Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/120203

Power-Models for Proportions with Zero/One Excess.

Guillermo Martínez-Flórez[1], *Heleno Bolfarine*[2] *and Héctor W. Gómez*[3,*]

[1] Departamento de Matemáticas y Estadística, Facultad de Ciencias Básicas, Universidad de Córdoba, Córdoba, Colombia.
[2] Departamento de Estatística, IME, Universidade de São Paulo. São Paulo, Brazil.
[3] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile.

Received: 6 Nov. 2017, Revised: 8 Jan. 2018, Accepted: 13 Feb. 2018
Published online: 1 Mar. 2018

Abstract: In this paper we consider power distributions for modeling proportions or rates with zero and/or one inflation as an alternative to beta regression. The model considered is a mixture between a Bernoulli type process for explaining the zero and/or one excess and a truncated power-normal model for explaining the continuous response. The maximum likelihood approach is considered for parameter estimation. Observed and expected information matrices are derived, illustrating interesting aspects of the likelihood approach. Given the flexibility of the power-normal distribution, we are able to show, in a practical scenario, the better performance of our proposal over the beta regression model.

Keywords: Censoring; power-normal distribution; maximum likelihood.

1 Introduction

Statistical models that are used to explain response variables in the $(0,1)$ interval has received increasing attention in the literature. Among others, we mention in [1], [2] and [3]. Extensions of these models to situations where response values are in the intervals $[0,1]$, $[0,1)$ or $(0,1]$ have also been reported in [4]. Response variables of these type include, for example, the proportion of deaths caused by smoking, the proportion of income tax spend in education, the proportion of family income spent on food and so on. The situation of a response variable with zero/one inflation is related to the data set on the percentage of infant (less than one year old) deaths with not well defined causes in Brasilian counties during the year 2000 collected by government agency DATASUS. Out of the 5561 observations collected, we have a total of 3367 zeros and 174 ones, which most certainly should be incorporated into the study. To deal with this more complicated scenario an extension of the beta regression model was considered in [4], leading to quite satisfactory results.

In this paper, we propose an alternative approach to deal with the data set described above. It is based on an extension of the tobit censored model (see, [5]) on the interval $[0,1]$, to incorporate zero/one inflation. It is considered that part of the zeros and/or ones come from a Bernoulli type model that links possible zero and/or one excess with a group of covariates that may have influence on probability of their occurrence. Moreover the continuous responses are modeled by using the power-normal distribution ([6], [7]), which is more flexible than the normal distribution in terms of asymmetry and kurtosis with well behaved maximum likelihood estimators, for which regularity conditions are satisfied since Fisher information matrix is nonsingular at vicinity of symmetry. Moreover, since the tobit model depends on the normal distribution function the extension we propose depends on the distribution function of the power-normal distribution which is almost as simple to deal with as is the normal distribution function. An alternative is to use the skew-normal distribution (see, [8]), which presents singular information matrix (dealt with properly in [9]) and presents a not so simple to deal with distribution function (see, [10]).

The paper is organized as follows. In Section 2 we present a brief review on asymmetric power distributions emphasizing the power-normal and log-power-normal distributions. Section 3 is devoted to extending models in Section 2 to censored situations with emphasis on the doubly censored power-normal and doubly censored log-power-normal models. In Section 4 the models in Section 3 are considered for the situation of zero and/or one inflation by considering mixtures between Bernoulli

* Corresponding author e-mail: hector.gomez@uantof.cl

and a doubly censored (log) power-normal (TPN) models. Parameter estimation is dealt with by using the maximum likelihood approach. Observed and expected (Fisher) information matrices are derived and shown to be nonsingular. A real data analysis is considered from where we concluded that the mixture Bernoulli/TPN model is significantly better than the modified Beta model.

2 Asymmetric distributions

Recent work by [6] and [7] reveal that power distributions can be a viable alternative for modeling asymmetric data. In its more general form, the density function for the standard version of this model can be written as

$$h(z;\alpha) = \alpha f(z)\{F(z)\}^{\alpha-1}, \quad z \in \mathrm{R}, \ \alpha \in \mathrm{R}^{+}, \quad (1)$$

where F is an absolutely continuous distribution function with density function $f = dF$. We refer to this model as the alpha-power distribution and use the notation $Z \sim AP(\alpha)$. This is an alternative to asymmetric models with asymmetry and kurtosis coefficients far from the ones observed with the normal distribution, as is the case with the skew-normal distribution (see, [8]). Parameter α (taken as a real positive number) is a shape parameter that controls the amount of asymmetry in the distribution. As shown in the papers above, parameter α is also in control of the amount of kurtosis in the model. A real data application reported in [7] illustrates the fact that the use of asymmetric models can be a viable alternative to data transformation, which typically makes it difficult to interpret the transformed variables so that it probably will lead to erroneous inference.

In the particular case where $F = \Phi(\cdot)$, we have the family of distributions called the power-normal (PN) distribution (see, [6]), and denoted $PN(\alpha)$. Earlier considerations of the above alpha-power model are reported in [11] (nonparametric testing) and in [12] in hydrological contexts.

If Z is a random variable from a standard $AP(\alpha)$ distribution, then the location-scale extension of Z is obtained from the transformation $X = \xi + \sigma Z$, where $\xi \in R$ and $\sigma \in R^{+}$, is a scale parameter.

For this extension we use the notation $X \sim AP(\xi,\eta,\alpha)$. In [7] have shown that the Fisher information matrix is not singular at $\alpha = 1$, indicating that large sample properties of the maximum likelihood estimators are valid, including the large sample behavior of the likelihood ratio statistics (LRS). That is, $\alpha = 1$ (normality) can be tested using the large sample chi-square distribution for the LRT. Notice that this model can be extended by considering $\xi_i = \mathbf{x}_i'\beta$ to replace ξ, where β is an unknown vector of regression coefficients and $\mathbf{x}_i$ a vector of known regressors hopefully correlated with the response vector.

In the special case where $f = \phi(\cdot)$ and $F = \Phi(\cdot)$, the density and distribution functions of the standard normal distribution, respectively, the standard power-normal distribution follows.

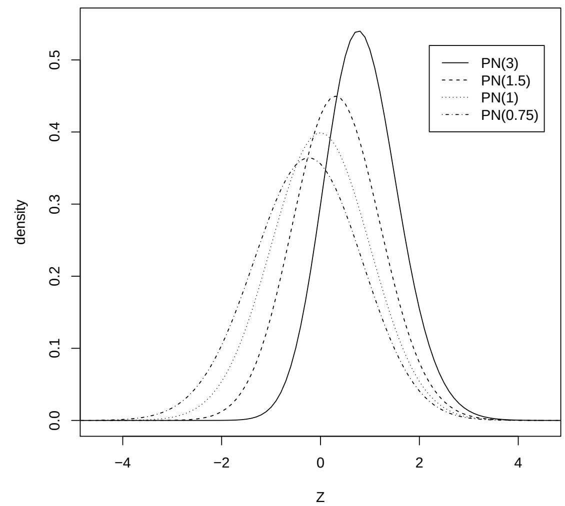


Fig. 1: Density $PN(\alpha)$ for $\alpha = 0.75$ (dashed line), 1 (dotted line), 1.5 (dashed-dotted line) and 3 (solid line).

