Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 285 a 301

Censored Bimodal Symmetric-Asymmetric Alpha-Power Model

Modelo bimodal censurado simétrico-asimétrico alpha-potencia

Hugo S. Salinas[1,a], Guillermo Martínez-Flórez[2,b], Germán Moreno-Arenas[3,c]

[1]Departamento de Matemáticas, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile

[2]Departamento de Matemáticas y Estadística, Facultad de Ciencias, Universidad de Córdoba, Córdoba, Colombia

[3]Escuela de Matemáticas, Universidad Industrial de Santander, Bucaramanga, Colombia

Abstract

We introduce the censored bimodal symmetric-asymmetric alpha-power models to adjust censored data with bimodality and high levels of skewness and kurtosis. The moments corresponding are computed, the maximum likelihood estimation for the model parameters is considered and the observed information matrix is derived. We show the appropriateness of the proposed models through two applications with censored real data related to HIV-1 RNA measurement.

***Key words*:** AART, alpha-power model, bimodality, censorship, cumulative distribution, HIV-1 RNA, limit of detection, power-normal model.

Resumen

Se introducen los modelos potencia alfa simétricos asimétricos bimodales censurados con el fin de ajustar datos censurados con bimodalidad y altos niveles de sesgo y curtosis. Los momentos correspondientes son calculados, se considera la estimación máximo verosímil para los parámetros del modelo y se deriva la matriz de información observada. Se muestra la utilidad de los modelos propuestos a través de dos aplicaciones con datos censurados reales relacionados con la medición de HIV-1 RNA.

***Palabras clave*:** AART, bimodalidad, censura, distribución acumulada, HIV-1 RNA, límite de detección, modelo alfa potencia, modelo normal potencia.

[a]Associate professor. E-mail: hugo.salinas@uda.cl
[b]Professor. E-mail: gmartinez@correo.unicordoba.edu.co
[c]Professor. E-mail: gmorenoa@uis.edu.co

1. Introduction

In epidemiological studies where biomarkers are the main outcomes, it is common to have detection limits below which it is not possible to determine the specific values. For instance, in highly active antiretroviral therapy (HAART), the number of viral load measurements in patients with HIV has a lower detection limit when using ultrasensitive tests.

The quantitative measurements in people with HIV may be highly left censored with a high percentage below the detection limit, depending on the method used for each measurement. For example, in Bucaramanga City, Colombia, the viral load measurements are conducted by different laboratories, and the HIV-1 RNA quantification is performed by three different methods: Versant bDNA 3.0® (Bayer), LCx HIV® (Abbott) and Amplicor HIV Monitor v1.5® (Roche), all of which have a detection limit of 50 copies per mL. In order to model the percentage of individuals below the detection limit, an asymmetric bimodal model may be necessary for this type of variable.

The analysis of viral load, HIV-RNA, (scale $\log_{10}$) is used to measure the effectiveness of HAART therapy which suppresses HIV-1 RNA to undetectable levels, thereby reducing the morbidity and mortality rates of HIV. Li, Chu, Gallant, Hoover, Mack, Chmiel & Muñoz (2006) found that $\log_{10}$(HIV-1 RNA) has two modal values in its distribution, corresponding to the optimal and suboptimal response to HAART, and it can be modeled with a mixture of two normal distributions in the presence of left censoring. In other cases, the bimodal behavior is also considered as the variable has a high (or low) degree of asymmetry and kurtosis in at least some partial distributions that compose the bimodal behavior.

In general a random variable Y, which has a part of its probability at discrete points and the rest spread over several intervals, has a mixture distribution.

When data are censored, the observed variable Y is a mixture of a continuous latent process Y^* and a selection mechanism (censoring or truncation) modeling in binary form. This idea was popularized by Tobin (1958) and the resulting model is known as the Tobit model, which is defined in terms of the latent variable $Y_i = Y_i^* I(Y_i^* > c)$, for some constant c, where $I(\cdot)$ is the indicator function and Y^* has a certain distribution, e.g., normal Tobin (1958) or Student-t of Arellano-Valle, Castro, González-Farías & Muñoz-Gajardo (2012) or generalized normal of Martínez-Flórez, Bolfarine & Gómez (2013).

Until the last two decades of the twentieth century, the inferential processes assumed the normality of the data under study. This assumption is unrealistic for many variables, and the inferential processes are inadequate. In these situations many authors choose to transform the variables in order to attain data symmetry or normality. This transformation leads to unsatisfactory results because the interpretation of results becomes cumbersome. The data becomes more difficult to interpret when there are several variables with different types of transformations. In view of these deficiencies, more flexible models have been developed that are able to accommodate different degrees of asymmetry and kurtosis. Previous work in this area include Azzalini (1985), Henze (1986), Durrans (1992), Fernández &

Steel (1998), Mudholkar & Hutson (2000), Pewsey (2000), Eugene, Lee & Famoye (2002), Jones (2004), Gómez, Venegas & Bolfarine (2007) and Arnold, Gómez & Salinas (2009).

For bimodal data, extensions for asymmetric cases have been studied by Kim (2005), Gómez et al. (2007) and Arnold et al. (2009), among others. Kim (2005) introduces the bimodal skew-normal called the *two-pieces skew-normal model*. An asymmetric extension of this model was presented by Arnold et al. (2009) who defined *the extended two-pieces skew-normal model*. Gómez, Elal-Olivero, Salinas & Bolfarine (2009) also studied a bimodal skew-normal model for certain values of the shape parameter, and this distribution is called *skew-flexible-normal*. Other works in this area have been published by Elal-Olivero, Gómez & Quintana (2009) and Bolfarine, Gómez & Rivas (2011).

In this paper, we present a new distribution for adjusted censored data with bimodality and high levels of skewness and kurtosis. The paper is structured as follows. In Section 2, we introduce the censored bimodal symmetric alpha-power distribution, moments, estimation and inference for model parameters. In Section 4, we introduce the censored bimodal asymmetric alpha-power distribution, moments, estimation and inference for model parameters; we derive the information matrix and discuss likelihood ratio tests for some hypotheses of interest. In Section 6, the appropriateness of this model is illustrated using two applications involving real data. Finally, some concluding remarks are presented in Section 7.

