

# Bayesian analysis based on the Jeffreys prior for the hyperbolic distribution

**Thais C. O. Fonseca** \*

UNIRIO

**Helio S. Migon**

IM & COPPE - UFRJ

**Marco A. R. Ferreira**

University of Missouri, USA

## Abstract

In this work we develop Bayesian analysis based on the Jeffreys prior for the hyperbolic family of distributions. It is usually difficult to estimate the four parameters in this class: improper prior distributions may lead to improper posterior distributions, whereas proper prior distributions may dominate the analysis. In addition, maximum likelihood estimation typically requires large sample sizes of the order of thousands of observations to be reliable. We derive the Jeffreys prior for the hyperbolic family of distributions. In addition, a simulation study shows that Bayesian methods based on Jeffreys prior provide reliable point and interval estimators. Moreover, this simulation study shows that Bayesian estimators compare favorably to maximum likelihood estimators. Finally, an application to real data shows that our Bayesian methodology allows for reliable parameter estimation even for small datasets.

---

\*Address for correspondence: Instituto de Matemática e PEP/COPPE - UFRJ, Caixa Postal 68507, CEP 21945-970, Rio de Janeiro, RJ, telephone: 55-21 - 2562 8290, ext. 207, fax: 55-21 - 22709702. (Email: thais.fonseca@uniriotec.br).

*Keywords: asymmetry; Bayesian analysis; heavy tails; hyperbolic distribution; non-informative prior.*

# 1 Introduction

The hyperbolic is a flexible distribution for data that may have heavy tails and skewness. The hyperbolic distribution heavy tails result from the fact that its log-density is a hyperbola (Barndorff-Nielsen, 1977). Since its introduction by Barndorff-Nielsen (1977), the hyperbolic distribution has been used with success in many areas of application such as turbulence (Barndorff-Nielsen, 1979), biology (Blæsild, 1981) and finance (Eberlein et al., 1998; Prause, 1999; Bauer, 2000; Bingham and Kiesel, 2001). Even though the hyperbolic distribution allows for both skewness and heavy tails through easily interpretable parameters, the task of parameter estimation is not trivial. Maximum likelihood estimation typically requires large sample sizes of the order of thousands of observations to be reliable. In addition, improper prior distributions may lead to improper posterior distributions, whereas proper prior distributions may dominate the analysis. As a solution to the estimation problem, we derive here the Jeffreys prior for the hyperbolic distribution. In addition, we show through a simulation study that Bayesian methods based on this Jeffreys prior provide reliable point and interval estimators even for small datasets.

In the context of robustness, Barndorff-Nielsen (1977) pointed out that the hyperbolic distribution can be represented as a normal-mean mixture with a generalized inverse Gaussian (GIG) as mixing distribution. More specifically, let  $Y|\sigma^2 \sim N(\mu + \beta\sigma^2, \sigma^2)$  and  $\sigma^2 \sim GIG(1, \alpha^2 - \beta^2, \delta^2)$ , where  $GIG(\lambda, \phi, \gamma)$  has density given by

$$f(x; \lambda, \phi, \gamma) = \frac{(\phi/\gamma)^{\lambda/2}}{2K_\lambda(\sqrt{\phi\gamma})} x^{\lambda-1} \exp\left\{-\frac{1}{2}(\gamma x^{-1} + \phi x)\right\}, \quad (1)$$

where  $\phi, \gamma \geq 0$ ,  $\lambda \in \Re$  and  $K_\lambda$  is the modified Bessel function of third-order and index  $\lambda$ . For additional information on the GIG distribution, see Jørgensen (1982), Silva et al. (2006) and references therein. Integrating  $\sigma^2$  out, we obtain that  $Y$  has a hyperbolic distribution with density given by

$$f_{hyp}(y; \alpha, \beta, \mu, \delta) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} \exp\{-\alpha\sqrt{\delta^2 + (y - \mu)^2} + \beta(y - \mu)\}, \quad y \in \Re, \quad (2)$$

where,  $\alpha, \beta, \mu$  e  $\delta$  are parameters, satisfying  $|\beta| < \alpha$ ,  $\mu \in \Re$  and  $\delta > 0$ . The parameters  $\alpha$  and  $\beta$  determine the shape, where  $\beta$  is responsible for skewness;  $\delta$  and  $\mu$  are scale and location parameters, respectively. Figure 1 presents the density function (2) for  $\alpha = 2.5$ ,  $\beta \in \{-2, 0, 2\}$ ,  $\delta = 1$  and  $\mu = 0$ . When  $\beta$  is negative we obtain positive asymmetry, when  $\beta$  is positive we obtain negative asymmetry and  $\beta = 0$  implies a symmetric density function. We use the notation  $Hyp(\alpha, \beta, \mu, \delta)$  to denote the hyperbolic distribution with parameters  $\alpha, \beta, \mu$  and  $\delta$ .

All the moments of the  $Hyp(\alpha, \beta, \mu, \delta)$  have explicit expressions and, in particular, the mean and the variance are

$$E(Y) = \mu + \frac{\beta\delta^2 K_2(\xi)}{\xi K_1(\xi)}, \quad (3)$$

$$var(Y) = \delta^2 \left( \frac{K_2(\xi)}{\xi K_1(\xi)} + \frac{\beta^2\delta^2}{\xi^2} \left[ \frac{K_3(\xi)}{K_1(\xi)} - \left( \frac{K_2(\xi)}{K_1(\xi)} \right)^2 \right] \right), \quad (4)$$

where  $\xi = \delta\sqrt{\alpha^2 - \beta^2}$ . Note that when  $\beta = 0$  the mean is simply  $\mu$ . The mathematical properties of these univariate distributions are well-known (see Blæsild, 1981). Blæsild and Sørensen (1992) provide maximum likelihood methods to estimate parameters of this model. The `HyperbolicDist` package within the R statistical environment implements maximum likelihood estimation based on a number of numerical maximization methods.

Despite the nice properties of the hyperbolic distribution, difficulties arise in the estimation of its parameters. More specifically, for some samples the likelihood function is maximized when a combination of the parameters goes to infinity. As a consequence, the MLE may not exist with positive probability. This probability of nonexistence of the MLE is higher for smaller samples. In addition, for any finite sample size the likelihood function does not vanish in the tails. As a result, Bayesian analysis based on improper priors may lead to useless improper posterior distributions.

