International Journal of Statistics and Systems
ISSN 0973-2675 Volume 12, Number 4 (2017), pp. 813–822

http://www.ripublication.com

Robust estimation of skew-normal distribution with location and scale parameters via log-regularly varying functions

Shintaro Hashimoto[1]
Department of Mathematics,
Graduate School of Science, Hiroshima University,
1-3-1 Kagamiyama, Higashi-Hiroshima, 739–8526, Japan.

Abstract

The method of robust parameter estimation for location and scale parameters in the skew-normal distribution is proposed. One of the methods of rejection of outliers is the use of heavy-tailed distributions in the sense of the regularly varying functions (e.g., Student's t-distribution). In this paper, we propose the mixture distribution of the skew-normal and log-regularly varying functions which have heavier tails than those of regularly varying functions, and consider the joint estimation for location and scale parameters in the presence of outliers. In simulation studies, we show the behaviors of the maximum likelihood estimators based on our proposal method for large outlier and also compare the mean square errors with other competitors.

AMS subject classification: Primary 62F35, Secondary 62F15.
Keywords: heavy-tailed distribution, location and scale parameters, log-regularly varying functions, outliers, skew-normal distribution.

1. Introduction

It has been important to construct the robust posterior distribution and likelihood function against outliers for the estimation of location and scale parameters in both Bayesian and frequentist contexts. In general, if we assume the symmetric probability distribution, we often use sample mean and standard deviation for the estimators of location and scale parameters, respectively. However, these estimators suffer from outliers for light-tailed

[1]Research of the author is supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 17K14233.

or asymmetric distributions. To overcome this problem, we often use the sample median as an estimator for location parameter. Unfortunately, it is known that it does not work well if outliers concentrate in one-sided tails. On the other hand, assuming the heavy-tailed model (e.g., Student's t-distribution) is one of the methods of robust parameter estimation for location and scale parameters. A research of Bayesian robustness based on the heavy-tailed distributions is at least dates back to Dawid (1973) and has been developed by many authors (Andrade and O'Hagan, 2006, 2011; O'Hagan and Pericchi, 2012; Andrade et al., 2013; Desgagné, 2013). However, these method also has serious problem. Though the influence of outliers of estimator for location parameter is completely removed, that of estimator for scale parameter can not be removed completely (we called it partial robustness) (see e.g., Andrade and O'Hagan, 2006, 2011; Andrade et al., 2013). It is well-known that the regularly varying function is a measure of the heaviness of the distribution and it often appears in probability theory and extreme value theory (Bingham et al., 1987; Resnick, 1987). From this measure, Student's t-distribution is heavy tailed in the sense of regularly varying. In this paper, we deal with the heavy-tailed distribution in the sense of the log-regularly varying function. It is known that the log-regularly varying function has heavier tails than that of the regularly varying one (Desgagné (2015)). The distribution which have log-regularly varying tails is also called super-heavy-tailed in some contexts. For example, we note that the log-Pareto function which is proportional to $|x|(\log|x|)^{-\beta}$ $(\beta > 1)$ is log-regularly varying.

Recently, Desgagné (2015) consider the robust parameter estimation based on the log-regularly varying function, and show the sufficient condition to obtain whole robustness against outliers for estimation of location and scale parameters. Further, the symmetric distribution which have the log-regularly varying tails is also constructed. Though they consider the estimation in Bayesian framework, the maximum a posteriori (MAP) estimator under the uniform prior is equivalent to the maximum likelihood estimator (MLE). Hence, we note that their result involves the result of the MLE. Robust Bayes estimations based on the log-regularly varying function are also studied by Gagnon et al. (2016) and Desgagné and Gagnon (2016) in linear regression setting.