2. Censored Bimodal Symmetric Alpha-Power Model

Based on the works by Durrans (1992) and Kim (2005), Bolfarine, Martínez-Flórez & Salinas (2012) introduced the bimodal symmetric alpha-power model, whose probability density is

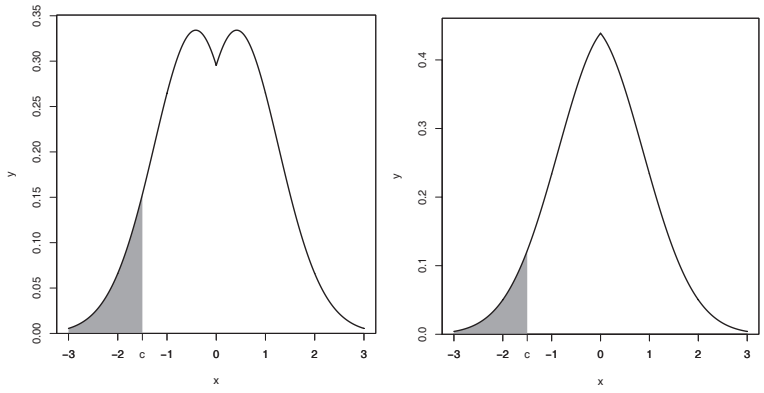
$$\varphi(z;\alpha) = \alpha c_\alpha f(z) \left\{F(|z|)\right\}^{\alpha-1}, \quad -\infty < z < \infty \tag{1}$$

where $\alpha \in \mathbb{R}^+$, F is an absolutely continuous density function with density function $f = dF$ symmetric around zero and $c_\alpha = \frac{2^{\alpha-1}}{2^\alpha - 1}$ is the normalizing constant. We use the notation $Z \sim BSP(\alpha)$.

Now, consider a random variable $Y^* \sim BSP(\alpha)$ where $(Y_1^*, Y_2^*, \ldots, Y_n^*)$ is a random sample of size n and point of censorship equal to c. Values of Y^* greater than the constant c are mapped to themselves, whereas values of Y^* less than or equal to the constant c are mapped to c. Then, without loss of generality for $c = 0$, the observed value is $Y_i = D_i Y_i^*$, $i = 1, 2, \ldots, n$, where $D_i = I(Y_i^* > 0)$. Here we have a random sample that is left censored. We say that Y follows a censored BSP distribution. We denote this variable by $Y \sim CBSP(\alpha)$. The generalization to right censoring or when the point of censorship is different from zero is trivial.

For a random variable $Y \sim CBSP(\alpha)$ with $\alpha \in \mathbb{R}^+$, the location-scale extension is defined as the distribution of the random variable $X = \xi + \eta Y$ for $\xi \in \mathbb{R}$ and $\eta > 0$. We use the notation $X \sim CBSP(\xi, \eta, \alpha)$.

From equation (1), when $f = \phi$ and $F = \Phi$ are the standard normal density and cumulative distribution functions, respectively, we obtain the bimodal power-normal density function and use the notation $Z \sim BPN(\alpha)$. Similarly, we obtain the censored bimodal power-normal density function $Y \sim CBPN(\alpha)$ and the location-scale extension $X \sim CBPN(\xi, \eta, \alpha)$. The density function of the random variable $Y \sim CBPN(\alpha)$ is symmetric and unimodal for $0 < \alpha \leq 1$ and bimodal for $\alpha > 1$. Figure 1 depicts plots for the random variable $Y \sim CBPN$ with a point of censorship $c \neq 0$ and two values of α.



(a) (b)

FIGURE 1: Densities of $CBPN(0, 1, \alpha)$ censored at the left (grey color): (a) $\alpha = 1.75$ and (b) $\alpha = 0.75$.

The moments of the random variable CBSP are given as functions of the incomplete moments of the alpha-power model which are defined as

$$\mu_r(x) = \int_x^\infty \alpha z^r f(z) \{F(z)\}^{\alpha-1} dz, \quad r = 0, 1, 2, \ldots,$$

The r-th moment of the random variable $X \sim CBSP(\xi, \eta, \alpha)$ is then given by

$$\mathbb{E}(X^r) = c_\alpha \sum_{k=0}^{r} \binom{r}{k} \xi^{r-k} \eta^k \mu_k(0)$$

3. Inference to CBSP Model

The contribution of the censored and uncensored observations to the log-likelihood function is as follows: if $Y_i = 0$, then $\mathbb{P}(Y_i = 0) = \mathbb{P}(X_i \leq 0) = c_\alpha \left[1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha\right]$, and for the non-nulls Y_i's we have that they are distributed as the respective X_i's.

Assume that n independent and identically distributed observations $x_1, x_2, \ldots, x_n$ are available from $BSP(\xi, \eta, \alpha)$. We denote by $\sum_0$ the sum over the censored

observations and by $\sum_1$ the sum over uncensored observations. The log-likelihood function of (ξ, η, α) based on $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is given by

$$\ell(\xi, \eta, \alpha; \boldsymbol{x}) = \sum_0 \left(\log(c_\alpha) + \log\left[1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha\right]\right) + \sum_1 [\log(\alpha) + \log(c_\alpha) - \log(\eta) + \log(f(z_i)) + (\alpha - 1)\log(F(|z_i|))]$$

where $z_i = \frac{x_i - \xi}{\eta}$. Hence, assuming that f' exists, the score function defined as the first derivative of the log-likelihood function, with respect to all parameters is given by:

$$U(\xi) = -\frac{\alpha}{\eta} \sum_0 \frac{\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1} f\left(\frac{\xi}{\eta}\right)}{1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha} - \frac{1}{\eta} \sum_1 \left\{ \frac{f'(z_i)}{f(z_i)} + (\alpha - 1)\text{sgn}(z_i) \frac{f(|z_i|)}{F(|z_i|)} \right\}$$

$$U(\eta) = \frac{\alpha\xi}{\eta^2} \sum_0 \frac{\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1} f\left(\frac{\xi}{\eta}\right)}{1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha} - \frac{1}{\eta} \sum_1 \left\{ 1 + z_i \frac{f'(z_i)}{f(z_i)} + (\alpha - 1)|z_i| \frac{f(|z_i|)}{F(|z_i|)} \right\}$$

and

$$U(\alpha) = \sum_0 \left\{ -\frac{\log(2)}{2^\alpha - 1} - \frac{\left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha \log\left[F\left(\frac{\xi}{\eta}\right)\right]}{1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha} \right\} + \sum_1 \left\{ \frac{1}{\alpha} - \frac{\log 2}{2^\alpha - 1} + \log[F(|z_i|)] \right\}$$

The score equations are obtained by equating the derivatives presented above to zero. The maximum likelihood estimators are the solutions of the score equations, and clearly depend on the functions f and F. Model parameters are estimated using iterative algorithms that can be implemented in any statistical package. The elements of the observed information matrix are given in Appendix.

