Scientific
Research
Publishing

Basic Functions for Computational Implementation of the Box-Cox Symmetric Class of Distributions

Giovana Fumes-Ghantous[1], José Eduardo Corrente[2]

[1]Basic Science Department, Faculty of Animal Science and Food Engineering, University of São Paulo, São Paulo, Brazil
[2]Research Support Office, Botucatu Medical School, University of São Paulo State, Botucatu, Brazil
Email: giovana.fumes@usp.br, jose.corrente@unesp.br

Abstract

A class of distributions called Box-Cox symmetric was proposed for random variables with asymmetric distributions. This class allows through its structure an interpretation of the parameters in terms of quantiles (in particular, the median), relative dispersion and skewness. This study presents the initial results of the computational development of basic functions of each of the distributions that make up the Box-Cox symmetric class. Four functions have been developed to compose a routine in software R up to now. These functions are related to random numbers generation, probability density function, cumulative distribution function, and quantile function associated to a given probability. Examples of implemented functions were presented. The gamlss routine was used to check the performance of developed functions.

1. Introduction

The Box-Cox symmetric class of distributions (BCS) emerges as an alternative for modeling highly asymmetric data even when outliers are present [1]. The main idea to develop this class of distributions has begun to solve problems related to analyze nutrient intake data in dietitian area. In this context and as the data consumption present an asymmetric distribution, it is common to apply Box-Cox transformation [2] in order to model the transformed data with the normal distribution approach. It just works well when the data present a reasonably well

behavior distribution. In the case of asymmetric distribution or outliers' presence, this propose doesn't work very well, since in the presence of outliers, Box-Cox transformation cannot give an adequate distribution for the data. So, an alternative approach for estimating usual nutrient distributions and prevalence of inadequate nutrient intakes was done through a Box-Cox t model with random intercept [3].

The BCS class allows through its structure an interpretation of the parameters in terms of quantiles (especially, the median), relative dispersion and skewness, which makes it interesting for regression models. This class includes as particular cases the log-normal [4], Box-Cox t [5], Box-Cox Cole-Green [6] and Box-Cox power exponential [7] distributions.

This work presents an initial study about the computational implementation of the distributions that make up the BCS class, through the presentation of four basic functions. Additionally, a study with the gamlss package in software R is presented to corroborate the proposal.

2. Methodology

Let Y be a positive and continuous random variable. The Box-Cox symmetric class of distributions is defined from the transformation

$$Z = Z(Y, \mu, \sigma, \nu) = \begin{cases} \dfrac{1}{\sigma \nu}\left[\left(\dfrac{Y}{\mu}\right)^{\nu} - 1\right], & \text{if } \nu \neq 0, \\ \dfrac{1}{\sigma}\log\left(\dfrac{Y}{\mu}\right), & \text{if } \nu = 0, \end{cases} \quad (1)$$

where $\mu > 0, \sigma > 0, -\infty < \nu < \infty$ and Z has a standard symmetric distribution truncated at the interval

$$A(\sigma, \nu) = \begin{cases} \left(-\dfrac{1}{\sigma \nu}, \infty\right), & \text{if } \nu > 0, \\ \left(-\infty, -\dfrac{1}{\sigma \nu}\right), & \text{if } \nu < 0, \\ (-\infty, \infty), & \text{if } \nu = 0, \end{cases}$$

and $Z \sim S(0, 1, A(\sigma, \nu); r)$, with $r(\cdot)$ being the density generating function. Thus, Y has a BCS distribution with parameters μ, σ and ν, so $Y \sim BCS(\mu, \sigma, \nu; r)$, if Z is given in (1) is such that $Z \sim S(0, 1, A(\sigma, \nu); r)$, with $r(\cdot)$ being the density generating function. The class of the symmetric distributions has a number of well-known distributions as special cases depending on the choice of r. It includes the normal distribution, the Student-t, power exponential, type I logistic, type II logistic and slash distributions among others. These densities have quite different tail behaviors, and some of them may have heavier or lighter tails than the normal distribution [1].

The first basic function generates random numbers from a probability density function. The function is initiated by r (*random*) followed by the required BCS

distribution, given by:

$$rBCS(n, \mu, \sigma, \nu) \tag{2}$$

where:

- n: number of observations to be simulated;
- μ: parameter related to the median in a distribution that belongs to the BCS class;
- σ: parameter related to the relative dispersion in a distribution that belongs to the BCS class;
- ν: parameter related to the skewness in a distribution that belongs to the BCS class.

Note that the proposed function requires only three parameters, but some distributions of the BCS class require a fourth parameter, that is related to kurtosis of the distribution. For instance, the Box-Cox t distribution has an additional parameter to model the tail decay, which is defined with the others as $BCT(\mu, \sigma, \nu, \tau)$.

