Article

Robust Mixture Modeling Based on Two-Piece Scale Mixtures of Normal Family

Mohsen Maleki ¹ , Javier E. Contreras-Reyes ^{2,*} and Mohammad R. Mahmoudi ³

¹ Department of Statistics, College of Sciences, Shiraz University, Shiraz 71946 85115, Iran; m.maleki.stat@gmail.com

² Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción 4081112, Chile

³ Department of Statistics, Faculty of Science, Fasa University, Fasa 74616 86131, Iran; mahmoudi.m.r@fasau.ac.ir

* Correspondence: jecontrr@uc.cl; Tel.: +56-9-7960-8218

Received: 18 February 2019; Accepted: 27 March 2019; Published: 1 April 2019

Abstract: In this paper, we examine the finite mixture (FM) model with a flexible class of two-piece distributions based on the scale mixtures of normal (TP-SMN) family components. This family allows the development of a robust estimation of FM models. The TP-SMN is a rich class of distributions that covers symmetric/asymmetric and light/heavy tailed distributions. It represents an alternative family to the well-known scale mixtures of the skew normal (SMSN) family studied by Branco and Dey (2001). Also, the TP-SMN covers the SMN (normal, t, slash, and contaminated normal distributions) as the symmetric members and two-piece versions of them as asymmetric members. A key feature of this study is using a suitable hierarchical representation of the family to obtain maximum likelihood estimates of model parameters via an EM-type algorithm. The performances of the proposed robust model are demonstrated using simulated and real data, and then compared to other finite mixture of SMSN models.

Keywords: ECME algorithm; finite mixture model; maximum likelihood estimates; scale mixtures of normal family; two-piece distributions

1. Introduction

Finite mixture models are highly demanded in machine-learning analysis, due to their properties, computational tractability, and for being a good approximation for continuous densities [1]. They are also an important statistical tool for many applications in clustering, discriminant analysis, image processing and satellite imaging [2]. Beyond the already known results provided for the finite mixture of normal distributions (FM-NOR) model in the literature [1], recent developments cover symmetric/asymmetric and light/heavy tailed distributions. One of these is the novel class of finite mixture of multivariate skew-normal mixture (FM-SN) models [3,4], which provides some advantages over the normal mixtures: the normal components allow an arbitrarily close modeling of any distribution by increasing the number of components, and, in the context of supervised learning, groups of observations represented by asymmetrically distributed data can lead to the wrong classification. The components of skew-normal mixture models, however, capture skewness due to their flexibility [1]. In addition, a robust extension of the FM-SN model to robust finite mixture of skew-t (FM-ST) has been done in the influential works of [3,5–7]. The FM-ST components, too, capture both skewness and extreme observations due to their flexibility [8].

The SMSN family is a rich and very strong flexible class of distributions which covers the light/heavy-tailed distributions; e.g., skew-normal (SN), skew-t (ST), skew-slash (SSL) and skew contaminated-normal (SCN) distributions, and has been widely considered in many statistical

models, especially FM models (see e.g., [5,9–15]). The SMSN family is an extension of the skewed version of the well-known symmetric scale mixtures of the normal (SMN) family which contains the light/heavy-tailed members: the normal (N), t (T), slash (SL) and contaminated-normal (CN) distributions [16]. Lange et al. [17], Lange and Sinsheimer [18], and Maleki and Nematollahi [19] used the SMN family in an application of robust statistical modeling. A two-piece distribution based on the symmetrical distributions with various scales is an alternative approach to model atypical data (see e.g., [10,20–24]). In our approach, we have used the two-piece distributions based on the SMN family. This family, called the two-piece distributions based on the scale mixtures of normal (TP-SMN), and analogy of the SMSN family, contains the light/heavy-tailed members: the two-piece normal (TP-N), two-piece t (TP-T), two-piece slash (TP-SL) and two-piece contaminated-normal (TP-CN) distributions as its members.

In this paper, we consider the TP-SMN family of distributions as a two-component mixture of truncated SMN distributions on a special two partition of the real domain ($\mathbb{R}$), and then propose the finite mixture of this family, called FM-TP-SMN models. It represents an alternative family to the well-known scale mixtures of skew normal (SMSN) family studied by [25]. We have also used a hierarchical representation of the FM-TP-SMN and implemented an expectation-maximization (EM)-type algorithm for finding the maximum likelihood (ML) estimates of the proposed model. Studies by [21,23], show that by truncating the distribution in two partitions, makes it possible to obtain a better fit of empirical distribution because, the subjacent process of the complete likelihood is modeled. This way, the “two-piece” modeling is a direct competitor against the FM-SMSN family of distributions [21].

