Statistics and Its Interface Volume 11 (2018) 237–249

Censored bimodal symmetric-asymmetric families

Guillermo Martínez-Flórez, Heleno Bolfarine, and Héctor W. Gómez*

In this paper, we introduce two new families of distributions that are suitable for fitting unimodal as well as bimodal symmetric and asymmetric censored data. The models extend the skew normal model to bimodal symmetric and asymmetric situations and typically involves less parameters to be estimated than mixtures of normal distributions. Maximum likelihood estimation (MLE) is discussed and Fisher information matrices are derived. Results of a simulation study indicate stable parameter recovery in moderate and large samples. Applications to two real data sets are reported. The first data set is related to results of a study on antiretroviral therapy (HAART) to AIDS patients with strong evidence of bimodality and asymmetry. The second data set (fetal weight of unborn children) presents bimodal symmetry, well captured by the model introduced.

Keywords and phrases: Bimodal distribution, Generalized Gaussian distribution, Kurtosis, Power-normal model, Skew-normal distribution, Skewness.

1. INTRODUCTION

Researchers are often confronted with data for which the recorded continuous response variable has a lower bound and takes on this boundary value for a sizeable fraction of sample observations. One of the circumstances in which this is the case is when the observed values are true zeros. An example is the amount of money spent on a new car last year by families in a certain community. Since some families did not buy a new car last year, their actual expenditure is zero. Another situation occurs when the response variable Y is antibody concentration, which can typically be measured by laboratory techniques, the choice depending on the particular antigen(s) of interest, the method of sample collection, the desired scaling, and the available technology. Regardless of the technique, there is always a concentration value T (lower detection limit or LDL) below which an exact measurement cannot be reported, and is a function of the assay that is employed. In general, T is assumed to be a known constant. When data from an assay is left-censored, the lower detection limit is known and can be used to substitute values for the censored observations, that is the value T. Related to this problem is the blood concentration of HIV-RNA in blood samples ($\log_{10}$ scale) of HIV patients. For some patients, HIV levels are below (or at) the detection limit which, using the Roche Amplicor assay, is 50 copies/ml. That is, values lower than (or equal to) that limit are recorded as 50 copies/ml.

As shown in this paper, an efficient analysis can be undertaken by combining continuous information with binary data (Moulton and Halsey, 1995). The continuous information comes from observations above LDL and the binary data correspond to observations below LDL. This idea was popularized by Tobin (1958) and the resulting model is typically referred to as the Tobit model. For a recent study on asymmetric extensions of the Tobit model see Martínez-Flórez et al. (2013a). Models for limited and censored data based on the mixture between the logpower- normal and Bernoulli-type distributions see Martínez-Flórez et al. (2013b). Another recent study on doubly censored regression models with inflation is Martínez-Flórez et al. (2015).

Recently, to study the relationship between serum antibody neutralization activity (determined by IC50) and the B cell immune response, Chen et al. (2016) used the (ordinary) Tobit model given that the IC50 values can not be observed when they are bellow the lower detection limitation (LDL).

The present paper focuses on extending the censored normal model with no covariates to symmetric and asymmetric bimodal and unimodal situations which, to the best of our knowledge is not in the literature. In the study of Li et al. (2006), the authors concluded that the distribution of HIV-RNA ($\log_{10}$) is bimodal and considered it to be a mixture of two normal distributions, reflecting different responses to highly active antiretroviral therapy. The mixture of two normal distributions can be written as

$$\frac{p}{\sigma_1}\phi\left(\frac{x-\mu_1}{\sigma_1}\right)+\frac{1-p}{\sigma_2}\phi\left(\frac{x-\mu_2}{\sigma_2}\right),$$

where ϕ is the density of the standard normal distribution and $0<p<1$. It is the case, however, that mixtures of distributions is a very controversial topic (Marin et al., 2005) mainly because one has to deal with nonidentifiability issues. For data sets with lower detection limit (Tobit-type censored model), we denote the Tobit extended two-normals mixture model as $\mathrm{CMN}(\mu_1,\sigma_1,\mu_2,\sigma_2,p)$. We consider then an alternative route, which is made possible by extending the usual normal and skew-normal (Azzalini, 1985) models to models

*Corresponding author.

that are able to incorporate a certain degree of asymmetry and bimodality. We use the maximum likelihood approach for parameter estimation and the BIC and CAIC criterions for model comparison.

The paper is organized as follows. In Section 2 we review basic results for bimodal symmetric and asymmetric models in the literature. Section 3 is devoted to an extension of the ordinary normal model to a censored (Tobit-type) flexible model, which can incorporate unimodal as well as bimodal censored normal distributions. Some properties of the distribution are presented. Location-scale extensions are considered and maximum likelihood estimation is discussed. Observed and expected (Fisher) information matrices are derived. The bimodal symmetric flexible extension is considered in Section 4, where maximum likelihood estimation is discussed. An asymmetric extension is considered for the symmetric model and maximum likelihood estimation is discussed and observed and expected information matrices are presented. Applications are considered in Section 5 to a data set from a Colombian Cancer Clinic Center, which illustrates the fact that the censored skew-normal bimodal model can present a much better fit than the solely unimodal models and symmetric bimodal models. A censored (Tobit type) mixture of two normal models is also studied. A second data set (related to the ecographic weight of unborn children) is studied indicating strong evidence of symmetric bimodality. Thus, the models proposed and studied in this paper seem to be useful in practical situations. A simulation study reported in Appendix 3, illustrates good performance of the maximum likelihood estimators under controlled situations.

2. EXISTING BIMODAL ASYMMETRIC MODELS

Azzalini (1985) considers a general representation for an asymmetric distribution, namely,

$$\varphi(z;\lambda) = 2f(z)G(\lambda z), \qquad z, \lambda \in \mathbb{R}, \tag{1}$$

where f is a probability density function (pdf) symmetric around zero and G is an absolutely continuous symmetric distribution function and λ is a parameter which controls the asymmetry in the model.

In the particular case where $f = \phi$ and $G = \Phi$, the density and distribution functions of the standard normal distribution, respectively, we obtain the density function of the "so-called" skew-normal distribution, namely

$$\phi_{SN}(z) = 2\phi(z)\{\Phi(\lambda z)\}, \qquad z \in \mathbb{R}, \tag{2}$$

which we denote by $Z \sim SN(\lambda)$. Some other results related to the skew-normal model appear in Azzalini (1986) and Pewsey (2000), where difficulties with maximum likelihood estimation are pointed out. In particular, for the location-scale extension, it is shown that the Fisher information matrix is singular at $\lambda = 0$. The cumulative distribution function for this model is given by

$$\Phi_{SN}(z) = \Phi(z) - 2T(z,\lambda), \tag{3}$$

where $T(\cdot,\cdot)$ is the Owen(1956) function defined as

$$T(h,a) = \int_0^a \frac{\phi(h)\phi(hx)}{1+x^2}dx,$$

where h is a real number and a is a positive real number.

Several bimodal extensions of the skew normal model (Azzalini, 1985) have been considered in the literature. Kim (2005) introduced the model

$$f(u;\lambda) = k_\lambda \phi(u)\Phi(\lambda|u|) \tag{4}$$

where λ is a real number and k_λ is a normalizing constant. For $\lambda > 0$, Kim shows that model (4) is bimodal. This model became known as the "two-pieces skew-normal model (TN)". Gómez et al. (2011), defines the skew-flexible-normal model with density function given by

$$f(u;\lambda,\delta) = c_\delta \phi(|u|+\delta)\Phi(\lambda u) \tag{5}$$

where δ is a real number, c_δ is a normalizing constant and λ as the asymmetry parameter. Gómez et al. (2011) showed that, for $\delta < 0$, model (5) is bimodal. Arnold et al. (2009), studied a bimodal asymmetric model called "*the extended two-pieces skew-normal model* (ETN)", with density function given by

$$f(u;\lambda,\beta) = 2k_\lambda \phi(u)\Phi(\lambda|u|)\Phi(\beta u) \tag{6}$$

where β and λ are real numbers and k_λ is a normalizing constant. It can be shown that this model is asymmetric and bimodal for certain values of λ and β. Arnold et al. (2009) show that the information matrix for a location-scale extension of this model is singular at $\lambda = \beta = 0$, that is, for the normal distribution, as is the case with Azzalini's skew-normal model. In the special case of $\lambda = \beta = 0$, Arnold et al. (2009) use the iterative approach proposed by Rotnitzky et al. (2000) to find a reparametrization that leads to a nonsingular information matrix. Another type of bimodal distribution, originally applied in a survival analysis context was introduced by Ramires et al. (2016). It was termed, exponentiated log-sinh distribution, which depends on four parameters. Its density function is given by

$$\begin{aligned} f(x,\mu,\sigma,\nu,\tau) = {} & \frac{\tau\nu}{x\sigma\pi}\frac{\cosh(w)}{[\nu^2\sinh^2(w)+1]} \\ & \times \left[\frac{1}{2} + \frac{1}{\pi}\arctan[\nu\sinh(w)]\right]^{\tau-1}, \\ & \tau,\nu,\sigma > 0, \quad x,\mu \in \mathbb{R}, \end{aligned}$$

where $w = \frac{\log(x)-\mu}{\sigma}$ and μ and σ are location-scale parameters τ is a shape parameter and ν controlling uni(bi)-modality of the distribution. Moreover, sinh(cosh) are the hyperbolic functions sin(cos).

