x;\theta,\lambda)}{\partial \lambda} f_{\mathrm{EP}}(x;\theta,\lambda)dx = 0,$$

$$n^{-1}\frac{\partial \ell_{\mathrm{GA}}^{(n)}(\bar{\beta}_n,\bar{\alpha}_n)}{\partial \bar{\beta}_n} \xrightarrow{a.s.} \int_0^\infty \frac{\partial \log f_{\mathrm{GA}}(x;\widetilde{\beta},\widetilde{\alpha})}{\partial \widetilde{\beta}} f_{\mathrm{GA}}(x;\widetilde{\beta},\widetilde{\alpha})dx = 0,$$

and

$$n^{-1}\frac{\partial \ell_{\mathrm{GA}}^{(n)}(\bar{\beta}_n,\bar{\alpha}_n)}{\partial \bar{\alpha}_n} \xrightarrow{a.s.} \int_0^\infty \frac{\partial \log f_{\mathrm{GA}}(x;\widetilde{\beta},\widetilde{\alpha})}{\partial \widetilde{\alpha}} f_{\mathrm{GA}}(x;\widetilde{\beta},\widetilde{\alpha})dx = 0.$$

Since Assumptions 2.1-2.6 of White (1982a) are satisfied, it follows from Theorem 3.3 of White (1982b) that $\sqrt{n}\{(\theta_n^*,\lambda_n^*) - (\theta,\lambda)\}$ and $\sqrt{n}\{(\beta_n^*,\alpha_n^*) - (\widetilde{\beta},\widetilde{\alpha})\}$ are asymptotically normal distributed.

The results discussed above may be applied in (4.20) and they ensure that $n^{-1/2}\{T_n - \mathrm{E}_{\mathrm{EP}}(T_n)\}$ is asymptotically equivalent to $n^{-1/2}\{\widetilde{T}_n^{\mathrm{EP}} - n\mathrm{AM}_{\mathrm{EP}}(\lambda)\}$. $\square$

Proof of the Lemma 7

We here prove the part (i) of the Lemma 7. Proof of the part (ii) follows in a similar way and therefore is omitted. From second equation of (4.11), we see that the study of the continuous function $h(x) = -x^{-1} + (1-e^{-x})^{-1}$, for $x \in \mathbb{R}$, will be of interest. One may check that $\lim_{x\to-\infty} h(x) = 0$ and $\lim_{x\to\infty} h(x) = 1$. Further, the derivative of $h(\cdot)$ is given by

$$h'(x) = e^{-x}\frac{e^x + e^{-x} - x^2 - 2}{x^2(1-e^{-x})^2} = \frac{4e^{-x}}{x^2(1-e^{-x})^2}\left\{\sinh^2(x/2) - x^2/4\right\}.$$

It is easy to see that $\sinh^2(x/2) \geq x^2/4$ for all $x \in \mathbb{R}$. Further, at $x = 0$ we have $h'(0) \equiv \lim_{x\to 0} h'(x) = 0$. Therefore $h(\cdot)$ is a strictly increasing function that maps $(-\infty,\infty)$ in $(0,1)$.

On the other hand, from the first equation of (4.11), we see that it is important to study the behaviour of the continuous function $z(x) = (1/x - \alpha/\beta)(\beta+x)^{\alpha+1}/(\alpha\beta^\alpha)$, for $x > 0$. It is easy to show that $\lim_{x\to 0} z(x) = \infty$ and $\lim_{x\to\infty} z(x) = -\infty$. Further, we have that

$$z'(x) = \frac{(\beta+x)^\alpha}{\alpha\beta^\alpha}\left\{-(\alpha+1)\frac{\alpha}{\beta} - \frac{1}{x}\left(\frac{\beta}{x} - \alpha\right)\right\}.$$

If $\beta/x \geq \alpha$, we easily see that $z'(x) \leq 0$. Suppose $1/x < \alpha/\beta$, hence it follows that $x^{-1}(\alpha - \beta/x) - (\alpha+1)\alpha/\beta < -\beta/x^2 - \alpha/\beta < 0$. With the results above, we obtain that $z(\cdot)$ maps $(0,\infty)$ in $(-\infty,\infty)$ and is strictly decreasing.

From second equation of (4.11), we have $\theta = \beta\{h(\lambda(\theta))^{-1/\alpha} - 1\}$, which is a continuous and strictly decreasing function (since $h(\cdot)$ is continuous and strictly increasing). Therefore we obtain that $z(\beta\{h(\lambda)^{-1/\alpha} - 1\})$ is a continuous and strictly decreasing function (with respect to λ). Then we conclude that $\widetilde{\lambda}$ exists and is unique and consequently $\widetilde{\theta}$ also exists and is unique.

It now remains to show that $(\widetilde{\theta},\widetilde{\lambda})$ is a global maximum point of the function $\omega_{\rm GA}(\theta,\lambda)$. One may check that

$$\frac{\partial^2\omega_{\rm GA}}{\partial\widetilde{\lambda}^2} = \frac{-4e^{-\widetilde{\lambda}}}{\widetilde{\lambda}^2(1-e^{-\widetilde{\lambda}})^2}\left\{\sinh^2(\widetilde{\lambda}/2) - \widetilde{\lambda}^2/4\right\} \leq 0, \quad \frac{\partial^2\omega_{\rm GA}}{\partial\widetilde{\theta}\partial\widetilde{\lambda}} = -\alpha\beta^\alpha(\beta+\widetilde{\theta})^{-(\alpha+1)} \leq 0$$

and

$$\frac{\partial^2\omega_{\rm GA}}{\partial\widetilde{\theta}^2} = -\frac{h(\widetilde{\lambda})^{2/\alpha}}{\beta^2(1-h(\widetilde{\lambda})^{1/\alpha})^2}\left\{1-\alpha(\alpha+1)\widetilde{\lambda}h(\widetilde{\lambda})(1-h(\widetilde{\lambda})^{1/\alpha})\right\} \leq 0.$$

With this, we have easily that the matrix $\partial^2\omega_{\rm GA}/\partial\widetilde{\theta}\partial\widetilde{\lambda}$ is negative definite and hence our claim follows.

To show a.s. convergence stated in the Lemma, we show that the three first regularity conditions given by White (1982a) holds. Assumptions 2.1, 2.2 and 2.3 (a) are easily verified and are left to the reader. The results on existence and uniqueness of the global maximum point of $\omega_{\rm GA}(\cdot,\cdot)$ proves that the Assumption 2.3 (b) is satisfied. Therefore, from Theorem 2.2 of White (1982b) it follows that $\widehat{\theta}_n \stackrel{a.s.}{\to} \widetilde{\theta}$ and $\widehat{\lambda}_n \stackrel{a.s.}{\to} \widetilde{\lambda}$ as $n \to \infty$. $\square$

Proof of the Theorem 8

By using Lemma 7, this proof may be obtained similarly as proof of the Theorem 6 and therefore it is omitted. $\square$

4.7 Concluding remarks

We proposed a likelihood ratio test to discriminate exponential-Poisson (EP) and gamma distributions based on Cox's statistic. We showed that our test statistic properly normalized is asymptotically normally distributed for both null hypotheses (data come from an EP or gamma distribution). With this, we discussed how to determine the minimum sample size to discriminate the EP and gamma distributions based on the probability of correct selection and a tolerance level (based on Kolmogorov-Smirnov or Hellinger distances). Future research would be, for example, to discriminate EP and Weibull distribuions. These distributions have several properties in common and have been compared in practical situations. To consider discrimination between these models under some type of censoring would also be of interest, as made by Dey and Kundu (2012), where the discrimination between Weibull and log-normal distributions is considered under Type-II censoring.