The rest of this paper is organized as follows. In Section 2, we review some main properties of the TP-SMN family and represent this family as a two-component mixture of the truncated SMN distributions. In Section 3, the FM-TP-SMN model is introduced and the ML estimates of the proposed model parameters via an EM-type algorithm are provided. In Section 4, numerical studies with an application of the proposed models and estimates are considered. Some conclusions and ideas for future research are offered in Section 5.

2. The Two-Piece Scale Mixtures of Normal Distributions

In this section, we analyze some necessary properties of the TP-SMN family of distributions for our proposed FM model.

The well-known SMN family introduced by [16] (the basis of the robust asymmetric TP-SMN family), has the following probability density function (PDF) and stochastic representation. Let $X \sim \text{SMN}(\mu, \sigma, \nu)$, then its PDF is

$$f_{\text{SMN}}(x|\mu, \sigma, \nu) = \int_0^\infty \phi(x|\mu, u^{-1}\sigma^2) dH(u|\nu), \quad x \in \mathbb{R}, \quad (1)$$

and its stochastic representation is

$$X = \mu + \sigma U^{-1/2} W, \quad (2)$$

where $\phi(\cdot|\mu, \sigma^2)$ represents the density of $N(\mu, \sigma^2)$ distribution, $H(\cdot|\nu)$ is the cumulative distribution function (CDF) of the scale mixing random variable U , which can be indexed by a scalar or vector of parameters ν , and W is a standard normal random variable that is independent of U .

The TP-SMN is a rich family of distributions that covers the asymmetric light-tailed TP-N (also called the epsilon-skew-normal; [26]), the asymmetric heavy-tailed TP-T, TP-SL and TP-CN distributions, and their corresponding symmetric members. Note that symmetric members of the TP-SMN and SMSN classes are the SMN family. In terms of density, for $y \in \mathbb{R}$ this family can be represented as

$$g(x|\mu, \sigma, \gamma, \nu) = \begin{cases} 2(1-\gamma)f_{\text{SMN}}(x|\mu, \sigma(1-\gamma), \nu), & y \leq \mu, \\ 2\gamma f_{\text{SMN}}(x|\mu, \sigma\gamma, \nu), & y > \mu, \end{cases} \quad (3)$$

where $0 < \gamma < 1$ is the slant parameter, $f_{SMN}(\cdot|\mu, \sigma, \nu)$ is given by (1) and is denoted by $Y \sim TP-SMN(\mu, \sigma, \gamma, \nu)$ with $E(Y) = \mu - b\sigma(1 - 2\gamma)$ and $Var(Y) = \sigma^2[c_2k_2(\nu) - b^2c_1^2]$, for which $b = \sqrt{2/\pi}k_1(\nu)$, $c_r = \gamma^{r+1} + (-1)^r(1 - \gamma)^{r+1}$, $k_r(\nu) = E(U^{-r/2})$, and U is the scale mixing random variable in (2).

Different TP-SMN member distributions in (3) are obtained by several distributions for scale mixing random variable U in (2), as follows:

- Two-piece normal (TP-N): $U = 1$ with probability one,
- Two-piece t (TP-T): $U \sim \text{Gamma}(\nu/2, \nu/2)$, i.e. $\nu = \nu$,
- Two-piece slash (TP-SL): $U \sim \text{Beta}(\nu, 1)$, i.e. $\nu = \nu$,
- Two-piece contaminated normal (TP-CN): $h(u|\nu) = \nu I_{(u=1)} + (1 - \nu)I_{(u=1)}$, i.e. $\nu = (\nu, \tau)^\top$.

For more details and statistical properties of the TP-SMN family, see [20,23].

Further, the two-piece distributions can be represented as the two-component mixture with separated supports, i.e., left and right half basic distributions [20] (Equation (4)), especially when $Y \sim TP-SMN(\mu, \sigma, \gamma, \nu)$ with PDF given in (3), two-component mixture left and right half SMN distributions with special component probabilities as follows:

$$g(x|\mu, \sigma_1, \sigma_2, \nu) = 2 \frac{\sigma_1}{\sigma_1 + \sigma_2} f_{SMN}(y|\mu, \sigma_1, \nu) I_{(-\infty, \mu]}(y) + 2 \frac{\sigma_2}{\sigma_1 + \sigma_2} f_{SMN}(x|\mu, \sigma_2, \nu) I_{(\mu, +\infty)}(y). \quad (4)$$

Note in Equation (4) that, the scale parameter σ and slant parameter γ in Equation (3) are recovered in the form of $\sigma = \sigma_1 + \sigma_2$ and $\gamma = \sigma_2 / (\sigma_1 + \sigma_2)$.