3. FLEXIBLE NORMAL MODEL

Normal distribution has been widely utilized in a multitude of scenarios to model continuous a data set that presents a symmetric behaviour about their mean value and with a unique maximum, that is, they are unimodal. For the bimodal case, the mixture of two normal distributions containing five parameters is typically used to estimate.

In this paper, we present a bimodal distribution containing less parameters than the normal mixture in which it is considered that the data comes from a unique population. This model has the characteristic of being symmetric about its location and has pdf given by

$$f(y;\delta) = c_\delta \phi(|y| + \delta) \tag{7}$$

where δ is a real number and $c_\delta = (2(1 - \Phi(\delta)))^{-1}$ is a normalizing constant. We call this model the flexible normal model and we denote it by $FN(\delta)$. Also, for $\delta = 0$ it is obtained the standard normal distribution.

Differentiating equation (7) with respect to y and equating to zero for $\delta < 0$, we obtain

$$\begin{cases} y_1 = \delta, & \text{if } y < 0, \\ y_2 = -\delta, & \text{if } y \geq 0, \end{cases}$$

where $y_1 \in \mathbb{R}^-$ and $y_2 \in \mathbb{R}^+$. Consequently, Y is a bimodal random variable. It then follows that the model is bimodal for δ lower than zero. This model is a special case of the model in (5) for $\lambda = 0$ and is suitable for data with symmetric bimodal behaviour.

The moments of the random variable flexible normal are given as functions of the incomplete moments of the normal distribution which are defined as

$$\mu_r(x) = \int_x^\infty z^r \phi(z) dz.$$

The r-th moment of the random variable Z follows a flexible normal distribution and is then given by

$$\mathbb{E}(Z^r) = c_\delta \sum_{k=0}^{r} \binom{r}{k} (\delta)^{r-k} \mu_0(\delta) \left(1 + (-1)^{r-k}\right). \tag{8}$$

From the normal model with mean μ and variance σ^2, we say that a random variable X follows a location-scale flexible normal distribution if its density function is given by

$$f(x;\mu,\sigma,\delta) = \frac{c_\delta}{\sigma} \phi\left(\left|\frac{x-\mu}{\sigma}\right| + \delta\right), \qquad x \in \mathbb{R}, \tag{9}$$

where μ is a location parameter and $\sigma > 0$ is a scale parameter. we denote it by $FN(\mu,\sigma,\delta)$.

For a sample of size n, $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ where $y_i \sim FN(\mu,\sigma,\delta)$, $i = 1, 2, \cdots, n$, the maximum likelihood estimators can be found by maximizing the log-likelihood function

$$\ell(\delta) = n\log(c_\delta) - n\log(\sigma) - \frac{1}{2}\sum_{i=1}^{n}\left(\left|\frac{y_i - \mu}{\sigma}\right| + \delta\right)^2.$$

Given the complexity of the model, maximization has to be performed numerically. The Fisher information matrix for the location-scale flexible normal distribution, defined as the expected values of the negative of the second derivative of the log-likelihood function, is given by:

$$I(\mu,\sigma,\delta) = \begin{pmatrix} \frac{1}{\sigma^2} & \frac{\delta}{2\sigma^2} & 0 \\ \frac{\delta}{2\sigma^2} & \frac{2}{\sigma^2} + \frac{\delta}{\sigma^2}(\delta + h(\delta)) & \frac{1}{\sigma}(\delta - h(\delta)) \\ 0 & \frac{1}{\sigma}(\delta - h(\delta)) & 1 + h(\delta)(\delta - h(\delta)) \end{pmatrix}.$$

where $h(\delta) = \frac{\phi(\delta)}{1-\Phi(\delta)}$. The columns of this matrix are linearly independent and for $\delta = 0$, which corresponds to the case normal of parameters (μ, σ^2), the determinant of $I(\delta) = \frac{2}{\sigma^3}\left(1 - \frac{3}{\pi}\right) \neq 0$, indicating that the information matrix is nonsingular.

Hence, the regularity conditions are satisfied in general and the usual $\sqrt{n}$-property for the maximum likelihood estimators hold for all μ, σ and δ. Therefore, for large sample sizes,

$$(\hat{\mu}, \hat{\sigma}, \hat{\delta})' \xrightarrow{A} N_3((\mu,\sigma,\delta)', I(\mu,\sigma,\delta)^{-1}),$$

so that the maximum likelihood estimators are consistent and asymptotically normally distributed with covariance matrix equals to the inverse of the Fisher information matrix.

3.1 Flexible censored normal model

In this section we extend the ordinary normally distributed Tobit model to the normal bimodal situation.

Consider now that y^* follows the standard flexible censored normal distribution and that $(y_1^*, y_2^*, ..., y_n^*)$ is a random sample where only values of y^* greater than a constant c are recorded. For values of $y^* \leq c$ only the value c is recorded. Hence, the observed values are

$$y_i = \begin{cases} y_i^*, & \text{if } y_i^* > c, \\ c, & \text{otherwise}, \end{cases}$$

for $i = 1, 2, ..., n$. The resulting sample is a left censored sample. In this case we say that random variable Y follows a flexible censored normal model and we denote it by $CFN(\delta)$.

The case $c = 0$ and censoring at the left can be seen as a special case of the general situation above. For $c = 0$, since $Pr[y_i = 0] = Pr[y_i^* < 0] = \frac{1}{2}$, we have that the density function of y is

$$f(y) = \begin{cases} \frac{1}{2}, & \text{if } y \leq 0, \\ c_\delta \phi(y + \delta), & \text{if } y > 0. \end{cases}$$

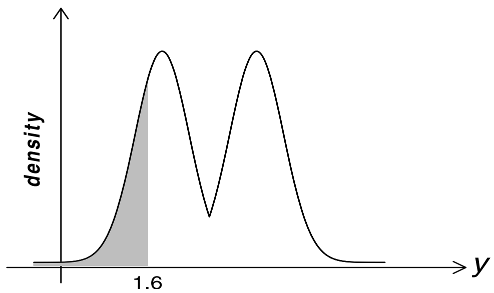


Figure 1. Density $CFN(1.5, 1.0, -1.75)$ censored at the left (grey color).

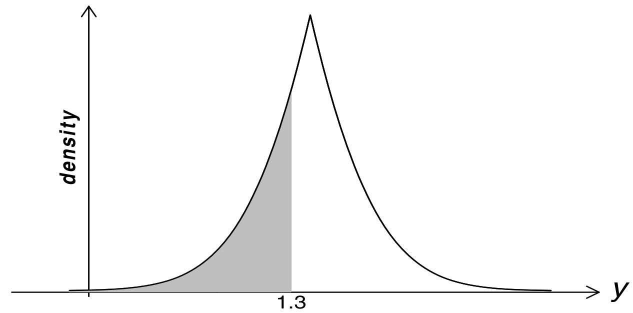


Figure 2. Density $CFN(1.5, 1.0, 1.5)$ censored at the left (grey color).

As demonstrated above, the density function of this random variable is bimodal for values of δ smaller than zero and unimodal for values of δ greater than zero. For $\delta = 0$ we have the ordinary Tobit normal model.

For $c = 0$, the r-th moment of the random variable Z following a flexible censored normal model is then given by

$$\mathbb{E}(Z^r) = c_\delta \sum_{k=0}^{r} \binom{r}{k} (-\delta)^{r-k} \mu_0(\delta). \tag{10}$$

It follows that the expectation and variance of the random variable Z are given by

$$\begin{aligned} \mathbb{E}(Z) &= c_\delta[\phi(\delta) - \delta(1 - \Phi(\delta))] \quad \text{and} \\ Var(Z) &= c_\delta \Bigg[\left(1 + \frac{1}{2}\delta^2\right)(1 - \Phi(\delta)) \\ &\quad + \phi(\delta)(1 - \delta - c_\delta \phi(\delta)) \Bigg]. \end{aligned}$$

3.1.1 The CFN location-scale extension

Defining

$$y_i = \begin{cases} x_i, & \text{if } x_i > 0, \\ 0, & \text{otherwise,} \end{cases}$$

one obtains the left censored flexible normal distribution, which we denote by $X \sim CFN(\mu, \sigma, \delta)$. Figure 1 and 2 depicts plots for the distribution density for two values of δ.

Moreover, the r-th moment of the random variable Y follows a location-scale flexible censored normal distribution is given by: $\mathbb{E}(Y^r) = \sum_{k=0}^{r} \binom{r}{k} \mu^k \delta^{r-k} \mathbb{E}(Z^{r-k})$, where Z follows the standard flexible censored normal distribution.

