THE SACRAMENTO DISTRIBUTION

Vinicius Pereira do SACRAMENTO [1]
Katia Pires Nascimento do SACRAMENTO[2]
Francisco Alixandre Avila RODRIGUES[1]

▪ ABSTRACT: In this article, we present the Sacramento distribution of two parameters. This model competes effectively with the distributions used in the fault data analysis because it has a non-monotonous hazard function that can shape many forms of hazard. Some mathematical properties of the new distribution are also presented, including hazard function, survival, general formula for moments. The maximum likelihood method is used to estimate the model parameters. We obtain the expected information matrix and discuss inference methods. Finally, two real data sets are analyzed and comparisons are made between the new distribution with the Burr XII, Burr III and Beta Prime distributions to show the flexibility and potential of the new distribution.

▪ KEYWORDS: Expected information matrix; general formula for moments; hazard function; maximum likelihood; survival.

1 Introduction

Generalized distribution models have become very popular and have been widely used in the last decades to model varied databases such as reliability, engineering, biological, among other applications. In general, a known theoretical distribution is used to generalize another distribution and through it this resulting new model gains more parameters. As is the case of generalized beta distributions. This distribution class gained popularity following the works of Eugene et al. (2002) and Jones (2004).

[1] Universidade Federal do Cariri - UFCA, Centro de Ciências e Tecnologia, CEP: 63048-080, Juazeiro do Norte, CE, Brasil, E-mail: *vinicius.sacramento@ufca.edu.br; alixandre.avila@ufca.edu.br*
[2] Universidade Regional do Cariri - URCA, Departamento de Matemática, CEP: 63041-145, Juazeiro do Norte, CE, Brasil, E-mail: *katia1pires@hotmail.com*

The advantage of these new models is that they model data better, but calculations for estimates, moments, etc., are not trivial in many cases. However, long before these new classes of distributions emerged, some articles studied properties and theories of the cumulative distribution function and thus proposed new distributions. For example, Burr (1942) created the Burr system of cumulative distributions (cdf) proposing twelve different forms of distribution functions. From this, several studies as Burr and Cislak (1968), Burr (1973), Rodriguez (1977), Tadikamalla (1980), Wingo (1993), Zimmer et al. (1998), Mousa and Jaheen (2002), Shao et al. (2004), Raqab and Kundu (2006), Gunasekera (2018) that explored some of these twelve forms. The Burr distribution has been used in physical applications, reliability, survival analysis, quality control, forest stands.

Thus, as in the work proposed by Burr and Cislak (1968), we introduce an analytical formulation that allowed to define a new cumulative distribution function (cdf). Based on the cdf, we proposed the probability density function named Sacramento (SACR) distribution, with two parameters. To evaluate the performance of the Sacramento distribution, we performed experiments on real database. The results the new distribution when compared to some distributions, used in modeling survival data, presented satisfactory results. Seem to be an important distribution used in this type of analysis.

In addition to proposing the SACR distribution, we present some important aspects of this distribution. Those aspects are presented in this article that has the following structure: Section 1 the introduction. Section 2 we define the SACR distribution and its respective cumulative distribution function (cdf). Section 3 show the moments. We provides the risk rate function, reliability function and provides graphical illustrations, in Section 4. Section 5 the quantile function and hazard quantile function. In Section 6, we discuss parameter estimation of the distribution using the maximum likelihood method and applications of the Sacramento distribution was discussed in Section 7. In Section 8 finally our conclusions.

2 The Sacramento distribution

Suppose X is a random variable that follows the Sacramento distribution SACR(σ, λ), with scale parameter $\sigma > 0$, shape parameter $\lambda > 1$. Hence, the cumulative distribution function (cdf) is given by

$$F(x) = \frac{2}{\pi} \arctan\left[\left(\frac{x}{\sigma}\right)^{\lambda}\right], \; x > 0; \lambda > 1, \sigma > 0. \tag{1}$$

The probability density function (pdf) of the SACR distribution for (1) is given by

$$f(x) = \frac{2\lambda \left(\frac{x}{\sigma}\right)^{-1+\lambda}}{\pi \left(1 + \left(\frac{x}{\sigma}\right)^{2\lambda}\right) \sigma}, \; x > 0. \tag{2}$$

Figures 1 and 2 illustrate some of the possible shapes of the cdf (1) and pdf (2), respectively, for selected parameter values.