By using auxiliary (latent) variables S_j , $j = 1, 2$; in terms of the components of the mixture in Equation (4), the TP-SMN random variable can have the following stochastic representation

$$\begin{cases} Y|S_1 = 1 \sim SMN(\mu, \sigma_1, \nu) I_A(y_i), \\ Y|S_2 = 1 \sim SMN(\mu, \sigma_2, \nu) I_{A^c}(y_i), \end{cases} \quad (5)$$

where $A = (-\infty, \mu)$ and $SMN(\cdot) I_A(\cdot)$ denotes the truncated SMN distribution on the interval A , and $\mathbf{S} = (S_1, S_2)^\top$ has a multinomial distribution with following probability mass function (PMF):

$$P(\mathbf{S} = \mathbf{s}) = \left(\frac{\sigma_2}{\sigma_1 + \sigma_2} \right)^{z_1} \left(\frac{\sigma_1}{\sigma_1 + \sigma_2} \right)^{z_2}; \quad s_1, s_2 = 0, 1, \quad (6)$$

and is denoted by $\mathbf{S} \sim M(1, \sigma_1 / (\sigma_1 + \sigma_2), \sigma_2 / (\sigma_1 + \sigma_2))$. Note that each component-label is a Bernoulli random variable $S_k \sim \text{Binomial}(1, \sigma_k / (\sigma_1 + \sigma_2))$; $k = 1, 2$, such that $S_1 + S_2 = 1$.

3. Finite Mixtures TP-SMN

In this section, we introduce the finite mixture of TP-SMN (FM-TP-SMN) model and obtain the ML estimates of this model's parameters.

3.1. FM-TP-SMN Model

Here, we consider a distribution represented as a g -component mixture of TP-SMN distributions. In terms of density, this mixture distribution is characterized by the following density:

$$f(y|\Theta) = \sum_{j=1}^g \pi_j g(y|\mu_j, \sigma_j, \gamma_j, \nu_j), \quad y \in \mathbb{R}, \quad (7)$$

where $\Theta = (\pi_1, \dots, \pi_g, \mu_1, \dots, \mu_g, \sigma_1, \dots, \sigma_g, \gamma_1, \dots, \gamma_g, \nu_1, \dots, \nu_g)^\top$, for which $\mathbf{\pi} = (\pi_1, \dots, \pi_g)^\top$ with $\pi_j > 0$, $j = 1, \dots, g$, $\sum_{j=1}^g \pi_j = 1$, and, for $j = 1, \dots, g$, $g(y|\mu_j, \sigma_j, \gamma_j, \nu_j)$ is an $TP-SMN(\mu_j, \sigma_j, \gamma_j, \nu_j)$ -component density as defined in (1). Also, we write $Y \sim FM-TP-SMN(\Theta)$ to say that a random variable Y has an FM-TP-SMN distribution as defined by (7).

Concerning the parameter ν_j of the mixing distribution $H(\cdot|\nu_j)$, for $j = 1, \dots, g$, it is worth noting that it can be a vector of parameters, e.g. the contaminated normal distribution. Thus, for computational convenience we assume that $\nu_1 = \dots = \nu_g = \nu$ (see also [5]).

In terms of the components of the mixtures, Equation (7) can be equivalently obtained by

$$Y|Z_j = 1 \sim TP - SMN(\mu_j, \sigma_j, \gamma_j, \nu), \quad j = 1, \dots, g, \quad (8)$$

where $Z = (Z_1, \dots, Z_g)^\top \sim \text{Multinomial}(1, \pi_1, \dots, \pi_g)$ is a multinomial (component-label) vector with probability mass function $P(Z_1 = z_1, \dots, Z_g = z_g) = \pi_1^{z_1} \pi_2^{z_2} \dots \pi_g^{z_g}$, $z_j = 0, 1$; $j = 1, \dots, g$, $\sum_{j=1}^g z_j = 1$.

Since only one component of Z can be equal to one (remaining ones are zero), events $\{Z_j = 1\}$ and $\{Z_j = 1, Z_r = 0; \forall j \neq r\}$ are equivalent, indicating thus that the distribution of Y corresponds to the i -th component of the mixture; for further details, see e.g., [1].

Remark 1. Let $Y \sim FM - TP - SMN(\Theta)$, then the mean and variance of Y are, respectively, given by

$$E[Y] = \sum_{j=1}^g \pi_j E[Y|Z_j = 1] = \sum_{j=1}^g \pi_j E[X_j],$$

and

$$\text{Var}[Y] = \sum_{j=1}^g \pi_j \{ \text{Var}[Y|Z_j = 1] + (E[Y|Z_j = 1] - E[Y])^2 \} = \sum_{j=1}^g \pi_j \{ \text{Var}[X_j] + (E[X_j] - \bar{\mu})^2 \},$$

where $X_j \sim TP - SMN(\mu_j, \sigma_j, \gamma_j, \nu)$, $j = 1, \dots, g$, and $\bar{\mu} = \sum_{j=1}^g \pi_j E[X_j]$ (see e.g., [2]).

