The Power Normal distribution

Rui Gonçalves
LIAAD-INESCTEC and Faculty of Engineering of the University of Porto, R. Dr. Roberto Frias, 4200-65 Porto.
E-mail: rjasg@fe.up.pt

Abstract. The Inverse Box-Cox (BC) transformation (see [1]) produces the Power Normal (PN) distribution family that includes the log-normal and the normal distributions, see [2], [4] and [3]. The Box-Cox power transformation aims to transform data to approximate normality (truncated normal (TN)) therefore the knowledge of the its inverse scale is of major importance. In this paper we consider the univariate PN distribution, and, because in many applications there are more than one variable and correlation we also consider the bivariate PN distribution (BPN), see [6] and [5]. We give some important results concerning both distributions. In [6] it is given a formula to approximate the ordinary moments of the PN distribution, herein, we give an exact formula for the ordinary integer moments of the PN distribution. Regarding the BPN distribution, we calculate the marginal probability density functions (pdf) and we note that they are not univariate PN distributed however the conditional pdf is PN univariate. This is also true for the corresponding functions of the transformed scale, the TN distribution. The correlation curve or curve of the conditional moments \( E(X_2|X_1 = x_1) \) that is useful to characterize the structure of the correlation is very well fitted by a power law model \( f(x) = ax^b + c \).

1. Introduction
The BC transformation[1] of a positive random variable (rv) is defined by, \( Y = (X^\lambda - 1)/\lambda \) when \( \lambda \neq 0 \) and \( Y = \log(X) \) if \( \lambda = 0 \), where \( \lambda \) is the transformation parameter. The BC is commonly used, for instance in time series analysis as a way to stabilize the sample variance. When BC transformation works well the transformed variable distribution is close to the normal or truncated normal distributions. A question may be asked, what kind of probability distribution is the one obtained by inverting the BC transformation? The answer is the PN distribution. The PN family of distributions was first noted by Goto and Inoue[2]. In that seminal paper the authors investigate some of its properties, namely they give an expression for the mean in terms of infinite series. Taylor [7] proposes a measure of location on the non symmetrical original distribution using the inverse Box-Cox transformation on the center of the transformed data.

A very important issue in Box-Cox literature has to do with parameter estimation. What transformation parameter works better for the data? In [4], the authors address this problem and conduct a simulation study considering two methods of estimation, one that does not take into account the truncation point and another method that includes the truncation. The authors conclude that estimating \( \lambda \) using the first method, normality was not rejected for samples of large size \( (n \geq 50) \). They also conclude that the standardized truncation parameter \( k \) and the size of the sample \( (n) \) were the most influencing factors on the mean bias (MB) and mean square error (MSE) on the estimates of \( \lambda, \mu \) and \( \sigma \) showing that no consistent estimators can be found using the first method. Using the second method was not possible to estimate for samples of...
small size and the author concludes that the probable cause for that is the correlation between the estimators of $\lambda$, $\mu$ and $\sigma$. The second method also revealed less MB and MSE than that of the first method except for the case $\lambda < 0$. Goto et. al. [5] extend their work on the PN distribution to the bivariate case (BPN) including expressions for the ordinary moments, marginal and conditional moments. Regarding the parameter estimation in the bivariate case they conclude that it seems difficult the estimation allowing truncation to be used in practice. More recently, in [6], the authors give expressions that approximates the ordinary moments of the PN distribution. In the same paper are also presented expressions for the cumulative distribution function (CDF) and quantile functions and it is proven that the correlation in the inverse scale is always smaller than the transformed one.

This paper is organized as follows. In the 2nd section we obtain an expression to the ordinary moments of $X$ when $r/\lambda$ is an integer. In the 3rd section we present the BPN distribution and we calculate the corresponding marginal and conditional distributions. The marginal distributions are found not to be univariate PN. In the 4th section we give an expression for the conditional mean and we note that the correlation curve for the BPN distribution has a remarkable fit to a power law model. In 5th (last) section we present the final remarks and we give an hint of future work on this theme.