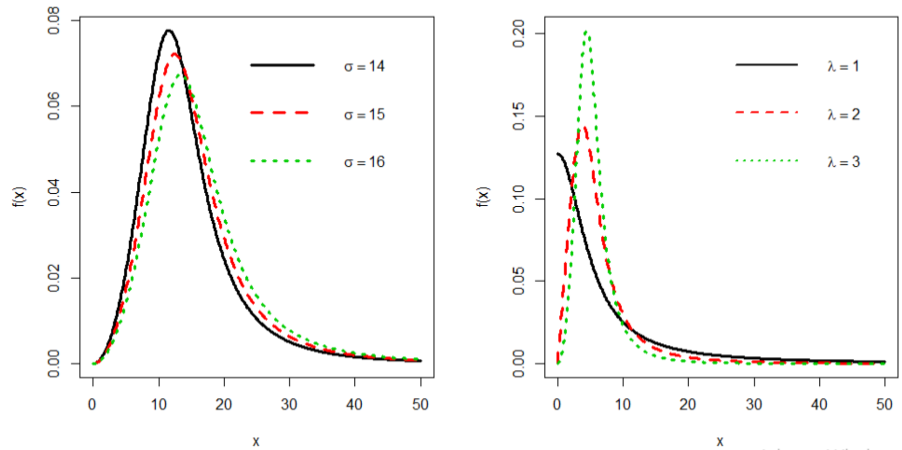


Figure 1 - Shape of Sacramento distribution for selected parameters.

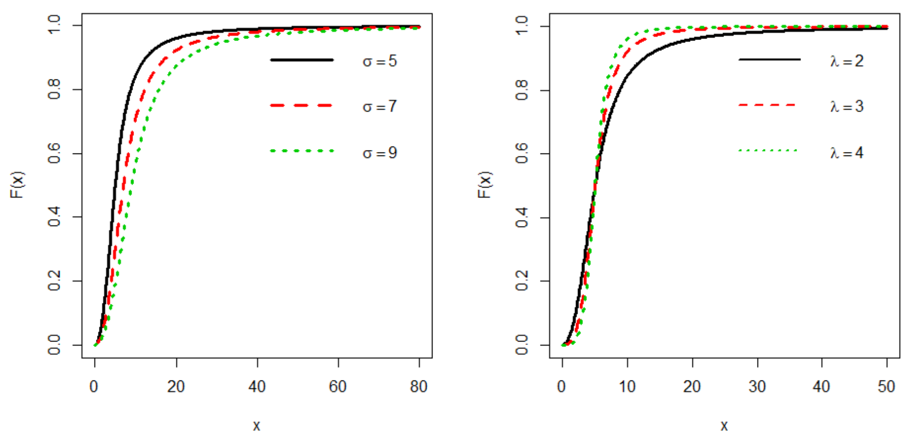


Figure 2 - cdf of the Sacramento distribution for selected parameter values.

3 Moments

In any statistical analysis the moments are of great importance, especially in the applied works. Through the moments one can verify asymmetry, tendency, for example. In this section, the general expression for the rth moment of the Sacramento distribution is presented now.

Let $X \sim Sacramento(\sigma, \lambda)$. The rth moment of the random variable X is defined as follows:

$$E\left[x^r\right] = \sigma^r \sec\left[\frac{\pi r}{2\lambda}\right], r < \lambda, r + \lambda > 1. \tag{3}$$

4 Reliability function

Let X be a continuous random variable with cdf given in (1), and pdf (2), the reliability function $S(x)$ and hazard function $h(x)$ are provided by equations (4) and (5), as follow:

