Bayesian Modeling Using a Class of Bimodal Skew-Elliptical Distributions

David Elal-Olivero* Héctor W. Gómez † Fernando A. Quintana‡

July, 2008

Abstract

We consider Bayesian inference using an extension of the family of skew-elliptical distributions studied by Azzalini (1985, 2005). This new class is referred to as bimodal skew-elliptical (BSE) distributions. The elements of the BSE class can take quite different forms. In particular, they can adopt both uni- and bimodal shapes. The bimodal case behaves similarly to mixtures of two symmetric distributions and we compare inference under the BSE family with the specific case of mixtures of two normal distributions. We study the main properties of the general class and illustrate its applications to two problems involving density estimation and linear regression.

*Departamento de Matemáticas, Facultad de Ingeniería, Universidad de Atacama, Chile, E-mail: delal@mat.uda.cl. Partially supported by Grant DIUDA 221140

†Departamento de Estadística, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Concepción, Chile. E-mail: hgomez@udec.cl. Partially supported by Grant FONDECYT 1060727.

‡Departamento de Estadística, Facultad de Matemática, Pontificia Universidad Católica de Chile, Santiago, Chile. E-mail: quintana@mat.puc.cl. Partially supported by Grant FONDECYT 1060729.

Key Words: Bimodality, Density Estimation, Linear Regression, Skew-Normal Distribution, Stochastic Representation.

1 Introduction

Azzalini and Capitanio (2003) proved that if g is a probability density function (pdf) that is symmetric about zero and if H is a cumulative distribution function (cdf) such that its density h is also symmetric about zero, then for any odd function $w(x)$

$$f_X(x) = 2g(x)H(w(x)), \qquad -\infty < x < \infty, \tag{1}$$

is a pdf on $\mathbb{R}$. This result has turned out to be quite useful in the process of constructing skewed distributions from symmetric ones. We denote a random variable X with pdf (1) as $X \sim S(g, h, w)$. In particular, if $g = \phi$, and $H = \Phi$ represent the standard normal pdf and cdf, respectively, then taking $w(x) = \lambda x$ for $\lambda \in \mathbb{R}$ we obtain the skew-normal (SN) distribution. This has been extensively studied in Azzalini (1985, 1986), Henze (1986), Pewsey (2000) and many others. The special case where $H' = g$ and g is the Laplace, logistic or uniform distribution, has been considered by Gupta, Chang and Huang (2002). Related constructions have been developed in Balakrishnan and Ambagaspitiya (1994) and in Arnold and Beaver (2000a, 2000b). Nadarajah and Kotz (2003) consider fixing $g = \phi$ and H given by one of the normal, Student-t, Cauchy, Laplace, logistic or uniform cdfs. Gómez, Venegas and Bolfarine (2007) consider the family of distributions generated by fixing $H = \Phi$ and letting g be any symmetric pdf. More generally, there has been a number of works exploring bimodality arising from skew-distributions. Azzalini and Capitanio (2003), Arnold, Castillo and Sarabia (2002) Ma and Genton (2004), Arellano-Valle, Gómez and Quintana (2005), and Arellano-Valle, Cortés and Gómez (2006).

From the Bayesian standpoint, inference for skew-elliptical models has been considered in Fernández and Steel (1998) and Sahu, Dey and Branco (2003). More recently, Ferreira and Steel (2006) described a general form to introduce skewness in symmetric models.

Our aim is to introduce a new family of distributions that is flexible enough to support both, uni- and bimodal shapes. Many datasets arising in practice can be adequately modeled this way and so our proposal plays a unifying role in this context.

To achieve our purpose, we resort to (1), adopting a specific form for $w(x)$ that yields the desired result. We will study properties of the proposed family of distributions as well as applications to some standard problems. In the theoretical discussion, the following stochastic representation of (1) turns out to be quite useful: let $X \sim g$ and $Y \sim h$ be independent random variables and define

$$Z = \begin{cases} X & \text{if } Y < w(X) \\ -X & \text{if } Y \geq w(X). \end{cases} \tag{2}$$

Then Z has distribution given by (1). This shows that these skew-distributions arise from a hidden truncation mechanism (Ferreira and Steel 2006).

The rest of this paper is organized as follows. Section 2 presents the new family, and develops its main property. In particular, we show how uni- and bimodal shapes are obtained. Section 3 discusses some practical issues in the general use of skew-bimodal distributions in Bayesian inference. Section 4 illustrates the use of skew-normal with bimodal shape distributions for density and linear regression problems. Special attention is given to comparing inference under the new family with mixtures of two normal distributions. A final discussion is presented in Section 5.

2 Adding Bimodality and Skewness to Symmetric Distributions

The result stated next is useful in the process of introducing skewness *and* bimodality in families of symmetric distributions.

Proposition 1 *Let f_0 be a unimodal pdf, H a cdf having pdf h, and assume both f_0 and h are symmetric about zero. If $\kappa = \int_{-\infty}^{\infty} x^2 f_0(x)\, dx < \infty$, then*

$$f(x) = 2\left(\frac{1+\alpha x^2}{1+\alpha\kappa}\right) f_0(x) H(w(x)), \qquad -\infty < x < \infty, \tag{3}$$

is a pdf for any odd function $w(\cdot)$ and any $\alpha \geq 0$.

