Use of Lèvy distribution to analyze longitudinal data with asymmetric distribution and presence of left censored data

Jorge A. Achcar^{1,a}, Emílio A. Coelho-Barros^b, José Rafael Tovar Cuevas^c,
Josmar Mazucheli^d

^aDepartamento de Medicina Social, Universidade de São Paulo, Brazil;

^bDepartamento de Matemática, Universidade Federal do Paraná, Brazil;

^cEscuela de Estadística, Universidad del Valle, Colombia;

^dDepartamento de Estatística, Universidade Estadual de Maringá, Brazil

Abstract

This paper considers the use of classical and Bayesian inference methods to analyze data generated by variables whose natural behavior can be modeled using asymmetric distributions in the presence of left censoring. Our approach used a Lèvy distribution in the presence of left censored data and covariates. This distribution could be a good alternative to model data with asymmetric behavior in many applications as lifetime data for instance, especially in engineering applications and health research, when some observations are large in comparison to other ones and standard distributions commonly used to model asymmetry data like the exponential, Weibull or log-logistic are not appropriate to be fitted by the data. Inferences for the parameters of the proposed model under a classical inference approach are obtained using a maximum likelihood estimators (MLEs) approach and usual asymptotical normality for MLEs based on the Fisher information measure. Under a Bayesian approach, the posterior summaries of interest are obtained using standard Markov chain Monte Carlo simulation methods and available software like SAS. A numerical illustration is presented considering data of thyroglobulin levels present in a group of individuals with differentiated cancer of thyroid.

Keywords: asymmetric data, covariates, left censoring, Lèvy distribution

1. Introduction

Stable distributions, usually denoted by $S_\alpha(\beta, \mu, \sigma)$, are a good alternative to generalize normal distributions. The normal distribution, the Lèvy distribution and the Cauchy distributions are special cases of this family of distributions. This family of distributions is described by four parameters α, β, μ , and σ , where the parameter α defines the fatness of the tails, $\alpha \in (0, 2]$, where $\alpha \approx 2$ indicates thin tails; β is a skewness parameter, $\beta \in [-1, 1]$ where $\beta = 0$ indicates symmetry; $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$ are, respectively, location and scale parameters (Lèvy, 1924).

A great problem associated to stable distributions $S_\alpha(\beta, \mu, \sigma)$, is that in general there is no simple closed form for their probability density functions (pdf) which makes difficult to obtain the estimates of the parameters using classical methods wherever, it is known that all pdfs of stable distributions are continuous (Gnedenko and Kolmogorov, 1968; Skorohod, 1961) and unimodal (Ibragimov and

¹ Corresponding author: Departamento de Medicina Social, Universidade de São Paulo – FMRP/USP, Ribeirão Preto, SP, Brazil. E-mail: achcar@fmrp.usp.br

Chernin, 1959; Kanter, 1976) which facilitates the use of Bayesian methods. A Bayesian analysis of this class of distributions was introduced by Buckle, (1995) using Markov chain Monte Carlo (MCMC) methods (Achcar *et al.*, 2013). The use of Bayesian methods with MCMC simulation techniques can have great flexibility by considering latent variables (Damien *et al.*, 1999; Tanner and Wang, 1987), where samples of latent variables are simulated in each step of the Gibbs or Metropolis-Hastings algorithms.

Stable distributions have been widely used in the analysis of economics and actuarial data sets (Kleiber and Kotz, 2003), in the theoretical and applied physics (Barkai *et al.*, 2000; Tsallis *et al.*, 1995), in genetics research (Yu *et al.*, 2010) and more recent in the modelling of the ultrasound images (Ranjani and Chithra, 2015). Many of these authors, worked with the Lèvy distribution to model the variability of the data in situations with heavy tails and asymmetry; however, we did not find any work that model censored longitudinal data in presence of asymmetry, extreme data and covariates.

In applied areas as in engineering and medical studies, it is possible to have data from variables whose behavior in the nature can be described using distribution with marked asymmetry and fat tails and additionally, there is the presence of some observations which can not be measured below some small threshold by the limitation of the measuring equipment. In this direction, there are situations as in the case of the treatment of individuals with differentiated thyroid cancer, which are necessary to follow up the patient by a fixed predetermined time, that is, we have in this case, longitudinal data with left censorship, asymmetry and heavy tails. In this paper, we develop a methodology that assumes the use of the Lèvy distribution as alternative to model data obtained during the follow up of the individuals, whose empirical distribution of frequencies shows a marked asymmetry in the presence of left censoring, extreme values and covariates. For this purpose, we first conducted a simulation study to evaluate the fit of the Lèvy distribution and compare it with results obtained using other existing common lifetime distributions; then we applied our procedure to a data set of individuals diagnosed with different forms of thyroid cancer, where all patients were treated with surgery in the thyroid gland and radioactive iodine therapy.

This paper is organized as follows: in Section 2, we present a background about the Lèvy distribution; in Section 3, we introduce the likelihood function; in Section 4, we introduce the Lèvy distribution in presence of censored data and covariates; in Section 5, we present an example with simulated data; in Section 6 we present an application with a real medical data; finally in Section 7 we introduce some comments and remarks.