In this paper, we try to extend the symmetric distribution proposed by Desgagné (2015) to the asymmetric one. In fact, we propose the methods which can be robust estimation for location and scale parameters in the skew-normal distribution which is typical asymmetric distribution proposed by Azzalini (1985). This paper is organized as follows: In section 2, we introduce the measures of heaviness of the tails of distribution, and some previous results given by Desgagné (2015). In section 3, a new skew probability distribution is proposed and some technical issues are discussed. In section 4, it is shown that the MLEs for location and scale parameters based on our proposal model is good robustness properties through some simulation studies.

2. Robust parameter estimation based on the log-regularly varying functions

First, we introduce the measures which characterize the heaviness of the tail of distributions. The following regularly varying function is well-known as a measure of the heaviness of the tails (see e.g., Resnick, 1987).

Definition 2.1. [Regular variation] The positive valued measurable function f on $[x_0, \infty)$ $(x_0 \in \mathbb{R})$ is regularly varying at ∞ with index ρ $(\in \mathbb{R})$, written $f \in R_\rho(\infty)$, if for all $\lambda > 0$ it holds

$$\frac{f(\lambda x)}{f(x)} \to \lambda^\rho \quad (x \to \infty).$$

For example, Student's t-distribution is called heavy-tailed distribution in the sense of regularly varying. Next, we define log-regularly varying function which characterizes the heavier tails than regularly varying one like Student's t-distribution.

Definition 2.2. [Log-regular variation] The positive valued measurable function f on $[x_0, \infty)$ $(x_0 \in \mathbb{R})$ is log-regularly varying at ∞ with index ρ $(\in \mathbb{R})$, written $f \in L_\rho(\infty)$, if for all $\nu > 0$ it holds

$$\frac{g(x^\nu)}{g(x)} \to \nu^{-\rho} \quad (x \to \infty).$$

We note that it is easy to show that $g \in L_\rho(\infty)$ is equivalent to $f \in R_{-\rho}(\infty)$ if we put $f(x) = g(e^x)$, that is, $g(x) = f(\log x)$.

Next, we introduce the setting of statistical model and some results of Desgagé (2015). Let $X_1, \ldots, X_n$ be random variables conditonally independent given μ and σ with their conditional densities given by $(1/\sigma)f((x_i - \mu)/\sigma)$ $(\mu \in \mathbb{R},\ \sigma > 0)$. We put $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ and let its realized values $\boldsymbol{x}_n = (x_1, \ldots, x_n)$. Let $\pi(\mu, \sigma)$ be joint prior density of (μ, σ) and we assume that $\sigma\pi(\mu, \sigma) < \infty$. Then the posterior distribution of (μ, σ) given $\boldsymbol{X}_n$ is given by

$$\pi(\mu, \sigma | \boldsymbol{x}_n) = [m(\boldsymbol{x}_n)]^{-1}\pi(\mu, \sigma) \prod_{i=1}^{n} \frac{1}{\sigma} f\left(\frac{x_i - \mu}{\sigma}\right),$$

where $m(\boldsymbol{x})$ is the marginal density of $\boldsymbol{X}$ given by

$$m(\boldsymbol{x}_n) = \int\int \pi(\mu, \sigma) \prod_{i=1}^{n} \frac{1}{\sigma} f\left(\frac{x_i - \mu}{\sigma}\right) d\mu d\sigma.$$

In this paper, we consider the joint estimation problem for μ and σ in the presence of outliers in data $\boldsymbol{x}_n$. For $i = 1, \ldots, n$, we define three binary functions k_i, l_i and r_i as follows:

$$k_i := \begin{cases} 1 & \text{if } x_i \text{ is a nonoutlying} \\ 0 & \text{otherwise} \end{cases}, \quad l_i := \begin{cases} 1 & \text{if } x_i \text{ is a left outlier} \\ 0 & \text{otherwise} \end{cases},$$

$$r_i := \begin{cases} 1 & \text{if } x_i \text{ is a right outlier} \\ 0 & \text{otherwise} \end{cases}.$$