4. Censored Bimodal Asymmetric Alpha-Power Model

The CBPN model is an alternative when data are censored and have a bimodal and symmetrical distribution; however, in case that the asymmetrical distributions are not adequate, we introduce another model for censored data whose distribution function is bimodal and asymmetric. The following lemma given by Azzalini (1985) will be essential to achieve this model.

Lemma 1. *Let f_0 be a probability density function symmetric about zero and G be a distribution function such that G' exists and is a probability density function symmetric about zero. Then $f_Z(z;\beta) = 2f_0(z)G(\beta z)$ is a probability density function for $z, \beta \in \mathbb{R}$.*

Based on this lemma and given that the density function of a random variable $BSP(\alpha)$ is symmetric about zero, then for G, which is a distribution function such that G' is a probability density function symmetric about zero, then

$$\varphi(z;\alpha,\beta) = 2\alpha c_\alpha f(z)\left\{F(|z|)\right\}^{\alpha-1} G(\beta z), \quad -\infty < z < \infty \tag{2}$$

is a probability density function, such that $\alpha \in \mathbb{R}^+$ and $\beta \in \mathbb{R}$. The parameter β controls asymmetric behavior. We denote by $Z \sim BAsP(\alpha,\beta)$.

The location-scale extension for a random variable $Z \sim BAsP(\alpha,\beta)$ is defined as the distribution of the random variable $X = \xi + \eta Z$, where $\xi \in \mathbb{R}$ is the location parameter and $\eta > 0$ for the scale parameter. We denote by $X \sim BAsP(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\xi, \eta, \alpha, \beta)$. Thus, redefining the random variable latent $Y_i = X_i I(X_i > 0)$ we obtain a censored random variable, which we denote by $Y \sim CBAsP(\boldsymbol{\theta})$.

When $F = G = \Phi$ in equation (2) naturally follows the *censored bimodal asymmetric alpha-power normal model*, which we denote by $CBAsN(\boldsymbol{\theta})$, this distribution is bimodal for $\alpha > 1$ and unimodal for $0 < \alpha \leq 1$, while the parameter β controls asymmetric behavior.

Figure 2 depicts plots for the random variable $Y \sim CBAsN(\boldsymbol{\theta})$ with point of censorship $c \neq 0$ and two values of β.

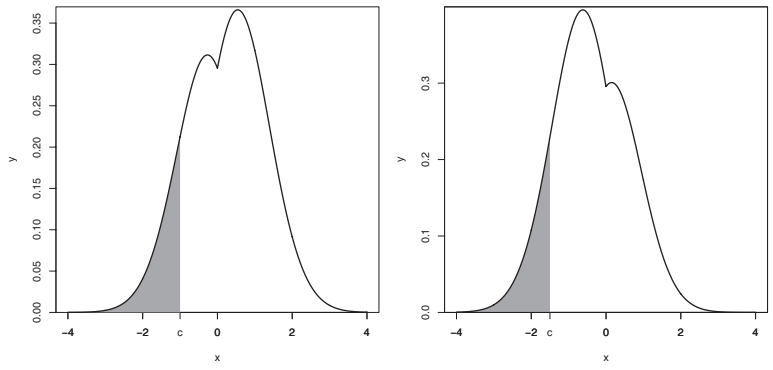


FIGURE 2: Density of $CBAsN(0, 1, 1.75, \beta)$ censored at the left (grey color). (a) $\beta = 0.25$ and (b) $\beta = -0.45$.

5. Inference to CBAsN Model

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n obtained from the $CBAsN(\boldsymbol{\theta})$ distribution with unknown parameter vector $\boldsymbol{\theta}$. The contribution of the i-th

observation to the likelihood is given by $\mathbb{P}(Y = 0) = \mathbb{P}(X \leq 0) = \alpha c_\alpha \, A_c(\boldsymbol{\theta})$ where $A_c(\boldsymbol{\theta}) = \int_{\frac{\xi}{\eta}}^{\infty} \phi(z)\{\Phi(z)\}^{\alpha-1}\{1-\Phi(\beta z)\}dz$.

The log-likelihood function of $\boldsymbol{\theta}$ based on $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ is given by

$$\begin{aligned}\ell(\boldsymbol{\theta};\boldsymbol{y}) &= \sum_1 [\log(2\alpha c_\alpha) - \log(\eta) + \log(\phi(z_i)) + (\alpha-1)\log(\Phi(|z_i|)) + \log(\Phi(\beta z_i))] \\ &+ \sum_0 [\log(\alpha c_\alpha) + \log A_c(\boldsymbol{\theta})]\end{aligned}$$

where $z_i = \frac{y_i - \xi}{\eta}$. The first derivatives of the log-likelihood function with respect to the parameters are given by:

$$\begin{aligned}U(\xi) &= -\frac{n_0 r_c(\boldsymbol{\theta})}{\eta A_c(\boldsymbol{\theta})} - \frac{1}{\eta}\sum_1 \left\{-z_i + (\alpha-1)\mathrm{sgn}(z_i)\frac{\phi(|z_i|)}{\Phi(|z_i|)} + \beta\frac{\phi(\beta z_i)}{\Phi(\beta z_i)}\right\} \\ U(\eta) &= \frac{n_0 r_c(\boldsymbol{\theta})\xi}{\eta^2 A_c(\boldsymbol{\theta})} - \frac{1}{\eta}\sum_1 \left\{1 - z_i^2 + (\alpha-1)|z_i|\frac{\phi(|z_i|)}{\Phi(|z_i|)} + \beta z_i\frac{\phi(\beta z_i)}{\Phi(\beta z_i)}\right\} \\ U(\alpha) &= n\left\{\frac{1}{\alpha} - \frac{\log 2}{2^\alpha - 1}\right\} + \frac{n_0 B_c(\boldsymbol{\theta})}{A_c(\boldsymbol{\theta})} + \sum_1 \{\log[\Phi(|z_i|)]\}\end{aligned}$$