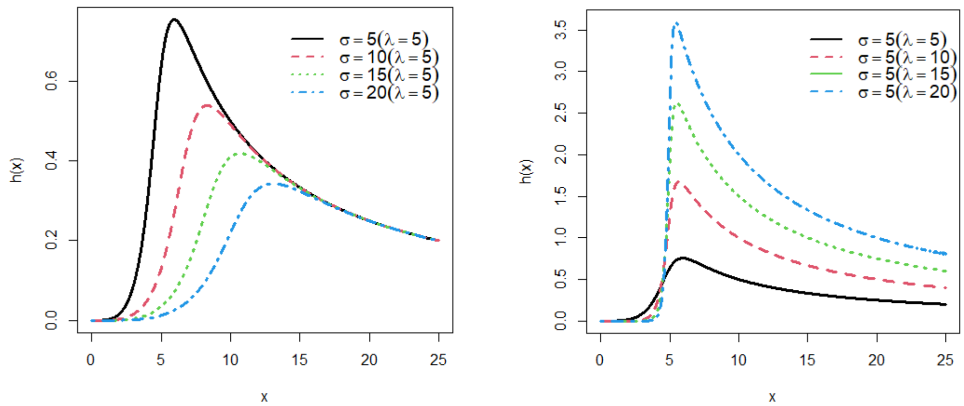
$$S(x) = 1 - \frac{2\arctan((x/\sigma)^\lambda)}{\pi},\ x > 0;\ \lambda > 1,\ \sigma > 0. \tag{4}$$

and

$$h(x) = \frac{2\lambda((x/\sigma)^\lambda)}{x(1 + (x/\sigma)^\lambda)(\pi - 2\arctan((x/\sigma)^\lambda))},\ x > 0;\ \lambda > 1, \sigma > 0, \tag{5}$$

The Figure 3 illustrates some of the possible shapes of $h(x)$ for selected values of (λ, σ): for fixed λ's, Figure 3 (a), increasing the values, for σ, flattens and shifts the maximum risk position to the right, for the case in which the σ's are fixed, Figure 3 (b), it shows that in increasing the λ values, the maximum values of the hazard functions are directly proportional and the non-variation of the σ parameter means that there is no translation of the hazard function. Furthermore, $h(x) \to 0$ as $x \to 0$ and $h(x) \to 0$ as $x \to \infty$, for all $\sigma > 0$ and $\lambda > 1$.

,



(a) Examples of Sacramento Hazard Functions for $\lambda = 5$.

(b) Examples of Sacramento Hazard Functions for $\sigma = 5$.

Figure 3 - Plot of Sacramento Hazard Functions.

4.1 Stress-strength reability

Let X and Y be independent random variables SACR with parameters (λ, σ_1) and (λ, σ_1), respectively. In this way, the component X represents the force of a component and Y the stress acting on it. The probability density functions (pdfs) of X and Y are given, respectively, by

$$f_x(x;\lambda,\sigma_1) = \frac{2\lambda\left(\frac{x}{\sigma_1}\right)^{-1+\lambda}}{\pi\left(1+\left(\frac{x}{\sigma_1}\right)^{2\lambda}\right)\sigma_1}, \quad x>0; \quad \lambda>1, \quad \sigma_1>0. \tag{6}$$

$$f_y(y;\lambda,\sigma_2) = \frac{2\lambda\left(\frac{y}{\sigma_2}\right)^{-1+\lambda}}{\pi\left(1+\left(\frac{y}{\sigma_2}\right)^{2\lambda}\right)\sigma_2}, \quad y>0; \quad \lambda>1, \quad \sigma_2>0. \tag{7}$$

where σ_1 and σ_2 are unknown parameters and λ is a known common parameter.

$$\begin{aligned} R = P[Y < X] &= \int_0^{\infty}\int_0^{x} \frac{2\lambda\left(\frac{x}{\sigma_1}\right)^{-1+\lambda}}{\pi\left(1+\left(\frac{x}{\sigma_1}\right)^{2\lambda}\right)\sigma_1} \frac{2\lambda\left(\frac{y}{\sigma_2}\right)^{-1+\lambda}}{\pi\left(1+\left(\frac{y}{\sigma_2}\right)^{2\lambda}\right)\sigma_2} dy\, dx \\ &= \int_0^{\infty} \frac{2^2\lambda\left(\frac{x}{\sigma_1}\right)^{-1+\lambda}}{\pi^2\left(1+\left(\frac{x}{\sigma_1}\right)^{2\lambda}\right)\sigma_1} \arctan\left[\left(\frac{x}{\sigma_2}\right)^{\lambda}\right] dx. \\ &= \frac{1}{2}\left[Sign\left[\left(\frac{1}{\sigma_2}\right)^{\lambda}\right]\left[-LerchPhi[\sigma_1^{-2\lambda}\sigma_2^{2\lambda}, 2, 1/2]\right.\right. \\ &+ \left.\left.\left[\pi^2 - 4\lambda\arctan\left[\left(\frac{\sigma_2}{\sigma_1}\right)^{\lambda}\right](\log[\sigma_1]-\log[\sigma_2])\right]\left(\frac{1}{\sigma_2}\right)^{\lambda}\right]\left(\frac{1}{\sigma_2}\right)^{-\lambda}\right], \end{aligned} \quad (8)$$