References

[1] Abramowitz, M., Stegun, I.A. Handbook of Mathematical Functions. New York: Dover, 1965.

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

[3] Adamidis, K., Dimitrakopoulou, T., Loukas, S. (2005). On a generalization of the exponential-geometric distribution. *Statistics and Probability Letters*, **73**, 259–269.

[4] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

[5] Al-Mutairi, D.K., Ghitany, M.E., Gupta, R.C. (2011). Estimation of reliability in a series system with random sample size. *Computational Statistics and Data Analysis*, **55**, 964–972.

[6] Bain, L.J., Engelhardt, M. (1980). Probability of correct selection of Weibull versus gamma based on likelihood ratio. *Communications in Statistics - Simulation and Computation*, **9**, 375–381.

[7] Balasooriya, U., Abeysinghe, T. (1994). On selecting between Gamma and Weibull distributions: An approach based on prediction of order statistics. *Journal of Applied Statistics*, **21**, 17–27.

[8] Barreto-Souza, W., Cribari-Neto, F. (2009). A generalization of the exponential-Poisson distribution. *Statistics and Probability Letters*, **79**, 2493–2500.

[9] Barreto-Souza, W., Simas, A.B. (2013). The exp-G family of probability distributions. *Brazilian Journal of Probability and Statistics*, **27**, 84–109.

[10] Bryson, M.C., Siddiqui, M.M. (1969). Some criteria for aging. *Journal of the American Statistical Association*, **64**, 1472-1483.

[11] Chahkandi, M., Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis*, **53**, 4433–4440.

[12] Chen, M.H., Ibrahim, J.G., Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**, 909–919.

[13] Cooner, F., Banerjee, S., Carlin, B.P., Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, **102**, 560–572.

[14] Cox, D.R. (1961). Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability, Berkeley, University of California Press. 105–123.

[15] Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society - Series B*, **24**, 406–424.

[16] Dey, A.K., Kundu, D. (2012). Discriminating between the Weibull and log-normal distributions for Type-II censored data. *Statistics*, **46**, 197–214.

[17] Dumonceaux, R. and Antle, C. E. (1973). Discriminating between the log-normal and Weibull distribution. *Technometrics*, **15**, 923–926.

[18] Dumonceaux, R., Antle, C.E., Haas, G. (1973). Likelihood ratio test for discrimination between two models with unknown location and scale parameters. *Technometrics*, **15**, 19–27.

[19] Fearn, D.H. and Nebenzahl, E. (1991). On the maximum likelihood ratio method of deciding between the Weibull and Gamma distributions. *Communications in Statistics – Theory and Methods*, **20**, 579–593.

[20] Gupta, R.D. and Kundu, D. (2003). Discrimination between Weibull and generalized exponential distributions. *Computational Statistics and Data Analysis*, **43**, 179–196.

[21] Gupta, R. D. and Kundu, D. (2004). Discrimination between gamma and generalized exponential distributions. *Journal of Statistical Computation and Simulation*, **74**, 107–121.

[22] Karlis, D. (2009). A note on the exponential Poisson distribution: A nested EM algorithm. *Computational Statistics and Data Analysis*, **53**, 894–899.

[23] Kundu, D., Gupta, R. D., Manglick, A. (2005). Discriminating between the log-normal and generalized exponential distributions. *Journal of Statistical Planning and Inference*, **127**, 213–227.

[24] Kus, C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis*, **51**, 4497–4509.

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

[26] Morais, A.L., Barreto-Souza, W. (2011). A compound class of Weibull and power series distributions. *Computational Statistics and Data Analysis*, **55**, 1410–1425.

[27] Nichols, M.D. and Padgett, W. J. (2006). A bootstrap control chart for Weibull percentiles. *Quality and reliability engineering international*, **22**, 141-151.

[28] Proschan, F. (1963) Theoretical explanation of observed decreasing failure rate. *T*echnometrics, **5**, 375-383.

[29] Quesenberry, C.P., Kent, J. (1982). Selecting among probability distributions used in reliability. *Technometrics*, **24**, 59–65.

[30] Ristić, M.M., Nadarajah, S. (2012). A new lifetime distribution. *Journal of Statistical Computation and Simulation*. In press. DOI: 10.1080/00949655.2012.697163.

[31] Wiens, B.L. (1999). When log-normal and gamma models give different results: A case study. *American Statistician*, **53**, 89–93.

[32] White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association*, **76**, 419–433.

[33] White, H. (1982a). Regularity conditions for Cox's test of non-nested hypotheses. *Journal of Econometrics*, **19**, 301–318.

[34] White, H. (1982b). Maximum likelihood estimation of misspecified Models. *Econometrica*, **50**, 1–26.

[35] Yakovlev, A.Y., Tsodikov, A.D. Stochastic Models of Tumor Latency and Their Biostatistical Applications. World Scientific, Singapore, 1996.

CHAPTER 5

Modified moment estimator for the precision parameter in a class of regression models

Artigo atualmente submetido para publicação.

Resumo

Neste capítulo, consideramos a estimação do parâmetro de precisão para uma extensa classe de modelos de regressão. Por exemplo, nossos resultados valem para os modelos lineares generalizados (McCullagh e Nelder, 1989), modelos de quase-verossimilhança (Wedderburn, 1974), entre outros. Sabemos que o estimador do parâmetro de dispersão baseado na estatística de Pearson (também conhecido como estimador de momentos do parâmetro de dispersão) funciona bem para os modelos lineares generalizados e de quase-verossimilhança, mas a versão para o parâmetro de precisão deste estimador é significativamente viesado para pequenas e médias amostras. Propomos, então, um método simples para a redução do viés deste estimador para o parâmetro de precisão. O uso prático da redução do viés é ilustrado através de estudos de simulação, onde simulações de Monte Carlo são usados para comparar a performance desses estimadores.

Palavras-chave: Estimador de momentos; Modelos de regressão; Parâmetro de precisão; Parâmetro de dispersão.

Abstract

In this chapter, we consider precision parameter estimation for an extensive class of regression models. For instance, our results hold for generalized linear models (see McCullagh and

Nelder, 1989), quasi-likelihood models (see Wedderburn, 1974), among others. The Pearson-based dispersion estimator (which is also known as the moment estimator of the dispersion parameter) is known to work well for both generalized linear models and quasi-likelihood models, but the precision parameter version of this estimator is significatively biased in small and medium sample sizes. We thus propose a simple bias-reduction method to reduce the bias of this precision parameter estimator. The practical use of such bias reduction is illustrated in a simulation study, where Monte Carlo simulation is used to compare the finite sample performances of all these estimators.

Keywords: Regression models; Precision parameter; Dispersion parameter; Moment estimator.

