Inference for grouped data with a truncated skew-Laplace distribution

F. J. Rubio* and M. F. J. Steel†

Abstract

The skew-Laplace distribution has been used for modelling particle size with point observations. In reality, the observations are truncated and grouped (rounded). This must be formally taken into account for accurate modelling, and it is shown how this leads to convenient closed-form expressions for the likelihood in this model. In a Bayesian framework, we specify "noninformative" benchmark priors which only require the choice of a single scalar prior hyperparameter. We derive conditions for the existence of the posterior distribution when rounding and various forms of truncation are considered in the model. We will focus mostly on modelling microbiological data obtained with flow cytometry using a skew-Laplace distribution. However, we also use the model on data often used to illustrate other skewed distributions, and we show that our modelling favourably compares with the popular and flexible skew-Student models. Further examples on simulated data illustrate the wide applicability of the model.

Key Words: Bayesian inference; flow cytometry data; glass fibre data; posterior existence; rounding

1 Introduction

We propose a truncated skew-Laplace distribution for use with coarse (in particular rounded) or set observations. Bayesian inference will be conducted using Markov chain

*UNIVERSITY OF WARWICK, DEPARTMENT OF STATISTICS, COVENTRY, CV4 7AL, UK. E-mail: F.J.Rubio@warwick.ac.uk

†UNIVERSITY OF WARWICK, DEPARTMENT OF STATISTICS, COVENTRY, CV4 7AL, UK. E-mail: M.F.Steel@stats.warwick.ac.uk]

Monte Carlo methods. Our leading example concerns microbiological data obtained with flow cytometry, in particular forward scatter (FS) data obtained for the Escherichia Coli (E. Coli) bacterium. Julià and Vives-Rego (2005, 2008) use a skew-Laplace distribution to model these data, which are truncated due to the sensitivity of the flow cytometer and are recorded as set data because the observations are presented as integers. Truncation and coarsening must be formally included in the model in order to conduct inference appropriately and to fit the data well. This application will be used throughout most of the paper, and will serve as an important motivating example. However, later we will use the same model for a data set on the breaking strength of glass fibres, which has frequently been used in the statistics literature for illustrating skewed distributions. Further examples on simulated data illustrate the general applicability of the model.

In order to define the skew-Laplace distribution, we use the general skewing framework of Fernández and Steel (1998a). This leads to a skew-Laplace distribution which is parameterised through a single skewness parameter. This skewness parameter has a nice interpretation in terms of the allocation of mass to the left and to the right of the mode. It also leads to inferential advantages as the skewing and scale parameters have clearly defined roles, which *e.g.* facilitates specification of the prior distribution.

Despite the introduction of skewness, rounding and truncation in the model, the likelihood has a relatively simple closed-form expression. This makes efficient likelihood-based inference feasible, and in this paper we will focus on Bayesian inference. Maximum likelihood estimates, profile likelihoods and confidence intervals are numerically very close to posterior modes, posterior density functions and Highest Posterior Density (HPD) credible intervals, respectively. For models with various degrees of truncation, we propose benchmark "non-informative" priors which require the choice of a scalar prior hyperparameter. As these priors are improper, we also derive sufficient conditions for the existence of the posterior. These conditions are quite mild and trivial to check. An important advantage of the Bayesian framework is that it naturally leads to formal model comparison on the basis of Bayes factors. We compute Bayes factors between the various models as a function of the single prior hyperparameter and also consider comparison based on predictive performance. For the glass fibre data, we compare the skew-Laplace model with commonly used skew-Student specifications and find the former does better in terms of Bayes factors and matches the best skew-Student model in terms of predictive performance. Inference with the skew-Laplace model is not substantially complicated by the use of set observations or truncation of the sample space, in contrast with skew-Student or skew-normal models, for which the likelihood is not available in closed form.

2 Set observations

Whenever we use a continuous model for the observations, the actually recorded values are necessarily rounded, as they are recorded to some finite precision. There has been an active literature on the quantitative effects of rounding (or grouping), as summarized in e.g. Heitjan (1989) and more recently in Schneeweiss et al. (2010). Within a Bayesian context, the explicit modelling of grouped data or set observations has been proposed by Fernández and Steel (1998b, 1999a) as a way to avoid pathological situations such as the nonexistence of a posterior with a proper prior. The reason for such behaviour is linked to the fact that any set of point observations has zero probability under a continuous sampling model. Set observations in our context of rounding are simple neighbourhoods (intervals of positive Lebesgue measure) of the recorded point observations that are chosen in accordance with the precision of the measuring process. Thus, for $i = 1, \ldots, n$ and some $d > 0$, we define

$$\mathbb{P}[\text{observing } y_j] = \mathbb{P}\left[y_j \in S_j\right] = \mathbb{P}\left[y_j - d < Y < y_j + d\right]. \tag{1}$$

3 The skew-Laplace distribution

In order to define the skew-Laplace distribution we use the skewness mechanism proposed in Fernández and Steel (1998a). Thus, we say that $X \sim$ skew-Laplace(μ, σ, γ) if the density function of X is

$$f_X(x|\mu,\sigma,\gamma) = \begin{cases} \frac{1}{\sigma\left(\gamma+\frac{1}{\gamma}\right)} \exp\left[\frac{\gamma(x-\mu)}{\sigma}\right] & \text{for } x < \mu\,, \\ \frac{1}{\sigma\left(\gamma+\frac{1}{\gamma}\right)} \exp\left(\frac{\mu-x}{\gamma\sigma}\right) & \text{for } x \geq \mu\,, \end{cases} \tag{2}$$

where $\mu \in \mathbb{R}$, $\sigma, \gamma > 0$. This model (with a different, less interpretable parameterisation, used in Julià and Vives-Rego, 2005, 2008) was called the two-piece double exponential distribution in Lingappaiah (1988). The allocation of mass to each side of the mode is given by

$$\frac{1 - F_X(\mu|\mu,\sigma,\gamma)}{F_X(\mu|\mu,\sigma,\gamma)} = \gamma^2,$$

which clearly highlights the role of γ as the skewness parameter, with μ the location parameter (which is always the mode) and σ a scale parameter. Of course, for $\gamma = 1$ we obtain the usual Laplace distribution, whereas right (positive) skewness corresponds to $\gamma > 1$ and left (negative) skewness to $\gamma < 1$. Inverting γ corresponds to mirroring the density function around the mode. If we measure skewness by the usual third centered moment divided by the cubed standard deviation, the difference between mean and mode divided by the standard deviation or the measure in Arnold and Groeneveld (1995) (defined as one minus twice the probability mass to the left of the mode), then γ and $1/\gamma$ lead to equal amounts of skewness with opposing signs. All these measures are strictly increasing functions of the skewness parameter γ.

The distribution function of X is given by

$$F_X(x|\mu,\sigma,\gamma) = \begin{cases} \frac{1}{1+\gamma^2}\exp\left(\frac{\gamma(x-\mu)}{\sigma}\right) & \text{for } x < \mu\,, \\ \frac{1}{1+\gamma^2}\left[1-\gamma^2\left(\exp\left(\frac{\mu-x}{\gamma\sigma}\right)-1\right)\right] & \text{for } x \geq \mu\,. \end{cases}$$

First we investigate the analysis with the skew-Laplace distribution in (2), taking into account the fact that the actual observations are rounded as described in Section 2.

3.1 Likelihood Function

Consider an independent sample of rounded observations $y_1, ..., y_n$ from (2). The rounding as in (1) implies that

$$\begin{aligned} \mathbb{P}[\text{observing } y_j] &= \mathbb{P}\left[y_j - d < Y < y_j + d\right] \\ &= F_X(y_j + d|\mu,\sigma,\gamma) - F_X(y_j - d|\mu,\sigma,\gamma). \end{aligned} \tag{3}$$

Suppose that there are k different observations $y^* = \{y_1^*, ..., y_k^*\}$ and $\{n_1, ..., n_k\}$ are the corresponding observed frequencies. The likelihood function for this sample is

$$\mathcal{L}(y|\mu,\sigma,\gamma) \propto \prod_{j=1}^{k}\left[F_X(y_j^* + d|\mu,\sigma,\gamma) - F_X(y_j^* - d|\mu,\sigma,\gamma)\right]^{n_j}.$$

The E.Coli dataset contains $n = 9,015$ observations, rounded to $k = 98$ integer values (so that $d = 1/2$), ranging from 47 to 165 with frequencies in between 1 and 306. The glass fibre data have $n = 63$ observations, rounded to the nearest one hundredth ($d = 0.005$), ranging from 0.55 to 2.24 with 49 repeated observations.

