Japanese Journal of Statistics and Data Science (2024) 7:701–738
https://doi.org/10.1007/s42081-024-00251-4

ORIGINAL PAPER

Alternative skew Laplace scale mixtures for modeling data exhibiting high-peaked and heavy-tailed traits

A. F. Otto[1] · A. Bekker[1,2] · J. T. Ferreira[1] · O. Arslan[3]

Received: 2 June 2023 / Revised: 7 May 2024 / Accepted: 8 May 2024 /
Published online: 27 May 2024

Abstract

The search and construction of appropriate and flexible models for describing and modelling empirical data sets incongruent with normality retains a sustained interest. This paper focuses on proposing flexible skew Laplace scale mixture distributions to model these types of data sets. Each member of the collection of distributions is obtained by dividing the scale parameter of a conditional skew Laplace distribution by a purposefully chosen mixing random variable. Highly-peaked, heavy-tailed skew models with relevance and impact in different fields are obtained and investigated, and elegant sampling schemes to simulate from this collection of developed models are proposed. Finite mixtures consisting of the members of the skew Laplace scale mixture models are illustrated, further extending the flexibility of the distributions by being able to account for multimodality. The maximum likelihood estimates of the parameters for all the members of the developed models are described via a developed EM algorithm. Real-data examples highlight select models' performance and emphasize their viability compared to other commonly considered candidates, and various goodness-of-fit measures are used to endorse the performance of the proposed models as reasonable and viable candidates for the practitioner. Finally, an outline is discussed for future work in the multivariate realm for these models.

Keywords Bodily injury claims data · Contaminated model · Donor ideology data · Finite mixtures · Heavy-tailed distributions · Scale mixtures

✉ J. T. Ferreira
johan.ferreira@up.ac.za

1 Department of Statistics, University of Pretoria, Pretoria, South Africa

2 Centre for Environmental Studies, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa

3 Department of Statistics, Ankara University, Ankara, Turkey

1 Introduction

The search for appropriate and flexible models for describing larger data sets, which frequently deviate from normality, remains a major focus in many research fields (Adcock & Azzalini, 2020). The Laplace distribution is among the most well-known parametric models for real valued data, and is preferred to the normal distribution when the empirical distribution is still symmetric but presents heavy-tails and a discernible peak (Kotz et al., 2001; Wilson, 1923). It is often referred to as the double exponential distribution, as it can be thought of as two exponential distributions spliced together back-to-back with an additional location parameter.

The probability density function (PDF) of the Laplace distribution is expressed in terms of the absolute difference from the mean, which is reminiscent of the normal distribution which is expressed in terms of the squared difference from the mean. As a result, the Laplace distribution has heavier tails than the normal distribution (DeCarlo, 1997). The excess kurtosis, defined as kurtosis minus 3 (i.e. the kurtosis of the univariate normal distribution), of the Laplace distribution is fixed at 3 and it is thus *leptokurtic*. The flexibility of the Laplace distribution is still, however, limited because of this fixed excess kurtosis.

In an attempt to make the tail behaviour of the Laplace more flexible, Punzo and Bagnato (2021) extended this branch of literature by introducing the *Laplace scale mixture* (LSM) family of distributions to make the Laplace tail behaviour more flexible in order to model the highly-peaked and heavy-tailed return distribution of cryptocurrencies. Due to its symmetric property, the LSM models are not able to model skewness. Punzo and Bagnato (2022a) circumvented this limitation by using the asymmetric Laplace distribution as reference distribution instead, and developed a family of asymmetric Laplace scale mixture (ALSM) distributions capable of modelling, not only highly-peaked and heavy-tailed data, but also, skewness.

Various types of skewed generalisations of the Laplace distribution (1) have sporadically appeared in literature over the past few decades. One of the earliest is introduced by McGill (1962) and a variation thereof is studied by Holla and Bhattacharya (1968). Another generalization is studied by Poiraud-Casanova and Thomas-Agnan (2000), which is a special case of the skewed Laplace distribution of Yu and Jin (2005).
Subbotin (1923) proposed a generalization of the Laplace distribution, which is sometimes referred to as the exponential power function distribution. The Laplace distribution is also a special case of several distributions: the asymmetric power (Komunjer, 2007), double Weibull (Balakrishnan & Kocherlakota, 1985), Sargan (Hadri, 1996), the Laplace normal mixture (Kanji, 1985) and beta Laplace distributions (Kozubowski & Nadarajah, 2008) for example. A variance-mean mixture of multivariate normal distributions in Doğru and Arslan (2023), seeks to define a flexible extension of the Laplace distribution, and an extensive review of the various Laplace distribution generalisations is given in Kozubowski and Nadarajah (2010). The alpha-skew Laplace distribution Harandi and Alamatsaz (2013) is a bimodal distribution, which contains the Laplace distribution as a special case, and is further extended to a discrete alpha skew Laplace (Harandi & Alamatsaz, 2015), the Balakrishnan alpha-skew Laplace (Shah et al., 2019) and the Balakrishnan alpha-beta-skew Laplace distribution (Shah et al., 2023). Other multimodal extensions are proposed

in Chakraborty et al. (2014), Doğru and Arslan (2017) and Doğru and Arslan (2021). Fernández and Steel (1998) introduced skewness to a symmetric distribution by converting a symmetric PDF into a skewed one by stipulating inverse scale factors in the positive and negative half lines. This resulted in the asymmetric Laplace distribution studied in Kotz et al. (2001), from which the ALSM distributions were formed upon.

In Sect. 2, we show that other generalizations of the Laplace distribution can be used as reference distributions (Kozubowski & Nadarajah, 2010) and be considered as valid alternatives to the ALSM models introduced in Punzo and Bagnato (2022a); literature contextualizes the choices of the Bernoulli-, shifted exponential-, unimodal gamma-, and power function distributions - which are considered as mixing variables in this paper. Convenient computational tools are reviewed such as moments and sampling schemes. The focus is on using previously unconsidered skew Laplace (SL) models as the reference (conditional) distribution that can leverage skewness. Different mixing random variables on all or part of the real line are considered to develop the skew Laplace scale mixture (SLSM) models in Sect. 3; in this way, the novelty of this paper is thus two-fold. Ultimately, a collection of flexible *skew Laplace scale mixture* (SLSM) distributions that can model data with peakedness, heavy-tails, skewness and multimodality is proposed. In Sect. 4, finite mixtures consisting of the members of the SLSM models are also explored - further extending the models ability to fit multimodal data. Methods considered for estimation are described in Sect. 5, followed by applications of the proposed models on the well-known bodily injury claims dataset (Rempala & Derrig, 2005) and the refractive index of glass fragments in Sect. 6. Furthermore, we employ finite mixtures of the members of the SLSM distributions fitting the campaign finance scores, which serve as indicators of the ideological leanings of donors to political candidates in the United States. Finally, Sect. 7 concludes the paper with a view towards future work in the multivariate setting.

2 Essential components

The PDF of the Laplace distribution is expressed as

$$f_L(x;\mu,\beta) = \frac{1}{2\beta}\mathrm{e}^{-\frac{|x-\mu|}{\beta}}, \quad -\infty < x < \infty, \tag{1}$$

where $\mu \in (-\infty, \infty)$ and $\beta > 0$ are location and scale parameters, respectively. If a random variable X has the Laplace distribution, we denote it by $X \sim Laplace(\mu, \beta)$. The LSM distribution introduced by Punzo and Bagnato (2021) is presented in Definition 2.1.

Definition 2.1 A random variable X is said to have the LSM distribution with location $\mu \in (-\infty, \infty)$, scale $\beta \in (0, \infty)$ and tailedness $\boldsymbol{\theta}$, if its PDF is given by

$$g_{LSM}(x;\mu,\beta,\boldsymbol{\theta}) = \int_{S_h} f_L(x;\mu,\beta/w)h(w;\boldsymbol{\theta})dw, \quad -\infty < x < \infty, \tag{2}$$

where $h(w; \boldsymbol{\theta})$ is the mixing PDF of W, with support $S_h \subseteq (0, \infty)$, depending on the vector of parameters $\boldsymbol{\theta}$.

In a similar manner, the SLSM distribution follows in Definition 2.2 which considers a SL distribution as the conditional reference distribution. The proposed SLSM distribution is constituted from two components: the reference (conditional) distribution $f_{SL}(\cdot)$ and the mixing distribution, $h(\cdot)$. This is similar to the ALSM distribution introduced by Punzo and Bagnato (2022a); and the acronym SLSM is used to avoid confusion with the models proposed and discussed in Punzo and Bagnato (2022a).

Definition 2.2 A random variable X is said to have the SLSM distribution with location $\mu \in (-\infty, \infty)$, scale $\beta \in (0, \infty)$, skewness λ, and tailedness $\boldsymbol{\theta}$, if its PDF is given by

$$g_{SLSM}(x; \mu, \beta, \boldsymbol{\theta}, \lambda) = \int_{S_h} f_{SL}(x; \mu, \beta/w, \lambda) h(w; \boldsymbol{\theta}) dw, \quad -\infty < x < \infty, \tag{3}$$

$$\equiv \sum_{w} f_{SL}(x; \mu, \beta/w, \lambda) h(w; \boldsymbol{\theta}) \tag{4}$$

where $h(w; \boldsymbol{\theta})$ is the mixing PDF of W, with support $S_h \subseteq (0, \infty)$, depending on the vector of parameters $\boldsymbol{\theta}$. This is denoted as $X \sim SLSM(\mu, \beta, \boldsymbol{\theta})$. (4) follows when W is a discrete random variable.

A SLSM distribution can be thought of as a compound distribution, with the same location μ and skewness parameter λ but with different scale β/w due to w (see also Punzo et al. 2018). The component skew Laplace distributions, as conditional reference distribution, are not taken uniformly from the set, but according to a set of "weights" determined by the probabilistic behaviour of W. Note that if W is degenerate in 1 (i.e. $W \equiv 1$ which implies that $P[W = 1] = 1$) then $f_{SL}(x; \mu, \beta, \lambda)$ is obtained, whereas in other cases the tails of f_{SLSM} are heavier compared to f_{SL}.

Formulation of skew Laplace distributions

Azzalini (1985) presented the skew-normal distribution by adding an additional parameter to introduce asymmetry, as described in Definition 2.3. Using this definition, two reference (conditional) candidates will be proposed.

Definition 2.3 Denote a PDF on $\mathbb{R}$ by $f_0(\cdot)$, a continuous cumulative distribution function (CDF) on $\mathbb{R}$ by $G_0(\cdot)$, and a real-valued function on $\mathbb{R}$ by $w(\cdot)$, such that $f_0(-x) = f_0(x)$, $G_0(-x) = 1 - G_0(x)$ and $w(-x) = -w(x)$ $\forall x \in (-\infty, \infty)$. Then

$$f_X(x) = 2 f_0(x) G_0(w(x)) \tag{5}$$

is a PDF on $\mathbb{R}$ (Azzalini 1985). Note that f_0 is termed the symmetric base PDF, $2G_0(w(x))$ is termed the skewing mechanism, and f_X is termed the skewed version of the symmetric base PDF.

2.1 An Azzalini type skew Laplace distribution

Azzalini investigated the case where $w(x) = \lambda x$ for $\lambda \in (-\infty, \infty)$ and f_0 is a standard-normal PDF, which yields the skew-normal distribution. Similarly, as illustrated in Aryal and Nadarajah (2005), a skewed Laplace distribution can be obtained by utilizing (5) and instead using a Laplace distribution as the base distribution; see also Gupta et al. (2002), Jagannathan (2005) and Kozubowski and Nolan (2008). This will be referred to as the skew Laplace type I (SL_I) distribution, with PDF given by

$$f_{SL_I}(x;\mu,\beta,\lambda) = \begin{cases} \frac{1}{2\beta}\mathrm{e}^{\frac{(\lambda+1)(x-\mu)}{\beta}}, & x \le \mu, \\ \frac{1}{\beta}\mathrm{e}^{-\frac{(x-\mu)}{\beta}} - \frac{1}{2\beta}\mathrm{e}^{-\frac{(\lambda+1)(x-\mu)}{\beta}}, & x > \mu, \end{cases} \tag{6}$$

where $\beta, \lambda > 0$. This is denoted by $X \sim SL_I(\mu, \beta, \lambda)$.

To obtain valid PDF's a restriction is placed on $\lambda > 0$. The corresponding PDF for $\lambda < 0$ can be obtained using the fact that $-X$ has the PDF $2f_L(x;\mu,\beta)F_L(-\lambda x;\mu,\beta)$, where $F_L(\cdot)$ denotes the CDF of the Laplace distribution. Thus, the $\lambda > 0$ restriction does not result in a loss of generality in the distribution (see also Kozubowski and Nolan (2008)). The main feature of the SL_I distribution is that the addition of the new parameter λ allows for a greater degree of flexibility. If $\lambda = 0$, (6) reduces to the PDF of the Laplace distribution (1).

