Semiparametric Scale Mixture of Skew Normal Distributions

Namhwa, Lee
Thesis advisor: Prof. Byungtae, Seo

Department of Statistics, Sungkyunkwan University

Dec 6, 2019

Outline

1. Mixture model & Skew Normal Distribution.
2. Scale Mixture of Skew Normal Distribution
3. Semiparametric Scale Mixture of Skew Normal
4. Simulation
5. Real Data Analysis : Estimation
6. Conclusion

Mixture Models

Finite Mixture Models

- We can manage numerous data structure using Mixture Models.
- From Lindsay (1995), finite mixture models can be written as

$$f(x; \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{j=1}^{k} \pi_j f(x; \theta_j) \tag{1}$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)^T$ is a mixing probabilities, constrained to be non-negative and $\sum_{j=1}^{k} \pi_j = 1$, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T$ is a vector of mixing parameters.

Mixture Models

Latent Distributions

- Consider θ_j in (1) as an random variable that describes the specific attributes of the j-th component population.
- Let Q be a latent distribution of the component parameter θ_j.
- Then, estimation of unknown parameter $\boldsymbol{\pi}, \boldsymbol{\theta}$ is equivalent to estimate the unknown distribution Q.

Mixture Models

Discrete Latent Distribution

- The discrete probability measure $\boldsymbol{Q}$ puts mass π_j at the component parameter θ_j. That is,

$$f(x; \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{j=1}^{k} \pi_j f(x; \theta_j) = \mathbb{E}_x[f(x; \Theta)]. \tag{2}$$

Continuous Latent Distribution

- We can extend (2) to a continuous density, say $d\boldsymbol{Q}(\theta) = q(\theta)d\theta$. Then the mixture model becomes

$$f(x; \boldsymbol{Q}) := \int f(x; \theta) d\boldsymbol{Q}(\theta) = \int f(x; \theta) q(\theta) d\theta.$$

Nonparametric Mixture

Nonparametric Mixture Model (Lindsay (1995))

- If the latent distribution Q is treated as completely unspecified, this will be called the ***nonparametric mixture model***.
- Since we do not assume anything on Q, we can expect to obtain more flexible distribution family.
- If there exists another parameter η which is differ from mixing parameter θ, this will be called ***semiparametric mixture model***.

Estimation of Mixture Models

Case1 : Fixed Support Size

- The number of components k is assumed to be known.
- In order to estimate appropriately, we should choose a proper k.
- $\widehat{Q}$ that maximizes the likelihood can be estimated by EM-type algorithm.

Estimation of Mixture Models

Case2 : Flexible Support Size

- There is no assumption about the number of components k.
- It is a reasonable approach since, in general, we do not know how many components the data has.
- $\widehat{Q}$ that maximizes the likelihood cannot be estimated by EM-type algorithm.
- Algorithms for the flexible support size case leads to the mixture NPMLE theorem (Bohning (1982), Lindsay et al. (1983)).

Directional Derivative

Estimation

- NPMLE for mixture models can be characterized in terms of *directional derivative.*

Directional Derivative

The directional derivative is given by

$$D_Q(\sigma) = \sum_{i=1}^{n} \frac{f(x_i,\sigma)}{f(x_i,Q)} - n = \sum_{i=1}^{n} \frac{f(x_i,\sigma)}{\sum_{l=1}^{k} \pi_l f(x_i,\sigma_l)} - n$$

where Q is a probability measure with finite support and σ is a support point of mixture model.

Nonparametric Mixture

Idea for Computing NPMLE

- If $D_Q(\sigma^+) > 0$ for some σ^+, then the likelihood can be increased over $\ell(Q)$ by using a distribution with additional mass at σ^+.
- If $\hat{Q}$ is NPMLE, $\ell(\hat{Q}) > \ell(Q)$ for all Q.
- There exists an π such that

$$\ell((1-\pi)Q + \pi Q_\sigma^+) > \ell(Q),$$

where Q_σ^+ is a distribution with mass 1 at σ^+.