3.2 Bayesian Inference

In order to come up with a reasonable "noninformative" prior of the parameters in our model (2), we first consider the fact that the three parameters have clearly distinct roles, so that a product structure for the prior seems a good choice. In the symmetric model (*i.e.* $\gamma = 1$) the (noninformative) full Jeffreys prior is given by $p(\mu, \sigma) \propto \sigma^{-2}$, as is the case for any location-scale model (Fernández and Steel, 1999b). We then modify this prior by bounding the parameter space of the location μ, which is important in ensuring that a posterior distribution exists (*i.e.* is a well-defined probability distribution). As we are dealing with necessarily positive observations with an internal mode in both of our applications, we use zero as a lower bound for the mode μ, whereas we introduce a single hyperparameter M as the upper bound. To elicit a prior for the skewness parameter γ, we consider the skewness measure of Arnold and Groeneveld (1995), which takes values in the interval (-1,1) and specify a uniform prior on this measure. This leads to the following prior for the model parameters:

$$\pi(\mu, \sigma, \gamma) \propto \frac{\gamma}{\sigma^2 \left(1 + \gamma^2\right)^2} I(0 < \mu \leq M). \tag{4}$$

Note that this density is improper in σ and the prior mass assigned to a range of positive skewness (say, $\gamma \in (a, b)$ with $b > a > 1$) is the same as that assigned to the corresponding range of negative skewness ($\gamma \in (1/b, 1/a)$). We take the upper bound M to be 1000 in the results presented in Sections 3-5.

We obtain the following sufficient condition for the existence of the posterior distribution.

Theorem 1 *The posterior distribution of* (μ, σ, γ) *for the model* (2) *and the prior distribution* (4) *is proper if the number of different observations is at least* 3*, i.e.* $k \geq 3$.

Proof. *see Appendix*

Inference for the E.Coli data was conducted using a Markov chain Monte Carlo (MCMC) algorithm. In particular, we simulated a chain of length $2,510,000$ from the posterior using the t-walk algorithm (Christen and Fox, 2010) and after a burn-in of $10,000$ we retained every 100th set of parameter values, leading to sample of 25,000 draws. Figure 1 shows the marginal posterior distributions of (μ, σ, γ). Inference is quite precise with 95% Highest Posterior Density (HPD) credible intervals given as follows: μ: $(69.75, 70.93)$, γ: $(1.03, 1.10)$ and σ: $(10.29, 10.73)$. It is clear that the relatively large dataset contains quite a lot of information on the three parameters in our model. The evidence indicates a relatively small but quite precisely determined amount of right skewness in the data. Prior density functions are also displayed in Figure 1, but they are virtually flat for the range of the parameter values shown (prior density values are quite small, so the prior for μ and γ is scaled up by the most convenient power of ten; for σ an arbitrary scaling is applied). Figure 2 shows the predictive distribution of the data (the sampling density in (2) with the parameters integrated out with the posterior distribution). However, comparing the data histogram with this predictive density indicates a rather poor fit of the data. For example, it seems that the slightly positive skewness is not consistent with the perhaps more pronounced left "shoulder" in the data when we limit ourselves to the range where data were actually observed. On the other hand, the far left tail of the predictive density is simply not matched by any data. Thus, it appears truncation of the data is an issue and we will now use a model that allows us to formally accommodate such truncation.

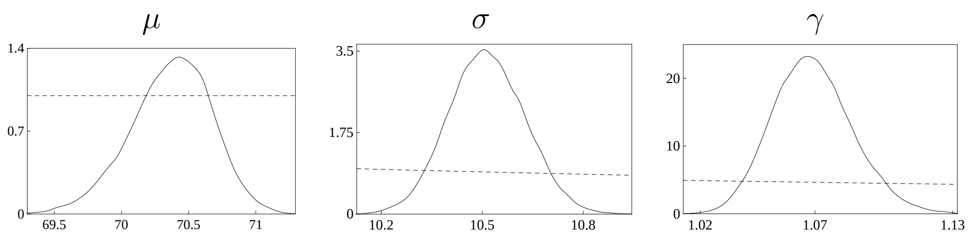

Figure 1: E. Coli data: Posterior (solid line) and scaled prior (dashed line) density functions.

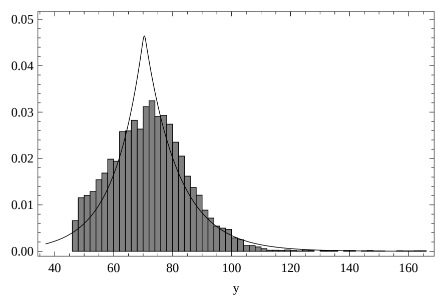

Figure 2: Histogram of E. Coli data and predictive density.

4 Doubly Truncated Model

Let us consider Y to be a version of the skew-Laplace distributed random variable X in (2), truncated to the interval $[\theta_1, \theta_2]$. The density function of Y is then

$$f_Y(y|\mu, \sigma, \gamma, \theta_1, \theta_2) = \frac{f_X(y|\mu, \sigma, \gamma) I_{[\theta_1, \theta_2]}(y)}{F_X(\theta_2|\mu, \sigma, \gamma) - F_X(\theta_1|\mu, \sigma, \gamma)}, \tag{5}$$

where $\theta_1, \theta_2 \in \mathbb{R}$ and $\theta_1 < \mu < \theta_2$. Note that μ is still a location parameter (the mode), σ is a scale parameter, γ is a skewness parameter and (θ_1, θ_2) are threshold or boundary parameters. The allocation of mass to each side of the mode is given by

$$\frac{1-F_Y(\mu|\mu,\sigma,\gamma,\theta_1,\theta_2)}{F_Y(\mu|\mu,\sigma,\gamma,\theta_1,\theta_2)} = \gamma^2 \frac{1-\exp\left(\frac{\mu-\theta_2}{\gamma\sigma}\right)}{1-\exp\left(\frac{\gamma(\theta_1-\mu)}{\sigma}\right)},$$

where F_Y is the distribution function of Y and is given by

$$F_Y(y|\mu,\sigma,\gamma,\theta_1,\theta_2) = \begin{cases} 0, & \text{for } y < \theta_1, \\ \frac{F_X(y|\mu,\sigma,\gamma)-F_X(\theta_1|\mu,\sigma,\gamma)}{F_X(\theta_2|\mu,\sigma,\gamma)-F_X(\theta_1|\mu,\sigma,\gamma)}, & \text{for } \theta_1 \leq y \leq \theta_2, \\ 1, & \text{for } y > \theta_2. \end{cases}$$

So the mass allocation both sides of the mode in this doubly truncated model is affected by γ as before but also by the boundary parameters. Of course, if $\theta_1 \to -\infty$ and $\theta_2 \to \infty$ we retrieve the previous model in the limit, but we will assume finite values for θ_1 and θ_2 in this section.

