

The Burr XII power series distributions: A new compounding family

Rodrigo B. Silva and Gauss M. Cordeiro

Universidade Federal de Pernambuco

Abstract. Generalizing lifetime distributions is always precious for applied statisticians. In this paper, we introduce a new family of distributions by compounding the Burr XII and power series distributions. The compounding procedure follows the key idea by Adamidis and Loukas (*Statist. Probab. Lett.* **39** (1998) 35–42) or, more generally, by Chahkandi and Ganjali (*Comput. Statist. Data Anal.* **53** (2009) 4433–4440) and Morais and Barreto-Souza (*Comput. Statist. Data Anal.* **55** (2011) 1410–1425). The proposed family includes as a basic exemplar the Burr XII distribution. We provide some mathematical properties including moments, quantile and generating functions, order statistics and their moments, Kullback–Leibler divergence and Shannon entropy. The estimation of the model parameters is performed by maximum likelihood and the inference under large sample. Two special models of the new family are investigated in details. We illustrate the potential of the new family by means of two applications to real data. It provides better fits to these data than other important lifetime models available in the literature.

1 Introduction

In many practical situations, classical probability distributions do not provide adequate fits to real data. For example, if the data are asymmetric, the normal distribution will not be a good choice. So, several methods for generating new probability distributions by adding one or more parameters has been studied in the statistical literature recently. Among these methods, the compounding of some discrete and important lifetime distributions has been in the vanguard of lifetime modeling. Adamidis and Loukas (1998) pioneered 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 by Kus (2007) and Tahmasbi and Rezaei (2008), respectively. Barreto-Souza et al. (2010) and Lu and Shi (2012) proposed the Weibull–geometric (WG) and Weibull–Poisson (WP) distributions, which naturally extend the EG and EP distributions, respectively. Further, Rodrigues et al. (2011) defined the Weibull negative binomial (WNB) distribution, which includes as sub-models the WG and WP distributions.

Key words and phrases. Burr XII distribution, information matrix, Kullback–Leibler divergence, order statistic, power series distribution.

Received June 2013; accepted November 2013.

In the same way, several families of distributions were proposed by compounding some useful lifetime and power series distributions in the last few years. [Chahkandi and Ganjali \(2009\)](#) introduced the exponential power series (EPS) family of distributions, which contains as special cases the EP, EG and EL distributions. [Morais and Barreto-Souza \(2011\)](#) defined the Weibull power series (WPS) family which includes as sub-models the EPS distributions. The WPS distributions can have an increasing, decreasing and upside down bathtub failure rate function. The generalized exponential power series (GEPS) distributions were proposed by [Mahmoudi and Jafari \(2012\)](#) following the same approach of [Morais and Barreto-Souza \(2011\)](#). In a very recent paper, [Silva et al. \(2013\)](#) studied the extended Weibull power series (EWPS) family, which includes as special models the EPS and WPS distributions.

The Burr XII (BXII) distribution has cumulative distribution function (c.d.f.) and probability density function (p.d.f.) (for $x > 0$) given by

$$G(x; c, k) = 1 - (1 + x^c)^{-k} \quad (1.1)$$

and

$$g(x; c, k) = ckx^{c-1}(1 + x^c)^{-k-1}, \quad (1.2)$$

respectively, where $k > 0$ and $c > 0$ are shape parameters. Its tractability advantage is that the c.d.f. and reliability function have closed-form, which simplify the computation of the percentiles and the likelihood function for censored data. It has algebraic tails which are effective for modeling failures that occur with lesser frequency than those with corresponding models based on exponential tails. Further, it has as special models the logistic and Weibull and it is a very popular distribution for modeling lifetime data and for modeling phenomenon with monotone failure rates. When modeling monotone hazard rates, the Weibull distribution may be an initial choice because of its negatively and positively skewed density shapes. However, it does not provide a reasonable parametric fit for modeling phenomenon with non-monotone failure rates such as the bathtub shaped and unimodal failure rates, which are common in reliability and biological studies. Such bathtub hazard curves have nearly middle portions and the corresponding densities have a positive anti-mode. Unimodal failure rates can be observed in course of a disease whose mortality reaches a peak after some finite period and then declines gradually. Examples of approximately BXII distributed phenomena may be found in flood frequency, software reliability, structural and wind engineering. [Shao \(2004\)](#) discussed maximum likelihood estimation of its parameters and [Shao et al. \(2004\)](#) studied models for extremes based on the BXII distribution with application to flood frequency analysis. According to [Soliman \(2005\)](#), this model covers the curve shape characteristics for a large number of distributions. Its versatility and flexibility turns it quite attractive as a tentative model for lifetime data. For this model, [Wu et al. \(2007\)](#) examined the estimation problems under progressive type II censoring with

random removals, where the number of units removed at each failure time has a discrete uniform distribution and [Silva et al. \(2008\)](#) proposed the log-Burr XII location-scale regression model as a feasible alternative to the log-logistic regression model.

In this paper, we define a new family of *Burr XII power series* (BXIIPS) models obtained by compounding the BXII and power series distributions. The compounding procedure follows the pioneering work of [Adamidis and Loukas \(1998\)](#). The new family includes as special models the EPS and WPS distributions in addition to those which arise as special models of the BXII distribution. The hazard function of the new distribution can be decreasing and upside-down bathtub shaped. We are motivated to study the BXIIPS distributions because of the wide usage of (1.1) and the fact that the current generalization provides means of its continuous extension to still more complex situations. A positive point of the current generalization is that the BXII distribution is a basic exemplar of the proposed family.

Furthermore, the new family is well-motivated for industrial applications and biological studies. As a first example, consider the time to relapse of cancer under the first-activation scheme. Suppose that the number, say N , of carcinogenic cells for an individual left active after the initial treatment follows a power series distribution and let X_i be the time spent for the i th carcinogenic cell to produce a detectable cancer mass, for $i \geq 1$. If $\{X_i\}_{i \geq 1}$ is a sequence of i.i.d. BXII random variables independent of N , then the time to relapse of cancer of a susceptible individual can be modeled by the BXIIPS family of distributions. Another example considers that the failure of a device occurs due to the presence of an unknown number, say N , of initial defects of the 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 i th defect, for $i \geq 1$. If we assume that the X_i 's are independent and identically distributed (i.i.d.) BXII random variables independent of N , which follows a power series distribution, then the time to the first failure is appropriately modeled by the BXIIPS family. For reliability studies, from $X = \min\{X_i\}_{i=1}^N$ and $Z = \max\{X_i\}_{i=1}^N$, the proposed models can be used in serial and parallel systems with identical components, which appear in many industrial applications and biological organisms. Further, as discussed by [Cooner et al. \(2007\)](#), the first activation scheme may be questionable for certain diseases. Consider that the number N of latent factors that must all be activated by failure follows a power series distribution and assume that X_i represents the time of resistance to a disease manifestation due to the i th latent factor has the BXII distribution. In the last-activation scheme, the failure occurs after all N factors have been activated. So, the new family of distributions is able for modeling the time to the failure under last-activation scheme.

This paper is organized as follows. In Section 2, the new family is defined by mixing the BXII and zero truncated power series distributions, where the mixing procedure was previously proposed by [Morais and Barreto-Souza \(2011\)](#) and [Silva et al. \(2013\)](#). In Section 3, some mathematical properties of the new family are

obtained. In Section 4, the estimation of the model parameters is performed by the method of maximum likelihood. In Sections 5 and 6, we introduce and study two special models of the BXIIPS family. In Section 7, two illustrative examples based on real data are provided. Finally, in Section 8, concluding remarks are addressed.

2 The new family of distributions

The new family of distributions is defined as follows. Given N , let X_1, \dots, X_N be i.i.d. random variables having the BXII distribution (1.1) with shape parameters $c > 0$ and $k > 0$, where N is a discrete random variable having a power series probability mass function (p.m.f.) (truncated at zero) given by

$$p_n = P(N = n) = \frac{a_n \theta^n}{C(\theta)}, \quad n = 1, 2, \dots \tag{2.1}$$

The coefficients a_n 's depend only on n , $C(\theta) = \sum_{n=1}^{\infty} a_n \theta^n$ and $\theta > 0$ is such that $C(\theta)$ is finite. It is important remark that the probability distributions of the form (2.1) have been considered in Boehme and Powell (1968) and Ostrovska (2007). In Table 1, we give some power series distributions (truncated at zero) defined by (2.1) such as the Poisson, logarithmic, geometric and binomial distributions. Let $X_{(1)} = \min\{X_i\}_{i=1}^N$. The conditional cumulative distribution of $X_{(1)}|N = n$ is given by

$$G_{X_{(1)}|N=n}(x) = 1 - (1 + x^c)^{-nk}$$

that is, $X_{(1)}|N = n$ has the BXII distribution with parameters c and nk . Hence, we obtain

$$P(X_{(1)} \leq x, N = n) = \frac{a_n \theta^n}{C(\theta)} [1 - (1 + x^c)^{-nk}], \quad x > 0, n \geq 1.$$

So, the marginal c.d.f. of $X_{(1)}$ reduces to

$$F(x; \theta, c, k) = 1 - \frac{C[\theta(1 + x^c)^{-k}]}{C(\theta)}, \quad x > 0. \tag{2.2}$$

We call (2.2) the BXIIPS family of distributions.