Figure 1 depicts plots for the PN density for several values of α.

If Z is a random variable from a standard $PN(\alpha)$ distribution then the location-scale extension of Z, $X = \xi + \eta Z$, where $\xi \in R$ and $\eta \in R^{+}$, has probability density function given by

$$\varphi_F(x;\xi,\eta,\alpha) = \frac{\alpha}{\eta}\phi\left(\frac{x-\xi}{\eta}\right)\left\{\Phi\left(\frac{x-\xi}{\eta}\right)\right\}^{\alpha-1}. \quad (2)$$

We will denote this extension by using the notation $X \sim PN(\xi,\eta,\alpha)$.

Asymmetry and kurtosis ranges reported in [7] reveals [-0.6115,0.9007] and [1.7170,4.3556], respectively, against [-0.9995,0.9995] and [3,3.8692] for the skew-normal model.

For the power-normal, the survival and risk functions are, respectively, given by

$$S(t) = 1 - \{\Phi(t)\}^{\alpha}$$

and

$$h(t) = \frac{\varphi_\Phi(t;\alpha)}{S(t)} = \alpha h_\phi(t)\frac{\{\Phi(t)\}^{\alpha-1} - \{\Phi(t)\}^{\alpha}}{1-\{\Phi(t)\}^{\alpha}},$$

where $h_\phi(t)$ is the risk index of the normal distribution and $\varphi_\Phi(t;\alpha)$ is the density function of the power normal distribution. Therefore, the cumulative risk function and the inverse risk index are given by

$$H(t,\alpha) = -\log[S(t)]$$

and

$$r(t) = \frac{\varphi_{\Phi}(t;\alpha)}{\mathscr{F}_T(t;\alpha)} = \alpha\frac{\phi(t)}{\Phi(t)} = \alpha r_{\phi}(t),$$

that is, the inverse risk index T is proportional to the inverse risk index of the normal distribution $r_{\phi}(t)$.

2.1. The log-alpha-power distribution

The distribution of the lifetime of an equipment and the concentration of a chemical element in soil (blood or water) samples is typically described by the log-normal distribution. In many of these situations, however, the asymmetry of the distribution as well as its kurtosis can be above or below the expected for the log-normal model, reason why it is necessary to think of a more flexible model that achieves such deviation in modeling positive data. In [10], extended the log-skew-normal (LSN) model to the analysis of continuous data with the possibility of a discrete response. However, for this type of data, the censored part involves the distribution function of the response variable, which makes likelihood maximization and Bayesian posterior derivation somewhat complicated to deal with. To contour this difficulty [10] proposed using Laguerre quadrature which may critically depends on the number of knots chosen and be sometimes unstable.

As an alternative to the log-skew-normal model to fit positive data from such experimental situations, we consider the log-power-normal (LPN) distribution, which contains as a special case the log-normal model. The advantage of such model is that it has an extra shape parameter which makes it more flexible for fitting data coming from experiments such as the ones described above.

We say that a random variable Y, with support in R^+, follows a univariate log-alpha-power distribution with parameter α, that we denote $Y \sim LAP(\alpha)$, if the transformed variable $X = \log(Y) \sim AP(\alpha)$.

The pdf for the random variable $Y \sim LAP(\alpha)$ can be written as

$$h_Y(y;\alpha) = \frac{\alpha}{y} f(\log(y))\{F(\log(y))\}^{\alpha-1}, \quad y \in \mathrm{R}^+, \quad (3)$$

where F is an absolutely continuous distribution function with density function $f = dF$. We refer to this model as the standard log-alpha-power distribution.

In the special case where $f = \phi(\cdot)$ and $F = \Phi(\cdot)$, the standard log-power-normal distribution follows, with density function given by

$$h_Y(y;\alpha) = \frac{\alpha}{y}\phi(\log(y))\{\Phi(\log(y))\}^{\alpha-1}, \quad y \in \mathrm{R}^+, \quad (4)$$

which we denote by $Y \sim LPN(\alpha)$. Its cumulative distribution function can be written as

$$H_Y(y;\alpha) = \{\Phi(\log(y))\}^{\alpha}, \quad y \in \mathrm{R}^+. \quad (5)$$

As shown in [13], the Fisher information matrix for the log-power normal location-scale (LPN) version of the model is not singular at $\alpha = 1.0$, the log-normal (LN) situation. Hence, the hypothesis $H_0 : LPN = LN$ can be tested using the chi-square distribution for the likelihood ratio statistics.

3 Censored and truncated PN distributions

It is common in clinical and biological experiments that response variables present upper and lower limits and takes on these limit values for a sizeable fraction of the observed data. For example, breast cancer studies, some patients may present relapses after radiation and hormonal therapies. On the other hand, other patients may never present relapses and become cured.

In a double censored situation, the response variable is restricted to take values on an interval and eventually may take on the boundary values for significant part of the data. The limiting values are typically called detection limits (lower detection limits (LDL) and upper detection limits (UDP)). In such situations we have the tobit doubly censored model.

The tobit model may however not be adequate in situations where the observed values for the continuous part of the data present asymmetry or kurtosis large than what is expected for the normal model. In such situations, the tobit-power-normal model can be a viable alternative to the ordinary tobit-normal model.

3.1. Doubly censored power-normal model

Suppose that $y^* \sim PN(\xi,\eta;\alpha)$. Consider a sample of size n, $(y_1^*, y_2^*, \ldots, y_n^*)$ and that only values of y^* between constants c_0 and c_2. For values of $y^* \le c_0$ only the value c_0 is reported while for value of $y^* \ge c_2$ only the value c_2 is reported. Then, observed data can be written as

$$y_i = \begin{cases} c_0, & \text{if } y^* \le c_1, \\ y_i^*, & \text{if } c_0 < y_i^* < c_2, \\ c_2, & \text{if } y^* \ge c_2 \end{cases}$$

$i = 1,2,\ldots,n$. The resulting sample is said to be a doubly censored PN sample. For observations $y_i = c_0$, we have that

$$Prob[y_i = c_0] = Prob[y_i^* \le c_0] = \{\Phi(z_0)\}^{\alpha},$$

where $z_0 = (c_0 - \xi)/\eta$, while for $y_i^* = c_2$ we have that

$$Prob[y_i = c_2] = Prob[y_i^* \ge c_2] = 1 - \{\Phi(z_2)\}^{\alpha},$$

where $z_2 = (c_2 - \xi)/\eta$. For continuous responses, that is, $c_0 < y_i^* < c_2$, we have that $y_i \sim PN(\xi,\eta,\alpha)$. This random variable, we denote by $PNDC(\xi,\eta,\alpha)$. Particularly, for $\alpha=1$, the model reduces to the doubly censored tobit model.