2. Power Normal distribution

Given the random variable $Y$ normally distributed, if we invert the BC transformation on this variable we obtain $X = (\lambda Y + 1)^{1/\lambda}$ if $\lambda \neq 0$ and $X = e^Y$ if $\lambda = 0$. Remember that $X > 0$ therefore $Y > -1/\lambda$ for $\lambda > 0$ and $Y < -1/\lambda$ for $\lambda < 0$, so only when $\lambda = 0$ $Y$ is normally distributed. The PN family of distributions has its pdf given by,

$$f_X(x) = \frac{x^{\lambda-1}}{\sigma A(\lambda, \mu, \sigma)} \phi \left( \frac{x^{(\lambda)} - \mu}{\sigma} \right), \quad x > 0. \tag{1}$$

where $x^{(\lambda)} = x^{\lambda-1}/\lambda$. Using the same notation of [5] we define,

$$A(\lambda, \mu, \sigma) = \begin{cases} \Phi(k), & \lambda > 0 \\ 1, & \lambda = 0 \\ \Phi(-k), & \lambda < 0 \end{cases}$$

where $k$ is the standardized truncation constant of the TN distribution, $k = (\lambda \mu + 1)/(\lambda \sigma)$, $\phi$ is the standard normal pdf and $\Phi$ is the standard normal CDF. Since $X > 0$ then $Y$ has a truncated normal (TN) distribution with pdf given by,

$$g(Y, \mu, \sigma, \lambda) = \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (Y - \mu)^2 \right]. \tag{2}$$

Let us consider the case $\lambda > 0$. Note that $\mu$ and $\sigma$ are the mean and standard deviation of a normal random variable associated to the TN with truncation at $-1/\lambda$. Treating $Y$ as a truncated normal variable may be avoidable if $\mu$ is large so that $P(Y < -1/\lambda)$ is negligible. If not then we can always add a sufficiently large constant $c$ to $X$ and consider the variable $X + c$ instead. However in [9], it is reported that this operation can lead to poorly behaved likelihood functions which are important for estimation purposes.
3. Ordinary moments of the PN distribution

Let $X$ be a PN distributed random variable with parameters $\lambda, \mu, \sigma$, it is straightforward to calculate the moments of $X$ of order $r$ for which $r/\lambda$ is an integer ($n = r/\lambda$). We have $E(X^r) = E[(\lambda Y + 1)^n]$ that is given by,

$$E[(\lambda Y + 1)^n] = \frac{1}{A(\lambda, \mu, \sigma)\sqrt{2\pi}} \int_{-1/\lambda}^{+\infty} (\lambda y + 1)^n \exp\left[-\frac{1}{2} \left(\frac{y - \mu}{\sigma}\right)^2\right] dy =$$

$$= \frac{1}{A(\lambda, \mu, \sigma)} \int_{-1/\lambda}^{+\infty} \phi(y) (\lambda y + 1)^n dy =$$

$$= \frac{1}{A(\lambda, \mu, \sigma)} \sum_{i=0}^{n} \frac{n!}{i!} \int_{-k}^{+\infty} (\lambda y + 1)^i \phi(y) dy =$$

$$= \frac{1}{A(\lambda, \mu, \sigma)} \sum_{i=0}^{n} \frac{n!}{i!} \left(\frac{\lambda}{\mu}\right)^i \int_{-k}^{+\infty} (\lambda z + 1)^i \phi(z) dz =$$

$$= \frac{1}{A(\lambda, \mu, \sigma)} \sum_{i=0}^{n} \frac{n!}{i!} \left(\frac{\lambda}{\mu}\right)^i \sum_{j=0}^{i} \binom{i}{j} \int_{-k}^{+\infty} z^j \phi(z) dz .$$

We used the substitution $z = \frac{y - \mu}{\sigma}$ so that $dz = \frac{dy}{\sigma}$. Note that $k = \frac{1 + \mu \lambda}{\mu \sigma}$. In [6] it is given an approximation in the form of an infinite series for any ordinary moment of $X$ using the Chebychev-Hermite polynomials.