Table 1 Useful quantities for some power series distributions

Distribution	a_n	$C(\theta)$	$C'(\theta)$	$C''(\theta)$	$C^{-1}(\theta)$	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	$\binom{m}{n}$	$(\theta + 1)^m - 1$	$m(\theta + 1)^{m-1}$	$\frac{m(m-1)}{(\theta+1)^{2-m}}$	$(\theta - 1)^{1/m} - 1$	$\theta \in (0, 1)$

The BXIIPS survival function becomes

$$S(x; \theta, c, k) = \frac{C[\theta(1+x^c)^{-k}]}{C(\theta)}, \quad x > 0. \quad (2.3)$$

The p.d.f. corresponding to (2.2) (for $x > 0$) is given by

$$f(x; \theta, c, k) = \theta c k x^{c-1} (1+x^c)^{-k-1} \frac{C'[\theta(1+x^c)^{-k}]}{C(\theta)}. \quad (2.4)$$

Hereafter, we denote a random variable X with density function (2.4) by $X \sim \text{BXIIPS}(\theta, c, k)$. The hazard rate function of X reduces to

$$h(x; \theta, c, k) = \theta c k x^{c-1} (1+x^c)^{-k-1} \frac{C'[\theta(1+x^c)^{-k}]}{C[\theta(1+x^c)^{-k}]}, \quad x > 0. \quad (2.5)$$

In Sections 5 and 6, we verify that the hazard rate function (2.5) can be decreasing and upside-down bathtub shaped for two special BXIIPS models. The BXIIPS family can be very useful in modeling lifetime data and practitioners may be interested in using some distributions of the proposed family of models.

3 General properties

3.1 A useful expansion

The following proposition reveals that the new family has the BXII distribution as a limiting distribution, whereas Proposition 3 provides a useful expansion for the density function (2.4).

Proposition 1. *The Burr XII distribution with parameters c and k is a limiting case of the BXIIPS family of distributions when $\theta \rightarrow 0^+$.*

Proof. The proof is given in the Appendix A. □

Consider now the convenient re-parametrization $(c, k) \rightarrow (\rho, \beta)$ with $\rho = k^{-1}$ and $\beta = k^{-1/c}$, for which the kernel of the BXII distribution (1.1) can be expressed as $[1 + \rho(x/\beta)^c]^{-1/\rho}$.

Proposition 2. *The WPS sub-family of distributions with parameters β^{-1} , c and θ is a limiting case of the BXIIPS family of distributions when $\rho \rightarrow 0^+$.*

Proof. If $\rho \rightarrow 0^+$, then

$$\lim_{\rho \rightarrow 0^+} F(x; \theta, \rho, \beta) = 1 - \lim_{\rho \rightarrow 0^+} \frac{C[\theta(1 + \rho(x/\beta)^c)^{-1/\rho}]}{C(\theta)} = 1 - \frac{C(\theta e^{-(x/\beta)^c})}{C(\theta)}.$$

So, the BXIIPS distributions converges to WPS distributions for small values of ρ . □

Remark. The EPS distributions can be directly obtained from the WPS distributions when $c = 1$. So, for $\rho \rightarrow 0^+$ and $c = 1$, we obtain the EPS distributions as limiting special models.

Proposition 3. *The BXIIPS density function can be expressed as an infinite mixture of BXII densities with parameters c and nk given by*

$$f(x; \theta, c, k) = \sum_{n=1}^{\infty} p_n g(x; c, nk). \tag{3.1}$$

Proof. The proof is given in the Appendix A. □

So, we can obtain some structural quantities of X such as the moments and generating function from those of the BXII distribution.

3.2 Quantiles and moments

Quantile functions are used in theoretical aspects, statistical applications and Monte Carlo methods. Monte Carlo simulations employ quantile functions to produce simulated random variables for classical and new continuous distributions. The quantile function, say $Q(u)$, of X is given by

$$x = Q(u) = \left\{ \left[\frac{C^{-1}[(1-u)C(\theta)]}{\theta} \right]^{-1/k} - 1 \right\}^{1/c}, \tag{3.2}$$

where u is a uniform random variable on the unit interval $(0, 1)$ and $C^{-1}(\cdot)$ is the inverse function of $C(\cdot)$.

An explicit expression for the s th moment of X follows from Proposition 3, for $s < ck$,

$$\mu'_s = E(X^s) = k \sum_{n=1}^{\infty} n p_n B\left(nk - \frac{s}{c}, 1 + \frac{s}{c}\right). \tag{3.3}$$

The central moments (μ_s) and cumulants (κ_s) of X can be determined from (3.3) as

$$\mu_s = \sum_{k=0}^p \binom{s}{k} (-1)^k \mu_1^s \mu'_{s-k} \quad \text{and} \quad \kappa_s = \mu'_s - \sum_{k=1}^{s-1} \binom{s-1}{k-1} \kappa_k \mu'_{s-k},$$

respectively, where $\kappa_1 = \mu'_1$. For lifetime models, it is of interest the s th incomplete moment of X defined by $T_s(y) = \int_0^y x^s f(x) dx$. The quantity $T_s(y)$ comes from (3.1) as

$$T_s(y) = ck \sum_{n=1}^{\infty} n p_n \int_0^y x^{s+c-1} (1+x)^{-nk-1} dx.$$

Setting $t = (1 + x^c)^{-1}$, we can write

$$T_s(y) = k \sum_{n=1}^{\infty} np_n \int_0^{1/(1+y^c)} t^{nk-s/c-1} (1-t)^{s/c} dt$$

and then

$$T_s(y) = k \sum_{n=1}^{\infty} np_n B_{1/(1+y^c)}(nk - sc^{-1}, sc^{-1} + 1), \tag{3.4}$$

where $B_z(a, b) = \int_0^z t^{a-1} (1-t)^{b-1} dt$ is the incomplete beta function.

An application of the incomplete moments refers to the Lorenz and Bonferroni curves. They are useful in fields like economics, reliability, demography, insurance and medicine. For a given probability π , they are defined by $L(\pi) = T_1(q)/\mu'_1$ and $B(\pi) = T_1(q)/(\pi\mu'_1)$, respectively, where $q = Q(\pi)$ comes from equation (3.2). In economics, if $\pi = F(q)$ is the proportion of units whose income is lower than or equal to q , $L(\pi)$ gives the proportion of total income volume accumulated by the set of units with an income lower than or equal to q . In a similar manner, the Bonferroni curve $B(\pi)$ gives the ratio between the mean income of this group and the mean income of the population.

3.3 Generating function

Let $M_k(t)$ be the moment generating function (m.g.f.) of the BXII(c, k) distribution. The m.g.f. $M(t)$ of X can be obtained from (3.1) as

$$M(t) = \sum_{n=1}^{\infty} p_n M_{nk}(t), \tag{3.5}$$

where $M_{nk}(t)$ is the BXII(c, nk) generating function. For $t < 0$, a simple expression for $M_k(t)$ follows as

$$M_k(t) = ck \int_0^{\infty} e^{tx} x^{c-1} (1+x^c)^{-k-1} dx.$$

This integral can be determined from the Meijer G -function defined by

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

where i is the imaginary unit and L denotes an integration path; see Section 9.3 in Gradshteyn and Ryzhik (2000) for a description of this path. See also Paranaíba et al. (2011, 2012). The Meijer G -function contains as particular cases many integrals with elementary and special functions (Prudnikov et al., 1986). Consider

the following result, which holds for positive integers m and k , $\mu > -1$ and $p > 0$ (Prudnikov et al., 1992, page 21):

$$\begin{aligned}
 I\left(p, \mu, \frac{m}{k}, \nu\right) &= \int_0^\infty \exp(-px)x^\mu(1+x^{m/k})^\nu dx \\
 &= \frac{k^{-\nu}m^{\mu+1/2}}{(2\pi)^{(m-1)/2}\Gamma(-\nu)p^{\mu+1}} \\
 &\quad \times G_{k+m,k}^{k,k+m}\left(\frac{m^m}{p^m} \mid \begin{matrix} \Delta(m, -\mu), \Delta(k, \nu+1) \\ \Delta(k, 0) \end{matrix}\right),
 \end{aligned}
 \tag{3.6}$$

where $\Delta(k, a) = \frac{a}{k}, \frac{a+1}{k}, \dots, \frac{a+k}{k}$. If we assume that $c = m/k$, where m and k are positive integers, we obtain using the integral (3.6),

$$M_k(t) = mI\left(-t, \frac{m}{k} - 1, \frac{m}{k}, -k - 1\right).
 \tag{3.7}$$

This condition is not restrictive since every positive real number can be approximated by a rational number. Hence, for $t < 0$, the generating function of X follows from (3.5) as

$$M(t) = m \sum_{n=1}^\infty p_n I\left(-t, \frac{m}{nk} - 1, \frac{m}{nk}, -nk - 1\right).
 \tag{3.8}$$