and

$$U(\beta) = \frac{n_0 D_c(\boldsymbol{\theta})}{A_c(\boldsymbol{\theta})} + \sum_1 z_i \frac{\phi(\beta z_i)}{\Phi(\beta z_i)}$$

where

$$\begin{aligned}B_c(\boldsymbol{\theta}) &= \int_{\frac{\xi}{\eta}}^{\infty} \phi(z)\{\Phi(z)\}^{\alpha-1}\log(\Phi(z))\{1-\Phi(\beta z)\}dz, \\ r_c(\boldsymbol{\theta}) &= \phi\left(\frac{\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1}\left\{1-\Phi\left(\frac{\beta\xi}{\eta}\right)\right\}, \\ D_c(\boldsymbol{\theta}) &= \int_{\frac{\xi}{\eta}}^{\infty} z\phi(z)\{\Phi(z)\}^{\alpha-1}\{1-\Phi(\beta z)\}dz\end{aligned}$$

The maximum likelihood estimate $\widehat{\boldsymbol{\theta}} = (\hat{\xi}, \hat{\eta}, \hat{\alpha}, \hat{\beta})$ of $\boldsymbol{\theta}$ is obtained by setting $U(\xi) = U(\eta) = U(\alpha) = U(\beta) = 0$ and solving these equations numerically using iterative techniques. The elements of the observed information matrix are given in Appendix.

6. Illustrations

In this section we illustrate the usefulness of the proposed models by fitting a CBAsP distribution to some data sets. We use two real data sets to compare the fit of this model with censored normal (CN), censored skew-normal (CSN) and censored bimodal skew-normal (CBSN) distributions and with the parent distribution itself.

6.1. HIV-1 RNA Data Obtained from the Secretariat of Health of Bucaramanga City

The database was provided by Secretariat of Health, Department of Santander, Colombia, and corresponds to persons who are reported to the SIVIGILA system. This database maintains the absolute confidentiality of patient identification and contains the age, gender, date of admission to the SIVIGILA system, presence or absence of HAART treatment, CD-4 count and HIV-1 RNA plasma levels (viral load) of some patients. The database corresponds to 1275 persons infected with HIV, and who have been officially reported to the Surveillance and Epidemiology Service of Bucaramanga City. Tests used for the diagnosis of 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 performed. The combination of these two methods is highly accurate. Patients are at different stages of treatment, 681 patients in the sample have had HAART therapy since 2007 and HIV-1 RNA plasma level (viral load) measurement, and there were 206 women and 475 men.

Because the measurements were performed at 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 of which have a detection limit of 50 copies per mL. Descriptive statistics for $\log_{10}$(HIV-1 RNA) observations above the detection limit of 475 men in the example are mean=1.7350 and variance=1.7397. The skewness=0.5258 and kurtosis=2.1346 correspond to sample values above $\log_{10}(50)$. These statistics show that the data have a high positive bias and a low kurtosis compared to the normal model, which is an indication that the censored normal model is not an alternative to adjusting for viral loads. In addition to these characteristics, the histogram of Figure 3 shows that the behavior of the $\log_{10}$(HIV-1 RNA) variable is bimodal, and therefore the censored bimodal skew-normal model can be used to adjust $\log_{10}$(HIV-1 RNA) data. Furthermore, we adjust the censored normal (CN), censored skew-normal (CSN), censored bimodal symmetric skew-normal (CBPN) and censored bimodal asymmetric skew-normal (CBAsPN) models.

As can be seen in Figure 3-(a), the CSN model can accommodate to some degree the asymmetry that occurs in the observations, but it fails to explain the bimodality of the variable if it is adjusted for the CBPN and CBAsPN models.

To compare between the models considered above, we use the Akaike Information Criterion (AIC; Akaike 1974) and Bayesian Information Criterion (BIC; Schwarz 1978). Table 1 shows maximum likelihood estimates for the four adjusted models. According to the AIC and BIC criterions, the CBAsPN is a better fit for $\log_{10}$(HIV-1 RNA) data.

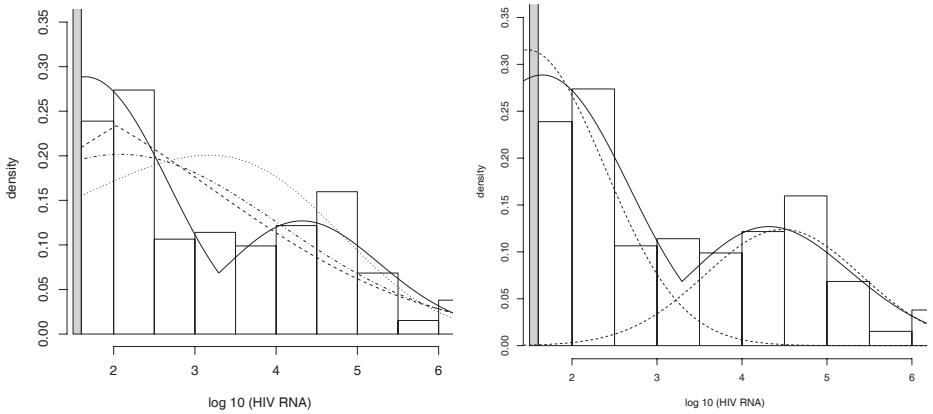
A parametric test to prove the bimodality hypothesis is given by $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$, which compares the CSN model with the CBAsPN model using the likelihood ratio statistics based on the ratio $\Lambda_1 = L_{CSN}(\widehat{\xi}, \widehat{\eta}, \widehat{\beta})/L_{CBAsPN}(\widehat{\xi}, \widehat{\eta},$

$\widehat{\alpha}, \widehat{\beta}$). Substituting the estimated values, we obtain $-2\log(\Lambda_1) = -2(-414.79 + 405.05) = 19.48$ which, when compared with the 95% critical value of $\chi^2_1 = 3.84$, indicate that the null hypotheses is clearly rejected. Then, the CBAsPN model is a good alternative for modeling $\log_{10}$(HIV-1 RNA) data.

Table 1: Maximum likelihood parameter estimates (Standard derivation in brackets) for CN, CSN, CBPN and CBAsPN models.