To prove Proposition 1 consider, for a given $\alpha \geq 0$,

$$g(x) = \left(\frac{1+\alpha x^2}{1+\alpha\kappa}\right) f_0(x), \tag{4}$$

which is a symmetric pdf. The result then follows immediately from (1).

An immediate consequence of Proposition 1 is the stochastic representation of the new family by means of hidden truncation: if $X \sim \left(\frac{1+\alpha x^2}{1+\alpha\kappa}\right) f_0(x)$ is independent of $Y \sim h$ then Z defined in (2) has pdf given by (3).

Because f_0 is symmetric and unimodal, the density $g(x)$ defined in (4) is symmetric and may adopt both, uni- and bimodal shapes. Besides, it is the $H(w(x))$ factor that breaks the symmetry. For later reference we denote a random variable X with pdf $g(x)$ given in (4) by $X \sim B(f_0, \alpha)$ and a random variable Y with pdf (3) as $Y \sim SB(f_0, h, w, \alpha)$. We refer to the corresponding families of distributions that follow by letting f_0 be any symmetric unimodal pdf as *symmetric bimodal* and *skew-elliptical bimodal*, respectively. We note that the $B(f_0, \alpha)$ family contains all the symmetric unimodal distributions (taking $\alpha = 0$), and similarly, the $SB(f_0, h, w, 0)$ sub-family agrees with the class of all unimodal skew-elliptical distributions.

The above construction can be seen to generalize some well-known classes of skewed distribution. For instance, taking $f_0(x) = h(x) = \phi(x)$ and $w(x) = \lambda x$ we extend the skew-normal model studied in Azzalini (1985). When $w(x) = \lambda x + \gamma x^3$ we obtain an extension of the model presented in Ma and Genton (2004). More generally, if $f_0 = \phi$ we generalize the family of distributions in Nadarajah and Kotz (2003) and when $h = \phi$, a generalization of Gómez, Venegas and Bolfarine (2007) follows. In all these cases, the original family is recovered by choosing $\alpha = 0$.

Figure 1 shows the effect of the α parameter on the symmetric and skewed cases, using $f_0(x) = h(x) = \phi(x)$, and a skewing factor defined in terms of $w(x) = x/2$, i.e., corresponding to the skew-normal density with parameter $\lambda = 1/2$. Because in general $H(w(x)) = 1 - H(-w(x))$, the α parameter can be seen as controlling the density value at the highest mode, i.e., the one concentrating most of the mass. Taking $w(x) = -x/2$ we obtain a reflection of Figure 1 with respect to $x = 0$.

As is the case of several other families of skewed distributions, such as Azzalini's skew-normal, the bimodal class satisfies a *perturbation invariance* property, which we state next.

Proposition 2 *If* $X \sim SB(f_0, h, w, \alpha)$ *then* $|X| \sim HalfB(f_0, \alpha)$ *for any symmetric pdf* h *and function* w*.*

Proof. Letting $Y = |X|$, it follows that for $y \geq 0$

$$\begin{aligned} f_Y(y) &= f_X(y) + f_X(-y) \\ &= 2\left(\frac{1+\alpha y^2}{1+\alpha\kappa}\right) f_0(y)(H(w(y)) + H(w(-y))) \\ &= 2\left(\frac{1+\alpha y^2}{1+\alpha\kappa}\right) f_0(y)(H(w(y)) + H(-w(y))) \\ &= 2\left(\frac{1+\alpha y^2}{1+\alpha\kappa}\right) f_0(y). \qquad \square \end{aligned}$$

Proposition 2 is useful to compute moments of (3). If $X \sim SB(f_0, h, w, \alpha)$ and $Z = |X|$ then $Z \sim HalfB(f_0, \alpha)$, and the even moments of X and Z coincide. The random variable Z has a much simpler distribution, which may be thought of as extending the family of half-symmetric distributions on the positive real line. On the other hand, the odd moments of X can be expressed as

$$E[X^{2k-1}] = \frac{1}{1+\alpha\kappa} E[Y^{2k-1}] + \frac{\alpha}{1+\alpha\kappa} E[Y^{2k+1}], \qquad k = 1, 2, 3, \ldots,$$

where $Y \sim S(f_0, h, w)$.

3 Bayesian Inference Under Skew Bimodal Models

The skew-bimodal family of distributions introduced in Section 2 is quite flexible in the sense of including all symmetric and unimodal skewed densities. Practical use of such models, however, requires making choices on the pdfs f_0, h and the skewing function w. We adopt here a pragmatic viewpoint and consider models with $h = \phi$ and $w(x) = \lambda x$, so the skewing factor is $\Phi(\lambda x)$, exactly as in Azzalini's skew-normal distribution.

In the above scenario, the choice of f_0 has a critical impact in model performance. As a default candidate we suggest taking $f_0 = \phi$, which yields bimodal distributions that resemble mixtures of two normals. See the bottom panel in Figure 1. The resulting family will be referred to as the *bimodal skew-normal distribution*, with pdf defined as

$$f(x) = \sqrt{\frac{2}{\pi}} \left(\frac{1+\alpha x^2}{1+\alpha} \right) \exp\left\{ -\frac{x^2}{2} \right\} \Phi(\lambda x), \qquad x \in \mathbb{R}, \tag{5}$$

where $\lambda \in \mathbb{R}$ is the skewness parameter with $\lambda = 0$ representing symmetry of (5). A random variable X with pdf (5) is denoted as $X \sim SBN(\alpha, \lambda)$. It is interesting to point out some basic features of the $SBN(\alpha, \lambda)$ distribution. For instance, when $\lambda \to \pm\infty$ then (5) approaches a distribution that can be described as (4) with $f_0 = \phi$, but conditioned to lie on $\mathbb{R}^+$ or $\mathbb{R}^-$, as determined by the sign adopted by λ. This limit distribution is never bimodal. We thus see that λ controls the skewness of the $SBN(\alpha, \lambda)$ family, and is also involved in the shape it adopts, because for large enough $|\lambda|$ the bimodality practically disappears.