Then for each $i = 1, \dots, n$, we have $k_i + l_i + r_i = 1$ and we put $\sum_{i=1}^{n} k_i =: k$, $\sum_{i=1}^{n} l_i =: l$ and $\sum_{i=1}^{n} r_i =: r$. Now, there are $k = n - r - l$ non-outliers and let $\boldsymbol{x}_k$ be nonoutlying observations. We can write $x_i = a_i + b_i\omega$ for $i = 1, \dots, n$, where a_i and b_i are some constant such that $a_i \in \mathbb{R}$ and

$$b_i = 0 \ (k_i = 1), \quad b_i < 0 \ (l_i = 1), \quad b_i > 0 \ (r_i = 1),$$

and we consider the asymptotic behavior when $\omega \to \infty$. Recently, the sufficient condition to whole robustness against outliers was shown by Desgagné (2015).

Theorem 2.3. [Desgagné, 2015] Under some regularity conditions, we assume that (i) $xf(x) \in L_\rho(\infty)$ and (ii) $k > \max(l, r)$. Then we have the following:

(a) $\displaystyle\lim_{\omega\to\infty} \frac{m(\boldsymbol{x}_n)}{\prod_{i=1}^{n}[f(x_i)]^{l_i+r_i}} = m(\boldsymbol{x}_k)$,

(b) $\displaystyle\lim_{\omega\to\infty} \pi(\mu, \sigma|\boldsymbol{x}_n) = \pi(\mu, \sigma|\boldsymbol{x}_k)$ uniformly on μ and σ,

(c) $\displaystyle\lim_{\omega\to\infty} \int_0^\infty \int_{-\infty}^\infty |\pi(\mu, \sigma|\boldsymbol{x}_n) - \pi(\mu, \sigma|\boldsymbol{x}_k)|\, d\mu d\sigma = 0$,

(d) $\mu, \sigma|\boldsymbol{x}_n \xrightarrow{d} \mu\,\sigma|\boldsymbol{x}_k, \quad \mu|\boldsymbol{x}_n \xrightarrow{d} \mu|\boldsymbol{x}_k, \quad \sigma|\boldsymbol{x}_n \xrightarrow{d} \sigma|\boldsymbol{x}_k \ (\omega \to \infty)$,

(e) $\displaystyle\lim_{\omega\to\infty} \mathcal{L}(\mu, \sigma|\boldsymbol{x}_n) = \mathcal{L}(\mu, \sigma|\boldsymbol{x}_k)$ uniformly on μ and σ,

where

$$\pi(\mu, \sigma|\boldsymbol{x}_k) = [m(\boldsymbol{x}_k)]^{-1}\pi(\mu, \sigma)\prod_{i=1}^{n}\left[\frac{1}{\sigma}f\left(\frac{x_i - \mu}{\sigma}\right)\right]^{k_i},$$

$$m(\boldsymbol{x}_k) = \int\int \pi(\mu, \sigma)\prod_{i=1}^{n}\left[\frac{1}{\sigma}f\left(\frac{x_i - \mu}{\sigma}\right)\right]^{k_i} d\mu d\sigma$$

and $\mathcal{L}(\mu, \sigma|\boldsymbol{x}_n)$ is the likelihood function of (μ, σ).

This theorem means that the marginal density, joint posterior density, marginal posterior density and likelihood function does not be affected by outliers ω for large ω. The condition (i) means that tails of underlying distribution are log-regularly varying and (ii) means that the number of nonoutlying observations is larger than $\max(l, r)$. Further, Desgagné (2015) constructed the symmetric distribution with log-regularly varying tails as a mixture of the symmetric density and log-Pareto-tailed functions. A random variable X have the log-Pareto-tailed symmetric distribution if the density function is given by

$$f(x|\boldsymbol{\phi}, \alpha, \beta) = K_{(\boldsymbol{\phi},\alpha,\beta)} \left\{ g(x|\boldsymbol{\phi}) 1_{[-\alpha,\alpha]}(x) + g(\alpha|\boldsymbol{\phi}) \frac{\alpha}{|x|} \left(\frac{\log \alpha}{\log |x|} \right)^{\beta} 1_{(\alpha,\infty)}(|x|) \right\}, \tag{2.1}$$