5.1 Introduction

The aim of this chapter is to study and reduce the bias of the Pearson-based precision parameter estimation for a large class of regression models. We are considering a large class of regression models, namely, the class of regression models in which the response variables are modelled through the mean, say μ, and which the variance has the form $\phi^{-1}V(\mu)$, where where $V(\mu)$ is a function of the mean, $\phi^{-1} = \sigma^2$ is called the precision parameter and σ^2 is the dispersion parameter. To be more specific, consider the independent random variables $Y_1, \ldots, Y_n$, n being the sample size, where for each i, $E(Y_i) = \mu_i$ and $Var(Y_i) = \sigma^2 V(\mu_i)$. Further, we consider a systematic component which is parametrized as $g(\mu_i) = \eta_i = h(x; \beta_i)$, where $g(\cdot)$ is a known one-to-one link function, $x_i = (x_{i_1}, \ldots, x_{i_q})^\top$ is a vector of known and fixed values of q explanatory variables for the ith unit of observation, $\beta = (\beta_1, \ldots, \beta_p)^\top$, with $p < n$, is a set of unknown parameters to be estimated and $h(\cdot;\cdot)$ is a continuously differentiable function such that the $n \times p$ derivative matrix $\tilde{X} = \partial\eta_i/\partial\beta$ has rank p for all β.

This class of regression models covers many important regression models, such as the generalized linear models (see McCullagh and Nelder, 1989), the exponential family nonlinear models (see Cordeiro and Paula, 1989), the quasi-likelihood models (see Wedderburn, 1974), the extended quasi-likelihood models (see Nelder and Pregibon, 1987), the beta regression models (see Ferrari and Cribari-Neto, 2004), the symmetrical models with mean $\mu \in (0, \infty)$ (see Lange et al., 1989), the dispersion models which are modelled through the mean (see Jørgensen, 1987), among others.

It is known in the literature that the Pearson-based dispersion parameter estimator, say $\tilde{\sigma}^2$ works well for both generalized linear models and the quasi-likelihood models. Nevertheless, the precision parameter version of this estimator, say $\tilde{\phi} = \tilde{\sigma}^{-2}$, has some drawbacks, for instance, by Jensen inequality, $E(\tilde{\phi}) \geq 1/E(\tilde{\sigma}^2)$, and thus, in general, the precision parameter is positively biased. Several models use the precision parameter, for instance, Cordeiro and McCullagh (1991) parametrized the generalized linear models in terms of the precision parameter and computed the $\mathcal{O}(n^{-1})$ bias of its maximum likelihood estimator, Jørgensen (1997) parametrizes the dispersion model in terms of a precision parameter, Ferrari and Cribari-Neto (2004) defined a class of regression models with beta distributed response variables in which the parameterization uses a precision parameter, among others.

The usage of the precision parameter becomes particularly useful when the dispersion parameter is small. This phenomenon is very common in practice, so that Jørgensen (1987) studied the asympotic properties of the dispersion models when the dispersion, σ^2, converges to zero, or equivalently, when the precision parameter, ϕ, tends to infinity.

It is known that the Pearson-based dispersion parameter estimator is consistent, which implies that it is asymptotically unbiased. Since the reciprocal function is continous, we have also that the precision parameter estimator is also consistent and thus asymptotically unbiased. This, however, does not imply that the finite sample behavior should be similar, that is, even though the Pearson-based dispersion parameter estimator has a very small bias, even in small samples, the precision parameter version is, generally, very biased. This can be easily seen, as remarked above, by Jensen inequality.

The chief goal of this paper is to obtain a simple modified estimator of the precision parameter, preserving all the good large sample asymptotic properties, and having less bias for small samples when compared to the original estimators. We begin by proposing a bias-reduction method which preserves the asymptotic properties of this estimator and, generally, reduce the bias. An alternative approach is given by considering, also, two bootstrap bias adjustment schemes.

The rest of this chapter unfolds as follows. In Section 5.2, we consider the Pearson-based dispersion and precision parameter estimators, we give a simple bias-adjusted estimator for the precision parameter and consider the bootstrap bias-correction schemes. In Section 5.3, we present simulation results that show that the proposed estimators have better performance in small samples, in terms of bias, than the original versions. Finally, the chapter is concluded in Section 5.4 with some final remarks.

5.2 Estimation and bias-reduction

The maximum likelihood estimator of σ^2 has, for some regression models, some undesirable properties, for instance, the generalized linear models (see McCullagh and Nelder, 1989) whose maximum likelihood estimator of σ^2 depends on the deviance function. For these models the Pearson-based dispersion parameter estimator

$$\tilde{\phi}^{-1} = \tilde{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}, \tag{5.1}$$

is prefered instead of the usual maximum likelihood estimate, where $\hat{\mu}_i$ is the maximum likelihood of μ_i. In fact, this follows a long tradition for linear normal models for which the estimator in equation (5.1) is always used and is unbiased. In fact, $\tilde{\sigma}^2$ is the default estimator of σ^2 in generalized linear model functions in the statistical programs `S-Plus` and `R`.

Therefore, the standard procedure for estimating $\phi = \sigma^{-2}$ is considering $\tilde{\phi} = \tilde{\sigma}^{-2}$. We now suggest an heuristic approach of bias-adjusting this estimator. To begin with, we would like to remark that, in general and under suitable regularity conditions, the Pearson statistic

$$X^2(\mu(\hat{\beta}),\sigma^2) = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\sigma^2 V(\hat{\mu}_i)},$$

which is in fact equal to the score statistic for goodness of fit for a generalized nonlinear model (see Smyth, 2003), is approximately distributed as χ^2_{n-p}. This explains why in expression (5.1) the quantity $n-p$ is used as denominator instead of n. We will now use an analogous argument to provide a bias-reduced version of $\tilde{\phi}$. Accordingly, we have that $X^2(\mu(\hat{\beta}),\sigma^2)$ is approximately χ^2_{n-k} which is in fact the Gamma distribution with parameters $(n-p)/2$ and 2, i.e., is approximately $\Gamma((n-p)/2,2)$. Therefore, $X^2(\mu(\hat{\beta}),\sigma^2)^{-2}$ is approximately distributed as $\Gamma^{-1}((n-p)/2,1/2)$, i.e., an inverse Gamma distribution with parameters $(n-p)/2$ and $1/2$. This implies that its mean is approximately $1/(n-p-2)$, which leads us to the following modified estimator for ϕ:

$$\breve{\phi} = (n-p-2)\left(\sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}\right)^{-1}. \tag{5.2}$$

Following the approximation argument, this implies that this estimator is less biased than the original estimator $\tilde{\phi}$, and further, it is obvious that it also has a less variance, which, heuristically, implies that this estimator has better mean-squared error performance. In the following Section we compare these estimators by simulation for a number of different models.