The problem of the likelihood function not vanishing in the tails also occurs for many other classes of distributions. For example, this problem occurs for the skew-normal distribution (Azzalini, 1985, 2005) and the Student-t distribution (Zellner, 1976). In the context of Bayesian inference for these distributions, these problems have been solved through the use of noninformative priors. For the skew-normal distribution, Liseo and Loperfido (2006) have proposed a default Bayesian solution based on the

reference prior for the parameters. For the Student-t distribution, Fonseca et al. (2008) have proposed a default Bayesian solution based on the Jeffreys prior. The proposals of Liseo and Loperfido (2006) and Fonseca et al. (2008) lead to valid proper posterior distributions. Here we obtain a default Bayesian solution based on the Jeffreys prior for the hyperbolic distribution. As we show in Section 4, our Bayesian proposal yields estimation procedures with good frequentist properties.

The remainder of this paper is organized as follows. In Section 2 we discuss the MLE difficulties associated with the hyperbolic model. In Section 3 we derive the Jeffreys prior for the parameters of the hyperbolic distribution. Section 4 presents the frequentist properties for the Bayesian and maximum likelihood estimators. Section 5 presents an application to real data that shows that our Bayesian methodology allows for reliable parameter estimation even for small datasets. Final discussions and some extensions are given in Section 6.

## 2 The hyperbolic model and MLE difficulties

Consider a random sample  $y = (y_1, \dots, y_n)$  from the hyperbolic distribution with density function given by Equation (2). Then the likelihood function is given by

$$L(\alpha, \beta, \mu, \delta; y) = \left\{ \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} \right\}^n \exp \left\{ -\alpha \sum_{i=1}^n \sqrt{\delta^2 + (y_i - \mu)^2} + \beta \sum_{i=1}^n (y_i - \mu) \right\}. \quad (5)$$

We denote the model parameters by  $\theta = (\alpha, \beta, \mu, \delta)'$ .

Maximum likelihood estimation for the hyperbolic distribution is problematic since several models are limiting or particular cases. For example, the normal distribution  $N(\mu, \sigma^2)$  is a limiting case when  $\beta = 0$ ,  $\alpha \rightarrow \infty$  and  $\delta/\alpha \rightarrow \sigma^2$ . In addition, the Laplace distribution is a limiting case when  $\beta = 0$  and  $\delta \rightarrow 0$ .

**Proposition 2.1** *The likelihood function given in Equation (5) satisfies,*

$$L(\alpha, \beta, \mu, \delta; y) = O(1), \text{ as } \delta \rightarrow \infty, \frac{\delta}{\alpha} \rightarrow \sigma^2, \beta = 0, \quad (6)$$

with  $L(\alpha, \beta, \mu, \delta; y) \rightarrow \prod_{i=1}^n \phi(y_i; \mu, \sigma^2)$ , where  $\phi(\cdot; \mu, \sigma^2)$  is the normal density with mean  $\mu$  and variance  $\sigma^2$ .

**Proof 2.1** When  $\beta \rightarrow 0$  we have that  $f(y_i) \rightarrow \frac{1}{2\delta K_1(\delta\alpha)} \exp\{-\alpha\sqrt{\delta^2 + (y_i - \mu)^2}\}$ . Moreover, for  $x$  large  $K_1(x) \rightarrow \sqrt{\frac{\pi}{2x}} \exp\{-x\}$  (Abramowitz and Stegun, 1972, p. 378, eq. 9.7.2). Thus for  $\delta \rightarrow \infty$  and  $\frac{\delta}{\alpha} \rightarrow \sigma^2$  we find  $f(y_i) \rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\alpha\delta - \alpha\sqrt{\delta^2 + (y_i - \mu)^2}\}$ . In addition,  $\alpha\delta - \alpha\sqrt{\delta^2 + (y_i - \mu)^2} = \frac{-\alpha^2(y_i - \mu)^2}{(\alpha\delta + \alpha\sqrt{\delta^2 + (y_i - \mu)^2})}$  which converges to  $\frac{-(y_i - \mu)^2}{2\sigma^2}$  as  $\delta \rightarrow \infty$  and  $\delta/\alpha \rightarrow \sigma^2$ .

**Proposition 2.2** The likelihood function given in Equation (5) satisfies,

$$L(\alpha, \beta, \mu, \delta; y) = O(1), \text{ as } \delta \rightarrow 0, \beta = 0, \quad (7)$$

with  $L(\alpha, \beta, \mu, \delta; y) \rightarrow \prod_{i=1}^n g(y_i; \alpha, \mu)$ , where  $g(\cdot; \alpha, \mu)$  is the Laplace density with parameters  $\alpha$  and  $\mu$ .

**Proof 2.2** When  $\beta \rightarrow 0$  we have that  $f(y_i) \rightarrow \frac{1}{2\delta K_1(\delta\alpha)} \exp\{-\alpha\sqrt{\delta^2 + (y_i - \mu)^2}\}$ . Moreover, for  $x \rightarrow 0$ ,  $K_1(x) \rightarrow x^{-1}$  (Abramowitz and Stegun, 1972, p. 375, eq. 9.6.9). Thus for  $\delta \rightarrow 0$  and  $\alpha$  finite constant we find  $f(y_i) \rightarrow \frac{\alpha}{2} \exp\{-\alpha|y_i - \mu|\}$ .

To illustrate the consequences of Proposition 2.1 in terms of estimation, we have simulated a dataset from  $Hyp(2, 0, 2, 1)$  with sample size  $n = 100$ . Figure 2(a) shows a contour plot of the likelihood function for  $\alpha$  and  $\delta$  holding  $\beta$  and  $\mu$  at their true values. Unfortunately, the maximum of this likelihood function is located far from the true values of  $\alpha$  and  $\delta$ . We have used the function `hyperbFit(.)` of the R-package `HyperbolicDist` to compute the MLE of the parameters using three optimization methods: Newton, quasi-Newton, and Nelder-Mead. Table 1 shows the results, that are far from the true values of the parameters. The problem is not with the optimization methods, but with the bad behavior of the likelihood function. Other authors such as for example Barndorff-Nielsen and Blæsild (1981) and Eberlein and Keller (1995) have also noticed this problematic behavior of the likelihood function. In the extreme case, the likelihood may be maximized when a combination of the parameters goes to infinity and then the MLE may not exist.