4.1 The likelihood function

An independent sample $y_1, ..., y_n$ from (5) rounded as in (1) leads to

$$\begin{aligned} \mathbb{P}[\text{observing } y_j] &= \mathbb{P}\left[y_j - d < Y < y_j + d\right] \\ &= F_Y(y_j + d|\mu,\sigma,\gamma,\theta_1,\theta_2) - F_Y(y_j - d|\mu,\sigma,\gamma,\theta_1,\theta_2). \end{aligned}$$

As before, we suppose that there are k different observations $y_1^*, ..., y_k^*$, of which the smallest is $y_{(1)}$ and the largest is $y_{(n)}$, and $n_1, ..., n_k$ are the corresponding observed frequencies. The likelihood function for this sample is

$$\begin{aligned} \mathcal{L}(y|\mu,\sigma,\gamma,\theta_1,\theta_2) &\propto \prod_{j=1}^{k} \left[F_Y(y_j^* + d|\mu,\sigma,\gamma,\theta_1,\theta_2) - F_Y(y_j^* - d|\mu,\sigma,\gamma,\theta_1,\theta_2)\right]^{n_j} \\ &= \left[F_X(\theta_2|\mu,\sigma,\gamma) - F_X(\theta_1|\mu,\sigma,\gamma)\right]^{-n} \\ &\times I_{\left(-\infty, y_{(1)}-d\right]}(\theta_1) I_{\left[y_{(n)}+d, \infty\right)}(\theta_2) \\ &\times \prod_{j=1}^{k} \left[F_X(y_j^* + d|\mu,\sigma,\gamma) - F_X(y_j^* - d|\mu,\sigma,\gamma)\right]^{n_j}. \end{aligned}$$

4.2 Bayesian Inference

Consider the following improper prior for the parameters of the sampling model in (5)

$$\pi(\mu,\sigma,\gamma,\theta_1,\theta_2) \propto \frac{\gamma}{\sigma^2\left(1+\gamma^2\right)^2} I(0 < \theta_1 < \mu < \theta_2 < M), \tag{6}$$

which is in line with the prior (4) used for the untruncated model, and is again improper only in σ. Note that the prior assumes that the mode is contained within the range of observed data. This may not always seem like a reasonable assumption, but we feel that the use of a skew-Laplace model would not be natural if we were faced with data that look like one tail of such a model (we would then simply use a version of an exponential model).

The existence of the posterior is warranted by the following result:

Theorem 2 *The posterior distribution of* $(\mu, \sigma, \gamma, \theta_1, \theta_2)$ *for the Bayesian model in* (5) *and* (6) *is proper if the number of different observations is at least* 4*, i.e.* $k \geq 4$*.*

Proof. *See Appendix.*

We have used the same value of M and the same MCMC algorithm (with the same runlength) as in Section 3. Figure 3 shows the marginal prior (scaled as before) and posterior distributions for the E. Coli data. It is interesting to note the dramatically

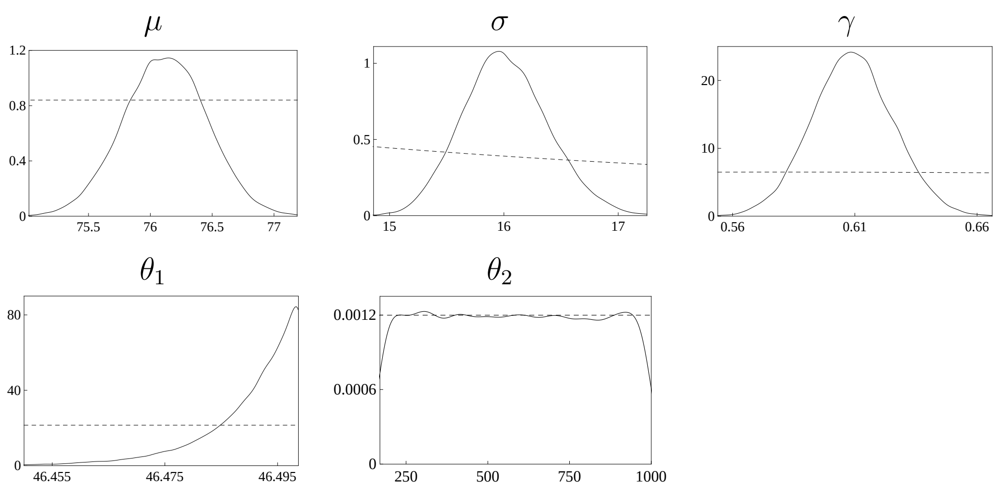

Figure 3: E. Coli data: Posterior (solid line) and scaled prior (dashed line) density functions.

different inference on the skewness parameter γ in this truncated model. As the data truncation is now being dealt with by the boundary parameters, we no longer need γ to reduce the mass in the left tail, and we get evidence for strong negative skewness instead, which is much more in line with the data histogram. The posterior distribution of θ_2 is flat (like the prior) over the range $(y_n + 1/2, M)$ indicating that the data carry no information about θ_2 within this range. This is in line with the classical analysis, where the profile likelihood of θ_2 has an asymptote of ≈ 0.7 times the maximum value for large θ_2. There is no real data evidence to distinguish between values of θ_2 above $y_n + 1/2$ and this suggests the use of a model with only left truncation. Estimated 95% HPD credibility intervals for the other parameters are μ: $(75.44, 76.77)$, γ: $(0.57, 0.64)$, σ: $(15.28, 16.75)$ and θ_1: $(46.47, 46.50)$.

Figure 4 shows the predictive density fit to the data, which is clearly much improved because of the truncation.

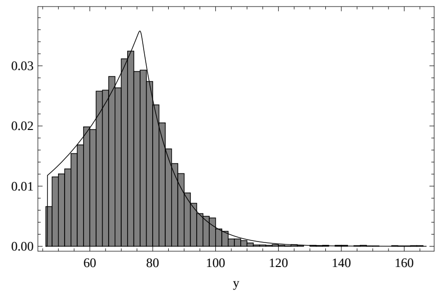

Figure 4: Histogram of E. Coli data and predictive density.

5 Left truncated model

As the particular data used here seem to indicate that truncation on the right is superfluous, we now consider a model with only left truncation. So, let Y be a truncated version of X in $[\theta_1, \infty)$. The density function of Y is

$$f_Y(y|\mu,\sigma,\gamma,\theta_1) = \frac{f_X(y|\mu,\sigma,\gamma) I_{[\theta_1,\infty)}(y)}{1 - F_X(\theta_1|\mu,\sigma,\gamma)}. \tag{7}$$

Now $\theta_1 \in \mathbb{R}$ is the only threshold parameter and we restrict $\theta_1 < \mu$. The allocation of mass to each side of the mode is given by

$$\frac{1 - F_Y(\mu|\mu,\sigma,\gamma,\theta_1)}{F_Y(\mu|\mu,\sigma,\gamma,\theta_1)} = \gamma^2 \frac{1}{1 - \exp\left(\frac{\gamma(\theta_1-\mu)}{\sigma}\right)},$$

where F_Y is the distribution function of Y and is given by

$$F_Y(y|\mu,\sigma,\gamma,\theta_1) \;=\; \begin{cases} 0 & \text{for } \; y < \theta_1, \\ \frac{F_X(y|\mu,\sigma,\gamma) - F_X(\theta_1|\mu,\sigma,\gamma)}{1 - F_X(\theta_1|\mu,\sigma,\gamma)}, & \text{for } \; \theta_1 \leq y. \end{cases}$$

5.1 The likelihood function

Consider an independent sample $y_1, ..., y_n$ from (7) with rounding as in (1).

The likelihood function for a sample of k different observations $y_1^*, ..., y_k^*$ with frequencies $n_1, ..., n_k$ is given by

$$\begin{aligned}\mathcal{L}(y^*|\mu,\sigma,\gamma,\theta_1) &\propto \prod_{j=1}^{k}\left[F_Y(y_j^*+d|\mu,\sigma,\gamma,\theta_1)-F_Y(y_j^*-d|\mu,\sigma,\gamma,\theta_1)\right]^{n_j}\\ &= \left[1-F_X\left(\theta_1|\mu,\sigma,\gamma\right)\right]^{-n} I_{\left(-\infty,y_{(1)}-d\right]}(\theta_1)\\ &\times \prod_{j=1}^{k}\left[F_X(y_j^*+d|\mu,\sigma,\gamma)-F_X(y_j^*-d|\mu,\sigma,\gamma)\right]^{n_j}.\end{aligned}$$

5.2 Bayesian Inference

Consider the following improper prior for the parameters of the model (5)

$$\pi(\mu,\sigma,\gamma,\theta_1) \propto \frac{\gamma}{\sigma^2\left(1+\gamma^2\right)^2} I(0<\theta_1<\mu<M), \tag{8}$$

which is the prior suggested by (6) for this reduced model.

Posterior existence is ensured by the following result:

Theorem 3 *The posterior distribution of* $(\mu,\sigma,\gamma,\theta_1)$ *for the model* (7) *and the prior distribution* (8) *is proper if the number of different observations is at least* 4*, i.e.* $k \geq 4$.

Proof. *See Appendix*

We used the same value for M and the same MCMC strategy to obtain posterior results. As expected, results are very close to the doubly truncated model, except that we do not have the right truncation parameter in the model. Marginal posterior density functions for μ, γ, σ and θ_1 are virtually identical as well as the predictive distribution.