Nonparametric Mixture

Mixture NPMLE theorem in Lindsay (1995)

1. $\hat{Q}$ is the NPMLE if and only if $D_Q(\sigma) \leq 0 \ \ \forall\sigma$.
2. $\hat{Q}$ is discrete and has at most n support points σ.
3. $D_Q(\sigma) = 0$ for all support points of $\hat{Q}$.

Skew Normal Distribution

Skew Normal Distribution

- As developed by Azzalini (1985), a random variable Y follows a univariate skew normal distribution with location parameter ξ, scale parameter σ, and skewness parameter $\lambda \in \mathbb{R}$ if it has the density

$$\psi(y|\xi, \sigma^2, \lambda) = \frac{2}{\sigma}\phi\left(\frac{y-\xi}{\sigma}\right)\Phi\left(\lambda\frac{y-\xi}{\sigma}\right),$$

where $\phi(\cdot)$, $\Phi(\cdot)$ denote the standard normal density and cummulative distribution function, respectively.

Skew Normal Distribution

Moments of Skew Normal Distribution

- Suppose $Y \sim SN(\xi, \sigma^2, \lambda)$, then

$$\mathbb{E}\,(Y) = \xi + \sqrt{\frac{2}{\pi}}\delta(\lambda)\sigma, \quad \mathbb{V}\,(Y) = \left\{1 - \frac{2}{\pi}\delta^2(\lambda)\right\}\sigma^2.$$

$$\gamma_Y = \frac{\sqrt{2}(4-\pi)\lambda^3}{\{\pi + (\pi-2)\lambda^2\}^{3/2}}, \quad \kappa_Y = 3 + \left.\frac{8(\pi-3)\lambda^4}{\pi + (\pi-2)\lambda^2}\right.^{2}.$$

where $\delta(\lambda) = \lambda/\sqrt{1+\lambda^2}$ and γ_Y, κ_Y are the measures of skewness and kurtosis, respectively.

- Range : $-0.9953 < \gamma_Y < 0.9953,$

  $3 < \kappa_Y < 3.8692$

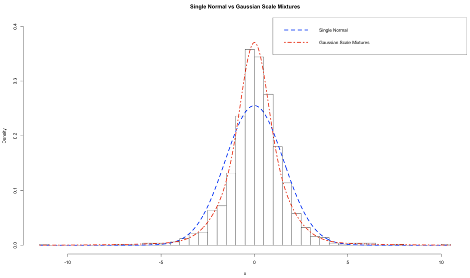
Scale Mixtures of Skew Normal

Gaussian Scale Mixtures

- Gaussian scale mixtures can manage the symmetric heavy tail family.

Examples: $X \sim t_3$ with $n = 1000$.

Single normal (blue) and gaussian scale mixture (red).



Scale Mixtures of Skew Normal

Scale Mixture of Skew Normal

- Branco and Dey (2001) suggested a continuous scale mixture of skew normal (SMSN) distribution.

Continous Scale Mixture of Skew Normal (da Silva Ferreira et al., 2011)

A random variable Y follows a distribution between the SMSN class with location parameter $\xi \in \mathbb{R}$, scale factor σ^2 and skewness parameter $\lambda \in \mathbb{R}$, if its pdf is given by

$$f(y) = 2\int_0^{+\infty} \phi(y|\xi, \sigma^2\kappa(u))\Phi_1(\lambda\frac{y-\xi}{\sigma})dQ(u), \tag{3}$$

where U is a positive random variable with cdf $Q(u; \boldsymbol{\nu})$.