To shed light on how the likelihood function problematic behavior depends on sample size, we have computed the probability of nonexistence of the MLE of  $\alpha$  for the symmetric ( $\beta = 0$ ) and asymmetric ( $\beta = 0.1\alpha$ ) cases. For both cases, we consider sample sizes  $n \in \{30, 50, 100, 200, 2000\}$ , and parameter values  $\delta = \alpha$ ,  $\mu = 0$ , and

$\alpha \in \{0.5, 1, 2, 3, 5\}$ . Tables 2 and 3 present the results for the symmetric and asymmetric cases, respectively. Both cases lead to similar behavior with respect to the probability of nonexistence of the MLE of  $\alpha$ . More specifically, this probability increases as  $\alpha$  increases and decreases as the sample size  $n$  increases. Therefore, it seems safe to use maximum likelihood estimation for datasets of size  $n = 2000$  or larger, but for smaller sample sizes the MLE does not seem to be adequate.

In a Bayesian context, the bad behavior of the likelihood function is also an issue and, as a consequence, the choice of the prior distribution for the parameters is extremely important. As a solution for the inference problems we propose the use of the Jeffreys prior as a calibration for the likelihood function. Figure 2(b) shows, for the same simulated dataset of Figure 2(a), the resulting posterior density for  $(\alpha, \delta)$  using the Jeffreys prior that we derive in the next section. The Jeffreys prior corrects the bad behavior of the likelihood function and leads to a posterior density located close to the true parameters values.

### 3 Jeffreys prior

In this section, we derive the Jeffreys prior for the parameters of the hyperbolic distribution. As shown in Firth (1993), the Jeffreys prior works as a calibration tool for the information provided by the likelihood function. In particular, inference for small samples is possible using this approach as the Jeffreys prior compensates the fact that the likelihood function in (5) does not vanish in the tails. As we shall see, this implies very different inferences obtained using the frequentist and default Bayesian estimation approaches for the parameters in this model, with results being in favor of the Bayesian approach.

**Theorem 3.1** *The Jeffreys prior associated with model (5) is*

$$P^J(\theta) \propto |I(\theta)|^{1/2}, \quad (8)$$

where  $I(\theta)$  given by

$$\begin{pmatrix} \frac{\alpha^2 \delta^4 S_1}{\rho^2} - \frac{\delta^2 R_1}{\rho} - \frac{1}{\alpha^2} & -\frac{\alpha \beta \delta^4 S_1}{\rho^2} & -\frac{\beta}{\alpha} & \alpha \delta S_1 + \frac{\rho R_1}{\alpha \delta} - \frac{2\alpha \delta R_1}{\rho} - \frac{2}{\alpha \delta} \\ & \frac{\delta^2 R_1}{\rho} + \frac{\beta^2 \delta^4 S_1}{\rho^2} & 1 & \frac{2\beta \delta R_1}{\rho} - \beta \delta S_1 \\ & & \alpha^4 \varphi_2 - 2\alpha^4 \mu \varphi_1 + \alpha^4 \mu^2 \varphi_0 & -\delta \alpha^4 \varphi_1 + \mu \delta \alpha^4 \varphi_0 + \frac{\beta \rho R_1}{\delta} - \frac{2\beta}{\delta} \\ & & & \delta^2 \alpha^4 \varphi_0 - \frac{\rho^2 R_1^2}{\delta^2} + \frac{4\rho R_1}{\delta^2} - \frac{4}{\delta^2} \end{pmatrix}$$

represents the Fisher expected information matrix,  $\theta = (\alpha, \beta, \mu, \delta)$ ,  $R_1 = R_1(\rho) = \frac{K_2(\rho)}{K_1(\rho)}$ ,  $S_1 = S_1(\rho) = \frac{K_3(\rho)}{K_1(\rho)} - R_1^2(\rho)$  and  $\varphi_k(\alpha, \beta, \mu, \delta) = \int_{-\infty}^{\infty} \frac{y^k}{\alpha^2 \delta^2 + \alpha^2 (y - \mu)^2} f(y|\theta) dy$ .

**Proof 3.1** Define  $\vartheta = \alpha \sqrt{\delta^2 + (y - \mu)^2}$ .

Using the property that  $E \left[ \frac{\partial}{\partial \theta_i} \log L(\theta; y) \right] = 0$ ,  $i = 1, 2, 3, 4$ , the first derivatives of the log likelihood are given by

$$\begin{aligned} (a) \quad \frac{\partial}{\partial \alpha} \log L(\theta; y) &= -\frac{1}{\alpha} \{ \vartheta - E[\vartheta] \} \\ (b) \quad \frac{\partial}{\partial \beta} \log L(\theta; y) &= y - E[y] \\ (c) \quad \frac{\partial}{\partial \delta} \log L(\theta; y) &= -\alpha^2 \delta \left\{ \frac{1}{\vartheta} - E \left[ \frac{1}{\vartheta} \right] \right\} \\ (d) \quad \frac{\partial}{\partial \mu} \log L(\theta; y) &= \alpha^2 \delta \left\{ \frac{y - \mu}{\vartheta} - E \left[ \frac{y - \mu}{\vartheta} \right] \right\} \end{aligned}$$

The Fisher information matrix is given by

$$I_{ij} = E \left\{ \left( \frac{\partial}{\partial \theta_i} \log L(\theta; y) \right) \left( \frac{\partial}{\partial \theta_j} \log L(\theta; y) \right) \right\}, \text{ for } i, j = 1, \dots, 4.$$

Thus,

$$I_{11} = E \left[ \left( \frac{\partial}{\partial \alpha} \log L(\theta; y) \right)^2 \right] = \frac{1}{\alpha^2} \text{Var}(\vartheta), \text{ which is obtained from the result } E[(y - \mu)^2] = \frac{\delta^2 R_1}{\rho} + \frac{\beta^2 \delta^4 S_1}{\rho^2} \text{ from (3) and (4)}.$$

$$I_{33} = E \left[ \left( \frac{\partial}{\partial \mu} \log L(\theta; y) \right)^2 \right] = \alpha^4 (\varphi_2 - 2\mu \varphi_1 + \mu^2 \varphi_0).$$

$$I_{44} = E \left[ \left( \frac{\partial}{\partial \delta} \log L(\theta; y) \right)^2 \right] = \alpha^4 \delta^2 \text{Var} \left( \frac{1}{\vartheta} \right) = \alpha^4 \delta^2 \{ \varphi_0 - E^2 \left[ \frac{1}{\vartheta} \right] \}, \text{ which follows from the expectation of (c).}$$