Thus, if the intention is to generate random numbers from a variable that has a Box-Cox Normal distribution (BCN), for example, you have to write $rBCN(n, \mu, \sigma, \nu)$ and to specify the number of observations to be simulated and the parameters values.

The second basic function generates the probability density function as result. The notation is initiated by d (*density*) followed by the required BCS distribution:

$$dBCS(y, \mu, \sigma, \nu)$$

where:

- y: value of the observed variable to obtain the corresponding point of the probability density function;
- μ, σ and ν: parameters related to a distribution that belongs to the BCS class, as described in the Equation (2).

Then, considering all possible values for y, a graph of the probability density function of a variable that follows a BCS distribution can be plotted. For instance, in order to obtain the density function of the Box-Cox Normal distribution (BCN), just specify the parameters in the function $dBCN(y, \mu, \sigma, \nu)$, for any possible values of y.

The third basic function returns the cumulative distribution function ($P(Y \leq y)$), which is initiated by p (*probability*) followed by the required BCS distribution, written as:

$$pBCS(y, \mu, \sigma, \nu),$$

where:

- y: any possible positive value of Y;
- μ, σ and ν: parameters related to a distribution that belongs to the BCS class, as described in the Equation (2).

For example, if the aim is to obtain the cumulative probability associated to

any value of a random variable with Box-Cox Normal distribution (BCN), just specify the parameters $pBCN(y,\mu,\sigma,\nu)$, for any value of y.

The fourth basic function returns the quantile of a BCS distribution. For that, the function initiates by q (*quantile*) followed by the distribution to obtain the quantile, which is written as follows:

$$qBCS(p,\mu,\sigma,\nu),$$

where:

- p: probability value, which varies between 0 and 1;
- μ, σ and ν: parameters related to a distribution that belongs to the BCS class, as described in the Equation (2).

For instance, to obtain the quantile associated with a probability value of a Box-Cox Normal distribution (BCN), just write $qBCN(p,\mu,\sigma,\nu)$, specifying the probability value and the parameters values.

In order to evaluate the implemented functions, the gamlss package present in R was used gamlss [8]. The Box-Cox Normal distribution was chosen to compare the results.

All implemented functions were done in the software R, version 4.0.2.

3. Results

3.1. The Use of the Implemented Functions

In order to generate random numbers, consider a Box-Cox Slash distribution. In this case, it can be used the command *rBCSlash* (*n, μ, σ, ν, q*) with $n = 100$, $\mu = 5$, $\sigma = 0.5$, $\nu = -0.5$; 0; 0.5 and $q = 2$. **Figure 1** presents the histogram. The skewness is the main feature of the functions that make up such a class of distributions.

To obtain the probability density function of a Box-Cox type II logistic distribution (BCLog2), it can be used the command $dBCLog2(y,\mu,\sigma,\nu)$.
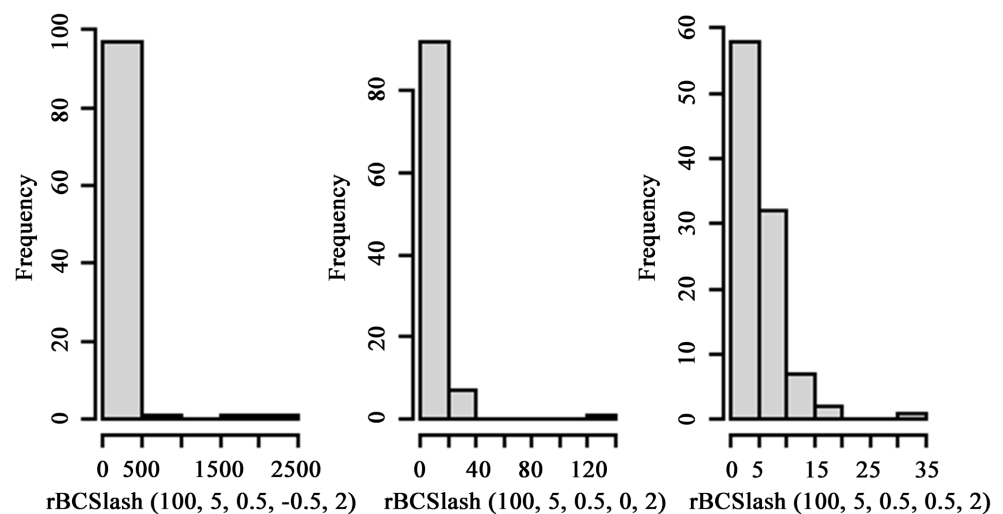

Figure 1. Histograms of data from $n = 100$ random numbers generated by *Box-Cox Slash* (*BCSlash*), $\mu = 5$, $\sigma = 0.5$, $\nu = -0.5$; 0; 0.5 and $q = 2$.