Denoting by $\sum_0$, $\sum_1$ and $\sum_2$, the sums corresponding to $y^*\le c_0$, $c_0<y_i^*<c_2$ and $y^*\ge c_2$ respectively, then, the log-likelihood function corresponding to a sample of size n for estimating $\theta=(\xi,\eta,\alpha)'$ can be written as

$$\ell(\theta;\mathbf{Y})=\alpha\sum_0\log[\Phi(z_0)]+\sum_2\log\left[1-\{\Phi(z_2)\}^{\alpha}\right]+\sum_1\{\log(\alpha)-\log(\eta)+\log(\phi(z_{1i}))+(\alpha-1)\log(\Phi(z_{1i}))\},$$

where $z_{1i}=(y_i-\xi)/\eta$, $i=1,\ldots,n$. Hence, the elements for the score function are given by

$$U(\xi)=-\frac{1}{\eta}\sum_0 r(z_0)+\frac{1}{\eta}\sum_1\{z_{1i}-(\alpha-1)w_{1i}\}+\frac{1}{\eta}\sum_2 h(z_2),$$

$$U(\eta)=-\frac{1}{\eta}\sum_0 r(z_0)z_0+\frac{1}{\eta}\sum_1\left\{-1+z_{1i}^2-(\alpha-1)z_{1i}w_{1i}\right\}+\frac{1}{\eta}\sum_2 z_2h(z_2),$$

$$U(\alpha)=\sum_0\log[\Phi(z_0)]+\sum_1\left\{\frac{1}{\alpha}+\log(\Phi(z_{1i}))\right\}-\alpha^{-1}\sum_2\log(\Phi(z_2))w_2^{-1}h(z_2),$$

where $w_2=\frac{\phi(z_2)}{\Phi(z_2)}$, $w_{1i}=\frac{\phi(z_{1i})}{\Phi(z_{1i})}$, and h,r are the risk and inverse risk functions, respectively.

Maximum likelihood estimators for the parameters are then solutions to the equations $U(\xi)=0$, $U(\eta)=0$ and $U(\alpha)=0$, which require numerical procedures. It can be shown that the observed information entries are then given by

$$j_{\xi\xi}=\frac{1}{\eta^2}\sum_0 r(z_0)\{z_0+\alpha^{-1}r(z_0)\}+\frac{1}{\eta^2}\sum_1\{1+(\alpha-1)[z_{1i}w_{1i}+w_{1i}^2]\}+\frac{1}{\eta^2}\sum_2\{h(z_2)[-z_2+(\alpha-1)w_2+h(z_2)]\},$$

$$j_{\eta\xi}=\frac{1}{\eta^2}\sum_0 r(z_0)\{-1+z_0^2+\alpha^{-1}z_0r(z_0)\}+\frac{1}{\eta^2}\sum_1\{2z_{1i}+(\alpha-1)[-w_{1i}+z_{1i}^2w_{1i}+z_{1i}w_{1i}^2]\}+\frac{1}{\eta^2}\sum_2\{h(z_2)[1-z_2^2+(\alpha-1)z_2w_2+z_2h(z_2)]\},$$

$$j_{\eta\eta}=\frac{1}{\eta^2}\sum_0 z_0r(z_0)\{-2+\alpha^{-1}z_0r(z_0)+z_0^2\}+\frac{1}{\eta^2}\sum_2\{z_2h(z_2)[2-z_2^2+(\alpha-1)z_2w_2+z_2h(z_2)]\}+\frac{1}{\eta^2}\sum_1\{-1+3z_{1i}^2+(\alpha-1)[-2z_{1i}w_{1i}+z_{1i}^2w_{1i}^2+z_{1i}^3w_{1i}]\},$$

$$j_{\alpha\xi}=\frac{1}{\alpha\eta}\sum_0 r(z_0)+\frac{1}{\eta}\sum_1 w_{1i}-\frac{1}{\eta}\sum_2\{h(z_2)[\alpha^{-1}+\log(\Phi(z_2))[1+\alpha^{-1}w_2^{-1}h(z_2)]]\},$$

$$j_{\alpha\eta}=\frac{1}{\alpha\eta}\sum_0 z_0r(z_0)+\frac{1}{\eta}\sum_1 z_{1i}w_{1i}-\frac{1}{\eta}\sum_2\{z_2h(z_2)[\alpha^{-1}+\log(\Phi(z_2))[1+\alpha^{-1}w_2^{-1}h(z_2)]]\},$$

$$j_{\alpha\alpha}=\frac{1}{\alpha^2}\sum_1 1+\alpha^{-2}\sum_2\{w_2^{-2}\log^2(\Phi(z_2))h(z_2)[\alpha w_2+h(z)]\}.$$

The expected (Fisher) information matrix follows then by taking expectations of the above components (multiplied by n^{-1}), it is important in the sense that the asymptotic distribution of the maximum likelihood estimators is asymptotically normal with the asymptotic variance as the inverse of the Fisher information matrix.

3.2. Doubly censored log-power-normal model

In cases the response variable takes only positive values with, we can consider the transformation $Z=log(Y)$, where $Z\sim N(\xi,\eta^2)$.

Considering now that $Z\sim PN(\xi,\eta,\alpha)$, we obtain the log-power-normal model with parameters ξ, η and α, $Y\sim LPN(\xi,\eta,\alpha)$. The density function for this model is given by: $\varphi_{LPN}(y;\xi,\eta,\alpha)=\varphi_{\Phi}(\log(y);\xi,\eta,\alpha)/y$, $y>0$. The corresponding distribution function is given by $\mathscr{F}_Y(y;\alpha)=\{\Phi((log(y)-\xi)/\eta)\}^{\alpha}$. If data are in the interval $[0,\infty)$, censored at zero and with high positive asymmetry we can replace y by $y+1$ given that logarithm of $c_0=0$ does not exists. For doubly censored data we use the notation $LPNDC(\xi,\eta,\alpha)$.

The log-likelihood function for model LPNDC with $c_0=0$ is given by $\ell_{LPN}(\theta;\mathbf{Y})=-\sum_1\log(y+1)+\ell(\theta;\log(\mathbf{Y}+\mathbf{1}))$, where $\ell(.)$ is the likelihood function for model PNDC, with $z_0=-\xi/\eta$, $z_{1i}=(\log(y_i+1)-\xi)/\eta$ and $z_2=(\log(c_2+1)-\xi)/\eta$. The score functions and the observed information matrix can be obtained from the corresponding ones from the PNDC model, replacing

$h(z_2)$ by $h_{LPN}(z_2) = h(log(c_2+1))/y$ and $r(z_0)$ by $r_{LPN}(z_0) = r(z_0)/y$ where $h(\cdot)$ and $r(\cdot)$ are risk and inverse risk functions for model PN.

4 The Bernoulli/doubly censored power-normal mixtures model

4.1. Mixture model

For the response variables distributed in the interval $[0,1]$, the zero and one ($c_0 = 0$ and $c_2 = 1$) doubly censored tobit model may not be the best choice because of zero/one inflation which may require asymmetric models to be able to capture such special features.

