

Robust estimation of location and scale parameters for skew-normal distribution using log-Pareto function

広島大学・理学研究科 橋本真太郎 (Shintaro Hashimoto)
(Department of Mathematics, Hiroshima University)

1 はじめに

位置尺度母数のベイズ推定において，データの外れ値に関してロバストな事後分布を構成することは重要である．一般に，位置尺度母数の推定においては対称分布であれば標本平均や標本標準偏差がよく用いられるが，正規分布などの裾の軽い分布を仮定したときにはこれらの推定量は外れ値の影響を大きく受けてしまう．また，外れ値が分布の両裾に均等にあると考えられるときは位置母数のメディアン推定を行うことがあるが，外れ値が分布の片方の裾に集中しているときは推定精度が悪くなることが知られている．外れ値に影響を受けにくい推定を行う一つの方法として， t -分布などの裾の重い分布を仮定して推測を行うことがある．この方法では，位置母数の推定に関しては外れ値の影響を十分に取り除けるが，尺度母数に関しては部分的に外れ値の影響が残ってしまうという問題点がある．分布の裾の重さに関する指標として，正則変動関数がよく知られており，確率論や極値統計学などでよく現れる (Resnick, 1987)．この指標によると t -分布は正則変動であるという意味で *heavy tailed* ということになる．本稿では，正則変動関数よりもさらに裾の重い対数正則変動関数を考える．これは， t -分布よりも裾が重いため *super-heavy tailed* とよばれることもある．例えば $|x|(\log |x|)^{-\beta}$ ($\beta > 1$) に比例するような対数パレート関数は対数正則変動である．

最近，Desgagné (2015) は位置尺度母数のベイズ推定に関して，対数パレート関数を裾にもつような対称分布を仮定したとき，左裾と右裾の外れ値の個数の大きい方が非外れ値の個数を超えさえしなければ，外れ値が十分裾にあるときに位置母数・尺度母数の同時事後分布とそれぞれの周辺事後分布，さらに尤度関数は外れ値の影響を全く受けないことを示した．さらに，対数パレート関数を裾に持つような対称分布を構成して，パラメータ推定の方法を提案している．なお，位置尺度母数の同時事前分布として一様分布を仮定したときの事後分布に基づく最大事後確率 (maximum a posteriori; MAP) 推定量は最尤推定量 (MLE) に一致することから，ベイズ推定を考えているが最尤推定における結果も含ん

でいることに注意する。

本稿では、対称分布でなく、非対称な分布の位置尺度母数のロバスト推定を Desgagné (2015) における結果に基づいて行う。具体的には、裾に対数正則変動である対数パレート関数を持つような非対称分布を構成し、MLE が外れ値に影響を受けないことを数値実験により示す。非対称分布としては Azzalini (1985) により提案された skew-normal 分布を本稿では考える。より一般的な非対称な分布族への拡張は今後の課題である。

2 対数正則変動関数に基づくロバスト推定

裾の重さを特徴付ける指標としてよく知られている正則変動関数は次で定義される。

定義 2.1 (Regular variation). $[x_0, \infty)$ ($x_0 \in \mathbb{R}$) 上の正值可測関数 f が $x \rightarrow \infty$ で指数 $\rho \in \mathbb{R}$ をもつ正則変動関数 (regular varying function) であるとは、任意の $\lambda > 0$ に対して次が成り立つことである：

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

これを $f \in R_\rho(\infty)$ と表す。

例えば、 t -分布は正則変動であるという意味で裾の重い分布であるといわれる。次に、 t -分布などよりもさらに裾の重い分布を特徴付けるものとして対数正則変動関数を定義する。

定義 2.2 (Log-regularly variation). $[x_0, \infty)$ ($x_0 \in \mathbb{R}$) 上の正值可測関数 g が $x \rightarrow \infty$ で指数 $\rho \in \mathbb{R}$ をもつ対数正則変動関数 (Log-regularly varying function) であるとは、任意の $\nu > 0$ に対して次が成り立つことである：