where $z \in \mathbb{R}$, $\alpha > 1$ and $\beta > 1$. Here, $1_A(\cdot)$ is an indicator function which is defined by $1_A(x) = 1$ $(x \in A)$; $= 0$ $(x \notin A)$, and $g(\cdot|\boldsymbol{\phi})$ is any symmetric at origin and continuous function on $[-\alpha, \alpha]$ with vector parameter $\boldsymbol{\phi} \in \boldsymbol{\Theta}$. In the case of location-scale family, we often consider $\boldsymbol{\phi} = (\mu, \sigma) \in \Theta \subset \mathbb{R} \times \mathbb{R}_{>0}$. The normalizing constant $K_{\boldsymbol{\phi},\alpha,\beta}$ is given by

$$K_{\boldsymbol{\phi},\alpha,\beta} = \frac{\beta - 1}{(2G(\alpha|\boldsymbol{\phi}) - 1)(\beta - 1) + 2g(\alpha|\boldsymbol{\phi})\alpha \log \alpha},$$

where $G(\alpha|\boldsymbol{\phi}) := \int_{-\infty}^{\alpha} g(u|\boldsymbol{\phi}) du$.

3. Construction of log-Pareto-tailed skew-normal distribution

In this section, we extend the symmetric mixture distribution (2.1) constructed by Desgagné (2015) to asymmetric case. In particular, we construct log-Pareto-tailed skew-normal distribution.

The most popular asymmetric family of distributions is skew family and its probability density function is defined by

$$f(x) = 2f_0(x)G_0(w(x)), \tag{3.1}$$

where we assume that the functions $f_0(\cdot)$, $G_0(\cdot)$ and $w(\cdot)$ satisfy the conditions $f_0(-x) = f_0(x)$, $G_0(-y) = 1 - G_0(y)$ and $w(-x) = -w(x)$, respectively. It is known that the skew family (3.1) has some good properties by putting $w(x) = \lambda x$ $(\lambda \in \mathbb{R})$ (Azzalini, 2014). Now, let $\varphi(x)$ and $\Phi(x)$ be the density function and distribution function of $N(0, 1)$, respectively. By putting $f_0(x) = \varphi(x)$, $G_0(x) = \Phi(x)$, and $w(x) = \lambda x$, we have

$$\varphi(x; \lambda) = 2\varphi(x)\Phi(\lambda x), \tag{3.2}$$

where $\lambda \in \mathbb{R}$ is called the skew parameter which is a measure of skewness of distributions. The distribution with the density (3.2) is called the skew-normal distribution written by $SN(0, 1; \lambda)$ and have been widely used in application in recent years. We note that if we put $\lambda = 0$ in (3.2), $\varphi(x; 0) = \varphi(x)$ is the density function of the standard normal distribution. This means that the skew-normal distribution involve the (symmetric) normal distribution as a special case. Further, we also note that the density (3.2) corresponds to the density of the half-normal distribution as $\lambda \to \infty$.

It is not easy to construct log-Pareto-tailed asymmetric distribution based on (3.1), so we construct log-Pareto-tailed skew-normal distribution in this paper. Hereafter, we assume that the skew parameter λ is known and we focus on the estimation for location and scale parameters in this paper.