3.2 Estimation

We denote by $\sum_0$ the sum over censored observations and $\sum_1$ the sum over noncensored observations. Therefore, for observations $y_i = 0$, we have that

$$Pr(y_i = 0) = Pr(x_i \le 0) = c_\delta \left\{ 1 - \Phi\left(\frac{\mu + \sigma\delta}{\sigma}\right) \right\}$$

and for $y_i > 0$, the distribution of y_i is equal to the distribution of x_i that is, $y_i \sim FN(\mu, \sigma, \delta)$. Hence, for a random sample of size n, namely $\mathbf{y} = (y_1, y_2, ..., y_n)$, the log-likelihood function for $\theta = (\mu, \sigma, \delta)'$ is given by

$$\begin{aligned} \ell(\theta; \mathbf{y}) &= \sum_0 \log\left[c_\delta \left(1 - \Phi\left(\frac{\mu + \sigma\delta}{\sigma} \right) \right) \right] \\ &\quad + \sum_1 \left[\log(c_\delta) - \log(\sigma) + \log(\phi(|z_i| + \delta)) \right], \end{aligned}$$

where $z_i = \frac{y_i - \mu}{\sigma}$, $i = 1, ..., n$. Therefore, the score function for the model parameters is given by

$$\begin{aligned} U(\mu) &= -\frac{n_0}{\sigma} \frac{\phi\left(\frac{\mu+\sigma\delta}{\sigma}\right)}{1 - \Phi\left(\frac{\mu+\sigma\delta}{\sigma}\right)} \\ &\quad + \frac{1}{\sigma}\sum_1 \frac{y_i - \mu}{\sigma} - \frac{\delta}{\sigma}\sum_1 sgn(y_i - \mu), \\ U(\sigma) &= \frac{n_0\mu}{\sigma} \frac{\phi\left(\frac{\mu+\sigma\delta}{\sigma}\right)}{1 - \Phi\left(\frac{\mu+\sigma\delta}{\sigma}\right)} \\ &\quad - \frac{n_1}{\sigma} + \frac{1}{\sigma}\sum_1 \left(\frac{y_i - \mu}{\sigma}\right)^2 + \frac{\delta}{\sigma}\sum_1 \left|\frac{y_i - \mu}{\sigma}\right|, \\ U(\delta) &= -n_0 \frac{\phi\left(\frac{\mu+\sigma\delta}{\sigma}\right)}{1 - \Phi\left(\frac{\mu+\sigma\delta}{\sigma}\right)} \\ &\quad + \frac{n\phi(\delta)}{1 - \Phi(\delta)} - \sum_1 \left|\frac{y_i - \mu}{\sigma}\right| - n_1\delta, \end{aligned}$$

where n_0 and n_1 is the number of censored and noncensored observations, respectively, and "sgn" is the *sign* function. The system of equations obtained by equating the scores to zero has no closed form solution and needs to be solved numerically. Therefore, the maximum likelihood estimator for θ can be obtained using iterative procedures such as a Newton-Raphson or quasi-Newton type algorithms. There are however, other numerical procedures based on the expected (Fisher) information matrix that can be used. One possibility is to use "optim" or "maxLik" in R (see R Development Core Team, 2012).

Procedures "optim" or "maxLik" optimize the log-likelihood function using the "optim" program in R functions based on the Nelder-Mead, quasi-Newton and conjugate-gradient algorithms. For smooth objective functions, those methods are faster, requiring less iterations for convergence. Some methods utilize the score function (derivative of the likelihood) and when this function is not provided they compute the gradients by using finite differences. The Nelder-Mead in "optim" works fairly well for nondifferentiable functions as is the case with the absolute value ($|\cdot|$) and the sign function. The observed and expected information matrices are given in Appendix 1.

3.2.1 Simulation study

For studying the behavior of the maximum likelihood estimator of the shape parameter $\hat{\delta}$ for small and moderate samples, a small scale Monte Carlo simulation study was conducted for the censored flexible normal model. Parameter values were fixed at $\delta = -5, \ -2, \ 2$ and 5, were sample sizes considered were $n = 50, \ 150$ and 1000. A total of 1000 repetitions were considered for each sample size. Censoring percentage $p\%$ considered were: 5%, 10%, 20% and 40%. To simulate deviates from the flexible normal model, the inversion method using the distribution function

$$\mathcal{F}_F(z;\delta) = \begin{cases} c_\delta \Phi(z-\delta), & \text{if } z < 0, \\ \frac{1}{2} + c_\delta \left[\Phi(z+\delta) - \Phi(\delta)\right], & \text{if } z \geq 0, \end{cases} \tag{11}$$

was employed.

To evaluate estimators performance for point estimates, the following quantities were considered: relative bias (RB), defined as (bias/true parameter value) and the squared root of the mean squared error ($\sqrt{MSE}$), which is the mean over all samples of the squared bias. Maximum likelihood parameter estimates were computed by using procedure "optim" (using "nlminb") of statistical package R.

Negative bias is noticed for parameter δ, which becomes small as the sample size increases (see Appendix 3). Bias is greater for $\delta < 0$ which seems to be explained by the bimodal character of the generated data.

It can also be noted that the $\sqrt{MSE}$ for maximum likelihood estimators (MLE) of δ decreases as sample size increases which is expected since estimators are consistent. To improve small sample performance, bias corrections such as bootstrap or jackknife could be tried.

4. CENSORED BIMODAL MODELS

We now study two bimodal type censored models, one symmetric and the other asymmetric, based on the two-pieces skew-normal model and the extended two-pieces skew-normal model proposed by Kim (2005) and Arnold et al. (2009), respectively.

4.1 The censored bimodal symmetric model

The model proposed by Kim (2005),

$$f(z;\lambda) = k_\lambda \phi(z)\Phi(\lambda|z|), \tag{12}$$

where λ is a real number and $k_\lambda = 2\pi/(\pi + 2\arctan(\lambda))$ is a normalizing constant, is a viable alternative for fitting symmetric bimodal data for $\lambda > 0$. We use the notation $TN(\lambda)$. We can extend the model (12) to the censored set up situation by considering the random variable

$$y_i = \begin{cases} z_i, & \text{if } z_i > 0, \\ 0, & \text{otherwise,} \end{cases}$$

which is left censored and which we denote by $CTN(\lambda)$. Hence, for $\lambda > 0$ we have a censored bimodal symmetric model.

The density function for the right truncated random variable Y is given by

$$f(y|y>c) = \frac{2k_\lambda \phi(y)\Phi(\lambda|y|)}{1 + k_\lambda[\Phi(c) - 0.5 + \pi^{-1}\arctan(\lambda) - 2T(c,\lambda)]}, \tag{13}$$

where $T(\cdot,\cdot)$ is the Owen (1956) function.

For $c = 0$, the moments of the random variable Y can be obtained from the moments of the random variable with density function $CTN(\lambda)$, leading to the following results

$$\begin{aligned}
\mathbb{E}(Y) &= \frac{k_\lambda}{2\sqrt{2\pi}}\left[\frac{\lambda}{\sqrt{1+\lambda^2}} + 1\right], \\
\mathbb{E}(Y^2) &= k_\lambda\left[\frac{1}{4} + \frac{1}{2\pi}\arctan\lambda + \frac{1}{2\pi}\frac{\lambda}{\sqrt{1+\lambda^2}}\right], \\
\mathbb{E}(Y^3) &= \frac{k_\lambda}{2\sqrt{2\pi}}\left[2 + \frac{3\lambda + 2\lambda^3}{(1+\lambda^2)^{3/2}}\right], \\
\mathbb{E}(Y^4) &= k_\lambda\left[\frac{3}{4} + \frac{3}{2\pi}\arctan\lambda + \frac{1}{2\pi}\frac{\lambda(2\lambda^2+5)}{(1+\lambda^2)^2}\right].
\end{aligned}$$

Hence, the variance of the random variable Y is given by

$$\begin{aligned}
\sigma^2 = {} & \frac{k_\lambda}{4\pi(\pi + 2\arctan\lambda)} \\
& \times \left((\pi + 2\arctan\lambda)^2 + \frac{4\lambda}{\sqrt{1+\lambda^2}}(\pi + \arctan\lambda)\right. \\
& \left. - \pi\left(\frac{2\lambda^2+1}{1+\lambda^2}\right)\right).
\end{aligned}$$