SCALE MIXTURES OF SKEW NORMAL

Example : Continuous SMSN ((da Silva Ferreira et al., 2011))

- *Skew-Student-t normal distribution* with d.f. ν.
  - $\kappa(u) = 1/u$
  - $U \sim Gamma(\nu, \nu), \nu > 0$
- *Skew-Slash distribution.*
  - $\kappa(u) = 1/u, \quad 0 < u < 1$
  - $U \sim Beta(\nu, 1), \quad \nu > 0$

SEMIPARAMETRIC SMSN

Semiparametric Scale Mixture of Skew Normal

- Do not make any parametric assumptions on Q.
  That is, consider ***nonparametric mixture***.
- In this case, scale parameter σ is the only mixing parameter, and location parameter ξ and skewness parameter λ is unknown with fixed value.
- ***Semiparametric Mixture Model.***

Semiparametric SMSN

Semiparametric Scale Mixture of Skew Normal

A random variable Y follows a distribution between the *Semiparametric SMSN* class with location parameter $\xi \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}^+$ and skewness parameter $\lambda \in \mathbb{R}$, if its pdf is given by

$$f_Y(y; \boldsymbol{Q}) = \int f(y|\xi, \sigma, \lambda) d\boldsymbol{Q}(\sigma), \tag{4}$$

where $f(y|\xi, \sigma, \lambda)$ is a density of $SN(\xi, \sigma, \lambda)$ and $\boldsymbol{Q}$ is unspecified.

Semiparametric SMSN

Semiparametric Scale Mixture of Skew Normal

- Does "Semiparametric SMSN" has wider range of skewness and kurtosis than single Skew Normal has?
- Suppose that

$$f_Y(y; \boldsymbol{\pi}, \xi, \boldsymbol{\sigma}, \lambda) = \sum_{j=1}^{m} \pi_j f_j(y)$$

where $f_j(y)$ is a density of $SN(\xi, \sigma_j, \lambda)$, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)^T$, and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_m)^T$.

SEMIPARAMETRIC SMSN

Moments of Finite Scale Mixture of Skew Normal

- Then, the moments of the random variable that has j-th component density are

$$\mu_j = \xi + \sqrt{\frac{2}{\pi}}\delta(\lambda)\sigma_j, \quad \nu_j^2 = \left\{1 - \frac{2}{\pi}\delta^2(\lambda)\right\}\sigma_j^2.$$

$$\gamma = \frac{\sqrt{2}(4-\pi)\lambda^3}{\{\pi + (\pi-2)\lambda^2\}^{3/2}}, \quad \kappa = 3 + \frac{8(\pi-3)\lambda^4}{\{\pi + (\pi-2)\lambda^2\}^2}.$$

where $\delta(\lambda) = \lambda/\sqrt{1+\lambda^2}$ and γ, κ are the measures of skewness and kurtosis, respectively.

Semiparametric SMSN

Moments of Finite Scale Mixture of Skew Normal

- Then, the moments of the random variable Y are

$$\mathbb{E}(Y) = \sum_{j=1}^{m} \pi_j \mu_j, \; V(Y) = \sum_{j=1}^{m} \pi_j \nu_j^2 + \sum_{j=1}^{m} \pi_j (\mu_j - \mu)^2,$$

$$\gamma_Y = \frac{1}{\nu^3} \sum_{j=1}^{m} \pi_j \{\gamma \nu_j^3 + (\mu_j - \mu)^3 + 3(\mu_j - \mu)\nu_j^2\}$$

$$\kappa_Y = \frac{1}{\nu^4} \sum_{j=1}^{m} \pi_j \{\kappa \nu_j^4 + (\mu_j - \mu)^4 + 4(\mu_j - \mu)\gamma \nu_j^3 + 6(\mu_j - \mu)^2 \nu_j^2\}.$$

SEMIPARAMETRIC SMSN

Moments of Finite Scale Mixture of Skew Normal

- Consider 2-component scale mixture of skew normal. That is,

$$Y \sim 0.6SN(0,1,5) + 0.4SN(0,3,5)$$

- Then, $\gamma_Y = 1.82502$, $\kappa_Y = 7.03653$, which is outside the range of skewness and kurtosis of single skew normal.
- Since semiparametric mixtures include finite mixtures, we can make wider distribution family that contains the asymmetric, heavy tail distributions.