For comparison purposes, we consider now a well-known approach to bias-reducing estimate of ϕ. It is based upon the numerical estimation of ϕ through the parametric bootstrap resampling scheme introduced by Efron (1979). Consider a random sample $\underset{\sim}{y} = (y_1,\ldots,y_n)^\top$ be a random sample of size n, where each element is a random draw from the random variable Y which has the distribution function $F = F(\phi)$. Here, ϕ is the parameter that indexes the distribution, and is viewed as a functional of F, i.e., $\phi = t(F)$. Finally, let $\tilde{\phi}$ be an estimator of ϕ based on $\underset{\sim}{y}$; we write $\tilde{\phi} = s(\underset{\sim}{y})$. The application of the bootstrap method consists in obtaining, from the original sample $\underset{\sim}{y}$, a large number of pseudo-samples $\underset{\sim}{y}^* = (y_1^*,\ldots,y_n^*)^\top$, and then extracting information from these samples to improve inference. The bootstrap samples are obtained from $F(\tilde{\phi})$, which we shall denote as $F_{\tilde{\phi}}$, through sampling with replacement.

Let $B_F(\tilde{\phi},\phi)$ be the bias of the estimator $\tilde{\phi} = s(\underset{\sim}{y})$, that is,

$$B_F(\tilde{\phi},\phi) = \mathrm{E}_F[\tilde{\phi} - \phi] = \mathrm{E}_F[s(\underset{\sim}{y})] - t(F),$$

where the subscript F indicates that expectation is taken with respect to F . The bootstrap estimators of the bias are obtained by replacing the true distribution F, which generated the original sample, with $F_{\tilde{\phi}}$ in the above expression. Therefore, the estimates of the bias are given by

$$B_{F_{\tilde{\phi}}}(\tilde{\phi},\phi) = \mathrm{E}_{F_{\tilde{\phi}}}[s(\underset{\sim}{y})] - t(F_{\tilde{\phi}}).$$

If B bootstrap samples $(\underset{\sim}{y}^{*1},\underset{\sim}{y}^{*2},\ldots,\underset{\sim}{y}^{*B})$ are generated independently from the original sample $\underset{\sim}{y}$, and the respective bootstrap replications $(\tilde{\phi}^{*1},\tilde{\phi}^{*2},\ldots,\tilde{\phi}^{*B})$ are calculated, where

$\tilde{\phi}^{*b} = s(\underset{\sim}{y}^{*b})$, $b = 1, \ldots, B$, then it is possible to approximate the bootstrap expectations $\mathrm{E}_{F_{\tilde{\phi}}}[s(\underset{\sim}{y})]$ by the mean $\tilde{\phi}^{*(\cdot)} = \frac{1}{B}\sum_{i=1}^{B}\tilde{\phi}^{*b}$. Therefore, the bootstrap bias estimates based on B replications of $\tilde{\phi}$ are

$$\tilde{B}_F(\tilde{\phi}, \phi) = \tilde{\phi}^{*(\cdot)} - s(\underset{\sim}{y}).$$

By using the bootstrap bias estimate presented above, we arrive at the following bias-corrected, to order $\mathcal{O}(n^{-1})$, estimators:

$$\phi_{\mathrm{boot}} = s(\underset{\sim}{y}) - \tilde{B}_F(\tilde{\phi}, \phi) = 2\tilde{\phi} - \tilde{\phi}^{*(\cdot)}.$$

The corrected estimate ϕ_{boot} were called constant-bias-correcting (CBC) estimates by MacKinnon and Smith (1998).

5.3 Numerical results

In this section, we perform some numerical experiments to observe how the result derived in Section 5.2 works for different sample sizes. We here are interested in comparing, via Monte Carlo simulation, the finite-sample performance of the modified precision parameter comparing to the true Pearson-based and the parametric bootstrap method.

The systematic component used in the numerical exercise is

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \quad i = 1, \ldots, n,$$

where the log link function was used for the lognormal, gamma and inverse gamma models. For the inverse Gaussian model, we set the reciprocal link function. The true values were taken as $\beta_0 = 2, \beta_1 = 3$ and $\beta_2 = 0.5$. The explanatory variables x_1 and x_2 were generated from an exponential distribution with mean equal to $1/3$ and an uniform distribution on $(0, 2)$ of size n, respectively. It can be done by using the inversion method. We emphasize that the values of X were held constant throughout the simulations. We defined the sample sizes as $n = 10, 20, 40, 60$ and 100. The total number of Monte Carlo replications was set at 5000 for each sample size. All simulations were performed using the `R` programming language (see, for instance, `R` Core Team, 2012).

In each of the 5000 replications, we fitted the model and computed the modified precision parameter ($\breve{\phi}$), its Pearson-based ($\tilde{\phi}$) and parametric bootstrap (ϕ_{boot}) versions. The number of bootstrap replications was set to 600. In order to compare the results, we computed, for each estimator, the mean and mean squared error (MSE) of estimates for all sample sizes. Tables 5.1–5.4 present the simulation results for $\phi = 20, 60, 100, 200$ and 1000.

Tables 5.1–5.4 present simulation results for the lognormal, inverse Gaussian, gamma and inverse gamma models. Initially, we note that, for all considered models and values of ϕ, the MSEs of the modified estimators are smaller than those of the original Pearson-based estimator, for all given sample sizes. By this fact, we conclude that the precision parameter is better estimated by the modified estimators than by the Pearson-based estimators. On the other hand, although, for some cases, the bootstrap estimates provide better results for large sample sizes, the modified estimator presents great superiority for small and moderate sample sizes. This phenomenon is observed in all models. Moreover, the bootstrap MSEs are very close to the MSE of the modified estimator, and the computation of the parametric bootstrap biases is computer intensive, whereas the modified estimator is not. Therefore, the proposed modification is effective.

Another step is to verify the impact of the proposed modified estimator on the model parameters. For that, we now turn to the evaluation of confidence intervals with nominal coverages $1-\alpha = 0.90$ and 0.99 for each parameter considering the three estimatives ($\hat{\phi}$, $\breve{\phi}$ and ϕ_{boot}). All confidence intervals were defined such that the probability that the true parameter value belongs to the interval is $1-\alpha$, the probability that the true parameter value is smaller than the lower limit of the interval is $\alpha/2$ and the probability that the true value of the parameter is greater than the upper limit of the interval is $\alpha/2$, for $0 < \alpha < 1/2$. Tables 5.5 - 5.8 contain the confidence intervals constructed from 5000 replications, for sample sizes $n = 10, 20, 40, 60, 80$ and 100. An important advantage that the modified estimator has over the others, is that the confidence intervals induced by the modified estimator has smaller lengths in comparison with the confidence intervals.

Table 5.1: Empirical means and mean squared errors (in parentheses) for the lognormal model.