Definition 3.1. [log-Pareto-tailed skew-normal distribution] Let $\lambda > 0$. A random variable X have the log-Pareto-tailed skew-normal distribution if the density function is given by

$$\begin{aligned} &f(x|\lambda, \alpha_1, \alpha_2, \beta) \\ &= C_{(\lambda,\alpha_1,\alpha_2,\beta)}\left\{\varphi(x; \lambda)1_{[-\alpha_1,\alpha_2]}(x) + \varphi(-\alpha_1; \lambda)\frac{1+\alpha_1}{|1-x|}\left(\frac{\log(1+\alpha_1)}{\log|1-x|}\right)^{\beta} 1_{(-\infty,-\alpha_1)}(x)\right. \\ &\quad \left. + \varphi(\alpha_2; \lambda)\frac{\alpha_2}{|x|}\left(\frac{\log\alpha_2}{\log|x|}\right)^{\beta} 1_{(\alpha_2,\infty)}(x)\right\}, \end{aligned} \tag{3.3}$$

where $x \in \mathbb{R}$, $\lambda > 0$, $0 < \alpha_1 < 1.96$, $\alpha_2 > 1.96$ and $\beta > 1$, and the normalizing constant $C_{(\lambda,\alpha_1,\alpha_2,\beta)}$ can be given by the following analytical form:

$$\begin{aligned} C_{(\lambda,\alpha_1,\alpha_2,\beta)} &= \left\{\int\left[\varphi(x; \lambda)1_{[-\alpha_1,\alpha_2]}(x) + \varphi(-\alpha_1; \lambda)\frac{1+\alpha_1}{|1-x|}\left(\frac{\log(1+\alpha_1)}{\log|1-x|}\right)^{\beta} 1_{(-\infty,-\alpha_1)}(x)\right.\right. \\ &\quad \left.\left. + \varphi(\alpha_2; \lambda)\frac{\alpha_2}{|x|}\left(\frac{\log\alpha_2}{\log|x|}\right)^{\beta} 1_{(\alpha_2,\infty)}(x)\right]dx\right\}^{-1} \\ &= \frac{\beta-1}{\{H(\alpha_1; \lambda) + H(\alpha_2; \lambda) - 1\}(\beta-1) + \varphi(-\alpha_1; \lambda)(1+\alpha_1)\log(1+\alpha_1) + \varphi(\alpha_2; \lambda)\alpha_2\log\alpha_2}, \end{aligned}$$

where $H(\alpha; \lambda) := \int_{-\infty}^{\alpha} \varphi(t; \lambda)dt$. For $\lambda < 0$, (3.3) is also defined by the same way.

Note that log-Pareto-tailed skew-normal distribution (3.3) is the mixture distribution of the skew-normal distribution and log-Pareto function, that is, the core part of the density function is the skew-normal distribution ($SN(0, 1; \lambda)$) and both tails is the function which is proportional to $|x|(\log|x|)^{-\beta}$ ($\beta > 1$). We also note that the log-Pareto function is also called super-heavy-tailed because its tail is heavier than that of like Student's t-distribution in general.

We now introduce the location and scale parameters (μ, σ) in the density (3.3), and consider the estimation problem for (μ, σ). However, we have the two problem, that is, (i) how to determine the mixture ratio and (ii) how to choose the parameters α_1, α_2 and β. In this paper, we determine the mixture ratio q in advance, e.g., $q = .95$. It may be unnatural assumption because the outliers are usually in the far away from the core