where Sign is the signum function and the LerchPhi function is defined as follows:

$$LerchPhi(x, lambda, \sigma) = \sum_{n=0}^{\infty} \frac{x^n}{(\sigma+n)^{\lambda}}, \quad (9)$$

for $|x| < 1$ or $|x| = 1$ and $\Re(\lambda) > 1$ McPhedran et al. (2006).

5 Sacramento quantile function

Let X be a random variable with distribution function F, and let $u \in [0, 1]$. A value of x such that $F(x^-) = P(X < x) \leq u$ and $F(x) = P(X \leq x) \geq u$ is called a quantile of order u for the distribution.
The quantile function of Sacramento distribution is obtained by solving the equation,

$$u = \frac{2}{\pi}\arctan[(\frac{Q(u)}{\sigma})^{\lambda}] \quad (10)$$

Thus, the quantile function is given by

$$Q(u) = \sigma\sqrt[\lambda]{\tan(\frac{\pi u}{2})} \quad , 0 \leq u \leq 1 \quad (11)$$

From the definition in (11) since F is continuous $FQ(u) = u$, where $FQ(u) = u$ represents the composite function $F(Q(u)) = q$ and $Q'(u) = q(u)$. We must

$$\frac{\partial F(Q(u))}{\partial u} = q(u)fQ(u) = 1$$

According to Parzen (1979), the density quantil function is defined as $fQ(u)$, where f is the density function of X. The hazard rate of X de (5) can be written in 11 terms as the hazard quantile function,

$$
\begin{aligned}
H(u) &= hQ(u) = (1-u)^{-1} fQ(u) \\
&= [(1-u)q(u)]^{-1} \\
&= \left[\frac{\pi(1-u)\sigma \sec\left[\frac{\pi u}{2}\right]^2 \tan\left[\frac{\pi u}{2}\right]^{-1+1/\lambda}}{2\lambda} \right]^{-1}.
\end{aligned} \tag{12}
$$

6 Inference

Consider that $X \sim Sacramento(\theta)$, in which $\theta = (\sigma, \lambda)^T$ is the parameter vector. The log-likelihood function of θ can be written as

$$
\ell(\theta) = n\log(2\lambda) - n\log(\pi\sigma) + (\lambda-1)\sum_{i=1}^{n}\log(x_i/\sigma) - \sum_{i=1}^{n}\log(1+(x_i/\sigma)^{2\lambda}) \tag{13}
$$

For a given random sample $\mathbf{x} = (x_1, \cdots, x_n)$ of X, and size n,the total log-likelihood is $\ell = \sum_{i=1}^{n} \ell^i$, in which ℓ^i is the log-likelihood for the i-th remark $(i = 1, \cdots, n)$. The maximum likelihood estimator (MLE) $\hat{\theta}$ de θ can be calculated numerically. The components of the vector score $U_\theta = (U_\sigma, U_\lambda)^T$ are given by

$$
U_\sigma = \frac{-n(-1+\lambda)}{\sigma} + \sum_{i=1}^{n} \frac{2\lambda x_i (x_i/\sigma)^{2\lambda-1}}{\sigma^2[1+(x_i/\sigma)^{2\lambda}]}
$$

$$
U_\lambda = \frac{n}{\lambda} + \sum_{i=1}^{n}\log(x_i/\sigma) - \sum_{i=1}^{n} \frac{2\log(x_i/\sigma)(x_i/\sigma)^{2\lambda}}{1+(x_i/\sigma)^{2\lambda}} \tag{14}
$$