2. Background about the Lèvy distribution

If a random variable $X \sim S_\alpha(\beta, \mu, \sigma)$, then $Z = (X - \mu)/\sigma \sim S_\alpha(\beta, 0, 1)$ (Lukacs, 1970; Nolan, 2015) and the support of all stable distributions are given in $(-\infty, \infty)$, except for $\alpha < 1$ and $|\beta| = 1$ where the support is $(-\infty, 0)$ for $\beta = 1$ and $(0, \infty)$ for $\beta = -1$ (Feller, 1971). If $\alpha < 1$, the variance is infinite and the mean of the stable distribution does not exists. The characteristic function $\Phi(\cdot)$ of a stable distribution is given by,

$$\log [\Phi(t)] = \begin{cases} i\mu t - |\sigma t|^\alpha \left[1 - i\beta \text{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right) \right], & \text{for } \alpha \neq 1, \\ i\mu t - |\sigma t|^\alpha \left[1 + i\beta \text{sign}(t) \left(\frac{2}{\pi}\right) \log |t| \right], & \text{for } \alpha = 1, \end{cases} \quad (2.1)$$

Figure 1: Plots of the standard Lèvy density assuming different values of the parameter σ .

where $i = \sqrt{-1}$ and $\text{sign}(\cdot)$ function is given by,

$$\text{sign}(x) = \begin{cases} -1, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases} \quad (2.2)$$

When $\alpha = 2$ and $\beta = 0$, we have the normal distribution with mean equals to μ and variance equals to $2\sigma^2$; if $\alpha = 1/2$ and $\beta = 1$, we have a Lèvy distribution with pdf given by,

$$f(t; \mu, \sigma) = \left(\frac{\sigma}{2\pi}\right)^{\frac{1}{2}} (t - \mu)^{-\frac{3}{2}} \exp\left[-\frac{\sigma}{2(t - \mu)}\right], \quad (2.3)$$

where $t \geq \mu$.

The standard Lèvy distribution is given when $\mu = 0$, that is, we have a one parameter distribution with pdf given by,

$$f(t; \sigma) = \left(\frac{\sigma}{2\pi}\right)^{\frac{1}{2}} t^{-\frac{3}{2}} \exp\left(-\frac{\sigma}{2t}\right), \quad (2.4)$$

where $t > 0$.

In Figure 1, we have the plots of the Lèvy pdf assuming $\mu = 0$ and different values of the scale parameter σ .

The cumulative distribution function (cdf) of the random variable T with a standard Lèvy distribution with pdf (2.4) is given by,

$$F(t; \sigma) = P(T < t) = \text{erfc}\left[\left(\frac{\sigma}{2t}\right)^{\frac{1}{2}}\right], \quad (2.5)$$

Figure 2: Plots of the hazard Lèvy distribution assuming different values of the parameter σ .

where $\text{erfc}(x) = 1 - \text{erf}(x)$ is the complementary error function and the error function $\text{erf}(x)$ is given by,

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-k^2) dk. \quad (2.6)$$

The Lèvy distribution with pdf (2.3) has undefined mean and undefined variance, but the median is given by,

$$\text{Median} = \mu + \frac{\sigma}{2 \left[\text{erfc}^{-1} \left(\frac{1}{2} \right) \right]^2}, \quad (2.7)$$

where the inverse complementary error function is defined as $\text{erfc}^{-1}(1-x) = \text{erf}^{-1}(x)$.

The hazard function $h(t)$ of the standard Lèvy distribution ($\mu = 0$) is given from the expression $h(t) = f(t)/S(t)$ where $f(t)$ is the pdf (2.4) and $S(t)$ is the survival function $S(t) = P(T > t) = 1 - F(t)$, where $F(t)$ is given in (2.5), that is,

$$h(t) = \frac{\left(\frac{\sigma}{2\pi} \right)^{\frac{1}{2}} t^{-\frac{3}{2}} \exp\left(-\frac{\sigma}{2t}\right)}{1 - \left\{ \text{erfc} \left[\left(\frac{\sigma}{2t} \right)^{\frac{1}{2}} \right] \right\}}. \quad (2.8)$$

In Figure 2, it is presented the plots of the standard Lèvy hazard function ($\mu = 0$) with different values for the parameter σ using the approximation (Abramowitz and Stegun, 1992),

$$\text{erf}(x) \approx 1 - \frac{1}{(1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4)^4}, \quad (2.9)$$

(maximum error: 5×10^{-4}), where $a_1 = 0.278393$, $a_2 = 0.230389$, $a_3 = 0.000972$, and $a_4 = 0.078108$.

3. Likelihood function

Let $T_1, \dots, T_n$ be a random sample of size n of a standard Lèvy distribution ($\mu = 0$) with pdf (2.4). The likelihood function for σ is given by,

$$L(\sigma; t) = \left(\frac{\sigma}{2\pi}\right)^{\frac{n}{2}} \prod_{i=1}^n t_i^{-\frac{3}{2}} \exp\left(-\frac{\sigma}{2} \sum_{i=1}^n t_i^{-1}\right). \quad (3.1)$$

The log likelihood function for σ is given by,

$$l(\sigma; t) = \frac{n}{2} [\log(\sigma) - \log(2\pi)] - \frac{3}{2} \sum_{i=1}^n \log(t_i) - \frac{\sigma}{2} \sum_{i=1}^n t_i^{-1}. \quad (3.2)$$

The maximum likelihood estimator (MLE) for σ is given by,

$$\hat{\sigma} = \frac{n}{\sum_{i=1}^n t_i^{-1}}. \quad (3.3)$$

Hypothesis tests and confidence intervals for the parameter σ could be obtained using standard asymptotic results for the maximum likelihood estimator for σ , as for example using the asymptotic normality of the MLE $\hat{\sigma} \sim N(\sigma, 2\hat{\sigma}^2/n)$. In this way a 95% confidence interval for σ is given by,

$$\left(\hat{\sigma} - 1.96\hat{\sigma} \sqrt{\frac{2}{n}}; \hat{\sigma} + 1.96\hat{\sigma} \sqrt{\frac{2}{n}}\right). \quad (3.4)$$

Assuming the Lèvy distribution with $\mu \neq 0$ and density (2.3), the MLE for μ and σ are given, respectively, by,

$$\hat{\mu} = X_{(1)} \quad \text{and} \quad \hat{\sigma} = \frac{n}{\sum_{i=1}^n (t_i - \hat{\mu})^{-1}}, \quad (3.5)$$

where $X_{(1)} = \text{minimum}(X_1, \dots, X_n)$.