Model	CN	CSN	CBPN	CBAsPN
$\widehat{\xi}$	0.477(0.137)	1.689(1.147)	0.431(0.186)	1.692(0.085)
$\widehat{\eta}$	1.978(0.121)	2.362(0.767)	2.139(0.226)	1.549(0.120)
$\widehat{\alpha}$			0.396(0.576)	4.007(0.629)
$\widehat{\beta}$		–0.861 (1.013)		–0.595(0.100)
AIC	833.615	835.587	834.337	818.108
BIC	854.268	848.076	846.826	834.761

Additionally, we carry out the parametric test: $H_0 : (\alpha, \beta) = (1, 0)$ versus $H_1 : (\alpha, \beta) \neq (1, 0)$, which compares the CN model with the CBAsPN model. Using the statistic likelihood of ratio, $\Lambda_2 = L_{CN}(\widehat{\xi}, \widehat{\eta})/L_{CBAsPN}(\widehat{\xi}, \widehat{\eta}, \widehat{\alpha}, \widehat{\beta})$ leading to $-2\log(\Lambda_2) = -2(-414.81 + 405.05) = 19.52$, which is greater than the value of the chi-square with a 5% significance, $\chi^2_1 = 3.84$. This confirms that the best model to fit $\log_{10}$(HIV-1 RNA) data is the CBAsPN model. We can also observe that to some degree, the model adjusts the bimodality, but cannot adjust the asymmetry present in the observations of the viral load.



(a) (b)

Figure 3: (a) Histogram for $\log_{10}$(HIV-1 RNA): CBAsPN (solid line), CBPN (dashed line), CSN (dotted line) and CN (dashed line with points), (b) CBAsPN (solid line) CMN (dashed line).

Another model widely applied in such situations is the mixture model of two normal distributions (see Teck-Onn, Bakri, Morad & Hamid (2002), Chu, Moulton, Mack, Passaro, Barroso & Muñoz (2005), Li et al. (2006), Schneider, Margolick, Jacobson, Reddy, Martinez-Maza & Muñoz (2012), among others). The normal

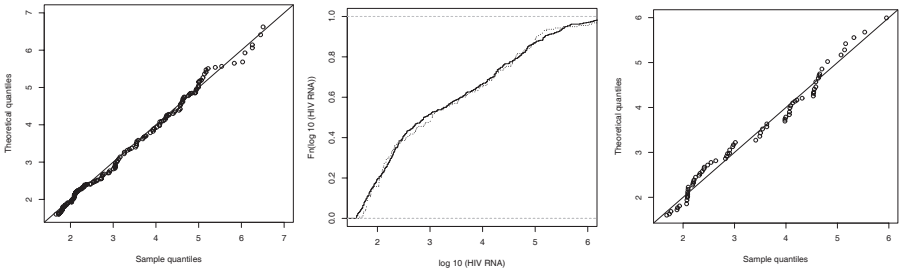
mixture model is given by

$$\rho(x;\mu_1,\sigma_1,\mu_2,\sigma_2,p) = pf_1(x,\mu_1,\sigma_1) + (1-p)f_2(x;\mu_2,\sigma_2)$$

where f_j is a normal distribution with parameters (μ_j,σ_j), $j=1,2$ and $0<p<1$. For data with detection limits, we denote them using the CMN$(\mu_1,\sigma_1,\mu_2,\sigma_2,p)$ model. Now we compare the CBAsPN with CMN$(\mu_1,\sigma_1,\mu_2,\sigma_2,p)$.

The estimated model is CMN$(1.48, 0.90, 4.48, 0.92, 0.71)$ with AIC=819.9 and BIC=840.7. This model has AIC and BIC greater than that of the CBAsPN model, so the CBAsPN model fits the data $\log_{10}$(HIV-1 RNA) better than the CMN model. Figure 3-(b) shows the estimated CBAsPN and CMN models. Furthermore, we studied the goodness of fit of the CBAsPN model getting the QQ-plot and cumulative distribution function from the MLE's.

The QQ-plot and the cumulative distribution function obtained from the estimated model are given in Figure 4(a)-(b): these show the good fit obtained in the estimated model. The total censored data corresponds to 39.92% of the sample under study, and the estimated percentage with the CBAsPN model is 39.50%, while in the CBPN model, it is 40.43%, which confirms the good fit of the CBAsPN model.



(a) (b) (c)

FIGURE 4: (a) QQ-plot men, (b) cumulative distribution function for men and (c) QQ-plot women.

These results indicate that the CBAsPN model is a suitable option for adjusting this type of information. In the case of HIV-infected women ($n = 106$) under HAART, 33.96% are below the detection limit. The estimated model was CBAsPN(1.6306, 1.8201, 2.8874, –0.5936), which estimated 32.95% of women below the detection limit. The QQ-plot given in Figure 4-(c) illustrates the good behavior of the CBAsPN model.

6.2. HIV-1 RNA Measuring by COBAS TaqMan

Plasma HIV-1 RNA was measured in 306 samples which were collected from 273 men in highly active antiretroviral therapy, with both Roche COBAS TaqMan (whose detection limit is 20 copies per mL) and Roche Amplicor (whose detection limit is 50 copies per mL) assays. See Schneider et al. (2012) for details.

The data used in this paper to illustrate the model are measurements made with the Roche TaqMan assay with $\log_{10}$(HIV-1 RNA). The non-censored information has a mean $\overline{y} = 1.3235$ and variance $s^2 = 1.5849$. Quantities $\sqrt{b_1} = 0.7012$ and $b_2 = 2.0054$ correspond to sample asymmetry and kurtosis coefficients for values above $\log_{10}(20)$, respectively. These statistics show that the data displays a high positive bias and a low kurtosis over the normal model. Figure 5 shows that the behavior of the variable under study is bimodal. Therefore, a censored bimodal asymmetric power-normal model may be used to adjust the $\log_{10}$(HIV-1 RNA) data. We adjusted the CSN and CBAsPN models.