The maximum likelihood estimate (MLE) $\hat{\theta}$ of θ is obtained by solving the nonlinear likelihood equations $U_\sigma(\theta) = 0$, $U_\lambda(\theta) = 0$. These equations can not be solved analytically and statistical software can be used to obtain the MLEs numerically. We can use iterative techniques such as Newton-Raphson, quasi-Newton algorithms, etc., to obtain the estimate $\hat{\theta}$. In addition, you can use routines in the R statistical package, such as maxLik, Nlminb, and others, to find the maximum of the function (13), with initial kicks for the parameter choices. For interval estimation and hypothesis tests on the model parameters, we require the 2×2 unit observed information matrix, given by $J_n(\theta) = \frac{-\partial^2 \ell(\theta)}{\partial(\theta)\partial\theta^\top} = -U_{ij}$, para $i, j = \sigma$ e λ.

Under conditions that are fulfilled for parameters in the interior of the parameter space but not on the boundary, the asymptotic distribution of

$$
\sqrt{n}(\widehat{\theta} - \theta) \text{ is } N_2(0, I(\theta)^{-1}),
$$

where $I(\theta)$ is the unit expected information matrix, i.e.,

$$I(\theta) = -\frac{1}{n}\begin{bmatrix} E(U_{\lambda\lambda}) & E(U_{\lambda\sigma}) \\ E(U_{\sigma\lambda}) & E(U_{\sigma\sigma}) \end{bmatrix},$$

whose elements are

$E(U_{\sigma\sigma}) = \frac{-n(\lambda^2-2\sigma+2\lambda(1+\sigma))}{2\sigma^2}$,

$E(U_{\lambda\lambda}) = \frac{-n\pi^2}{8\lambda^2}$,

$E(U_{\lambda\sigma}) = E(U_{\sigma\lambda}) = 0.$

This asymptotic behaviour holds if $I(\theta)$ is replaced by $J(\widehat{\theta})$, i.e. the observed information matrix evaluated at $\widehat{\theta}$. The asymptotic multivariate normal $N_2(0, I(\theta)^{-}1)$ distribution can be used to construct approximate confidence intervals for the individual parameters and for the hazard and survival functions.

7 Applications

In this section, the Sacramento distribution was adjusted using real databases and then compared with the Burr III, Burr XII and Beta Prime distributions in order to compare them and verify their potentiality and robustness. The first data set consist of the number of successive failures for the air conditioning system of each member in a fleet of 13 Boeing 720 jet airplanes Proschan (1963). These data with 214 observations were also discussed by Dahiya and Gurland (1972), Gleser (1989), Barreto-Souza et al. (2011) and Kehinde et al. (2018). The second data set consists of 63 observations, and are the strengths of 1.5 cm glass fibers, measured at the National Physical Laboratory, England. These data were analyzed by Smith and Naylor (1987). To obtain the maximum likelihood estimates (MLEs) for the distribution parameters, the maxLik function of the statistical software R was used, and the iteration method was Newton Raphson. The estimated values of the parameters, -2log-Likelihood statistic, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) are presented in the Tables 1 and 2. The distributions compared to the Sacramento distribution were: Burr III, Burr XII and Beta Prime.

The Figure 4 shows the fits of the Sacramento, Burr III, Bur XII and Beta Prime models for the first set of data. According to the illustration, the good fit of the Sacramento distribution is observed.

The plots of estimated densities of Sacramento, Burr III, Burr IX and Beta Prime in Figures 4 and 5 show that the Sacramento distribution provides a good fit for both sets of data.