A technique for generating SL_I random variables with PDF (6) is the inversion transform sampling method (see Aryal & Nadarajah 2005). Therefore, in this case,

$$X \overset{d}{=} \begin{cases} \frac{\beta}{1+\lambda}\ln(2U(1+\lambda)) + \mu, & 0 \le U \le \frac{1}{2(1+\lambda)}, \\ \mu - \beta\ln\left\{1 - U + (1-U)\sum_{k=1}^{\infty}\binom{(1+\lambda)k}{k-1}\frac{(1-U)^{\lambda k}}{k2^k(1+\lambda)^k}\right\}, & \frac{1}{2(1+\lambda)} \le U \le 1, \end{cases} \tag{7}$$

where $U \sim Uniform(0,1)$. Without loss of generality, the kth moment of $X \sim SL_I(0, \beta, \kappa)$ is given by

$$\mathrm{E}(X^k) = \begin{cases} \beta^k\Gamma(k+1), & \text{if } k \text{ is even}, \\ \beta^k\Gamma(k+1)\left\{1 - \frac{1}{(1+\lambda)^{k+1}}\right\}, & \text{if } k \text{ is odd}. \end{cases} \tag{8}$$

From (8) and the binomial expansion, the kth central moment of X can be calculated as

$$\mathrm{E}_{SL_{II}}\left[(X-\mu)^k\right] = \begin{cases} \mu^k + \sum_{j=1}^{k/2} \binom{k}{2j} \mu^{k-2j} \beta^{2j} \Gamma(2j+1) \\ - \sum_{j=1}^{k/2} \binom{k}{2j-1} \mu^{k-2j+1} \beta^{2j-1} \Gamma(2j) \left\{1 - \frac{1}{(1+\lambda)^{2j}}\right\}, & \text{if } k \text{ is even,} \\ -\mu^k - \sum_{j=1}^{\frac{k-1}{2}} \binom{k}{2j} \mu^{k-2j} \beta^{2j} \Gamma(2j+1) \\ + \sum_{j=1}^{\frac{k+1}{2}} \binom{k}{2j-1} \mu^{k-2j+1} \beta^{2j-1} \Gamma(2j) \left\{1 - \frac{1}{(1+\lambda)^{2j}}\right\}, & \text{if } k \text{ is odd.} \end{cases} \tag{9}$$

From (8) and (9) it follows that

$$\mathrm{E}_{SL_I}(X) = \mu + \beta\left(1 - \frac{1}{(1+\lambda)^2}\right),$$

$$\mathrm{Var}_{SL_I}(X) = \frac{\beta^2(2+8\lambda+8\lambda^2+4\lambda^3+\lambda^4)}{(1+\lambda)^4},$$

$$\mathrm{Skewness}_{SL_I}(X) = \frac{2\lambda(6+15\lambda+20\lambda^2+15\lambda^3+6\lambda^4+\lambda^5)}{(2+8\lambda+8\lambda^2+4\lambda^3+\lambda^4)^{3/2}}, \tag{10}$$

$$\mathrm{Kurtosis}_{SL_I}(X) = \frac{3(8+64\lambda+176\lambda^2+272\lambda^3+276\lambda^4+192\lambda^5+88\lambda^6+24\lambda^7+3\lambda^8)}{(2+8\lambda+8\lambda^2+4\lambda^3+\lambda^4)^2}. \tag{11}$$

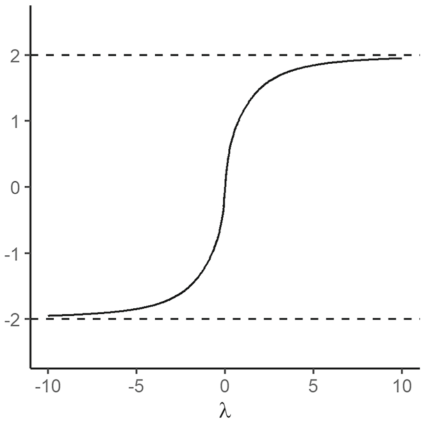
A comprehensive description of the mathematical properties of the SL_I distribution is derived in Aryal and Nadarajah (2005). From (10) and (11) it is evident that the skewness and kurtosis only depend on λ. Figure 1 illustrates the skewness and kurtosis for varying values of λ with skewness between $(-2, 2)$ and kurtosis between $(5.81, 9)$, as given in Aryal and Nadarajah (2005).

2.2 A Subbotin type skew Laplace distribution

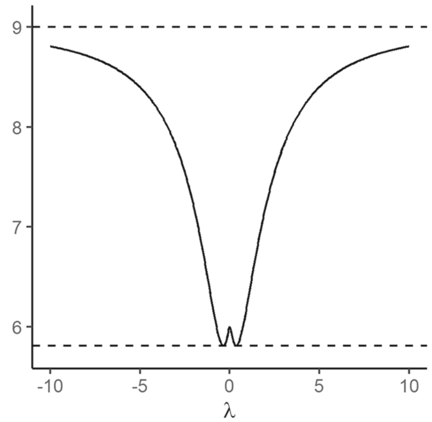
By applying Azzalini's skewing method to a generalised normal distribution attributed to Subbotin (1923) as the base distribution, with $w(x) = \sqrt{2}\lambda x$, the PDF of a random variable X with the skew generalised normal (SGN) distribution is given by

$$g_{SGN}(x;\mu,\beta,\delta,\lambda) = \frac{\delta}{\beta\Gamma\left(\frac{1}{\delta}\right)} \mathrm{e}^{-\left|\frac{x-\mu}{\beta}\right|^\delta} \Phi\left(\sqrt{2}\lambda\left(\frac{x-\mu}{\beta}\right)\right), \quad -\infty < x < \infty, \tag{12}$$

where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution, $\mu, \lambda \in (-\infty, \infty)$ and $\beta, \delta > 0$.



(a) Skewness.



(b) Kurtosis.

Fig. 1 Skewness and kurtosis of SL_I with PDF (6) as a function of λ

By setting δ to 1 in (12), we obtain the Skew Laplace type II distribution with the PDF given by

$$
\begin{aligned}
f_{SL_{II}}(x;\mu,\beta,\lambda) &= \frac{1}{\beta}\mathrm{e}^{-|\frac{x-\mu}{\beta}|}\Phi\left(\sqrt{2}\lambda\left(\frac{x-\mu}{\beta}\right)\right)\\
&= 2f_L(x;\mu,\beta)\Phi\left(\sqrt{2}\lambda\left(\frac{x-\mu}{\beta}\right)\right), \quad -\infty < x < \infty, \qquad (13)
\end{aligned}
$$

where $\mu, \lambda \in (-\infty, \infty)$ and $\beta > 0$, denoted as $X \sim SL_{II}(\mu, \beta, \lambda)$.

(13) is often referred to as the skew Laplace normal (SLN) distribution in literature (Gómez et al., 2007), with finite mixtures of the SLN distribution considered in Doğru and Arslan (2017) and Doğru and Arslan (2021). As noted in Bekker et al. (2020), the acceptance-rejection method described in Lange et al. (1999) cannot draw appropriate samples from certain parameter structures from a SGN distribution. Following a similar approach to that of Sadeghkhani and Ghosh (2018), a stochastic representation is given in Algorithm 1 for generating random numbers from a SL_{II} distribution with PDF, $f_{SL_{II}}(x;\mu,\beta,\lambda)$, as given in (13).

Algorithm 1 Generating Skew Laplace Type II Random Variables

1: Generate $U \sim Laplace(0, 1)$ by

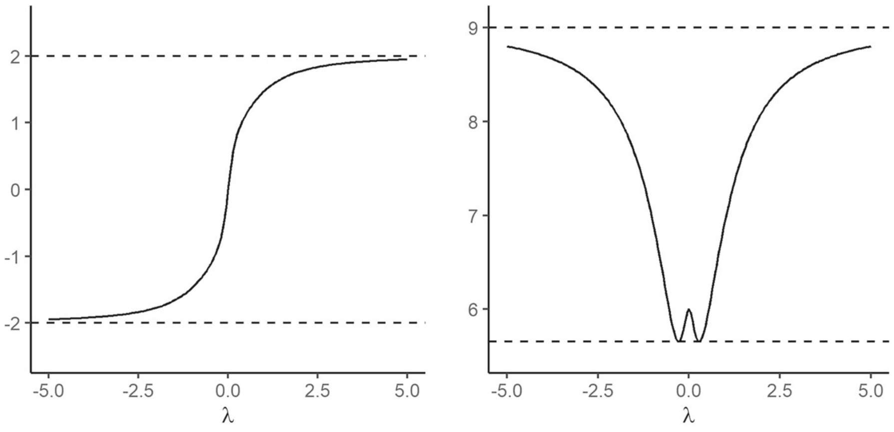
$$U \stackrel{d}{=} \begin{cases} Z, & \text{with probability } \frac{1}{2} \\ -Z, & \text{with probability } \frac{1}{2} \end{cases}$$

where $Z \sim EXP(1)$;

2: Generate $U_1 \sim N(0, 1)$, with U and U_1 independent;

3: If $U_1 \le \sqrt{2}\lambda U$, set $X = \mu + \beta U$, otherwise return to step 1;

4: **return** X.



(a) Skewness. (b) Kurtosis.

Fig. 2 Skewness and kurtosis of SL_{II} with PDF (13) as a function of λ

If $Y \sim SL_{II}(0, 1, \lambda)$, the rth moment is given as

$$\mathrm{E}[Y^r] = \begin{cases} \Gamma(r+1), & \text{if } r \text{ is even,} \\ \Gamma(r+1)\left(2\mathrm{E}_{X_1}\left[\Phi\left(\sqrt{2}\lambda X_1\right)\right]-1\right), & \text{if } r \text{ is odd,} \end{cases} \tag{14}$$

where $X_1 \sim Gamma(r+1, 1)$. Therefore, the kth moment of $X = \mu + \beta Y$ is

$$\begin{aligned} \mathrm{E}_{SL_{II}}[X^k] = \mathrm{E}[(\mu+\beta Y)^k] &= \sum_{r=0}^{k} \binom{k}{r} \mu^{k-r}\beta^r \mathrm{E}\left[Y^r\right] \\ &= \mu + \beta\left(2\mathrm{E}_{X_2}\left[\Phi\left(\sqrt{2}\lambda X_2\right)\right]-1\right), \end{aligned} \tag{15}$$

where $X_2 \sim Gamma(2, 1)$. The other centralised moments follow in a similar manner. As illustrated in Fig. 2, the skewness is bounded between $(-2, 2)$, while the kurtosis is bounded by (5.6558,9), with the lower bound being calculated numerically.

3 Skew Laplace scale mixtures

In this section, we introduce some members of the SLSM models (Definition 2.2) by considering different convenient mixing PDFs, denoted by $h(w; \boldsymbol{\theta})$. These members have the ability to nest their respective reference distribution under a suitable choice of $\boldsymbol{\theta}$, which allows the use of the likelihood ratio test to determine whether the SLSM models are significant improvements over the respective SL distributions. Furthermore, the resulting SLSM distributions have closed-form PDFs and we prove that the members are unimodal. The sampling schemes described for the two SL models can also be used to generate SLSM random variables.

3.1 Mode

The SLSM distribution is unimodal, with mode μ. By taking the first derivative with respect to x of $g_{SLSM}(x;\mu,\beta,\boldsymbol{\theta},\lambda)$

$$g'_{SLSM}(x;\mu,\beta,\boldsymbol{\theta},\lambda)=\int_{S_h} f'_{SL}(x;\mu,\beta/w,\lambda)h(w;\boldsymbol{\theta})dw,\quad x>0$$
$$\equiv\sum_{w} f'_{SL}(x;\mu,\beta/w_i,\lambda)h(w_i;\boldsymbol{\theta})$$

where

$$f'_{SL_I}(x;\mu,\beta/w,\lambda)=\begin{cases}\dfrac{w^2\,(\lambda+1)\,\mathrm{e}^{\frac{w(\lambda+1)(x-\mu)}{\beta}}}{2\beta^2}, & x\le\mu,\\ \dfrac{w^2\,(\lambda+1)\,\mathrm{e}^{-\frac{w(\lambda+1)(x-\mu)}{\beta}}}{2\beta^2}-\dfrac{w^2\mathrm{e}^{-\frac{w(x-\mu)}{\beta}}}{\beta^2}, & x>\mu,\end{cases}$$

and

$$f'_{SL_{II}}(x;\mu,\beta/w,\lambda)=\begin{cases}\frac{w^2}{\beta^2}\left(\mathrm{e}^{\frac{(x-\mu)}{\beta/w}}\Phi\left(\sqrt{2}\lambda\frac{x-\mu}{\beta/w}\right)+\sqrt{2}\lambda\mathrm{e}^{\frac{(x-\mu)}{\beta/w}}\phi\left(\sqrt{2}\lambda\frac{x-\mu}{\beta/w}\right)\right), & x\le\mu,\\ \frac{w^2}{\beta^2}\left(-\mathrm{e}^{-\frac{(x-\mu)}{\beta/w}}\Phi\left(\sqrt{2}\lambda\frac{x-\mu}{\beta/w}\right)+\sqrt{2}\lambda\mathrm{e}^{-\frac{(x-\mu)}{\beta/w}}\phi\left(\sqrt{2}\lambda\frac{x-\mu}{\beta/w}\right)\right), & x>\mu,\end{cases}$$

where $\phi(\cdot)$ denotes the PDF of the standard normal distribution, and since $f'_{SL}(x;\mu,\beta/w,\lambda)>0$ for $x<\mu$, $f'_{SL}(x;\mu,\beta/w,\lambda)<0$ for $x>\mu$, and $f'_{SL}(x;\mu,\beta/w,\lambda)=0$ for $x=\mu$ for both SL_I and SL_{II}, and recalling $h(w;\boldsymbol{\theta})>0$ for $w>0$, it follows directly the SLSM distribution is unimodal, with mode μ.

Definition 3.1 If Y has a SL distribution with either PDF (6) or (13) and is independent of W, then

$$X\stackrel{d}{=}\frac{Y}{W}\tag{16}$$

has a SLSM distribution.