Letting $\alpha \to \infty$, (5) converges to the density function given by $2x^2\phi(x)\Phi(\lambda x)$, which is again bimodal, with one mode at each size of zero. In this case, simple calculations show that λ controls the location of the modes and the density value at the modes. When $\lambda \to 0$ the resulting density is symmetric and the modes approach $\pm\sqrt{2}$. As $\lambda > 0$ increases, the density has higher values at the right mode than at the left one, as the probability mass starts shifting towards the positive values. The reverse is observed as $\lambda < 0$ decreases. Thus the perturbation factor $\Phi(\lambda x)$ above controls the skewness with respect to the symmetric bimodal model. In the general case, the two modes are also located at each side of zero. Positive values of λ push the left mode towards zero, and in the limit case $\lambda \to \infty$, the left mode disappears and the right mode converges to $\sqrt{2}$. A similar behavior occurs when λ is negative and/or approaches $-\infty$. Thus, as in the general case, this model approaches unimodality as $|\lambda| \to \infty$.

The inferential target for the SBN model includes learning about the skewness parameter λ, and the bimodality parameter α. In addition, $\lambda = 0$ and $\alpha = 0$ have specific meanings attached (symmetry and unimodality, respectively). In defining appropriate prior distributions for (α, λ) we note that improper flat priors are not an option because they lead to improper posteriors. The information matrix corresponding to the family (5) is quite involved and not available in closed form, although straightforwardly seen to be of diagonal form. Thus, it seems adequate to think of independent priors for α and λ.

We will consider here two separate scenarios: testing and general estimation. In the first case we assume each prior distribution to be the mixture of an atomic distribution at 0 and a suitable continuous distribution. This yields posterior distributions that are of the same nature, thus allowing to specifically quantify the posterior probability of symmetry and/or unimodality. We suggest the following default prior for λ:

$$p(\lambda) = p_{0,l}\delta_\lambda(0) + (1 - p_{0,l})\sigma_l^{-1}\phi(\lambda\sigma_l^{-1}), \tag{6}$$

for $0 < p_{0,l} < 1$ and $\sigma_l > 0$. To enforce bimodality in the prior, it is convenient to choose a moderately small value of σ_l, e.g. $\sigma_l = 1$. This is because $|\lambda| > 3$ implies practically unimodal shapes. Similarly, for α we propose

$$p(\alpha) = p_{0,a}\delta_\alpha(0) + 2(1 - p_{0,a})\sigma_a^{-1}\phi(\alpha\sigma_a^{-1})I\{\alpha > 0\}, \tag{7}$$

i.e. another mixture of an atom at 0 and a normal distribution truncated at 0 from below. An alternative prior would replace the truncated normal term by $A^{-1}I\{0 < \alpha < A\}$, i.e., a uniform distribution with a large enough value of the upper limit A. This choice would give slightly less prior weight to values of α close to 0, thus discouraging symmetry.

In the second case, i.e. estimation, we drop the atoms at 0 in both (6) and (7), leading to default priors that are given by standard normals (truncated to the positive numbers whenever appropriate) and/or uniform distributions.

4 Data Illustrations

In this section we apply the bimodal skew-normal model to the estimation and testing scenarios described earlier.

4.1 Estimation for simulated data

For the purpose of model fitting we need to consider a location-scale variation of (5), i.e. we adopt

$$\frac{y_i - \mu}{\sigma} \mid \alpha, \lambda, \mu, \sigma \sim SNB(\alpha, \lambda), \tag{8}$$

with the estimation priors described in Section 3. The model specification is completed by selecting adequate priors for μ and σ. We note that for any fixed value of

α, λ and σ we have $1/\sigma \int f((x-\mu)/\sigma)\, d\mu < \infty$, where $f(x)$ is the density function (5), which follows immediately from the fact that $f(x)$ can be bounded by a constant times $(1+\alpha x^2)\exp(-x^2/2)$. Therefore a non-informative flat prior may be placed on μ, which would lead to a proper posterior. However, we do not advocate the use of non-probability prior models and choose instead a vague prior $\mu \sim N(0, 10^4)$. Similarly, the non-informative prior $\pi(\sigma) \propto 1/\sigma$ for the scale parameter again leads to a proper posterior, but we choose instead $\sigma \sim U(0, 10^4)$.