6 Model Comparison

One advantage of Bayesian methods is that model comparison can formally be conducted by Bayes factors. Here Bayes factors can be computed between all three models despite the arbitrary integrating constant (improperness) of the prior, since the prior has a product structure with an improper factor (in σ) which is common to all models, and the factor corresponding to model-specific parameters is integrable and thus properly normalised. The marginal likelihoods needed in the calculation of Bayes factors are estimated using importance sampling, with an importance function chosen to resemble the posterior but with fatter tails. Results with reciprocal importance sampling (Gelfand and Dey, 1994) are very close. Table 2 contains values for the logarithm of the Bayes factors. Information-based criteria are typically a lot easier to compute and we also present values for the BIC (Schwarz, 1978) and the DIC (Deviance Information Criterion) of Spiegelhalter et al. (2002). An alternative approach to model comparison is through the predictive performance of the models; we compute the log predictive score (LPS; see e.g. Gneiting and Raftery, 2007) based on how well the predictive distribution matches a randomly chosen prediction subsample, not used in the posterior inference. We use 20 prediction subsamples of 450 observations each and compute the LPS as the average over the 20 subsamples (smaller values are better).

	Model		
Criterion	untruncated	doubly trunc.	left trunc.
BIC	73020.9	71553.8	71545.4
DIC	72999.6	71516.9	71517.1
log Bayes factor	0	733	732
LPS	1822.1	1785.1	1785.8

Table 1: E. Coli data: Various criteria for model comparison. In the prior for the truncated models we choose $M = 1000$. Bayes factors are computed through importance sampling and we state the logarithm of the Bayes factor in favour of the model in the column versus the untruncated model. Log predictive scores (LPS) are computed on the basis of 20 partitions, each retaining 450 observations in the prediction sample.

From the results in Table 2 we immediately deduce that the truncated models are much preferred to the untruncated version. The relative support for both truncated models is in favour of the left truncated model if we consider BIC. The DIC, LPS and the Bayes factor all favour the doubly truncated model. Only in the case of the Bayes factor can this be interpreted in terms of posterior model probabilities: if we assume unitary prior odds, the posterior probability attached to the doubly truncated model is 2.5 times as large as that of the left truncated model. Clearly, the posterior mass assigned to the untruncated model is negligible.

Finally, we remind ourselves that the prior hyperparameter M must be selected in specifying the prior, and we know that Bayes factors can be quite sensitive to the choice of prior (Kass and Raftery, 1995). Therefore, we now investigate the sensitivity of the Bayes factor to the choice of M. Figure 5 shows how estimates for the marginal likelihoods of all models and the (most relevant) Bayes factor between the truncated models vary with M. For each value of M (in the range from 200, just above the largest observation, to 2000) we run ten importance sampling estimates and the results are indicated through boxplots. Clearly, estimates are quite precise for all three models. As expected, marginal likelihood values are affected by the choice of M, since the prior domain for μ is extended beyond areas with appreciable likelihood values as M grows, so that the only real effect of larger M is that we average the likelihood with smaller prior density values, thus leading to a smaller marginal likelihood. However, the ratio of marginal likelihoods (the Bayes factor) is relatively stable as M varies. As a consequence, we consistently get slightly more support for the doubly truncated model for reasonable values of M, say $M > 300$.

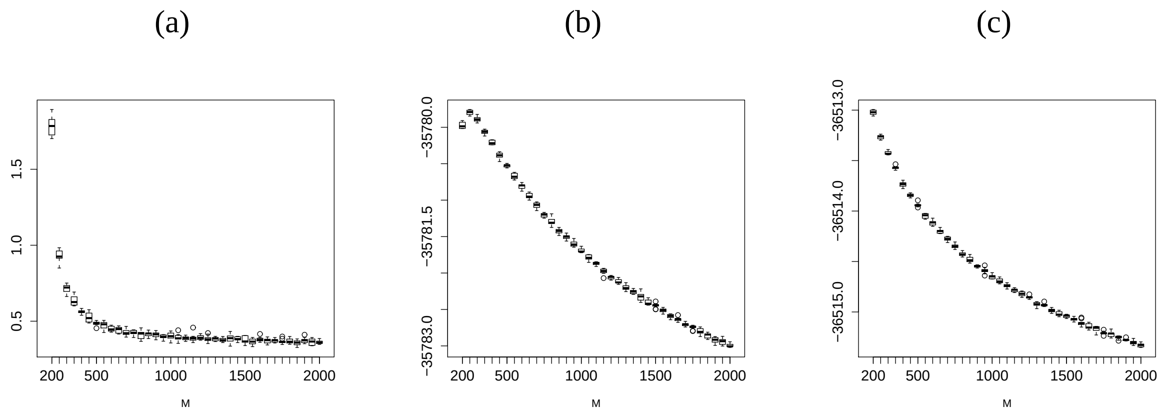

Figure 5: E.Coli data: Box plots based on 10 posterior samples using importance sampling. In all graphs results are given as a function of M. (a) Bayes factor in favour of the left truncated model versus the doubly truncated one. (b) Marginal likelihood for the doubly truncated model. (c) Marginal likelihood for the untruncated model.

7 Glass fibre data

Consider the data reported in Smith and Naylor (1987) about the breaking strength of $n = 63$ glass fibres. These data were used repeatedly in the literature with a variety of skewed distributions (Jones and Faddy, 2003; Ferreira and Steel, 2006). We compare the skew-Laplace model with the more commonly used skew-Student model (with the inverse scale factor skewing of Fernández and Steel, 1998a), on the basis of set observations. This skew-Student sampling model is given by

$$f_t(x|\mu,\sigma,\gamma,\nu) = \begin{cases} \frac{2c_\nu}{\sigma\left(\gamma+\frac{1}{\gamma}\right)}\left[1+\frac{1}{\nu}\left(\frac{\gamma(x-\mu)}{\sigma}\right)^2\right]^{-(\nu+1)/2} & \text{for } x<\mu\,, \\ \frac{2c_\nu}{\sigma\left(\gamma+\frac{1}{\gamma}\right)}\left[1+\frac{1}{\nu}\left(\frac{(x-\mu)}{\gamma\sigma}\right)^2\right]^{-(\nu+1)/2} & \text{for } x\geq\mu\,, \end{cases} \tag{9}$$

where $c_\nu = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]}\sqrt{\frac{1}{\nu\pi}}$ and $\nu > 0$ is the degrees of freedom parameter. Ferreira and Steel (2006) find that the skew-Student with $\nu = 2$ performs well for these data, but we will focus on the skew-Student with unknown degrees of freedom, as this retains the flexibility to adapt the tails to the data.

These data are breaking strengths, and therefore are subject to the physical constraint that they can not be negative. Thus, the first skew-Laplace model we consider is the left truncated one in (7), but with θ_1 fixed to be zero. In combination with the prior for the three model parameters in (4), this leads to the following result:

Theorem 4 *The posterior distribution of* (μ,σ,γ) *for the skew-Laplace model left truncated at zero, i.e.* (7) *with* $\theta_1 = 0$*, and the prior distribution* (4) *is proper if the number of different set-observations is at least* 4*, i.e.* $k \geq 4$*.*

Proof. *See Appendix*

For the skew-Student model in (9) we adopt the prior based on (4) with an extra factor for the degrees of freedom parameter

$$\pi(\mu,\sigma,\gamma,\nu) \quad \propto \quad \frac{\gamma}{\sigma^2(1+\gamma^2)^2} I(0<\mu<M)P_\nu, \tag{10}$$

for which we can derive the following result on posterior existence:

Theorem 5 *The posterior distribution of* (μ,σ,γ,ν) *for the skew-Student model in (9) and the prior distribution* (10) *is proper if the number of different set-observations is at least* 3*, i.e.* $k \geq 3$ *and* P_ν *is a proper distribution with zero mass on* $(-\infty, 1+\epsilon)$ *for any* $\epsilon > 0$*.*

Proof. *See Appendix*

The restriction on the prior support means that we want the predictive mean to exist, which may not be a very unreasonable assumption. Note that very small values of ν are typically associated with problems in classical likelihood inference or Bayesian inference on the basis of point observations (Fernández and Steel, 1999a). Theorem 5 also covers the case where we fix ν at any value larger than or equal to one, simply by taking P_ν to be Dirac. For the prior P_ν in the case of unknown ν we consider two possibilities: firstly, a thin-tailed gamma prior with shape parameter 2 and scale parameter 0.1, restricted to $[1+\epsilon, \infty)$ which covers a large range of values. Secondly, we adopt a hierarchical prior constructed from putting an exponential prior on the scale parameter of the gamma with shape parameter 2; this leads to the gamma-gamma prior, given by $\pi(\nu) \propto \nu/(\nu+d)^3$ with $d > 0$ and defined for $\nu \geq 1+\epsilon$. This prior has a very fat tail with no mean and shares the right-tail behaviour of the Jeffreys prior derived for the symmetric Student-t model in Fonseca at al. (2008). Here we adopt $d = 2$ which means the mode is at the boundary for $\nu = 1+\epsilon$. Throughout, we take ϵ to be machine precision (the results are the same for any $\epsilon \leq 0.0001$).