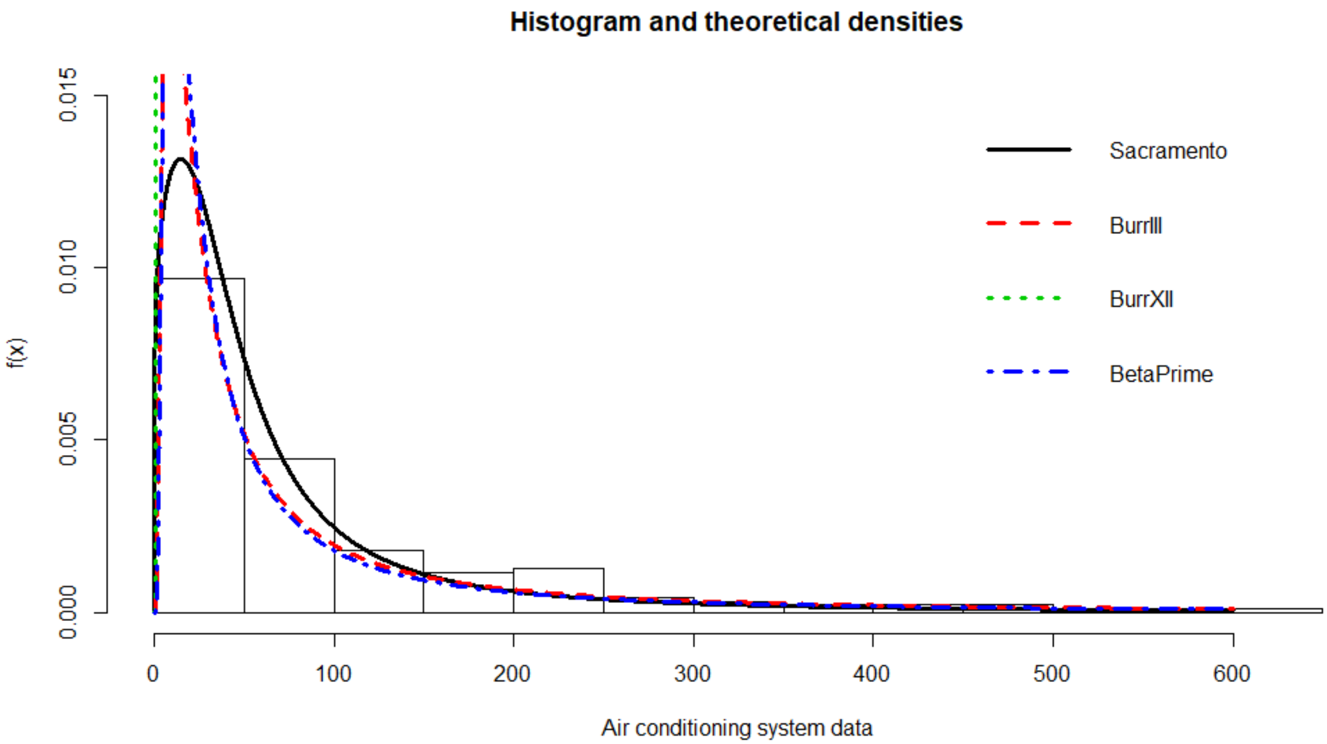


Figure 4 - Fitted densities of the Sacramento, Burr III, Burr XII and Beta Prime distributions for the first data set.

Table 1 - The statistics -2Log-Likelihood, AIC and BIC and estimates of the model parameters for the Boeing 720 data, the corresponding standard error (given in parentheses)

MODEL	-2Log-Likelihood	AIC	BIC	α	β
Sacramento	2080.792	2084.784	2091.256	1.17102	50.83174
				(0.07617)	(2.09715)
Burr XII	2345.973	2350,18	2356.491	11.649259	0.022139
				(0.864568)	(0.002306)
Burr III	2116.623	2116.56	2123.033	0.80032	14.15020
				(0.03878)	(1.73821)
Beta Prime	2128.322	2132.322	2138.795	17.0888	0.7571
				(0.6503)	(0.0478)