For the special cases $c = 1$ and $c = 2$, we obtain simple expressions for $M_k(t)$, and consequently for $M(t)$, using equations (1) (on page 16) and (2) (on page 20) of the book by Prudnikov et al. (1992). For $c = 1$ and $t < 0$, we have

$$M_k(t) = k(-t)^k e^{-t} \Gamma(-k, -t),$$

where $\Gamma(v, x) = \int_x^\infty t^{v-1} e^{-t} dt$ is the complementary incomplete gamma function. For $c = 2$ and $t < 0$, we obtain

$$\begin{aligned}
 M_k(t) &= {}_1F_2\left(1; \frac{1}{2}; 1 - k; \frac{t^2}{4}\right) + \frac{t}{2} B\left(2, k - \frac{1}{2}\right) {}_1F_2\left(1; \frac{3}{2}; k + \frac{7}{2}; \frac{-t^2}{4}\right) \\
 &\quad + \frac{\Gamma(-2k)}{(-t)^{-2k}},
 \end{aligned}$$

where

$${}_1F_2(a; b, c; x) = \sum_{r=0}^\infty \frac{(a)_r}{(b)_r (c)_r} \frac{x^r}{r!}$$

is a generalized hypergeometric function and $(a)_r = a(a + 1) \cdots (a + r - 1)$ denotes the ascending factorial.

3.4 Order statistics

Now, let X_1, \dots, X_m be a random sample with density function (2.4) and $X_{i:m}$ be the i th order statistic. The density function of $X_{i:m}$, say $f_{i:m}(x; \theta, c, k)$, can be expressed from (2.3) as

$$\begin{aligned} f_{i:m}(x; \theta, c, k) &= \frac{1}{B(i, m - i + 1)} f(x; \theta, c, k) F(x; \theta, c, k)^{i-1} [1 - F(x; \theta, c, k)]^{m-i} \quad (3.9) \\ &= \frac{f(x; \theta, c, k)}{B(i, m - i + 1)} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \left[\frac{C[\theta(1+x^c)^{-k}]}{C(\theta)} \right]^{m+j-i}. \end{aligned}$$

Now, we derive a useful expansion for (3.9). First, we obtain an expansion for $C[\theta(1+x^c)^{-k}]^j$

$$\begin{aligned} C[\theta(1+x^c)^{-k}]^j &= \left[\sum_{n=1}^{\infty} a_n \theta^n (1+x^c)^{-nk} \right]^j \\ &= \left\{ a_1 \theta (1+x^c)^{-k} \left[1 + \frac{a_2}{a_1} \theta (1+x^c)^{-k} + \frac{a_3}{a_1} \theta^2 (1+x^c)^{-2k} + \dots \right] \right\}^j \\ &= a_1^j \theta^j (1+x^c)^{-jk} \left[\sum_{m=0}^{\infty} b_m \theta^m (1+x^c)^{-mk} \right]^j, \end{aligned}$$

where $b_m = a_{m+1}/a_1$ for $m = 1, 2, 3, \dots$. Using the identity (Gradshteyn and Ryzhik, 2000) $(\sum_{m=0}^{\infty} b_m z^m)^j = \sum_{m=0}^{\infty} d_{j,m} z^m$ for a positive integer j , we can write

$$C[\theta(1+x^c)^{-k}]^j = a_1^j \sum_{m=0}^{\infty} d_{j,m} \theta^{j+m} (1+x^c)^{-k(j+m)}, \quad (3.10)$$

where $d_{j,0} = 1$ and the coefficients for $t \geq 1$ can be obtained from the recurrence equation $d_{j,t} = t^{-1} \sum_{m=1}^t [m(j+1) - t] b_m d_{j,t-m}$.

Secondly, we derive an expansion for $C'(\theta) = \sum_{n=1}^{\infty} n a_n \theta^{n-1}$ in $f(x; \theta, c, k)$. We have

$$\begin{aligned} C'[\theta(1+x^c)^{-k}] &= \sum_{r=1}^{\infty} r a_r \theta^{r-1} (1+x^c)^{-k(r-1)} \\ &= a_1 \sum_{r=0}^{\infty} (r+1) b_r \theta^r (1+x^c)^{-rk}, \end{aligned} \quad (3.11)$$

where b_r was defined before.

Inserting equations (3.10) and (3.11) in (3.9), we obtain

$$f_{i:m}(x; \theta, c, k) = \frac{1}{B(i, m - i + 1)} \times \sum_{j=0}^{i-1} \sum_{r=0}^{\infty} \delta_{j,m,r} g(x; c, k(j + 2m + r)), \tag{3.12}$$

where $g(x; c, k(j + 2m + r))$ denotes the BXII density function with parameters c and $k(j + 2m + r)$ and

$$\delta_{j,m,r} = \frac{(-1)^j (r + 1) c \theta^{j+2m+r} a_1^{m+j} b_r d_{j,m}}{(j + 2m + r) C(\theta)^{j+m-2}} \binom{i-1}{j}.$$

Equation (3.12) reveals that the density function of the BXIIPS order statistics is a triple linear combination of BXII densities, where the quantities $\delta_{j,m,r}$ depend only on the discrete distribution in the power family. So, some mathematical properties of $X_{i:m}$ can be immediately obtained from those BXII properties.

An explicit expression for the s th moment $X_{i:m}$ can also be obtained from a result due to Barakat and Abdelkader (2004)

$$\begin{aligned} E(X_{i:m}^s) &= s \sum_{j=m-i+1}^m (-1)^{j-m+i-1} \binom{j-1}{m-i} \binom{m}{j} \int_0^\infty x^{s-1} S(x, \theta, c, k)^j dx \\ &= s \sum_{j=m-i+1}^m \frac{(-1)^{j-m+i-1}}{C(\theta)^j} \binom{j-1}{m-i} \binom{m}{j} \int_0^\infty x^{s-1} C[\theta(1+x^c)^{-k}]^j dx, \end{aligned} \tag{3.13}$$

for $i = 1, \dots, m$. Using (3.10) and the $(s - c)$ th moment of the BXII distribution, the last integral (for $s > c$) can be reduced to

$$\int_0^\infty x^{s-1} C[\theta(1+x^c)^{-k}]^j dx = \frac{a_1^j}{c} \sum_{r=0}^{\infty} \frac{\theta^{j+r} d_{j,r}}{[k(j+r) - 1]} B[k(j+r) - s, s].$$

In Sections 5 and 6, explicit expressions for (3.13) are presented for some special cases.

3.5 A characterization

Shannon (1948) introduced the probabilistic definition of entropy following a dual concept in statistical mechanics. The Shannon entropy plays a central role in information theory and sometimes is referred to as a 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. Let X be a random variable with continuous distribution and density function f . Then, the Shannon entropy of X is defined by

$$\mathbb{H}_{\text{Sh}}(f) = - \int_{\mathbb{R}} f(x; \theta, c, k) \log f(x; \theta, c, k) dx. \tag{3.14}$$

Jaynes (1957) pioneered 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 family of density functions

$$\mathbb{F} = \{f(x; \theta, c, k) : E_f(T_i(X)) = \alpha_i, i = 0, \dots, m\},$$

where $T_1(X), \dots, T_m(X)$ 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 (3.14) subject to the information constraints defined in the family \mathbb{F} . Shore and Johnson (1980) treated axiomatically the maximum entropy method. This method has been successfully applied to 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 family \mathbb{F} , denoted by f^{ME} , determined as the solution of the optimization problem

$$f^{\text{ME}}(x; \theta, c, k) = \arg \max_{f \in \mathbb{F}} \mathbb{H}_{\text{Sh}}.$$

Jaynes (1957, page 623) demonstrated 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 family (2.4). For this purpose, the next result plays an important role.