Let $X\sim SLSM(\mu,\beta,\lambda,\boldsymbol{\theta})$. Subsequently, from the representation given in (16), noting that Y and W are stochastically independent, the kth moment is given by

$$\mathrm{E}_{SLSM}(X^k)=\mathrm{E}_h[\mathrm{E}_{SL}\left(X^k|W=w\right)].\tag{17}$$

3.2 Skew Bernoulli Laplace distributions

Consider the following mixing Bernoulli distribution

$$W = \begin{cases} 1 & \text{with probability } \theta_1 \\ 1/\theta_2 & \text{with probability } 1-\theta_1 \end{cases} \tag{18}$$

where $\theta_1 \in (0, 1)$ and $\theta_2 > 1$, and corresponding probability mass function (PMF)

$$h_B(w; \boldsymbol{\theta}) = \theta_1^{\frac{w-1/\theta_2}{1-1/\theta_2}} (1-\theta_2)^{\frac{1-w}{1-1/\theta_2}}, \tag{19}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)$ Punzo and Bagnato (2021). It follows that the kth moment of $1/W$ is

$$\mathrm{E}\left(\frac{1}{W^k}\right) = \theta_1 + (1-\theta_1)\theta_2^k. \tag{20}$$

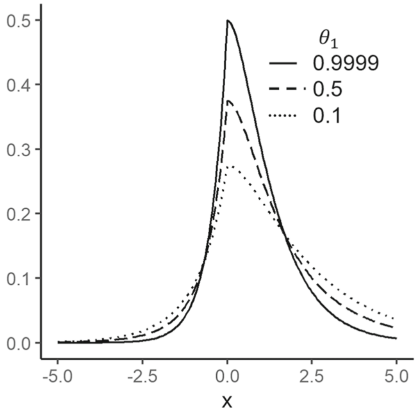
3.2.1 Skew Bernoulli Laplace type I distribution

By applying the SLSM to the SL_I distribution with PDF (6) and the Bernoulli mixing PMF $h_B(w; \boldsymbol{\theta})$ in (19), a random variable X has the skew Bernoulli Laplace type I distribution if its PDF is given by

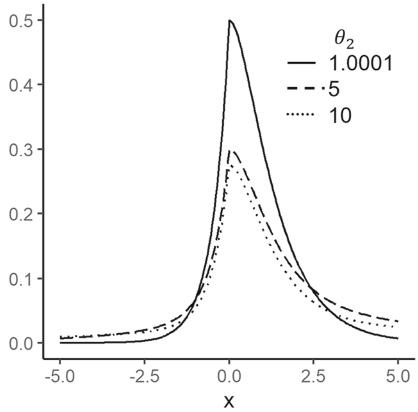
$$g_{SBL_I}(x; \mu, \beta, \theta_1, \theta_2, \lambda) = \begin{cases} \frac{\theta_1}{2\beta} \mathrm{e}^{\frac{(\lambda+1)(x-\mu)}{\beta}} + \frac{1-\theta_1}{2\beta\theta_2} \mathrm{e}^{\frac{(\lambda+1)(x-\mu)}{\beta\theta_2}}, & x \le \mu, \\ \theta_1 \left(\frac{1}{\beta} \mathrm{e}^{-\frac{(x-\mu)}{\beta}} - \frac{1}{2\beta} \mathrm{e}^{-\frac{(\lambda+1)(x-\mu)}{\beta}} \right) & \\ +(1-\theta_1) \left(\frac{1}{\beta\theta_2} \mathrm{e}^{-\frac{(x-\mu)}{\beta\theta_2}} - \frac{1}{2\beta\theta_2} \mathrm{e}^{-\frac{(\lambda+1)(x-\mu)}{\beta\theta_2}} \right), & x > \mu, \end{cases} \tag{21}$$

where $x, \mu \in (-\infty, \infty)$, $\beta > 0$, $\lambda \ge 0$, $\theta_1 \in (0, 1)$ and $\theta_2 > 1$. This is denoted by $X \sim SBL_I(\mu, \beta, \theta_1, \theta_2, \lambda)$. Figure 3 illustrates (21) for different values of θ_1 and θ_2, with expressions for the moments given in Appendix A.

The behaviour of the skewness and kurtosis, as a function of λ, are illustrated in Fig. 4 for varying values of θ_1 and θ_2. From Fig. 4a it is evident that values of θ_1 close to 1 tend to produce the skewness curve obtained for the SL_I distribution with PDF (21), as illustrated in Fig. 1a. It can be observed that as θ_1 decreases the skewness increases. Similarly, from Fig. 4c, the kurtosis curve is produced when θ_1 is close to the plot obtained for the SL_I distribution. $\theta_1 \to 1$ can thus be considered as bound for the possible range of skewness, i.e. when $\theta_1 \to 1$ the skewness lies between $(-2, 2)$, and also acts as a lower bound for the kurtosis. Figure 4b and d demonstrate the skewness and kurtosis behaviour for varying values of θ_2, respectively. When θ_2 is close to one the skewness and kurtosis of the SL_{II} distribution are obtained, and can thus be considered a practical lower bound.



(a) Varying θ_1 if $\mu = 0, \beta = 1, \lambda = 1, \theta_2 = 2$



(b) Varying θ_2 if $\mu = 0, \beta = 1, \lambda = 1, \theta_1 = 0.5$

Fig. 3 SBL_I with PDF (21) for $\mu = 0$, $\beta = 1$, $\lambda = 1$ and different values of θ_1 and θ_2

3.2.2 Skew Bernoulli Laplace type II distribution

Next, the SL_{II} distribution (22) is considered as the conditional reference distribution. A random variable X has the skew Bernoulli type II distribution if its PDF is given by

$$g_{SBL_{II}}(x; \mu, \beta, \theta_1, \theta_2, \lambda) = 2\theta_1 f_L(x; \mu, \beta)\Phi\left(\sqrt{2}\lambda\left(\frac{x-\mu}{\beta}\right)\right) + 2(1-\theta_1) f_L(x; \mu, \theta_2\beta)\Phi\left(\sqrt{2}\lambda\left(\frac{x-\mu}{\theta_2\beta}\right)\right), \tag{22}$$

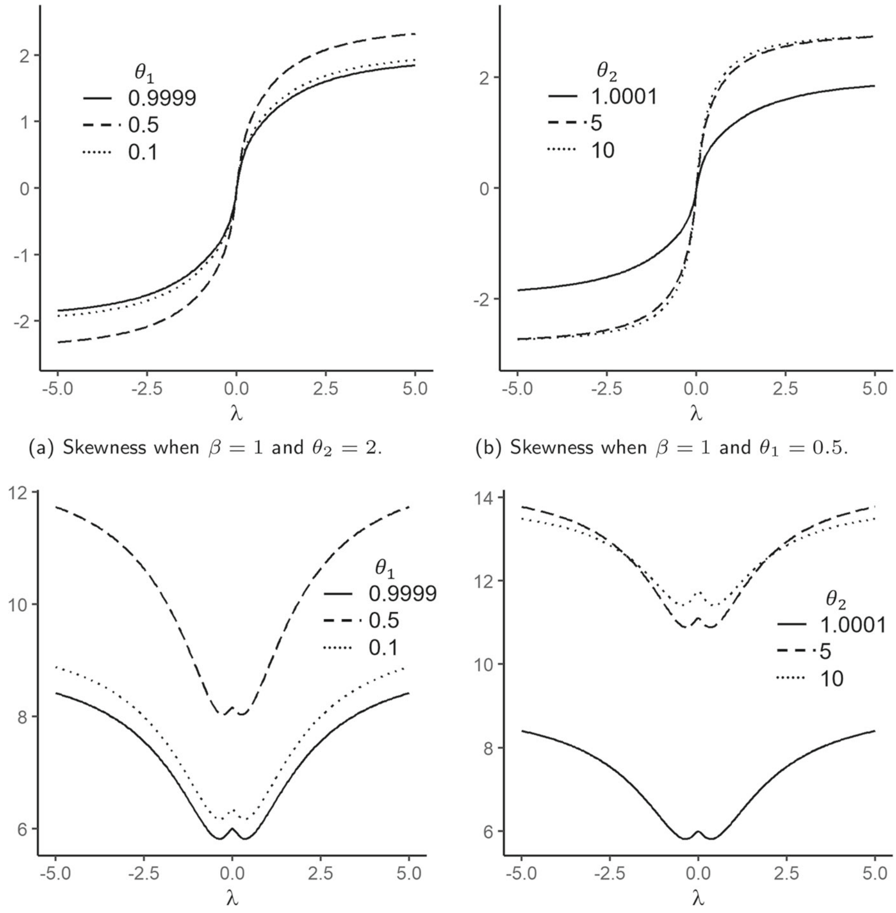
where $x, \mu, \lambda \in (-\infty, \infty)$, $\beta > 0$, $\theta_1 \in (0, 1)$, $\theta_2 > 1$ and where $f_L(\cdot)$ is given by (1). This is denoted by $X \sim SBL_{II}(\mu, \beta, \theta_1, \theta_2, \lambda)$ and illustrated for varying values of θ_1 and θ_2 in Fig. 5.

In this case, the kth moment is given as

$$\mathrm{E}_{SBL_{II}}[X^k] = \theta_1 \sum_{r=0}^{k} \binom{k}{r} \mu^{k-r} \beta^r \mathrm{E}\left[Y^r\right] + (1-\theta_1) \sum_{r=0}^{k} \binom{k}{r} \mu^{k-r} (\beta\theta_2)^r \mathrm{E}\left[Y^r\right], \tag{23}$$

where $\mathrm{E}[Y^r]$ is given in (14). Using (23), expressions for the mean, variance, skewness and kurtosis can be obtained. Figure 6 illustrates the skewness and kurtosis for varying values of θ_1 and θ_2.

The PDFs in (21) and (22) can be viewed as *contaminated* SL models, i.e. the SL counterpart of the contaminated normal distribution (Tukey, 1960). It should be noted the additional parameters θ_1 and θ_2 have an interpretation of practical interest. θ_1 is the proportion of points from the reference distribution, while the parameter θ_2 denotes the degree of contamination. Since it is common practice to assume that at least half of the observations are considered "good" in robustness studies (Punzo & Bagnato,



(a) Skewness when $\beta = 1$ and $\theta_2 = 2$.

(b) Skewness when $\beta = 1$ and $\theta_1 = 0.5$.

(c) Kurtosis when $\beta = 1$ and $\theta_2 = 2$.

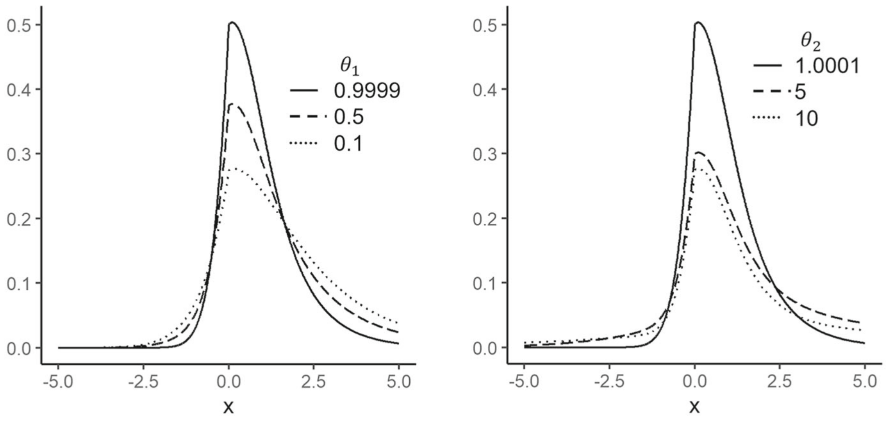
(d) Kurtosis when $\beta = 1$ and θ_1 =0.5.

Fig. 4 Skewness and kurtosis of SBL_I with PDF (21) as a function of λ

2021), we will assume that $\theta_1 \in (0.5, 1)$. Because of the assumption that $\theta_2 > 1$, it can be meant as the increase in variability due to the points which do not come from the reference distribution. Both PDFs reduce to their respective SL distributions given in (6) and (13) when $\theta_1 \to 1$ and $\theta_2 \to 1$. In addition, if $\lambda = 0$, (21) and (22) simplifies to the Bernoulli Laplace distribution developed in Punzo and Bagnato (2021). The Laplace distribution with PDF (1) is obtained if $\theta_1 \to 1$, $\theta_2 \to 1$ and $\lambda = 0$.

As discussed in Punzo and Bagnato (2021), an advantage of using a contaminated model is that, given μ, β, λ, θ_1 and θ_2, it is possible to establish whether a data point x comes from the reference distribution or not via the *a posteriori* probability

$$P(x \text{ comes from the reference distribution}) = \frac{\theta_1 f_{SL_i}(x; \mu, \beta, \lambda)}{g_{SBL_i}(x; \mu, \beta, \theta_1, \theta_2, \lambda)} \qquad (24)$$



(a) Varying θ_1 if $\mu = 0, \beta = 1, \lambda = 1, \theta_2 = 2$ (b) Varying θ_2 if $\mu = 0, \beta = 1, \lambda = 1, \theta_1 = 0.5$

Fig. 5 SBL_{II} with PDF (22) for $\mu = 0$, $\beta = 1$, $\lambda = 1$ and different values of θ_1 and θ_2

for $i = I, II$. If the *a posteriori* probability in (24) is greater than 0.5 then x can be considered to come from the respective SL distribution in (6) or (13).

3.3 Skew shifted exponential Laplace distributions

As a next candidate for the mixing distribution, consider the shifted exponential distribution with PDF given by

$$h_{SE}(w;\theta) = \theta \mathrm{e}^{-\theta(w-1)}, \quad w > 1, \tag{25}$$

where $\theta > 0$ Punzo and Bagnato (2021) and the kth moment of $1/W$ is

$$\mathrm{E}\left(\frac{1}{W^k}\right) = \theta \mathrm{e}^{\theta} \int_1^{\infty} \mathrm{e}^{-\theta t}/t^k dt = \theta \mathrm{e}^{\theta} E_k(\theta),$$

where $E_k(\theta)$ is the exponential integral function.

3.3.1 Skew shifted exponential Laplace type I distribution

If (25) is considered as the mixing PDF in (3), the skew shifted exponential Laplace type I (SSEL_I) distribution is proposed as

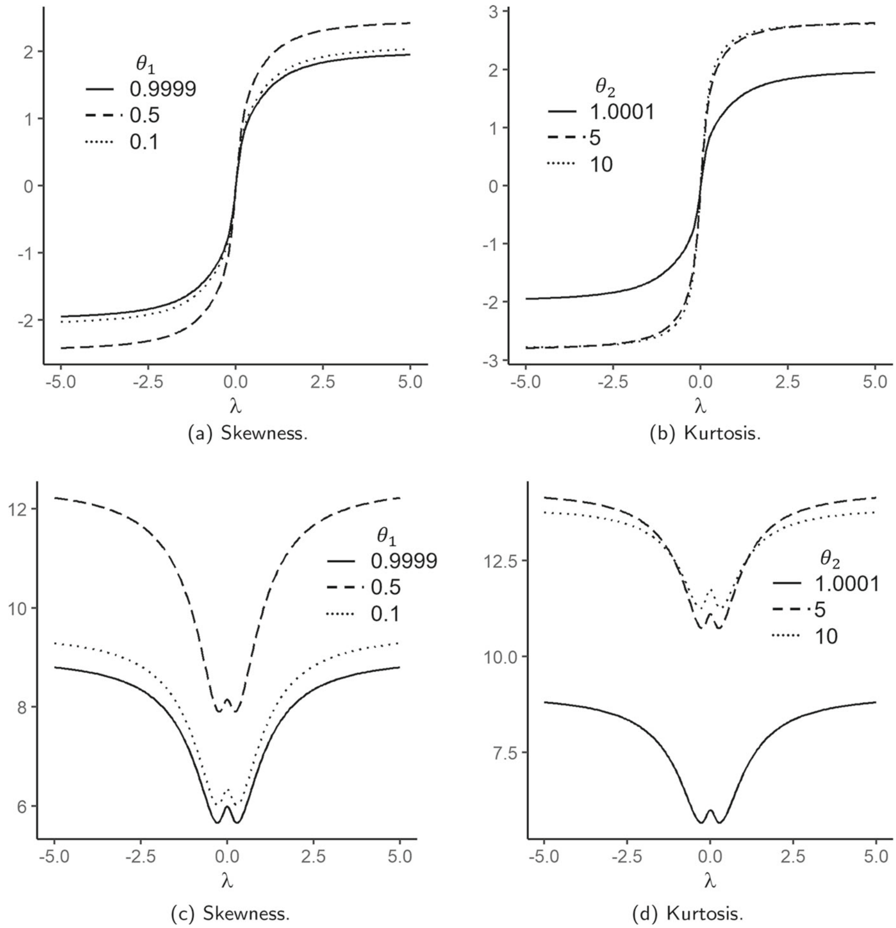


Fig. 6 Skewness and Kurtosis of SBL_{II} with PDF (22) as a function of λ

$$
g_{SSEL_I}(x;\mu,\beta,\theta,\lambda)=\begin{cases}-\dfrac{\theta\left[(x-\mu)(\lambda+1)-\beta(\theta+1)\right]\mathrm{e}^{\frac{(\lambda+1)(x-\mu)}{\beta}}}{2\left[(x-\mu)(\lambda+1)-\beta\theta\right]^2}, & x\le\mu,\\ \dfrac{\theta\left(x-\mu+\beta(\theta+1)\right)\mathrm{e}^{-\frac{x-\mu}{\beta}}}{(x-\mu+\beta\theta)^2} & \\ -\dfrac{\theta\left[(x-\mu)(\lambda+1)+\beta(\theta+1)\right]\mathrm{e}^{-\frac{(x-\mu)(\lambda+1)}{\beta}}}{2\left[(x-\mu)(\lambda+1)+\beta\theta\right]^2}, & x>\mu,\end{cases} \tag{26}
$$

where $x,\mu\in(-\infty,\infty)$, $\beta,\theta>0$ and $\lambda\ge 0$, with (6) used as the conditional (reference) distribution. Figure 7 illustrates (26) for varying values of θ, where it can be observed that the $SSEL_I$ distribution reduces to the SL_I distribution (6) as $\theta\to\infty$.