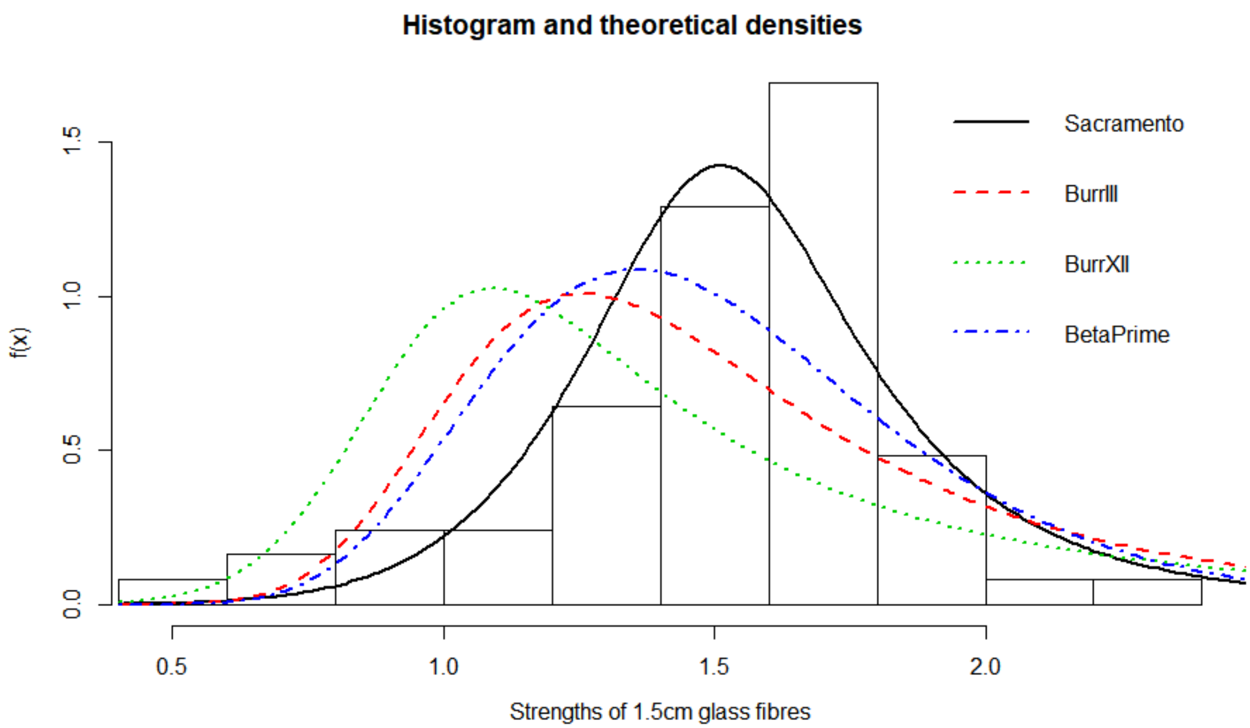


Figure 5 - Fitted densities of the Sacramento, Burr III, Burr XII and Beta Prime distributions for the second data set.

Table 2 - AIC, BIC and (MLEs) of the distributions for the second data set

MODEL	-2Log-Likelihood	AIC	BIC	α	β
Sacramento	41.63148	45.63148	49.88575	6.84883	1.54479
				(0.78313)	(0.03819)
Burr XII	95.68713	99.68713	103.9414	7.45920	0.32274
				(1.29985)	(0.06661)
Burr III	72.8382	76.8382	81.09247	4.0818	3.4083
				(0.3387)	(0.4507)
Beta Prime	56.79804	60.79804	65.05231	36.3642	25.0472
				(0.6704)	(0.5868)

Conclusion

In this work, we define a new model called Sacramento distribution. It is observed that the new distribution of two parameters is quite similar in nature to the other distributions like Burr III, Burr XII and Beta Prime.

Some mathematical properties are derived and plots of the pdf, cdf and hazard functions are presented to show the flexibility of the new distribution. We obtained the quantile function, general formula for the moments, the fdp of the order statistics, and two measures of entropy. The estimation of maximum likelihood, the expected and observed information matrices are discussed.

Finally, we fit Sacramento model to two different types real data sets e we observed that the new distribution presented a more flexible behavior in relation to the others models compared.

SACRAMENTO, V. P.; SACRAMENTO, K. P. N.; RODRIGUES, F. A. A. A distribuição Sacramento. *Rev. Bras. Biom.,* Lavras, v.39, n.3, p.434-446, 2021.