Proposition 4. *If X is a random variable with density function (2.4) and $Y \sim \text{BXII}(c, k)$, we have*

$$\text{C1} \quad E\{\log C'[\theta(1 + X^c)^{-k}]\} = \frac{\theta}{C(\theta)} E\{C'[\theta(1 + Y^c)^{-k}] \log C'[\theta(1 + Y^c)^{-k}]\},$$

$$\text{C2} \quad E[\log(X)] = \frac{\theta}{C(\theta)} E\{C'[\theta(1 + Y^c)^{-k}] \log(Y)\},$$

$$\text{C3} \quad E[\log(1 + X^c)] = \frac{\theta}{C(\theta)} E\{C'[\theta(1 + Y^c)^{-k}] \log(1 + Y^c)\}.$$

Proof. The constraints C1, C2 and C3 are easily demonstrated and the proofs are omitted. \square

In the next proposition, we demonstrate that the BXIIIPS family has maximum entropy in the family of all probability distributions specified by the constraints stated therein.

Proposition 5. *The density function $f(\cdot)$ of a random variable X , given by (2.4), is the unique solution of the optimization problem*

$$f = \arg \max_{h \in \mathbb{F}} \mathbb{H}_{\text{Sh}}(h),$$

under the constraints C1, C2 and C3 given in Proposition 4.

Proof. The proof is given in the Appendix A. □

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

$$\begin{aligned} \mathbb{H}_{\text{Sh}}(f) &= \log C(\theta) - \log(\theta ck) - \frac{\theta(c-1)}{C(\theta)} \mathbb{E}\{C'[\theta(1+Y^c)^{-k}] \log(Y)\} \\ &\quad - \frac{\theta}{C(\theta)} \mathbb{E}\{C'[\theta(1+Y^c)^{-k}] \log C'[\theta(1+Y^c)^{-k}]\} \\ &\quad + \frac{\theta(k-1)}{C(\theta)} \mathbb{E}\{C'[\theta(1+Y^c)^{-k}] \log(1+Y^c)\}, \end{aligned}$$

where Y has p.d.f. given by (1.2).

Some results in this section can be obtained numerically in any symbolic software such as MAPLE, MATLAB, MATHEMATICA, OX and R. 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.

4 Maximum likelihood estimation and inference

The unit log-density of X with observed value x is given by

$$\begin{aligned} \ell &= \ell(\theta, c, k) \\ &= \log(\theta ck) + (c-1) \log x - (k+1) \log(1+x^c) \\ &\quad + \log C'[\theta(1+x^c)^{-k}] - \log C(\theta) \end{aligned}$$

and the corresponding score function is $U = (\partial \ell / \partial \theta, \partial \ell / \partial c, \partial \ell / \partial k)^\top$, where

$$\begin{aligned} U_\theta &= \frac{\partial \ell}{\partial \theta} = \frac{1}{\theta} + C'[\theta(1+x^c)^{-k}]^{-1} \frac{\partial C'[\theta(1+x^c)^{-k}]}{\partial \theta} - \frac{C'(\theta)}{C(\theta)}, \\ U_c &= \frac{\partial \ell}{\partial c} = \frac{1}{c} + \log x - c(k+1) \frac{x^{c-1}}{1+x^c} + C'[\theta(1+x^c)^{-k}]^{-1} \frac{\partial C'[\theta(1+x^c)^{-k}]}{\partial c} \end{aligned}$$

and

$$U_k = \frac{\partial \ell}{\partial k} = \frac{1}{k} - \log(1+x^c) + C'[\theta(1+x^c)^{-k}]^{-1} \frac{\partial C'[\theta(1+x^c)^{-k}]}{\partial k}.$$

For a random sample $x = (x_1, \dots, x_n)$ of size n from X and $\Theta = (\theta, c, k)^\top$, the total log-likelihood for Θ is

$$\ell_n = \ell_n(\Theta) = \sum_{i=1}^n \ell_n^{(i)},$$

where $\ell_n^{(i)}$ is the log-likelihood for the i th observation ($i = 1, \dots, n$) as given before. The total score function is $U_n = U_n(\Theta) = \sum_{i=1}^n U^{(i)}$, where $U^{(i)}$ (for $i = 1, \dots, n$) has the form given before. The observed information matrix is

$$J_n(\Theta) = n \begin{pmatrix} J_{\theta,\theta} & J_{\theta,c} & J_{\theta,k} \\ \cdot & J_{c,c} & J_{c,k} \\ \cdot & \cdot & J_{k,k} \end{pmatrix},$$

whose elements are obtained from standard calculations. They are listed in Appendix C. The maximum likelihood estimator (MLE) $\hat{\Theta}$ of Θ is obtained numerically from the solution of the non-linear system of equations $U_n = 0$.

Often with lifetime data and reliability studies, one encounters censoring. A very simple random censoring mechanism that is 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 $\delta_i = I(X_i \leq C_i)$ is such that $\delta_i = 1$ if X_i is a time to event and $\delta_i = 0$ if it is right censored for $i = 1, \dots, n$. The censored likelihood $L(\Theta)$ for the model parameters is

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

where $f(x_i; \theta, c, k)$ and $S(x_i; \theta, c, k)$ are given by (2.4) and (2.3), respectively.

Under conditions that are fulfilled for the parameter Θ in the interior of the parameter space but not on the boundary, the asymptotic result holds $\sqrt{n}(\hat{\Theta} - \Theta) \overset{A}{\sim} N_3(0, K(\Theta)^{-1})$, where ' $\overset{A}{\sim}$ ' stands for the asymptotic distribution, $K(\Theta) = \lim_{n \rightarrow \infty} n^{-1} J_n(\Theta)$ is the unit information matrix. The approximate multivariate normal distribution $N_3(0, J_n(\hat{\Theta})^{-1})$, where the observed matrix $J_n(\Theta)$ is evaluated at $\hat{\Theta}$, can be used to construct confidence intervals for the model parameters. The well-known likelihood ratio (LR) statistic can be adopted for testing hypotheses on the parameters in the usual way. In particular, this statistic is useful to check if the fit using the BXIIPS distribution is statistically superior to a fit using other distributions for a given data set.

5 The Burr XII Poisson distribution

We consider the composition defined by the BXII (with parameters $c > 0$ and $k > 0$) and zero truncated Poisson (with parameter $\theta > 0$) distributions. This

model is called the *Burr XII Poisson* (BXIIP) distribution. This distribution comes from (2.2) using the function $C(\theta) = e^\theta - 1$, $\theta > 0$, which corresponds to the truncated Poisson distribution. Thus, the BXIIP cumulative function reduces to

$$F(x; \theta, c, k) = 1 - \frac{\exp\{\theta(1 + x^c)^{-k}\} - 1}{e^\theta - 1}, \quad x > 0. \quad (5.1)$$

From the general expressions (2.4) and (2.5), the density and hazard rate functions reduce to

$$f(x; \theta, c, k) = \frac{\theta ck}{e^\theta - 1} x^{c-1} (1 + x^c)^{-k-1} \exp\{\theta(1 + x^c)^{-k}\}, \quad x > 0 \quad (5.2)$$

and

$$h(x; \theta, c, k) = \frac{\theta ck x^{c-1} (1 + x^c)^{-k-1}}{1 - \exp\{-\theta(1 + x^c)^{-k}\}}, \quad x > 0.$$

Here, a random variable T following (5.1) and (5.2) is denoted by $T \sim \text{BXIIP}(\theta, c, k)$. From Propositions 1 and 2, the BXII and WP distributions are limiting special cases of the BXIIP distribution. Further, the EP distribution is obtained directly from the WP distribution for $c = 1$. Plots of the density and hazard rate functions of the BXIIP distribution are displayed in Figure 1 to show its flexibility to model lifetime data.