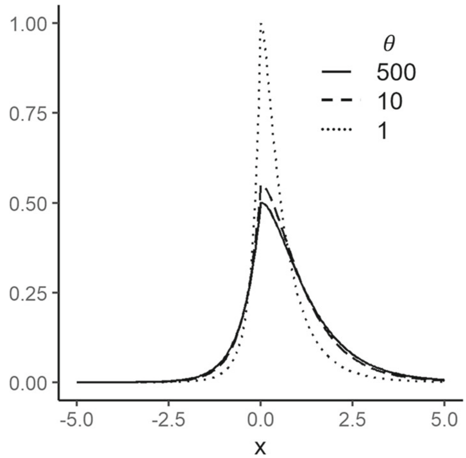


Fig. 7 $SSEL_I$ with PDF (26) for $\mu = 0, \beta = 1, \lambda = 1$ and different values of θ

Figure 17a in Appendix C illustrates the skewness bounded between (-2,2) and behaves similarly to the skewness of SL_I when $\theta \to \infty$; it is evident that the kurtosis increases as $\theta \to 0$.

3.3.2 Skew shifted exponential type II distribution

Suppose the PDF in (13) is the reference distribution in (3), then we propose the skew shifted exponential Laplace type II (SSEL$_{II}$) distribution with PDF

$$g_{SSEL_{II}}(x;\mu,\beta,\theta,\lambda) = \frac{2\theta f_L(x;\mu,\beta)}{\left(\frac{|x-\mu|}{\beta}+\theta\right)}\left(\frac{\mathrm{E}_{X_3}\left[\Phi\left(\frac{\sqrt{2}\lambda(X_3+1)(x-\mu)}{\beta}\right)\right]}{\left(\frac{|x-\mu|}{\beta}+\theta\right)} + \mathrm{E}_{X_4}\left[\Phi\left(\frac{\sqrt{2}\lambda(X_4+1)(x-\mu)}{\beta}\right)\right]\right), \tag{27}$$

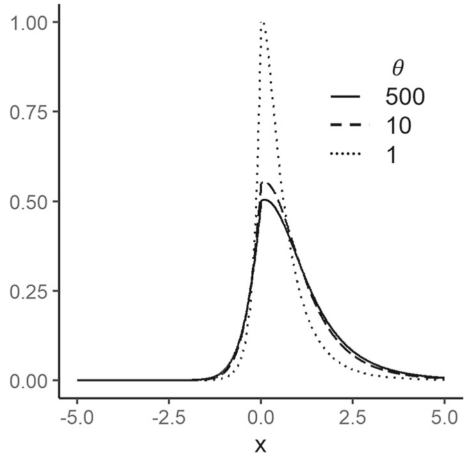
where $x, \mu, \lambda \in (-\infty, \infty)$, $\beta, \theta > 0$, $X_3 \sim Gamma\left(2, \frac{|x-\mu|}{\beta}+\theta\right)$ and $X_4 \sim Exp\left(\frac{|x-\mu|}{\beta}+\theta\right)$.

Figure 8 illustrates (27) for varying values of θ, while the skewness and kurtosis are shown in Fig. 17b and 18b, respectively. These are interpretable in a similar manner as for the $SSEL_I$ distribution. It follows that (26) and (27) reduce to (21) and (22) respectively as $\theta \to \infty$. It follows from Fig. 18b that for small values of θ, the kurtosis of the SSEL models is higher.

3.4 Skew unimodal gamma Laplace distributions

As mixing distribution we consider the unimodal gamma (UG) distribution (Chen, 2000; Punzo, 2019) with PDF

Fig. 8 $SSEL_{II}$ with PDF (27) for $\mu = 0$, $\beta = 1$, $\lambda = 1$ and different θ



$$h_{UG}(w;\theta) = \frac{w^{\frac{1}{\theta}} \mathrm{e}^{-\frac{w}{\theta}}}{\theta^{\frac{1}{\theta}+1}\Gamma\left(\frac{1}{\theta}+1\right)}, \quad w > 0, \tag{28}$$

where $\theta > 0$ and $\Gamma(\cdot)$ denotes the gamma function. It follows that the kth moment of $1/W$ is

$$\mathrm{E}\left(\frac{1}{W^k}\right) = \frac{\Gamma\left(\frac{1}{\theta}+1-k\right)}{\theta^{k-1}\Gamma\left(\frac{1}{\theta}\right)}$$

if $\theta < 1/(k-1)$.

3.4.1 Skew unimodal gamma Laplace type I distribution

From (6) and (28) it follows that X has the skew unimodal gamma Laplace type I (SUGL_I) distribution with PDF

$$g_{SUGL_I}(x;\mu,\beta,\theta,\lambda) = \begin{cases} \frac{1}{2}(\theta+1)\beta^{\frac{1}{\theta}+1}\left[\beta-\theta\left(\lambda+1\right)\left(x-\mu\right)\right]^{-\frac{1}{\theta}-2}, & x \le \mu, \\ (\theta+1)\beta^{\frac{1}{\theta}+1}\Big(\left[\beta+\theta(x-\mu)\right]^{-\frac{1}{\theta}-2} & \\ -\frac{1}{2}\left[\beta+\theta(\lambda+1)(x-\mu)\right]^{-\frac{1}{\theta}-2}\Big), & x > \mu, \end{cases} \tag{29}$$

where $x, \mu \in (-\infty, \infty)$, $\beta, \theta, > 0$ and $\lambda \ge 0$. Figure 9 illustrates the PDF (29).

From Figs. 17c and 18c it is evident that the range of possible values of both the skewness and kurtosis of (29) are increasing functions of θ and behave in a similar manner to that of the SL_I distribution when $\theta \to 0$.

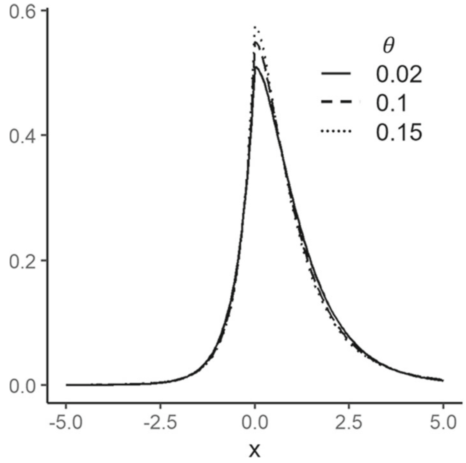


Fig. 9 $SUGL_I$ with PDF (29) for $\mu = 0, \beta = 1, \lambda = 1$ and different values of θ

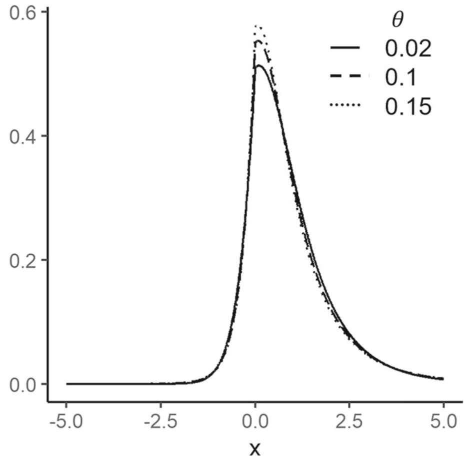


Fig. 10 $SUGL_{II}$ with PDF (30) for $\mu = 0, \beta = 1, \lambda = 1$ and different values of θ

3.4.2 Skew unimodal gamma Laplace type II distribution

Next, from (13) and (28), a random variable X has the skew unimodal gamma Laplace type II (SUGL_{II}) distribution with PDF

$$g_{SUGL_{II}}(x; \mu, \beta, \theta, \lambda) = \frac{\theta + 1}{\beta\theta^{\frac{1}{\theta}+2}\left(\frac{1}{\theta} + \frac{|x-\mu|}{\beta}\right)^{\frac{1}{\theta}+2}} \mathrm{E}_{X_5}\left[\Phi\left(\sqrt{2}\lambda X_5\left(\frac{x-\mu}{\beta}\right)\right)\right], \tag{30}$$

where $x, \mu, \lambda \in (-\infty, \infty)$, $\beta, \theta > 0$ and $X_5 \sim Gamma\left(\frac{1}{\theta} + 2, \frac{1}{\theta} + \frac{|x-\mu|}{\beta}\right)$. Figure 10 illustrates (30) for varying values of θ, and Figs. 17d and 18d the corresponding skewness and kurtosis as a function of λ.

3.5 Skew power-function Laplace distributions

Lastly, the power function (PF) distribution is considered as mixing distribution with PDF given as

$$h_{PF}(w;\theta) = \theta w^{\theta-1}, \quad 0 < w < 1, \tag{31}$$

where $\theta > 0$ Ahsanullah and Kabir (1974). The kth moment of $1/W$ is

$$\mathrm{E}\left(\frac{1}{W^k}\right) = \frac{\theta}{\theta - k},$$

which exists for $\theta > k$.

3.5.1 Skew power-function Laplace type I distribution

From (31) and (6), it follows that a random variable X has the skew power-function Laplace type I (SPFL_I) distribution with PDF given by

$$g_{SPFL_I}(x;\mu,\beta,\theta) = \begin{cases} \dfrac{\theta\left(\Gamma(\theta+1)-\Gamma\left(\theta+1,-\frac{(\lambda+1)(x-\mu)}{\beta}\right)\right)}{2\beta\left(-\frac{(\lambda+1)(x-\mu)}{\beta}\right)^{\theta+1}}, & x \le \mu, \\ \frac{\beta^\theta\theta}{(x-\mu)^{\theta+1}}\left[\Gamma(\theta+1) - \Gamma\left(\theta+1, \frac{x-\mu}{\beta}\right)\right. \\ \left. - \dfrac{\left(\Gamma(\theta+1)-\Gamma\left(\theta+1,\frac{(\lambda+1)(x-\mu)}{\beta}\right)\right)}{2(\lambda+1)^{\theta+1}}\right], & x > \mu, \end{cases} \tag{32}$$

where $x, \mu \in (-\infty, \infty)$, $\beta, \theta > 0$, $\lambda \ge 0$, $\Gamma(\cdot,\cdot)$ denotes the upper incomplete gamma function (Gradshteyn & Ryzhik, 2014). Figure 11 illustrates (32) for varying values of θ. From Figs. 17e and 18e, it is evident that as $\theta \to \infty$, the skewness and kurtosis reduces to that of the SL_I distribution with PDF (6). Conversely, the range of possible values of the skewness and kurtosis increase as $\theta \to 0$; $\theta \to \infty$ acts as a lower bound.

3.5.2 Skew power-function Laplace type II distribution

From (13) and (31) a random variable X has the SPFL_{II} distribution with PDF given by

$$g_{SPFL_{II}}(x;\mu,\beta,\theta_1,\theta_2,\lambda) = \frac{\theta\beta^\theta\gamma(\theta+1, \frac{|x-\mu|}{\beta})}{|x-\mu|^{\theta+1}}\mathrm{E}_{X_6}\left[\Phi\left(\sqrt{2}\lambda X_6\left(\frac{x-\mu}{\beta}\right)\right)\right], \tag{33}$$

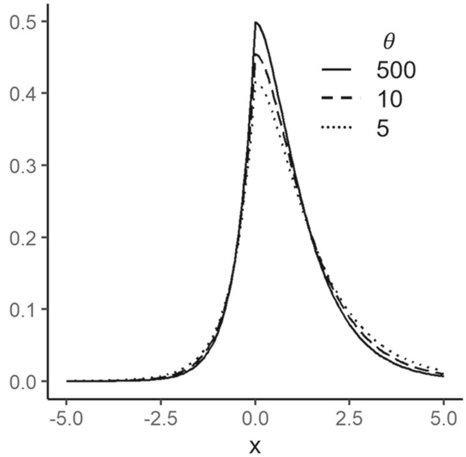


Fig. 11 $SPFL_I$ with PDF (32) for $\mu = 0, \beta = 1, \lambda = 1$ and different values of θ

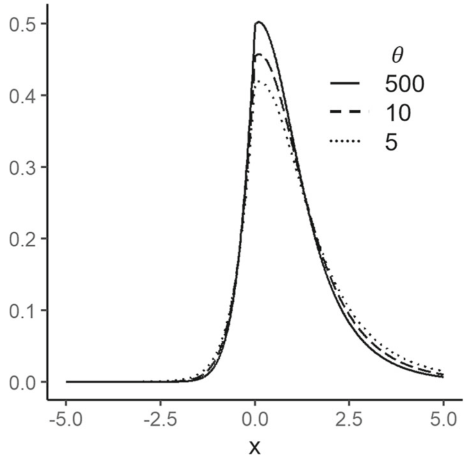


Fig. 12 $SPFL_{II}$ with PDF (33) for $\mu = 0, \beta = 1, \lambda = 1$ and different values of θ

where $x, \mu, \lambda \in (-\infty, \infty)$ and $\beta, \theta > 0$, $X_6 \sim TGamma\left(\theta + 1, \frac{\beta}{|x-\mu|}, 1\right)$. Note if Z has a truncated gamma distribution its PDF is given as

$$f(z; k, \theta, a) = \frac{z^{k-1} \mathrm{e}^{-\frac{z}{\theta}}}{\theta^k \gamma\left(k, \frac{a}{\theta}\right)}$$

where $z \in (0, a], \kappa, \theta, a > 0$ and $\gamma(\cdot, \cdot)$ denotes the lower incomplete gamma function (Gradshteyn & Ryzhik, 2014). This is denoted by $Z \sim TGamma(k, \theta, a)$.