SEMIPARAMETRIC SMSN

Estimation of Semiparametric SMSN

- Sometimes, σ that is closed to 0 makes likelihood unbounded.
- That is, likelihood is maximized when the scale parameter becomes 0.
- Then, 0 is included in the set of support point, but we do not want this kind of estimate.
- Support point σ near to 0 is called ***spurious support***.

SEMIPARAMETRIC SMSN

Estimation of Semiparametric SMSN

- The *spurious support* problem can be solved by estimating the model after adding some constant $c > 0$ to σ (Seo et al., 2017). That is, consider our model (4) as

$$f_Y(y; \boldsymbol{Q}) = \int f(y|\xi, \sigma + c, \lambda) d\boldsymbol{Q}(\sigma).$$

  with constant $c > 0$.

- Then, directional derivatives become

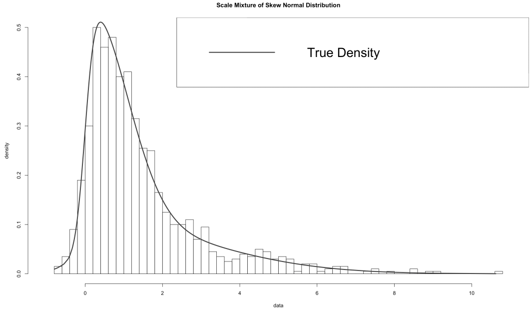
$$\begin{aligned} D_Q(\sigma) &= \sum_{i=1}^{n} \frac{f(x_i, \sigma + c)}{f(x_i, Q)} - n \\ &= \sum_{i=1}^{n} \frac{f(x_i, \sigma + c)}{\sum_{l=1}^{k} \pi_l f(x_i, \sigma_l + c)} - n \end{aligned}$$

Simulation

Simulation Setting

- We fixed the location parameter ($\xi = 0$) and the skewness parameter ($\lambda = 5$). The data follows

$$X \sim 0.6\, SN(\xi, \sigma = 1, \lambda) + 0.4\, SN(\xi, \sigma = 3, \lambda), n = 1000.$$



ISDM Algorithm

Intra Simplex Direction Method

- Lesperance and Kalbfleisch (1992) suggested *Intra Simplex Direction Method*(ISDM) as a faster algorithm than VDM.
- Instead of taking the only global maxima as support points, add all local maxima.
- Find an appropriate convex combination of the selected support points.

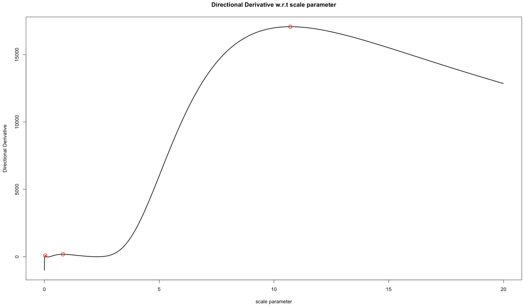
ISDM Algorithm

Intra Simplex Direction Method

- **Step 1** : Assign the initial values, say $\xi_0, \sigma_0, \lambda_0$, to the parameter ξ, σ, and λ, respectively.
- **Step 2** : With fixed $\xi^{(t)}, \lambda^{(t)}$, find all local maxima and add them to the set of current support points, $\boldsymbol{Q}^{(t)}$.
- **Step 3** : Find proper mixing probabilities corresponding to each support points and set this as $\boldsymbol{Q}^{(t+1)}$.
- **Step 4** : With fixed $\boldsymbol{Q}^{(t+1)}$, find $\xi^{(t+1)}, \lambda^{(t+1)}$ that maximize the likelihood.
- **Step 5** : Go back to **Step 2**.