The s th moment of T has a closed-form expression obtained from (3.3) as

$$E(T^s) = \frac{k}{e^\theta - 1} \sum_{n=1}^{\infty} \frac{\theta^n}{\Gamma(n)} B\left(nk - \frac{s}{c}, 1 + \frac{s}{c}\right), \quad s < ck.$$

The moments of the order statistics $T_{1:m}, \dots, T_{m:m}$ from a random sample of the BXIIP distribution are given by

$$\begin{aligned} E(T_{i:m}^s) &= s \sum_{j=m-i+1}^m \frac{(-1)^{j-m+i-1}}{C(\theta)^j} \binom{j-1}{m-i} \binom{m}{j} \\ &\quad \times \int_0^\infty z^{s-1} \{\exp[\theta(1+z^c)^{-k}] - 1\}^j dz. \end{aligned}$$

They are easily obtained numerically. Further, after some algebraic calculations, the Shannon entropy for the BXIIP distribution reduces to

$$\begin{aligned} \mathbb{H}_{\text{Sh}}(f) &= \log\left(\frac{e^\theta - 1}{\theta ck}\right) - \frac{\theta(c-1)}{e^\theta - 1} E\{C'[\theta(1+Y^c)^{-k}] \log(Y)\} \\ &\quad - \frac{\theta}{e^\theta - 1} E\{C'[\theta(1+Y^c)^{-k}] \log C'[\theta(1+Y^c)^{-k}]\} \\ &\quad + \frac{\theta(k-1)}{e^\theta - 1} E\{C'[\theta(1+Y^c)^{-k}] \log(1+Y^c)\}. \end{aligned}$$

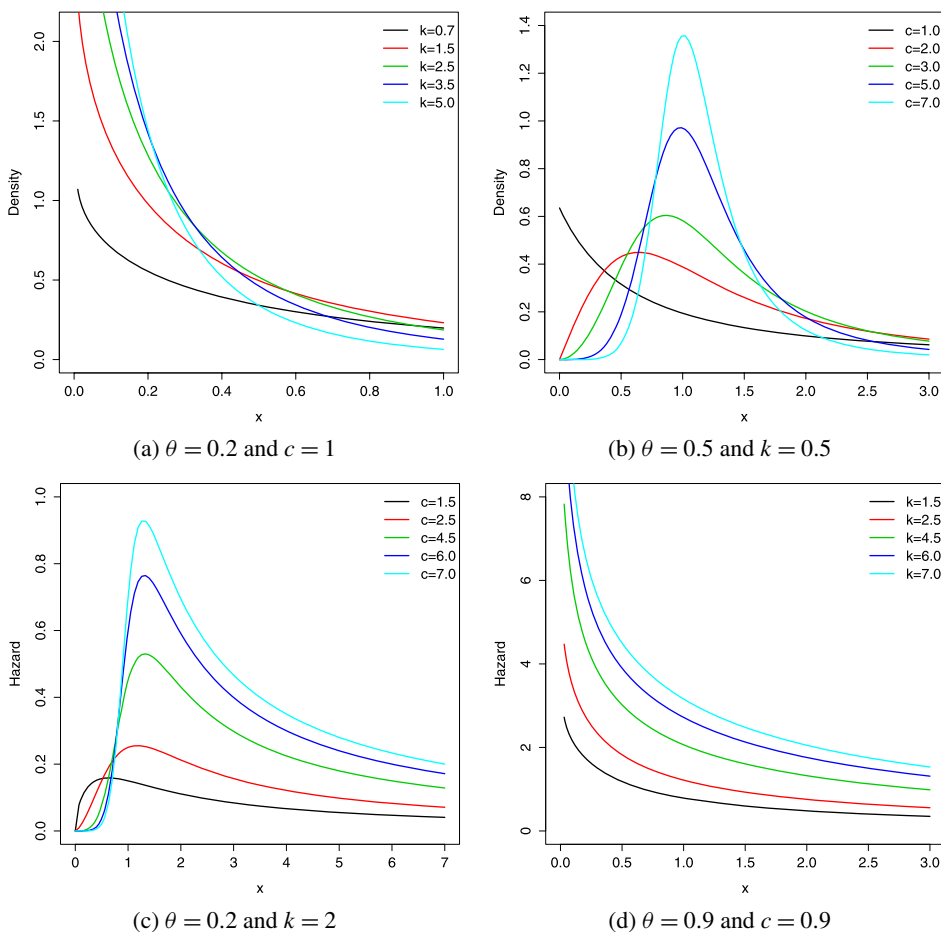


Figure 1 Plots of the BXIIP density and hazard functions for some parameter values.

Setting $u = (1 + y^c)^{-k}$, we obtain

$$E\{C'[\theta(1 + Y^c)^{-k}] \log(Y)\} = \frac{1}{c} \int_0^1 e^{\theta u} \log(u^{-1/k} - 1) du$$

and

$$E\{C'[\theta(1 + Y^c)^{-k}] \log C'[\theta(1 + Y^c)^{-k}]\} = \int_0^1 \theta u e^{\theta u} du = \frac{e^\theta(\theta - 1) + 1}{\theta}.$$

By expanding $e^{\theta u}$ in power series and using (for $a > -1$) $\int_0^1 u^a \log(u) du = -(a + 1)^{-2}$, we have

$$E\{C'[\theta(1 + Y^c)^{-k}] \log(1 + Y^c)\} = -\frac{1}{k} \int_0^1 e^{\theta u} \log(u) du = \frac{1}{k} \sum_{j=0}^{\infty} \frac{\theta^j}{(j + 1)^2 j!}.$$

Hence,

$$\begin{aligned} \mathbb{H}_{\text{Sh}}(f) = & \log\left(\frac{e^\theta - 1}{\theta ck}\right) - \frac{\theta(c-1)}{c(e^\theta - 1)} \int_0^1 e^{\theta u} \log(u^{-1/k} - 1) du \\ & - \frac{e^\theta(\theta - 1) + 1}{e^\theta - 1} + \frac{\theta(k-1)}{e^\theta - 1} \sum_{j=0}^{\infty} \frac{\theta^j}{(j+1)^2 j!}. \end{aligned} \quad (5.3)$$

The integral in (5.3), say J_1 , is determined in the Appendix B as

$$J_1 = -k(a+1) \sum_{k=0}^{\infty} \frac{\theta^n H_{k(n+1)-k^2+1}}{n!} - k^{-1} \sum_{n=0}^{\infty} \frac{\theta^n}{(n+1)^2 n!},$$

here H_{a+1} is the harmonic number.

6 The Burr XII geometric distribution

The *Burr XII geometric* (BXIIG) distribution arises by taking $C(\theta) = \theta(1 - \theta)^{-1}$ corresponding to the geometric distribution in (2.2). We denote a random variable Z with the BXIIG distribution by $T \sim \text{BXIIG}(\theta, c, k)$. The c.d.f. and p.d.f. of T are given by

$$F(x; \theta, c, k) = \frac{1 - (1 + x^c)^{-k}}{1 - \theta(1 + x^c)^{-k}}, \quad x > 0$$

and

$$f(x; \theta, c, k) = \frac{ck(1 - \theta)x^{c-1}(1 + x^c)^{-k-1}}{[1 - \theta(1 + x^c)^{-k}]^2}, \quad x > 0. \quad (6.1)$$

Its hazard rate function becomes

$$h(x; \theta, c, k) = \frac{ckx^{c-1}}{(1 + x^c)[1 - \theta(1 + x^c)^{-k}]}, \quad x > 0.$$

It is quite clear that the BXII distribution is the limiting case when $\theta \rightarrow 0^+$. For β and $\theta \rightarrow 0^+$, we obtain the Weibull distribution as a limiting case. So, the WG distribution is also a special case of the BXIIG distribution when $\beta \rightarrow 0$. The EG distribution is immediately obtained from the WG distribution for $c = 1$. Plots of the BXIIG density and hazard rate functions are displayed in Figure 2.

The s th moment of T is given by

$$\mathbb{E}(T^s) = k(1 - \theta) \sum_{n=1}^{\infty} n\theta^{n-1} B\left(nk - \frac{s}{c}, 1 + \frac{s}{c}\right).$$

An explicit expression for the s th moment of the i th order statistic $T_{i:m}$ is given by

$$\mathbb{E}(T_{i:m}^s) = s \sum_{j=m-i+1}^m \frac{(-1)^{j-m+i-1} \binom{j-1}{m-i} \binom{m}{j}}{C(\theta)^j} \int_0^\infty z^{s-1} \left(\frac{1}{\theta(1+z^c)^k - 1}\right)^j dz.$$