Figure 6 shows the inference on the parameters of the zero truncated skew-Laplace model using $M = 10$, and it is clear that skewness is again an important aspect of the data. As in other studies with this application, we find clear evidence of negative skewness. Posterior predictive density functions are overplotted with a histogram (chosen

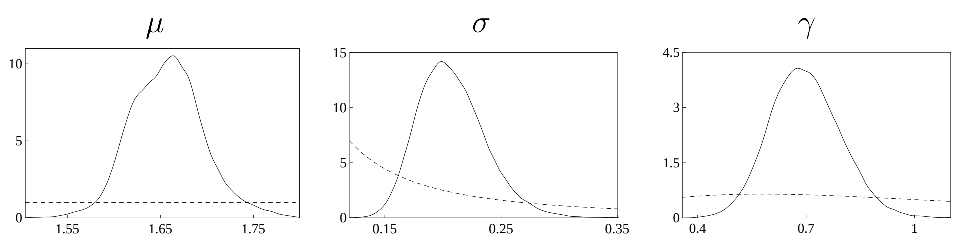

Figure 6: Glass data: Posterior (solid line) and prior (dashed line) density functions for the Laplace model.

according to Sturges' formula) of the data in Figure 7. All models seem to fit the data reasonably well, but there are some differences between the predictives. It is interesting to note that the skew-Laplace model does not lead to such a sharp peak as in the application with the E. Coli data. The fact that the data are not very peaked means there is some posterior uncertainty regarding the mode (see Figure 6), and this is reflected in the posterior predictive (which is simply the sampling model integrated out with the posterior). As a consequence, the skew-Laplace and the skew-Student model with $\nu = 2$ are actually very similar. Thus, the simple skew-Laplace model adapts to the data at hand.

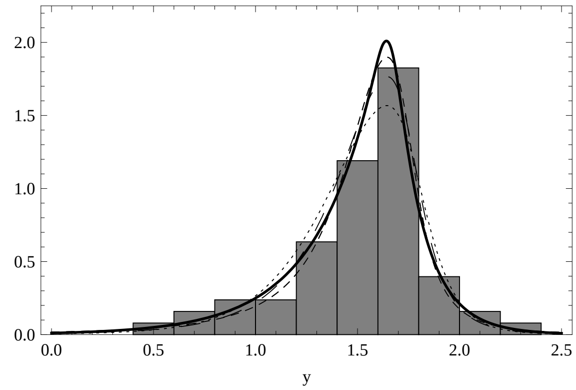

Figure 7: Histogram of glass data, predictive density for the skew-Laplace (bold line), skew-t_2 (short dashes), skew-t_ν with gamma prior (dotted line), skew-t_ν with gamma-gamma prior (long dashes).

7.1 Model comparison

In order to have a more formal comparison of the different models, we can again compute the Bayes factors. Marginal likelihood estimates depend on M, as discussed in Section 6, and this leads to the Bayes factors displayed in Figure 8. These are Bayes factors in favour of the zero truncated skew-Laplace model as a function of M and the boxplots correspond to ten importance sampling estimates. Clearly, the skew-Laplace

(a) (b) (c)

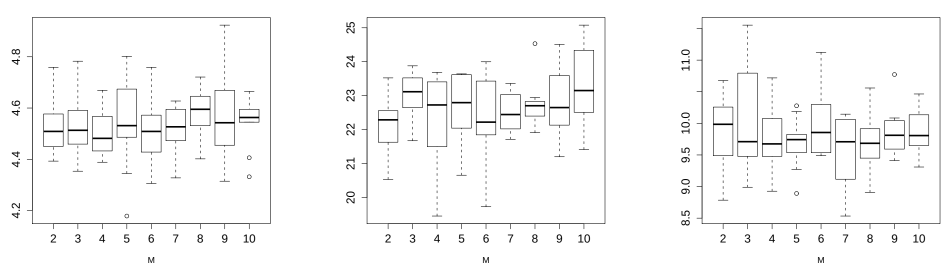

Figure 8: Glass data: Bayes factors as a function of M in favour of the zero truncated skew-Laplace model versus (a) skew-t_2 model; (b) skew-t_ν model with gamma prior; (c) skew-t_ν model with gamma-gamma prior

model beats the skew-Student models. Among the skew-Student models, it seems best to fix ν to be a suitable value for these data, namely $\nu = 2$. The value of M does not seem to have a systematic effect on these Bayes factors. Of course, truncation is not built into the skew-Student models, but this aspect is not that important for the Bayes factors, as the untruncated skew-Laplace model does almost equally well with these data (*e.g.* the Bayes factor is around 1.24 in favour of the zero truncated skew-Laplace model for $M = 10$). Truncation is, however, not that easily implemented in the skew-Student models, both in terms of computational ease and proving results such as Theorem 5.

To assess the impact of the different priors on ν, we overplot posterior and prior density functions for ν in Figure 9. Despite its fatter right tail, the gamma-gamma prior has a mode closer to zero and leads to more posterior mass concentration on small values of ν. Thus, the predictive and the marginal likelihood are closer to that of the case with $\nu = 2$ than with the gamma prior.

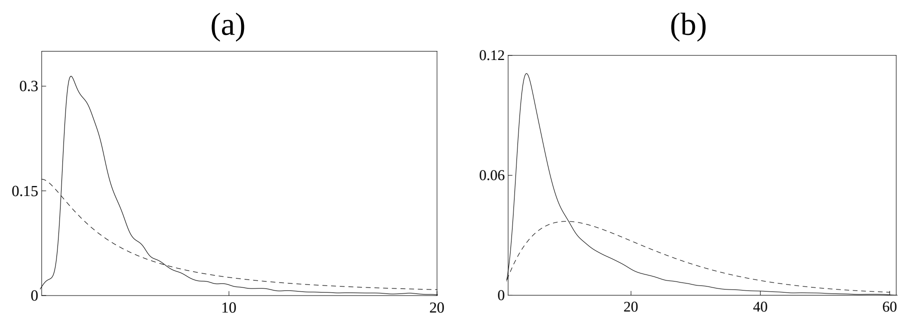

Figure 9: Glass data: degrees of freedom parameter ν for skew-Student (a) Posterior distribution of ν (solid line) and gamma-gamma prior (dashed line). (b) Posterior distribution of ν (solid line) and gamma prior (dashed line).

Model	Skew-Laplace	Skew-t gamma-gamma	Skew-t gamma	Skew-t_2
LPS	95.76	96.28	96.08	95.74

Table 2: Glass data: Log predictive scores (LPS), computed on the basis of 20 partitions, each retaining 20 observations in the prediction sample.

We compare the models in terms of their predictive performance by computing log predictive scores, averaged over 20 partitions of the data where 20 randomly chosen observations are used in the prediction subsample, and the results are presented in Table 2. The skew-Laplace and skew-t_2 models predict best and are roughly equally good.