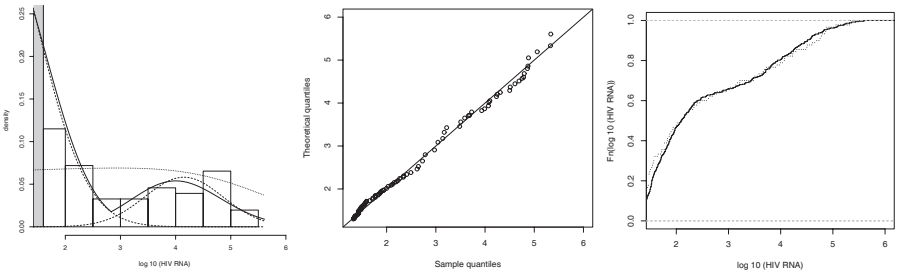
Table 2 shows maximum likelihood estimates of the proposed model. According to the AIC criterion, the model that best fits the $\log_{10}$(HIV-1 RNA) data is the CBAsPN normal model. The CSN model fails to capture the bimodality of the $\log_{10}$(HIV-1 RNA) data.

TABLE 2: Maximum likelihood parameter estimates (with (SD)) for CSN and CBAsPN models.

Model	$\widehat{\xi}$	$\widehat{\eta}$	$\widehat{\alpha}$	$\widehat{\beta}$	AIC
CSN	4.355(0.379)	11.121(1.371)		–9.637(3.274)	685.373
CBAsPN	1.531(0.090)	1.729(0.174)	6.400(0.901)	–1.175(0.148)	585.669

We can see that the estimate of α in the CSN model is significantly different from zero, which verifies the high degree of asymmetry present in the observations. Figure 5 shows that the CSN model adjusts to some extent the asymmetry present in the observations, but fails to explain the natural bimodality of the variable under study.

Again, we can prove the bimodality hypothesis $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$. Then, using the statistic likelihood of ratios, $\Lambda_3 = L_{CSN}(\widehat{\xi}, \widehat{\eta}, \widehat{\beta})/L_{CBAsPN}(\widehat{\xi}, \widehat{\eta}, \widehat{\alpha}, \widehat{\beta})$ and substituting the estimated values, we obtain $-2\log(\Lambda_3) = -2(-339.69 + 288.83) = 101.72$, which is greater than the value of the chi-square with 5% significance, $\chi^2_1 = 3.84$. Then the CBAsPN model is a good alternative for modelling $\log_{10}$(HIV-1 RNA) data.



(a) (b) (c)

FIGURE 5: (a) Histogram for $\log_{10}$(HIV-1 RNA), models: CBAsPN (solid line), CSN (dotted line) and CMN (dashed line), (b) QQ-plot and (c) cumulative distribution function for uncensored values.

We also obtained the estimate for the $CMN(0.577, 0.903, 4.15, 0.706, 0.897)$ model with AIC=585.27 (see Figure 5-(a)). There is no statistical difference between the AIC of the two models, and therefore, the two models have a similar fit. However, the CBAsPN model has fewer parameters, and is therefore less suitable than the CMN model.

Figure 5-(b)-(c) illustrate the QQ-plot and cumulative distribution function from the estimated model for uncensored data: these show the good fit of the estimated model. The total censored data corresponds to 70.58% of the study population, and the percentage estimated with the CBAsPN model is 70.69%, while with the CMN model, it is 70.74%, which illustrates the good fit of the CBAsPN model.

7. Concluding Remarks

We proposed two new distributions called the censored bimodal symmetric alpha-power and censored bimodal asymmetric alpha-power distributions. These distributions can adjust the skewness and bimodality of censored data. The inclusion of a new parameter can explain the asymmetric and bimodal behavior of an extended family of distributions, allowing a more flexible model than the censored normal, censored skew-normal models and censored mixture normal. The parameter estimation is approached by the maximum likelihood ratio and the observed information matrix is derived. Two real applications using data from HIV-infected persons illustrate the usefulness of the new models. The first application compares the censored normal, censored skew-normal and censored mixture normal with the two proposed models. The second application compares the censored skew-normal model and censored mixture normal with the CBAsPN model. The results show that the CBAsPN model fits much better to the viral load. The usefulness of the new models is tested with the likelihood ratio statistics and formal goodness-of-fit tests. The CBAsPN model has the potential to attract wider applications for censored data.

Acknowledgements

The authors acknowledge the comments and suggestions of the referees that helped to improve significantly our work. We also want to especially thank the Editor of this journal for having given suggestions and corrections of our manuscript. Moreno-Arenas thanks the Mobility Program of the Industrial University of Santander (Colombia) and the research of Salinas was supported by DIUDA 221229 (Chile).

[Recibido: enero de 2013 — Aceptado: agosto de 2013]

References

Akaike, H. A. (1974), 'A new look at statistical model identification', *IEEE Transaction on Automatic Control* **19**(6), 716–723.

Arellano-Valle, R., Castro, L., González-Farías, G. & Muñoz-Gajardo, K. (2012), 'Student-t censored regression model: Properties and inference', *Statistical Methods and Applications* **21**(4), 453–473.

Arnold, B. C., Gómez, H. W. & Salinas, H. S. (2009), 'On multiple constraint skewed models', *Statistics* **43**(3), 279–293.

Azzalini, A. (1985), 'A class of distributions which includes the normal ones', *Scandinavian Journal of Statistics* **12**(2), 171–178.

Bolfarine, H., Gómez, H. W. & Rivas, L. I. (2011), 'The log-bimodal-skew-normal model. A geochemical application', *Journal of Chemometrics* **25**(6), 329–332.

Bolfarine, H., Martínez-Flórez, G. & Salinas, H. S. (2012), 'Bimodal symmetric-asymmetric power-normal families', *Communications in Statistics-Theory and Methods* . DOI:10.1080/03610926.2013.765475.

Chu, H., Moulton, L. H., Mack, W. J., Passaro, D. J., Barroso, P. F. & Muñoz, A. (2005), 'Correlating two continuous variables subject to detection limits in the context of mixture distributions', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **54**, 831–845.

Durrans, S. R. (1992), 'Distributions of fractional order statistics in hydrology', *Water Resources Research* **28**(6), 1649–1655.

Elal-Olivero, D., Gómez, H. W. & Quintana, F. A. (2009), 'Bayesian modeling using a class of bimodal skew-elliptical distributions', *Journal of Statistical Planning and Inference* **139**, 1484–1492.

Eugene, N., Lee, C. & Famoye, F. (2002), 'Beta-normal distribution and its applications', *Communications in Statistics-Theory and Methods* **31**(4), 497–512.