The FM-TP-ESN densities in (7) are an extremely flexible class which includes the finite mixtures of SMN densities as special case, when $\varepsilon_j = 0$, $j = 1, \dots, g$.

For each i.i.d. sample in the form of $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$, by considering the PDF (7), the log-likelihood function is

$$\ell(\Theta|\mathbf{Y}) = \sum_{i=1}^n \log \left(\sum_{j=1}^g \pi_j g(y|\mu_j, \sigma_j, \gamma_j, \nu) \right). \quad (9)$$

3.2. ML Estimates of Model Parameters

We can utilize a (latent) indicator (allocation) variables $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ig})^\top$, $i = 1, \dots, n$, to assign observations belonging to different components of the mixture ($j = 1, \dots, g$), so in terms of Z_{ij} , we can conclude that

$$Y_i|Z_{ij} = 1 \sim^{ind.} TP - SMN(\mu_j, \sigma_j, \gamma_j, \nu), \quad P(Z_{ij} = 1) = \pi_j; \quad i = 1, \dots, n, \quad j = 1, \dots, g,$$

and so using Equations (2) and (5) with $S_{ij} = (S_{ij1}, S_{ij2})^\top$, $i = 1, \dots, n$, we have that

$$\begin{aligned} Y_i|U_{ij}, Z_{ij} = 1, S_{ijk} = 1 &\sim^{ind.} N(\mu_j, u_{ij}^{-1/2} \sigma_j^2) I_{A_j}(y_i)^{2-k} I_{A_j^c}(y_i)^{k-1}, \\ U_{ij}|Z_{ij} = 1, S_{ijk} = 1 &\sim^{ind.} H(u_{ij}|\nu), \\ S_{ij}|Z_{ij} = 1 &\sim^{iid.} M\left(1, \frac{\sigma_{1j}}{\sigma_{1j} + \sigma_{2j}}, \frac{\sigma_{2j}}{\sigma_{1j} + \sigma_{2j}}\right), \\ Z_i &\sim^{iid.} M(n, \pi_1, \dots, \pi_g), \end{aligned} \quad (10)$$

for $i = 1, \dots, n; j = 1, \dots, g; k = 1, 2$, $A_j = (-\infty, \mu_j]$ and $N(\cdot)I_A(\cdot)$ denotes the truncated normal distribution on the interval A .

The above hierarchical representation of the FM-TP-SMN model will be used to obtain the ML estimates via an ECME-algorithm. This algorithm is a generalization of the ECM-algorithm introduced by [27], which is an extension of the EM-algorithm [28]. It can be obtained by replacing some CM-steps, which maximize the constrained expected complete-data log-likelihood function, with steps that maximize the corresponding constrained actual likelihood function. As [27,29] indicated, the joint ML estimates obtained by ECME-algorithms are much more efficient than other EM-type algorithms.

Let $\mathbf{C} = \{\mathbf{Y}, \mathbf{S}, \mathbf{Z}\}$ denotes the complete data, where $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ is the observed sample $\mathbf{S} = (\mathbf{S}_1^\top, \dots, \mathbf{S}_n^\top)^\top$ and $\mathbf{Z} = (\mathbf{Z}_1^\top, \dots, \mathbf{Z}_n^\top)^\top$ are the latent or unobserved variables from the FM-TP-SMN model with vector of parameters $\boldsymbol{\Theta} = (\pi_1, \dots, \pi_g, \mu_1, \dots, \mu_g, \sigma_{11}, \dots, \sigma_{1g}, \sigma_{21}, \dots, \sigma_{2g}, \nu)^\top$. Considering the hierarchical representation (10), the completed (augmented) likelihood function is given by

$$L_C(\boldsymbol{\Theta}) = \prod_{i=1}^n \prod_{j=1}^g \prod_{k=1}^2 (\phi(y_i | \mu_j, u_{ijk}^{-1} \sigma_k^2) h(u_{ijk} | \nu) p(\mathbf{z}_i | \pi_1, \dots, \pi_g) p(\mathbf{s}_i | \sigma_{1j}, \sigma_{2j}) I_{A_j}(y_i)^{2-k} I_{A_j^c}(y_i)^{k-1})^{z_{ij} s_{ijk}},$$

where $A_j = (-\infty, \mu_j]$. After ignoring constants and using auxiliary (latent) variables the completed log-likelihood function is in the form:

$$\ell(\boldsymbol{\Theta}) = -\sum_{i=1}^n \sum_{j=1}^g z_{ij} \log(\sigma_{1j} + \sigma_{2j}) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^g \sum_{k=1}^2 \frac{z_{ij} s_{ijk} u_{ijk}}{\sigma_{kj}^2} (y_i - \mu_j)^2 + \sum_{i=1}^n \sum_{j=1}^g \sum_{k=1}^2 z_{ij} s_{ijk} \log h(u_{ijk} | \nu). \quad (11)$$