Figure 2 presents the probability density function for some different values of the parameters.

Using a Box-Cox Cauchy distribution to calculate the cumulative distribution function, the command was $pBCCauchy$ (y, $\mu = 5$, $\sigma = 0.5$, $\nu = 0.5$). Considering several values of y, Figure 3 presents the result.

As an application of the fourth function (quantile) using the *Box-Cox Double Exponential distribution* (*BCDE*), it was considered $\mu = 5$, $\sigma = 0.5$ and $\nu = 0.5$. The function $qBCDE$ ($p = 0.95$, $\mu = 5$, $\sigma = 0.5$, $\nu = 0.5$) provided the value 9.957671, that correspond the 95% of probability.

3.2. Comparison between the Implemented Functions and the Gamlss Package

The gamlss package presented in software R was used to analyze the implemented functions. This package presents some particular cases of the BCS distributions. The Box-Cox Normal distribution was chosen to make the comparisons.

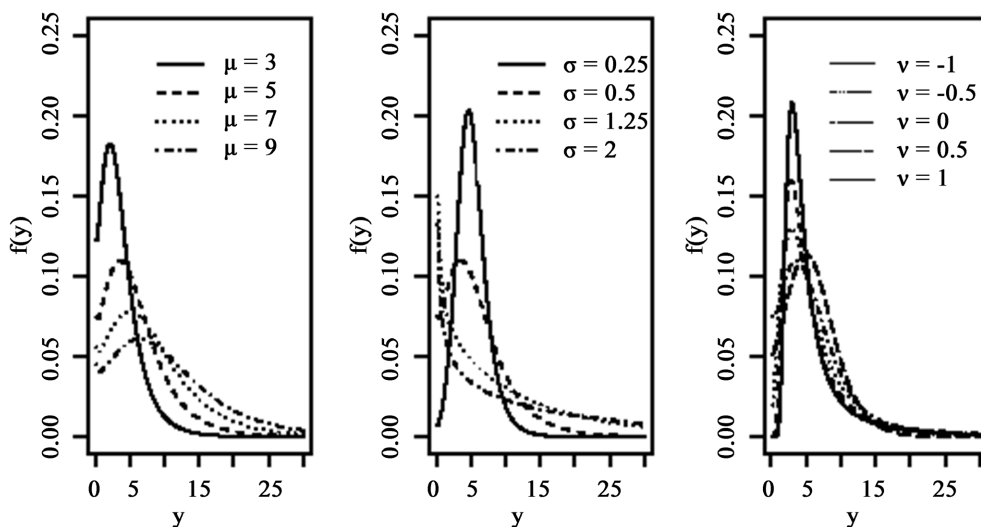Figure 4 shows 1000 random numbers that were generated from a Box-Cox

Figure 2. Probability density functions of the *Box-Cox type II logistic* (*BCLog2*). On the left, values of μ varying, $\sigma = 0.5$ and $\nu = 0.5$; in the center $\mu = 5$, values of σ varying and $\nu = 0.5$; on the right, $\mu = 5$; $\sigma = 0.5$ and ν varying.
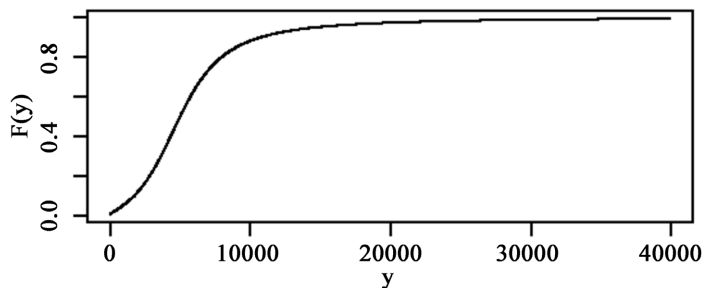

Figure 3. Cumulative distribution function of a *Box-Cox Cauchy* (*BCCauchy*), $\mu = 5$, $\sigma = 0.5$ and $\nu = 0.5$.

Normal distribution (BCN) for $\mu = 1$, $\sigma = 0.5$ and $v = 0.5$ using the implemented function and the gamlss package. In the gamlss, the Box-Cox Normal is known as Box-Cox Cole Green (BCCG).

Figure 5 shows a graph of a probability density function generated from the Box-Cox Normal (BCN) with the parameters $\mu = 5$, $\sigma = 0.5$ and $v = 0.5$ by the implemented function and by the gamlss package.