Simulation

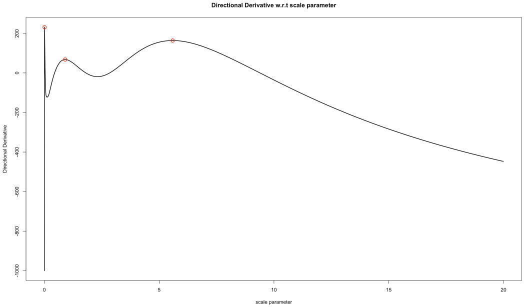
ISDM Algorithm : 1st Iteration



Support	$\sigma_0 = 2.2997$	$\sigma_{11} = 0.1384$	$\sigma_{12} = 0.9087$	$\sigma_{13} = 10.8099$
Mixing Probability	$\pi_0 = 0.7459$	$\pi_{11} = 2.23\text{e-}14$	$\pi_{12} = 0.2331$	$\pi_{13} = 0.0210$

Simulation

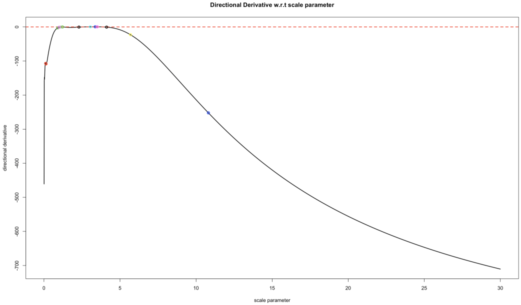
ISDM Algorithm : 2nd Iteration



σ	2.2997	0.1384	0.9087	10.8099	0.11	1.0071	5.6891
π	0.5296	7.73e-232	2.24e-07	7.23e-308	8.46e-231	0.4061	0.0643

SIMULATION

ISDM : Directional Derivative Plot



- If $\hat{Q}$ is NPMLE, then, $D_{\hat{Q}}(\sigma) \leq 0$ for all scale parameter σ.

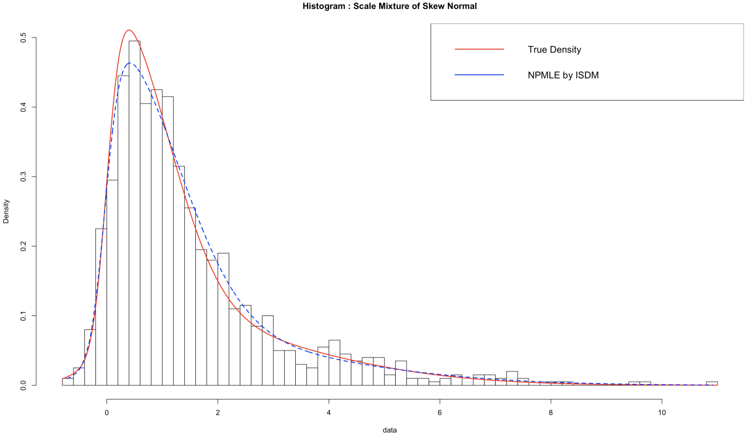
Nonparametric Mixture

Mixture NPMLE theorem in Lindsay (1995)

1. $\hat{Q}$ is the NPMLE if and only if $D_Q(\sigma) \leq 0 \ \ \forall \sigma$.
2. $\hat{Q}$ is discrete and has at most n support points σ.
3. $D_Q(\sigma) = 0$ for all support points of $\hat{Q}$.

Simulation

ISDM Algorithm : Result



Simulation

ISDM Algorithm

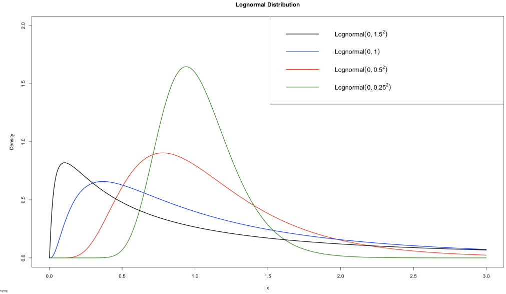
Type	Iteration(Support)	Location(ξ)	Skewness(λ)
ISDM	6(14)	-0.0293	5.7306

- We can check that NPMLE is similar to the true density.
- That is, we can estimate finite scale mixture of skew normal by using ***Semiparametric Scale Mixture of Skew Normal***.