Quantities $\hat{z}_{ij} = E[Z_{ij} | \hat{\boldsymbol{\Theta}}, y_i]$, $\hat{s}_{ijk} = E[S_{ijk} | \hat{\boldsymbol{\Theta}}, y_i, Z_{ij} = 1]$ and $\hat{w}_{ijk} = E[Z_{ij} S_{ijk} U_{ijk} | \hat{\boldsymbol{\Theta}}, y_i]$, must be defined, and using known properties of conditional expectation and PDF in (4), we obtain $\hat{w}_{ijk} = \hat{z}_{ij} \hat{s}_{ijk} \hat{\kappa}_{ijk}$, where $\hat{\kappa}_{ijk} = E[U_{ijk} | \hat{\boldsymbol{\Theta}}, y_i, Z_{ij} = 1, S_{ijk} = 1]$, $i = 1, \dots, n$, $j = 1, 2$, and

$$\hat{z}_{ij} = E[Z_{ij} | \hat{\boldsymbol{\Theta}}, y_i] = \frac{\pi_j g(y_i | \hat{\mu}_j, \hat{\sigma}_{1j}, \hat{\sigma}_{2j}, \hat{\nu})}{\sum_{j=1}^g \pi_j g(y_i | \hat{\mu}_j, \hat{\sigma}_{1j}, \hat{\sigma}_{2j}, \hat{\nu})}; \quad i = 1, \dots, n, \quad j = 1, \dots, g, \quad (12)$$

$$\hat{s}_{ij1} = \frac{2 \left(\frac{\sigma_{1j}}{\hat{\sigma}_{1j} + \hat{\sigma}_{2j}} \right) f_{SMN}(y_i | \hat{\mu}_j, \hat{\sigma}_{1j}, \hat{\nu}) I_{(-\infty, \hat{\mu}_j]}(y_i)}{g(y_i | \hat{\mu}_j, \hat{\sigma}_{1j}, \hat{\sigma}_{2j}, \hat{\nu})} = I_{(-\infty, \hat{\mu}_j]}(y_i), \quad (13)$$

where $g(\cdot | \cdot)$ is the TP-SMN PDF defined in Equation (4), and $\hat{s}_{ij2} = 1 - \hat{s}_{ij1}$, and the conditional expectation $\hat{\kappa}_{ij}$ for the TP-SMN distribution members are given by:

- Two-piece normal (TP-N): $\hat{\kappa}_{ijk} = 1$,
- Two-piece t (TP-T): $\hat{\kappa}_{ijk} = \frac{\hat{\nu} + 1}{\hat{\nu} + d_{ijk}}$,
- Two-piece slash (TP-SL): $\hat{\kappa}_{ij} = \frac{2\hat{\nu} + 1}{d_{ijk}} \frac{P_1(\hat{\nu} + 3/2, d_{ijk}/2)}{P_1(\hat{\nu} + 1/2, d_{ijk}/2)}$,
- Two-piece contaminated Normal (TP-CN): $\hat{\kappa}_{ij} = \frac{\hat{\tau}^{3/2} \hat{\nu} e^{-\hat{\tau} d_{ijk}/2} + (1 - \hat{\nu}) e^{-d_{ijk}/2}}{\hat{\tau}^{1/2} \hat{\nu} e^{-\hat{\tau} d_{ijk}/2} + (1 - \hat{\nu}) e^{-d_{ijk}/2}}$,

where $d_{ijk} = \left(\frac{y_i - \hat{\mu}_j}{\hat{\sigma}_{kj}} \right)^2$ and $P_x(a, b)$ denotes the distribution function of the *Gamma*(a, b) distribution evaluated at x .