Therefore, we introduce the mixture between a discrete and a continuous response variable so that the continuous variable follows a power-normal distribution. We consider that the point mass at zero can be modeled by a Bernoulli random variable with parameter γ, namely $Ber(y;\gamma)$, and that responses between zero and one can be modeled by alpha-power (or log-alpha-power) distribution with parameter $\theta = (\xi,\eta,\alpha)'$. The density function corresponding to this model is given by

$$g(y_i) = \begin{cases} p(1-\gamma), & \text{if } y_i = 0, \\ (1-p)\frac{\varphi_F(y_i,\xi,\eta,\alpha)}{\{F(z_2)\}^\alpha - \{F(z_0)\}^\alpha}, & \text{if } 0 < y_i < 1, \\ p\gamma, & \text{if } y_i = 1. \end{cases}$$

where $0 < p,\gamma < 1$, $\eta,\alpha > 0$ and $\xi \in \mathbb{R}$. $\varphi_F(y_i,\xi,\eta,\alpha)$ denotes the density function of a power-normal distribution. As a consequence of the above construction, it can be noted that $Prob[y=0] = p(1-\gamma)$ and $Prob[y=1] = p\gamma$. The cumulative distribution function of y_i can then be written as:

$$\mathscr{F}_Y(y_i;\xi,\eta,\alpha) = \begin{cases} p(1-\gamma), & \text{if } y_i \le 0, \\ p(1-\gamma) + (1-p)\frac{\{F(z_i)\}^\alpha - \{F(z_0)\}^\alpha}{\{F(z_2)\}^\alpha - \{F(z_0)\}^\alpha}, & \text{if } 0 < y_i < 1, \\ 1, & \text{if } y_i \ge 1. \end{cases}$$

4.1.1. Estimation

We consider initially that $F = \Phi$, the distribution function of the normal distribution, so that we have a mixture between a discrete Bernoulli random variable with parameter γ and the $PN(\xi,\eta,\alpha)$ distribution. We denote this model by $MBPN(p,\gamma,\xi,\eta,\alpha)$. We consider a random sample of size n, $\mathbf{y} = (y_1,\ldots,y_n)^T$, from the distribution $MBPN(p,\gamma,\xi,\eta,\alpha)$. Denoting by $n_0 = \sum_{i=1}^n I_0(y)$, $n_1 = \sum_{i=1}^n I_1(y)$ and $n_{01} = \sum_{i=1}^n I_{\{0,1\}}(y)$, where $I_A(y)$ is the indicator function for the set A. Then, the log-likelihood function for $\theta = (p,\gamma,\xi,\eta,\alpha)$ given $\mathbf{Y}$ can be written as:

$$\begin{aligned}\ell(\theta;\mathbf{Y}) = {} & n_{01}\log(p) + (n-n_{01})\log(1-p) + n_1\log(\gamma) + \\ & n_0\log(1-\gamma) + \sum_1 \{\log(\alpha) - \log(\eta) + \log(\phi(z_i)) + \\ & (\alpha-1)\log(\Phi(z_i)) - \log(\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha)\}\end{aligned}$$

where, $z_i = (y_i - \xi)/\eta$, $i = 1,\ldots,n$. Hence, using an approach similar as that in [7], the elements for the score function are given by

$$U(p) = \frac{n_{01}}{p} - \frac{n-n_{01}}{1-p}, \qquad U(\gamma) = \frac{n_1}{\gamma} - \frac{n_0}{1-\gamma},$$

$$\begin{aligned}U(\xi) = (n-n_{01})\left\{\frac{\overline{z} - (\alpha-1)\overline{w}}{\eta} + \right. \\ \left. \frac{\varphi_\Phi(c_2,\theta) - \varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\},\end{aligned}$$

$$\begin{aligned}U(\eta) = -(n-n_{01})\left\{\frac{1-\overline{z^2} + (\alpha-1)\overline{zw}}{\eta} - \right. \\ \left. \frac{z_2\varphi_\Phi(c_2,\theta) - z_0\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\},\end{aligned}$$

$$\begin{aligned}U(\alpha) = (n-n_{01})\left\{\overline{u} + \frac{1}{\alpha} - \right. \\ \left. \frac{\{\Phi(z_2)\}^\alpha\log(\Phi(z_2)) - \{\Phi(z_0)\}^\alpha\log(\Phi(z_0))}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\},\end{aligned}$$

where $w_i = \phi(z_i)/\Phi(z_i)$ and $u_i = \log\{\Phi(z_i)\}$.
Hence, the MLE for parameter $\theta = (\xi,\eta,\alpha)'$, is obtained by solving the system of equations which follows by equating above derivatives (scores) to zero. Then, we obtain the solutions $\hat{p} = \frac{n_{01}}{n}$ and $\hat{\gamma} = \frac{n_1}{n_{01}}$, corresponding, respectively, to the proportions of zeros and ones and the proportions of ones in the subsample of zeros and ones. It can be shown that $\hat{p}$ is an unbiased estimator for p.

To $\theta_1 = (\xi,\eta,\alpha)'$ the system of equations has no analytical solution and has to solved by numerical methods such as the Newton-Raphson procedure.

4.2. Observed information matrix

The observed information matrix follows from the second derivatives of the log-likelihood function which are denoted by $j_{pp}, j_{\gamma p}, j_{\gamma\gamma}, j_{\xi\xi}, j_{\xi\eta},\ldots, j_{\alpha\alpha}$, and are given by

$$j_{pp} = \frac{n_{01}(1-2p) + np^2}{p^2(1-p)^2}, \qquad j_{\gamma\gamma} = \frac{n_1(1-2\gamma) + n_{01}\gamma^2}{\gamma^2(1-\gamma)^2},$$

$$j_{\xi p}=j_{\eta p}=j_{\alpha p}=j_{\xi\gamma}=j_{\eta\gamma}=j_{\alpha\gamma}=j_{\gamma p}=0,$$

$$\begin{aligned}j_{\xi\xi}=&\frac{n-n_{01}}{\eta}\left\{\frac{1+(\alpha-1)\left[\overline{w^2}+\overline{zw}\right]}{\eta}-\right.\\&\left.\eta\left[\frac{\varphi_\Phi(c_2,\theta)-\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}\right]^2\right\}-\\&\frac{n-n_{01}}{\eta}\frac{[z_2-(\alpha-1)w_2]\varphi_\Phi(c_2,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}+\\&\frac{n-n_{01}}{\eta}\frac{[z_0-(\alpha-1)w_0]\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha},\end{aligned}$$

$$\begin{aligned}j_{\xi\eta}=&\frac{n-n_{01}}{\eta^2}\left\{2\overline{z}+(\alpha-1)\left[\overline{zw^2}-\overline{z^2w}-\overline{w}\right]\right\}\\&-\frac{n-n_{01}}{\eta}\frac{[z_2^2-(\alpha-1)w_2z_2-1]\varphi_\Phi(c_2,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}\\&+\frac{n-n_{01}}{\eta}\frac{[z_0^2-(\alpha-1)w_0z_0-1]\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}\\&-(n-n_{01})\frac{(\varphi_\Phi(c_2,\theta)-\varphi_\Phi(c_0,\theta))}{[\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha]^2}\cdot\\&\cdot(z_2\varphi_\Phi(c_2,\theta)-z_0\varphi_\Phi(c_0,\theta)),\end{aligned}$$

$$\begin{aligned}j_{\eta\eta}=&\frac{n-n_{01}}{\eta^2}\left\{3\overline{z^2}-1-(\alpha-1)\left[2\overline{zw}-\overline{z^3w}-\overline{z^2w^2}\right]\right.\\&\left.-\eta^2\left[\frac{z_2\varphi_\Phi(c_2,\theta)-z_0\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}\right]^2\right\}\\&-\frac{n-n_{01}}{\eta}\frac{[z_2^3-(\alpha-1)w_2z_2^2-2z_2]\varphi_\Phi(c_2,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}\\&+\frac{n-n_{01}}{\eta}\frac{[z_0^3-(\alpha-1)w_0z_0^2-2z_0]\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha},\end{aligned}$$