Simulation

Log normal distribution

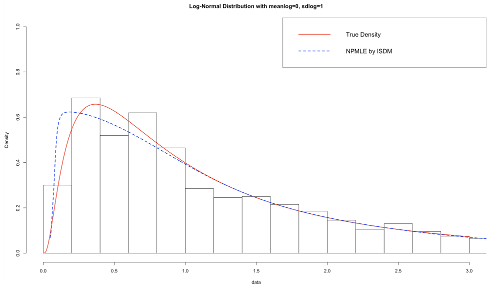
- Log-normal distribution is well-known for its asymmetry and heavy tail.



distribution.png

Simulation

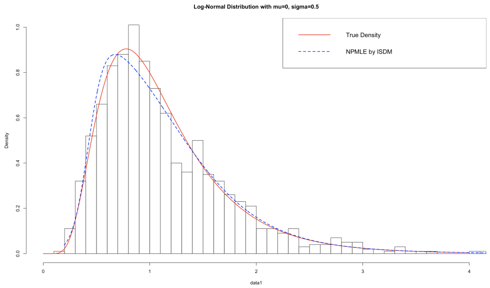
Case 1: Log normal with $\mu = 0, \sigma = 1$



- With $\mu = 0, \sigma = 1$, log-normal density becomes highly asymmetric and heavy tailed distribution
- We can check that NPMLE is similar to the true density.

SIMULATION

Case 2: Log normal with $\mu = 0, \sigma = 0.5$

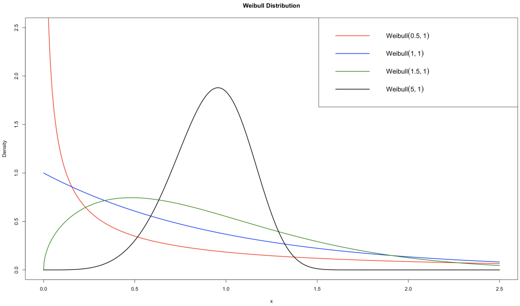


- With $\mu = 0, \sigma = 0.5$, log-normal density becomes quite asymmetric.
- We can check that NPMLE is similar to the true density.

SIMULATION

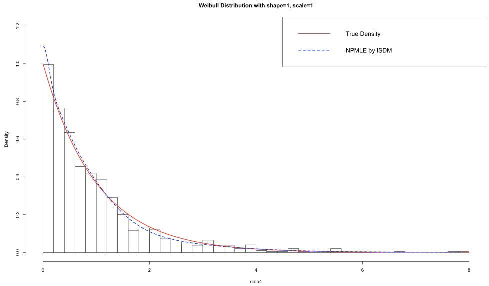
Weibull distribution

- Weibull distribution is also well-known for its asymmetry and mainly used in Survival Analysis.



Simulation

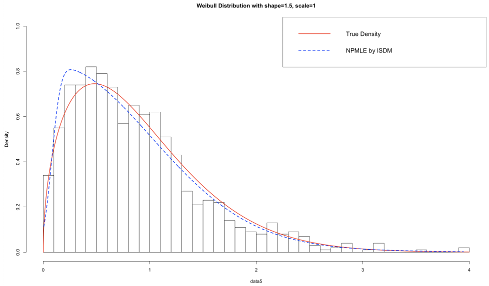
Case 1: Weibull$(1, 1)$



- With shape parameter $k = 1$, and scale parameter $\lambda = 1$, weibull density becomes exponential distribution with mean 1.
- We can check that NPMLE is similar to the true density.

Simulation

Case 2: Weibull$(1.5, 1)$



- With shape parameter $k = 1.5$, and scale parameter $\lambda = 1$, weibull density becomes highly asymmetric and heavy tailed distribution.
- We can check that NPMLE is similar to the true density.