ϕ	Scheme	$n=10$	$n=20$	$n=40$	$n=60$	$n=100$
20	$\tilde{\phi}$	28.8564 (574.06)	23.2527 (95.3690)	21.4939 (33.2830)	21.0774 (20.2240)	20.7005 (11.2170)
	$\breve{\phi}$	20.6117 (253.24)	20.5171 (66.2800)	20.3321 (27.8950)	20.3378 (17.8630)	20.2737 (10.3630)
	ϕ_{boot}	16.2350 (200.57)	19.5138 (64.4910)	19.8911 (27.6500)	20.0425 (17.7020)	20.1013 (10.3170)
60	$\tilde{\phi}$	86.1223 (6766)	68.3346 (787.30)	63.7242 (247.88)	62.7463 (166.89)	61.7247 (90.4200)
	$\breve{\phi}$	61.5160 (3106)	60.2952 (558.95)	60.2796 (221.81)	60.5446 (148.66)	60.4521 (84.0810)
	ϕ_{boot}	50.6719 (2333)	58.5641 (544.39)	59.7283 (220.63)	60.1997 (148.16)	60.2588 (84.2180)
100	$\tilde{\phi}$	139.74 (13406)	114.38 (2269)	105.35 (709.71)	103.46 (422.50)	101.85 (234.68)
	$\breve{\phi}$	99.8178 (6034)	100.93 (1607)	99.6568 (609.54)	99.8342 (382.22)	99.7521 (221.87)
	ϕ_{boot}	82.7706 (4675)	98.4285 (1556)	98.9649 (607.17)	99.4564 (382.81)	99.5337 (222.56)
200	$\tilde{\phi}$	285.57 (64944)	227.82 (9149)	211.73 (2876)	207.39 (1673)	204.55 (928.64)
	$\breve{\phi}$	203.98 (29414)	201.02 (6521)	200.28 (2451)	200.11 (1506)	200.33 (870.98)
	ϕ_{boot}	170.27 (22368)	196.81 (6317)	199.28 (2444)	199.53 (1504)	200.08 (873.08)
1000	$\tilde{\phi}$	1407 (1435631)	1127 (208646)	1061 (70940)	1033 (41116)	1020 (23215)
	$\breve{\phi}$	1005 (647896)	995.24 (149720)	1004 (60089)	997.04 (37258)	999.33 (21870)
	ϕ_{boot}	842.03 (492631)	976.68 (145371)	1000 (59860)	995.58 (37325)	998.75 (21888)

Table 5.2: Empirical means and mean squared errors (in parentheses) for the inverse Gaussian model.

ϕ	Scheme	$n = 10$	$n = 20$	$n = 40$	$n = 60$	$n = 100$
20	$\tilde{\phi}$	29.8287 (665.39)	22.7279 (81.3944)	20.8347 (29.7146)	20.5684 (14.2895)	20.3233 (8.2507)
	$\breve{\phi}$	21.3062 (291.90)	20.0540 (57.5788)	19.7085 (26.0505)	19.8467 (13.0270)	19.9043 (7.8228)
	ϕ_{boot}	17.5704 (263.94)	19.4785 (58.4670)	19.5625 (26.7176)	19.7587 (13.3772)	19.8648 (8.0879)
60	$\tilde{\phi}$	83.5283 (4171)	68.0762 (820.63)	63.9492 (245.94)	62.6698 (157.65)	60.8253 (84.3160)
	$\breve{\phi}$	59.6631 (1846)	60.0672 (588.12)	60.4925 (206.36)	60.4709 (140.36)	59.5712 (80.4051)
	ϕ_{boot}	49.7992 (1799)	58.9642 (596.27)	60.1954 (209.41)	60.2092 (142.44)	59.4888 (81.6685)
100	$\tilde{\phi}$	130.45 (8709)	115.90 (2381)	104.79 (669.57)	103.59 (410.85)	102.82 (208.11)
	$\breve{\phi}$	93.1804 (4017)	102.27 (1662)	99.1280 (579.35)	99.9547 (370.53)	100.70 (192.48)
	ϕ_{boot}	79.6873 (4398)	100.34 (1739)	98.1418 (583.88)	99.6447 (381.20)	100.57 (198.67)
200	$\tilde{\phi}$	286.82 (68230)	225.04 (8870)	212.79 (2701)	207.97 (1668)	203.19 (878.27)
	$\breve{\phi}$	204.87 (30989)	198.56 (6420)	201.29 (2272)	200.67 (1494)	199.00 (833.65)
	ϕ_{boot}	173.69 (30585)	194.94 (6606)	201.38 (2296)	200.56 (1507)	198.89 (842.49)
1000	$\tilde{\phi}$	1430 (1350082)	1148 (223342)	1069 (80918)	1034 (39495)	1018 (19531)
	$\breve{\phi}$	1021 (594621)	1013 (156910)	1011 (68197)	998.41 (35653)	997.17 (18426)
	ϕ_{boot}	848.73 (496510)	991.62 (153001)	1010 (70552)	995.24 (36007)	995.57 (18703)

Table 5.3: Empirical means and mean squared errors (in parentheses) for the gamma model.

ϕ	Scheme	$n = 10$	$n = 20$	$n = 40$	$n = 60$	$n = 100$
20	$\tilde{\phi}$	27.8171 (468.71)	22.7339 (85.0776)	21.2704 (30.2228)	20.7840 (17.6732)	20.4582 (9.5883)
	$\breve{\phi}$	19.8694 (207.98)	20.0594 (60.4212)	20.1206 (25.6142)	20.0547 (15.8854)	20.0364 (8.9969)
	ϕ_{boot}	16.6226 (162.86)	19.5557 (59.1499)	19.9629 (25.4913)	19.9596 (15.8518)	19.9843 (8.9971)
60	$\tilde{\phi}$	85.9499 (6099)	68.4750 (808.96)	63.4458 (268.77)	62.0315 (152.95)	61.1101 (79.3470)
	$\breve{\phi}$	61.3928 (2770)	60.4191 (574.07)	60.0162 (229.87)	59.8550 (138.59)	59.8501 (74.9491)
	ϕ_{boot}	51.5405 (2069)	59.2205 (555.83)	59.7254 (228.69)	59.6980 (138.80)	59.7881 (74.9527)
100	$\tilde{\phi}$	138.69 (12213)	113.19 (2001)	105.98 (727.74)	104.06 (434.99)	102.14 (228.01)
	$\breve{\phi}$	99.0652 (5468)	99.8750 (1422)	100.25 (619.23)	100.41 (389.79)	100.04 (214.29)
	ϕ_{boot}	83.1933 (4264)	97.9658 (1381)	99.8159 (617.34)	100.18 (387.92)	99.9528 (214.25)
200	$\tilde{\phi}$	281.06 (56322)	227.75 (8839)	211.55 (2772)	207.04 (1705)	204.55 (941.40)
	$\breve{\phi}$	200.76 (25384)	200.96 (6283)	200.11 (2361)	199.77 (1542)	200.33 (883.25)
	ϕ_{boot}	168.60 (18874)	197.28 (6103)	199.40 (2358)	199.39 (1540)	200.21 (883.65)
1000	$\tilde{\phi}$	1406 (1825164)	1128 (211485)	1052 (68277)	1035 (41367)	1018 (23268)
	$\breve{\phi}$	1004 (846977)	995.39 (151895)	995.93 (58614)	999.24 (37338)	997.92 (21979)
	ϕ_{boot}	842.81 (625975)	977.08 (147072)	992.54 (58439)	997.99 (37328)	997.45 (22061)

Table 5.4: Empirical means and mean squared errors (in parentheses) for the inverse gamma model.