$$\begin{aligned}j_{\xi\alpha}=&(n-n_{01})\left\{\frac{\overline{w}}{\eta}\right.\\&+\frac{(\varphi_\Phi(c_2,\theta)-\varphi_\Phi(c_0,\theta))\{\Phi(z_2)\}^\alpha\log(\Phi(z_2))}{[\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha]^2}\\&\left.-\frac{(\varphi_\Phi(c_2,\theta)-\varphi_\Phi(c_0,\theta))\{\Phi(z_0)\}^\alpha\log(\Phi(z_0))}{[\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha]^2}\right\}\\&-(n-n_{01})\frac{[\alpha^{-1}+\log(\Phi(z_2))]\varphi_\Phi(c_2,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}\\&+(n-n_{01})\frac{[\alpha^{-1}+\log(\Phi(z_0))]\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha},\end{aligned}$$

$$\begin{aligned}j_{\eta\alpha}=&(n-n_{01})\left\{\frac{\overline{wz}}{\eta}+\right.\\&\frac{(z_2\varphi_\Phi(c_2,\theta)-z_0\varphi_\Phi(c_0,\theta))(\{\Phi(z_2)\}^\alpha\log(\Phi(z_2))}{[\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha]^2}\\&\left.-\frac{(z_2\varphi_\Phi(c_2,\theta)-z_0\varphi_\Phi(c_0,\theta))\{\Phi(z_0)\}^\alpha\log(\Phi(z_0))}{[\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha]^2}\right\}\\&-(n-n_{01})\frac{z_2[\alpha^{-1}+\log(\Phi(z_2))]\varphi_\Phi(c_2,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}\\&+(n-n_{01})\frac{z_0[\alpha^{-1}+\log(\Phi(z_0))]\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha},\end{aligned}$$

$$\begin{aligned}j_{\alpha\alpha}=&(n-n_{01})\left\{\frac{1}{\alpha^2}\right.\\&\left.-\frac{\{\Phi(z_2)\}^\alpha\{\Phi(z_0)\}^\alpha[\log(\Phi(z_2))-\log(\Phi(z_0))]^2}{[\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha]^2}\right\}.\end{aligned}$$

4.3. Expected information matrix

The expected (or Fisher) information matrix follows by taking expectation of the elements of the observed information matrix. Considering the quantities $a_{kj}=E\{z^k(\phi(z)/\Phi(z))^j\}$ for $k=0,1,2,3$ and $j=1,2$, $\theta_1=p$, $\theta_2=\gamma$, $\theta_3=\xi$, $\theta_4=\eta$ and $\theta_5=\alpha$, we have that the elements of the Fisher information matrix, denoted

$$i_{\theta_r\theta_p}=n^{-1}E\left\{-\frac{\partial^2\ell(\theta;\mathbf{y})}{\partial\theta_r\partial\theta_s}\right\},\quad r,s=1,2,3,4,5$$

are given by

$$i_{pp}=\frac{1}{p(1-p)},\qquad i_{\gamma\gamma}=\frac{p}{\gamma(1-\gamma)},$$

$$i_{\xi p}=i_{\eta p}=i_{\alpha p}=i_{\xi\gamma}=i_{\eta\gamma}=i_{\alpha\gamma}=i_{\gamma p}=0,$$

$$\begin{aligned}i_{\xi\xi}=&\frac{1-p}{\eta}\left\{\frac{1+(\alpha-1)[a_{02}+a_{11}]}{\eta}\right.\\&\left.-\eta\left[\frac{\varphi_\Phi(c_2,\theta)-\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}\right]^2\right\}\\&-\frac{1-p}{\eta}\frac{[z_2-(\alpha-1)w_2]\varphi_\Phi(c_2,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha}\\&+\frac{1-p}{\eta}\frac{[z_0-(\alpha-1)w_0]\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha-\{\Phi(z_0)\}^\alpha},\end{aligned}$$

$$
\begin{aligned}
i_{\xi\eta} &= \frac{1-p}{\eta^2}\{2a_{10}+(\alpha-1)[a_{12}+a_{21}-a_{01}]\}\\
&-\frac{1-p}{\eta}\frac{[z_2^2-(\alpha-1)w_2z_2-1]\varphi_{\Phi}(c_2,\theta)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}}\\
&+\frac{1-p}{\eta}\frac{[z_0^2-(\alpha-1)w_0z_0-1]\varphi_{\Phi}(c_0,\theta)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}}\\
&-(1-p)\frac{(\varphi_{\Phi}(c_2,\theta)-\varphi_{\Phi}(c_0,\theta))z_2\varphi_{\Phi}(c_2,\theta)}{[\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}]^2}\\
&+(1-p)\frac{(\varphi_{\Phi}(c_2,\theta)-\varphi_{\Phi}(c_0,\theta))z_0\varphi_{\Phi}(c_0,\theta)}{[\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}]^2},
\end{aligned}
$$

$$
\begin{aligned}
i_{\eta\eta} &= \frac{1-p}{\eta^2}\{3a_{20}-1-(\alpha-1)[2a_{11}-a_{31}-a_{22}]\}\\
&-(1-p)\left[\frac{z_2\varphi_{\Phi}(c_2,\theta)-z_0\varphi_{\Phi}(c_0,\theta)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}}\right]^2\\
&-\frac{1-p}{\eta}\frac{[z_2^3-(\alpha-1)w_2z_2^2-2z_2]\varphi_{\Phi}(c_2,\theta)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}}\\
&+\frac{1-p}{\eta}\frac{[z_0^3-(\alpha-1)w_0z_0^2-2z_0]\varphi_{\Phi}(c_0,\theta)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}},
\end{aligned}
$$

$$
\begin{aligned}
i_{\xi\alpha} &= (1-p)\left\{\frac{a_{01}}{\eta}\right.\\
&+\frac{(\varphi_{\Phi}(c_2,\theta)-\varphi_{\Phi}(c_0,\theta))\{\Phi(z_2)\}^{\alpha}\log(\Phi(z_2))}{[\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}]^2}\\
&\left.-\frac{(\varphi_{\Phi}(c_2,\theta)-\varphi_{\Phi}(c_0,\theta))\{\Phi(z_0)\}^{\alpha}\log(\Phi(z_0))}{[\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}]^2}\right\}\\
&-(1-p)\frac{[\alpha^{-1}+\log(\Phi(z_2))]\varphi_{\Phi}(c_2,\theta)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}}\\
&+(1-p)\frac{[\alpha^{-1}+\log(\Phi(z_0))]\varphi_{\Phi}(c_0,\theta)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}},
\end{aligned}
$$

$$
\begin{aligned}
i_{\eta\alpha} &= (1-p)\left\{\frac{a_{11}}{\eta}\right.\\
&+\frac{(z_2\varphi_{\Phi}(c_2,\theta)-z_0\varphi_{\Phi}(c_0,\theta))\{\Phi(z_2)\}^{\alpha}\log(\Phi(z_2))}{[\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}]^2}\\
&\left.-\frac{(z_2\varphi_{\Phi}(c_2,\theta)-z_0\varphi_{\Phi}(c_0,\theta))\{\Phi(z_0)\}^{\alpha}\log(\Phi(z_0))}{[\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}]^2}\right\}\\
&-(1-p)\frac{z_2[\alpha^{-1}+\log(\Phi(z_2))]\varphi_{\Phi}(c_2,\theta)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}}\\
&+(1-p)\frac{z_0[\alpha^{-1}+\log(\Phi(z_0))]\varphi_{\Phi}(c_0,\theta)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}},
\end{aligned}
$$

$$
\begin{aligned}
i_{\alpha\alpha} &= (1-p)\left\{\frac{1}{\alpha^2}\right.\\
&\left.-\frac{\{\Phi(z_2)\}^{\alpha}\{\Phi(z_0)\}^{\alpha}[\log(\Phi(z_2))-\log(\Phi(z_0))]^2}{[\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}]^2}\right\},
\end{aligned}
$$

The above expressions have to be computed numerically.