Figure 12 illustrates (33) for varying values of θ and Figs. 17f and 18f shows the corresponding skewness and kurtosis as a function of λ.

4 Finite mixtures

In this section, the focus shifts to a finite mixture approach that is able to model multimodal data, while still being flexible enough to model skewness and heavy-tailed data. Finite mixture models provide a flexible framework for analysing a variety of data, with numerous applications, including classification, clustering, and data mining, image analysis, pattern recognition, latent class analysis, density estimation (Bishop & Nasrabadi, 2006; Frühwirth-Schnatter, 2006; McLachlan & Basford, 1988; McLachlan et al., 2019). According to Titterington et al. (1985), a broad motivation for using a finite mixture of univariate densities is that we assume that there are k underlying groups/components and that each observation belongs to one of the components. The aim is to infer the distribution for each component separately. The PDF of a k-component SLSM finite mixture is given in Definition 4.1.

Definition 4.1 The finite mixture model of k unimodal SLSM PDFs, where $g_{SLSM}(x; \mu_i, \beta_i, \boldsymbol{\theta}_i, \lambda_i)$ is the unimodal PDF of the ith component with parameters $\mu_i, \beta_i, \boldsymbol{\theta}_i, \lambda_i$ (as defined in Definition 2), is given as

$$m(x; \boldsymbol{\psi}) = \sum_{j=1}^{k} \pi_j g_{SLSM}(x; \mu_j, \beta_j, \boldsymbol{\theta}_j, \lambda_j), \quad -\infty < x < \infty, \tag{34}$$

where $\pi_j \in (0, 1]$ is the mixing proportion for the jth component and $\sum_{j=1}^{k} \pi_j = 1$, and $\boldsymbol{\psi} = (\boldsymbol{\pi}', \boldsymbol{\mu}', \boldsymbol{\beta}', \boldsymbol{\theta}', \boldsymbol{\lambda}')'$, with $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_k')'$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)'$, contains all the parameters of the finite mixture.

Finite mixtures consisting of members of the SLSM models described in Sect. 3 are fitted to a real-world data set in Sect. 6.

5 Maximum likelihood estimation

The EM algorithm developed by Dempster et al. (1977) is based on finding the maximum likelihood estimate (MLE) of the parameters and weights of a distribution. A brief description of the EM algorithm follows: for the application of the EM algorithm, it is convenient to view the observed data, denoted by x_i, as incomplete. An unobserved, also known as *latent*, random variable is introduced. Let a vector of unknown parameters be denoted by $\boldsymbol{\psi}$ and the complete data likelihood function, denoted by $L_c(\boldsymbol{\psi})$. The EM algorithm seeks to find the MLE of the data likelihood function by iteratively applying two steps until convergence, namely the expectation and maximization step.

The expectation step requires the calculation of the conditional expectation of the complete-data log-likelihood function. For the $(r+1)$th iteration, this is given as

$$Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(r)}) = \mathrm{E}[\ln L_c(\boldsymbol{\psi})].$$

In the maximization step, the parameters are found that maximize $Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(r)})$, i.e.

$$\boldsymbol{\psi}^{(r+1)} = \arg\max_{\boldsymbol{\psi}} Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(r)}).$$

A full description and further extensions of the EM algorithm can be found in McLachlan and Krishnan (2007).

The computational ease and availability of optimization routines in R software, make direct numerical maximization (DNM) a practical choice to obtain parameter estimates when the distributions are unimodal. See MacDonald (2014) and MacDonald (2021) for several examples to show the advantages of DNM over the EM algorithm. While estimates for the SSEL, SUGL and SPFL distributions in Sect. 3 can also be estimated by making use of the EM algorithm, we opted to use the DNM method to numerically optimize the parameter estimates using the `optim()` function in R R Core Team (2013), included in the **stats** package. A simulation study is conducted in Example 6.1 to investigate the parameter recovery of the DNM method.

The EM algorithm is further expanded for the SBL_I and SBL_{II} distributions with PDFs (21) and (22) in Sect. 5.1 as an illustration and in Appendix B for the finite mixture models consisting of the SLSM models since they are not unimodal.

5.1 EM algorithm for SBL distributions

If the latent variable is denoted by W, the complete data are given by $(x_1, w_1), \ldots, (x_i, w_i), \ldots, (x_n, w_n)$. It follows that the complete data likelihood function, $L_c(\mu, \beta, \theta_1, \theta_2, \lambda)$, can be factored into the product of the conditional PDFs of $X_i|w_i$ and the joint marginal PDFs of W_i

$$L_c(\mu, \beta, \theta_1, \theta_2, \lambda) = \prod_{i=1}^{n} f_{SL}(x_i; \mu, \beta/w_i, \lambda) h_B(w_i; \theta_1, \theta_2), \tag{35}$$

where $f_{SL}(\cdot)$ is given in either (6) or (13) and $h_B(\cdot)$ in (19). From (35), the complete-data log-likelihood function can be written as

$$l_c(\mu, \beta, \theta_1, \theta_2, \lambda) = \sum_{i=1}^{n} \ln\left[f_{SL}(x_i; \mu, \beta/w_i, \lambda) h_B(w_i; \theta_1, \theta_2)\right]. \tag{36}$$

Consider the linear transformation of the missing variable W in (18)

$$\begin{aligned} V &= \frac{W - 1/\theta_2}{1 - 1/\theta_2} \\ &= \begin{cases} 1 & \text{with probability } \theta_1 \\ 0 & \text{with probability } 1 - \theta_1. \end{cases} \end{aligned}$$

In this case, v_i acts as an indicator variable where $v_i = 1$ if x_i comes from the reference distribution and $v_i = 0$ otherwise. The complete data are thus given by

$(x_1, v_1), \ldots, (x_i, v_i), \ldots, (x_n, v_n)$ and the complete-data likelihood function in (35) can be rewritten as

$$L_c(\mu, \beta, \theta_1, \theta_2, \lambda) = \prod_{i=1}^{n} [\theta_1 f_{SL}(x_i; \mu, \beta, \lambda)]^{v_i} [(1-\theta_1) f_{SL}(x_i; \mu, \beta\theta_2, \lambda)]^{1-v_i} \tag{37}$$

and the complete-data log likelihood function in (36) as

$$\begin{aligned} l_c(\mu, \beta, \theta_1, \theta_2, \lambda) &= \sum_{i=1}^{n} [v_i \ln\theta_1 + v_i \ln f_{SL}(x_i; \mu, \beta, \lambda) \\ &\quad + (1-v_i)\ln(1-\theta_1) + (1-v_i)\ln f_{SL}(x_i; \mu, \beta\theta_2, \lambda)] \\ &= l_{1c}(\mu, \beta, \theta_2, \lambda) + l_{2c}(\theta_1) \end{aligned} \tag{38}$$

where

$$l_{1c}(\mu, \beta, \theta_2, \lambda) = \sum_{i=1}^{n} [v_i \ln f_{SL}(x_i; \mu, \beta, \lambda) + (1-v_i)\ln f_{SL}(x_i; \mu, \beta\theta_2, \lambda)]$$

and

$$l_{2c} = \sum_{i=1}^{n} [v_i \ln\theta_1 + (1-v_i)\ln(1-\theta_1)].$$

In the expectation step, the conditional expectation of the complete-data log likelihood function given in (38), is denoted by

$$\begin{aligned} & Q(\mu, \beta, \theta_1, \theta_2, \lambda | \mu^{(r)}, \beta^{(r)}, {\theta_1}^{(r)}, {\theta_2}^{(r)}, \lambda^{(r)}) \\ &= Q_1(\mu, \beta, \theta_2, \lambda, | \mu^{(r)}, \beta^{(r)}, {\theta_1}^{(r)}, {\theta_2}^{(r)}, \lambda^{(r)}) \\ &\quad + Q_2(\theta_1 | \mu^{(r)}, \beta^{(r)}, {\theta_1}^{(r)}, {\theta_2}^{(r)}, \lambda^{(r)}) \end{aligned} \tag{39}$$

for the $(r+1)th$ iteration, which is in the same order as (38). From (24), it follows that the expected *a posteriori* probability for x_i to come from the reference distribution is given as

$$\begin{aligned} \mathrm{E}(V_i | X_i = x_i) &= \frac{{\theta_1}^{(r)} g_{SL}(x_i; \mu^{(r)}, \beta^{(r)}, \lambda^{(r)})}{{\theta_1}^{(r)} g_{SL}(x_i; \mu^{(r)}, \beta^{(r)}, \lambda^{(r)}) + (1-{\theta_1}^{(r)}) g_{SL}(x_i; \mu^{(r)}, \beta^{(r)}{\theta_2}^{(r)}, \lambda^{(r)})} \\ &=: {v_i}^{(r)}. \end{aligned} \tag{40}$$

By substituting ${v_i}^{(r)}$ in (38), the expected complete log likelihood in (39) is obtained.

In the maximization step, the update for ${\theta_1}^{(r+1)}$ is calculated independently by maximizing

$Q_2(\theta_1|\mu^{(r)}, \beta^{(r)}, \lambda^{(r)}, \theta_1{}^{(r)}, \theta_2{}^{(r)})$ with respect to θ_1. It follows that $\hat{\theta}_1^{(r+1)} = \frac{1}{n}\sum_{i=1}^{n} v_i{}^{(r)}$. Similarly, updates for $\mu^{(r+1)}$, $\beta^{(r+1)}$, $\theta_2{}^{(r+1)}$ and $\lambda^{(r+1)}$ are calculated by numerically maximizing
$Q_1(\mu, \beta, \theta_2, \lambda, |\mu^{(r)}, \beta^{(r)}, \lambda^{(r)}, \theta_1{}^{(r)}, \theta_2{}^{(r)})$.

An observation x is considered to come from the reference distribution if $v_i > 0.5$. Such a point can be referred to as a good point and as a bad point otherwise. This can be equivalently defined in terms of discriminant functions

$$D_{good}(x; \mu, \beta, \theta_1, \lambda) = \theta_1 g_{SL}(x; \mu, \beta, \lambda) \tag{41}$$

and

$$D_{bad}(x; \mu, \beta, \theta_1, \theta_2, \lambda) = (1-\theta_1) g_{SL}(x; \mu, \beta\theta_2, \lambda) \tag{42}$$

so that x will be classified as good if

$$D_{good}(x; \mu, \beta, \theta_1, \lambda) > D_{bad}(x; \mu, \beta, \theta_1, \theta_2, \lambda)$$

and bad otherwise. Solving the intersection points of the discriminant functions in (41) and (42), delimits the good and bad regions.

5.2 Initialization of the estimation process

It is well known that both the EM algorithm and DNM suffer from locally optimal solutions and the final estimates are dependent on the initial starting values, as pointed out by Mahdavi et al. (2023). To address this, we suggest using the following as initial values

$$\begin{aligned}
\mu^{(0)} &= \text{median}(x), \\
\beta^{(0)} &= \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \mu^{(0)}\right|, \\
\lambda^{(0)} &= \text{empirical skewness}(x).
\end{aligned}$$

For $\boldsymbol{\theta}$, we recommend initial values that make the SLSM model tend to the corresponding SL model. In the case of SBL_I for example, we suggest initializing $\theta_1^{(0)} \to 1^-$ and $\theta_2^{(0)} \to 1^+$.

6 Application

In this section a simulation study is conducted to investigate the parameter recovery of the DNM described in Sect. 5. The SLSM distributions developed in Sect. 3 are then applied to real-world data sets, namely the bodily injury claims data set (Rempala & Derrig, 2005) and the refractive index of glass fragments. The SLSM are compared

Table 1 Simulation results for generated SUGL_I data

	n	100	200	500
Empirical mean	$\hat{\mu}$	0.0023	– 0.0077	– 0.0067
	$\hat{\beta}$	1.9972	2.0198	2.0051
	$\hat{\lambda}$	2.2420	2.1691	2.0951
	$\hat{\theta}$	0.5097	0.5127	0.4942
Bias	$\hat{\mu}$	0.0023	– 0.0077	– 0.0067
	$\hat{\beta}$	– 0.0028	0.0198	0.0051
	$\hat{\lambda}$	0.2420	0.1691	0.0951
	$\hat{\theta}$	0.0097	0.0127	– 0.0058
Standard error	$\hat{\mu}$	0.1656	0.1151	0.0677
	$\hat{\beta}$	0.3443	0.2470	0.1585
	$\hat{\lambda}$	1.1184	0.7672	0.4189
	$\hat{\theta}$	0.3159	0.2302	0.1358
MSE	$\hat{\mu}$	0.0274	0.0133	0.0046
	$\hat{\beta}$	0.1186	0.0614	0.0251
	$\hat{\lambda}$	1.3093	0.6171	0.1846
	$\hat{\theta}$	0.0999	0.0532	0.0185

to the corresponding ALSM distributions (Punzo & Bagnato, 2022a) formed with the same mixing distributions, and to other flexible distributions: skew normal (SN) (Azzalini, 1985), SGN, skew t (ST) and skew generalised t (SGT) (Theodossiou, 1998) distribution. Finite mixtures consisting of the members of the SLSM distributions are fit to campaign finance scores that measure the ideology of donors to American political candidates. The Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978) are employed for model comparison and are ranked to aid the reader in comparing the performance of the different fitted models - in the spirit of Morris et al. (2019).