4.1.1 Maximum likelihood estimation

The location-scale extension of Kim (2005) can be written as

$$f(x;\xi,\eta,\lambda) = \frac{k_\lambda}{\eta}\phi\left(\frac{x-\xi}{\eta}\right)\Phi\left(\lambda\left|\frac{x-\xi}{\eta}\right|\right),$$

where $\xi \in R$ is a location parameter, $\eta \in R^+$, is a scale parameter and $k_\lambda = 2\pi/(\pi + 2\arctan(\lambda))$ is the normalizing constant. Being $\sum_0$ the sum over censored observations and $\sum_1$ the sum over noncensored observations, the log-likelihood function is given by

$$\ell(\theta;\mathbf{Y}) = \sum_0 \log\left[k_\lambda \Phi_{CTN}\left(-\frac{\xi}{\eta}\right)\right] + \sum_1 [\log(k_\lambda) - \log(\eta) + \log(\phi(z_i)) + \log(\Phi(\lambda|z_i|))],$$

where $\Phi_{CTN}(\cdot) = \Phi(\cdot) - 0.5\Phi_{SN}(\cdot)$ and $z_i = \frac{y_i-\xi}{\eta}$. Hence, the elements of the score function are given by

$$U(\xi) = -\frac{n_0}{\eta}\frac{\phi_{CTN}\left(\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)} + \frac{1}{\eta}\sum_1 \frac{y_i-\xi}{\eta} + \frac{\lambda}{\eta}\sum_1 sgn(y_i-\xi)\frac{\phi\left(\frac{y_i-\xi}{\eta}\right)}{\Phi\left(\left|\frac{y_i-\xi}{\eta}\right|\right)},$$

$$U(\eta) = \frac{n_0\xi}{\eta^2}\frac{\phi_{CTN}\left(\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)} - \frac{n_1}{\eta} + \frac{1}{\eta}\sum_1\left(\frac{y_i-\xi}{\eta}\right)^2 - \frac{\lambda}{\eta}\sum_1 \frac{y_i-\xi}{\eta}\frac{\phi\left(\frac{y_i-\xi}{\eta}\right)}{\Phi\left(\left|\frac{y_i-\xi}{\eta}\right|\right)},$$

$$U(\lambda) = -\frac{nk_\lambda}{\pi(1+\lambda^2)} + \frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{n_0}{(1+\lambda^2)}\frac{\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)} + \sum_1\left|\frac{y_i-\xi}{\eta}\right|\frac{\phi\left(\frac{y_i-\xi}{\eta}\right)}{\Phi\left(\left|\frac{y_i-\xi}{\eta}\right|\right)},$$

where $\phi_{CTN}(\cdot)$ is as above, and n_0 and n_1 are the number of censored and uncensored observations, respectively. The system of equations obtained by equating scores to zero has no closed form solution and has to be solved numerically, using iterative procedures such as a Newton-Raphson or quasi-Newton type algorithms.

Again a possibility is to use the package "optim" or "maxLik" in R. The expected and observed information matrices to obtain the estimates for the model parameters are given in Appendix 2.

4.2 The censored bimodal asymmetric model

As mentioned in the previous section, the censored bimodal model presented in that section has, as its main feature, the ability to adjust to symmetric bimodal data being then not adequate to situations where data is asymmetric and bimodal. For situations of the latter type we propose using the model studied in Arnold et al. (2009), which we denote by $ETN(\lambda,\beta)$ so that for the location-scale situation

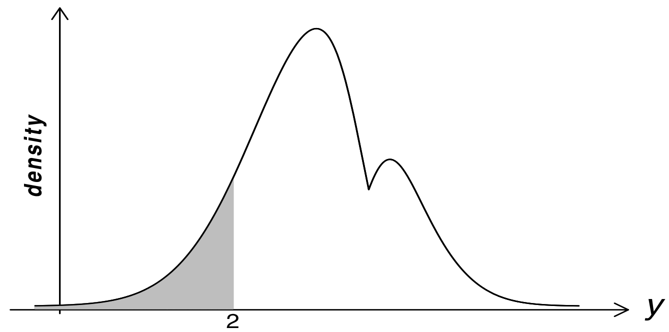


Figure 3. Density $CETN(7.5, 3.5, 3.5, -1.0)$ censored at the left (grey color).

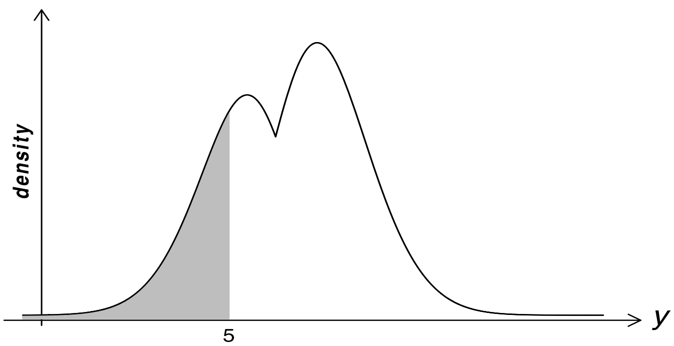


Figure 4. Density $CETN(7.5, 3.5, 1.5, 0.25)$ censored at the left (grey color).

we have that $X \sim ETN(\xi,\eta,\lambda,\beta)$. Considering the censored situation where

$$y_i = \begin{cases} x_i, & \text{if } x_i > 0,\\ 0, & \text{otherwise,}\end{cases}$$

we use the notation $Y \sim CETN(\xi,\eta,\lambda,\beta)$. Figures 3 and 4, depicts density functions for the censored random variable Y for some parameter values.

Hence, for $c = 0$, the contribution for the likelihood of observations smaller than or equal to zero are given by

$$\begin{aligned}\Psi(0) &= Pr(y=0) = Pr(x\le 0)\\ &= 2k_\lambda\left[\frac{1}{2}\left(1-\Phi\left(\frac{\beta\xi}{\eta}\right)\Phi\left(\frac{\xi}{\eta}\right)\right) - T\left(\frac{\xi}{\eta},\beta\right) + T\left(\frac{\xi}{\eta},\lambda\right)\right]\\ &\quad + 2k_\lambda\left[-T\left(\frac{\beta\xi}{\eta},\frac{1}{\beta}\right) + S\left(\frac{\beta\xi}{\eta},\frac{1}{\beta},\lambda\right) - \frac{1}{2\pi}\arctan\left(\frac{\beta\lambda}{\sqrt{1+\beta^2+\lambda^2}}\right)\right]\end{aligned} \tag{14}$$

where S is Steck's function defined as (see Owen, 1956):

$$S(h,a,b) = \int_{-\infty}^{h} T(ax,b)\phi(x)dx,$$

where h is a real number and a, b are positive real numbers.

4.3 The log-likelihood function

For a random sample of size n, $X_1,\ X_2,\ ..., X_n$, the log-likelihood function for the parameter vector $\theta = (\xi, \eta, \lambda, \beta)'$ given the sample $Y_1,\ Y_2, ..., Y_n$ is given by:

$$\ell(\theta; \mathbf{Y}) = \sum_{0} \log(\Psi(0)) + \sum_{1} [\log(2) + \log(k_\lambda) - \log(\eta) + \log(\phi(z_i)) + \log(\Phi(\lambda|z_i|)) + \log(\Phi(\beta z_i))],$$

where $z_i = \frac{y_i - \xi}{\eta}$. The score function and observed and expected information matrices can be obtained by proceeding similarly as in the previous cases. Maximum likelihood estimators obtained by equating score functions to zero have to be obtained using numerical procedures. Behaviour of the information matrix for $\lambda = 0$ and $(\lambda, \beta)' = (0, 0)'$ are similar to those in Arnold et al. (2009). Thus we have a singular information matrix and so model fitness can be performed by using AIC score type statistics.

5. REAL DATA ILLUSTRATIONS

Results of fitting the models discussed in this paper to two real data sets are now reported. The first seems to be better explained by an asymmetric bimodal model while the second by a symmetric bimodal model. Results indicate that models proposed can present better performance than the popular mixture of two normal distributions.

5.1 Illustration I

To illustrate the potential for applications of the models studied in this paper, we consider a sample of 369 HIV patients of which 263 are males and 106 are females, and which have been treated with the HAART therapy for less than one year at the Santander-Colombia Medical service hospital.

This data set contains age, date entering the program, and viral load for the patients. Tests used for diagnosing HIV infection in a particular person require a high degree of both sensitivity and specificity. In Colombia, this is achieved using an algorithm combining two tests for HIV antibodies. If antibodies are detected by an initial test based on the ELISA method, then a second test using the Western blot procedure is used.

Since the measurements come from different laboratories, the HIV-1-RNA quantification could be performed by three different methods: Versant bDNA 3.0® (Bayer), LCx HIV® (Abbott) and Amplicor HIV Monitor v1.5® (Roche), all with lower detection limits (LDL) of 50 copies per ml.