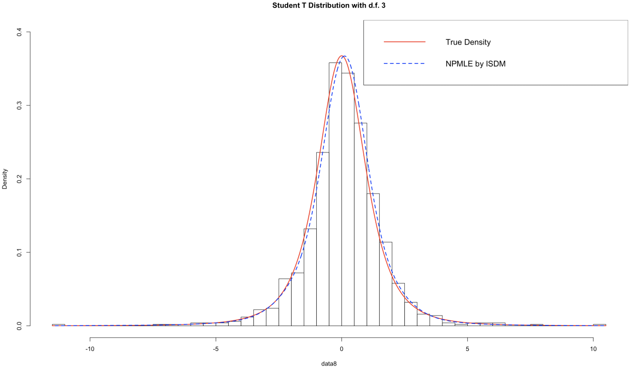
Simulation

Symmetric Distribution

- ***Semiparametric Scale Mixture of Normal(SMN)*** class contains almost all of unimodal and symmetric distributions.
- *t* distribution with d.f. 3 and laplace distribution are included in *semiparametric SMN*.
- ***Semiparametric SMSN*** class contains *semiparametric SMN*.

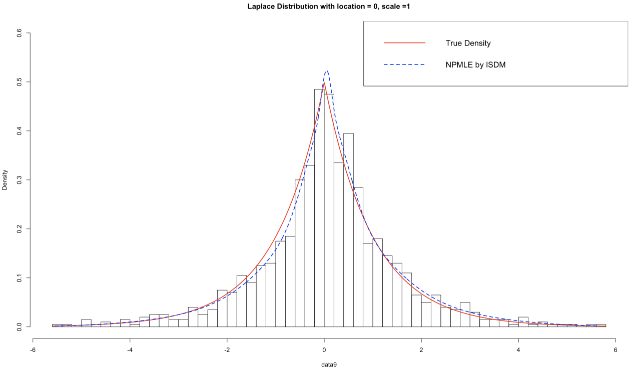
Simulation

Symmetric Distribution



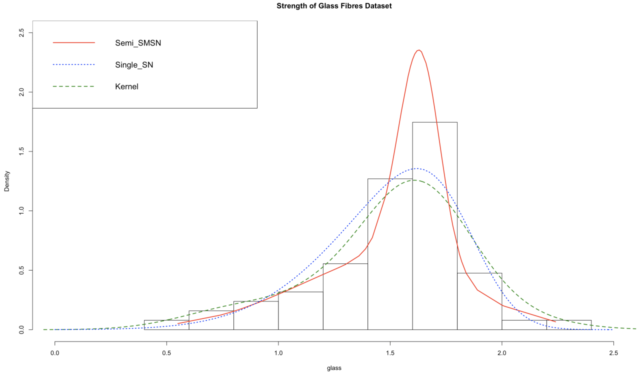
Simulation

Symmetric Distribution



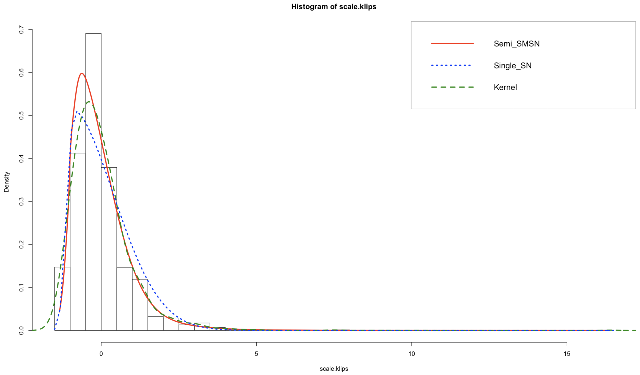
Real Data Analysis

Density Estimation : Strength of Glass Fibres Data



Real Data Analysis

Density Estimation : KLIPS Data



Conclusion

Conclusion

- We suggested ***Semiparametric Scale Mixture of Skew Normal*** distribution class which is wider than Semiparametric Scale Mixture of Normal class.
- We can manage asymmetric and heavy tailed data, using ***Semiparametric SMSN***.
- Now, we are identifying the theoretical coverage of ***Semiparametric SMSN*** by cheking the range of skewness and kurtosis.
- In the future, we plan to apply ***Semiparametric SMSN*** to the ***Modal Regression***.