4. The bivariate Power Normal distribution

For the bivariate case of the PN distribution we consider a pair of correlated left truncated normal random variables $Y_1$ and $Y_2$ with truncation parameters $\lambda_1 > 0$ and $\lambda_2 > 0$ respectively and with variance-covariance matrix $\Sigma$ given by,

$$\Sigma = \begin{pmatrix} \sigma^2_1 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma^2_2 \end{pmatrix}$$

where $\rho$ is the correlation parameter between $Y_1$ and $Y_2$. The random vector $(X_1, X_2)$ has a BPN distribution and its joint pdf is,

$$g(x_1, x_2) = \frac{x_1^{\lambda_1-1} x_2^{\lambda_2-1}}{A(\lambda, \mu, \Sigma)} f(x_1^{(\lambda_1)}, x_2^{(\lambda_2)})$$

for $x_1, x_2 > 0$. We will consider the case where $0 < \lambda_1, \lambda_2 < 1$. The function $f$ is the BTN pdf,

$$f(y_1, y_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp\left\{-\frac{Q(y_1, y_2)}{2}\right\}$$

where,

$$Q(y_1, y_2) = \frac{1}{1 - \rho^2} \left\{\left(\frac{y_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho \left(\frac{y_1 - \mu_1}{\sigma_1}\right) \left(\frac{y_2 - \mu_2}{\sigma_2}\right) + \left(\frac{y_2 - \mu_2}{\sigma_2}\right)^2\right\} .$$

The truncated proportional constant term $A(\lambda, \mu, \Sigma)$ depends on the truncation values for $Y_1, -1/\lambda_1$ and $Y_2, -1/\lambda_2$. For the interval $0 < \lambda_1 < 1$ and $0 < \lambda_2 < 1$ using Table 1 of [5] $A$ is given by,
\[ A(\lambda, \mu, \Sigma) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi_2(x_1, x_2) dx_1 dx_2. \]

where the standardized truncation points \( k_1 \) and \( k_2 \) are given by,

\[ k_j = \frac{\lambda_j \mu_j + 1}{\lambda_j \sigma_j}, \quad j = 1, 2. \]

Note that \( \phi_2 \) is the bivariate standard normal but in this paper we assume that the variables are not independent, they are correlated with correlation \( 0 < |\rho| < 1 \). Therefore,

\[ \phi_2(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)} \right\} . \]

Integrating 3 on each variable respectively we obtain the marginal pdf’s,

\[ g_j(x_j) = \frac{x_j^{\lambda_j-1}}{\sigma_j A(\lambda, \mu, \Sigma)} \phi \left( \frac{x_j^{(\lambda_j)} - \mu_j}{\sigma_j} \right) \Phi \left[ \frac{1}{\sqrt{1-\rho^2}} \left( \rho \frac{x_j^{(\lambda_j)} - \mu_j}{\sigma_j} + k_j \right) \right]. \]

When we integrate in \( x_1 \) then \( j = 2 \) and \( i = 1 \) and when we integrate in \( x_2 \) we have \( j = 1 \) and \( i = 2 \). As we can see, the marginal pdf’s are not PN univariate distributed. If we look to the other scale, the BTN distribution, its marginal distributions are also not TN distributed. The conditional pdf associated to the BPN pdf is,

\[ g_{x_2|x_1}(x_2) = \frac{g(x_1, x_2)}{g_1(x_1)} = \frac{x_2^{\lambda_2-1} f(x_2^{(\lambda_2)} | x_1^{(\lambda_1)})}{\Phi \left[ \frac{1}{\sqrt{1-\rho^2}} \left( \rho \frac{x_2^{(\lambda_2)} - \mu_2}{\sigma_2} + k_2 \right) \right]} \]

where,

\[ f(x_2^{(\lambda_2)} | x_1^{(\lambda_1)}) = \frac{1}{\sqrt{2\pi \sigma_2 \sqrt{1-\rho^2}}} \exp \left\{ -\frac{1}{2\sigma_2^2(1-\rho^2)} \left( x_2^{(\lambda_2)} - \mu_2 - \rho \frac{\sigma_2}{\sigma_1}(x_1^{(\lambda_1)} - \mu_1) \right)^2 \right\} . \]