Table 1. Summary statistics based on the variable $\log_{10}$ *HIV-1-RNA for the 157 noncensored men*

$\overline{y}$	s_y^2	$\sqrt{b_1}$	b_2
1.6488	1.7328	0.5213	2.1315

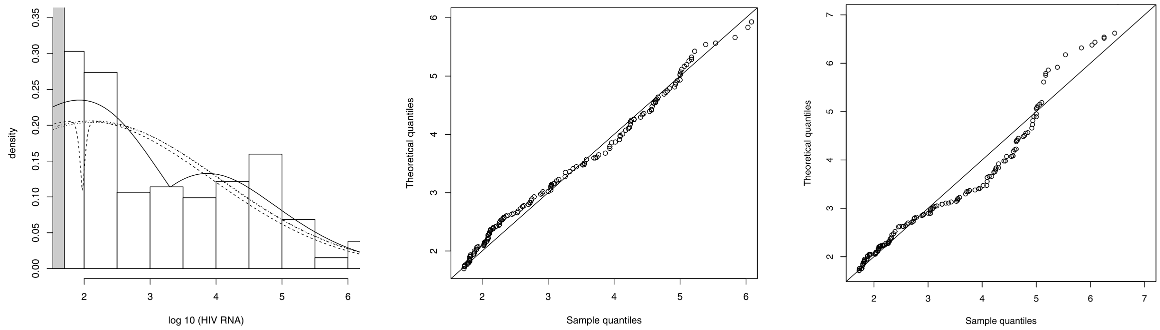
For the male dataset, the average age is 36.19 years, where the youngest male is less than one year old and the oldest is 83 years old. Descriptive statistics for the observations above the detection limit (157 observations or 60% of the full male data) are presented in Table 1. Quantities $\sqrt{b_1}$ and b_2 correspond to sample asymmetry and kurtosis coefficients for values above $\log_{10}(50)$. Statistics indicate that the data set presents high positive asymmetry and low kurtosis compared to the normal model giving indication that the censored-normal model may not be the best choice for fitting the data set. Moreover, Figure 5,-(a) gives strong evidence that the behavior of the variable HIV-1-RNA is bimodal so that a censored bimodal skew-normal model may be the best choice for fitting the HIV data set. To implement a more complete study we consider fitting the following censored models: normal (CN), skew-normal (CSN), bimodal symmetric skew-normal (CTN), flexible-normal (CFN) and bimodal asymmetric skew-normal (CETN).

It can be depicted from Figure 5-(a) that the CSN model seems to adequately fit the asymmetry in the data but not the bimodal nature of the data set, which seems to be best captured by models CTN and CETN. To compare models fit, we use the BIC and CAIC criterions (see Hastie and Tibshirani, (1990)), namely

$$\mathrm{BIC} = -2 * \hat{\ell}(\cdot) + p\log(n) \quad \text{and}$$
$$\mathrm{CAIC} = -2 * \hat{\ell}(\cdot) + p(\log(n) + 1),$$

where p is the number of parameter for the model being considered. The best model is the one with the smallest BIC or CAIC scores. Table 2 presents maximum likelihood estimators, BIC and CAIC values for models CN, CSN, CTN, CFN and CETN, which is the one corresponding to the best (smallest BIC or CAIC) model fitting. Figure 5-(b), presents the QQ-plot for the estimated CETN model indicating an excellent fit for most observations and the Figure 5,-(c), presents the QQ-plot for the estimated CFN model.

For the n=106 women (HIV) infected and under treatment with HAART therapy for no longer than one year, the average age is 34.73 years. Descriptive statistics for the observations above the detection limit are presented in Table 3. Values indicate that the data set presents high positive asymmetry and low kurtosis compared to the normal model giving indication that the censored-normal model may not be the best choice for fitting the viral load data set. Figure 7-(a), presents the QQ-plot for the estimated CN model.



(a) (b) (c)

Figure 5. (a) Histogram for $\log_{10}$ HIV-1-RNA to the 263 team men observation. Models: CETN (solid line), CTN (dashed line), CSN(dotted line) and CN (dashed and dotted line), (b) QQ-plot CETN model and (c) QQ-plot CFN model.

Table 2. Parameter estimates (standard errors) for the fitting censored models

Parameter	CN model	CSN	CTN	CFN	CETN
ξ	0.419(0.135)	1.314(1.515)	0.275(0.023)	0.322(0.006)	1.603(0.120)
η	1.936(0.118)	2.158(0.769)	1.949(0.110)	11.778(1.060)	2.031(0.154)
λ		-0.619 (1.188)	31.694(11.483)	7.273(0.005)	2.232(0.865)
β					-0.766(0.146)
BIC	829.39	834.99	836.15	842.59	826.18
CAIC	831.39	837.99	839.15	845.59	830.19

Table 3. Summary statistics for $\log_{10}$ HIV-1-RNA for the 69 non-censored women observations

$\overline{y}$	s_y^2	$\sqrt{b_1}$	b_2
1.7112	1.4249	0.3549	1.9836

Additionally, the unimodality testing statistics defined by Hartigan (1985), computationally implemented in Hartigan and Hartigan (1985) presents a value of $D = 0.0686$ with corresponding $pvalue = 0.0133$, so that the hypothesis that the viral load is unimodal is rejected at the 5% level. This is corroborated by the Figure 6-(a), which indicates the asymmetric bimodal behavior of the variable viral load. For the sake of model comparison, we also fitted the CFN and CETN models and the mixture of two normals.

The estimated model parameters using maximum likelihood turned out to be $\hat{\mu} = 1.006(0.137)$, $\hat{\sigma} = 1.079(0.213)$ and $\hat{\delta} = -0.987(0.379)$ with BIC= 348.95 and CAIC= 351.95 for the CFN model and $\hat{\xi} = 1.587(0.160)$, $\hat{\eta} = 1.840(0.213)$, $\hat{\lambda} = 2.261(1.508)$ and $\hat{\beta} = -0.588(0.199)$ with BIC= 349.29 and CAIC= 353.29 for the CETN model.

The QQ-plots obtained by using the estimated parameters are depicted in Figure 6-(b) and (c) providing an illustration of the good performance of the models under study.

Now we compare the CETN model with the $\mathrm{CMN}(\mu_1, \sigma_1, \mu_2, \sigma_2, p)$ model. The estimated model is

$$\mathrm{CMN}(1.675, 0.847, 4.404, 0.748, 0.711)$$

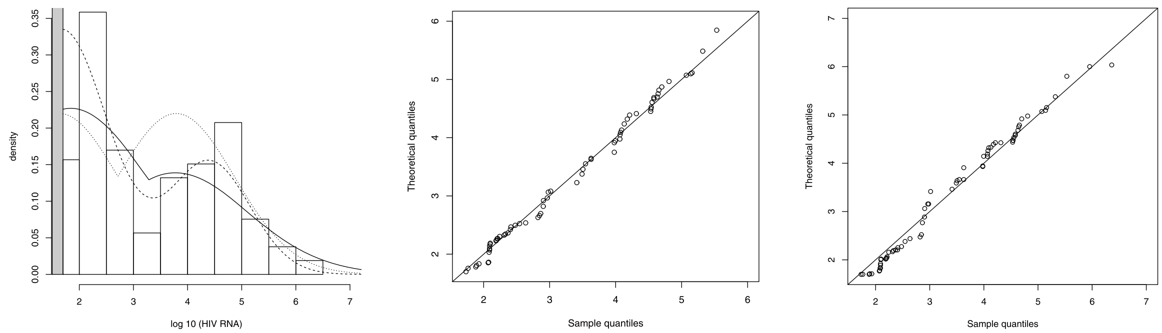
with BIC= 351.08 and CAIC= 356.08. This model has BIC and CAIC greater than the CETN and CFN models, so the CETN and CFN models fit the data better than the CMN model. Figure 6-(a) depicts the estimated CETN, CFN and CMN models. The QQ-plot obtained from the estimated CMN model is given in Figure 7-(b), this shows the fit obtained with the estimated models.

Note that in spite of the large standard error of the estimate of parameter λ (rendering thus the hypothesis $\lambda = 0$ nonsignificant) we prefer to use Hartigan and Hartigan (1985) approach indicating that we should prefer model CETN rather than model CSN. BIC also indicates smaller values for models CFN and CETN.

The total censored data corresponds to 34.90% of the sample under study. Further, the area under the estimated CETN density is 35.33%, under the CFN model is 28.41% and under the estimated density for the CMN model is 36.37%, with the best performance for the CETN model again.

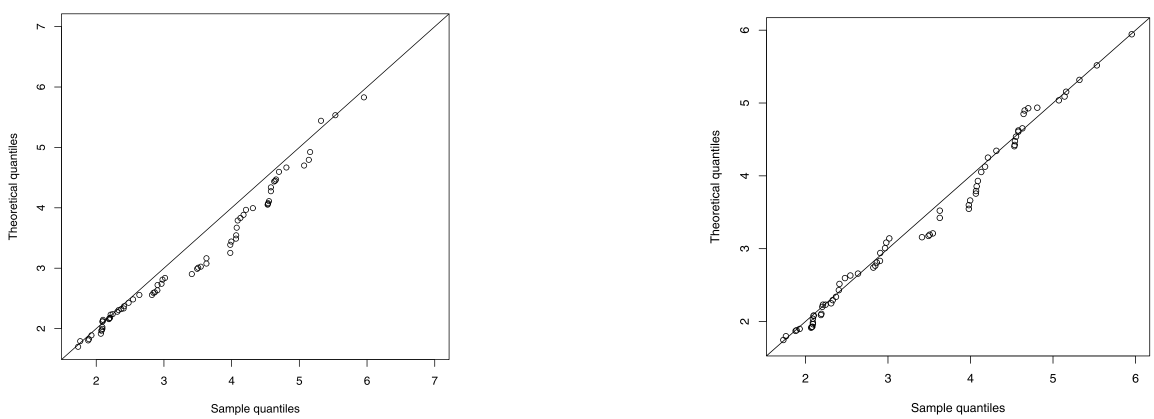
5.2 Illustration II

To illustrate the relevance of model FN in fitting symmetric bimodal data, we deal in this section with the ecographic weight (fetal weight in grams - variable b.weighgt)



(a) (b) (c)

Figure 6. (a) Histogram for $\log_{10}$ HIV-1-RNA for the 106 women noncensored observations. Models: CETN (solid line), CMN (dashed line) and CFN(dotted line), (b) QQ-plot CETN model and (c) QQ-plot CFN model.