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

これを $g \in L_\rho(\infty)$ と表す。

正則変動関数と対数正則変動関数の間には $f(x) = g(e^x)$ 、つまり $g(x) = f(\log x)$ とおくと、次の関係が成り立つことに注意する：

$$g \in L_\rho(\infty) \iff f \in R_{-\rho}(\infty).$$

いま、 $X_1, \dots, X_n$ を互いに独立に確率密度関数 $(1/\sigma)f((x_i - \mu)/\sigma)$ ($\mu \in \mathbb{R}$, $\sigma > 0$) に従う確率変数列とし、 $\mathbf{X}_n = (X_1, \dots, X_n)$ とおき、その実現値を $\mathbf{x}_n = (x_1, \dots, x_n)$ と

する。また、 $\pi(\mu, \sigma)$ を (μ, σ) の同時事前密度とし、 $\sigma\pi(\mu, \sigma) < \infty$ であることを仮定する。このとき、 $\mathbf{X}_n$ を与えたもとの (μ, σ) の事後分布は次のようになる：

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

ただし、 $m(\mathbf{x})$ は $\mathbf{X}$ の周辺分布であり

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

と表される。本稿では、データ $\mathbf{x}_n$ に外れ値が存在している場合の位置・尺度母数 (μ, σ) の同時推定問題を考える。各 i に対して、2 値関数 k_i, l_i, r_i を x_i が非外れ値なら $k_i = 1$ 、左裾の外れ値なら $l_i = 1$ 、右裾の外れ値なら $r_i = 1$ と定義する。このとき、各 i に対して $k_i + l_i + r_i = 1$ であり、 $\sum_{i=1}^n k_i = k$ 、 $\sum_{i=1}^n l_i = l$ 、 $\sum_{i=1}^n r_i = r$ とおく。つまり、 n 個の観測値 $\mathbf{x}_n = (x_1, \dots, x_n)$ において、左裾に l 個、右裾に r 個の外れ値があるということを表している。このとき、非外れ値は $k = n - r - l$ 個であり、外れ値を含まない観測値を $\mathbf{x}_k$ とかく。観測値 x_i はある定数 a_i と b_i を用いて

$$x_i = a_i + b_i \omega \quad (i = 1, \dots, n)$$

とかける。ただし、 $a_i \in \mathbb{R}$ であり、 b_i は

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

とし、 $\omega \rightarrow \infty$ とする。事後分布が外れ値に完全に影響を受けないための十分条件に関して次の定理が与えられている。

定理 2.1 (Desgagné, 2015). 適当な正則条件のもとで (i) $xf(x) \in L_\rho(\infty)$, (ii) $k > \max(l, r)$ であれば、次が成り立つ：

- (a) $\lim_{\omega \rightarrow \infty} \frac{m(\mathbf{x}_n)}{\prod_{i=1}^n [f(x_i)]^{l_i + r_i}} = m(\mathbf{x}_k),$
- (b) $\lim_{\omega \rightarrow \infty} \pi(\mu, \sigma | \mathbf{x}_n) = \pi(\mu, \sigma | \mathbf{x}_k)$ uniformly on μ and σ ,
- (c) $\lim_{\omega \rightarrow \infty} \int_0^\infty \int_{-\infty}^\infty |\pi(\mu, \sigma | \mathbf{x}_n) - \pi(\mu, \sigma | \mathbf{x}_k)| d\mu d\sigma = 0,$
- (d) $\mu, \sigma | \mathbf{x}_n \xrightarrow{d} \mu \sigma | \mathbf{x}_k, \quad \mu | \mathbf{x}_n \xrightarrow{d} \mu | \mathbf{x}_k, \quad \sigma | \mathbf{x}_n \xrightarrow{d} \sigma | \mathbf{x}_k \quad (\omega \rightarrow \infty),$
- (e) $\lim_{\omega \rightarrow \infty} \mathcal{L}(\mu, \sigma | \mathbf{x}_n) = \mathcal{L}(\mu, \sigma | \mathbf{x}_k)$ uniformly on μ and σ .

ただし,

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

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

であり, $\mathcal{L}(\mu, \sigma | \mathbf{x}_n)$ は (μ, σ) の尤度関数とする.