ϕ	Scheme	$n = 10$	$n = 20$	$n = 40$	$n = 60$	$n = 100$
20	$\tilde{\phi}$	30.4941 (684.36)	24.0735 (121.44)	22.0650 (42.9141)	21.2897 (24.0600)	20.8374 (13.6660)
	$\check{\phi}$	21.7815 (296.15)	21.2414 (83.1656)	20.8723 (35.3454)	20.5427 (21.1470)	20.4078 (12.6019)
	ϕ_{boot}	15.9675 (231.00)	19.4838 (79.6613)	20.0314 (34.5121)	19.9745 (20.9068)	20.0533 (12.4849)
60	$\tilde{\phi}$	86.0169 (5051)	69.3353 (841.40)	64.1442 (287.87)	62.7063 (172.44)	61.7765 (94.5649)
	$\check{\phi}$	61.4407 (2233)	61.1782 (588.60)	60.6770 (242.68)	60.5061 (153.98)	60.5027 (87.9312)
	ϕ_{boot}	49.2257 (1698)	58.7376 (568.69)	59.7042 (241.02)	59.8940 (153.48)	60.1533 (87.8196)
100	$\tilde{\phi}$	142.39 (15507)	114.32 (2251)	106.53 (777.96)	104.05 (447.63)	102.00 (238.87)
	$\check{\phi}$	101.71 (6998)	100.87 (1593)	100.77 (658.60)	100.40 (401.63)	99.8965 (225.30)
	ϕ_{boot}	82.9138 (5259)	97.7712 (1546)	99.6890 (655.32)	99.7520 (400.71)	99.5288 (225.94)
200	$\tilde{\phi}$	279.77 (50629)	225.57 (8950)	212.62 (2969)	207.22 (1765)	204.75 (949.44)
	$\check{\phi}$	199.84 (22584)	199.01 (6460)	201.13 (2516)	199.95 (1595)	200.53 (889.34)
	ϕ_{boot}	165.81 (17299)	194.10 (6287)	199.72 (2504)	199.17 (1592)	200.15 (889.92)
1000	$\tilde{\phi}$	1391 (1211563)	1138 (221281)	1052 (72199)	1034 (42606)	1025 (23791)
	$\check{\phi}$	993.90 (539996)	1004 (157416)	995.43 (62177)	998.25 (38560)	1004 (22222)
	ϕ_{boot}	831.95 (419369)	985.07 (152770)	991.48 (61924)	996.42 (38510)	1003 (22232)

5.4 Concluding remarks

We reduce the bias of the Pearson-based precision parameter estimation for a large class of regression models. Our results holds for generalized linear models (see McCullagh and Nelder, 1989), quasi-likelihood models (see Wedderburn, 1974), among others. We used simulation to conclude the superiority of the modified moment estimator over the usual and bootstrap methods, with regard to both bias reduction and mean square error. Further, an important advantage that the modified estimator has over the others, is that the confidence intervals induced by the modified estimator has smaller lengths in comparison with the confidence intervals. For future research, we can apply the result obtained in Section 2 for a general class of beta regression models.

Table 5.5: Empirical confidence intervals for the parameters of the lognormal regression parameters.

Parameter	Scheme	$n = 10$		$n = 20$		$n = 40$	
		90%	99%	90%	99%	90%	99%
β_0	$\tilde{\phi}$	[1.7084, 2.3047]	[1.5453, 2.4702]	[1.8090, 2.1928]	[1.7013, 2.3006]	[1.8628, 2.1330]	[1.7978, 2.2140]
	$\check{\phi}$	[1.7264, 2.2739]	[1.5009, 2.4842]	[1.8391, 2.1570]	[1.7066, 2.2816]	[1.8885, 2.1046]	[1.8097, 2.1954]
	ϕ_{boot}	[1.6838, 2.3165]	[1.4256, 2.5596]	[1.8344, 2.1617]	[1.6982, 2.2900]	[1.8872, 2.1059]	[1.8075, 2.1976]
β_1	$\tilde{\phi}$	[2.7093, 3.2829]	[2.5928, 3.4353]	[2.8317, 3.1607]	[2.7372, 3.2686]	[2.8890, 3.1087]	[2.8287, 3.1785]
	$\check{\phi}$	[2.7304, 3.2573]	[2.5601, 3.4583]	[2.8586, 3.1308]	[2.7443, 3.2541]	[2.9101, 3.0858]	[2.8390, 3.1638]
	ϕ_{boot}	[2.7004, 3.2872]	[2.5096, 3.5087]	[2.8564, 3.1330]	[2.7402, 3.2582]	[2.9096, 3.0863]	[2.8382, 3.1646]
β_2	$\tilde{\phi}$	[0.3681, 0.6276]	[0.2973, 0.7082]	[0.4103, 0.5851]	[0.3673, 0.6373]	[0.4390, 0.5587]	[0.4046, 0.5920]
	$\check{\phi}$	[0.3785, 0.6172]	[0.2838, 0.7217]	[0.4253, 0.5701]	[0.3726, 0.6319]	[0.4509, 0.5468]	[0.4113, 0.5852]
	ϕ_{boot}	[0.3650, 0.6306]	[0.2590, 0.7465]	[0.4242, 0.5712]	[0.3705, 0.6340]	[0.4506, 0.5470]	[0.4108, 0.5857]

Parameter	Scheme	$n = 60$		$n = 80$		$n = 100$	
		90%	99%	90%	99%	90%	99%
β_0	$\tilde{\phi}$	[1.8890, 2.1036]	[1.8332, 2.1712]	[1.9021, 2.0862]	[1.8523, 2.1428]	[1.9158, 2.0814]	[1.8733, 2.1320]
	$\check{\phi}$	[1.9104, 2.0804]	[1.8449, 2.1552]	[1.9209, 2.0661]	[1.8632, 2.1288]	[1.9330, 2.0632]	[1.8835, 2.1193]
	ϕ_{boot}	[1.9098, 2.0810]	[1.8437, 2.1564]	[1.9205, 2.0665]	[1.8624, 2.1295]	[1.9327, 2.0635]	[1.8829, 2.1199]
β_1	$\tilde{\phi}$	[2.9109, 3.0837]	[2.8678, 3.1378]	[2.9223, 3.0716]	[2.8826, 3.1165]	[2.9335, 3.0656]	[2.8991, 3.1056]
	$\check{\phi}$	[2.9282, 3.0651]	[2.8772, 3.1254]	[2.9375, 3.0553]	[2.8915, 3.1052]	[2.9471, 3.0511]	[2.9072, 3.0956]
	ϕ_{boot}	[2.9280, 3.0653]	[2.8768, 3.1257]	[2.9374, 3.0555]	[2.8913, 3.1054]	[2.9470, 3.0512]	[2.9070, 3.0957]
β_2	$\tilde{\phi}$	[0.4518, 0.5479]	[0.4260, 0.5763]	[0.4600, 0.5424]	[0.4338, 0.5646]	[0.4626, 0.5377]	[0.4408, 0.5571]
	$\check{\phi}$	[0.4618, 0.5380]	[0.4321, 0.5702]	[0.4686, 0.5338]	[0.4394, 0.5591]	[0.4706, 0.5297]	[0.4459, 0.5520]
	ϕ_{boot}	[0.4617, 0.5381]	[0.4319, 0.5704]	[0.4685, 0.5338]	[0.4392, 0.5592]	[0.4705, 0.5298]	[0.4458, 0.5521]

Table 5.6: Empirical confidence intervals for the parameters of the inverse gaussian regression parameters.