(a) (b)

Figure 7. (a) QQ-plot CN model and (b) QQ-plot CMN model.

for a set of 500 observations available for downloading at http://www.mat.uda.cl/hsalinas/cursos/2011/R/weight.rar.

Table 4. Descriptive statistics for variable b.weight

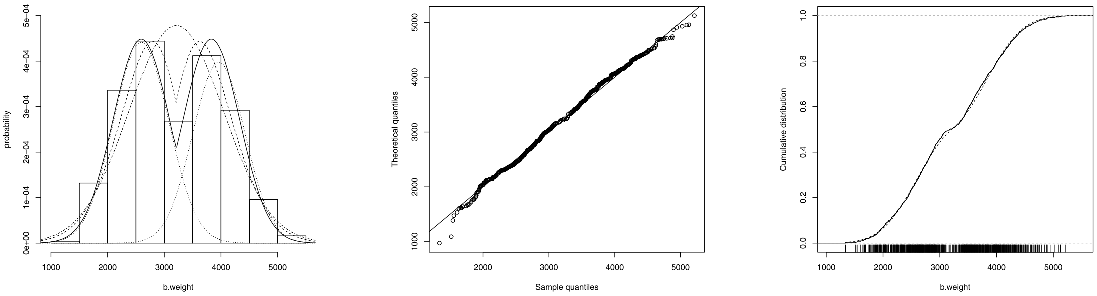
n	$\overline{x}$	S^2	Median
500	3210.356	695710.6	3175

The asymmetry coefficient for the data set is $\sqrt{b_1} = 0.0712$, indicating fair symmetry and, moreover, as the histogram in Figure 7 indicates, the data presents bimodal behavior. Table 5 presents maximum likelihood estimates for the following models: Normal, TN, MN and FN, and the corresponding BIC and CAIC scores.

We also fitted the mixture of two normal distributions, obtaining the following estimates: $\hat{\mu}_1 = 2592.206(74.865)$, $\hat{\mu}_2 = 3963.666(63.544)$, $\hat{\sigma}_1 = 487.577(51.844)$, $\hat{\sigma}_2 = 462.484(35.953)$ and $\hat{p} = 0.539(0.049)$, with BIC$= 8120.967$ and CAIC$= 8125.967$. Results indicate that models TN and FN are better than the ordinary normal model and model FN is better than models TN and MN. Figure 8-(a) illustrates the data histogram with the fitted distributions using the maximum likelihood estimates. Figures 8-(b) and (c)

Table 5. Maximum likelihood estimators for the fetal weight data and the corresponding standard errors (in parenthesis), BIC and CAIC values

Estimates	N	TN	FN
$\hat{\mu}$	3210.356(37.301)	3207.422(26.083)	3212.829(22.294)
$\hat{\sigma}$	834.092(26.415)	772.688(25.760)	498.498(24.879)
$\hat{\delta}$	–	1.771(0.539)	−1.239(0.116)
BIC	8156.713	8122.318	8100.083
CAIC	8158.713	8125.315	8103.083



(a) (b) (c)

Figure 8. (a) Histogram for b.weight. Models: Normal(dashed and dotted line), TN (dashed line), MN(dotted line) and FN(solid line), (b) QQ-plot for FN model, (c) cumulative distribution function FN model, empirical(solid line) and FN(dashed line).

depicts qqplot and the cumulative distribution function for model FN, corroborating the satisfactory fit of the model for variable b-weight.

6. FINAL DISCUSSION

This paper presented a series of models that can be used under censored data situations with possible bimodality. Therefore, models proposed extend the ordinary normal Tobit model, which was firstly designed for unimodal situations. As seen in the applications it is not uncommon in real data sets such anomalies to occur. As shown in this paper, an efficient analysis can be undertaken by combining continuous information with binary data by using asymmetric models. We consider then an alternative route, which is made possible by extending the usual normal and skew-normal (Azzalini, 1985) models to models that are able to incorporate a certain degree of asymmetry and bimodality. Moreover, models considered involve less parameters to be estimated than the ordinary mixtures of normals model. Estimation was discussed by using the maximum likelihood approach which requires numerical implementation given the complexity of the models under study. Fisher and observed information matrices were presented for some of the models. Model fitting is implemented by using the BIC and CAIC scores. A hypothesis testing for unimodality (Hartigan, 1985) is also implemented. Real data applications provide strong support for the new model (CETN), showing that it is a viable alternative to existing models in the literature, including the mixture of two normal models.

ACKNOWLEDGEMENT

The research of G. Martinez-Florez was supported by FCA-18-15 (Colombia). The research of H. W. Gómez was supported by SEMILLERO UA-2017 (Chile). The research of H. Bolfarine was supported by CNPq and Fapesp (Brasil).

APPENDIX 1

In this section we present the observed and expected information matrices for the $CNF(\mu,\sigma,\delta)$ model. Starting with the observed information matrix, namely the negative of the second derivative of the log-likelihood function with respect to the model parameters, for which we use the notation $j_{\mu\mu},\ j_{\mu\sigma},\ j_{\mu\delta}\ j_{\sigma\sigma},\ j_{\sigma\delta}$ and $j_{\delta\delta}$, we arrive at the following expressions, where $\mu^*=\frac{\mu+\sigma\delta}{\sigma}$ and $h(\cdot)=\frac{\phi(\cdot)}{1-\Phi(\cdot)}$

$$
\begin{aligned}
j_{\mu\mu} &= \frac{n_1}{\sigma^2}+\frac{n_0}{\sigma^2}h(\mu^*)\left[-\mu^*+h(\mu^*)\right]\\
j_{\mu\sigma} &= \frac{n_0}{\sigma^2}\left((\mu^*-\delta)\mu^*-1\right)h(\mu^*)-\frac{n_0}{\sigma^2}(\mu^*-\delta)h^2(\mu^*)\\
&\quad+\frac{2}{\sigma^2}\sum_{1}\frac{y_i-\mu}{\sigma}-\frac{\delta}{\sigma^2}\sum_{1}sgn(y_i-\mu),\\
j_{\sigma\sigma} &= \frac{n_0\mu}{\sigma^2}\left(1-(\mu^*-\delta)\mu^*\right)h(\mu^*)+\frac{n_0}{\sigma^2}(\mu^*-\delta)h^2(\mu^*)\\
&\quad-\frac{n_1}{\sigma^2}+\frac{3}{\sigma^2}\sum_{1}\left(\frac{y_i-\mu}{\sigma}\right)^2+\frac{2\delta}{\sigma^2}\sum_{1}\left|\frac{y_i-\mu}{\sigma}\right|,\\
j_{\mu\delta} &= -\frac{n_0}{\sigma}\left(\mu^*-h(\mu^*)\right)h(\mu^*)+\frac{1}{\sigma}\sum_{1}sgn(y_i-\mu),\\
j_{\sigma\delta} &= \frac{n_0\mu}{\sigma}\left(\mu^*-h(\mu^*)\right)h(\mu^*)-\frac{1}{\sigma}\sum_{1}\left|\frac{y_i-\mu}{\sigma}\right|,\\
j_{\delta\delta} &= -n_0\left(\mu^*-h(\mu^*)\right)h(\mu^*)+n\left(\delta-h(\delta)\right)h(\delta)+n_1.
\end{aligned}
$$

The expected (Fisher) information matrix follows then by taking expectations of the above components (multiplied by n^{-1}) (assuming that regularity conditions are satisfied).