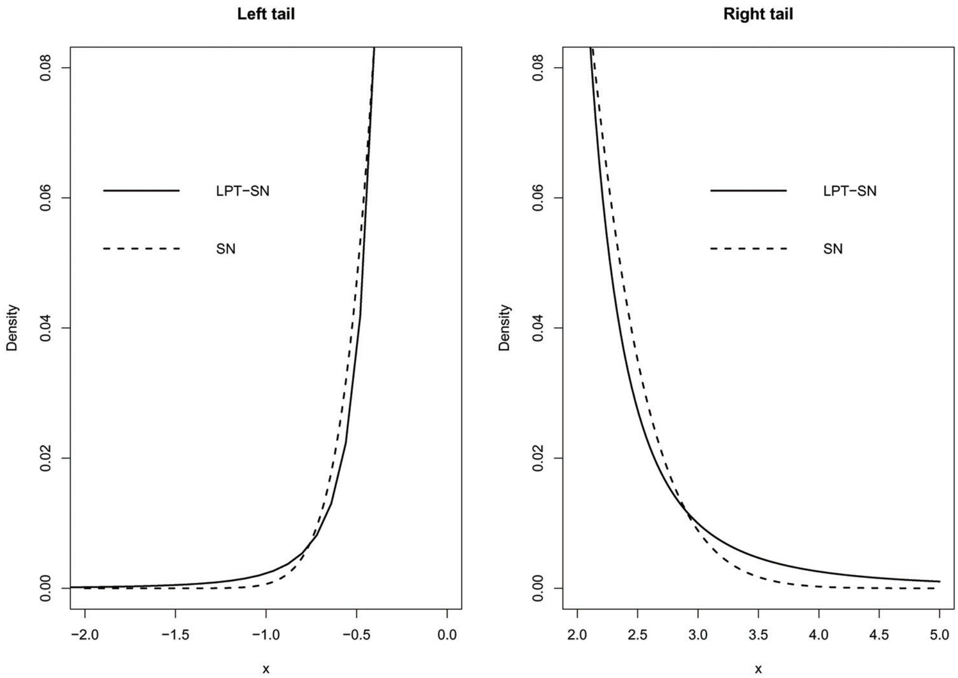

Figure 1: Tail behaviors of log-Pareto-tailed skew-normal (LPT-SN) distribution and skew-normal (SN) distribution for $\lambda = 3$.

part. Then we can propose the algorithm to determine the parameters α_1, α_2 and β by following way.

1. Determine the mixture ratio q $(0 < q < 1)$ and find k (> 0) satisfies

$$\int_{\{x:\ f(x)\geq k\}} \phi(x;\lambda)dx = q,$$

   where $\phi(x;\lambda) = 2\phi(x)\Phi(\lambda x)$ is the density function of skew-normal distribution with skew parameter λ.

2. By using k and quantile points of $\phi(x;\lambda)$, find α_1, α_2.

3. By solving $C_{(\lambda,\alpha_1,\alpha_2,\beta)} = 1$, we obtain $\beta > 0$.

This algorithm is slightly different from that of Desgagné (2015). In fact, we use the concept of the highest posterior density (HPD) interval in step 1 because the skew-normal distribution is asymmetric. For $\lambda = 3$ and $q = .95$, we can determine the parameters $\alpha_1 \approx 0.41$, $\alpha_2 \approx 2.08$, and $\beta \approx 4.57$, for example.

4. Simulation studies

In this section, we show some simulation studies for estimation of (μ, σ) in the presence of outliers. We consider the following mixture model:

$$
\begin{aligned}
& f(x|\boldsymbol{\phi}, \alpha_1, \alpha_2, \beta) \\
&= C_{(\boldsymbol{\phi},\alpha_1,\alpha_2,\beta)} \left\{ \varphi\left(\frac{x-\mu}{\sigma}; \boldsymbol{\phi}\right) 1_{[-\alpha_1,\alpha_2]}\left(\frac{x-\mu}{\sigma}\right) \right. \\
&\quad + \varphi(-\alpha_1; \boldsymbol{\phi}) \frac{1+\alpha_1}{|1-((x-\mu)/\sigma)|} \left(\frac{\log(1+\alpha_1)}{\log|1-((x-\mu)/\sigma)|}\right)^{\beta} 1_{(-\infty,-\alpha_1)}\left(\frac{x-\mu}{\sigma}\right) \\
&\quad \left. + \varphi(\alpha_2; \boldsymbol{\phi}) \frac{\alpha_2}{|(x-\mu)/\sigma|} \left(\frac{\log\alpha_2}{\log|(x-\mu)/\sigma|}\right)^{\beta} 1_{(\alpha_2,\infty)}\left(\frac{x-\mu}{\sigma}\right) \right\}, \qquad (4.1)
\end{aligned}
$$

where $\boldsymbol{\phi} = (\mu, \sigma, \lambda)$ and $\mu \in \mathbb{R}$, $\sigma > 0$. We assume that λ is known. Without loss of generality, we choose the improper and uninformative joint prior density $\pi(\mu, \sigma) \propto 1/\sigma$. So, both the Bayesian and frequentist approaches can be used. We note that under this prior, the MLEs and posterior modes are very similar behavior. So, hereafter, we consider the MLEs for (μ, σ).