Now, the expectation step (E-step) at the $(r + 1)$ th iteration of the ECME-algorithm requires the calculation of $Q(\boldsymbol{\Theta} | \hat{\boldsymbol{\Theta}}^r) = E_{\boldsymbol{\Theta}}[\ell_c(\boldsymbol{\Theta}) | \hat{\boldsymbol{\Theta}}^r, \mathbf{y}]$. So,

E-step.

$$Q(\boldsymbol{\Theta} | \hat{\boldsymbol{\Theta}}^r) = -n \sum_{j=1}^g \log(\sigma_{1j} + \sigma_{2j}) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^g \sum_{k=1}^2 \frac{\hat{w}_{ijk}}{\sigma_{kj}^2} (y_i - \mu_j)^2 + \sum_{i=1}^n \sum_{j=1}^g \sum_{k=1}^2 E[Z_{ij} S_{ijk} \log h(U_{ijk} | \nu) | \hat{\boldsymbol{\Theta}}^r, y_i].$$

For the conditionally maximizing steps (CM-steps) at the $(r + 1)$ -th iteration of the ECME-algorithm we have:

CM-steps.

Update $\pi_j, j = 1, \dots, g$, as:

$$\pi_j^{(r+1)} = \frac{\sum_{i=1}^n \hat{z}_{ij}^{(r)}}{n}. \quad (14)$$

Update $\mu_j, j = 1, \dots, g$, as:

$$\mu_j^{(r+1)} = \frac{\sum_{i=1}^n \hat{\alpha}_{ij}^{(r)} y_i}{\hat{\alpha}_{ij}^{(r)}}, \quad (15)$$

where $\hat{\alpha}_{ij}^{(r)} = \sum_{k=1}^2 \hat{w}_{ijk}^{(r)} / \sigma_{kj}^{2(r)}$.

Update $\hat{\sigma}_{kj}^{(k+1)}; k = 1, 2; j = 1, \dots, g$, by solving the following stressed cubic equations

$$\sigma_{kj}^3 + p\sigma_{kj} + q = 0; \quad k = 1, 2, \quad (16)$$

where $p = -\frac{\sum_{i=1}^n \hat{w}_{ijk}^{(r)} (Y_i - \mu_j^{(r+1)})^2}{\hat{z}_{ij}^{(r)}}$, for which $q = p\sigma_{2j}I_{(k=1)} + p\sigma_{1j}I_{(k=2)}$. Note that $p, q < 0$, so this cubic equation has unique just root in the $(0, +\infty)$ interval.

CML-step of the ECME-algorithm.

$$\mathbf{v}^{(k+1)} = \operatorname{argmax}_{\mathbf{v}} \ell(\hat{\Theta}_{-\mathbf{v}}^{(r+1)}, \mathbf{v} | \mathbf{Y}), \quad (17)$$

where $\ell(\cdot | \mathbf{Y})$ is the log-likelihood function given in (9) and $\hat{\Theta}_{-\mathbf{v}}^{(r+1)}$ denotes the $(r + 1)$ -th update of $\hat{\Theta}$ except $\mathbf{v}$.

The ECME-algorithm iterates until a sufficient convergence rule is satisfied, e.g., if $|\ell(\hat{\Theta}^{(r+1)} | \mathbf{y}) / \ell(\hat{\Theta}^{(r)} | \mathbf{y}) - 1| \leq \epsilon$, under the determined tolerance ϵ .

4. Numerical Studies

In this section, we assess the performance of the proposed FM model using simulated and real datasets. The implementations of the algorithms were based on the R software [30] version 3.5.1 with a core i7 760 processor 2.8 GHz, and a relative tolerance of 10^{-5} was used for convergence of the ECME-algorithms. A sample copy of the R code is available up on request from the authors and will be available in an R package specialize to this proposed model.

4.1. Simulations

In this section, we have three simulations. In the first, we showed the robustness of the FM-TP-SMN models to classify heterogeneous data; in the second, we showed the misspecification of the proposed FM-TP-SMN models; and in the third simulation we considered suitability of the asymptotic properties for proposed model estimates.

Table 1. Mean of true allocations rates for fitted finite mixture two-piece distributions based on the scale mixtures of normal (FM-TP-SMN) models.

True Model	Sample Size	Fitted Model				
		FM-Normal	FM-TP-N	FM-TP-T	FM-TP-CN	FM-TP-SL
FM-TP-T	100	0.3745	0.7026	0.7674	0.8026	0.7902
	350	0.2836	0.7693	0.8372	0.8462	0.8450
	800	0.2231	0.7990	0.8418	0.8490	0.8469
FM-TP-CN	100	0.6023	0.6512	0.7835	0.7941	0.7829
	350	0.6258	0.7654	0.8542	0.8622	0.8510
	800	0.6395	0.7810	0.8599	0.8665	0.8563
FM-TP-SL	100	0.5374	0.7735	0.7810	0.7747	0.7840
	350	0.5903	0.8235	0.8420	0.8250	0.8491
	800	0.6034	0.8264	0.8437	0.8281	0.8507

4.1.1. Clustering

The FM models are useful for clustering the observations by allocating them into groups of observations that are similar in some sense. In fact, by considering the estimated (posterior) probabilities, we can assign such observation points to given groups. However, some atypical data have an undesirable effect to suitable clustering (see e.g., [1,2,8]). In our models, we consider the skewness and use the clustering as a base on them to show the robustness on the clustering of atypical data in components. We generated 1000 samples from the FM-TP-SMN with two components and for each sample, and considered the k-means clustering while we have ignored the true classification on these classifications.