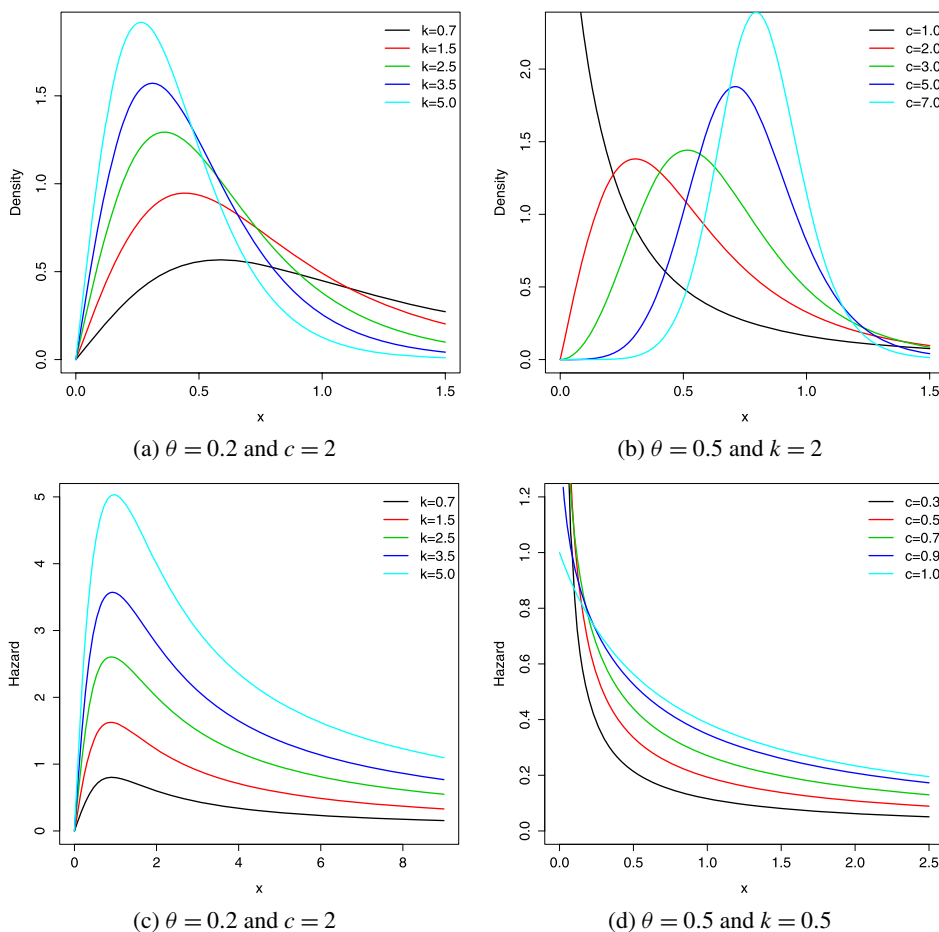


Figure 2 Plots of the BXIIG density and hazard functions for some parameter values.

Further, we obtain an expression for the BXIIG Shannon entropy

$$\begin{aligned}
 \mathbb{H}_{\text{Sh}}(f) &= -\log[ck(1-\theta)] - 2[\theta + \log(1-\theta)] - (1-\theta)(k-1)\log(1-\theta) \\
 &\quad - (1-\theta)(c-1) \int_0^1 \frac{\log(z^{-1/k} - 1)}{(1-\theta z)^2} dz.
 \end{aligned}
 \tag{6.2}$$

The integral in (6.2), say J_2 , is determined in the Appendix B as

$$J_2 = k \sum_{n,j=1}^{\infty} \frac{n\theta^{n-1}}{j(nk+j)} - \frac{\log(1-\theta)}{k\theta}.$$

7 Empirical illustrations

In this section, we compare the fits of some special models of the BXIIPS class by means of two real data sets to show the potential of the new class. 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. We also fit the BXII and WP distributions to make a comparison with the BXIIPS models. The c.d.f. of the WP distribution (for $x > 0$) is given by

$$F(x; \theta, \alpha, \beta) = [e^{\theta \exp(-\beta x^\alpha)} - e^\theta](1 - e^\theta)^{-1},$$

where θ , α and β are positive parameters. For these distributions, we estimate the model parameters and compare the values of the Kolmogorov–Smirnov (K–S) statistic, $-2\ell(\hat{\Theta})$, Akaike information criterion (AIC) and Bayesian information criterion (BIC).

The first data set consists of the 1519 observations of budget share for fuel expenditure of british households. They were drawn from the 1980–1982 British Family Expenditure Surveys (FES) and studied by [Blundell et al. \(1998\)](#) which has been concerned with investigating the ‘shape’ of consumer preferences using semi-parametric methods. In Table 2, we list the MLEs of the parameters (with corresponding standard errors in parentheses), $-2\ell(\hat{\Theta})$, the K–S, AIC and BIC statistics for the BXIIP, BXIIG, BXII and WP models.

Roughly, we conclude that all competing distributions can be used to model these data. However, the K–S test and the current statistics indicate that the BXIIPS distributions outperform the other distributions. Plots of the estimated p.d.f. and c.d.f. of the fitted BXIIP, BXIIG, BXII and WP models to the first data set are displayed in Figure 3, confirming the superiority of the BXIIG model.

Table 2 Parameter estimates, K–S, AIC and BIC statistics for the first data set

Distribution	$\hat{\theta}$	\hat{c}	\hat{k}	$\hat{\alpha}$	$\hat{\beta}$	K–S	$-2\ell(\hat{\Theta})$	AIC	BIC
BXIIP	4.8519 (0.4989)	2.3229 (0.0442)	49.4268 (7.8702)	–	–	0.2216	–5219	–5213	–5197
BXIIG	0.9997 (0.0001)	3.2964 (0.0697)	1.3124 (1.0877)	–	–	0.2161	–5276	–5270	–5254
BXII	–	1.9063 (0.0336)	76.8212 (5.4854)	–	–	0.2617	–5082	–5078	–5067
WP	4.9389 (0.4968)	–	–	2.4126 (0.0446)	47.1347 (7.4834)	0.2278	–5215	–5209	–5193

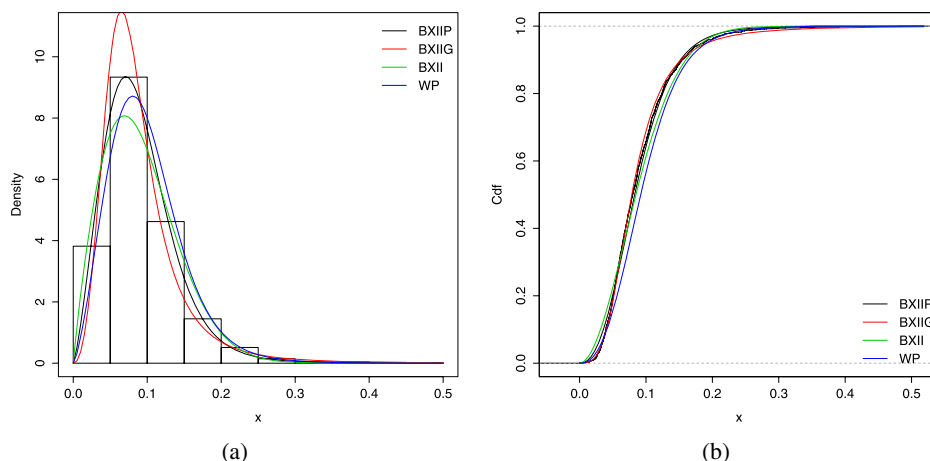


Figure 3 Estimated (a) *p.d.f.* and (b) *c.d.f.* for the BXIIP, BXIIG, BXII and WP models for the first data set.

Table 3 Parameter estimates, *K-S*, *AIC* and *BIC* statistics for the second data set

Distribution	$\hat{\theta}$	\hat{c}	\hat{k}	$\hat{\alpha}$	$\hat{\beta}$	K-S	$-2\ell(\hat{\Theta})$	AIC	BIC
BXIIP	7.5291 (2.2724)	2.4189 (0.3279)	30.8878 (17.9114)	–	–	0.1075	–61.6	–55.6	–52.6
BXIIG	0.9999 (0.0000)	4.1000 (0.9625)	1.2937 (3.3203)	–	–	0.0945	–72.4	–66.4	–63.4
BXII		1.6924 (0.2235)	29.9045 (12.5340)	–	–	0.3087	–53.9	–49.9	–49.2
WP	5.1247 (2.2736)			2.3989 (0.3326)	27.3059 (16.8773)	0.2395	–61.1	–55.1	–52.2

As a second example, we consider a data set from Murthy et al. (2004) consisting of the failure times of 20 mechanical components. The data are: 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 and 0.485. In Table 3, we display the MLEs of the parameters (with corresponding standard errors in parentheses) and the values of the K-S, $-2\ell(\hat{\Theta})$, AIC and BIC statistics for the BXIIP, BXIIG, BXII and WP models. They indicate that the BXIIP and BXIIG distributions are superior to the other distributions in terms of model fitting. From the figures in Table 3, we conclude that the BXIIPs distributions provide better fits to these data than the BXII and WP models. Plots of the estimated *p.d.f.* and *c.d.f.* of the fitted BXIIP, BXIIG, BXII and WP models to the second data set are displayed in Figure 4. They indicate that the BXIIG yields the best fit.