References I

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178.

Bohning, D. (1982). Convergence of simar's algorithm for finding the maximum likelihood estimate of a compound poisson process. *The Annals of Statistics*, pages 1006–1008.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79(1):99–113.

da Silva Ferreira, C., Bolfarine, H., and Lachos, V. H. (2011). Skew scale mixtures of normal distributions: properties and estimation. *Statistical Methodology*, 8(2):154–171.

Lesperance, M. L. and Kalbfleisch, J. D. (1992). An algorithm for computing the nonparametric mle of a mixing distribution. *Journal of the American Statistical Association*, 87(417):120–126.

REFERENCES II

Lindsay, B. G. (1995). Mixture models: theory, geometry and applications. In *NSF-CBMS regional conference series in probability and statistics*, pages i–163. JSTOR.

Lindsay, B. G. et al. (1983). The geometry of mixture likelihoods: a general theory. *The annals of statistics*, 11(1):86–94.

Seo, B., Noh, J., Lee, T., and Yoon, Y. J. (2017). Adaptive robust regression with continuous gaussian scale mixture errors. *Journal of the Korean Statistical Society*, 46(1):113–125.

Appendix

Directional Derivative

- Let P and Q are probability measures with finite support. For $\alpha \in [0,1]$, define

$$\ell^*(\alpha) := \ell((1-\alpha)Q + \alpha P) = \sum_{i=1}^{n} \log f(x_i, (1-\alpha)Q + \alpha P)$$
$$= \sum_{i=1}^{n} \log((1-\alpha) f(x_i, Q) + \alpha f(x_i, P)).$$

- Taking the derivative of $\ell^*(\alpha)$ at $\alpha = 0$ gives

$$\frac{\partial}{\partial \alpha} \ell^*(\alpha) = \Phi(Q, P) = \lim_{\alpha \to 0} \frac{\ell((1-\alpha)Q + \alpha P) - \ell(Q)}{\alpha}. \tag{5}$$

Appendix

Directional Derivative

- Applying l'Hôpital's rule to (4) leads to

$$\Phi(Q,P) = \sum_{i=1}^{n} \frac{f(x_i,P) - f(x_i,Q)}{f(x_i,Q)}.$$

- In particular, for the one point mass P_σ at σ the directional derivative is given by

$$\begin{aligned} D_Q(\sigma) &= \sum_{i=1}^{n} \frac{f(x_i,\sigma)}{f(x_i,Q)} - n \\ &= \sum_{i=1}^{n} \frac{f(x_i,\sigma)}{\sum_{l=1}^{k} \pi_l f(x_i,\sigma_l)} - n \end{aligned}$$

Appendix

ISDM : Pseudocode

From an initial estimate, G_1 and $j = 1$,

Step 1 : Compute all local maxima $\theta^*_{j1}, \ldots, \theta^*_{jm_j}$ of $g(\theta, G_j), \theta \in \Omega$, such that $g(\theta^*_{js}, G_j) \geq 0 (s = 1, \ldots, m_j < k)$. If $max_s g(\theta^*_{js}, G_j) = 0$, stop.

Step 2 : Compute $\epsilon_{j0}, \ldots, \epsilon_{jm_j}$, the values that maximize

$$\ell\left\{\epsilon_0 G_j + \sum_{s=1}^{m_j} \epsilon_s \delta_{\theta^*_{js}}\right\} = \sum_r n_r \log\left\{\epsilon_0 L_r(G_j) + \sum_{s=1}^{m_j} \epsilon_s L_r(\theta^*_{js})\right\}$$

subject to $\sum_{s=0}^{m_j} \epsilon_s = 1$ and $\epsilon_s \geq 0 (s = 0, \ldots, m_j)$.

Step 3 : $G_{j+1} = \epsilon_{j0} G_j + \sum_{s=1}^{m_j} \epsilon_{js} \delta_{\theta^*_{js}}$. Set $j = j + 1$ and go to **Step 1.**