8 Conclusions

In this paper, we describe inference with the skew-Laplace model, a flexible model for use with unimodal data sets where rounding and truncation of the data are possibly important issues. We formally incorporate rounding of the data and truncation of the support in the analysis. For four versions of the model (untruncated support, finite support with unknown boundaries, left truncated support with unknown boundary, left truncated at zero), we specify a fairly noninformative and sensible prior which only depends on a single hyperparameter M and we derive sufficient conditions for the existence of the posterior. These conditions refer to the number of different observations in the sample, are trivial to check and are very likely to be satisfied in samples of practical interest. The particularly tractable nature of the skew-Laplace model makes it easy to introduce rounding and truncation, both for computational implementations and for proofs of posterior existence. In particular, the likelihood of the model is available in closed form, in contrast with many other models, such as the skew-normal or skew-Student (e.g. using the skewing ideas of Azzalini, 1985, Fernández and Steel, 1998a, or Jones and Faddy, 2003).

The skew-Laplace model behaves well in the motivating application on flow cytometry data, as could perhaps be expected. However, it also beats the skew-Student in the glass fibre data set, an application for which skew-Laplace modelling does not seem the most appropriate at first sight, given the shape of the data histogram. In order to further illustrate the applicability of the skew-Laplace model to various datasets, Figure 10 shows the predictive distribution obtained with the skew-Laplace model for three simulated samples. We have drawn $n = 100$ observations from the skew-normal distribution of Azzalini (1985) (panel (a)) and a symmetric Student-t with 2 degrees of freedom (panel (c)). The data in panel (b) were generated from a Gamma(2,5) distribution ($n = 1000$) and analysed with a skew-Laplace truncated at zero. In all cases we have recorded data up to one decimal place (so that $d = 0.05$). Clearly, the skew-Laplace fits the rather different shapes of these three data sets quite well.

Acknowledgements: We are grateful to Olga Julià for kindly providing us with the E. Coli data. F. Javier Rubio acknowledges support from CONACYT.

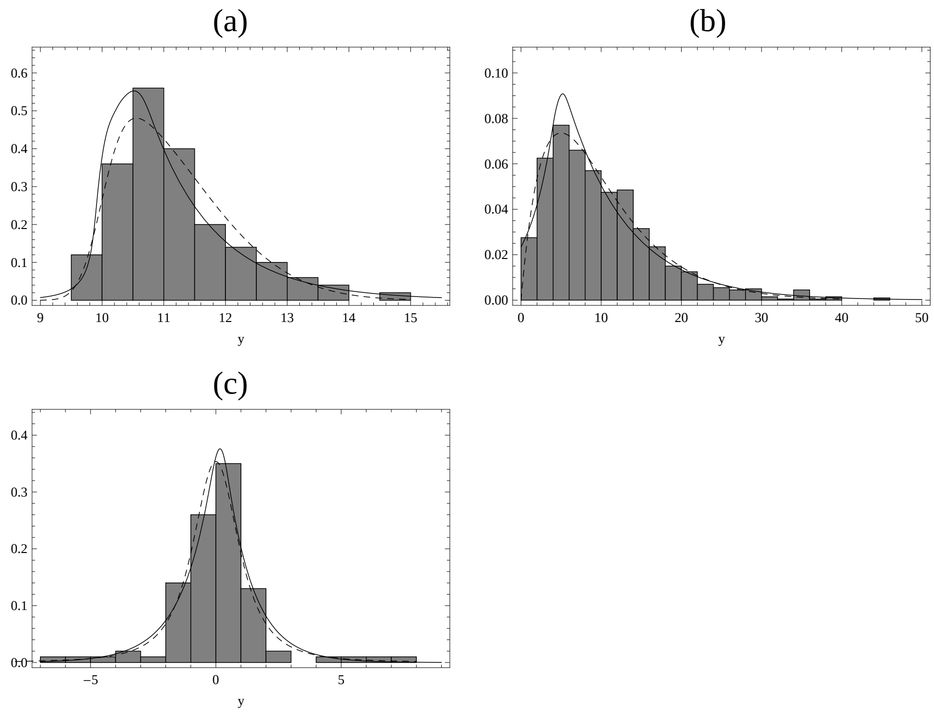

Figure 10: Simulated data: Skew-Laplace predictive (solid line) and data-generating density (dashed line) with data histogram in grey. Data generated from (a) Azzalini skew-normal ($n = 100$) (b) Gamma(2,5) with zero truncated skew-Laplace ($n = 1000$) ; (c) t_2 ($n = 100$).

Appendix: Proofs

Throughout, the order statistics of the observations will be denoted by $y^*_{(1)} < y^*_{(2)} \cdots < y^*_{(k)}$.

Proof of Theorem 1

If $0 \le \mu < y^*_{(2)}$ then

$$\mathcal{L}(y|\mu,\sigma,\gamma) \le F_X\left(y^*_{(k)} + d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(k)} - d\middle|\mu,\sigma,\gamma\right) = 2df_X(\zeta_k|\mu,\sigma,\gamma),$$

where $\zeta_k \in \left(y^*_{(k)} - d, y^*_{(k)} + d\right)$ then

$$\mathcal{L}(y|\mu,\sigma,\gamma) < 2df_X\left(y^*_{(k)} - d\middle|\mu,\sigma,\gamma\right) < 2df_X\left(y^*_{(k)} - d\middle|y^*_{(2)},\sigma,\gamma\right).$$

If $y^*_{(2)} \le \mu \le M$ then

$$\mathcal{L}(y|\mu,\sigma,\gamma) \;\; \le \;\; F_X\left(y^*_{(1)} + d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(1)} - d\middle|\mu,\sigma,\gamma\right) = 2df_X(\zeta_1|\mu,\sigma,\gamma),$$

where $\zeta_1 \in \left(y^*_{(1)} - d, y^*_{(1)} + d\right)$ then

$$\mathcal{L}(y|\mu,\sigma,\gamma) < 2df_X\left(y^*_{(1)} + d\middle|\mu,\sigma,\gamma\right) < 2df_X\left(y^*_{(1)} + d\middle|y^*_{(2)},\sigma,\gamma\right).$$

Therefore we have, for some finite and positive constant C, that

$$\begin{aligned}\int_0^M\int_0^\infty\int_0^\infty \mathcal{L}(y|\mu,\sigma,\gamma)\pi(\mu,\sigma,\gamma)\,d\sigma d\gamma d\mu &= \int_0^{y^*_{(2)}}\int_0^\infty\int_0^\infty \mathcal{L}(y|\mu,\sigma,\gamma)\pi(\mu,\sigma,\gamma)\,d\sigma d\gamma d\mu \\ + \int_{y^*_{(2)}}^M\int_0^\infty\int_0^\infty \mathcal{L}(y|\mu,\sigma,\gamma)\pi(\mu,\sigma,\gamma)\,d\sigma d\gamma d\mu &\leq C\left(\frac{y^*_{(2)}}{y^*_{(k)} - y^*_{(2)} - d} + \frac{M - y^*_{(2)}}{y^*_{(2)} - y^*_{(1)} - d}\right),\end{aligned}$$

which is finite provided we have at least three distinct observations (*i.e.* $k \geq 3$).

Proof of Theorem 2

First of all, note that for all $K_1 \geq \mu$ and $K_2 \geq K_1 + \epsilon$

$$\frac{F_X(K_2+\epsilon|\mu,\sigma,\gamma) - F_X(K_2|\mu,\sigma,\gamma)}{F_X(K_1+\epsilon|\mu,\sigma,\gamma) - F_X(K_1|\mu,\sigma,\gamma)} = \exp\left[-\frac{K_2 - K_1}{\gamma\sigma}\right],$$

and for all $L_2 \leq \mu - \epsilon$ and $L_1 \leq L_2 - \epsilon$

$$\frac{F_X(L_1+\epsilon|\mu,\sigma,\gamma) - F_X(L_1|\mu,\sigma,\gamma)}{F_X(L_2+\epsilon|\mu,\sigma,\gamma) - F_X(L_2|\mu,\sigma,\gamma)} = \exp\left[-\frac{\gamma(L_2 - L_1)}{\sigma}\right],$$

If $y^*_{(1)} - d \leq \mu \leq y^*_{(2)} + d$ and $\epsilon = 2d$ then

$$\begin{aligned}\mathcal{L}(y|\mu,\sigma,\gamma,\theta_1,\theta_2) &\leq \frac{F_X\left(y^*_{(k)} + d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(k)} - d\middle|\mu,\sigma,\gamma\right)}{F_X\left(y^*_{(k)} + d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(1)} - d\middle|\mu,\sigma,\gamma\right)} \\ &\leq \frac{F_X\left(y^*_{(k)} + d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(k)} - d\middle|\mu,\sigma,\gamma\right)}{F_X\left(y^*_{(k-1)} + d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(k-1)} - d\middle|\mu,\sigma,\gamma\right)} \\ &\leq \exp\left[-\frac{y^*_{(k)} - y^*_{(k-1)}}{\gamma\sigma}\right],\end{aligned}$$