Parameter	Scheme	$n = 10$		$n = 20$		$n = 40$	
		90%	99%	90%	99%	90%	99%
β_0	$\tilde{\phi}$	[1.7311, 2.3268]	[1.5626, 2.4954]	[1.8137, 2.2069]	[1.7025, 2.3181]	[1.8684, 2.1354]	[1.7962, 2.2086]
	$\check{\phi}$	[1.7545, 2.3034]	[1.5307, 2.5272]	[1.8474, 2.1732]	[1.7146, 2.3060]	[1.8950, 2.1088]	[1.8110, 2.1939]
	ϕ_{boot}	[1.7274, 2.3305]	[1.4816, 2.5763]	[1.8456, 2.1750]	[1.7114, 2.3092]	[1.8947, 2.1091]	[1.8104, 2.1944]
β_1	$\tilde{\phi}$	[2.4226, 3.5866]	[2.0932, 3.9161]	[2.6398, 3.3267]	[2.4454, 3.5211]	[2.7610, 3.2250]	[2.6405, 3.3719]
	$\check{\phi}$	[2.4686, 3.5407]	[2.0316, 3.9777]	[2.6985, 3.2680]	[2.4664, 3.5002]	[2.8072, 3.1788]	[2.6666, 3.3458]
	ϕ_{boot}	[2.4166, 3.5926]	[1.9372, 4.0720]	[2.6956, 3.2709]	[2.4612, 3.5053]	[2.8067, 3.1794]	[2.6657, 3.3467]
β_2	$\tilde{\phi}$	[0.2649, 0.7532]	[0.1267, 0.8914]	[0.3453, 0.6571]	[0.2570, 0.7454]	[0.3931, 0.6122]	[0.3334, 0.6724]
	$\check{\phi}$	[0.2842, 0.7339]	[0.1008, 0.9172]	[0.3719, 0.6304]	[0.2666, 0.7358]	[0.4150, 0.5904]	[0.3456, 0.6603]
	ϕ_{boot}	[0.2626, 0.7554]	[0.0618, 0.9563]	[0.3706, 0.6318]	[0.2642, 0.7382]	[0.4147, 0.5906]	[0.3452, 0.6606]

Parameter	Scheme	$n = 60$		$n = 80$		$n = 100$	
		90%	99%	90%	99%	90%	99%
β_0	$\tilde{\phi}$	[1.8975, 2.1132]	[1.8328, 2.1713]	[1.9080, 2.0925]	[1.8560, 2.1470]	[1.9155, 2.0818]	[1.8686, 2.1283]
	$\check{\phi}$	[1.9199, 2.0909]	[1.8465, 2.1576]	[1.9274, 2.0730]	[1.8683, 2.1347]	[1.9332, 2.0641]	[1.8800, 2.1169]
	ϕ_{boot}	[1.9197, 2.0910]	[1.8463, 2.1578]	[1.9274, 2.0731]	[1.8682, 2.1349]	[1.9331, 2.0642]	[1.8800, 2.1170]
β_1	$\tilde{\phi}$	[2.8084, 3.1846]	[2.7121, 3.2913]	[2.8407, 3.1594]	[2.7415, 3.2462]	[2.8578, 3.1449]	[2.7824, 3.2293]
	$\check{\phi}$	[2.8473, 3.1456]	[2.7355, 3.2679]	[2.8742, 3.1259]	[2.7630, 3.2246]	[2.8884, 3.1143]	[2.8019, 3.2098]
	ϕ_{boot}	[2.8471, 3.1458]	[2.7351, 3.2683]	[2.8741, 3.1260]	[2.7628, 3.2248]	[2.8883, 3.1144]	[2.8018, 3.2099]
β_2	$\tilde{\phi}$	[0.4120, 0.5873]	[0.3605, 0.6339]	[0.4261, 0.5762]	[0.3872, 0.6202]	[0.4309, 0.5644]	[0.3965, 0.6083]
	$\check{\phi}$	[0.4301, 0.5692]	[0.3715, 0.6228]	[0.4419, 0.5604]	[0.3971, 0.6103]	[0.4451, 0.5502]	[0.4058, 0.5990]
	ϕ_{boot}	[0.4300, 0.5693]	[0.3714, 0.6230]	[0.4418, 0.5604]	[0.3970, 0.6104]	[0.4451, 0.5502]	[0.4057, 0.5991]

Table 5.7: Empirical confidence intervals for the parameters of the gamma regression parameters.

Parameter	Scheme	$n = 10$		$n = 20$		$n = 40$	
		90%	99%	90%	99%	90%	99%
β_0	$\tilde{\phi}$	[1.8278, 2.1841]	[1.7301, 2.2883]	[1.8885, 2.1134]	[1.8263, 2.1785]	[1.9259, 2.0798]	[1.8830, 2.1240]
	$\check{\phi}$	[1.8400, 2.1675]	[1.7064, 2.3010]	[1.9069, 2.0931]	[1.8310, 2.1691]	[1.9408, 2.0640]	[1.8906, 2.1141]
	ϕ_{boot}	[1.8238, 2.1836]	[1.6771, 2.3303]	[1.9059, 2.0941]	[1.8292, 2.1708]	[1.9407, 2.0641]	[1.8903, 2.1144]
β_1	$\tilde{\phi}$	[2.7300, 3.2868]	[2.5749, 3.4474]	[2.8323, 3.1695]	[2.7390, 3.2672]	[2.8868, 3.1107]	[2.8248, 3.1754]
	$\check{\phi}$	[2.7505, 3.2626]	[2.5417, 3.4713]	[2.8597, 3.1391]	[2.7458, 3.2530]	[2.9082, 3.0875]	[2.8351, 3.1607]
	ϕ_{boot}	[2.7256, 3.2875]	[2.4966, 3.5165]	[2.8581, 3.1408]	[2.7428, 3.2560]	[2.9080, 3.0877]	[2.8347, 3.1610]
β_2	$\tilde{\phi}$	[0.3625, 0.6378]	[0.2846, 0.7158]	[0.4116, 0.5842]	[0.3627, 0.6330]	[0.4398, 0.5595]	[0.4059, 0.5934]
	$\check{\phi}$	[0.3734, 0.6268]	[0.2702, 0.7301]	[0.4264, 0.5694]	[0.3681, 0.6277]	[0.4517, 0.5476]	[0.4126, 0.5867]
	ϕ_{boot}	[0.3608, 0.6394]	[0.2472, 0.7530]	[0.4256, 0.5702]	[0.3667, 0.6291]	[0.4516, 0.5477]	[0.4124, 0.5869]