$$I_{13} = E \left[ \left( \frac{\partial}{\partial \alpha} \log L(\theta; y) \right) \left( \frac{\partial}{\partial \mu} \log L(\theta; y) \right) \right] = -\alpha \text{Cov} \left( \vartheta, \frac{y - \mu}{\vartheta} \right), \text{ which follows from the expectation of } y - \mu \text{ in (3), (a) and (d).}$$

$$I_{14} = E \left[ \left( \frac{\partial}{\partial \alpha} \log L(\theta; y) \right) \left( \frac{\partial}{\partial \delta} \log L(\theta; y) \right) \right] = \alpha \delta \text{Cov} \left( \vartheta, \frac{1}{\vartheta} \right), \text{ which follows from the expectation of (a) and (c).}$$

$$\begin{aligned} I_{34} &= E \left[ \left( \frac{\partial}{\partial \mu} \log L(\theta; y) \right) \left( \frac{\partial}{\partial \delta} \log L(\theta; y) \right) \right] = -\alpha^4 \delta \text{Cov} \left( \frac{1}{\vartheta}, \frac{y - \mu}{\vartheta} \right) \\ &= -\alpha^4 \delta \{ \varphi_1 - \mu \varphi_0 - E \left[ \frac{1}{\vartheta} \right] E \left[ \frac{y - \mu}{\vartheta} \right] \}, \text{ which follows from the expectation of (c) and (d).} \end{aligned}$$

The terms  $I_{12}$ ,  $I_{22}$ ,  $I_{23}$ ,  $I_{24}$  of  $I$  were computed using direct  $I_{ij} = E \left[ -\frac{\partial^2}{\partial\theta_i\partial\theta_j} \log L(\theta; y) \right]$  as they do not depend on  $y$ .  $\square$

Note that  $P^J(\theta)$  is not obtained in closed form and numerical Newton-Cotes methods are used to perform the integrations.

## 4 Frequentist properties

This section presents frequentist properties of the maximum likelihood estimator (MLE) and of the posterior median (Bayesian estimator) of  $\theta$  based on the Jeffreys prior proposed here. We focus on the bias, the frequentist mean squared error (MSE) and the frequentist coverage of 95% credible intervals for two different sample sizes,  $n = 50$  and  $n = 100$ . We have computed the bias, the MSE and the frequentist coverage via Monte Carlo simulation. More specifically, for each true value of  $\alpha \in \{0.5, 1, 3\}$ ,  $\beta \in \{0, 0.1\alpha\}$  and  $\delta \in \{0.5, 3\}$  and each sample size  $n \in \{50, 100\}$  we have simulated 1000 samples, computed the two estimates and the credibility interval for each sample and then estimated the bias, the MSE of each estimator and the frequentist coverage. The results obtained when  $\beta = 0.1\alpha$  are very similar to the results when  $\beta = 0$  and are not presented here. In all simulations, we have assumed  $\mu = 2$ . The MLE was obtained using the Nelder-Mead method in the R-package HyperbolicDist.

A Metropolis-Hastings algorithm was used for the computation of the posterior median and the credible intervals. We sample in blocks  $(\beta, \mu)$  and  $(\alpha, \delta)$  since the parameters are very correlated and the proposal distributions are  $\mu^{(prop)} \sim N(\mu^{(k)}, d_1^2)$ ,  $\beta^{(prop)} \sim TN_{(-\alpha^{(k)}, \alpha^{(k)})}(\beta^{(k)}, d_2^2)$ ,  $\alpha^{(prop)} \sim TN_{(|\beta^{(k+1)}|, \infty)}(\alpha^{(k)}, d_3^2)$  and  $\log(\delta^{(prop)}) \sim N(\log(\delta^{(k)}), d_4^2)$ , where  $TN_A(\mu, \sigma^2)$  means the Gaussian distribution with location  $\mu$  and scale  $\sigma^2$  truncated to the region  $A$ . The variances  $d_1^2, \dots, d_4^2$  were tuned to give an acceptance rate around 30%.

Figure 3 shows, for the four parameters in the model, the square root of the MSE for each combination of  $(\alpha, \delta, n)$ . As expected, the MSE decreases as the sample size increases. Moreover, the MLEs of  $\alpha$  and  $\beta$  have MSEs two orders of magnitude larger than those of the competing Bayesian estimators. In addition, the MLEs of  $\delta$  and  $\mu$  have much larger MSEs than the competing Bayesian estimators. Figure 4 shows the absolute value of the bias of the maximum likelihood and Bayesian estimators. The



MLE of  $\mu$  has smaller bias, but the MLEs of  $\alpha$ ,  $\beta$  and  $\delta$  have much larger bias than the corresponding Bayesian estimators. Figure 5 shows the frequentist coverage, as a function of the true value of  $\alpha$ , of 95% credible intervals for  $\alpha$ ,  $\beta$ ,  $\delta$  and  $\mu$ . The performance of the credible intervals decreases for larger  $\delta$ . Finally, the frequentist coverage becomes closer to the nominal level as the sample size increases.

The use of the Jeffreys prior amends the bad behavior of the likelihood function, and as a result the Bayesian estimator has much better frequentist properties than the MLE for small sample sizes.

## 5 Illustrative example

The dataset used in this example corresponds to the size of gravels collected from a sandbar in the Mamquam River, British Columbia, Canada. This dataset is available with the `HyperbolicDist` package. Gravel sizes are determined by passing clasts through templates of particular sizes. This gives a range in which the size of each clast lies. Sizes (in mm) are then converted into psi units by taking the base 2 logarithm of the size. The midpoints specified are the midpoints of the psi unit ranges, and counts gives the number of observations in each size range. The classes are of length 0.5 psi units. There are 3574 observations as described in Rice and Church (1996). Table 4 shows the maximum likelihood estimates of the model parameters obtained using the following methods implemented in *The HyperbolicDist Package*: Quasi-Newton, Nelder and Mead and Newton Raphson.

It is worth pointing out that the estimates obtained via the three alternative optimization methods are very similar, mainly due to the fact that the dataset is very large. A relevant question is: what would happen if these methods are applied to a small dataset? In order to answer this question we next present some results using a subset of the original data (Figure 6).