It then follows that the Fisher information matrix for $\theta=(p,\gamma,\xi,\eta,\alpha)$ is given by

$$
I(\theta)=(1-p)\begin{pmatrix}
\frac{1}{p(1-p)^2} & 0 & 0 & 0 & 0\\
0 & \frac{p}{\gamma(1-\gamma)(1-p)} & 0 & 0 & 0\\
0 & 0 & i_{\xi\xi} & i_{\xi\eta} & i_{\xi\alpha}\\
0 & 0 & i_{\xi\eta} & i_{\eta\eta} & i_{\eta\alpha}\\
0 & 0 & i_{\xi\alpha} & i_{\eta\alpha} & i_{\alpha\alpha}
\end{pmatrix}
$$

This result shows that the set of parameters $(p,\gamma)'$ and (ξ,η,α) are orthogonal, so that the information matrix is block orthogonal and can be written as $I(\theta) = Diag\{I_{p,\gamma},I_{\xi,\eta,\alpha}\}$, where $I_{p,\gamma}=Diag\left\{\frac{1}{p(1-p)},\frac{p}{\gamma(1-\gamma)}\right\}$.

Hence, for n large,

$$
\hat{\theta}\xrightarrow{A}N_5(\theta,\Sigma_{\theta\theta}),
$$

meaning that $\hat{\theta}$ is consistent and asymptotically normally distributed with $\Sigma_{\theta\theta}=I(\theta)^{-1}=Diag\{I_{p,\gamma}^{-1},I_{\xi,\eta,\alpha}^{-1}\}=Diag\{\Sigma_{p,\gamma},\Sigma_{\xi,\eta,\alpha}\}$ as the large sample variance. Given that the Fisher information matrix is orthogonal in two blocks, the corresponding parameters can be estimated separately.

The normal approximation $N_5(\theta,\Sigma_{\theta\theta})$ can be used for constructing $(\tau=1-\kappa)$ confidence intervals for θ_r, which are given by $\hat{\theta}_r\mp z_{1-\alpha/2}\sqrt{\hat{\sigma}(\hat{\theta}_r)}$, where $\hat{\sigma}(\cdot)$ is given by the corresponding r-th diagonal element of the matrix $\hat{\Sigma}_{\hat{\theta}\hat{\theta}}$ and $-z_{1-\kappa/2}$ is the $100(1-\kappa/2)$%-quantile of the standard normal distribution.

Considering the parametrization $\delta_1=p\gamma$ and $\delta_0=p-\delta_1$ we can rewrite the previous model as

$$
g(y_i)=\begin{cases}
\delta_0, & \text{if } y_i=0,\\
(1-\delta_0-\delta_1)\frac{\varphi_{\Phi}(y_i,\xi,\eta,\alpha)}{\{\Phi(z_2)\}^{\alpha}-\{\Phi(z_0)\}^{\alpha}}, & \text{if } 0<y_i<1,\\
\delta_1, & \text{if } y_i=1.
\end{cases}
$$

where $0<\delta_0=Prob(y_i=0),\delta_1=Prob(y_i=1)<1$ and $0<\delta_0+\delta_1<1$.

Then the log-likelihood function of $\theta = (\delta_0, \delta_1, \xi, \eta, \alpha)$ given $\mathbf{Y}$ can be written as:

$$\ell(\theta;\mathbf{Y}) = n_0 \log(\delta_0) + n_1 \log(\delta_1) + (n - n_{01}) \log(1 - \delta_0 - \delta_1) + \sum_1 \{\log(\alpha) - \log(\eta) + \log(\phi(z_i)) + (\alpha - 1)\log(\Phi(z_i)) - \log(\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha)\},$$

so that it is found that the elements of the score matrix are given by

$$U(\delta_0) = \frac{n_0}{\delta_0} - \frac{n - n_{01}}{1 - \delta_0 - \delta_1}, \quad U(\delta_1) = \frac{n_1}{\delta_1} - \frac{n - n_{01}}{1 - \delta_0 - \delta_1},$$

$$U(\xi) = (n - n_{01})\left\{\frac{\bar{z} - (\alpha - 1)\overline{w}}{\eta} + \frac{\varphi_\Phi(c_2, \theta) - \varphi_\Phi(c_0, \theta)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\},$$

$$U(\eta) = -(n - n_{01})\left\{\frac{1 - \overline{z^2} + (\alpha - 1)\overline{zw}}{\eta} - \frac{z_2\varphi_\Phi(c_2, \theta) - z_0\varphi_\Phi(c_0, \theta)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\},$$

$$U(\alpha) = (n - n_{01})\left\{\overline{u} + \frac{1}{\alpha} - \frac{\{\Phi(z_2)\}^\alpha \log(\Phi(z_2)) - \{\Phi(z_0)\}^\alpha \log(\Phi(z_0))}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\}.$$

From the first two equations, we obtain the estimator $\hat{\delta}_0 = n_0/n$, the proportion of zeros in the sample and $\hat{\delta}_1 = n_1/n$, the proportions of ones in the sample. The remaining parameters have to be estimated numerically.

In this new model, the Fisher information matrix can be written as

$I(\theta) = Diag\{I_{\delta_0,\delta_1}, I_{\xi,\eta,\alpha}\}$, where the elements of I_{δ_0,δ_1} are given by $i_{\delta_0\delta_0} = \frac{1-\delta_1}{\delta_0(1-\delta_0-\delta_1)}$, $i_{\delta_1\delta_0} = \frac{1}{1-\delta_0-\delta_1}$ and $i_{\delta_1\delta_1} = \frac{1-\delta_0}{\delta_1(1-\delta_0-\delta_1)}$, with $I_{\xi,\eta,\alpha}$ as computed for model $MBPN(p, \gamma, \xi, \eta, \alpha)$. Given this new parametrization, the parameters for the censored and noncensored parts of the model are again orthogonal so that the corresponding MLEs are asymptotically orthogonal and the parameters can be estimated separately.

For n large,

$$\hat{\theta} \xrightarrow{A} N_5(\theta, \Sigma_{\theta\theta}),$$

meaning that $\hat{\theta}$ is consistent and asymptotically normally distributed with $\Sigma_{\theta\theta} = I(\theta)^{-1} = Diag\{I^{-1}_{\delta_0,\delta_1}, I^{-1}_{\xi,\eta,\alpha}\} = Diag\{\Sigma_{\delta_0,\delta_1}, \Sigma_{\xi,\eta,\alpha}\}$ as the large sample variance.