4.1. Behaviors of the MLEs for (μ, σ)

For comparison, we consider the following three models with location and scale parameters, that is, (a) skew-normal (SN) distribution, (b) log-Pareto-tailed skew-normal (LPT-SN) distribution and (c) skew-*t* (St) distribution with degree of freedom $\nu = 10$. In all models, we assume that the skew parameter $\lambda = 3.0$ and we set $q = .95$ in model (b). As we mentioned in previous section, we can set $(\alpha_1, \alpha_2, \beta) = (0.41, 2.08, 4.57)$ in model (b). We generate the data $x_1, \dots, x_{20}$ (sample size $n = 20$) from the skew-normal distribution $SN(0, 1; 3)$ and let $\boldsymbol{x}_{21}(\omega) = (x_1, \dots, x_{20}, \omega)$. We note that the MLEs for (μ, σ) based on $x_1, \dots, x_{20}$ are (0.035, 0.809). We study the impact of moving ω from 0 to 100 on estimation for location and scale parameters based on the maximum likelihood method calculated from different three models in the above. In these models, it is not easy to calculate the MLEs in analytical forms, so we calculate these by numerical computation. We show the result in Figure 2. From this figure, we can find that the MLEs based on the skew-normal and skew-*t* is unstable for large ω. On the other hand, the case of log-Pareto-tailed skew-normal distribution is stable even if an outlier ω is large.

4.2. Comparison of the mean square errors of the MLEs

In this subsection, we consider the following three data generating distributions, that is, (a) $SN(0, 1, 3)$, (b) $0.90 \times SN(0, 1, 3) + 0.10 \times N(0, 6)$ and (c) $0.95 \times SN(0, 1, 3) + 0.05 \times N(8, 1)$. We generate the sample (sample size $n = 30$), and we calculate the mean squared error (MSE) for MLEs of (μ, σ) by using 25, 000 times Monte Carlo simulation (Table 1, 2). From these tables, we can find that our proposal model has minimum MSEs among three models in the presence of outliers. In particular, for the estimation of scale