Firstly, we have obtained a subsample ( $n = 200$ ) from the original dataset ( $n = 3574$ ). The sampling was done by dividing the domain of the original data in subintervals and by sampling observations in each subinterval with probability equal to the relative frequency observed in each subinterval. Next, we compare the estimates obtained using maximum likelihood estimation and the posterior median ( $MD[\theta|y]$ )

obtained using the proposed Jeffreys prior. The results are presented in Table 5.

Note that the Bayesian estimates are very good indeed. For instance, the posterior median obtained using the small dataset is very similar to the MLE obtained using the complete dataset. This can be seen in Figure 7. On the other hand, the maximum likelihood estimates obtained using only 200 observations have very bad performance when compared to the estimates obtained using the complete dataset. This example illustrates the superiority of the proposed Bayesian approach in the estimation of parameters of the hyperbolic distribution when the sample size is small.

## 6 Discussion

In this paper we have developed Bayesian analysis for the hyperbolic family of distributions using the noninformative Jeffreys prior. We have shown that our proposed methodology solves the problems associated with the bad behavior of the likelihood function. Moreover, a simulation study has shown that, when compared to the maximum likelihood estimator, our Bayesian estimator is superior.

Barndorff-Nielsen and Blæsild (1981) raised three relevant questions that we answer in our paper. Their first question was "For what sample size it is reasonable to consider a 4 parameter distribution?" For the hyperbolic distribution, the answer depends on which estimation method is used. Whereas the MLE needs a sample size of the order of thousands to be reliable, our Bayesian approach based on the Jeffreys prior provides reliable results with sample size as small as 50. Their second question was "Which parametrization of the distribution gives the most tractable form of the loglikelihood?" The second question suggests reparametrization in order to achieve tractability of the likelihood function. This is not an issue in our proposed Bayesian analysis because the Jeffreys prior is invariant under reparametrization. Their third question was "Which numerical procedure is optimal?" In the case of small samples, no numerical procedure will save the MLE; the problem is not the numerical optimization procedure but the bad behavior of the likelihood function. Conversely, for the Bayesian analysis we propose the MCMC algorithm we briefly describe in Section 4 works well even for small samples.

## Acknowledgement

The work of Fonseca was supported in part by CAPES and CRISM Grants. The work of Migon was supported in part by CNPq and CAPES Grants. The work of Ferreira was supported in part by National Science Foundation Grant DMS-0907064.

## References

- Abramowitz, M. and Stegun, A. S. (1972). *Handbook of Mathematical Functions*. Dover publications, New York.
- Azzalini, A. (1985). “A class of distributions which includes the normal ones.” *Scandinavian Journal of Statistics*, 12, 171–178.
- (2005). “The Skew-normal Distribution and Related Multivariate Families.” *Scandinavian Journal of Statistics*, 32, 159–188.
- Barndorff-Nielsen, O. (1977). “Exponentially decreasing distributions for the logarithm of particles size.” *Proc. Roy. Soc. London A*, 353, 401–419.
- (1979). “Models for Non-Gaussian Variation, with Applications to Turbulence.” *Proc. Roy. Soc. London A*, 368, 501–520.
- Barndorff-Nielsen, O. and Blæsild, P. (1981). “Hyperbolic distributions and ramifications: contributions to theory and application.” In *Statistical distributions in scientific work*, eds. C. Taillie, G. P. Patil, and B. A. Baldessari, vol. 4, 19–44. Dordrecht: Reidel.
- Bauer, C. (2000). “Value at risk using hyperbolic distributions.” *Journal of Economics and Business*, 52, 455–467.
- Bingham, N. H. and Kiesel, R. (2001). “Modelling asset returns with hyperbolic distributions.” In *Asset return distributions*, eds. J. Knight and S. Satchell, 1–20. Butterworth-Heinemann.
- Blæsild, P. (1981). “The two dimensional hyperbolic distribution and related distributions, with an application to Johanssen’s bean data.” *Biometrika*, 68, 251–263.
- Blæsild, P. and Sørensen, M. (1992). “Hyp – a computer program for analysing data by means of the hyperbolic distribution.” Research Report 248, Department of Theoretical Statistics, Aarhus University, Aarhus, Denmark.
- Eberlein, E. and Keller, U. (1995). “Hyperbolic distributions in finance.” *Bernoulli*, 1, 281–299.

- Eberlein, E., Keller, U., and Prause, K. (1998). “New insights into smile, mispricing and Value at Risk: The hyperbolic model.” *Journal of Business*, 71, 371–406.
- Firth, D. (1993). “Bias reduction of maximum likelihood estimates.” *Biometrika*, 80, 27–38.
- Fonseca, T. C. O., Ferreira, M. A. R., and Migon, H. S. (2008). “Objective Bayesian analysis for the Student-t regression model.” *Biometrika*, 95, 325–333.
- Jørgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution*, vol. 9 of Lecture Notes in Statistics. Heidelberg: Springer.
- Liseo, B. and Loperfido, N. (2006). “Default Bayesian analysis of the skew-normal distribution.” *Journal of Statistical Planning and Inference*, 136, 373–389.
- Prause, K. (1999). “The generalized hyperbolic model: estimation, financial derivatives, and risk measures.” Ph.D. thesis, University of Freiburg.
- Rice, S. and Church, M. (1996). “Sampling superficial gravels: the precision of size distribution percentile estimates.” *Journal of Sedimentary Research*, 66, 654–665.
- Silva, R. S., Lopes, H. F., and Migon, H. S. (2006). “The extended generalized inverse Gaussian distribution for log-linear and stochastic volatility models.” *Brazilian Journal of Probability and Statistics*, 20, 67–91.
- Zellner, A. (1976). “Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term.” *Journal of the American Statistical Association*, 71, 400–405.

$\theta$	Newton	Q-Newton	Nelder-Mead
$\alpha = 2$	27.4663	21.1749	27.4605
$\beta = 0$	-0.4910	-13.4426	0.1944
$\mu = 2$	2.4747	9.5139	1.8321
$\delta = 1$	25.6837	9.0368	25.6855

Table 1: Maximum likelihood estimates for  $\theta = (\alpha, \beta, \mu, \delta)$  obtained using 100 observations simulated from the  $Hyp(\alpha, \beta, \mu, \delta)$  with  $\alpha = 2$ ,  $\beta = 0$ ,  $\mu = 2$  and  $\delta = 1$ .