Fernández, C. & Steel, M. F. J. (1998), 'On Bayesian modeling of fat tails and skewness', *Journal of the American Statistical Association* **93**(441), 359–371.

Gómez, H. W., Elal-Olivero, D., Salinas, H. S. & Bolfarine, H. (2009), 'Bimodal extension based on the skew-normal distribution with application to pollen data', *Environmetrics* **22**(1), 50–62.

Gómez, H. W., Venegas, O. & Bolfarine, H. (2007), 'Skew-symmetric distributions generated by the distribution function of the normal distribution', *Environmetrics* **18**, 395–407.

Henze, N. (1986), 'A probabilistic representation of the skew-normal distribution', *Scandinavian Journal of Statistics* **13**(4), 271–275.

Jones, M. C. (2004), 'Families of distributions arising from distributions of order statistics', *TEST* **13**(1), 1–43.

Kim, H. J. (2005), 'On a class of two-piece skew-normal distribution', *Statistic* **39**(6), 537–553.

Li, X., Chu, H., Gallant, J. E., Hoover, D. R., Mack, W. J., Chmiel, J. S. & Muñoz, A. (2006), 'Bimodal virological response to antiretroviral therapy for HIV infection: an application using a mixture model with left censoring', *Journal of Epidemiology and Community Health* **60**(9), 811–818.

Martínez-Flórez, G., Bolfarine, H. & Gómez, H. W. (2013), 'The alpha-power tobit model', *Communications in Statistics-Theory and Methods* **42**(4), 633–643.

Mudholkar, G. S. & Hutson, A. D. (2000), 'The epsilon-skew-normal distribution for analyzing near-normal data', *Journal of Statistical Planning and Inference* **83**(2), 291–309.

Pewsey, A. (2000), 'Problems of inference for Azzalini's skew normal distribution', *Journal of Applied Statistics* **27**(7), 859–870.

Schneider, M., Margolick, J., Jacobson, L., Reddy, S., Martinez-Maza, O. & Muñoz, A. (2012), 'Improved estimation of the distribution of suppressed plasma HIV-1 RNA in men receiving effective antiretroviral therapy', *Journal of Acquired Immune Deficiency Syndromes* **59**(4), 389–392.

Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics* **6**(2), 461–464.

Teck-Onn, L., Bakri, R., Morad, Z. & Hamid, M. A. (2002), 'Bimodality in blood glucose distribution', *Diabetes Care* **25**(12), 2212–2217.

Tobin, J. (1958), 'Estimation of relationships for limited dependent variables', *Econometrica (The Econometric Society)* **26**(1), 24–36.

Appendix

Appendix A. Observed Information Matrix for CBSP Model

As is well known, the elements of the observed information matrix are computed as minus the second partial derivatives with respect to all parameters and are denoted by $j_{\xi\xi}, j_{\xi\eta}, \ldots, j_{\alpha\alpha}$. Assuming that f'' exists and making $w_i = \frac{f(|z_i|)}{F(|z_i|)}$, $s_i = \frac{f'(|z_i|)}{F(|z_i|)}$, $t_i = \frac{f''(z_i)}{f(z_i)}$, $v_i = \frac{f'(z_i)}{f(z_i)}$, $w_c = \frac{f\left(\frac{\xi}{\eta}\right)}{F\left(\frac{\xi}{\eta}\right)}$, $s_c = \frac{f'\left(\frac{\xi}{\eta}\right)}{F\left(\frac{\xi}{\eta}\right)}$, $p_c = \frac{\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha}}{1-\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha}}$, $q_c = \frac{f\left(\frac{\xi}{\eta}\right)}{1-\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha}}$ and $u_c = \log\left(F\left(\frac{\xi}{\eta}\right)\right)$.

The elements of the observed information matrix are given by

$$j_{\xi\xi} = \frac{\alpha n_0 p_c}{\eta^2}\left[\alpha p_c w_c^2 + (\alpha-1)w_c^2 + s_c\right] + \frac{1}{\eta^2}\sum_1\left\{(v_i^2 - t_i) + (\alpha-1)\left[w_i^2 - s_i\right]\right\}$$

$$j_{\eta\xi} = -\frac{\alpha n_0}{\eta^3}\left[w_c^2(\alpha\xi(p_c+1)-\xi) + \eta w_c + \alpha\xi s_c\right]$$
$$+ \frac{1}{\eta^2}\sum_1\left\{(v_i + t_i - v_i^2) + (\alpha-1)\left[\mathrm{sgn}(z_i)|z_i|w_i^2 - \mathrm{sgn}(z_i)|z_i|s_i - \mathrm{sgn}(z_i)w_i\right]\right\}$$

$$j_{\eta\eta} = \frac{\alpha\xi n_0}{\eta^4}\left[w_c^2(\alpha\xi(p_c+1)-\xi) + 2\eta w_c + \alpha\xi s_c\right]$$
$$- \frac{1}{\eta^2}\sum_1\left\{1 + \frac{1}{\eta^2}\left[2z_i v_i + z_i^2 t_i - z_i^2 v_i^2\right] + (\alpha-1)\left[2|z_i|w_i + z_i^2 s_i - z_i^2 w_i^2\right]\right\}$$

$$j_{\alpha\xi} = -\frac{n_0 p_c w_c}{\eta}\left[\alpha u_c(1+p_c)+1\right] - \frac{1}{\eta}\sum_1 \mathrm{sgn}(z_i)w_i,$$

$$j_{\alpha\eta} = \frac{n_0 p_c w_c \xi}{\eta^2}\left[\alpha u_c(1+p_c)+1\right] + \frac{1}{\eta}\sum_1 |z_i|w_i$$

and

$$j_{\alpha\alpha} = n\left[\alpha^{-2} - 2^{\alpha}(2^{\alpha}-1)^{-2}(\log 2)^2\right] + n_0 p_c u_c^2(1+p_c)$$

The elements of the expected (Fisher information matrix) are computed as n^{-1} times the expectation of the corresponding elements of the observed information matrix. This matrix clearly depends on the function f, and 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.