4. Presence of censored data and covariates

In presence of left censored data, the likelihood function for σ (standard Lèvy distribution) is given by,

$$L(\sigma; t, \delta) = \sum_{i=1}^n [f(t_i; \sigma)]^{\delta_i} [F(t_i; \sigma)]^{1-\delta_i}, \quad (4.1)$$

where $f(t; \sigma)$ is given in (2.4), $F(t; \sigma)$ is given in (2.5) and δ is an indicator of censored data ($\delta = 1$ for uncensored data and $\delta = 0$ for censored data).

Assuming the presence of a vector of covariates $\mathbf{z} = (z_1, z_2, \dots, z_p)$ we assume the following regression model:

$$\sigma = \exp(\beta_0 + \beta_1 z_1 + \dots + \beta_p z_p). \quad (4.2)$$

Inferences for the parameters σ and the regression parameters $\beta_0, \beta_1, \dots, \beta_p$ are obtained using MLE and standard MCMC methods (Casella and George, 1992; Chib and Greenberg, 1995).

Recently, problems involving censored data were developed by Kim (2016, 2017), Khan *et al.* (2016), Khan and King (2016), and Aryal and Elbatal (2015).

Table 1: Simulated values ($\sigma = 5$)

3289.97	4.51	18.57	45.59	112.29
0.75	91.05	40.21	4456.61	4.00
36.17	3.13	1.75	8.44	3.75
21.24	3.04	1386.99	326.37	79.96
1.87	0.71	6.05	13.38	2.00
2120.27	39.47	5.19	2.66	40.64
9.06	6.58	25.40	1.65	4.08
2.88	45.63	13.49	30.31	162.06
0.56	4.68	49.61	32.08	6.17
136.35	4.80	8.91	2.81	2.35

5. An example with simulated data

Under a survival study scenario, we simulated data from a standard Lèvy distribution. Using the inversion method, we have from (2.4) and (2.5),

$$F(t; \sigma) = P(T \leq t) = 1 - \frac{2}{\sqrt{\pi}} \int_0^{\frac{\sqrt{\sigma}}{\sqrt{t}}} \exp(-v^2) dv. \quad (5.1)$$

Using the transformation $v = z/\sqrt{2}$, we have $dv = dz/\sqrt{2}$, that is, $dz = \sqrt{2}dv$ and,

$$F(t; \sigma) = P(T \leq t) = 1 - \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\frac{\sqrt{\sigma}}{\sqrt{t}}} \exp\left(-\frac{z^2}{2}\right) dz = 1 - 2\Phi\left[\frac{\sqrt{\sigma}}{\sqrt{t}}\right], \quad (5.2)$$

where $\Phi(t)$ is the cdf of a standard $N(0, 1)$ normal distribution.

In this way, considering a given fixed value of σ , we use the fact that $F(t; \sigma)$ has an uniform $U[0, 1]$ distribution to simulate a value u from the uniform distribution $U[0, 1]$ to get a value t of the standard Lèvy distribution, from the relation,

$$\Phi\left[\frac{\sqrt{\sigma}}{\sqrt{t}}\right] = \frac{1-u}{2}, \quad (5.3)$$

that is,

$$t = \frac{\sigma}{\left(\Phi^{-1}\left[\frac{1-u}{2}\right]\right)^2}. \quad (5.4)$$

Considering $\sigma = 5$, we simulated 50 values from a standard Lèvy distribution using the R Software (R Core Team, 2014) (Table 1).

The sample mean, sample median and sample standard-deviation for the simulated lifetimes in Table 1 are given, respectively by, 254.32, 8.98, and 833.80. Assuming the standard Lèvy distribution with pdf (2.4), the MLE for σ is given by 4.2092 (see (3.3)). A 95% confidence interval for σ (see (3.4)) is given by (2.5592; 5.8592). The MLE of the median (2.7) using the approximation (2.9), is given by 9.2530.

In Figure 3, it is presented the histogram of the simulated lifetime data with the fitted standard Lèvy pdf curve from where it is observed a good fit of the model for the dataset. Using some existing popular lifetime distributions, we computed the MLEs for the median lifetime obtaining a value of 176.28 with the exponential distribution, 20.2076 with the Weibull distribution and 12.8717 with the

Figure 3: Histogram of the simulated lifetime data versus the fitted standard Lèvy probability density function.

Table 2: Goodness-of-fit statistics

Distribution	$-2 \times \log [L(\theta)]$	AIC	BIC
Lèvy	392.3	394.3	396.2
Exponential	653.9	655.9	657.8
Weibull	512.0	516.0	519.9
Log-Logistic	491.9	495.9	499.7

AIC = Akaike information criterion; BIC = Bayesian information criterion.

log-logistic distribution. It is therefore possible to get different estimates for the median lifetime using these standard lifetimes. In addition, we obtained probability plots, the empirical cdf and the goodness-of-fit statistics for each one of these distributions (Table 2, Figure 4, and Figure 5). From these plots we observe that the fits of these distributions are inadequate for the data set introduced in Table 1, wherever it was observed a good fit of the data to the standard Lèvy distribution (smaller goodness-of-fit statistics).