		$\alpha$				
		0.5	1	2	3	5
	30	0.214	0.347	0.524	0.585	0.647
	50	0.085	0.168	0.416	0.527	0.601
n	100	0.022	0.061	0.290	0.417	0.547
	200	0.000	0.006	0.159	0.301	0.463
	2000	0.000	0.000	0.000	0.015	0.182

Table 2: Probability of nonexistence of the maximum likelihood estimator of  $\alpha$ . The sample with size  $n$  was generated from  $Hyp(\alpha, \beta, \mu, \delta)$  with  $\beta = 0$ ,  $\mu = 0$  and  $\delta = \alpha$ .

		$\alpha$				
		0.5	1	2	3	5
	30	0.199	0.306	0.517	0.558	0.597
	50	0.091	0.169	0.409	0.505	0.535
n	100	0.016	0.057	0.269	0.423	0.524
	200	0.001	0.010	0.144	0.309	0.472
	2000	0.000	0.000	0.001	0.016	0.237

Table 3: Probability of nonexistence of the maximum likelihood estimator of  $\alpha$ . The sample with size  $n$  was generated from  $Hyp(\alpha, \beta, \mu, \delta)$  with  $\beta = 0.1\alpha$ ,  $\mu = 0$  and  $\delta = \alpha$ .

$\theta$	Newton	Q-Newton	Nelder-Mead
$\alpha$	5.619	5.402	5.618
$\beta$	-3.908	-3.706	-3.907
$\delta$	2.340	2.325	2.340
$\mu$	7.754	7.682	7.754

Table 4: Maximum likelihood estimates in the model  $Hyp(\alpha, \beta, \delta, \mu)$  using the complete Mamquam river dataset with  $n = 3574$  observations.

$\theta$	$SD[\theta y]$	$MD[\theta y]$	q0.025	q0.975	N	Q-N	N-M
$\alpha$	1.3364	3.7802	1.9358	6.7346	31446.898	44.691	63.980
$\beta$	1.126	-2.4233	-5.0103	-1.0207	-31445.088	-42.916	-62.186
$\delta$	0.8305	1.7831	0.8251	3.8229	0.048	1.201	1.032
$\mu$	0.5989	7.1576	6.2404	8.5026	10.071	9.694	9.827

Table 5: Posterior distribution summary and maximum likelihood estimates for the model  $Hyp(\alpha, \beta, \delta, \mu)$ , using a subsample of 200 observations from the Mamquam river dataset.

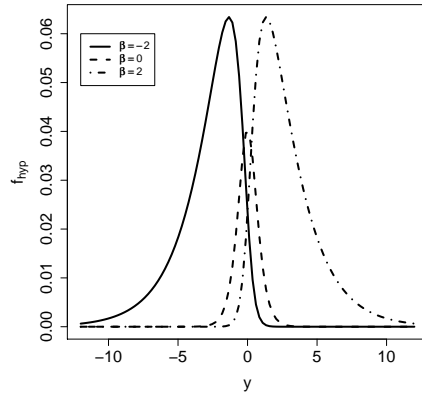
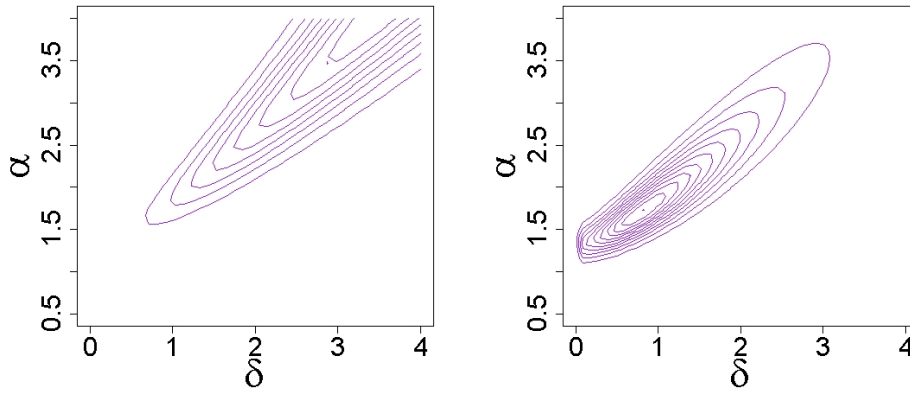


Figure 1: Density function as presented in (2) for  $\alpha = 2.5$ ,  $\beta \in \{-2, 0, 2\}$ ,  $\delta = 1$  and  $\mu = 0$ .



(a) Likelihood function.

(b) Posterior density.

Figure 2: Contour plots of the (a) likelihood function and (b) posterior density for  $(\alpha, \delta)$ , holding  $\beta$  and  $\mu$  at their true values, for a dataset of size  $n = 100$  simulated from model (2) with  $\alpha = 2$ ,  $\beta = 0$ ,  $\delta = 1$  and  $\mu = 2$ .