We simulated 1000 samples with sample sizes $n = 100, 350, 800$, from the FM-TP-SMN models with parameters $\pi_1 = 0.75$, $\mu_1 = 10$, $\mu_2 = 15$, $\sigma_1 = 2$, $\sigma_2 = 4$, $\gamma_1 = 0.1$, $\gamma_2 = 0.7$ and $\gamma_1 = \gamma_2 = 0.5$ (FM-Normal model), for which $\nu = 4$ for TP-T and TP-SL, and $\nu = (0.3, 0.3)$ for TP-CN. According to the FM-TP-SMN estimated (posterior) probabilities given in (12) and the threshold value 0.5, we allocated the observations to some specific component. For each sample $t = 1, \dots, 1000$, the mean value rate of the correct allocations are given in Table 1, which shows that clustering based on the FM-TP-T, FM-TP-SL and FM-TP-CN are more reasonability than the ordinary FM-Normal model clustering, in the presence of atypical data. Also note that in the case of the true model (FM-TP-T), the FM-TP-CN also outperforms the other models.

4.1.2. Misspecification

For this section, we simulated 2000 samples with lengths $n = 150$ from FM-SN (asymmetric and light tailed components) and FM-ST (asymmetric and heavy tailed components) separately, with parameters of the previous simulation structure and with $(\lambda_1, \lambda_2) = (-2, 3)$. Then, we fitted various proposed FM-TP-SMN models to these data. In Table 2, various FM-TP-SMN models were first compared with the ordinary FM-NOR model (symmetric and light tailed components) and then various competitors within the FM-TP-SMN models (asymmetric components). The results in the first of four rows of Table 2 demonstrate that the number of preferred models belongs to the class of FM-TP-SMN models against the FM-NOR model. Also, the number of preferred models to fit the FM-SN is FM-TP-N, and in this case other preferred models except the FM-TP-N model are models which have similarities with it (for example FM-TP-T with large values of degree of freedom ν), i.e., preferred fitted models to the FM-SN with asymmetric and light tailed components are the FM-TP-SMN models with light tailed components. In the cases of FM-ST with asymmetric but heavy tailed components, also the FM-TP-SMN models with heavy tailed components were preferred. In this and the real application parts, the model selection criteria to choose the best model are: logarithm of

the maximized likelihood function (log-like) which is $\ell(\hat{\Theta}|\mathbf{y})$, Akaike information criteria (AIC); [31], Bayesian information criteria (BIC); [32], in the form of

$$AIC = 2k - 2\ell(\hat{\Theta}|\mathbf{y}) \quad \text{and} \quad BIC = k \log n - 2\ell(\hat{\Theta}|\mathbf{y}),$$

respectively, where k is the number of the model parameters.

Table 2. The number of times (out of 2000) the true FM models chosen under seven proposed hypotheses.

Condition Examined	True Model:		FM-SN		FM-ST	
	Criteria:		AIC	BIC	AIC	BIC
FM-TP-N vs. FM-Normal			1934	1958	1986	1975
FM-TP-T vs. FM-Normal			1511	1645	2000	2000
FM-TP-SL vs. FM-Normal			1387	1410	1823	1865
FM-TP-CN vs. FM-Normal			1503	1532	1746	1732
FM-TP-T vs. FM-TP-N			83	92	1861	1883
FM-TP-SL vs. FM-TP-N			112	123	1821	1793
FM-TP-CN vs. FM-TP-N			143	196	1732	1746

4.1.3. Asymptotical results

For this section, we simulated 400 samples each one with sample sizes $n = 150, 600, 1000, 2000, 4000$, from some FM-TP-T models with two components which are weak separated (WS), medium separated (MS) and strong separated (SS) of components, i.e., little, medium and large overlap of components respectively (see Figure 1), for which $\pi_1 = 0.4$, $\mu_1 = 10$, $\mu_2 = 15$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\gamma_1 = 0.65$, $\gamma_2 = 0.35$, $\nu = 4$ and $(\mu_1, \mu_2) = (-0.2, 0.2)$ for WS data; $(\mu_1, \mu_2) = (-1.5, 1.5)$ for MS data; and $(\mu_1, \mu_2) = (-5, 5)$ for SS data.