The elements of this matrix, which we denote by $i_{\mu\mu},\ i_{\mu\sigma},\ i_{\mu\delta}\ i_{\sigma\sigma},\ i_{\sigma\delta}$ and $i_{\delta\delta}$, can be written as

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

with $\theta_1=\mu$, $\theta_2=\sigma$ and $\theta_3=\delta$. Its elements can be written as

$$
\begin{aligned}
i_{\mu\mu} &= \frac{1}{\sigma^2}\left[1 - c_\delta\left(1 - \Phi\left(\mu^*\right)\right)\right] + \frac{c_\delta}{\sigma^2}\phi\left(\mu^*\right)\left[-\mu^* + h(\mu^*)\right], \\
i_{\mu\sigma} &= \frac{c_\delta}{\sigma^2}\phi\left(\mu^*\right)\left[\left(\mu^* - \delta\right)\left(\mu^* - h(\mu^*)\right) - 1\right] \\
&\quad - \frac{\delta c_\delta}{\sigma^2}\left(1 - \Phi\left(\mu^*\right)\right) \\
&\quad + \frac{2c_\delta}{\sigma^2}\left[\phi\left(\mu^*\right) + \phi(\delta) + \delta\left(\Phi\left(\mu^*\right) + \Phi(\delta) - \frac{3}{2}\right) - \frac{1}{\sqrt{2\pi}}\right], \\
i_{\sigma\sigma} &= \frac{\mu c_\delta}{\sigma^2}\phi\left(\mu^*\right)\left[1 + \left(\mu^* - \delta\right)\left(-\mu^* + h(\mu^*)\right)\right] - \frac{1}{\sigma^2} \\
&\quad + \frac{c_\delta}{\sigma^2}\Big[-2\delta\phi(\delta) + (1 + 2\delta^2)\left(1 - \Phi\left(\mu^*\right)\right) \\
&\quad - 4\delta^2(1 - \Phi(\delta))\Big] \\
&\quad + \frac{c_\delta}{\sigma^2}\Big[\left(3\mu^* - 4\delta\right)\phi\left(\mu^*\right) \\
&\quad + 3(1 + \delta^2)\left(1 - 2\Phi(\delta) + \Phi\left(\mu^*\right)\right)\Big], \\
i_{\mu\delta} &= \frac{c_\delta}{\sigma}\phi\left(\mu^*\right)\left[-\mu^* + h(\mu^*)\right] + \frac{c_\delta}{\sigma}\left(1 - \Phi\left(\mu^*\right)\right), \\
i_{\sigma\delta} &= \frac{c_\delta \mu}{\sigma}\phi\left(\mu^*\right)\left[\mu^* - h(\mu^*)\right] - \frac{\delta c_\delta}{\sigma}\left(1 - \Phi\left(\mu^*\right)\right) \\
&\quad + \frac{c_\delta}{\sigma}\left[2\delta\left(1 - \Phi(\delta)\right) - 2\phi(\delta) + \phi\left(\mu^*\right)\right], \\
i_{\delta\delta} &= c_\delta\phi\left(\mu^*\right)\left[-\mu^* + h(\mu^*)\right] + h(\delta)\left[\delta - h(\delta)\right] \\
&\quad + 1 - c_\delta\left(1 - \Phi\left(\mu^*\right)\right).
\end{aligned}
$$

APPENDIX 2

In this appendix we present the observed and expected information matrices for the $CTN(\xi, \eta, \lambda)$ model. Starting with the observed information matrix. Considering: $\Phi_{CTN}(\cdot) = \Phi(\cdot) - 0.5\Phi_{SN}(\cdot)$ and $\phi_{CTN}(\cdot) = \phi(\cdot) - 0.5\phi_{SN}(\cdot)$ we arrive at the following expressions:

$$
\begin{aligned}
j_{\xi\xi} &= \frac{n_0}{\eta^2}\left\{\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\left[-\frac{\xi}{\eta} + \frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right]\right. \\
&\quad \left. + \frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{\lambda\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right\} \\
&\quad + \frac{1}{\eta^2}\sum_1\left\{1 + \lambda^2\left[\lambda|z_i|\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)} + \left(\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\right)^2\right]\right\}, \\
j_{\eta\xi} &= \frac{n_0}{\eta^2}\left\{\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\left[-1 + \left(\frac{\xi}{\eta}\right)^2 - \frac{\xi}{\eta}\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right]\right. \\
&\quad \left. + \frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{\xi}{\eta}\frac{\lambda\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right\} \\
&\quad + \frac{1}{\eta^2}\sum_1\left[2z_i + \lambda sgn(y_i - \mu)\left\{(1 - \lambda^2 z_i^2)\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\right.\right. \\
&\quad \left.\left. - \lambda|z_i|\left(\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\right)^2\right\}\right], \\
j_{\lambda\xi} &= \frac{n_0}{2\eta}\sqrt{\frac{2}{\pi}}\frac{\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\left[\frac{\xi}{\eta} - \frac{1}{1+\lambda^2}\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right] \\
&\quad + \frac{1}{\eta}\sum_1\left[-sgn(y_i - \mu)\left(\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\right)\left(\lambda^2 z_i^2 - 1\right)\right. \\
&\quad \left. + \lambda z_i\left(\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\right)^2\right] \\
j_{\eta\eta} &= \frac{n_0\xi}{\eta^2}\left\{\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\left[\frac{1}{\eta}\left(2 - \left(\frac{\xi}{\eta}\right)^2\right)\right.\right. \\
&\quad \left.\left. + \left(\frac{\xi}{\eta}\right)^2\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right] - \frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{\lambda\xi}{\eta^2}\frac{\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right\} \\
&\quad + \frac{1}{\eta^2}\sum_1\left[-1 + 3z_i^2 + \lambda z_i\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\left(\lambda^2 z_i - 2\right)\right. \\
&\quad \left. + \lambda^2|z_i|\left(\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\right)^2\right], \\
j_{\lambda\eta} &= \frac{n_0\xi}{2\eta^2}\sqrt{\frac{2}{\pi}}\frac{\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\left[-\frac{\xi}{\eta} + \frac{1}{1+\lambda^2}\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right] \\
&\quad + \frac{1}{\eta}\sum_1\left[z_i\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\left(1 - \lambda z_i|z_i|\right) + \lambda z_i^2\left(\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\right)^2\right], \\
j_{\lambda\lambda} &= \frac{n_0}{2}\sqrt{\frac{2}{\pi}}\frac{1}{1+\lambda^2}\frac{\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)} \\
&\quad \times\left[\frac{2\lambda}{1+\lambda^2} + \lambda\left(\frac{\xi}{\eta}\right)^2 + \frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{1}{1+\lambda^2}\frac{\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right] \\
&\quad + \frac{nk_\lambda}{\pi^2(1+\lambda^2)^2}\left(k_\lambda + 2\pi\lambda\right) \\
&\quad + \sum_1\left[z_i^2\frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\left(\lambda|z_i| + \frac{\phi(\lambda z_i)}{\Phi(\lambda|z_i|)}\right)\right].
\end{aligned}
$$

Expected information matrix

Taking expected values of the negative of the second derivative of the log-likelihood function, we arrive at the following entries

$$
I_{\theta_r\theta_p} = \mathbb{E}\left\{-\frac{\partial^2\ell(\boldsymbol{\theta};\mathbf{y})}{\partial\theta_r\partial\theta_p}\right\}, \quad r, p = 1, 2, 3,
$$

with $\theta_1 = \xi$, $\theta_2 = \eta$ and $\theta_3 = \lambda$. Considering: $b_j = \mathbb{E}\{z^j\left(\phi(\lambda z)/\Phi(\lambda|z|)\right])^2\}$, $b_j^* = \mathbb{E}\{|z|^j\left(\phi(\lambda z)/\Phi(\lambda|z|)\right])^2\}$ and $\gamma = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

$$
\begin{aligned}
i_{\xi\xi} &= \frac{k_\lambda}{\eta^2}\left\{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)\left[-\frac{\xi}{\eta}+\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right]\right.\\
&\quad\left.+\frac{1}{2}\sqrt{\frac{2}{\pi}}\lambda\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\right\}\\
&\quad+\frac{1}{\sigma^2}\left\{\left(1-k_\lambda\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)\right)\right.\\
&\quad\left.+\frac{\lambda\gamma k_\lambda}{\pi}\left[\sqrt{\frac{2}{\pi}}\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)-1\right]+\lambda^2 b_0\right\},
\end{aligned}
$$

$$
\begin{aligned}
i_{\eta\xi} &= \frac{k_\lambda}{\eta^2}\left\{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)\left[-1+\left(\frac{\xi}{\eta}\right)^2-\frac{\xi}{\eta}\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right]\right.\\
&\quad\left.+\frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{\xi}{\eta}\lambda\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\right\}\\
&\quad-\sqrt{\frac{2}{\pi}}\frac{k_\lambda}{\eta^2}\left[(1-\gamma)+\frac{3}{2}\gamma\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)+\sqrt{\frac{2}{\pi}}\phi_{SN}\left(\frac{\xi}{\eta}\right)\right]\\
&\quad-\sqrt{\frac{2}{\pi}}\frac{\lambda\gamma^2 k_\lambda}{2\eta^2}\left[\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)-\sqrt{\frac{2}{\pi}}\right]-\frac{\lambda^2}{\eta^2}b_1
\end{aligned}
$$

$$
\begin{aligned}
i_{\lambda\xi} &= \frac{k_\lambda}{2\eta}\sqrt{\frac{2}{\pi}}\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\\
&\quad\times\left[\frac{\xi}{\eta}(1+\gamma^2)+\frac{1}{1+\lambda^2}\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right]\\
&\quad+\frac{k_\lambda}{2\eta}\sqrt{\frac{2}{\pi}}\frac{1}{(1+\lambda^2)^{3/2}}\left[\lambda^2+\Phi\left(\sqrt{1+\lambda^2}\frac{\mu}{\sigma}\right)\right]+\frac{\lambda}{\sigma}b_1
\end{aligned}
$$