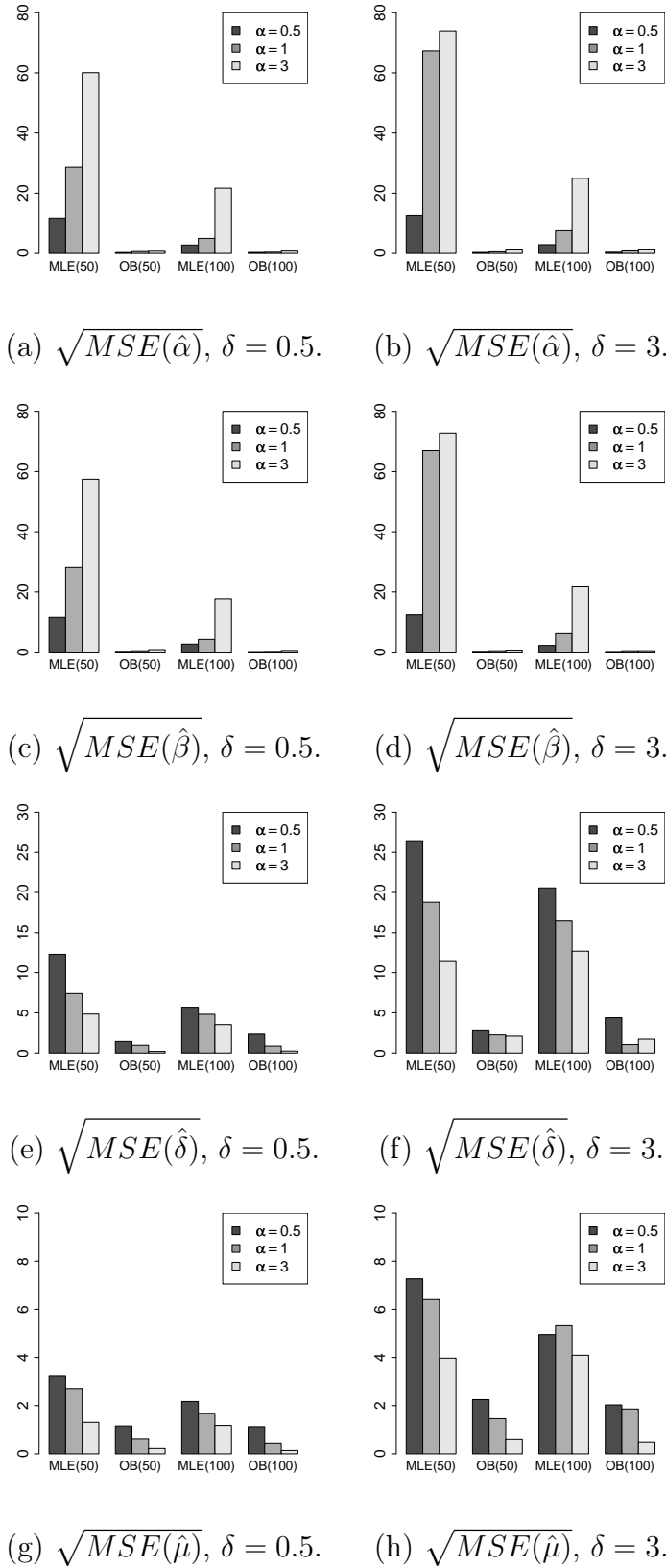
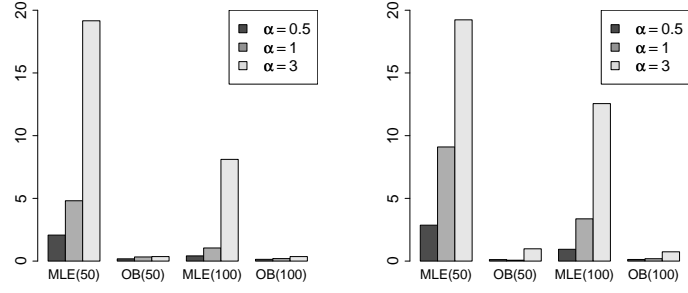
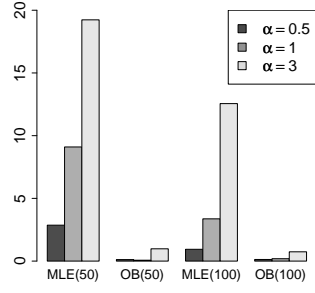


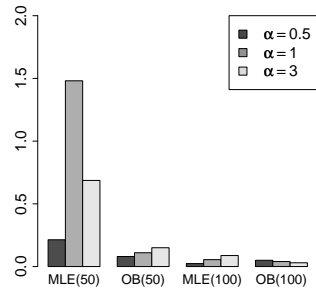
Figure 3: Square root of the mean squared error for the maximum likelihood (MLE) and Bayesian (OB) estimators of  $\alpha$ ,  $\beta$ ,  $\delta$  and  $\mu$  when  $\beta = 0$ .



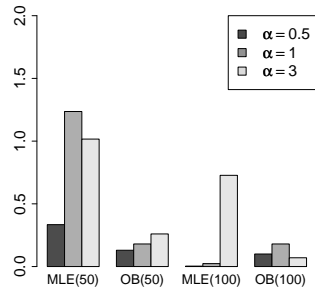
(a)  $|Bias(\hat{\alpha})|$ ,  $\delta = 0.5$ .



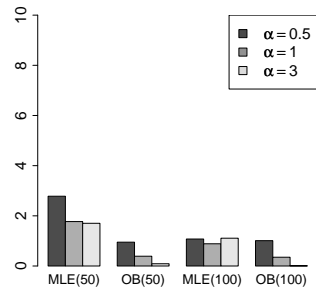
(b)  $|Bias(\hat{\alpha})|$ ,  $\delta = 3$ .



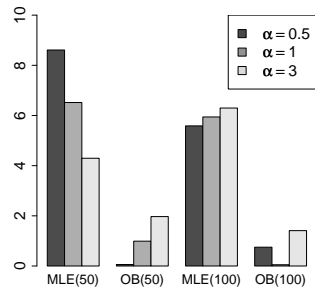
(c)  $|Bias(\hat{\beta})|$ ,  $\delta = 0.5$ .



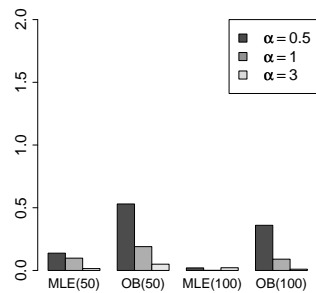
(d)  $|Bias(\hat{\beta})|$ ,  $\delta = 3$ .



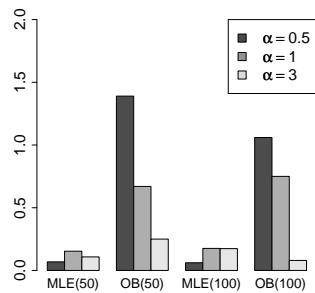
(e)  $|Bias(\hat{\delta})|$ ,  $\delta = 0.5$ .



(f)  $|Bias(\hat{\delta})|$ ,  $\delta = 3$ .

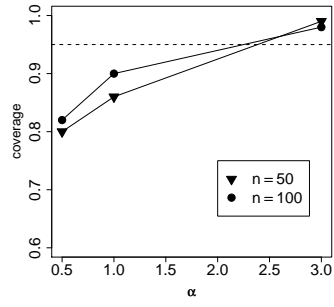


(g)  $|Bias(\hat{\mu})|$ ,  $\delta = 0.5$ .

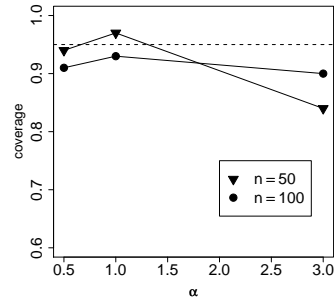


(h)  $|Bias(\hat{\mu})|$ ,  $\delta = 3$ .