4.4. Censored models for 0 or 1 inflation

Particular cases of the previous model with 0 and 1 inflation, is the situation of zero or one inflation. For the case of zero inflation, the density function is given by:

$$g(y_i) = \begin{cases} \delta_0, & \text{if } y_i = 0, \\ (1 - \delta_0)\frac{\varphi_\Phi(y_i, \xi, \eta, \alpha)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}, & \text{if } 0 < y_i \le 1. \end{cases}$$

where $0 < \delta_0 = Prob(y_i = 0)$ and $0 < \delta_0 < 1$.

Then the log-likelihood function of $\theta = (\boldsymbol{\delta_0}, \xi, \eta, \alpha)$ given $\mathbf{y}$ can be written as:

$$\ell(\theta;\mathbf{Y}) = n_0 \log(\delta_0) + (n - n_0)\log(1 - \delta_0) + \sum_1 \{\log(\alpha) - \log(\eta) + \log(\phi(z_i)) + (\alpha - 1)\log(\Phi(z_i)) + -\log(\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha)\},$$

so that the elements of score function are given by

$$U(\delta_0) = \frac{n_0}{\delta_0} - \frac{n - n_0}{1 - \delta_0},$$

$$U(\xi) = (n - n_0)\left\{\frac{\overline{z} - (\alpha - 1)\overline{w}}{\eta} + \frac{\varphi_\Phi(c_2, \theta) - \varphi_\Phi(c_0, \theta)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\},$$

$$U(\eta) = -(n - n_0)\left\{\frac{1 - \overline{z^2} + (\alpha - 1)\overline{zw}}{\eta} - \frac{z_2\varphi_\Phi(c_2, \theta) - z_0\varphi_\Phi(c_0, \theta)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\},$$

$$U(\alpha) = (n - n_0)\left\{\overline{u} + \frac{1}{\alpha} - \frac{\{\Phi(z_2)\}^\alpha \log(\Phi(z_2)) - \{\Phi(z_0)\}^\alpha \log(\Phi(z_0))}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\}.$$

From the first equation, we obtain the estimator $\hat{\delta}_0 = n_0/n$, the proportion of zeros in the sample. The remaining parameters require numerical procedures for their estimation. For the case of one inflation, the likelihood function is given by:

$$g(y_i) = \begin{cases} \delta_1, & \text{if } y_i = 1, \\ (1 - \delta_1)\frac{\varphi_\Phi(y_i, \xi, \eta, \alpha)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}, & \text{if } 0 \le y_i < 1. \end{cases}$$

where $0 < \delta_1 = Prob(y_i = 1)$ and $0 < \delta_1 < 1$. Then the log-likelihood function of $\theta = (\boldsymbol{\delta_1}, \xi, \eta, \alpha)$ given $\mathbf{y}$ can be written as:

$$\ell(\theta;\mathbf{Y}) = n_1 \log(\delta_1) + (n - n_1)\log(1 - \delta_1) + \sum_1 \{\log(\alpha) - \log(\eta) + \log(\phi(z_i)) + (\alpha - 1)\log(\Phi(z_i)) - \log(\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha)\},$$

so that the elements for the score function are given by

$$U(\delta_1) = \frac{n_1}{\delta_1} - \frac{n-n_1}{1-\delta_1},$$

$$U(\xi) = (n-n_1)\left\{\frac{\overline{z}-(\alpha-1)\overline{w}}{\eta} + \frac{\varphi_\Phi(c_2,\theta)-\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\},$$

$$U(\eta) = -(n-n_1)\left\{\frac{1-\overline{z^2}+(\alpha-1)\overline{zw}}{\eta} - \frac{z_2\varphi_\Phi(c_2,\theta)-z_0\varphi_\Phi(c_0,\theta)}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\},$$

$$U(\alpha) = (n-n_1)\left\{\overline{u}+\frac{1}{\alpha} - \frac{\{\Phi(z_2)\}^\alpha\log(\Phi(z_2)) - \{\Phi(z_0)\}^\alpha\log(\Phi(z_0))}{\{\Phi(z_2)\}^\alpha - \{\Phi(z_0)\}^\alpha}\right\}.$$

From the first equation, we obtain the estimator $\hat{\delta}_1 = n_1/n$, the proportions of ones in the sample. The remaining parameters have to be estimated numerically.

4.5. Bernoulli-LPN mixture

Taking now $\varphi_F(y_i,\xi,\eta,\alpha)$ as the density function corresponding to the LPN model, the Bernoulli/LPN model follows, that we denote by $MBLPN(p,\gamma,\xi,\eta,\alpha)$. This model is of great importance in modeling data with positive asymmetry and kurtosis above that expected for the ordinary normal distribution. The log-likelihood function for the reparametrized model can be written as

$$\ell_{MBLPN}(\theta;\mathbf{Y}) = -\sum_{1}\log(y_i) + \ell(\theta;\log(\mathbf{Y})),$$

where $\ell(\cdot)$ is the log-likelihood function for the MBPN model and $\log(Y) = (\log(y_1),\ldots,\log(y_n))'$. The score function and observed and expected information matrices are as given for the MBPN model, where $z_i = (\log(y_i)-\xi)/\eta$.

5 Numerical illustrations

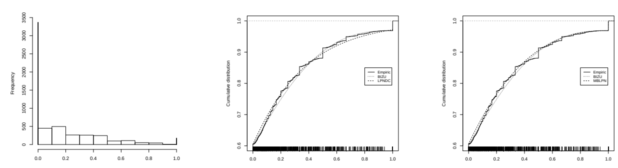
5.1. Real data illustration

In this section we shall illustrate the usefulness of the proposed models by fitting the LPNDC and MBLPN distributions to a real data set. The data set analyzed corresponds to the proportions of infant deaths not well explained in in 5561 Brasilian counties. Data is available for downloading at http:www.datasus.gov.br. The data set contains 3367 zeros (explained deaths) and 174 ones (all unexplained deaths). Data histogram is given in Figure 2 (a) revealing the data behavior.

In [4] studies the type of data described above using a beta regression model with zero ad/or one inflation. It is considered a mixture. It is assumed a mixture of a Bernoulli random variable for the discrete part and a beta regression for the continuous (between zero and one), which is denoted by $BIZU(\delta_0,\delta_1,\xi,\eta)$. For estimating the parameters for the BIZU model, the library GAMLSS available at the R package can be used. We consider fitting also the model LPNDC and the reparameterized model MBLPN.

Given orthogonality for the parameters of the mixed models, maximum likelihood for parameters δ_0 and δ_1 for models BIZU and MBLPN coincide and are given by $\hat{\delta}_0 = 0.6055(0.0066)$ and $\hat{\delta}_1 = 0.0313(0.0023)$. For the continuous part of the BIZU model, maximum likelihood estimates are given by $\hat{\xi} = 0.2974(0.0043)$ and $\hat{\eta} = 0.4562(0.0050)$. On the other hand, for model MBLPN maximum likelihood estimates are given by $\hat{\xi} = 0.6606(0.0066)$, $\hat{\eta} = 0.0215(0.0030)$ and $\hat{\alpha} = 0.0005(0.0004)$.