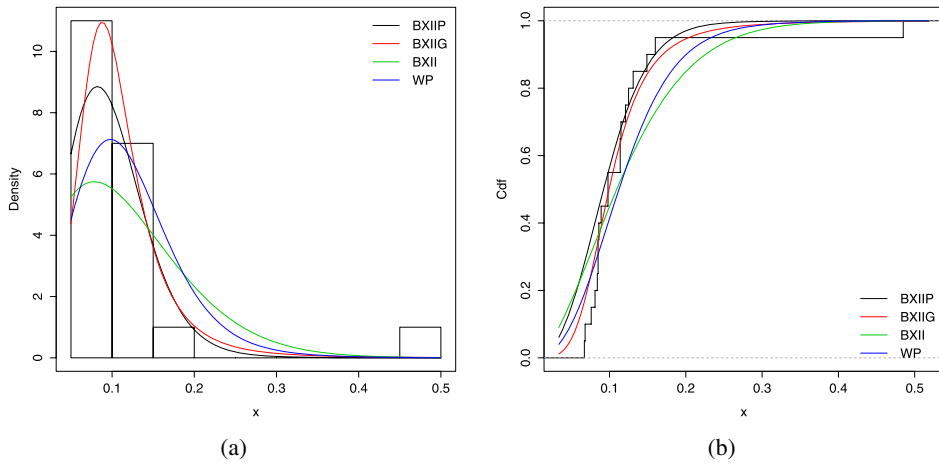


Figure 4 Estimated (a) *p.d.f.* and (b) *c.d.f.* for the BXIIP, BXIIG, BXII and WP models for the second data set.

8 Concluding remarks

We define a new class of lifetime models called the *Burr XII power series* (BXIIPS) family of distributions. It includes the Weibull power series distributions (Morais and Barreto-Souza, 2011) as limiting cases. The BXIIPS density function can be expressed as a mixture of BXII density functions. This property is important to explore several new results for special models. Explicit expressions are derived for the ordinary and incomplete moments, quantile and generating functions, order statistics and their moments and Shannon entropy. The estimation of the model parameters was performed using maximum likelihood. The observed information matrix is determined. We fit the BXIIPS models to two real data sets to show the potential of the proposed family. Finally, we hope that this generalization may attract more complex applications in the literature of lifetime models.

Appendix A: Proofs of Propositions 1, 2 and 3

Proof of Proposition 1. For $x > 0$, we have

$$\begin{aligned}
 \lim_{\theta \rightarrow 0^+} F(x) &= 1 - \lim_{\theta \rightarrow 0^+} \frac{\sum_{n=1}^{\infty} a_n [\theta(1+x^c)^{-k}]^n}{\sum_{n=1}^{\infty} a_n \theta^n} \\
 &= 1 - \lim_{\theta \rightarrow 0^+} \frac{(1+x^c)^{-k} + a_1^{-1} \sum_{n=2}^{\infty} a_n \theta^{n-1} (1+x^c)^{-nk}}{1 + a_1^{-1} \sum_{n=2}^{\infty} a_n \theta^{n-1}} \\
 &= 1 - (1+x^c)^{-k}. \quad \square
 \end{aligned}$$

Proof of Proposition 3. We have $C'(\theta) = \sum_{n=1}^{\infty} na_n\theta^{n-1}$. By using this result in (2.4), we obtain

$$\begin{aligned} f(x; \theta, c, k) &= \frac{ck\theta}{C(\theta)} x^{c-1} (1+x^c)^{-k-1} \sum_{n=1}^{\infty} na_n\theta^n (1+x^c)^{-k(n-1)} \\ &= \sum_{n=1}^{\infty} \frac{a_n\theta^n}{C(\theta)} cnkx^{c-1} (1+x^c)^{-nk-1} = \sum_{n=1}^{\infty} p_n g(x; c, nk), \end{aligned}$$

where $g(x; c, nk)$ is the BXII density function with parameters c and nk . □

Proof of Proposition 5. Let $z(\cdot)$ be a p.d.f. which satisfies the constraints C1, C2 and C3. The Kullback–Leibler divergence between the densities z and f is

$$D(z, f) = \int_{\mathbb{R}} z \log\left(\frac{z}{f}\right) dx.$$

Following Cover and Thomas (1991), we obtain

$$0 \leq D(z, f) = \int_{\mathbb{R}} z \log z \, dx - \int_{\mathbb{R}} z \log f \, dx = -\mathbb{H}_{\text{Sh}}(z) - \int_{\mathbb{R}} z \log f \, dx.$$

Let Y have p.d.f. given by (1.2). From the definition of f and based on the constraints C1, C2 and C3, we have

$$\begin{aligned} \int_{\mathbb{R}} z \log f \, dx &= \log(\theta ck) - \log C(\theta) + \frac{\theta(c-1)}{C(\theta)} \mathbb{E}\{C'[\theta(1+Y^c)^{-k}] \log(Y)\} \\ &\quad + \frac{\theta}{C(\theta)} \mathbb{E}\{C'[\theta(1+Y^c)^{-k}] \log C'[\theta(1+Y^c)^{-k}]\} \\ &\quad - \frac{\theta(k-1)}{C(\theta)} \mathbb{E}\{C'[\theta(1+Y^c)^{-k}] \log(1+Y^c)\} \\ &= \int_{\mathbb{R}} f \log f \, dx = -\mathbb{H}_{\text{Sh}}(f). \end{aligned}$$

So, we obtain $\mathbb{H}_{\text{Sh}}(z) \leq \mathbb{H}_{\text{Sh}}(f)$ with equality if and only if $z(x) = f(x)$ for all x , except for a null measure set, thus proving the uniqueness. □

Appendix B: Integrals J_1 and J_2

B.1 Integral J_1

By writing $J_1 = \int_0^1 e^{\theta u} [\log(1-u^{1/k}) - k^{-1} \log(u)] du$, we can determine the first integral, say J_{11} , after setting $x = 1-u^{1/k}$, from the power series for the exponential function and the integral (for $a > -1$) $\int_0^1 (1-u)^a \log(u) du = -H_{a+1}/(a-1)$ given by Mathematica. Here, H_{a+1} is the harmonic number defined for real

$a > 0$ by $H_{a+1} = (a + 1) \sum_{k=1}^{\infty} [k(a + k + 1)]^{-1}$ (see http://en.wikipedia.org/wiki/Harmonic_number). We obtain $J_{11} = -k(a + 1) \sum_{k=0}^{\infty} n!^{-1} \theta^n H_{k(n+1)-k^2+1}$. For the second integral J_{12} , we expand the exponential function in power series and use $\int_0^1 z^a \log(z) dz = -(a + 1)^{-2}$ (for $a > -1$). After some algebra, we have $J_{12} = -k^{-1} \sum_{n=0}^{\infty} \theta^n / [(n + 1)^2 n!]$. Combining the results, we obtain J_1 given in Section 5.

B.2 Integral J_2

The integral J_2 in (6.2) can be determined by writing

$$J_2 = \int_0^1 \frac{\log(1 - z^{1/k})}{(1 - \theta z)^2} dz - \int_0^1 \frac{\log(z^{1/k})}{(1 - \theta z)^2} dz.$$

By using $(1 - \theta z)^{-2} = \sum_{n=1}^{\infty} n(\theta z)^{n-1}$ and setting $z = x^k$, the first integral J_{21} in J_2 can be written as

$$J_{21} = k \sum_{n=1}^{\infty} n \theta^{n-1} \int_0^1 x^{nk-1} \log(1 - x) dx.$$

Since $\log(1 - x) = -\sum_{j=1}^{\infty} x^j / j$, we obtain $J_{21} = k \sum_{n,j=1}^{\infty} [n \theta^{n-1}] / [j(nk + j)]$. The second integral J_{22} in J_2 is computed using Mathematica as $J_{22} = (k\theta)^{-1} \log(1 - \theta)$. Then, it follows J_2 as given in Section 6.