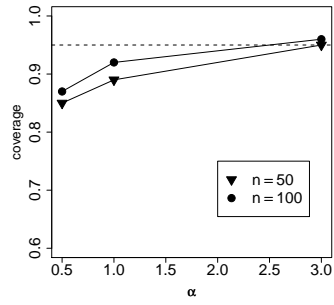
Figure 4: Absolute value of the bias for the maximum likelihood (MLE) and Bayesian estimators (OB) of  $\alpha$ ,  $\beta$ ,  $\delta$  and  $\mu$  when  $\beta = 0$ .



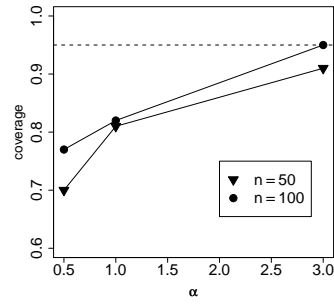
(a)  $Coverage(\hat{\alpha}), \delta = 0.5$ .



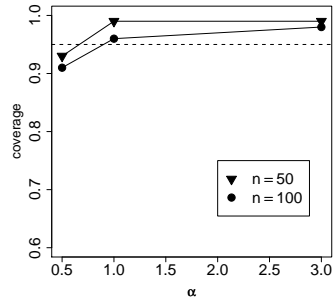
(b)  $Coverage(\hat{\alpha}), \delta = 3$ .



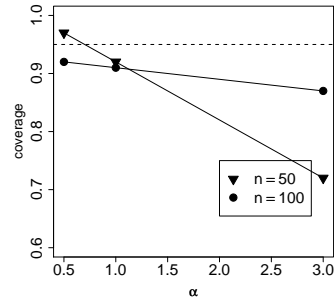
(a)  $Coverage(\hat{\beta}), \delta = 0.5$ .



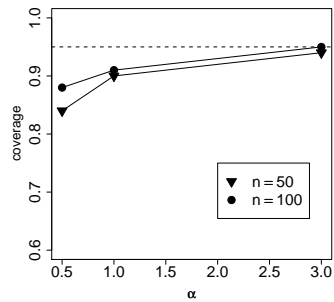
(b)  $Coverage(\hat{\beta}), \delta = 3$ .



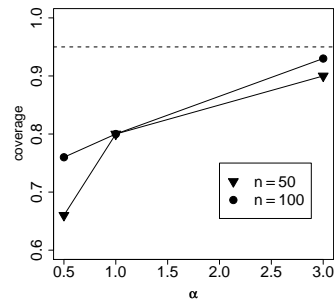
(a)  $Coverage(\hat{\delta}), \delta = 0.5$ .



(b)  $Coverage(\hat{\delta}), \delta = 3$ .

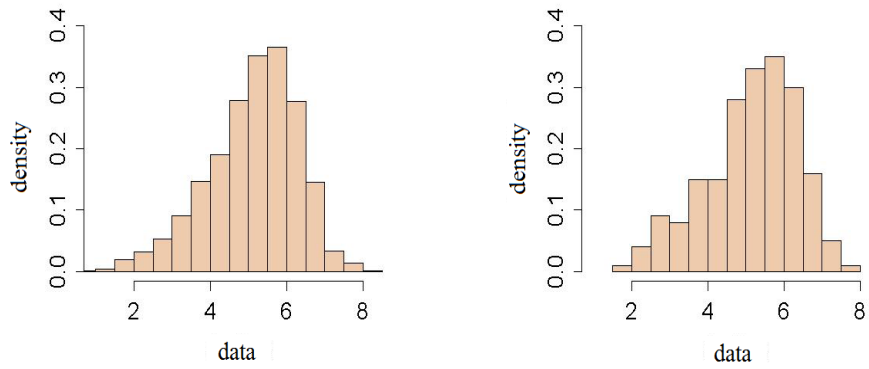


(a)  $Coverage(\hat{\mu}), \delta = 0.5$ .



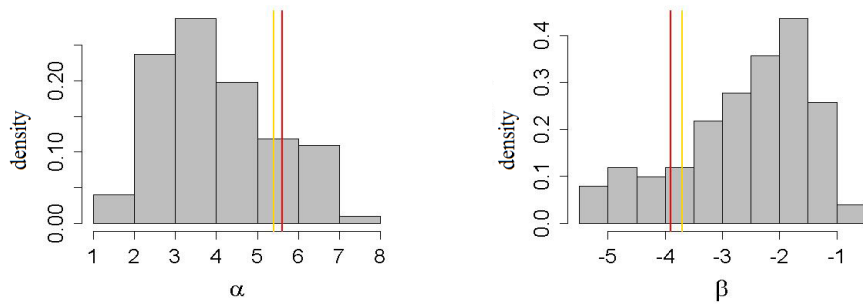
(b)  $Coverage(\hat{\mu}), \delta = 3$ .

Figure 5: Frequentist coverage, as a function of the true value of  $\alpha$ , of 95% credible intervals for  $\alpha$ ,  $\beta$ ,  $\delta$  and  $\mu$  when  $\beta = 0$ .

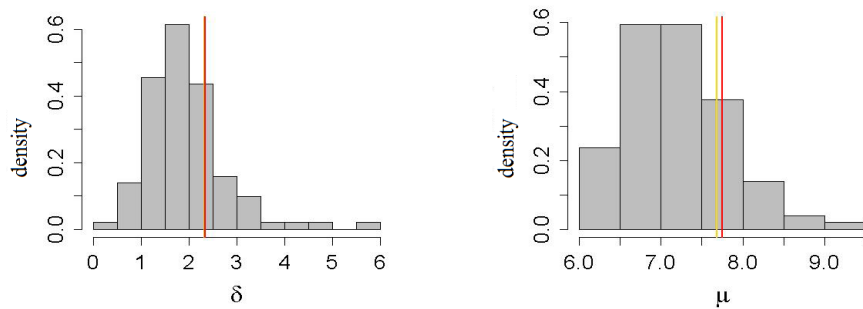


(a) Complete dataset. (b) Subsample of 200 observations.

Figure 6: Size of small stones at Mamquam river.



(a) Marginal posterior of  $\alpha$ . (b) Marginal posterior of  $\beta$ .



(c) Marginal posterior of  $\delta$ . (d) Marginal posterior of  $\mu$ .

Figure 7: Histograms of the marginal posterior distribution for the parameters in the model obtained using a subsample of size 200. The vertical lines represent the estimates obtained for the *complete dataset* using the Quasi-Newton, Nelder-Mead and Newton methods.