Now, assuming a Bayesian approach for the standard Lèvy distribution with a gamma non-informative prior distribution, $\text{Gamma}(0.000001, 0.000001)$, for σ , we find (use of SAS/MCMC procedure (SAS, 2010a), with a “burn-in sample” of size 5,000 and a final Gibbs sample of size 1,000) a Bayesian estimator (posterior mean) given by 4.1899 and 95% credible interval for σ given by (2.7004; 6.0862).

Assuming left censored data in $t < 5$ for the simulated values in Table 1, we obtained the goodness-of-fit statistics for each distributions (Lèvy, Exponential, Weibull, and Log-Logistic). These values are presented in Table 3.

The results of Table 3 it indicated a good fit of the left censored data to the standard Lèvy distribution (smaller goodness-of-fit statistics).

In this section we also study the performance of the MLE estimation for σ given in (3.3). We have

Figure 4: Probability plots assuming some usual lifetime distributions.

taken sample sizes $n = 10, 20, 30, 40, \dots, 100$ and $\sigma = 0.5, 1.0, 1.5$, and 2.0 . For each combination (n, σ) we have generated $B = 10,000$ pseudorandom samples from the Lèvy distribution using inverse cdf method (5.4) from where it was obtained a sample $\mathbf{t} = (t_1, \dots, t_n)$ to be used in the study.

For each combination (n, σ) we compute the bias and the root mean-squared error (RMSE) given, respectively, by,

$$\text{Bias}(\hat{\sigma}) = \frac{1}{B} \sum_{i=1}^B (\hat{\sigma}_i - \sigma) \quad \text{and} \quad \text{RMSE}(\hat{\sigma}) = \sqrt{\frac{1}{B} \sum_{i=1}^B (\hat{\sigma}_i - \sigma)^2}, \quad (5.5)$$

Figure 5: Plot of empirical cumulative distribution function versus the estimated cumulative distribution function.

Table 3: Goodness-of-fit statistics (left censored data in $t < 5$)

Distribution	$-2 \times \log [L(\theta)]$	AIC	BIC
Lèvy	204.1	206.1	207.3
Exponential	327.5	329.5	330.7
Weibull	222.6	226.6	229.1
Log-Logistic	225.9	229.9	232.3

AIC = Akaike information criterion; BIC = Bayesian information criterion.

where $\hat{\sigma}_i$ is the value of σ_i obtained in the i^{th} iteration.

In Figure 6 it is presented the behavior of the MLE, for both bias and the root of the mean-squared error. We observe from these charts that: 1) the biases of the MLE for the parameter σ tend to zero for large sample sizes n , that is, the estimator is asymptotically unbiased; 2) the RMSE of the MLE for the parameter σ tend to zero for large sample sizes n , that is, the estimator is consistent; 3) larger biases and RMSE are observed for large values of σ .

6. An application with medical data

To illustrate our proposed methodology, we used a medical data set related with differentiated cancer of thyroid. The data set consists of 91 patients treated by the head and neck surgeon Gabriel Sánchez de Guzmán who works in a level IV hospital located in Bogotá Colombia. This data set was also used by López *et al.*, (2014), in a descriptive study to evaluate the relationship between initial thyroglobulin levels and presence of recurrence of cancer one year after application of the treatment. Each patient had surgery to remove the thyroid gland and was then treated with radioiodine I-131. A blood sample was obtained to measure the thyroglobulin level ($Tg1$) before starting therapy with iodine. Two more measures were obtained after the application of the therapy, the first approximately six months

Figure 6: Estimated bias and RMSE for σ considering different sample sizes. RMSE = the root of the mean-squared error.

after the last therapy session ($Tg2$) and the second one year later ($Tg3$). In addition doctor Sánchez authorized us to use four covariates: age, sex, tumor size, dose of radioiodine used in therapy. The information on the thyroglobulin levels data equal to 0.1 are considering left censored since the measuring instrument does not detect values less than 0.1. Under our view, the main goal of this study was to check if the three responses are affected by the covariates and obtain predictions on thyroglobulin levels in the three time moments. Figure 7 presents the histograms of the thyroglobulin levels data in the three measurement times.

From the histograms in Figure 7, it is observed that some observations are large when compared to others values (very asymmetric data), that is, standard lifetime distributions as Weibull are not appropriate to analyze this data set. It is also observed the presence of left censored data in $Tg2$ and $Tg3$ (values equal to zero are left censored at $t = 0.1$). That is,

$$F(0.1) = P(T < 0.1) = \text{erfc}\left(\sqrt{\frac{\sigma}{2 \times 0.1}}\right) = \text{erfc}\left(\sqrt{\frac{\sigma}{0.2}}\right). \quad (6.1)$$

To analyze the Thyroglobulin data set, it was assumed three standard Lévy distributions with parameters σ_1 , σ_2 , and σ_3 in presence (or not) of covariates. We consider three situations in our analysis: 1) not considering the presence of covariates and considering independent responses, 2) considering the presence of covariates and independent responses, 3) considering the presence of covariates and dependent responses.

6.1. Analysis not considering the presence of covariates and assuming independent responses

Considering the first case (not presence of covariates and independent responses) to analyze the proposed medical data set, we get inferences for parameter σ under a classical approach (maximum likelihood estimates) and under a Bayesian approach. These results are presented in Table 4.