- *RESUMO: Neste artigo, apresentamos a distribuição Sacramento de dois parâmetros. Este modelo compete efetivamente com as distribuições usadas nas análises de dados de falha porque tem uma função de risco não monótona, que pode moldar muitas formas de risco. Algumas propriedades matemáticas da nova distribuição são também apresentadas, incluindo as funções de risco, de sobrevivência e a fórmula geral para momentos. O método de máxima verossimilhança foi usado para estimar os parâmetros do modelo. Obtivemos a matriz de informação esperada e discutimos métodos inferênciais. Finalmente, dois conjuntos de dados reais são analisados e comparações são feitas entre a nova distribuição com as distribuições Burr XII, Burr III* e *Beta Prime, para mostrar a flexibilidade e o potencial da nova distribuição.*

- *PALAVRAS-CHAVE: Fórmula geral para momentos; função de risco; matriz de informação esperada; máxima verossimilhança; sobrevivência.*

References

BARRETO-SOUZA, W.; DE MORAIS, A. L.; CORDEIRO, G. M. The Weibull-geometric distribution. *Journal of Statistical Computation and Simulation*, v.81, n.5, p.645-657, 2011.

BURR, I. W. Cumulative frequency functions. *Annals of mathematical statistics*, v.13, n.2, p.215-232, 1942.

BURR, I. W. Pameters for a general system of distributions to match a grid of $\alpha 3$ and $\alpha 4$. *in Statistics-Theory and Methods*, v.2, n.1, p.1-21, 1973.

BURR, I. W.; CISLAK, P. J. On a general system of distributions I. Its curve-shape characteristics II. The sample median. *Journal of the American Statistical Association*, v.63, n.322, p.627-635, 1968.

DAHIYA, R. C.; GURLAND, J. Goodness of fit tests for the gamma and exponential distributions. *Technometrics*, v.14, n.3, p.791-801, 1972.

EUGENE, N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods*, v.31, n.4, p.497-512, 2002.

GLESER, L. J. The gamma distribution as a mixture of exponential distributions. *The American Statistician*, v.43, n.2, p.115-117, 1989.

GUNASEKERA, S. Inference for the Burr XII reliability under progressive censoring with random removals. *Mathematics and Computers in Simulation*, v.144, p.182-195, 2018.

JONES, M. Families of distributions arising from distributions of order statistics. *Test*, v.13, n.1, p.1-43, 2004.

KEHINDE, O.; OSEBI, A.; GANIYU, D. A New Class of Generalized Burr III Distribution for Lifetime Data. *International Journal of Statistical Distributions and Applications*, v.4, n.1, p.6-21, 2018.

McPHEDRAN, R. C.; BOTTEN, L. C., NICOROVICI, N. A. P.; JOHN ZUCKER, I. Symmetrization of the Hurwitz zeta function and Dirichlet L functions. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v.463, p.281-301, 2007.

MOUSA, M. A.; JAHEEN, Z. F. Statistical inference for the Burr model based on progressively censored data. *Computers* & *Mathematics with Applications*, v.43. n.10, p.1441-1449, 2002.

PARZEN, E. Nonparametric statistical data modeling. *Journal of the American statistical association*, v.74, n.365, p.105-121, 1979.

PROSCHAN, F. Theoretical explanation of observed decreasing failure rate. *Technometrics*, v.5, n.3, p.375-383, 1963.

RAQAB, M. Z.; KUNDU, D. Burr type X distribution: revisited. *Journal of Probability and Statistical Sciences*, v.4, n.2, p.179-193, 2006.

RODRIGUEZ, R. N. A guide to the Burr type XII distributions. *Biometrika*, v.64, n.1, p.129-134, 1977.

SHAO, Q.; WONG, H.; XIA, J.; IP, W. C. Models for extremes using the extended three-parameter Burr XII system with application to flood frequency analysis/Modèles d´extrêmes utilisant le système Burr XII étendu à trois paramètres et application à l´analyse fréquentielle des crues. *Hydrological Sciences Journal*, v.49, n.4, p.685-702, 2004.

SMITH, R. L.; NAYLOR, J. A comparison of maximum likelihood and Bayesian estimators for the three?parameter Weibull distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, v.36, n.3, p.358-369, 1987.

WINGO, D. R. Maximum likelihood methods for fitting the burr type xii distribution to multiply (progressively) censored life test data. *Metrika*, v.40, n.1, p.203-210, 1993.

ZIMMER, W. J.; KEATS, J. B.; WANG, F. K. The burr xii distribution in reliability analysis. *Journal of quality technology*, v.30, n.4, p.386-394, 1998.

Received on 18.08.2020.

Approved after revised on 21.01.2021.