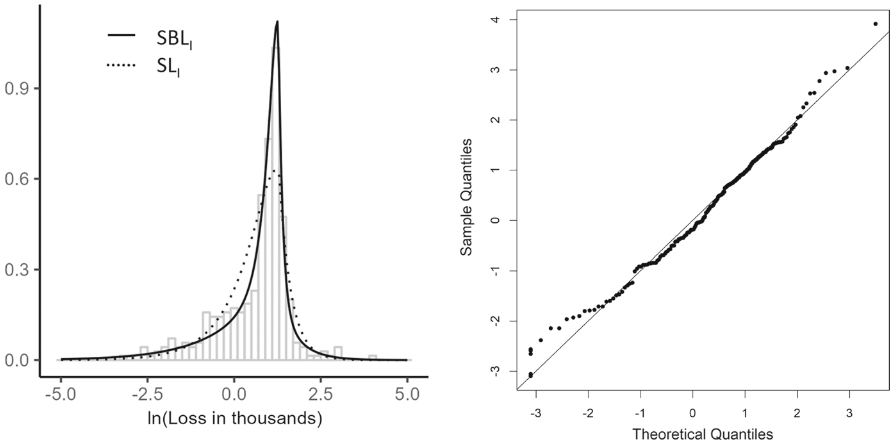
6.1 SUGL simulation study

As an illustration of the ability of the DNM approach to recover parameter estimates of the SLSM distributions, a simulation study on the $SUGL_I$ distribution is considered. Samples were generated using the sampling scheme described in Definition 3.1 for sample sizes of $n = 100, 200, 500$. Each of these instances was repeated 500 times in order to determine and observe the bias, standard error and mean squared error (MSE) of the estimates. For the simulation we consider $\mu = 0$, $\beta = 2$, $\lambda = 2$ and $\theta = 0.5$. The empirical results are displayed in Table 1. As expected, the bias is consistently small and the MSE decreases as the sample size increases.

Similar studies can be performed for the other SLSM models. This example proves that DNM can be used successfully for estimating the parameters of the different models.

Table 2 Descriptive statistics of the log bodily injury claim amounts

Number of observations	348
Mean	0.623
Standard deviation	1.028
Skewness	− 1.057
Kurtosis	4.678
Minimum	− 3.101
Maximum	3.912



(a) Histogram of the log bodily injury claim amounts with fitted SBL_I and SL_I PDFs superimposed.

(b) Quantile-quantile SBL_I plot.

Fig. 13 Histogram of log bodily injury claim amounts and SBL_I quantile-quantile plot

6.2 Bodily injury claims

The data set consists of 348 observations, i.e. claim amounts, given in thousands and is available in the **CASdatasets** package in R Dutang and Charpentier (2018). Insurance data are frequently skewed and leptokurtic (Ibragimov et al., 2015), once again emphasizing the need for flexible distributions. In analogy with Rempala and Derrig (2005), we considered the log of the claim amounts. From the descriptive statistics in Table 2 there is an indication that the data exhibit heavy-tail behaviour in the log of the claim amounts, as well as skewness. Furthermore, peakedness in the data is also evident from the histogram in Fig. 13a.

Table 3 shows that both the AIC and BIC indicate that (21) fits the best to the data (represented by bold values). Furthermore, 3 members of the SLSM models presented in this paper claim the first three spots. The fitted PDF (21) and the corresponding conditional SL_I distribution (13) are overlaid in Fig. 13a, while the quantile–quantile (Q–Q) SBL_I is given in Fig. 13b. From the linearity of the points in the Q–Q plot, it is evident that the SBL_I can be considered a valid candidate.

Table 3 Results of fitted models to the log bodily injury claim amounts

Model	#par	Log likelihood	AIC	Rank	BIC	Rank
Laplace	2	– 471.811	– 947.622	19	– 955.326	19
SL_I	3	– 435.377	– 876.754	17	– 888.311	17
SL_{II}	3	– 439.962	– 885.924	18	– 897.481	18
*SBL_I	5	– 414.422	**– 838.844**	1	**– 858.105**	1
*SBL_{II}	5	– 414.791	– 839.583	2	– 858.844	3
*$SSEL_I$	4	– 417.710	– 843.421	3	– 858.830	2
*$SSEL_{II}$	4	– 418.334	– 844.669	5	– 860.077	5
*$SUGL_I$	4	– 423.057	– 854.115	10	– 869.523	10
*$SUGL_{II}$	4	– 424.100	– 856.200	12	– 871.609	12
*$SPFL_I$	4	– 428.961	– 865.922	14	– 881.331	14
*$SPFL_{II}$	4	– 430.909	– 869.819	15	– 885.228	15
AL	3	– 427.411	– 860.821	13	– 872.378	13
ABL	5	– 418.135	– 846.270	6	– 865.531	7
ASEL	4	– 418.097	– 844.194	4	– 859.603	4
AUGL	4	– 420.764	– 849.528	8	– 864.937	6
APFL	4	– 423.543	– 855.087	11	– 870.496	11
SN	3	– 476.978	– 959.957	20	– 971.513	20
SGN	4	– 422.487	– 852.974	9	– 868.383	9
ST	4	– 432.210	– 872.419	16	– 887.828	16
SGT	5	– 418.455	– 846.911	7	– 866.172	8

6.3 Forensic glass fragments

Various types of glass are made of different elements and have distinct refractive indices. The refractive index (RI) determines how much light is bent, or refracted, when entering a material and plays an important role in differentiating types of glass. The SLSM models are fit to the RI of 214 glass fragments that are available in the fgl dataset in the **MASS** package in R Ripley et al. (2013) (summary statistics reported in Table 4).

Table 5 shows that both the AIC and BIC indicate that the $SSEL_{II}$ (27) fits the data the best (represented by bold values), supported by the Q–Q plot. The fitted PDF (26) and the corresponding conditional SL_{II} are superimposed in Fig. 14a.

6.4 Measures of Donor Ideology

Ideology plays a central role in American politics, influencing the preferences and behaviours of political contributors. Generally, donors are more inclined to support candidates whose ideological positions align closely with their own. This phenomenon can be quantified using what are termed common-space campaign finance scores (CFs-

Table 4 Descriptive statistics of the forensic glass fragments refractive index

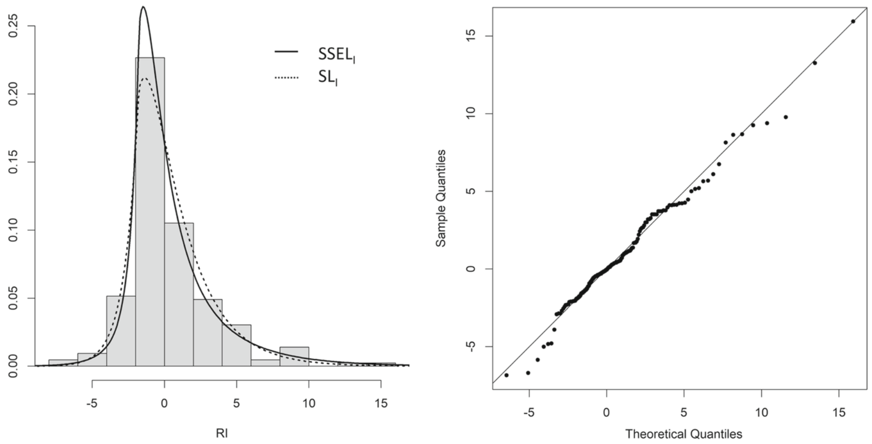
Number of observations	214
Mean	0.365
Standard deviation	3.037
Skewness	1.614
Kurtosis	7.789
Minimum	– 6.850
Maximum	15.930

Table 5 Results of fitted models to the forensic glass fragments refractive index

Model	#par	Log likelihood	AIC	Rank	BIC	Rank
Laplace	2	– 511.983	– 1027.966	19	– 1034.698	19
SL_I	3	– 496.895	– 999.790	15	– 1009.888	5
SL_{II}	3	– 500.214	– 1006.429	18	– 1016.526	17
*SBL_I	5	– 492.739	– 995.478	4	– 1012.307	14
*SBL_{II}	5	– 492.531	– 995.062	2	– 1011.891	12
*$SSEL_I$	4	– 493.319	– **994.637**	1	– **1008.101**	1
*$SSEL_{II}$	4	– 493.898	– 995.796	5	– 1009.260	4
*$SUGL_I$	4	– 493.580	– 995.159	3	– 1008.623	3
*$SUGL_{II}$	4	– 494.550	– 997.099	8	– 1010.563	8
*$SPFL_I$	4	– 494.445	– 996.891	7	– 1010.355	7
*$SPFL_{II}$	4	– 495.336	– 998.672	14	– 1012.136	13
AL	3	– 496.018	– 998.035	11	– 1008.133	2
ABL	5	– 494.898	– 999.795	16	– 1016.625	18
ASEL	4	– 495.129	– 998.258	12	– 1011.722	11
AUGL	4	– 494.695	– 997.390	9	– 1010.854	9
APFL	4	– 495.008	– 998.017	10	– 1011.481	10
SN	3	– 522.070	– 1050.141	20	– 1060.238	20
SGN	4	– 497.200	– 1002.399	17	– 1015.863	16
ST	4	– 494.381	– 996.762	6	– 1010.226	6
SGT	5	– 494.202	– 998.404	13	– 1015.234	15

cores), as described in Bonica (2014). These scores help estimate the ideal ideological point of donors. Data from the Database on Ideology, Money in Politics, and Elections (DIME): Public version 3.0 Bonica (2023) reveal notable patterns in the ideological distribution of political contributors across various industries. Some industries have donors that predominantly lean to the left, while others lean to the right. Historically, industries aligning with the Democratic Party have tended to support left-leaning policies, while those aligned with the Republican Party have favoured right-leaning ones.

Such industry-level ideological measures can shed light on the broader political behaviours of industries. For example, if we were to examine the CFscores of donors



(a) Histogram of the refractive index of forensic glass fragments with fitted SSEL$_I$ and SL$_I$ PDFs superimposed.

(b) Quantile-quantile SSEL$_I$ plot.

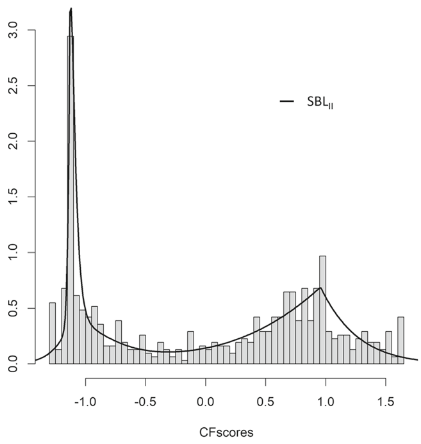
Fig. 14 Histogram of the refractive index of forensic glass fragments refractive index and SSEL$_I$ quantile-quantile plot

within the "pharmaceuticals" industry, we might gain insights into their general political inclinations. In order to better analyze and distinguish these patterns, we fit a SLSM finite mixture model with $k = 2$ components. This approach allows for clearer differentiation between the left and right-leaning tendencies of the industries under observation. The AIC and BIC values of the fitted finite mixtures are shown in Table 6. From this table, it's evident that the ALSM models outperformed our proposed models (represented by bold values). However, the SBL$_{II}$ model ranked second based on the AIC and fourth based on the BIC and the Q-Q plot further indicates that the SBL$_{II}$ remains a viable alternative model. It should be noted that the ALSM models discussed in the paper by Punzo and Bagnato (2022a) have not been previously examined within the framework of finite mixture models (Fig. 15).

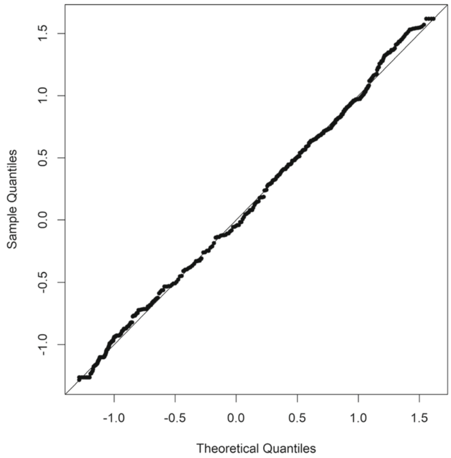
7 Final thoughts

7.1 Conclusion

Real data are often skewed and "contaminated" by outliers, highlighting the need for flexible models that are able to account for their unique characteristics. In this paper, a collection of flexible skew Laplace scale mixture (SLSM) distributions is developed by considering different skew Laplace candidates, formed with Azzalini's method of introducing skewness to symmetric distributions, and also various convenient discrete and continuous mixing variables. The resulting distributions have closed-form probability distribution functions and are able to model data with heavy-tails, skewness and peakedness. The flexibility is further extended by developing finite mixture models that are able to model multimodal data. The maximum likelihood estimates are obtained via an expectation-maximization algorithm and direct optimization and compared via



(a) Histogram of CFscores with fitted finite mixture of two SBL$_{II}$ PDFs superimposed.



(b) Quantile-quantile plot of finite mixture of two SBL$_{II}$.

Fig. 15 Histogram of CFscores and corresponding quantile–quantile plot of finite mixture of two SBL$_{II}$ PDFs

Table 6 Results of fitted mixtures with $k = 2$ components to the CFscores of donors within the pharmaceutical industry

Model	#*par*	Log likelihood	AIC	Rank	BIC	Rank
SL$_I$	7	− 546.492	− 1106.985	14	− 1137.970	14
SL$_{II}$	7	− 539.617	− 1093.234	13	− 1124.220	10
*SBL$_I$	11	− 521.220	− 1064.439	4	− 1113.131	7
*SBL$_{II}$	11	− 519.837	− 1061.675	2	− 1110.366	4
*SSEL$_I$	9	− 529.025	− 1076.050	8	− 1115.889	8
*SSEL$_{II}$	9	− 526.989	− 1071.977	7	− 1111.816	6
*SUGL$_I$	9	− 534.372	− 1086.744	9	− 1126.583	11
*SUGL$_{II}$	9	− 525.947	− 1069.893	5	− 1109.732	3
*SPFL$_I$	9	− 551.976	− 1121.952	15	− 1161.791	15
*SPFL$_{II}$	9	− 535.720	− 1089.439	10	− 1129.278	12
AL	7	− 539.279	− 1092.558	12	− 1123.544	9
ABL	11	− 517.079	**− 1056.157**	1	− 1104.848	2
ASEL	9	− 522.255	− 1062.255	3	**− 1102.348**	1
AUGL	9	− 526.436	− 1070.871	6	− 1110.709	5
APFL	9	− 536.302	− 1090.604	11	− 1130.442	13
SN	7	− 575.436	− 1146.872	17	− 1177.857	17
ST	7	− 567.616	− 1131.231	16	− 1166.643	16

the AIC, BIC and the likelihood ratio test. The SLSM distributions represent valid alternatives to other flexible distributions, as shown in Sect. 6, where several of the proposed models outperform the ALSM developed in Punzo and Bagnato (2022a).