Desgagné(2015)によると, 確率変数 X が log-Pareto-tailed symmetric 分布に従うとは, 密度関数が次のようにかけることである:

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

ただし, $z \in \mathbb{R}$, $\alpha > 1$, $\beta > 1$, $1_A(\cdot)$ は定義関数であり, $g(\cdot|\phi)$ はベクトルパラメータ $\phi \in \Phi$ を持つ $[-\alpha, \alpha]$ 上の任意の原点对称かつ連続関数とする. ϕ としては, 例えば位置尺度母数 $\phi = (\mu, \sigma)$ などを考えることが多い. また, 正規化定数は次で与えられる:

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

ただし, $G(\alpha|\phi) = \int_{-\infty}^{\alpha} g(u|\phi) du$ とする.

3 Log-Pareto-tailed skew-normal distribution

本節では, 対数正則変動関数を裾にもつ非対称な正規分布を構成し, そのモデルに基づく位置尺度母数のロバスト推定を考える. また, 数値実験を通して得られる推定量が他のモデルに比べ, 外れ値の影響を受けにくいことを示す.

3.1 分布の構成

Desgagné (2015) が提案した log-Pareto-tailed symmetric 分布を対称な分布を含む非対称な分布族へと拡張したい. 対称な分布を含むような非対称な分布族としては Azzalini (1985) によって提案された, skew 分布族があり, 確率密度関数は次のように表される:

$$f(x) = 2f_0(x)G_0(w(x)) \quad (1)$$

ただし, $f_0(-x) = f_0(x)$, $G_0(-y) = 1 - G_0(y)$, $w(-x) = -w(x)$ を満たすものとする. また, (1) で $w(x) = \lambda x$ ($\lambda \in \mathbb{R}$) とおくといくつかの良い性質を持つことが知られている

(Azzalini (2014)). 特に, $\varphi(x)$, $\Phi(x)$ をそれぞれ $N(0, 1)$ の密度関数, 分布関数として, $f_0(x) = \varphi(x)$, $G_0(x) = \Phi(x)$, $w(x) = \lambda x$ とおくと,

$$\varphi(x; \lambda) = 2\varphi(x)\Phi(\lambda x) \quad (2)$$

となり, これは歪正規分布 (skew normal distribution) とよばれており, 近年幅広く用いられている. ここで, 実数 λ は skew parameter と呼ばれ, 分布の歪み度合いを表す. また, (2) で $\lambda = 0$ とすると, $\varphi(x; 0) = \varphi(x)$ となりこれは標準正規分布の密度関数になり, 歪正規分布は特別な場合として対称な分布も含んでいることを意味する. また, $\lambda \rightarrow \infty$ とすると (2) は半正規分布 (half-normal distribution) になる. いま, Desgagné (2015) の log-Preto-tailed symmetric と同様の方法で log-Pareto-tailed asymmetric 分布を skew 分布族 (1) を次で定義してみる:

$$f(x|\phi, \alpha_1, \alpha_2, \beta) \quad (3)$$

$$= C_{(\phi, \alpha_1, \alpha_2, \beta)} \left\{ h(x|\phi) 1_{[-\alpha_1, \alpha_2]}(x) + h(\alpha_1|\phi) \frac{\alpha_1}{|x|} \left(\frac{\log \alpha_1}{\log |x|} \right)^\beta 1_{(-\infty, -\alpha_1)}(x) \right. \\ \left. + h(\alpha_2|\phi) \frac{\alpha_2}{|x|} \left(\frac{\log \alpha_2}{\log |x|} \right)^\beta 1_{(\alpha_2, \infty)}(x) \right\}. \quad (4)$$

ただし, $\alpha_1 > 1$, $\alpha_2 > 1$, $\beta > 1$ であり, $h(x|\phi)$ は (1) で与えられる skew 分布族. しかし, 非対称な分布のときは閾値 α_1, α_2 は必ずしも 1 より大きくなるとは限らない. そのため, 密度関数に幾らかの工夫が必要である. 例えば, (2) の場合は skew parameter が 0 のときは標準正規分布になることより, 95% 最高密度区間を考えると $\alpha_1 = \alpha_2 = 1.96$ になるため, skew parameter を $\lambda > 0$ に限定して考えると, $0 < \alpha_1 < 1.96$, $\alpha_2 > 1.96$ であることが予想できる (逆に $\lambda < 0$ なら, $\alpha_1 > 1.96$, $0 < \alpha_2 < 1.96$).