A natural competitor for this 4-parameter distribution is the 5-parameter mixture of 2 normal distributions, i.e.

$$y_i \mid \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p \sim pN(\mu_1, \sigma_1^2) + (1-p)N(\mu_1+\mu_2, \sigma_2^2), \tag{9}$$

where $0 < p < 1$, $\sigma_1^2, \sigma_2^2 > 0$, $\mu_1 \in \mathbb{R}$ and $\mu_2 > 0$, which is related to a parametrization discussed in, e.g. Mengersen and Robert (1996). See also de "Eyes" example in the Examples Volume II of WinBUGS (Spiegelhalter et al. 2003). The prior structure for this mixture model is assumed to be independent with $\mu_1 \sim N(0, 10^4)$, $\mu_2 \sim N(0, 10^4)I\{\mu_2 > 0\}$ (i.e. a normal distribution truncated to the positive real numbers), $p \sim U(0,1)$ and $\sigma_i^{-2} \sim \Gamma(0.01, 0.01)$ for $i = 1, 2$.

We applied both models to a set of $n = 1000$ simulated observations from the mixture with equal weights of a $\Gamma(2,2)$ and a normal distribution with mean -2 and variance 1. The simulated data histogram is presented in Figure 2 together with the posterior predictive densities for models (8) and (9). The two model produce similar curves, but neither provides, as was to be expected, a very close fit to the data. To compare the results of these fits we computed the pseudo-Bayes factors based on the *conditional predictive ordinate* (CPO) for model (8) against model (9), as explained in, e.g., Chen, Shao and Ibrahim (2000). The respective values of the log-CPO statistic were -1810.162 and -1815.977, leading to a pseudo-Bayes factor of 335.1611, which is interpreted as strong evidence in favor of our proposed model (8). As an additional comparison, we computed the *Bayesian Information Criterion* (BIC) values (Schwarz 1978) for the models, obtaining -3634.428 and -3655.831, respectively, which leads to identical conclusion.

4.2 Testing for a linear regression model

We turn our attention now to the testing problem under a linear regression model with skew-normal bimodal errors. As mentioned earlier, our motivation is to provide a flexible class of parametric distributions for the error term, i.e., capable of supporting various types of different behavior, including bimodality, skewness, and extended ranges of kurtosis. We specifically focus now on testing the hypotheses of symmetry and unimodality for the geyser data available on line in the R system, which consists of 272 observations on waiting times between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. Letting x_i represent the ith recorded duration of eruption and y_i the waiting time to the next eruption (in minutes) we have $n = 272$ observations. These data have been used in many places, including Härdle (1991). Azzalini and Bowman (1990) and Venables and Ripley (1999) present a slightly more complete version, and additional discussion and references can be found in the former. A histogram of the waiting times can be seen in Figure 4, which clearly suggests a bimodal shape.

We assume the following linear model for these data:

$$\frac{y_i - \beta_0 - \beta_1(x_i - \bar{x})}{\sigma} \mid \alpha, \lambda, \beta_0, \beta_1, \sigma \sim SBN(\alpha, \lambda), \tag{10}$$

i.e., the likelihood factor from the ith observation is a location-scale transformation of the $SBN(\alpha, \lambda)$, with scale parameter σ and location $\beta_0 + \beta_1(x_i - \bar{x})$, where $\bar{x} = n^{-1}\sum_{j=1}^{n} x_j$. Thus, the effect of eruption times is taken as deviations from its empirical mean.

As a preliminary analysis we ignore the information on eruption durations, and consider only the waiting times y_i, i.e., we assume the special case $P(\beta_1 = 0) = 1$ in (10). This is equivalent to the model (8). We use vague prior distributions and assume $p_{0,l} = 0.3$ and $\sigma_l = 10$ in (6), thus not enforcing bimodality in the prior. Also, we take

$p_{0,a} = 0.3$ and $\sigma_a = 5$ in (7), and further assume

$$\beta_0 \sim N(0, 10^4), \quad \text{and} \quad \sigma \sim U(0, 10^4).$$

Posterior computation was based on a random walk-type of Metropolis-Hastings algorithm, as described in. e.g. Chen, Shao and Ibrahim (2000), using normal or truncated normal proposal distributions whenever appropriate.

The upper portion of Table 1 presents some posterior summaries for individual parameters. Figure 3 shows the marginal posterior densities for all parameters in the model. We note that $p(\alpha \mid \boldsymbol{y})$ is mostly concentrated away from 0, and that $p(0 < \lambda < 0.5 \mid \boldsymbol{y})$ is nearly one, where $\boldsymbol{y}$ denotes the entire collection of observed responses. These combined facts suggest the presence of a bimodal and skewed to the right distribution. In fact, the estimated posterior probabilities $p(\alpha = 0 \mid \boldsymbol{y})$ and $p(\lambda = 0 \mid \boldsymbol{y})$ are both practically equal to 0, which was also the case even when changing $p_{0,l}$ and $p_{0,a}$ to be both 0.5. This is also corroborated by the posterior predictive density presented in Figure 4, which is overlaid on top of the histogram of $\boldsymbol{y}$ responses (solid line). Some additional numerical experimentation showed that we need to push $p_{0,l}$ and $p_{0,a}$ to be considerably high (about 0.9999 and 0.8, respectively) to observe non-negligible values for the corresponding posterior probabilities. This reveals the strong support for bimodality and skewness from the data, which would be overwhelmed only in the case of extremely informative priors. As a comparison, we fitted a 5-parameter mixture of normals to the same data, with results presented in Figure 4 (dashed line). This time the pseudo-Bayes factor for model (8) versus the competing mixture (9) was 0.1977, revealing actually some support for the alternative mixture. This is confirmed by the BIC values -2097.154 for model (8) versus -2090.427 for the mixture. As a further comparison we used the estimation priors for model (5), obtaining again similar results.