Appendix B. Observed Information Matrix for CBAsN Model

Similarly, as done before, it follows that the elements of the observed information matrix are given by

$$j_{\xi\xi} = \frac{n_0 r_c(\boldsymbol{\theta})}{\eta^2 A_c(\boldsymbol{\theta})}\left\{\frac{r_c(\boldsymbol{\theta})}{A_c(\boldsymbol{\theta})} - \frac{\xi}{\eta} + (\alpha-1)w_c\right\} - \frac{\beta}{\eta n_0 A_c(\boldsymbol{\theta})}\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)$$
$$\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1} + \frac{1}{\eta^2}\sum_1\left\{1 + (\alpha-1)\left[w_i^2 - \mathrm{sgn}(z_i)z_i w_i\right] + \beta^2\left[\beta z_i w_{1i} + w_{1i}^2\right]\right\}$$

$$
\begin{aligned}
j_{\eta\xi} &= -\frac{n_0 r_c(\boldsymbol{\theta})}{\eta^2 A_c^2(\boldsymbol{\theta})}\left[A_c(\boldsymbol{\theta}) + \frac{\xi}{\eta} r_c(\boldsymbol{\theta})\right] - \frac{n_0 \xi E_c(\boldsymbol{\theta})}{\eta^2 A_c(\boldsymbol{\theta})} + \frac{1}{\eta^2}\sum_1 \beta\big[\beta^2 z_i^2 w_{1i} + \beta z_i w_{1i}^2 \\
&\quad - w_{1i}\big], + \frac{1}{\eta^2}\sum_1 \left\{2z_i + (\alpha-1)\left[-z_i w_i^2 - \operatorname{sgn}(z_i) z_i^2 w_i + \operatorname{sgn}(z_i) w_i\right]\right\} \\
j_{\beta\xi} &= -\frac{n_0 \xi}{\eta^2}\frac{\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1}}{A_c(\boldsymbol{\theta})} - \frac{n_0}{\eta}\frac{r_c(\boldsymbol{\theta} B_c(\boldsymbol{\theta}))}{A_c^2(\boldsymbol{\theta})} \\
&\quad + \frac{1}{\eta^2}\sum_1 \left\{\eta w_{1i} - \beta\left[\beta z_i^2 w_{1i} + z_i w_{1i}^2\right]\right\} \\
j_{\alpha\xi} &= -\frac{n_0 r_c(\boldsymbol{\theta})}{\eta\, A_c^2(\boldsymbol{\theta})}\left[B_c(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta})\log\left(\Phi\left(\frac{\xi}{\eta}\right)\right)\right] - \frac{1}{\eta}\sum_1 \operatorname{sgn}(z_i) w_i \\
j_{\eta\eta} &= \frac{n_0 r_c(\boldsymbol{\theta})}{\xi\eta^4 A_c(\boldsymbol{\theta})}\left[2\eta - \xi\left(\frac{\xi}{\eta} - (\alpha-1)w_c\right) + \xi\frac{r_c(\boldsymbol{\theta})}{A_c(\boldsymbol{\theta})}\right] \\
&\quad - \frac{n_0\beta\xi^2}{\eta^4 A_c(\boldsymbol{\theta})}\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1} \\
&\quad + \frac{1}{\eta^2}\sum_1 \left\{-1 + 3z_i^2 + (\alpha-1)\left[-2|z_i| w_i + z_i^2 w_i^2 + |z_i|^3 w_i\right] - \beta\eta z_i w_{1i}\right\} \\
&\quad + \frac{\beta}{\eta^2}\sum_1 \left[\beta^2 z_i^3 w_{1i} + \beta z_i^2 w_{1i}^2 - 2 z_i w_{1i}\right] \\
j_{\beta\eta} &= \frac{n_0\xi}{\eta^3 A_c(\boldsymbol{\theta})}\left[\eta r_c(\boldsymbol{\theta}) D_c(\boldsymbol{\theta}) + \xi\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1}\right] \\
&\quad + \frac{1}{\eta}\sum_1 [z_i w_{1i} - \beta^2 z_i^3 w_{1i} - \beta z_i^2 w_{1i}^2] \\
j_{\alpha\eta} &= \frac{n_0 \xi r_c(\boldsymbol{\theta})}{\eta^2 A_c^2(\boldsymbol{\theta})}\left[B_c(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta})\log\left(\Phi\left(\frac{\xi}{\eta}\right)\right)\right] + \frac{1}{\eta}\sum_1 |z_i| w_i \\
j_{\beta\beta} &= \frac{n_0}{A_c^2(\boldsymbol{\theta})}\left[D_c^2(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta}) M_c(\boldsymbol{\theta})\right] + \sum_1 [\beta z_i^3 w_i + z_i^2 w_{1i}^2] \\
j_{\alpha\beta} &= \frac{n_0}{A_c^2(\boldsymbol{\theta})}\left[B_c(\boldsymbol{\theta}) D_c(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta}) H_c(\boldsymbol{\theta})\right] \\
j_{\alpha\alpha} &= n\left[\alpha^{-2} - 2^{\alpha}(2^{\alpha}-1)^{-2}(\log 2)^2\right] + \frac{n_0}{A_c^2(\boldsymbol{\theta})}\left[B_c^2(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta}) N_c(\boldsymbol{\theta})\right]
\end{aligned}
$$

where $w_{1i} = \phi(\beta z_i)/\Phi(\beta z_i)$,

$$
E_c(\boldsymbol{\theta}) = \frac{r_c(\boldsymbol{\theta})}{\eta^2}\left[-\xi + (\alpha-1)\eta w_c\right] - \frac{\beta}{\eta}\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1}
$$

$$H_c(\boldsymbol{\theta}) = -\int_{\frac{\xi}{\eta}}^{\infty} z\phi(z)\{\Phi(z)\}^{\alpha-1}\log(\Phi(z))\phi(\beta z)\,dz$$
$$M_c(\boldsymbol{\theta}) = \beta\int_{\frac{\xi}{\eta}}^{\infty} z^3\phi(z)\{\Phi(z)\}^{\alpha-1}\phi(\beta z)\,dz$$
$$N_c(\boldsymbol{\theta}) = \int_{\frac{\xi}{\eta}}^{\infty} \phi(z)\{\Phi(z)\}^{\alpha-1}\log^2(\Phi(z))\{1-\Phi(\beta z)\}\,dz$$

The elements of the expected information matrix are computed numerically and depend on the functions ϕ and Φ. The MLE distribution is asymptotically normal with the variance as the inverse of the Fisher information matrix.