For the LPNDC model, we have the following estimates: $\hat{\xi} = -0.7145(0.1566)$, $\hat{\eta} = 0.5621(0.0367)$ and $\hat{\alpha} = 4.6830(1.7087)$. The percentage of zeros and ones in the samples are 0.6055 and 0.0313, respectively, so that using the cumulative distribution function are 0.6047 and 0.0284, respectively, revealing good model fit. Furthermore, Figure 2 (b) and (c), depicts the cumulative distribution function for the three models, illustrating the fact that the models considered present good fit for the data set studied.



(a) (b) (c)

Fig. 2: (a) Histogram for death proportions, (b) Graphs: empiric (solid line in stars), LPNDC (dashed line), BIZU (doted line), (c) Graphs: empiric (solid line in stars), MBLPN (dashed line), BIZU (doted line)

5.1.1. Testing nested models

A test to compare the MBLPN and LPNDC models against the BIZU model requires a non-nested approach. Being, F_θ and G_γ two non-nested models, $f(y_i|x_i,\theta)$ and $g(y_i|x_i,\beta)$ two densities corresponding to the non-nested models, the likelihood ratio statistic to compare both models is given by

$$LR(\hat{\theta},\hat{\beta}) \equiv \ell_f(\hat{\theta}) - \ell_g(\hat{\beta}) = \sum_{i=1}^{n} \log \frac{f(y_i|\mathbf{x}_i,\hat{\theta})}{g(y_i|\mathbf{x}_i,\hat{\beta})},$$

which does not follow a chisquare distribution.

To contour this problem, in [14] proposed an alternative approach based on the Kullback-Liebler information criterion, (see, [15]). Based on the distance between each model and the true process generating the data, namely the model $h_0(\mathbf{y}_i, X_i)$, he arrived at the statistic

$$T_{LR,NN} = \frac{1}{\sqrt{n}} \frac{LR(\hat{\theta},\hat{\beta})}{\hat{\omega}^2},$$

where

$$\hat{\omega}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{f(y_i|\mathbf{x}_i,\hat{\theta})}{g(y_i|\mathbf{x}_i,\hat{\beta})}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{f(y_i|\mathbf{x}_i,\hat{\theta})}{g(y_i|\mathbf{x}_i,\hat{\beta})}\right)\right)^2,$$

is an estimator for the variance of $\frac{1}{\sqrt{n}}LR(\hat{\theta},\hat{\beta})$.

Hence, it was shown that, as $n \to \infty$,

$$T_{LR,NN} \xrightarrow{d} N(0,1)$$

under

$$H_0 : \mathbb{E}\left[\log\frac{f(y_i|\mathbf{x}_i,\theta)}{g(y_i|\mathbf{x}_i,\beta)}\right] = 0,$$

that is, models are equivalent. At the 5% level, being $z_{0.975}$ the critical value, we reject model equivalency if $T_{LR,NN} > z_{0.975}$, (that is, $T_{LR,NN} < -z_{0.025}$).

For the data set under study, being F_θ the MBLPN model and G_β, the LPNDC model, Vuong's approach lead to the observed value $T_{LR,NN} = 31.5079$ which is greater than the critical value $z_{0.975} = 1.96$ and hence, the distribution MBLPN is the better model.

Similarly, for comparing models MBLPN and BIZU, we have $T_{LR,NN} = 54.9704$ which favors model MBLPN, leading then to the conclusion that the best model to fit the data is model MBLPN.

6 Concluding remarks

Paper discusses an alternative to the beta regression model in the situation of zero and (or) one inflation. The approach is based on an extension of the tobit model with zero excess considered in [16]. Estimation is based on the likelihood approach and the Fisher information matrix is derived revealing orthogonality between the parameters, which simplifies large sample properties of the maximum likelihood estimators. An study with real data reveals that the model proposed can do even better than extensions of the beta regression model considered in [4].

References

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

[2] Branscum AJ, Johnson WO, Thurmond MC. Bayesian beta regression: applications to household expenditure data and genetic distance between foot-and-mouth deseases viruses. *Australian and New Zealand Journal of Statistics* **49**, 287-301 (2007).

[3] Bayes C, Bazan J, García C. A New Robust Regression Model for Proportions. *Bayesian Analysis* **7**, 841-866 (2008).

[4] Ospina R. Modelo de Regresso Beta Inflacionados. Tese (Doutorado em Estatstica) - Universidade de So Paulo. (2008).

[5] Tobin J. Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24–36 (1958).

[6] Gupta D, Gupta RC. Analyzing skewed data by power normal model. *Test* **17**, 197–210 (2008).

[7] Pewsey A, Gómez, HW, Bolfarine H. Likelihood-based inference for power distributions. *Test* **21**(4), 775-789 (2012).

[8] Azzalini A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178 (1985).

[9] Arellano-Valle RB, Azzalini A. The centred parameterizatiion and related quantities of the skew-t distribution. *Journal of Multivariate Analysis* **99**, 1362-1382 (2008).

[10] Chai H, Bailey K. Use of log-normal distribution in analysis of continuous data with a discrete component at zero. *Statistics in Medicine* **27**, 3643–3655 (2008).

[11] Lehmann EL. The power of rank tests. *Annals of Mathematical Statistics* **24**, 23–43 (1953).

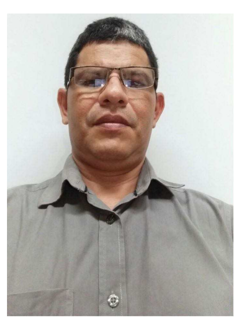
[12] Durrans SR. Distributions of fractional order statistics in hydrology. *Water Resources Research* **28**, 1649–1655 (1992).

[13] Martínez-Flórez G, Bolfarine H, Gómez, HW. The log alpha-power asymmetric distribution with application to air pollution. *Environmetrics* **25**, 44-56 (2014).

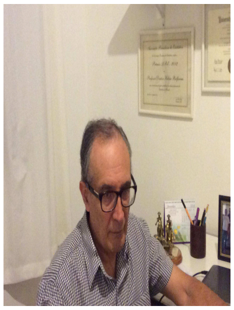
[14] Vuong Q. Likelihood ratio tests for model selection and nonested hypotheses. *Econometrica* **57**, 307-333 (1989).

[15] Kullback S, Leiber RA. On information and sufficiency. *Annals of Mathematical Statistics*. **22**(1), 79-86 (1951).

[16] Moulton L, Halsey NA. A Mixture Model with Detection Limits for Regression Analysses of Antibody Response to Vaccine. *Biometrics*. **51**, 1570–1578 (1995).



Guillermo Matínez-Flórez received the PhD degree in Statitics at the University of São Paulo (2011). He is currently a Titular Professor at the Universidad de Córdoba - Colombia. His research is focused in distribution theory and probability sampling.



Heleno Bolfarine received his PhD degree in Statistics at the University of California, Berkley (1982). He is currently a Titular Professor at the University of São Paulo. His research is focused in measurement error models, survival analysis, Bayesian analysis and distribution theory, among others.

Héctor W. Gómez received the PhD degree in Statistics at the Pontificia Universidad Católica of Chile (2004). He is currently a Titular Professor at the Universidad de Antofagasta - Chile. His research is focused in distribution theory.