If $y^*_{(2)} + d < \mu \leq y^*_{(k)} + d$ and $\epsilon = 2d$ then

$$\begin{aligned}
\mathcal{L}(y|\mu,\sigma,\gamma,\theta_1,\theta_2) &\leq \frac{F_X\left(y^*_{(1)}+d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(1)}-d\middle|\mu,\sigma,\gamma\right)}{F_X\left(y^*_{(k)}+d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(1)}-d\middle|\mu,\sigma,\gamma\right)} \\
&\leq \frac{F_X\left(y^*_{(1)}+d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(1)}-d\middle|\mu,\sigma,\gamma\right)}{F_X\left(y^*_{(2)}+d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(2)}-d\middle|\mu,\sigma,\gamma\right)} \\
&\leq \exp\left[-\frac{\gamma(y^*_{(2)}-y^*_{(1)})}{\sigma}\right].
\end{aligned}$$

We can then write, for some finite positive C

$$\int_0^{y^*_{(1)}-d}\int_{y^*_{(k)}+d}^{M}\int_{y^*_{(1)}-d}^{y^*_{(k)}+d}\int_0^\infty\int_0^\infty \mathcal{L}(y|\mu,\sigma,\gamma,\theta_1,\theta_2)\pi(\mu,\sigma,\gamma,\theta_1,\theta_2)\, d\sigma d\gamma d\mu d\theta_1 d\theta_2$$

$$\begin{aligned}
&\leq C\int_{y^*_{(1)}-d}^{y^*_{(2)}+d}\int_0^\infty\int_0^\infty \exp\left[-\frac{y^*_{(k)}-y^*_{(k-1)}}{\gamma\sigma}\right]\frac{1}{\sigma^2}\frac{\gamma}{(1+\gamma^2)^2}\,d\sigma d\gamma d\mu \\
&+ C\int_{y^*_{(2)}+d}^{y^*_{(k)}+d}\int_0^\infty\int_0^\infty \exp\left[-\frac{\gamma(y^*_{(2)}-y^*_{(1)})}{\sigma}\right]\frac{1}{\sigma^2}\frac{\gamma}{(1+\gamma^2)^2}\,d\sigma d\gamma d\mu \\
&\propto \frac{y^*_{(2)}-y^*_{(1)}+2d}{y^*_{(k)}-y^*_{(k-1)}} + \frac{y^*_{(k)}-y^*_{(2)}}{y^*_{(2)}-y^*_{(1)}} < \infty, \text{ provided } k \geq 4.
\end{aligned}$$

Proof of Theorem 3

The proof is analogous to the proof of Theorem 2 using the fact that

$$1 - F_X\left(y^*_{(1)}-d\middle|\mu,\sigma,\gamma\right) \geq F_X\left(y^*_{(k)}+d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(1)}-d\middle|\mu,\sigma,\gamma\right).$$

Proof of Theorem 4

The proof is analogous to the proof of Theorem 2 using the fact that

$$1 - F_X\left(0\middle|\mu,\sigma,\gamma\right) \geq F_X\left(y^*_{(k)}+d\middle|\mu,\sigma,\gamma\right) - F_X\left(y^*_{(1)}-d\middle|\mu,\sigma,\gamma\right).$$

Proof of Theorem 5

First we will prove that this result is equivalent to the properness of the posterior distribution for $\gamma = 1$ and then we will prove the result for $\gamma = 1$.

Without loss of generality let us assume that $S_1 \cap S_2 = \emptyset$, this assumption is reasonable given that $k \geq 3$. Then writing the Student's t as a scale mixture of normals with mixing parameters $\lambda = (\lambda_1, \ldots, \lambda_n)'$ and applying Fubini's theorem we get an upper bound for $\mathbb{P}[y_1 \in S_1, ..., y_n \in S_n]$ which is proportional to

$$\int_{1+\epsilon}^{\infty} \int_{\mathbb{R}_+^n} \int_{S_n \times \ldots \times S_1} \int_0^{\infty} \int_0^{\infty} \int_0^{M} \left(\prod_{j=1}^{n} \lambda_j^{\frac{1}{2}} \right) \frac{\sigma^{-n}}{(\gamma + 1/\gamma)^n}$$
$$\times \quad \exp\left[-\frac{1}{2\sigma^2 h(\gamma)^2} \sum_{j=1}^{n} \lambda_j (y_j - \mu)^2 \right] \frac{\gamma \sigma^{-2}}{(1+\gamma^2)^2}\, d\mu d\sigma d\gamma dy_1 ... dy_n dP_{\lambda|\nu} dP_{\nu},$$

where $h(\gamma) = \max\{\gamma, 1/\gamma\}$. Consider the change of variable $\vartheta = h(\gamma)\sigma$ we can rewrite this upper bound as follows

$$\int_0^{\infty} \frac{h(\gamma)^{n+1} \gamma^{n+1}}{(1+\gamma^2)^{n+2}} d\gamma \int_{1+\epsilon}^{\infty} \int_{\mathbb{R}_+^n} \int_{S_n \times \ldots \times S_1} \int_0^{\infty} \int_0^{M} \left(\prod_{j=1}^{n} \lambda_j^{\frac{1}{2}} \right) \frac{1}{\vartheta^{n+2}}$$
$$\times \quad \exp\left[-\frac{1}{2\vartheta^2} \sum_{j=1}^{n} \lambda_j (y_j - \mu)^2 \right] d\mu d\vartheta dy_1 ... dy_n dP_{\lambda|\nu} dP_{\nu}. \tag{11}$$

The first integral is finite and the second integral is equivalent to the marginal distribution when $\gamma = 1$. Now we will prove the properness of the posterior distribution for $\gamma = 1$. Defining $S^2(\lambda, y) = \sum_{1 \leq i < j \leq n} \lambda_i \lambda_j (y_i - y_j)^2$ and $\gamma = 1$ we have

$$\int_{1+\epsilon}^{\infty} \int_{\mathbb{R}_+^n} \int_{S_n \times \ldots \times S_1} \int_0^{\infty} \int_0^{M} \left(\prod_{j=1}^{n} \lambda_j^{\frac{1}{2}} \right) \frac{1}{\sigma^{n+2}} \exp\left[-\frac{1}{2\sigma^2} \frac{S^2(\lambda, y)}{\sum_{j=1}^{n} \lambda_j} \right]$$
$$\times \quad \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^{n} \lambda_j \left(\mu - \frac{\sum_{j=1}^{n} \lambda_j y_j}{\sum_{j=1}^{n} \lambda_j} \right)^2 \right] d\mu d\sigma dy_1 ... dy_n dP_{\lambda|\nu} dP_{\nu}$$
$$\leq \quad \int_{1+\epsilon}^{\infty} \int_{\mathbb{R}_+^n} \int_{S_n \times \ldots \times S_1} \int_0^{\infty} \int_{-\infty}^{\infty} \left(\prod_{j=1}^{n} \lambda_j^{\frac{1}{2}} \right) \frac{1}{\sigma^{n+2}} \exp\left[-\frac{1}{2\sigma^2} \frac{S^2(\lambda, y)}{\sum_{j=1}^{n} \lambda_j} \right]$$
$$\times \quad \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^{n} \lambda_j \left(\mu - \frac{\sum_{j=1}^{n} \lambda_j y_j}{\sum_{j=1}^{n} \lambda_j} \right)^2 \right] d\mu d\sigma dy_1 ... dy_n dP_{\lambda|\nu} dP_{\nu}$$
$$\propto \quad \int_{1+\epsilon}^{\infty} \int_{\mathbb{R}_+^n} \int_{S_n \times \ldots \times S_1} \int_0^{\infty} \left(\prod_{j=1}^{n} \lambda_j^{\frac{1}{2}} \right) \left(\sum_{j=1}^{n} \lambda_j \right)^{-\frac{1}{2}} \frac{1}{\sigma^{n+1}}$$
$$\times \quad \exp\left[-\frac{1}{2\sigma^2} \frac{S^2(\lambda, y)}{\sum_{j=1}^{n} \lambda_j} \right] d\sigma dy_1 ... dy_n dP_{\lambda|\nu} dP_{\nu} \tag{12}$$

$$\propto \int_{1+\epsilon}^{\infty} \int_{\mathbb{R}_+^n} \int_{S_n \times \ldots \times S_1} \left(\prod_{j=1}^{n} \lambda_j^{\frac{1}{2}} \right) \left(\sum_{j=1}^{n} \lambda_j \right)^{\frac{n-1}{2}} S^2(\lambda, y)^{-\frac{n}{2}} \, dy_1 \ldots dy_n dP_{\lambda|\nu} dP_{\nu}. \quad (13)$$