Figure 7: Histograms for $Tg1$, $Tg2$, and $Tg3$.Table 4: Inference results for Lèvy σ parameter assuming $Tg1$, $Tg2$, and $Tg3$ independent

Variable	MLE (SE)	95% confidence interval	PM (SD)	95% credible interval
$Tg1$	0.8790 (0.1303)	(0.6202; 1.1379)	0.8777 (0.1324)	(0.6399; 1.1535)
$Tg2$	0.1358 (0.0244)	(0.0874; 0.1843)	0.1359 (0.0252)	(0.0933; 0.1912)
$Tg3$	0.0747 (0.0155)	(0.0439; 0.1056)	0.0752 (0.0159)	(0.0462; 0.1090)

MLE = maximum likelihood estimates; SE = standard error; PM = posterior mean; SD = standard deviation.

Maximum likelihood estimates were obtained in SAS/NLMIXED procedure (SAS, 2010b), by applying the Newton-Raphson algorithm. To obtain the Bayesian estimates we have used MCMC methods available in SAS software 9.2, SAS/MCMC (SAS, 2010a). A single chain has been used in

Figure 8: Fitted models for the data.

the simulation of sample. We simulated other 10,000 Gibbs samples taking every 10^{th} sample, to get approximated uncorrelated values that result in a final chain of size 1,000. Usual existing convergence diagnostics available in the literature as traceplots of the simulated Gibbs samples for a single chain using the SAS/MCMC procedure indicated convergence for σ .

In Figure 8, it is presented the plots of estimated cdfs based on the maximum likelihood estimates and the plots of the empirical cdfs for censored data using the Kaplan-Meier method. In addition, we obtained goodness-of-fit statistics for the standard Lévy distribution, for the exponential distribution, for the Weibull distribution and for the log-logistic distribution (Table 5).

Table 5: Values of model discrimination criteria

Variable	Distribution	$-2 \times \log [L(\theta)]$	AIC	BIC
$Tg1$	Lèvy	590.2	592.2	594.8
	Exponential	1030.0	1032.0	1034.0
	Weibull	761.6	765.6	770.6
	Log-Logistic	739.2	743.2	748.2
$Tg2$	Lèvy	225.9	227.9	230.4
	Exponential	697.3	699.3	701.8
	Weibull	375.6	379.6	384.7
	Log-Logistic	360.4	364.4	369.4
$Tg3$	Lèvy	223.7	225.7	228.3
	Exponential	803.6	805.6	808.1
	Weibull	345.2	349.2	354.2
	Log-Logistic	334.3	338.3	343.3

AIC = Akaike information criterion; BIC = Bayesian information criterion.

From the fitted models (Figure 8 and Table 5), we conclude that the thyroglobulin data are better fitted by the standard Lèvy distribution. From the results of Table 4, we observe similar inference results using classical or Bayesian approaches. For the Bayesian approach, we have used a non-informative $\text{Gamma}(0.000001, 0.000001)$ prior distribution for the parameter σ .

6.2. Analysis considering the presence of covariates and independent responses

Considering the regression model (4.2) we have three independent models for σ_1 ($Tg1$), σ_2 ($Tg2$), and σ_3 ($Tg3$), given by,

$$\sigma_{1i} = \exp(\beta_{01} + \beta_{11}z_{1i} + \beta_{21}z_{2i} + \beta_{31}z_{3i}), \quad (6.2)$$

$$\sigma_{2i} = \exp(\beta_{02} + \beta_{12}z_{1i} + \beta_{22}z_{2i} + \beta_{32}z_{3i} + \beta_{42}z_{4i}), \quad (6.3)$$

$$\sigma_{3i} = \exp(\beta_{03} + \beta_{13}z_{1i} + \beta_{23}z_{2i} + \beta_{33}z_{3i} + \beta_{43}z_{4i}), \quad (6.4)$$

where $i = 1, \dots, n$ ($n = 91$); z_{1i} denotes the patient age (in years); z_{2i} is a binary variable denoting the patient sex (1 = woman; 0 = man); z_{3i} denotes the tumor size (in cm); and z_{4i} denotes the dose of I-131 (in mCi). Observe that for the variable $Tg1$ the covariate dose of radioiodine used in therapy was omitted, since it is not possible to observe the radioiodine dose for $Tg1$. The value of this covariate was only observed for $Tg2$ and $Tg3$.

Tables 6 and 7 indicate maximum likelihood estimates and the inference results considering a Bayesian approach for regression models for each response of interest. In the Bayesian analysis it was assumed for the parameters $\beta_{l1}, \beta_{l2}, \beta_{l3}$, and β_{l4} , $l = 1, 2, 3$, independent non-informative normal priors with mean 0 and variance 1000, that is, $N(0, 1000)$. A single chain has been used in the simulation of samples for parameters considering a “burn-in-sample” of size 10,000 to eliminate the possible effect of the initial values. After this “burn-in” period, we simulated other 100,000 Gibbs samples taking every 100th sample, to obtain approximated uncorrelated values which result in a final chain of size 1,000. Usual existing convergence diagnostics available in the literature for a single chain using the SAS/MCMC procedure indicated convergence for all parameters.

From the obtained results of Table 6, it is observed that the covariate Z_2 (gender) affects the response $Tg1$ since the zero value is not included in the 95% confidence intervals for the associated regression parameters. In the same way, Z_3 (tumor size) affects $Tg2$ and $Tg3$.

The results of Table 7 indicate that the covariates Z_1 (age) and Z_2 (gender) affect the response $Tg1$ since the zero value is not included in the 95% credible intervals for associated regression parameters.