5. Fitting the Correlation curve of the PN distribution

The graphical representation of set of points of the form \( (x_1, E[X_2|X_1 = x_1]) \) is usually called correlation curve (see [10]) in this case of \( X_2 \) given \( X_1 = x_1 \). As it is well known, if the variables were joint normal then the correlation curves would be lines. In the case of the joint PN distribution the calculation of the correlation curve can only be done numerically. That is exactly what we do in this section. In Figure 1, we show a representation of the correlation curve for a pair of PN random variables with parameters, \( \mu_1 = 2.5, \mu_2 = 1.2, \sigma_1 = 1, \sigma_2 = 0.5 \) and for \( \lambda_1 = 0.55 \) and \( \lambda_2 = 0.7 \). The correlation value is \( \rho = 0.75 \).

The used model is a power law function with two parameters, \( ax^b + c \). The estimated coefficients and their 95% confidence bounds are given in Table 1.

Concerning the goodness of fit statistics we have obtained \( SSE = 0.07194, R^2 = 1 \), Adjusted \( R^2 = 1 \) and \( RMSE = 0.01203 \). The statistics of the fitting reveal a root mean square error (RMSE) of 0.01203. Since that RMSE is so low we can say that the fit is quite good. Consequently we can say that the functional relation between \( X_2 \) and \( X_1 \) is very close to a Power Law.
Figure 1. Conditional mean $E(X_2|X_1 = x_1)$, fit and 95% confidence bounds.

Table 1. Parameters estimates for the General Power model: $f(x) = ax^b + c$.

| Coefficients | estimate   | 95% confidence bounds |
|--------------|------------|------------------------|
| $a$          | 0.5177     | (0.5152, 0.5202)       |
| $b$          | 0.7496     | (0.7185, 0.7207)       |
| $c$          | 0.7797     | (0.7738, 0.7856)       |

6. Summary and future work
In this paper we consider the PN distribution that is obtained using the inverse BC transformation. We present some characteristics of the PN family of distributions. We give a formula to calculate ordinary moments when $r/\lambda$ is integer. The bivariate PN is also considered and we calculate the marginal and conditional distributions associated to the BPN. We observe that the marginal distributions are not univariate PN distributed. We also calculate the conditional mean associated to the conditional distribution and we conclude that the correlation curve, $(x_1, E(X_2|X_1) = x_1)$ is very close to a power law function. This fact may be used to evaluate the possibility of applying the Box-Cox transformation. The future work about the PN distribution will be, for example, like in the work of Goto et al. [4] to conduct a simulation study and use the Maximum Likelihood estimation of the Power Normal’s parameters both in the univariate and multivariate cases. Another line of research is to find a better functional model for the correlation curve since although the fitting statistics are very good the residual plot shows some tendency along the curve.

References
[1] Box G E and Cox D R 1964 J. Roy. Stat. Soc. Ser. B, 26 211.
[2] Goto, M and Inoue, T 1980 Journal of Biometrics 1 28.
[3] Goto, M., Matsubara, Y. and Tsuchiya, Y. 1983 Rep. of Stat. Appl. Res. 30(3), 8-28.
[4] Goto, M. and Inoue, T. and Tsuchiya, Y. 1984 Bull. of Infor. and cyber., 21 1.
[5] Goto, M. and Hamasaki, T. 2002 Bull. of Infor. and Cyber. 34 1 29.
[6] Freeman, J and Modarres, R 2006 Stat. and Prob. Lett. 76 764.
[7] Taylor, J 1985 J. of the Ame. Stat. Asso. 80:390, 427-432.
[8] Taylor, J 1986 J. of the Ame. Stat. Asso. 81:393, 114.
[9] Atkinson, A, Pericchi, R and Smith, R 1991 J. Roy. Stat. Soc. B 53 2 473.
[10] 1990 Murteira, B, Probabilidade e Estatística (McGraw-Hill: Lisboa).