Using the proof of Theorem 4 in Fernández and Steel (1998b)

$$S^2(\lambda, y) = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \left(\sum_{j=1}^{n} \lambda_j \right) \eta_2^2 + (\eta_3 - \rho, \ldots, \eta_n - \rho) Q (\eta_3 - \rho, \ldots, \eta_n - \rho)',$$

where $\eta_i = y_1 - y_i$ for $i = 2, \ldots, n$, $\rho = \lambda_2 \eta_2 / (\lambda_1 + \lambda_2)$ and $Q = (q_{ij})_{i,j=3}^{n}$ with diagonal elements $q_{ii} = \lambda_i \sum_{j \neq i} \lambda_j$ and off-diagonal elements $q_{ij} = q_{ji} = -\lambda_i \lambda_j$. Defining $\alpha = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \left(\sum_{j=1}^{n} \lambda_j \right) \eta_2^2$ we get

$$\begin{aligned}
S^2(\lambda, y)^{-\frac{n}{2}} &= \alpha^{-\frac{n}{2}} \left[1 + (\eta_3 - \rho, \ldots, \eta_n - \rho) \frac{Q}{\alpha} (\eta_3 - \rho, \ldots, \eta_n - \rho)' \right]^{-\frac{n}{2}} \\
&\leq \alpha^{-\frac{n}{2}} \left[1 + (\eta_3 - \rho, \ldots, \eta_n - \rho) \frac{Q}{\alpha} (\eta_3 - \rho, \ldots, \eta_n - \rho)' \right]^{-\frac{n-1}{2}} \\
&= \alpha^{-\frac{1}{2}} S^2(\lambda, y)^{-\frac{n-1}{2}} \\
&\leq \left(\lambda_1^{-\frac{1}{2}} + \lambda_2^{-\frac{1}{2}} \right) \left(\sum_{j=1}^{n} \lambda_j \right)^{-\frac{1}{2}} |\eta_2|^{-1} S^2(\lambda, y)^{-\frac{n-1}{2}}.
\end{aligned}$$

Integrating $(\eta_3, \ldots, \eta_n)'$ over the whole of $\mathbb{R}^{n-2}$ as in Fernández and Steel (1998b) we get the following upper bound

$$\begin{aligned}
&\int_{S_n \times \ldots \times S_1} S^2(\lambda, y)^{-\frac{n}{2}} \, dy_1 \ldots dy_n \leq \left(\lambda_1^{-\frac{1}{2}} + \lambda_2^{-\frac{1}{2}} \right) \left(\sum_{j=1}^{n} \lambda_j \right)^{-\frac{1}{2}} |\eta_2|^{-1} \\
\times &\int_{S_n \times \ldots \times S_1} S^2(\lambda, y)^{-\frac{n-1}{2}} \, dy_1 \ldots dy_n \leq \left(\prod_{j=1}^{n} \lambda_j^{-\frac{1}{2}} \right) \left(\sum_{j=1}^{n} \lambda_j \right)^{-\frac{n-1}{2}} \\
\times &\left(\lambda_1^{-\frac{1}{2}} + \lambda_2^{-\frac{1}{2}} \right) \int_{\{y_1 \in S_1, y_1 - \eta_2 \in S_2\}} |\eta_2|^{-2} \, dy_1 d\eta_2. \quad (14)
\end{aligned}$$

Combining (13) and (14) we get

$$\begin{aligned}
&\int_{1+\epsilon}^{\infty} \int_{\mathbb{R}_+^n} \int_{S_n \times \ldots \times S_1} \left(\prod_{j=1}^{n} \lambda_j^{\frac{1}{2}} \right) \left(\sum_{j=1}^{n} \lambda_j \right)^{\frac{n-1}{2}} S^2(\lambda, y)^{-\frac{n}{2}} \, dy_1 \ldots dy_n dP_{\lambda|\nu} dP_{\nu} \\
\leq &\int_{1+\epsilon}^{\infty} \int_{\mathbb{R}_+^n} \left(\lambda_1^{-\frac{1}{2}} + \lambda_2^{-\frac{1}{2}} \right) dP_{\lambda|\nu} dP_{\nu} \int_{\{y_1 \in S_1, y_1 - \eta_2 \in S_2\}} |\eta_2|^{-2} \, dy_1 d\eta_2 \\
\propto &\int_{1+\epsilon}^{\infty} \int_{\mathbb{R}_+} \lambda_1^{-\frac{1}{2}} dP_{\lambda_1|\nu} dP_{\nu} \int_{\{y_1 \in S_1, y_1 - \eta_2 \in S_2\}} |\eta_2|^{-2} \, dy_1 d\eta_2.
\end{aligned}$$

The third integral is finite since $S_1 \cap S_2 = \emptyset$. Now, considering that $\lambda_j|\nu \sim Ga\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ for $j = 1, ..., n$

$$\int_{\mathbb{R}_+} \lambda_1^{-\frac{1}{2}} dP_{\lambda_1|\nu} = \frac{\sqrt{2}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\nu}\Gamma\left(\frac{\nu}{2}\right)} \leq \frac{\sqrt{2}\Gamma\left(\frac{\epsilon}{2}\right)}{\sqrt{\epsilon+1}\Gamma\left(\frac{\epsilon+1}{2}\right)}, \text{ given that } \nu \geq 1+\epsilon.$$

References

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *The American Statistician* 49: 34–38.

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

Christen, J. A. and Fox, C. (2010). A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis* 5: 1–20.

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

Fernández, C. and Steel, M. F. J. (1998b). On the dangers of modelling through continuous distributions: A Bayesian perspective, in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. eds., *Bayesian Statistics 6*, Oxford University Press (with discussion), pp. 213–238.

Fernández, C. and Steel, M. F. J. (1999a). Multivariate Student-t regression models: Pitfalls and inference. *Biometrika* 86: 153–167.

Fernández, C. and Steel, M. F. J. (1999b). Reference priors for the general location-scale model. *Statistics and Probability Letters* 43: 377–384.

Ferreira, J. T. A. S. and Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association*, 101: 823-829.

Fonseca, T., Ferreira, M. and Migon, H.. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika* 95: 325–333.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society*, Ser. B 56: 501-514.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association*, 102:360–378.

Heitjan, D. F. (1989). Inference from grouped continuous data: A review. *Statistical Science* 4: 164-183.

Jones, M. C. and Faddy, M. J. (2003). A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society* Ser. B 65: 159–174.

Julià, O. and Vives-Rego, J. (2005). Skew-Laplace distribution in Gram-negative bacterial axenic cultures: new insights into intrinsic cellular heterogeneity. *Microbiology* 151: 749–755.

Julià, O. and Vives-Rego, J. (2008). A microbiology application of the skew-Laplace distribution. *SORT* 32: 141–150.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association* 90: 773–795.

Lingappaiah, G. S. (1988). On two-piece double exponentail distributions. *Journal of the Korean Statistical Society* 17: 46–55.

Schneeweiss, H., Komlos, J. and Ahmad, A.S. (2010). Symmetric and asymmetric rounding: A review and some new results. *Advances in Statistical Analysis*, forthcoming.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics* 6: 461-464

Smith, R. L. and Naylor, J. C. (1987). A Comparison of Maximum Likelihood and Bayesian Estimators for the Three- Parameter Weibull Distribution. *Journal of the Royal Statistical Society*, Ser. C 36: 358–369.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society*, Ser. B 64 : 583-639.