一般の skew 分布族について考えることは容易ではないため, 本稿では $h(x|\phi)$ を (2) で与えられる skew normal 分布 $\varphi(x; \lambda)$ に限定して考える. $\lambda > 0$ のとき, log-Pareto-tailed skew normal (LPT-SN) 分布を次で定義する:

定義 3.1. 確率変数 X が log-Pareto-tailed skew normal (LPT-SN) 分布に従うとは, 確率密度関数が次で与えられることをいう:

$$f(x|\lambda, \alpha_1, \alpha_2, \beta) \\ = C_{(\lambda, \alpha_1, \alpha_2, \beta)} \left\{ \varphi(x; \lambda) 1_{[-\alpha_1, \alpha_2]}(x) + \varphi(-\alpha_1; \lambda) \frac{1 + \alpha_1}{|1 - x|} \left(\frac{\log(1 + \alpha_1)}{\log |1 - x|} \right)^\beta 1_{(-\infty, -\alpha_1)}(x) \right. \\ \left. + \varphi(\alpha_2; \lambda) \frac{\alpha_2}{|x|} \left(\frac{\log \alpha_2}{\log |x|} \right)^\beta 1_{(\alpha_2, \infty)}(x) \right\}. \quad (5)$$

ただし, $x \in \mathbb{R}$, $\lambda > 0$, $0 < \alpha_1 < 1.96$, $\alpha_2 > 1.96$, $\beta > 1$ であり, $C_{(\lambda, \alpha_1, \alpha_2, \beta)}$ は (5) を確率密度関数にするための正規化定数である. 正規化定数は解析的に求めることができ, 次のようになる:

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

ただし, $H(\alpha; \lambda) := \int_{-\infty}^{\alpha} \varphi(t; \lambda) dt$ とする. $\lambda < 0$ の場合も同様に定義できることに注意する.

3.2 パラメータの決め方

前節で見てきたように, (5) で与えられる LPT-SN 分布は skew normal 分布と対数パレート関数 (log-Pareto function) の混合分布の形になっている. つまり, 密度関数 (5) のコア (核) の部分は skew normal 分布 ($SN(0, 1; \lambda)$) で, 両裾に $|x|(\log|x|)^{-\beta}$ ($\beta > 1$) に比例する関数をつなぎ合わせた形で表現される. また, 対数パレート関数は t -分布よりも裾が重いため super-heavy tailed とよばれることもある. この密度関数に位置母数や尺度母数を導入してパラメータ推定をする際に, (5) においてあらかじめ決めておかなければならないことが2つある:

- skew normal 分布と裾の部分 ($\propto |x|(\log|x|)^{-\beta}$ ($\beta > 1$)) の混合比.
- パラメータ $\alpha_1, \alpha_2, \beta$ をどのように決めるか.

これら3つのパラメータは次の手順により決めることができる. ただし, 本稿では skew-normal 分布の歪み度合いを表す母数 λ は既知とする.

1. 混合比 q ($0 < q < 1$) を決め, 以下を満たすような k (> 0) を求める:

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

ただし, $\phi(x; \lambda) = 2\phi(x)\Phi(\lambda x)$ である.

2. 上で求めた k を用いて, $\phi(x; \lambda)$ の分位点から α_1, α_2 を求める.