Parameter	Scheme	$n = 60$		$n = 80$		$n = 100$	
		90%	99%	90%	99%	90%	99%
β_0	ϕ	[1.9368, 2.0621]	[1.9018, 2.0980]	[1.9470, 2.0541]	[1.9171, 2.0848]	[1.9511, 2.0466]	[1.9243, 2.0739]
	$\check{\phi}$	[1.9495, 2.0488]	[1.9090, 2.0893]	[1.9581, 2.0427]	[1.9236, 2.0771]	[1.9611, 2.0363]	[1.9304, 2.0670]
	ϕ_{boot}	[1.9495, 2.0489]	[1.9089, 2.0894]	[1.9581, 2.0427]	[1.9236, 2.0771]	[1.9611, 2.0363]	[1.9304, 2.0670]
β_1	$\tilde{\phi}$	[2.9159, 3.0893]	[2.8678, 3.1393]	[2.9231, 3.0737]	[2.8811, 3.1171]	[2.9327, 3.0649]	[2.8959, 3.1030]
	$\check{\phi}$	[2.9333, 3.0707]	[2.8773, 3.1267]	[2.9385, 3.0573]	[2.8900, 3.1058]	[2.9465, 3.0504]	[2.9041, 3.0928]
	ϕ_{boot}	[2.9333, 3.0707]	[2.8772, 3.1268]	[2.9384, 3.0574]	[2.8900, 3.1058]	[2.9464, 3.0504]	[2.9041, 3.0928]
β_2	$\tilde{\phi}$	[0.4517, 0.5482]	[0.4243, 0.5755]	[0.4592, 0.5432]	[0.4354, 0.5670]	[0.4648, 0.5386]	[0.4439, 0.5594]
	$\check{\phi}$	[0.4616, 0.5382]	[0.4304, 0.5694]	[0.4681, 0.5344]	[0.4410, 0.5614]	[0.4726, 0.5307]	[0.4490, 0.5544]
	ϕ_{boot}	[0.4616, 0.5382]	[0.4303, 0.5695]	[0.4680, 0.5344]	[0.4500, 0.5614]	[0.4726, 0.5307]	[0.4489, 0.5544]

Table 5.8: Empirical confidence intervals for the parameters of the inverse gamma regression parameters.

Parameter	Scheme	$n = 10$		$n = 20$		$n = 40$	
		90%	99%	90%	99%	90%	99%
β_0	$\tilde{\phi}$	[1.8255, 2.1616]	[1.7304, 2.2567]	[1.8993, 2.1152]	[1.8382, 2.1763]	[1.9167, 2.0716]	[1.8728, 2.1154]
	$\breve{\phi}$	[1.8386, 2.1484]	[1.7123, 2.2748]	[1.9177, 2.0968]	[1.8447, 2.1698]	[1.9321, 2.0562]	[1.8815, 2.1067]
	ϕ_{boot}	[1.8128, 2.1743]	[1.6654, 2.3216]	[1.9156, 2.0993]	[1.8401, 2.1745]	[1.9314, 2.0569]	[1.8802, 2.1080]
β_1	$\tilde{\phi}$	[2.7371, 3.2658]	[2.5613, 3.4463]	[2.8221, 3.1593]	[2.7267, 3.2547]	[2.8867, 3.1027]	[2.8255, 3.1639]
	$\breve{\phi}$	[2.7578, 3.2452]	[2.5310, 3.4766]	[2.8509, 3.1305]	[2.7369, 3.2445]	[2.9082, 3.0812]	[2.8376, 3.1518]
	ϕ_{boot}	[2.7129, 3.2901]	[2.5502, 3.4500]	[2.8461, 3.1353]	[2.7283, 3.2531]	[2.9074, 3.0820]	[2.8363, 3.1531]
β_2	$\tilde{\phi}$	[0.3669, 0.6369]	[0.2904, 0.7134]	[0.4132, 0.5884]	[0.3637, 0.6380]	[0.4345, 0.5552]	[0.4003, 0.5893]
	$\breve{\phi}$	[0.3774, 0.6264]	[0.2759, 0.7279]	[0.4282, 0.5735]	[0.3690, 0.6327]	[0.4465, 0.5432]	[0.4071, 0.5826]
	ϕ_{boot}	[0.3567, 0.6589]	[0.3026, 0.7199]	[0.4260, 0.5757]	[0.3649, 0.6368]	[0.4461, 0.5435]	[0.4065, 0.5832]
Parameter	Scheme	$n = 60$		$n = 80$		$n = 100$	
		90%	99%	90%	99%	90%	99%
β_0	$\tilde{\phi}$	[1.9340, 2.0636]	[1.9050, 2.0986]	[1.9456, 2.0549]	[1.9147, 2.0859]	[1.9561, 2.0514]	[1.9291, 2.0784]
	$\breve{\phi}$	[1.9527, 2.0508]	[1.9128, 2.0908]	[1.9572, 2.0434]	[1.9220, 2.0786]	[1.9662, 2.0413]	[1.9356, 2.0719]
	ϕ_{boot}	[1.9523, 2.0512]	[1.9120, 2.0915]	[1.9569, 2.0437]	[1.9215, 2.0790]	[1.9660, 2.0415]	[1.9352, 2.0722]
β_1	$\tilde{\phi}$	[2.9148, 3.0973]	[2.8631, 3.1489]	[2.9194, 3.0688]	[2.8771, 3.1111]	[2.9359, 3.0684]	[2.8984, 3.1059]
	$\breve{\phi}$	[2.9336, 3.0784]	[2.8746, 3.1374]	[2.9351, 3.0530]	[2.8870, 3.1011]	[2.9500, 3.0543]	[2.9075, 3.0969]
	ϕ_{boot}	[2.9333, 3.0787]	[2.8741, 3.1380]	[2.9348, 3.0533]	[2.8866, 3.1016]	[2.9499, 3.0544]	[2.9074, 3.0970]
β_2	$\tilde{\phi}$	[0.4542, 0.5495]	[0.4273, 0.5764]	[0.4557, 0.5401]	[0.4318, 0.5640]	[0.4585, 0.5330]	[0.4374, 0.5541]
	$\breve{\phi}$	[0.4641, 0.5396]	[0.4333, 0.5704]	[0.4646, 0.5312]	[0.4374, 0.5584]	[0.46640.5251]	[0.4425, 0.5490]
	ϕ_{boot}	[0.4639, 0.5398]	[0.4330, 0.5707]	[0.4646, 0.5313]	[0.4374, 0.5585]	[0.4663, 0.5251]	[0.4424, 0.5491]

References

[1] Cordeiro, G.M., McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society: Series B*, **53**, 629–643.

[2] Cordeiro, G.M. and Paula, G.A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika*, **76**, 93–100.

[3] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Annals of Statistics, **7**, 1–26.

[4] Ferrari S.L.P, Cribari-Neto, F. (2004) Beta regression for modeling rates and proportions. *Journal Applied Statistics*, **31**, 799–815.

[5] Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society: Series B*, **49**, 127–162.

[6] Jørgensen, B. (1997). The theory of dispersion models. Chapman & Hall, London.

[7] Lange, K.L., Little, R.J.A., Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, **84**, 881–896.

[8] MacKinnon, J.G., Smith Jr., A.A. (1998). Approximate bias correction in econometrics. Journal of Econometrics, **85**, 205–230.

[9] McCullagh, P., Nelder, J. (1989). Generalized Linear Models, second ed. Chapman & Hall, London.

[10] Nelder, J.A., Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, **74**, 221–32.

[11] R Core Team. (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. `http://www.R-project.org`

[12] Smyth, G.K. (2003). Pearson's goodness of fit statistic as a score test statistic. *Science and Statistics: A Festschrift for Terry Speed*. IMS Lecture Notes - Monograph Series, **40**.

[13] Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, **61**, 439–447.