Table 6: Maximum likelihood estimates (MLE) and standard error estimates (SE) for the regression models

Variable	Parameter	MLE	SE	95% confidence interval
$Tg1$	$\hat{\beta}_{01}$	-0.0010	0.6555	(-1.3032; 1.3011)
	$\hat{\beta}_{11}$	0.0215	0.0119	(-0.0021; 0.0451)
	$\hat{\beta}_{21}$	-1.6025	0.4818	(-2.5596; -0.6454)
	$\hat{\beta}_{31}$	0.0146	0.0115	(-0.0083; 0.0375)
$Tg2$	$\hat{\beta}_{02}$	-5.1193	1.5545	(-8.2072; -2.0314)
	$\hat{\beta}_{12}$	0.0309	0.0203	(-0.0093; 0.0712)
	$\hat{\beta}_{22}$	-0.7994	0.5373	(-1.8667; 0.2679)
	$\hat{\beta}_{32}$	0.0348	0.0119	(0.0112; 0.0584)
	$\hat{\beta}_{42}$	0.0109	0.0098	(-0.0085; 0.0303)
$Tg3$	$\hat{\beta}_{03}$	-4.3124	1.8805	(-8.0477; -0.5771)
	$\hat{\beta}_{13}$	0.0156	0.0226	(-0.0293; 0.0606)
	$\hat{\beta}_{23}$	-0.6396	0.6163	(-1.8637; 0.5846)
	$\hat{\beta}_{33}$	0.0348	0.0128	(0.0094; 0.0602)
	$\hat{\beta}_{43}$	0.0050	0.0117	(-0.0182; 0.0283)

Table 7: Posterior means (PM) and standard deviation (SD) for the regression models

Variable	Parameter	PM	SD	95% credible interval
$Tg1$	$\hat{\beta}_{01}$	0.6235	0.5923	(-0.3279; 1.4727)
	$\hat{\beta}_{11}$	0.0192	0.0102	(0.0014; 0.0386)
	$\hat{\beta}_{21}$	-1.9977	0.2580	(-2.4019; -1.2970)
	$\hat{\beta}_{31}$	0.0047	0.0118	(-0.0191; 0.0259)
$Tg2$	$\hat{\beta}_{02}$	-5.5631	0.3562	(-6.1059; -4.9325)
	$\hat{\beta}_{12}$	0.0317	0.0180	(-0.0030; 0.0662)
	$\hat{\beta}_{22}$	-0.7634	0.1842	(-1.0771; -0.4025)
	$\hat{\beta}_{32}$	0.0346	0.0116	(0.0108; 0.0552)
	$\hat{\beta}_{42}$	0.0130	0.0063	(0.0003; 0.0249)
$Tg3$	$\hat{\beta}_{03}$	-3.8062	1.4488	(-6.4211; -1.0299)
	$\hat{\beta}_{13}$	0.0119	0.0219	(-0.0299; 0.0551)
	$\hat{\beta}_{23}$	-0.4974	0.6063	(-1.6931; 0.7043)
	$\hat{\beta}_{33}$	0.0336	0.0137	(0.0043; 0.0581)
	$\hat{\beta}_{43}$	0.0017	0.0094	(-0.0151; 0.0205)

In the same way, Z_2 (gender), Z_3 (tumor size) and Z_4 (dose) affect $Tg2$ and the covariate Z_3 (tumor size) affects $Tg3$. That is, the Bayesian model is more sensible to detect covariate effects.

6.3. Analysis considering the presence of covariates and dependent responses

Assuming dependent responses for $Tg1$, $Tg2$, and $Tg3$ it is assumed the presence of a latent variable w_i which captures the dependence among the responses:

$$\sigma_{ji} = w_i \exp(\beta_{0j} + \beta_{1j}z_{1i} + \beta_{2j}z_{2i} + \beta_{3j}z_{3i} + \beta_{4j}z_{4i}); \quad (6.5)$$

where $j = 1, 2, 3$ (dependent responses); $i = 1, \dots, n$ ($n = 91$); z_{1i} denotes the patient age (in years); z_{2i} is a binary variable denoting the patient sex (1 = woman; 0 = man); z_{3i} denotes de tumor size (in cm); and z_{4i} denote the dose of I-131 (in mCi). For the parameters β_{4j} we have $j = 2, 3$ since it is not possible observe this covariate for $Tg1$ ($j = 1$).

Table 8 presents the inference results considering a Bayesian approach for the regression model. We consider for the parameters β_{j1} , β_{j2} , β_{j3} , and β_{j4} , $j = 1, 2, 3$, independent normal non-informative priors with mean 0 and variance 1,000, $N(0, 1000)$, and for the latent variable w_i we assume independent uniform non-informative distributions $U[0, 100]$. R2jags library (Su and Yajima, 2015)

Table 8: Posterior means (PM) and standard deviation (SD) for dependent responses regression model

Variable	Parameter	PM	SD	95% credible interval
$Tg1$	$\hat{\beta}_{01}$	-0.7663	0.6358	(-2.6072; 0.3472)
	$\hat{\beta}_{11}$	0.0009	0.0134	(-0.0183; 0.0361)
	$\hat{\beta}_{21}$	-2.6465	0.4021	(-3.3359; -1.9144)
	$\hat{\beta}_{31}$	0.0489	0.0130	(0.0224; 0.0772)
$Tg2$	$\hat{\beta}_{02}$	-8.4149	1.7018	(-11.0656; -4.5569)
	$\hat{\beta}_{12}$	0.0030	0.0648	(-0.1272; 0.0970)
	$\hat{\beta}_{22}$	-1.1078	1.3358	(-3.4245; 1.6122)
	$\hat{\beta}_{32}$	0.0997	0.0247	(0.0432; 0.1382)
	$\hat{\beta}_{42}$	-0.0109	0.0183	(-0.0371; 0.0195)
$Tg3$	$\hat{\beta}_{03}$	19.0035	7.1519	(7.6302; 28.5281)
	$\hat{\beta}_{13}$	0.0058	0.1010	(-0.1325; 0.2615)
	$\hat{\beta}_{23}$	-3.2657	1.9517	(-6.7785; 1.1160)
	$\hat{\beta}_{33}$	0.3152	0.1490	(0.1142; 0.6747)
	$\hat{\beta}_{43}$	-0.2662	0.1088	(-0.5338; -0.1402)