Next we turn to the regression problem. We use the same prior assumptions as in the case of no covariates, but choosing this time

$$\begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix} \sim N_2\left[\begin{pmatrix}0 \\ 0\end{pmatrix}, 10^4\begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}\right],$$

where the prior independence on β_0 and β_1 is justified by the centering of the eruption durations in (10). Posterior summaries of all parameters are presented in the lower part of Table 1. Marginal posterior distributions for α, λ, σ and contour plots of the joint distribution of β_0 and β_1 are given in Figure 5. Compared to the no covariates case, the posterior distribution of α is now concentrated on a region slightly closer to 0, and that of λ shifted notoriously to the right, which suggests unimodality by the argument given in the discussion following (5). This is confirmed by looking at the posterior predictive distribution at various choices of covariate values (not shown). Also, the joint marginal posterior distribution of the regression coefficients suggests independence in the regression parameters (estimated posterior correlation is -0.022), which can be seen from the contour plots in Figure 5.

The analysis suggests that bimodality is adequate for a model of the waiting times response ignoring covariates. When bringing the covariate information into the model, the bimodality no longer appears to be necessary. However, the unimodality a posteriori in this example was controlled by large values of λ. To investigate this issue further, we considered again the same model but imposing a small prior variance on λ. Taking $\sigma_l = 1$ we found that $p(\alpha = 0 \mid \boldsymbol{y}) = 0.413$ and $p(\lambda = 0 \mid \boldsymbol{y}) = 0.066$. The Bayes factor in favor of unimodality thus becomes 1.642 and 0.165 for symmetry, suggesting that unimodality and skewness are both indeed supported by the data.

5 Discussion

We have proposed a new class of skewed and bimodal distributions that contains the entire classes of symmetric, asymmetric and unimodal distributions, also including the case of bimodality. We have studied some of its basic properties and discussed the application of a special case to the Old Faithful data. We concluded that obtaining bimodal or unimodal shapes depended on whether the covariate information was ignored or not. In that sense, it was useful to have prior distributions that allow assessing specific values of parameters, e.g. $\alpha = 0$ or $\lambda = 0$, which correspond to imposing unimodality and symmetry, respectively.

We have considered explicit comparisons with a model consisting of the mixture of two normals. Our experience suggests that the mixture may perform better or worse than our proposed model, depending on specific features of the data. For instance, in the simulated data, one of the components is actually non-normal, and our model was seen to outperform the mixture of two normals. In any case, the bimodal model has the conceptual advantage of being able to yield exact unimodality for ranges of the α and λ parameters.

Acknowledgments

We thank the Editor, Associate Editor and Referees for their valuable comments that helped to improve the original version of this manuscript.

References

Arellano-Valle, R. B., Gómez, H. W. and Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions, *Journal of Statistical Planning and Inference* **128**(2): 427–443.

Arellano-Valle, R., Cortés, M. A. and Gómez, H. (2006). Uni-bimodal flexibility based on the epsilon-skew-normal distribution. Submitted.

Arnold, B. C. and Beaver, R. J. (2000a). The skew-Cauchy distribution, *Statistics & Probability Letters* **49**(3): 285–290.

Arnold, B. C. and Beaver, R. J. (2000b). Some skewed multivariate distributions, *American Journal of Mathematical and Management Sciences* **20**(1-2): 27–38. Statistical inference, combinatorics and related areas (Varanasi, 1997).

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2002). Conditionally specified multivariate skewed distributions, *Sankhyā. The Indian Journal of Statistics. Series A* **64**(2): 206–226. Selected articles from San Antonio Conference in honour of C. R. Rao (San Antonio, TX, 2000).

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

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones, *Statistica* **46**(2): 199–208.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics. Theory and Applications* **32**(2): 159–200. With discussion by Marc G. Genton and a rejoinder by the author.

Azzalini, A. and Bowman, A. W. (1990). A look at some data on the Old Faithful geyser, *Applied Statistics* **39**: 357–365.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **65**(2): 367–389.

Balakrishnan, N. and Ambagaspitiya, R. (1994). On skew-Laplace distributions, *Technical report*, Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada.

Chen, M.-H., Shao, Q.-M. and Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian computation*, Springer Series in Statistics, Springer-Verlag, New York.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness, *J. Amer. Statist. Assoc.* **93**(441): 359–371.

Ferreira, J. T. A. S. and Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions, *J. Amer. Statist. Assoc.* **101**(474): 823–829.

Gómez, H., Venegas, O. and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution, *Environmetrics* **18**(4): 395–407.

Gupta, A. K., Chang, F. C. and Huang, W. J. (2002). Some skew-symmetric models, *Random Operators and Stochastic Equations* **10**(2): 133–140.

Härdle, W. (1991). *Smoothing techniques*, Springer Series in Statistics, Springer-Verlag, New York. With implementation in S.

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

Ma, Y. and Genton, M. G. (2004). Flexible class of skew-symmetric distributions, *Scandinavian Journal of Statistics. Theory and Applications* **31**(3): 459–468.

Mengersen, K. L. and Robert, C. P. (1996). Testing for mixtures: a Bayesian entropic approach, *in* J. Berger, J. Bernardo, A. Dawid, D. Lindley and A. Smith (eds), *Bayesian statistics, 5 (Alicante, 1994)*, Oxford Sci. Publ., Oxford Univ. Press, New York, pp. 255–276.