Appendix C: Elements of the observed information matrix

The elements of the 3×3 information matrix $J_n(\Theta)$ are

$$J_{\theta,\theta} = -\frac{n}{\theta^2} + \sum_{i=1}^n \frac{z_{3i}}{z_{2i}} [(1 + x_i^c)^{-k}]^2 - \sum_{i=1}^n \frac{z_{2i}}{z_{1i}} [(1 + x_i^c)^{-k}]^2 - nC(\theta)^{-1} \frac{\partial^2 C(\theta)}{\partial \theta^2} + nC(\theta)^{-2} \left[\frac{\partial C(\theta)}{\partial \theta} \right]^2,$$

$$J_{\theta,c} = \theta k \sum_{i=1}^n \left[\left(\frac{z_{2i}}{z_{1i}} \right)^2 - \frac{z_{2i}^2}{z_{1i}} \right] x_i^c (1 + x_i^c)^{-2k-1} \log x_i - k \sum_{i=1}^n \frac{z_{2i}^2}{z_{1i}} x_i^c (1 + x_i^c)^{-k-1} \log x_i,$$

$$J_{\theta,k} = \theta \sum_{i=1}^n \left(\frac{z_{2i}}{z_{1i}^2} - \frac{z_{3i}}{z_{1i}} \right) (1 + x_i^c)^{-2k} \log(1 + x_i^c) - \sum_{i=1}^n \frac{z_{2i}}{z_{1i}} (1 + x_i^c)^{-k} \log(1 + x_i^c),$$

$$\begin{aligned}
J_{c,c} &= -\frac{n}{c} - (k+1) \sum_{i=1}^n \frac{x_i^c (\log x_i)^2}{1+x_i^c} \\
&\quad + (\theta k)^2 \sum_{i=1}^n \left(\frac{z_{3i}}{z_{1i}} - \frac{z_{2i}}{z_{1i}} \right) x_i^{2c} (1+x_i^c)^{-2(k+1)} (\log x_i)^2 \\
&\quad + \theta k (1+k) \sum_{i=1}^n \frac{z_{2i}}{z_{1i}} x_i^{2c} (1+x_i^c)^{-k-2} (\log x_i)^2 \\
&\quad - \theta k \sum_{i=1}^n \frac{z_{2i}}{z_{1i}} x_i^c (1+x_i^c)^{-k-1} (\log x_i)^2, \\
J_{c,k} &= \theta^2 k \sum_{i=1}^n \left(\frac{z_{3i}}{z_{1i}} - \frac{z_{2i}}{z_{1i}} \right) x_i^c (1+x_i^c)^{-2k-1} \log(x_i) \log(1+x_i^c) - \sum_{i=1}^n \frac{x_i^c \log x_i}{1+x_i^c} \\
&\quad + \theta (k-1) \sum_{i=1}^n \frac{z_{2i}}{z_{1i}} x_i^c (1+x_i^c)^{-k-1} \log(x_i) \log(1+x_i^c)
\end{aligned}$$

and

$$\begin{aligned}
J_{k,k} &= -\frac{n}{k} + \theta^2 \sum_{i=1}^n \left(\frac{z_{3i}}{z_{1i}} - \frac{z_{2i}}{z_{1i}} \right) (1+x_i^c)^{-2k} \log^2(1+x_i^c) \\
&\quad + \theta \sum_{i=1}^n \frac{z_{2i}}{z_{1i}} (1+x_i^c)^{-k} \log^2(1+x_i^c),
\end{aligned}$$

where $z_{1i} = C'(\theta(1+x_i^c)^{-k})$, $z_{2i} = C''(\theta(1+x_i^c)^{-k})$ and $z_{3i} = C'''(\theta(1+x_i^c)^{-k})$ for $i = 1, \dots, n$.

References

- Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics and Probability Letters* **39**, 35–42. [MR1649319](#)
- Barakat, H. M. and Abdelkader, Y. H. (2004). Computing the moments of order statistics from non-identical random variables. *Statistical Methods and Applications* **13**, 15–26. [MR2081962](#)
- Barreto-Souza, W., Morais, A. L. and Cordeiro, G. M. (2010). The Weibull-geometric distribution. *Journal of Statistical Computation and Simulation* **81**, 645–657. [MR2788571](#)
- Blundell, R., Duncan, A. and Pendakur, K. (1998). Semiparametric estimation and consumer demand. *Journal of Applied Econometrics* **13**, 435–462.
- Boehme, T. K. and Powell, R. E. (1968). Positive linear operators generated by analytic functions. *SIAM Journal on Applied Mathematics* **16**, 510–519. [MR0226266](#)
- Chahkandi, M. and Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis* **53**, 4433–4440. [MR2744336](#)
- Cooner, F., Banerjee, S., Carlin, B. P. and Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association* **102**, 560–572. [MR2370853](#)

- Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. New York: Wiley. [MR1122806](#)
- Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Table of Integrals, Series and Products*. San Diego: Academic Press. [MR1773820](#)
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Reviews* **106**, 620–630. [MR0087305](#)
- Kapur, J. N. (1989). *Maximum Entropy Models in Science and Engineering*. New York: Wiley. [MR1079544](#)
- Kus, C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis* **51**, 4497–4509. [MR2364461](#)
- Lu, W. and Shi, D. (2012). A new compounding life distribution: The Weibull–Poisson distribution. *Journal of Applied Statistics* **39**, 21–38. [MR2872325](#)
- Mahmoudi, E. and Jafari, A. A. (2012). Generalized exponential power series distributions. *Computational Statistics and Data Analysis* **56**, 4047–4066. [MR2957854](#)
- Morais, A. L. and Barreto-Souza, W. (2011). A Compound family of Weibull and power series distributions. *Computational Statistics and Data Analysis* **55**, 1410–1425. [MR2741424](#)
- Murthy, D. N. P., Xie, M. and Jiang, R. (2004). *Weibull Models*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley. [MR2013269](#)
- Ostrowska, S. (2007). Positive linear operators generated by analytic functions. *Proceedings of the Indian Academy of Sciences* **117**, 485–493. [MR2374844](#)
- Paranaíba, P. F., Ortega, E. M. M., Cordeiro, G. M. and Pascoa, M. A. R. (2012). The Kumaraswamy Burr XII distribution: Theory and practice. *Journal of Statistical Computation and Simulation* **82**, 1–27.
- Paranaíba, P. F., Ortega, E. M. M., Cordeiro, G. M. and Pescim, R. R. (2011). The beta Burr XII distribution with application to lifetime data. *Computational Statistics and Data Analysis* **55**, 1118–1136. [MR2736499](#)
- Prudnikov, A. P., Brychkov, Y. A. and Marichev, O. I. (1986). *Integrals and Series, Vol. 1*. Amsterdam: Gordon and Breach Science Publishers. [MR0888165](#)
- Prudnikov, A. P., Brychkov, Y. A. and Marichev, O. I. (1992). *Integrals and Series, Vol. 4: Direct Laplace Transforms*. New York: Gordon and Breach Science Publishers.
- Rodrigues, C., Cordeiro, G. M., Demétrio, C. G. B. and Ortega, E. M. M. (2011). The Weibull negative binomial distribution. *Advances and Applications in Statistics* **22**, 25–55. [MR2859071](#)
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* **27**, 379–432. [MR0026286](#)
- Shao, Q. (2004). Notes on maximum likelihood estimation for the three-parameter Burr XII distribution. *Computational Statistics and Data Analysis* **45**, 675–687. [MR2050262](#)
- Shao, Q. X., Wong, H., Xia, J. and Ip, W. C. (2004). Models for extreme using the extended three-parameter Burr XII system with application to flood frequency analysis. *Hydrological Sciences Journal* **49**, 685–702.
- Shore, J. E. and 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. [MR0560389](#)
- Silva, R. B., Bourguignon, M., Dias, C. R. B. and Cordeiro, G. M. (2013). The compound family of extended Weibull power series distributions. *Computational Statistics and Data Analysis* **58**, 352–367. [MR2997948](#)
- Silva, G. O., Ortega, E. M. M., Cancho, V. G. and Barreto, M. L. (2008). Log-Burr XII regression models with censored data. *Computational Statistics and Data Analysis* **52**, 3820–3842. [MR2427383](#)
- Soliman, A. A. (2005). Estimation of parameters of life from progressively censored data using Burr XII model. *IEEE Transaction on Reliability* **54**, 34–42.

- Soofi, E. S. (2000). Principal information theoretic approaches. *Journal of the American Statistical Association* **95**, 1349–1353. [MR1825292](#)
- Tahmasbi, R. and Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics and Data Analysis* **52**, 3889–3901. [MR2432214](#)
- Wu, J. W., Chen, Y. I. and Chang, C. T. (2007). Statistical Inference based on progressively censored samples with random removals from the Burr type XII distribution. *Journal of Statistical Computation and Simulation* **77**, 19–27. [MR2343409](#)
- Zografos, K. and Balakrishnan, N. (2009). On families of beta-and generalized gamma-generated distributions and associated inference. *Statistical Methodology* **6**, 344–362. [MR2751078](#)

Universidade Federal de Pernambuco
Cidade Universitária, 50740-540
Recife-PE
Brasil
E-mail: rodrigobs29@gmail.com
gauss@de.ufpe.br