Table 9: Deviance information criterion (DIC)

Model	DIC
Independent responses ($Tg1$)	583.36
Independent responses ($Tg2$)	220.00
Independent responses ($Tg3$)	225.50
Dependent responses	1472.93

Table 10: Conditional predictive ordinate (average)

Model	Independent responses	Dependent responses
$Tg1$	0.3916	0.3663
$Tg2$	5.6547	0.3560
$Tg3$	5.3591	0.3426

in R software (R Core Team, 2014) was chosen since the extructure of dependence is more simple to implement in R software compared to SAS software. Then a single chain has been used considering the simulation for each parameter for 15,000 times with a burn sample size of 5,000 to eliminate the possible effect of the initial values. We choose each 10^{th} simulated sample observation to obtain approximated uncorrelated values that result in a final chain of size 1,000. Convergence of the Gibbs Sampling algorithm was monitored using standard graphical methods, as the trace plots of the simulated samples (Brooks and Gelman, 1998)

The results of Table 8 indicate that the covariates Z_2 (gender) and Z_3 (tumor size) affect the response $Tg1$ since the zero value is not included in the 95% credible intervals for the associated regression parameters; Z_3 (tumor size) affects $Tg2$ and the covariates Z_3 (tumor size) and Z_4 (dose) affect $Tg3$. The obtained inferences are therefore different from the obtained inferences assuming independence among the responses.

Different model selection methods to choose the most adequate model could be adopted under the Bayesian paradigm (Berg *et al.*, 2004). We consider the deviance information criterion denoted by DIC (Spiegelhalter *et al.*, 2002) and the conditional predictive ordinate denoted by CPO (Gelfand *et al.*, 1992).

We used the CPO to choose the most adequate model because the DIC is not indicated to compare the independent responses and dependent responses models. We conclude, based on the results of Table 10, that the model with independent response is better fitted by the data (larger CPO), that is,

we use this model to obtain our final inferences of interest.

7. Concluding remarks

Usually, in the statistical analysis of real data sets, we could have observations generated by variables whose natural behavior is asymmetric and where some sample units show large values in comparison to other ones. Additionally, it is possible to have the presence of left censored data and covariates that make a quite complex scenario of the estimation process. To analyze this kind of data, we could use a Lèvy distribution as a good alternative for standard existing lifetime distributions available in the literature commonly used to model data with asymmetry behavior.

As a special application and motivation for the present paper, it was considered a real medical data set, that contains few individuals having very large values of the response variable in comparison with the other observations and where the equipment that measures the variable does not detect very small values below a given threshold. From the obtained results of a simulation study it was observed that the standard lifetime distributions usually used as exponential, Weibull, gamma, log-normal, log-logistic or generalized gamma are not good candidates to fit this type of data. From the results obtained by fitting the Lèvy distribution to a real data set, we concluded that this type of model can be effective to fit asymmetric data in the presence of left censored data and extreme values. It is important to point out that this paper extended the work of Achcar *et al.* (2013) by considering censored data and maximum likelihood estimation and we observe that, although the Lèvy distribution is a special case of the stable distributions considered by Achcar *et al.* (2013), in this work, we introduced a new approach to analyze data with marked asymmetry in presence of left censored data and covariates concluding that, the Lèvy distribution is a good alternative to fit data with the characteristics observed in the thyroglobulin levels data set.

Section 6 verified the good fit of the Lèvy distribution with the medical data set, where standard Bayesian diagnostic criteria as DIC or CPO were used to show the best fit of the Lèvy distribution for the thyroglobulin levels data set. From this analysis, it was possible to verify the covariates with significant effects on the response of interest and discover the best dependence structure for the data. In the same section, it was observed that, the model assuming independent responses on the time shows the best fit to the data (smaller DIC and larger CPO). Comparing the observed results under both approaches (classical and Bayesian), we concluded that the Bayesian inference approach could detect more covariate effects than a classical inference approach. Under the maximum likelihood approach we detected an effect of the covariate gender (first measure of the follow up) and tumor size (second and third evaluation) on the thyroglobulin levels. Under the Bayesian approach, it was possible to detect that the age and gender have important effects on thyroglobulin levels in the first evaluation, but in the second moment of follow up the covariate age is not important but it was observed the effect of tumor size besides the covariate gender. Finally, at the end of the follow up, the variable which presents more effect in the thyroglobulin levels is the tumor size. We conclude that these obtained inference results can be important for the medical knowledge about the thyroglobulin levels behavior during the follow up time after treatment for different forms of thyroid cancer.

In according with the obtained results from our simulation study, the more popular distributions (exponential, Weibull and log-logistic) usually used to fit asymmetry data, show a poor fit when the data are generated by a process that naturally can be model by a Lèvy distribution. This result implies that if the frequency distribution of the observed data shows marked asymmetry and extreme values, necessarily the data analyst should consider the Lèvy distribution as a candidate among the other commonly used distributions in their inference procedure.

Finally, it is important to point out that for a Bayesian analysis using standard MCMC simulation methods as the Gibbs sampling or Metropolis-Hastings algorithms (see for example, Casella and George (1992) or Chib and Greenberg (1995)) we have many available software that does not require a high computational expertise to get the posterior summaries of interest and could be used in any applied study.