Nadarajah, S. and Kotz, S. (2003). Skewed distributions generated by the normal kernel, *Statist. Probab. Lett.* **65**(3): 269–277.

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

Parameter	Mean	Standard Deviation	95% HPD
α	7.21	2.064	(3.934,11.810)
β_0	66.63	0.528	(65.568,67.649)
λ	0.24	0.057	(0.128,0.347)
σ	8.60	0.266	(8.103,9.164)
α	4.90	2.543	(1.293,11.175)
β_0	59.19	0.917	(57.688,61.388)
β_1	10.80	0.255	(10.29,11.279)
λ	10.58	5.442	(3.421,24.157)
σ	8.15	0.350	(7.424,8.819)

Table 1: Posterior summaries for density estimation of waiting times (upper part) and for linear regression model using eruption durations as a covariate (lower part).

Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models, *The Canadian Journal of Statistics. La Revue Canadienne de Statistique* **31**(2): 129–150.

Schwarz, G. (1978). Estimating the dimension of a model, *Ann. Statist.* **6**(2): 461–464.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). *WinBUGS 1.4 Manual.*

Venables, W. N. and Ripley, B. D. (1999). *Modern applied statistics with S-Plus, Third Edition*, Springer-Verlag, New York.

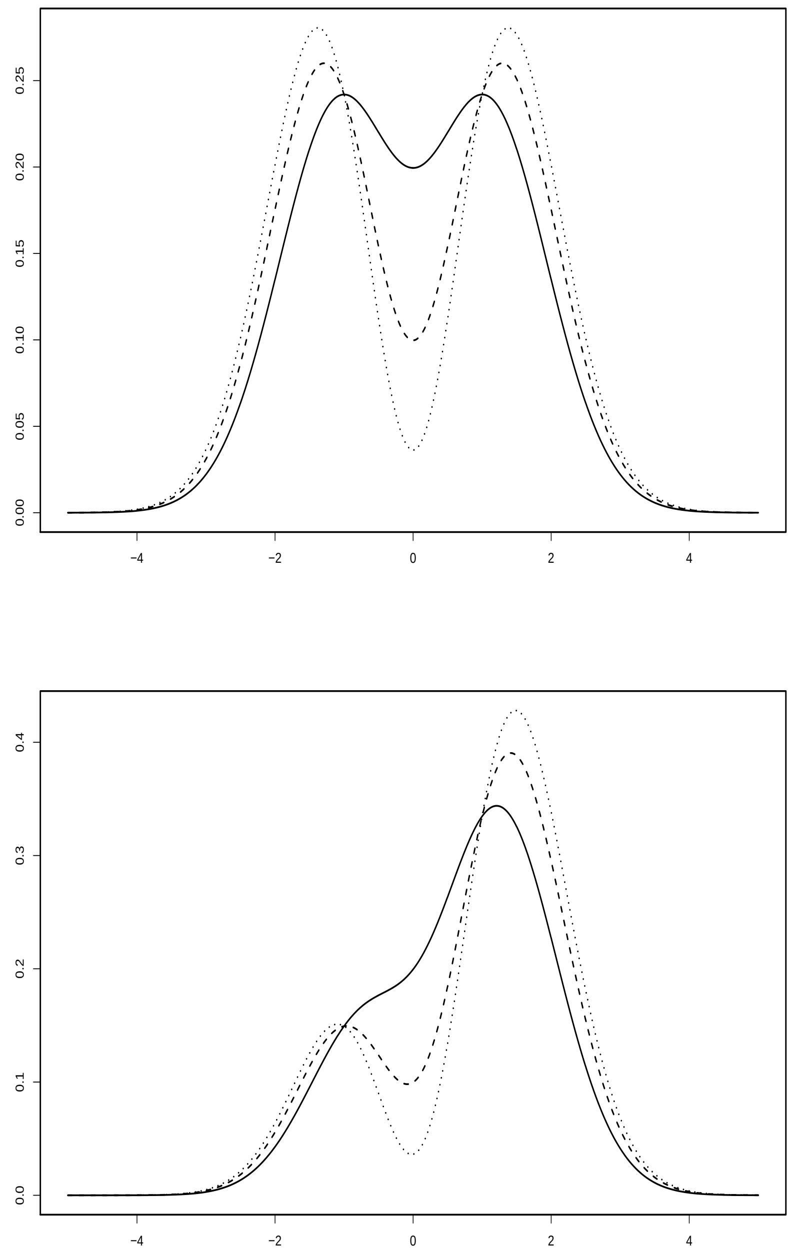

Figure 1: *The upper panel represents plots of the $B(f_0, \alpha)$ density when f_0 is the standard normal pdf ϕ and $\alpha = 1$ (solid line), $\alpha = 3$ (dashed line) and $\alpha = 10$ (dotted line). The bottom panel plots the $SB(f_0, h, w, \alpha)$ pdf for $f_0 = h = \phi$, $w(x) = x/2$ and $\alpha = 1$ (solid line), $\alpha = 3$ (dashed line) and $\alpha = 10$ (dotted line).*