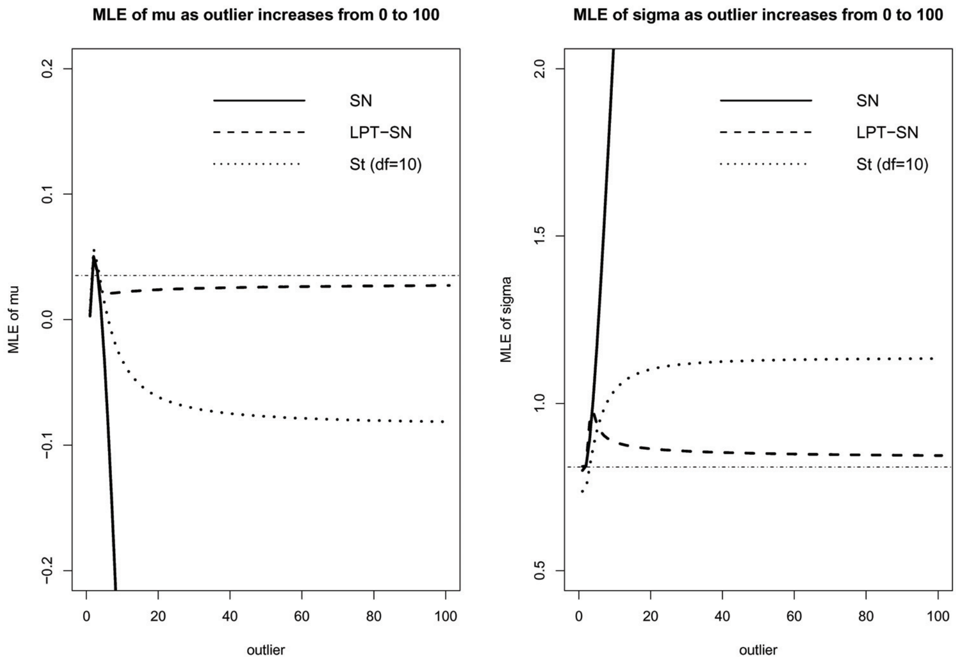

Figure 2: Behaviors of the maximum likelihood estimators of (μ, σ) under the three models as moving an outlier ω from 0 to 100.The horizontal lines are values of MLEs without ω (outlier).

parameter, MLEs based on the log-Pareto-tailed skew-normal distribution outperform the others. Hence, it can be found that our proposal model leads to the whole robust estimators for both location and scale parameters.

Table 1: MSEs for MLEs of location parameter μ ($n = 30$)

Model	100% $SN(0, 1; 3)$	10% $N(0, 6)$	5% $N(8, 1)$
log-Pareto-tailed SN	0.01	0.08	0.01
skew-t ($\nu = 10$)	0.01	0.17	0.02
SN	0.01	0.66	0.09

5. Concluding remarks

A skew-normal distribution with log-regularly varying function was constructed. It was shown that the MLEs for location and scale parameters are more robust against outliers than other competitors through some numerical studies.

As future works, we will extend our proposal distribution to more general skew family of distributions. Further, the joint estimation for (μ, σ, λ) with unknown λ should be also considered.

Table 2: MSEs for MLEs of scale parameter σ ($n = 30$)

Model	100% $SN(0, 1; 3)$	10% $N(0, 6)$	5% $N(8, 1)$
log-Pareto-tailed SN	0.02	0.13	0.06
skew-t ($\nu = 10$)	0.02	0.28	0.23
SN	0.02	1.18	1.31

References

[1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**: 171–178.

[2] Azzalini, A. (2014). *The Skew-Normal and Related Families*. Cambridge Univ. Press, New York.

[3] Andrade, J. A. A., Dorea, C. C. Y. and Guevara Otiniano, C. E. (2013). On the robustness of Bayesian modelling of location and scale structure using heavy-tailed distributions. *Communications in Statistics – Theory and Methods*, **42**: 1502–1514.

[4] Andrade, J. A. A. and O'Hagan, A. (2006). Bayesian robustness modelling using regularly varying distributions. *Bayesian Analysis*, **1**: 169–188.

[5] Andrade, J. A. A. and O'Hagan, A. (2011). Bayesian robustness modelling of location and scale parameters. *Scandinavian Journal of Statistics*, **38**: 691–711.

[6] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular variation*. In Encyclopedia of Mathematics and Its Applications, vol **27**. Cambridge: Cambridge University Press.

[7] Dawid, A. P. (1973). Posterior expectations for large observations. *Biometrika*, **60**: 664–667.

[8] Desgagné, A. (2013). Full robustness in Bayesian modelling of a scale parameter. *Bayesian Analysis*, **8**: 187–219.

[9] Desgagné, A. (2015). Robustness to outliers in location-scale parameter model using log-regularly varying distributions. *Ann. Statist.*, **43**: 1568–1595.

[10] Desgagné, A. and Gagnon, P. (2016). Bayesian robustness to outliers in linear regression and ratio estimation. arXiv: 1612.05307v1.

[11] Gagnon, P., Desgagné, A. and Bédard, M. (2016). Bayesian robustness to outliers in linear regression. arXiv: 1612.06198v1.

[12] O'Hagan, A. and Pericchi, L. (2012). Bayesian heavy-tailed models and conflict resolution: A review. *Brazilian Journal of Probability and Statistics*, **26**: 372–401.

[13] Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*, Springer, New York.