The novelty of the SLSM models is not limited to the illustrated distributions in this paper; see Kozubowski and Nadarajah (2010) for other unconsidered generalised Laplace distributions that can be considered as a conditional reference distribution. Similarly, other mixing distributions can be considered to further extend the collection of distributions. An example includes the reparameterised inverse Gaussian distribution introduced in Punzo (2019), with PDF

$$h(w, \theta) = \sqrt{\frac{3\theta + 1}{2\pi\theta w^3}} e^{-\frac{(w-\sqrt{3\theta+1})^2}{2\theta w}}, \quad w > 0$$

where $\theta > 0$, although a multitude of other mixing distributions may be considered in future.

A Bayesian parameter estimation scheme is worth investigating as a valuable alternative to the EM algorithm in future work. By fitting mixtures of the shifted asymmetric Laplace distribution, Fang et al. (2023) illustrated that a Bayesian approach is particularly useful as it solves the "infinite likelihood problem" that might arise and thus produces better estimates in certain cases.

7.2 Future work

As a starting point for future work, a multivariate extension of the skew Bernoulli Laplace distribution is developed that is capable of modelling data when the dependence in the sample is non-negligible. Arslan (2010) proposed a p-dimensional multivatiate skew Laplace distribution with PDF

$$f_{MSL}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}) = \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{2^p \pi^{(p-1)/2} \alpha \Gamma\left(\frac{p+1}{2}\right)} e^{-\alpha\sqrt{(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} + (\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}} \tag{43}$$

where $\boldsymbol{x} : p \times 1$, $\boldsymbol{\mu} : p \times 1 \in \mathbb{R}^p$, $p \geq 1$, $\boldsymbol{\Sigma} : p \times p$ is a positive definite matrix, $\boldsymbol{\gamma} : p \times 1 \in \mathbb{R}^p$ is the skewness parameter and $\alpha = \sqrt{1 + \boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}}$. This is denoted by $\boldsymbol{X} \sim MSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$.

Definition 7.1 Let X be a p-dimensional continuous random vector. The distribution of X is in the multivariate SLSM class with location parameter $\boldsymbol{\mu}$, dispersion matrix $\boldsymbol{\Sigma}$, skewness parameter $\boldsymbol{\gamma}$, and if there is a random variable $W \sim h(\cdot|\boldsymbol{w})$, a univariate CDF indexed by the parameter vector $\boldsymbol{w}$, such that $\boldsymbol{X}|W = w \sim MSLSM_p\left(\boldsymbol{\mu}, \kappa(w)\boldsymbol{\Sigma}, \boldsymbol{\gamma}\right)$ with PDF

$$f_{MSLSM}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \boldsymbol{w}) = \int_{S_h} f_{MSL}(\boldsymbol{x}; \boldsymbol{\mu}, \kappa(w)\boldsymbol{\Sigma}, \boldsymbol{\gamma}) h(\boldsymbol{w}; \boldsymbol{\theta}) dw \tag{44}$$

for all $\boldsymbol{x}, \boldsymbol{\gamma} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} > 0$ and some positive scale function $\kappa(w)$ Punzo and Bagnato (2022b).

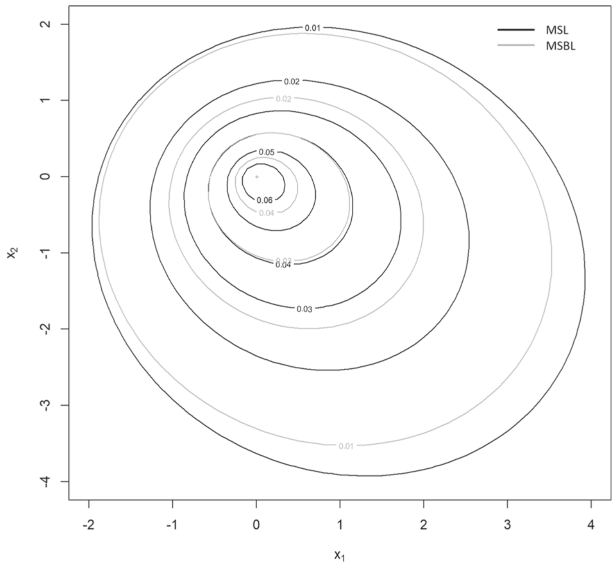


Fig. 16 Contour plots of (45) and (43) where $\boldsymbol{\mu} = (0, 0)^T$, $\boldsymbol{\Sigma} = 2\boldsymbol{I}$, $\boldsymbol{\gamma} = (-0.5, 0.5)^T$, $\theta_1 = 0.6$ and $\theta_2 = 4$

From (43) and Definition 7.1 and considering the Bernoulli as a mixing distribution, the multivariate skew Bernoulli Laplace distribution follows with PDF

$$\begin{aligned} f_{MSBL}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \theta_1, \theta_2) &= \theta_1 f_{MSL}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}) + (1-\theta_1) f_{MSL}(\boldsymbol{x}; \boldsymbol{\mu}, \theta_2\boldsymbol{\Sigma}) \\ &= \frac{\theta_1 |\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{2^p \pi^{(p-1)/2} \alpha \Gamma\left(\frac{p+1}{2}\right)} \\ &\quad \mathrm{e}^{-\alpha\sqrt{(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}+(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}} \\ &\quad + \frac{(1-\theta_1)|\theta_2\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{2^p \pi^{(p-1)/2} \alpha \Gamma\left(\frac{p+1}{2}\right)} \\ &\quad \mathrm{e}^{-\alpha\sqrt{(\boldsymbol{x}-\boldsymbol{\mu})^T(\theta_2\boldsymbol{\Sigma})^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}+(\boldsymbol{x}-\boldsymbol{\mu})^T(\theta_2\boldsymbol{\Sigma})^{-1}\boldsymbol{\gamma}} \end{aligned} \tag{45}$$

where $\boldsymbol{x} : p \times 1$, $\boldsymbol{\mu} : p \times 1 \in \mathbb{R}^p$, $p \geq 1$, $\boldsymbol{\Sigma} : p \times p$ is a positive definite matrix, $\theta_1 \in (0, 1)$ and $\theta_2 > 1$, $\boldsymbol{\gamma} : p \times 1 \in \mathbb{R}^p$ is the skewness parameter and $\alpha = \sqrt{1 + \boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}$, denoted by $\boldsymbol{X} \sim MSBL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \theta_1, \theta_2)$. Figure 16 illustrates (45) and (43) for $p = 2$.

It is clear that the MSBL distribution seems to be a promising starting point for future studies, while an investigation into multivariate extensions for other mixing distributions and their parameter estimation schemes, may also be considered. See Morris et al. (2019) for mixtures of multivariate contaminated shifted asymmetric Laplace distributions.

Appendix A: *SBL$_I$* moments

From (8), (17) and (20) it follows that

$$\mathrm{E}_{SBL_I}(X) = \beta\,(\theta_1 + (1-\theta_1)\,\theta_2)\left(1 - \frac{1}{(\lambda+1)^2}\right) + \mu,$$

$$\mathrm{Var}_{SBL_I}(X) = \beta^2\left[2\left(\theta_1 + (1-\theta_1)\,\theta_2^2\right) - (\theta_1 + (1-\theta_1)\,\theta_2)^2\left(1 - \frac{1}{(\lambda-1)^2}\right)^2\right],$$

$$\begin{aligned}\mathrm{Skewness}_{SBL_I}(X) = 2\Bigg(&(\theta_1 + (1-\theta_1)\,\theta_2)^3\left(1 - \frac{1}{(\lambda+1)^2}\right)^3\\ &+3\left((1-\theta_1)\,\theta_2^3 + \theta_1\right)\left(1 - \frac{1}{(\lambda+1)^4}\right)\\ &-3\left((1-\theta_1)\,\theta_2^2 + \theta_1\right)(\theta_1 + (1-\theta_1)\,\theta_2)\left(1 - \frac{1}{(\lambda+1)^2}\right)\Bigg)\\ &\times \frac{1}{\left(2\left((1-\theta_1)\,\theta_2^2 + \theta_1\right) - (\theta_1 + (1-\theta_1)\,\theta_2)^2\left(1 - \frac{1}{(\lambda+1)^2}\right)^2\right)^{3/2}},\end{aligned}$$

$$\begin{aligned}\mathrm{Kurtosis}_{SBL_I}(X) = &-3\,(\theta_1 + (1-\theta_1)\,\theta_2)^4\left(1 - \frac{1}{(\lambda+1)^2}\right)^4\\ &+ 12\left((1-\theta_1)\,\theta_2^2 + \theta_1\right)(\theta_1 + (1-\theta_1)\,\theta_2)^2\left(1 - \frac{1}{(\lambda+1)^2}\right)^2\\ &- 24\left((1-\theta_1)\,\theta_2^3 + \theta_1\right)\\ &(\theta_1 + (1-\theta_1)\,\theta_2)\left(1 - \frac{1}{(\lambda+1)^4}\right)\left(1 - \frac{1}{(\lambda+1)^2}\right)\\ &+ 24\left((1-\theta_1)\,\theta_2^4 + \theta_1\right)\\ \times &\frac{1}{(\theta_1 + (1-\theta_1)\,\theta_2)^4\left(1 - \frac{1}{(\lambda+1)^2}\right)^4 - 4\left((1-\theta_1)\,\theta_2^2 + \theta_1\right)(\theta_1 + (1-\theta_1)\,\theta_2)^2\left(1 - \frac{1}{(\lambda+1)^2}\right)^2 + 4\left((1-\theta_1)\,\theta_2^2 + \theta_1\right)^2}.\end{aligned}$$

Appendix B: EM algorithm for finite mixtures

Estimates for the finite mixtures described in Sect. 5 can be calculated by the EM algorithm. It follows that by denoting the latent variable as $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{ik})$, which acts as an indicator vector, where $z_{ij} = 1$ if x_i comes for the jth component and $z_{ij} = 0$ otherwise, the complete-data likelihood can be written as

$$L_c(\boldsymbol{\psi}) = \prod_{i=1}^{n}\prod_{j=1}^{k}\left[\pi_j\, g_{SLSM}(x_i; \boldsymbol{\theta}_j)\right]^{z_{ij}} \tag{46}$$

and from (46), the complete-data log likelihood as

$$l_c(\boldsymbol{\psi}) = l_{1c}(\boldsymbol{\theta}) + l_{2c}(\boldsymbol{\pi}) \tag{47}$$

where

$$l_{1c}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij}\ln[g_{SLSM}(x_i;\boldsymbol{\theta}_j)]$$

and

$$l_{2c}(\boldsymbol{\pi}) = \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij}\ln\pi_j.$$

In order to calculate the conditional expectation of the complete-data log likelihood function

$$Q(\boldsymbol{\theta},\boldsymbol{\pi}|\boldsymbol{\psi}^{(r)}) = Q_1(\boldsymbol{\theta}|\boldsymbol{\psi}^{(r)}) + Q_2(\boldsymbol{\pi}|\boldsymbol{\psi}^{(r)})$$

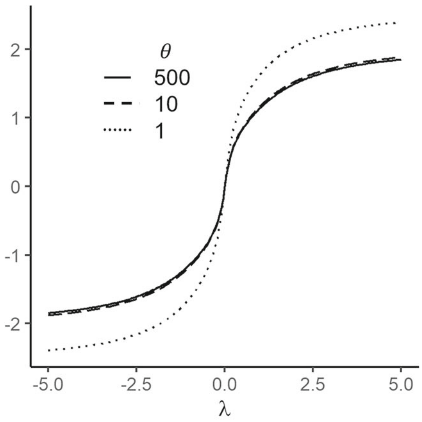
we need to substitute

$$\begin{aligned}\mathrm{E}(Z_{ij}|x_i) &= \frac{\pi_j^{(r)} g_{SLSM}(x_i;\boldsymbol{\theta}_j^{(r)})}{m(\boldsymbol{\psi}^{(r)})}\\ &=: z_{ij}^{(r)}\end{aligned}$$

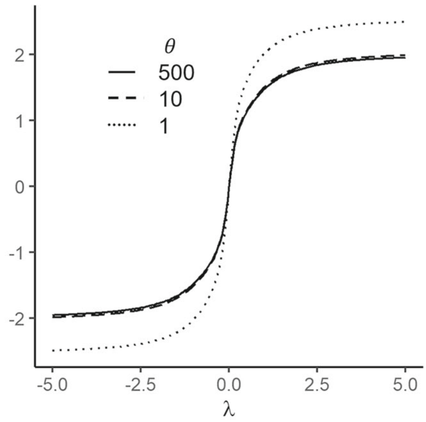
in (47). Thus, by maximizing $Q_2(\boldsymbol{\pi}|\boldsymbol{\psi}^{(r)})$ with respect to $\boldsymbol{\pi}$, an update for π_j is given as $\hat{\pi}_j^{(r+1)} = \frac{1}{n}\sum_{i=1}^{n} z_{ij}^{(r)}$, while updates for $\boldsymbol{\theta}$ are estimated by numerically maximizing $Q_2(\boldsymbol{\theta}|\boldsymbol{\psi}^{(r)})$.

Appendix C: Skewness and kurtosis of SLSM models

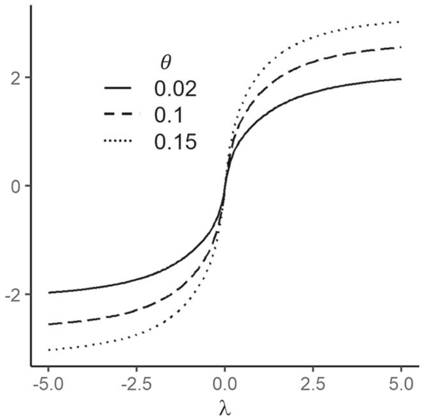
See Figs. 17 and 18.



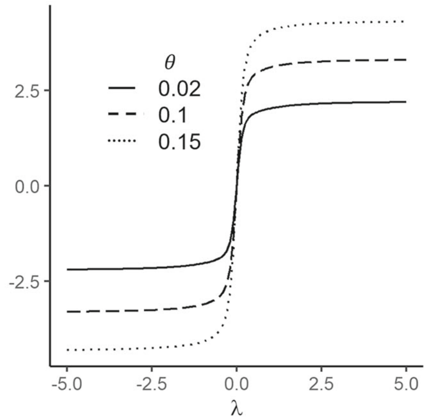
(a) $SSEL_I$ (26).