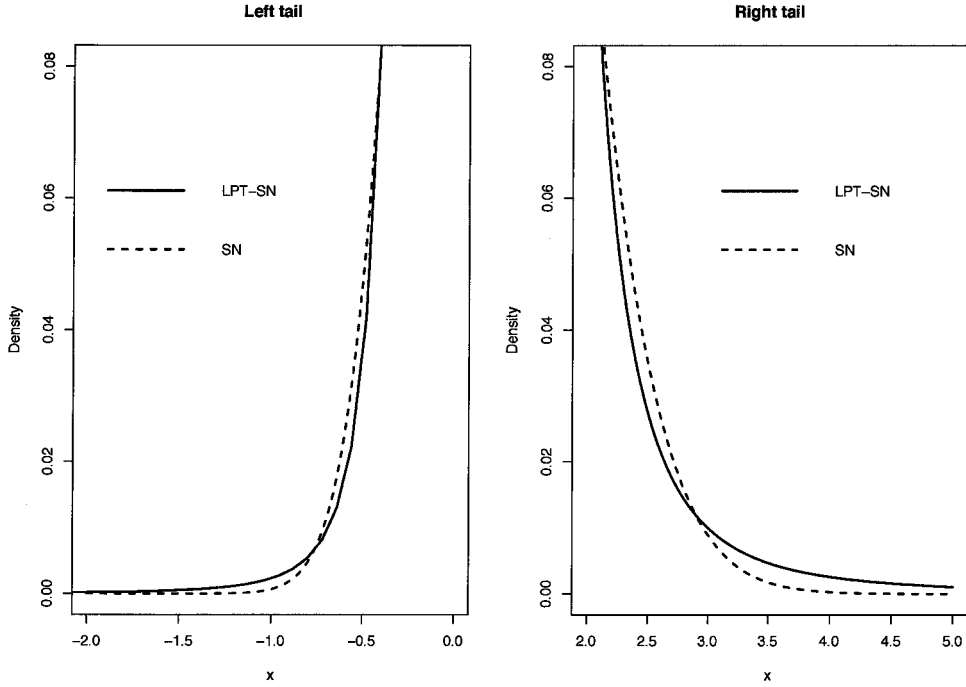


図 1 $\lambda = 3$ のときのと log-Pareto tailed skew normal(LPT-SN) 分布と skew normal(SN) 分布の裾の挙動.

3. 正規化定数 $C_{(\lambda, \alpha_1, \alpha_2, \beta)} = 1$ を解くことにより β を求める.

例えば, $\lambda = 3$, $q = 0.95$ のときは $\alpha_1 \approx 0.41$, $\alpha_2 \approx 2.08$ が得られる. そして, $C_{(\lambda, \alpha_1, \alpha_2, \beta)} = 1$ に $\lambda, \alpha_1, \alpha_2$ を代入して β について解くと $\beta \approx 4.57$ が得られる.

3.3 数値実験

前節で構成した LPT-SN 分布に位置母数 $\mu (\in \mathbb{R})$ と尺度母数 $\sigma (> 0)$ を導入して, 外れ値が存在する場合の (μ, σ) の推定を考える. 考えるモデルは以下の通りである:

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

$$+ \varphi(\alpha_2; \phi) \frac{\alpha_2}{|(x - \mu)/\sigma|} \left(\frac{\log \alpha_2}{\log |(x - \mu)/\sigma|} \right)^\beta 1_{(\alpha_2, \infty)} \left(\frac{x - \mu}{\sigma} \right) \Big\}. \quad (6)$$

ただし, $\phi = (\mu, \sigma, \lambda)$ とし, $\mu \in \mathbb{R}$, $\sigma > 0$ とする.

3.3.1 数値実験その1

比較のために以下の位置尺度母数をもつ3つのモデルを考える:

1. skew normal (SN) 分布 (non-robust)
2. log-Pareto tailed skew-normal (LPT-SN) 分布 (whole robust)
3. 自由度 $\nu = 10$ の skew- t (St) 分布 (partial robust)

位置母数 μ と尺度母数 σ の最尤推定量 (MLE) $\hat{\mu}(\omega), \hat{\sigma}(\omega)$ の ω を変化させたときの挙動をみる. ただし, skew normal 分布の skew parameter $\lambda = 3$ と固定する. $SN(\mu, \sigma, 3)$ を用いて μ と σ の最尤推定量 (MLE) を求めたいが, skew normal 分布の MLE は陽に書けないことが知られているため, 数値的に求めた^{*1}. 図2は数値実験の結果を表しているが, skew normal, skew- t 分布を仮定した場合は外れ値が大きくなるにつれて MLE は不安定になっていくが, LPT-SN は外れ値が大きくなったとしても非常に安定していることがわかる.