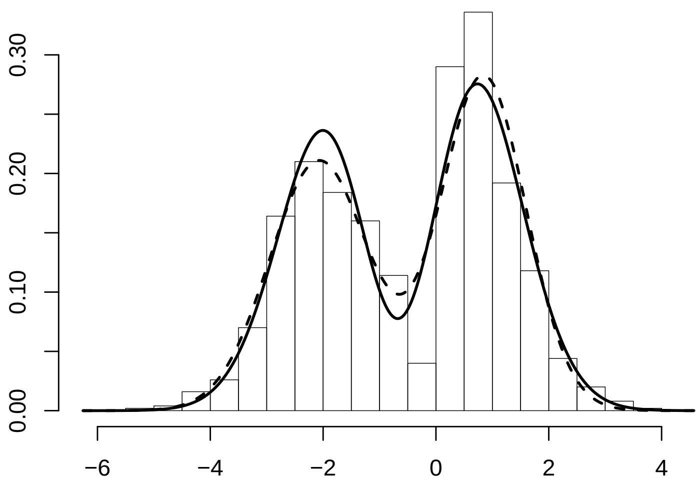

Figure 2: *Histogram of simulated data overlaid with posterior predictive densities for location-scale bimodal skew-normal model (solid line) and mixture of two normals model (dashed line)*

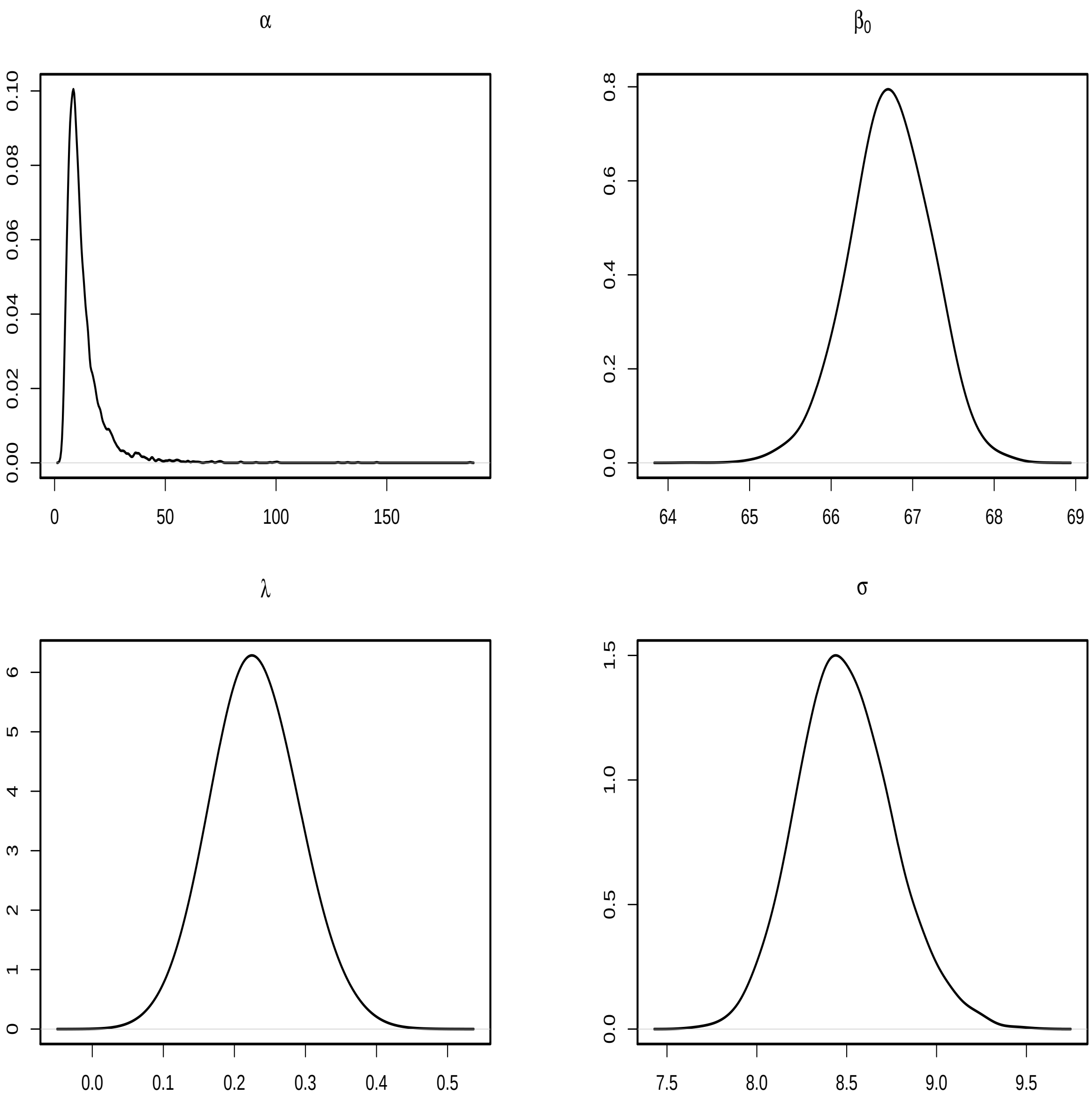

Figure 3: Posterior marginal densities for all parameters in density estimation problem.