$$
\begin{aligned}
i_{\eta\eta} &= \frac{k_\lambda\xi}{\eta^2}\left\{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)\left[\frac{1}{\eta}\left(2-\left(\frac{\xi}{\eta}\right)^2\right.\right.\right.\\
&\quad\left.\left.+\left(\frac{\xi}{\eta}\right)^2\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right]-\frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{\lambda\xi}{\eta^2}\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\right\}\\
&\quad-\frac{1}{\eta^2}+\frac{k_\lambda}{\eta^2}\left[\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)+\frac{3}{\pi}\left(\arctan\lambda+\frac{\gamma}{\sqrt{1+\lambda^2}}\right)\right.\\
&\quad\left.-\frac{3}{2}\left(\frac{\xi}{\eta}\phi_{SN}\left(\frac{\xi}{\eta}\right)-\Phi_{SN}\left(\frac{\xi}{\eta}\right)\right)\right]\\
&\quad+\frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{k_\lambda}{\eta^2}\left[-\gamma\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\left(\frac{3}{\sqrt{1+\lambda^2}}+\lambda\gamma\frac{\xi}{\eta}\right)\right.\\
&\quad\left.+\gamma^3\Phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\right]+\frac{\lambda^2}{\eta^2}b_1^*
\end{aligned}
$$

$$
\begin{aligned}
i_{\lambda\eta} &= \frac{\xi k_\lambda}{2\eta^2}\sqrt{\frac{2}{\pi}}\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\left[-\frac{\xi}{\eta}+\frac{1}{1+\lambda^2}\frac{\phi_{CTN}\left(-\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right]\\
&\quad+\frac{\lambda}{\eta}b_2+\sqrt{\frac{2}{\pi}}\frac{k_\lambda}{\eta(1+\lambda^2)}\left[\frac{1}{2}\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\right.\\
&\quad\left.\times\left(1+\lambda\left(\frac{\xi}{\eta}\right)^2+\frac{2\gamma}{\sqrt{1+\lambda^2}}\right)-\sqrt{\frac{2}{\pi}}\frac{\gamma}{\sqrt{1+\lambda^2}}\right],
\end{aligned}
$$

$$
\begin{aligned}
i_{\lambda\lambda} &= \frac{k_\lambda}{2}\sqrt{\frac{2}{\pi}}\frac{1}{1+\lambda^2}\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\\
&\quad\times\left[\frac{2\lambda}{1+\lambda^2}+\lambda\left(\frac{\xi}{\eta}\right)^2+\frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{1}{1+\lambda^2}\frac{\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)}{\Phi_{CTN}\left(-\frac{\xi}{\eta}\right)}\right]\\
&\quad+\frac{k_\lambda}{\pi^2(1+\lambda^2)^2}(k_\lambda+2\pi\lambda)+\frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{\gamma k_\lambda}{(1+\lambda^2)^{3/2}}\\
&\quad\times\left[2\sqrt{\frac{2}{\pi}}-\phi\left(\sqrt{1+\lambda^2}\frac{\xi}{\eta}\right)\left((1+\lambda^2)\left(\frac{\xi}{\eta}\right)^2+2\right)\right]+b_2.
\end{aligned}
$$

APPENDIX 3

Results of a simulation study to verify parameter recovery for the parameter δ with the flexible censored normal model of Section 3 are reported in Table 6. It can be depicted that

Table 6. Performance evaluation for MLEs of parameter δ under the CNF model

		$p=5\%$		$p=10\%$		$p=20\%$		$p=40\%$	
δ	n	RB	$\sqrt{MSE}$	RB	$\sqrt{MSE}$	RB	$\sqrt{MSE}$	RB	$\sqrt{MSE}$
	50	-0.022	0.157	-0.035	0.162	-0.055	0.174	-0.054	0.178
-5	150	-0.019	0.090	-0.035	0.103	-0.054	0.100	-0.058	0.121
	1000	-0.015	0.041	-0.033	0.041	-0.054	0.036	-0.060	0.057
	50	-0.062	0.175	-0.101	0.184	-0.156	0.190	-0.119	0.193
-2	150	-0.055	0.100	-0.100	0.110	-0.151	0.113	-0.132	0.126
	1000	-0.050	0.045	-0.098	0.046	-0.149	0.041	-0.149	0.057
	50	0.152	0.483	0.209	0.529	0.246	0.566	0.210	0.598
2	150	0.122	0.286	0.129	0.298	0.159	0.344	0.164	0.373
	1000	0.047	0.109	0.052	0.119	0.070	0.147	0.085	0.156
	50	0.094	0.757	0.119	0.740	0.131	0.797	0.111	0.962
5	150	0.066	0.417	0.070	0.412	0.065	0.388	0.078	0.530
	1000	0.017	0.119	0.013	0.093	0.014	0.107	0.022	0.148

the maximum likelihood estimation approach implemented performs well for moderate and large sample sizes.

Received 25 October 2016

REFERENCES

Arnold, B. C., Gómez, H. W., Salinas, H. S. (2009) On multiple constraint skewed models. Statistics, 43:279–293. MR2535732

Azzalini, A. (1985) A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 12:171–178. MR0808153

Azzalini, A. (1986) Further results on a class of distributions which includes the normal ones. Statistica, 46:199–208. MR0877720

Chen, T., Ma, S., Kobie, J., Rosenberg, A., Sanz, I., Liang, H. (2016) Identification of significant B cell associations with undetected observations using a Tobit model. Statistics and Its Interface, 9(1):79–91. MR3423313

Gómez, H. W., Elal-Olivero, D., Salinas, H. S., Bolfarine, H. (2011) Bimodal extension based on the skew-normal distribution with application to pollen data. Environmetrics, 22:50–62. MR2843336

Hartigan, J. A. and Hartigan, P. M. (1985) The dip test of unimodality, Ann. Stat., 13:70–84. MR0773153

Hartigan, P. M. (1985) Computation of the dip statistics to test for unimodality, Appl. Stat., 34:320–325.

Hastie, T. J., Tibshirani, R. J. (1990) Generalized additive models. London: Chapman & Hall. MR1082147

Kim, H. J. (2005) On a class of two-piece skew-normal distributions. Statistics, 39:537–553. MR2211732

Li, X., Chu, H., Gallant, J. E., Hoover, D. R., Mack, W. J., Chmiel, J. S., Muñoz, A. (2006) Bimodal virologic response to antiretroviral therapy for HIV infection: an application using a mixture model with left censoring. J. Epidemiol Commun Health, 60:811–818.

Marin, J. M., Mengersen, K., Robert, C. P. (2005) Bayesian modelling and inference on mixtures of distributions. In Rao, C. and Dey, D. (eds.), Handbook of Statistics, volume 25. Springer-Verlag, New York. MR2490536

Martínez-Flórez, G., Bolfarine, H., Gómez, H. W. (2013a) The alpha-power Tobit model. Communications in Statistics, Theory and Methods, 42:633–643. MR3211940

Martínez-Flórez, G., Bolfarine, H., Gómez, H. W. (2013b) Asymmetric regression models with limited responses with an application to antibody response to vaccine. Biometrical Journal, 55:156–172. MR3045838

Martínez-Flórez, G., Bolfarine, H., Gómez, H. W. (2015) Doubly censored power-normal regression models with inflation. TEST, 24:265–286. MR3348211

Moulton, L., Halsey, N. (1995) A Mixture Model with Detection Limits for Regression Analysis of Antibody Response to Vaccine. Biometrics, 51:1570–1578.

Owen, D. B. (1956) Tables for computing bi-variate normal probabilities. Annals Mathematical Statistics, 27:1075–1090. MR0127562

Pewsey, A. (2000) Problems of inference for Azzalini's skew-normal distribution. Journal of Applied Statistics, 27:859–870.

Ramires, T. G., Ortega, E. M. M., Cordeiro, G. M., Hens, N. (2016) A bimodal flexible distribution for lifetime data. Journal of Statistical Computation and Simulation, 86(12):2450–2470. MR3502174

R (2012). A language and environment for statistical computing. R Foundation for Statistical Computing, ISBN 3-900051-07-0, Vienna, Austria. http://www.Rproject.org.

Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000) Likelihood-based inference with singular information matrix, Bernoulli, 6:243–284. MR1748721

Tobin, J. (1958) Estimation of relationships for limited dependent variables. Econometrica, 26:24–36. MR0090462

Guillermo Martínez-Flórez
Departamento de Matemáticas y Estadística
Facultad de Ciencias Básicas
Universidad de Córdoba
Córdoba
Colombia
E-mail address: gmartinez@correo.unicordoba.edu.co

Heleno Bolfarine
Departamento de Estatítica
IME
Universidade de São Paulo
São Paulo
Brazil
E-mail address: hbolfar@ime.usp.br

Héctor W. Gómez
Departamento de Matemáticas
Facultad de Ciencias Básicas
Universidad de Antofagasta
Antofagasta
Chile
E-mail address: hector.gomez@uantof.cl