Acknowledgements

The authors are grateful to Gabriel Sánchez Guzmán, a head and neck surgeon of the Clinica Cardio Infantil de Bogotá, who kindly provided the data set used to fit the models proposed in our methodology. The third author had partial financial support by the Coordenação de Aperfeiçoamento de pessoal de Nível superior (CAPES) of Brazil

References

- Abramowitz M and Stegun IA (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, Reprint of the 1972 edition.
- Achcar JA, Lopes SRC, Mazucheli J, and Linhares RR (2013). A Bayesian approach for stable distributions: some computational aspects, *Open Journal of Statistics*, **3**, 268–277.
- Aryal G and Elbatal I (2015). On the exponentiated generalized modified Weibull distribution, *Communications for Statistical Applications and Methods*, **22**, 333–348.
- Barkai E, Silbey R, and Zumofen G (2000). Lévy distribution of single molecule line shape cumulants in glasses, *Physical Review Letters*, **84**, 5339–5342.
- Berg A, Meyer R, and Yu J (2004). Deviance information criterion for comparing stochastic volatility models, *Journal of Business and Economic Statistics*, **22**, 107–120.
- Brooks SP and Gelman A (1998). General methods for monitoring convergence of iterative simulations, *Journal of Computational and Graphical Statistics*, **7**, 434–455.
- Buckle DJ (1995). Bayesian inference for stable distributions, *Journal of the American Statistical Association*, **90**, 605–613.
- Casella G and George EI (1992). Explaining the Gibbs sampler, *The American Statistician*, **46**, 167–174.
- Chib S and Greenberg E (1995). Understanding the Metropolis-Hastings algorithm, *The American Statistician*, **49**, 327–335.
- Damien P, Wakefield J, and Walker S (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**, 331–344.
- Feller W (1971). *An Introduction to Probability Theory and Its Applications. Vol. II* (2nd ed), John Wiley & Sons, New York.
- Gelfand AE, Dey DK, and Chang H (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics*, (pp. 147–167), Oxford University Press, New York.
- Gnedenko BV and Kolmogorov AN (1968). *Limit Distributions for Sums of Independent Random Variables* (Revised ed), Addison-Wesley, Reading Mass.
- Ibragimov IA and Chernin KE (1959). On the unimodality of geometric stable laws, *Teoriya Veroyatnostei i ee Primeneniya*, **4**, 453–456.
- Kanter M (1976). On the unimodality of stable densities, *The Annals of Probability*, **4**, 1006–1008.
- Khan MS and King R (2016). New generalized inverse Weibull distribution for lifetime modeling.

- Communications for Statistical Applications and Methods* **23**, 147–161.
- Khan MS, King R, and Hudson IL (2016). Transmuted new generalized Weibull distribution for lifetime modeling, *Communications for Statistical Applications and Methods* **23**, 363–383.
- Kim N (2016). On the maximum likelihood estimators for parameters of a Weibull distribution under random censoring, *Communications for Statistical Applications and Methods*, **23**, 241–250.
- Kim N (2017). Goodness-of-fit tests for randomly censored Weibull distributions with estimated parameters, *Communications for Statistical Applications and Methods*, **24**, 519–531.
- Kleiber C and Kotz S (2003). Pareto distributions. In *Statistical Size Distributions in Economics and Actuarial Sciences*, (pp. 59–106), John Wiley & Sons, New York.
- Lévy P (1924). Théorie des erreurs. La loi de gauss et les lois exceptionnelles, *Bulletin de la Société Mathématique de France*, **52**, 49–85.
- López M, Cuevas JRT, and Gutiérrez CT (2014). Niveles de tiroglobulina previo a la ablacion y persistencia / recurrencia precoz del cancer diferenciado de tiroides, *Revista Ciencias de la Salud*, **12**, 9–21.
- Lukacs E (1970). *Characteristic Functions* (2nd ed), Hafner Publishing, New York.
- Nolan JP (2015). *Stable Distributions - Models for Heavy Tailed Data*, Birkhauser, Boston.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Ranjani JJ and Chithra MS (2015). Bayesian denoising of ultrasound images using heavy-tailed Levy distribution, *IET Image Processing*, **9**, 338–345.
- SAS Institute Inc (2010a). *The MCMC Procedure, SAS/STAT 9.22 User's Guide*, Cary, AS Institute Inc, NC.
- SAS Institute Inc (2010b). *The NLMIXED Procedure, SAS/STAT 9.22 User's Guide*, Cary, AS Institute Inc, NC.
- Skorohod AV (1961). On a theorem concerning stable distributions. In *Selected Translations in Mathematical Statistics and Probability*, Institute of Mathematical Statistics and American Mathematical Society, Providence, **1**, 169–170.
- Spiegelhalter DJ, Best NG, Carlin BP, and van der Linde A (2002). A Bayesian measure of model complexity and fit (with discussion), *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 583–639.
- Su Y and Yajima M (2015). *R2jags: A Package for Running jags from R*, R package version 0.05-01.
- Tanner MA and Wong HW (1987). The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**, 528–550.
- Tsallis C, Levy S, Souza A, and Maynard R (1995). Statistical-mechanical foundation of the ubiquity of Lévy distributions in Nature, *Physical Review Letters*, **75**, 3589–3593.
- Yu ZG, Anh V, Wang Y, Mao D, and Wanliss J. (2010). Modeling and simulation of the horizontal component of the geomagnetic field by fractional stochastic differential equations in conjunction with empirical mode decomposition, *Journal of Geophysical Research: Space Physics*, **115**, A10219.