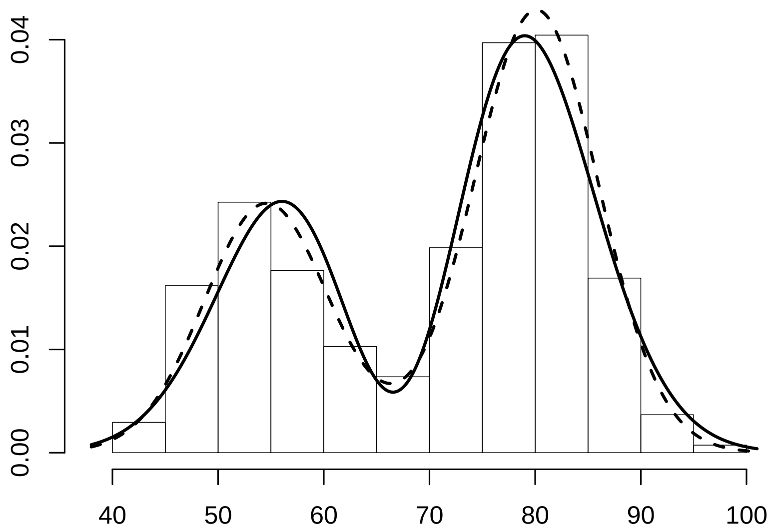

Figure 4: Histogram of waiting times until the next eruption (from the Old Faithful data) overlaid with posterior predictive densities for location-scale bimodal skew-normal model (solid line) and mixture of two normals model (dashed line).

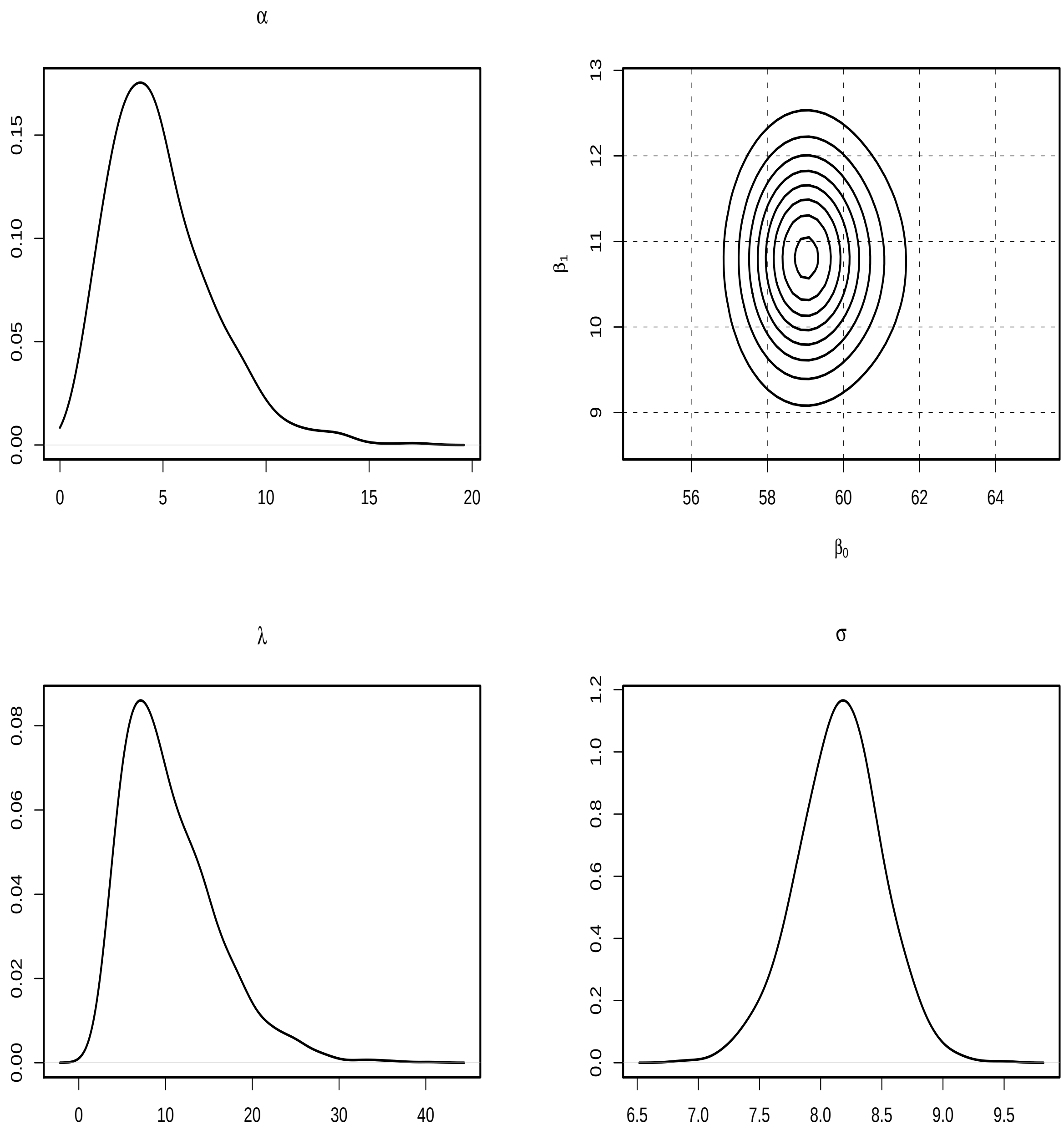

Figure 5: Posterior marginal densities and a bivariate contour plot for the linear regression problem.