Table 1 presents a comparison of the quantiles of BCN using implemented functions and gamlss package. As can be seen, results were the same.
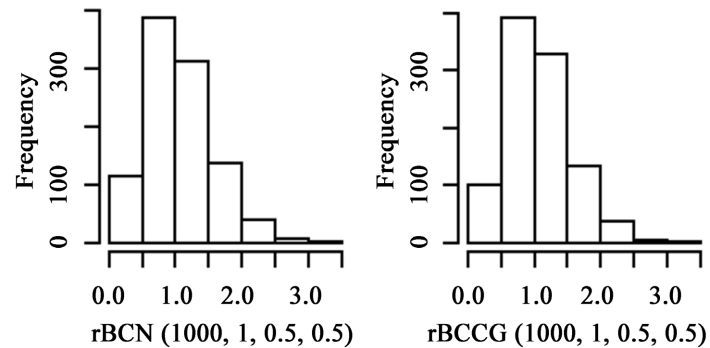

Figure 4. Histograms from 1000 random numbers generated by the probability density function of the *Box-Cox Normal* (*BCN*) for $\mu = 1$, $\sigma = 0.5$ and $v = 0.5$. On the left, by the implemented function, and on the right, by the gamlss package, using the same seed.
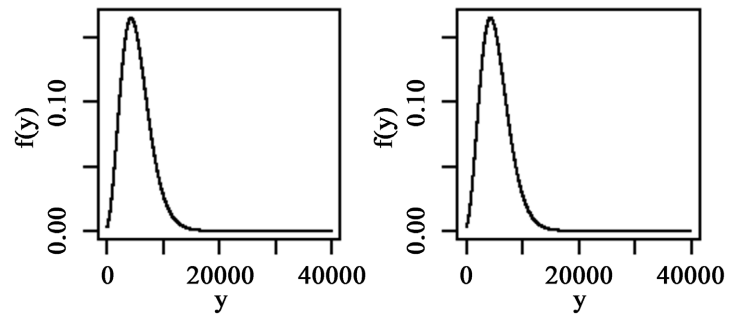

Figure 5. Densities of the Box-Cox Normal (BCN) for $\mu = 5$, $\sigma = 0.5$ *and* $v = 0.5$. On the left, by the function implemented, and on the right, by the package gamlss.

Table 1. Quantiles of the Box-Cox Normal distribution (BCN) for $\mu = 5$, $\sigma = 0.5$ *and* $v = 0.5$ by the implemented function (proposal) and by the gamlss package.

Quantiles	BCN (proposal)	BCCG (gamlss)
5%	1.733778	1.733778
10%	2.309639	2.309639
25%	3.456099	3.456099
50%	5.000099	5.000099
75%	6.828465	6.828465
90%	8.717180	8.717180
95%	9.957671	9.957671

4. Conclusions

In this work, functions for random numbers generation, probability density function, cumulative distribution function, and quantile function associated to a given probability of a distribution of the BCS class were presented.

Comparisons made using the gamlss routine, which has particular cases of some distributions of the BCS class, provided the same values.

Thus, these functions will be used to elaborate a routine for data analysis with asymmetric distributions in order to fit both experimental and regression models. The authors are using the R package and routines to make functions available.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Ferrari, S.L.P. and Fumes, G. (2017) Box-Cox Symmetric Distributions and Applications to Nutritional Data. *AStA—Advances in Statistical Applications to Nutrition Data*, **101**, 321-344. https://doi.org/10.1007/s10182-017-0291-6

[2] Box, G.E.P. and Cox, D.R. (1964) An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B*, **26**, 211-243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

[3] Fumes-Ghantous, G., Ferrari, S.L.P. and Corrente, J.E. (2018) Box-Cox t Random Intercept Model for Estimating Usual Nutrient Intake Distributions. *Statistical Methods & Applications*, **27**, 715-734. https://doi.org/10.1007/s10260-018-00438-6

[4] Vanegas, L.H. and Paula, G.A. (2016) Log-Symmetric Distributions: Statistical Properties and Parameter Estimation. *Brazilian Journal of Probability and Statistics*, **30**, 196-220. https://doi.org/10.1214/14-BJPS272

[5] Rigby, R.A. and Stasinopoulos, D. (2006) Using the Box-Cox t Distribution in GAMLSS to Model Skewness and Kurtosis. *Statistical Modelling*, **6**, 209-229. https://doi.org/10.1191/1471082X06st122oa

[6] Cole, T. and Green, P.J. (1992) Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood. *Statistics in Medicine*, **11**, 1305-1319. https://doi.org/10.1002/sim.4780111005

[7] Voudouris, V., Gilchrist, R., Rigby, R.A., Sedgwick, J. and Stasinopoulos, D.M. (2012) Modelling Skewness and Kurtosis with BCPE Density in GAMLSS. *Journal of Applied Statistics*, **39**, 1279-1293. https://doi.org/10.1080/02664763.2011.644530

[8] Stasinopoulos, D.M., Rigby, R.A. and Akantziliotou, C. (2008) Instructions on How to Use the GAMLSS Package in R. http://www.gamlss.com/wp-content/uploads/2013/01/gamlss-manual.pdf