3.3.2 数値実験その2

外れ値の割合も変えることを考える. ここでは, データ発生分布として以下の3つのモデルを考える:

1. $SN(0, 1, 3)$
2. $0.90 \times SN(0, 1, 3) + 0.10 \times N(0, 6)$
3. $0.95 \times SN(0, 1, 3) + 0.05 \times N(8, 1)$

各モデルからそれぞれ $n = 30$ の標本を生成し, このデータを用いて数値実験1で用いた3つのモデルの μ と σ の MLE の平均二乗誤差 (MSE) を 25,000 回のモンテカルロシミュレーションにより計算し比較を行った (表1,2). LPT-SN を用いると, 外れ値がない場合には他のモデルと同様の結果を示し, 外れ値が存在する場合には3つのモデルの中で MSE が最小となることがわかる. 特に, 尺度母数の推定の場合に LPT-SN に基づく推定量が圧倒していることがよくわかる.

^{*1} なお, モーメント推定量は陽にかけることが知られている (Azzalini, 2014).

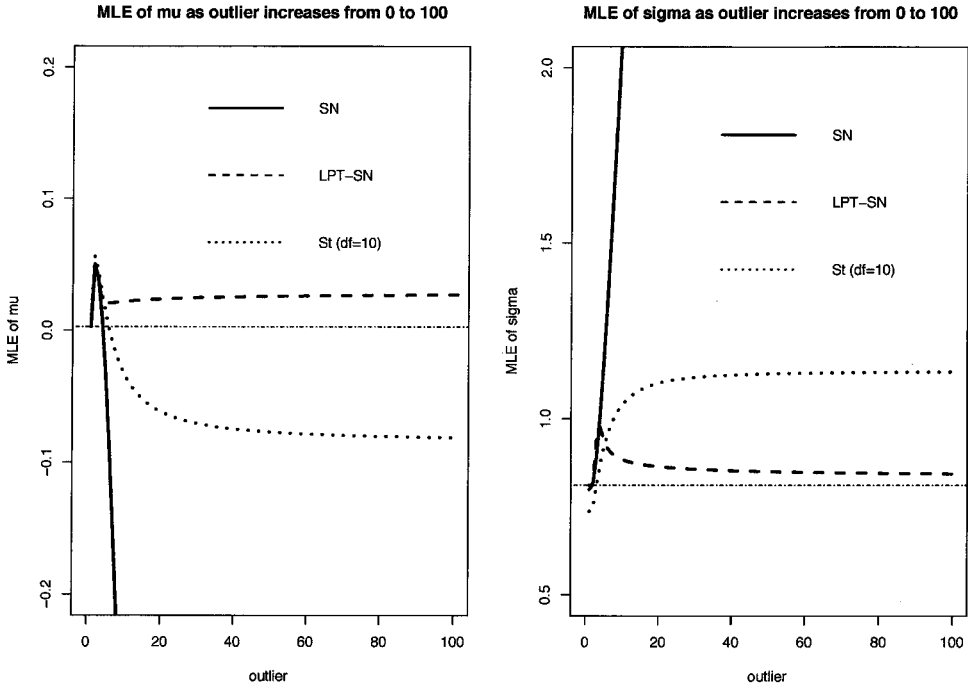


図2 3つのモデル (SN, LPT-SN, 自由度 10 の St) に対する, 位置・尺度母数の MLE $\hat{\mu}(\omega)$, $\hat{\sigma}(\omega)$ の一つの外れ値 ω を大きくしていったときの挙動.

表 1 MSE for MLE of 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

4 おわりに

本稿では, 対数正則変動関数に基づいた非対称な正規分布を構成し, 数値実験を通して位置尺度母数の MLE (あるいは MAP 推定量) が外れ値に関してロバストであることを示した. 対数正則変動関数を用いることにより, 推定量がロバストになることは Desgagné

表 2 MSE for MLE of 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

(2015) により示されているが、外れ値が十分大きい場合での結果である。実際に外れ値がそこまで大きな値でない場合には、外れ値の影響は部分的に残ることが予想される。また、ロバスト性を何らかの尺度によって特徴付ける必要がある。これらは今後の課題となってくる。

参考文献

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

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