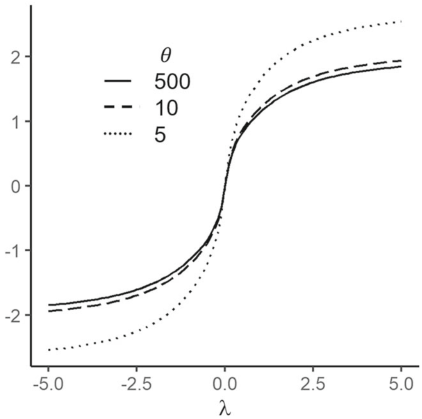
(b) $SSEL_{II}$ (27).



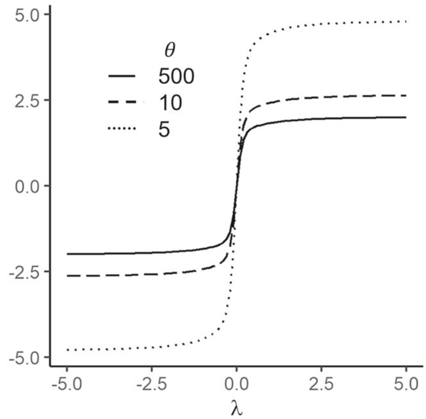
(c) $SUGL_I$ (29).



(d) $SUGL_{II}$ (30).

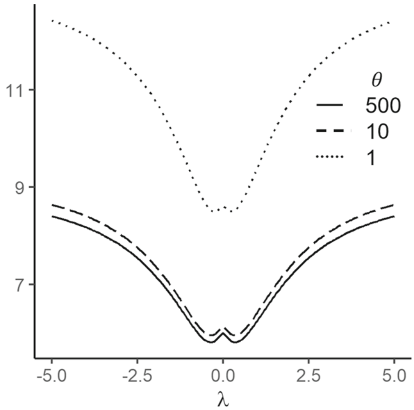


(e) $SPFL_I$ (32).

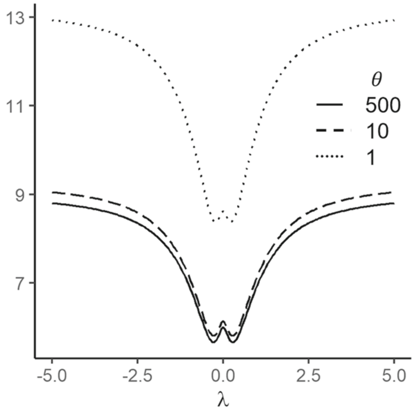


(f) $SPFL_{II}$ (33).

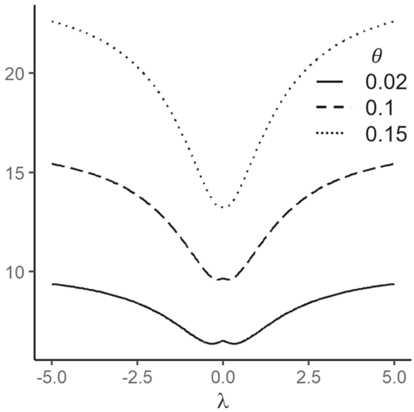
Fig. 17 Skewness of SLSM models as a function of λ



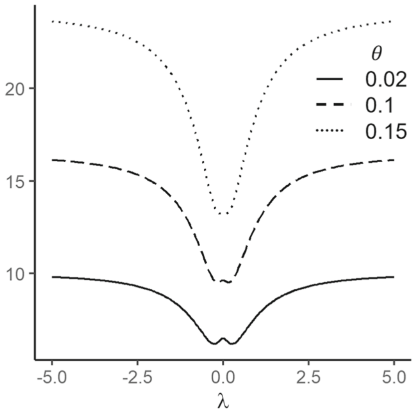
(a) $SSEL_I$ (26).



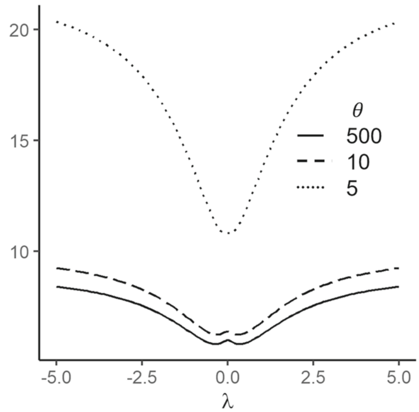
(b) $SSEL_{II}$ (27).



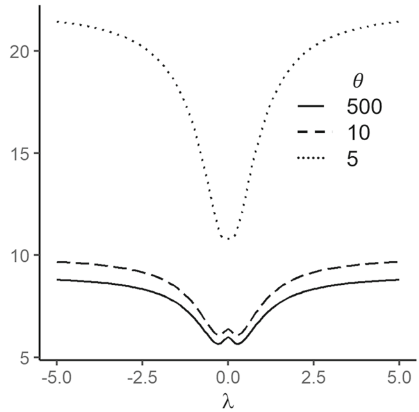
(c) $SUGL_I$ (29).



(d) $SUGL_{II}$ (30).



(e) $SPFL_I$ (32).



(f) $SPFL_{II}$ (33).

Fig. 18 Kurtosis of SLSM models as a function of λ

Acknowledgements The authors are grateful to the associate editor as well as two anonymous reviewers whose comments greatly assisted in presenting an improved manuscript. This work was based upon research supported in part by the National Research Foundation (NRF) of South Africa (SA), grant RA201125576565, nr 145681; NRF ref. SRUG2204203865 nr. 120839; NRF ref. MND210525603756, nr 114613; the Department of Research and Innovation at the University of Pretoria (SA), as well as the Centre of Excellence in Mathematical and Statistical Sciences grant nr 2022-047-STA, based at the University of the Witwatersrand (SA). The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

Funding Open access funding provided by University of Pretoria.

Data availability All datasets considered in this paper are freely available on the internet.

Declarations

Conflict of interest No potential conflict of interest was reported by the author(s).

References

Adcock, C., & Azzalini, A. (2020). A selective overview of skew-elliptical and related distributions and of their applications. *Symmetry, 12*(1), 118.

Ahsanullah, M., & Kabir, A. B. M. L. (1974). A characterization of the power function distribution. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique, 2*(1), 95–98.

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

Arslan, O. (2010). An alternative multivariate skew Laplace distribution: Properties and estimation. *Statistical Papers, 51*(4), 865–887.

Aryal, G., & Nadarajah, S. (2005). On the skew Laplace distribution. *Journal of Information and Optimization Sciences, 26*(1), 205–217.

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

Balakrishnan, N., & Kocherlakota, S. (1985). On the double Weibull distribution: Order statistics and estimation. *Sankhyā: The Indian Journal of Statistics, Series B, 47*(2), 161–178.

Bekker, A., Ferreira, J. T., Arashi, M., & Rowland, B. W. (2020). Computational methods applied to a skewed generalized normal family. *Communications in Statistics-Simulation and Computation, 49*(11), 2930–2943.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.

Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science, 58*(2), 367–386.

Bonica, A. (2023). *Database on ideology, money in politics, and elections: Public version 3.0 [computer file]*. Stanford University Libraries.

Chakraborty, S., Hazarika, P. J., & Ali, M. M. (2014). A multimodal skew Laplace distribution. *Pakistan Journal of Statistics, 30*(2), 253–264.

Chen, S. X. (2000). Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics, 52*(3), 471–480.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*(3), 292.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 39*(1), 1–22.
Doğru, F. Z., & Arslan, O. (2017). Parameter estimation for mixtures of skew Laplace normal distributions and application in mixture regression modeling. *Communications in Statistics-Theory and Methods, 46*(21), 10879–10896.
Doğru, F. Z., & Arslan, O. (2021). Finite mixtures of skew Laplace normal distributions with random skewness. *Computational Statistics, 36*(1), 423–447.
Doğru, F. Z., & Arslan, O. (2023). Variance-mean mixture of multivariate normal distribution and weighted gamma distribution: Properties and applications. *Journal of the Korean Statistical Society, 52*(1), 185–222.
Dutang, C., & Charpentier, A. (2018). *Casdatasets R package vignette. Reference manual. Version 1.0-8*. Technical report, packaged 2018-05-20.
Fang, Y., Franczak, B. C., & Subedi, S. (2023). Tackling the infinite likelihood problem when fitting mixtures of shifted asymmetric Laplace distributions. arXiv preprint arXiv:2303.14211.
Fernández, C., & Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association, 93*(441), 359–371.
Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models* (Vol. 425). Springer.
Gómez, H. W., Venegas, O., & Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics: The Official Journal of the International Environmetrics Society, 18*(4), 395–407.
Gradshteyn, I. S., & Ryzhik, I. M. (2014). *Table of integrals, series, and products*. Academic Press.
Gupta, A. K., Chang, F. C., & Huang, W. J. (2002). Some skew-symmetric models. *Random Operators and Stochastic Equations, 10*(2), 133–140.
Hadri, K. (1996). A note on Sargan densities. *Journal of Econometrics, 71*(1–2), 285–290.
Harandi, S. S., & Alamatsaz, M. H. (2013). Alpha-skew-Laplace distribution. *Statistics & Probability Letters, 83*(3), 774–782.
Harandi, S. S., & Alamatsaz, M. H. (2015). Discrete alpha-skew-Laplace distribution. *SORT: Statistics and Operations Research Transactions, 39*(1), 071–084.
Holla, M. S., & Bhattacharya, S. K. (1968). On a compound Gaussian distribution. *Annals of the Institute of Statistical Mathematics, 20*(1), 331–336.
Ibragimov, M., Ibragimov, R., & Walden, J. (2015). *Heavy-tailed distributions and robustness in economics and finance* (Vol. 214). Springer.
Jagannathan, K. (2005). *Statistical inference and goodness-of-fit tests for skewed double exponential models*. Bowling Green State University.
Kanji, G. K. (1985). A mixture model for wind shear data. *Journal of Applied Statistics, 12*(1), 49–58.
Komunjer, I. (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics, 22*(5), 891–921.
Kotz, S., Kozubowski, T., & Podgórski, K. (2001). *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance* (Vol. 183). Springer Science & Business Media.
Kozubowski, T. J., & Nadarajah, S. (2008). The beta Laplace distribution. *Journal of Computational Analysis & Applications*, *10*(1).
Kozubowski, T. J., & Nadarajah, S. (2010). Multitude of Laplace distributions. *Statistical Papers, 51*(1), 127–148.
Kozubowski, T. J., & Nolan, J. P. (2008). Infinite divisibility of skew Gaussian and Laplace laws. *Statistics & Probability Letters, 78*(6), 654–660.
Lange, K., Chambers, J., & Eddy, W. (1999). *Numerical analysis for statisticians* (Vol. 2). Springer.
MacDonald, I. L. (2014). Numerical maximisation of likelihood: A neglected alternative to EM? *International Statistical Review, 82*(2), 296–308.
MacDonald, I. L. (2021). Is EM really necessary here? Examples where it seems simpler not to use EM. *AStA Advances in Statistical Analysis, 105*(4), 629–647.
Mahdavi, A., Desmond, A. F., & Jamalizadeh, A. (2023). An EM algorithm for estimating the parameters of the skew generalized t-normal distribution with application to robust finite mixture modeling. *Communications in Statistics-Simulation and Computation*. https://doi.org/10.1080/03610918.2023.2263182
McGill, W. J. (1962). Random fluctuations of response rate. *Psychometrika, 27*(1), 3–17.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering* (Vol. 38). M. Dekker.

McLachlan, G. J., & Krishnan, T. (2007). *The EM algorithm and extensions* (Vol. 382). Wiley.

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and its Application, 6*, 355–378.

Morris, K., Punzo, A., McNicholas, P. D., & Browne, R. P. (2019). Asymmetric clusters and outliers: Mixtures of multivariate contaminated shifted asymmetric Laplace distributions. *Computational Statistics & Data Analysis, 132*, 145–166.

Poiraud-Casanova, S., & Thomas-Agnan, C. (2000). About monotone regression quantiles. *Statistics & Probability Letters, 48*(1), 101–104.

Punzo, A. (2019). A new look at the inverse Gaussian distribution with applications to insurance and economic data. *Journal of Applied Statistics, 46*(7), 1260–1287.

Punzo, A., & Bagnato, L. (2021). Modeling the cryptocurrency return distribution via Laplace scale mixtures. *Physica A: Statistical Mechanics and its Applications, 563*, 125354.

Punzo, A., & Bagnato, L. (2022a). Asymmetric Laplace scale mixtures for the distribution of cryptocurrency returns. arXiv:2209.12848.

Punzo, A., & Bagnato, L. (2022b). Dimension-wise scaled normal mixtures with application to finance and biometry. *Journal of Multivariate Analysis,191*, 105020.

Punzo, A., Bagnato, L., & Maruotti, A. (2018). Compound unimodal distributions for insurance losses. *Insurance: Mathematics and Economics, 81*, 95–107.

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Rempala, G. A., & Derrig, R. A. (2005). Modeling hidden exposures in claim severity via the EM algorithm. *North American Actuarial Journal, 9*(2), 108–128.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *CRAN R, 538*, 113–120.

Sadeghkhani, A., & Ghosh, I. (2018). A new generalized Balakrishnan type skewed-normal distribution: Properties and associated inference. *Communications in Statistics—Theory and Methods, 47*(18), 4483–4492.

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

Shah, S., Hazarika, P. J., & Chakraborty, S. (2019). The Balakrishnan alpha skew Laplace distribution: Properties and its applications. arXiv preprint arXiv:1910.01084.

Shah, S., Hazarika, P. J., Chakraborty, S., & Alizadeh, M. (2023). The Balakrishnan-alpha-beta-skew-Laplace distribution: Properties and applications. *Statistics, Optimization & Information Computing, 11*(3), 755–772.

Subbotin, M. T. (1923). On the law of frequency of error. *Matematicheskii Sbornik, 31*(2), 296–301.

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Management Science, 44*(12–part–1), 1650–1661.

Titterington, D. M., Smith, A. F. M., Makov, U. E., et al. (1985). *Statistical analysis of finite mixture distributions* (Vol. 198). Wiley.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to probability and statistics* (pp. 448–485). Stanford University Press.

Wilson, E. B. (1923). First and second laws of error. *Journal of the American Statistical Association, 18*(143), 841–851.

Yu, K., & Jin, Z. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics—Theory and Methods, 34*(9–10), 1867–1879.