Figure 1. An artificial simulated finite mixture two-piece distributions based on the scale mixtures of normal (FM-TP-SMN) data of length $n = 400$ with two components: weakly separated components (WS); medium separated components (MS) and strongly separated components (SS), with curved probability density function (PDF) that datasets extracted from them.

Using the proposed ECME algorithm to find the ML estimates we focus on the evaluation of Monte-Carlo average of biasness (MC-bias) and mean squared error (MSE) defined as of the ML estimates in each j -th sample, $j = 1, \dots, 400$, respectively given in Tables 3–5 by

$$MC_{Bias}(\xi) = \left(\frac{1}{400} \sum_{j=1}^{400} \hat{\xi}_j^{(i)} \right) - \xi_j, \quad MSE(\xi) = \frac{1}{400} \sum_{j=1}^{400} (\xi_j - \hat{\xi}_j^{(i)})^2,$$

where $\hat{\xi}_j^{(i)}$ is the ML estimate of the parameter ξ_j in the i -th sample.

These results in Tables 3–5 are obtained from the different fitted FM-TP-SMN models and show the performance of the proposed models as well as their parameters estimates. As the sample size increased we naturally observed that the Monte-Carlo average bias of ML estimates and MSE were tending toward zero.

Table 3. Monte-Carlo average bias (MC-bias) and mean squared error (MSE) for maximum likelihood (ML) estimates in the weak separated components (WS) FM-TP-true (T) model.

Measure	Parameter	Sample Size				
		$n = 150$	$n = 600$	$n = 1000$	$n = 2000$	$n = 4000$
MC-Bias	π_1	-1.64533×10^{-2}	2.50755×10^{-3}	-1.849×10^{-5}	8.427×10^{-6}	3.452×10^{-7}
	μ_1	3.48475×10^{-2}	-9.48375×10^{-4}	-2.287×10^{-3}	7.427×10^{-4}	6.584×10^{-4}
	μ_2	3.93636×10^{-1}	-1.98365×10^{-2}	-3.276×10^{-3}	9.4978×10^{-4}	-6.436×10^{-4}
	σ_1	-4.98656×10^{-1}	3.44827×10^{-2}	6.487×10^{-3}	-6.775×10^{-3}	-7.864×10^{-3}
	σ_2	4.57463×10^{-1}	4.03741×10^{-2}	-6.284×10^{-3}	-6.103×10^{-3}	9.903×10^{-4}
	γ_1	1.57363×10^{-2}	1.93845×10^{-3}	1.574×10^{-6}	1.427×10^{-6}	5.047×10^{-7}
	γ_2	1.40384×10^{-2}	1.73644×10^{-3}	2.037×10^{-6}	6.948×10^{-7}	4.765×10^{-7}
	ν	1.14024×10^2	1.36253×10^{-1}	-6.017×10^{-2}	1.284×10^{-2}	-1.201×10^{-2}
MSE	π_1	7.15248×10^{-3}	7.35733×10^{-4}	3.854×10^{-4}	6.036×10^{-5}	6.729×10^{-5}
	μ_1	1.10375×10^{-1}	8.99164×10^{-2}	3.927×10^{-3}	2.889×10^{-4}	2.960×10^{-4}
	μ_2	1.97364×10^0	1.23473×10^{-1}	4.920×10^{-2}	3.276×10^{-3}	3.328×10^{-3}
	σ_1	1.69475×10^0	1.11763×10^0	2.118×10^{-1}	3.2849×10^{-3}	3.453×10^{-3}
	σ_2	1.67568×10^1	1.43855×10^1	4.548×10^{-1}	6.786×10^{-2}	6.903×10^{-2}
	γ_1	8.00264×10^{-3}	3.49566×10^{-4}	2.801×10^{-4}	6.104×10^{-5}	6.003×10^{-5}
	γ_2	1.98374×10^{-2}	8.20183×10^{-3}	2.102×10^{-4}	6.352×10^{-5}	6.102×10^{-5}
	ν	0.97803×10^2	5.46093×10^{-1}	1.112×10^{-1}	1.684×10^{-2}	1.521×10^{-2}

Table 4. MC-Bias and MSE for ML estimates in the medium separated components (MS) FM-TP-T model.

Measure	Parameter	Sample Size				
		$n = 150$	$n = 600$	$n = 1000$	$n = 2000$	$n = 4000$
MC-Bias	π_1	5.03746×10^{-3}	-4.21374×10^{-3}	-1.40712×10^{-3}	-5.83744×10^{-4}	6.10927×10^{-5}
	μ_1	2.18723×10^{-2}	-1.72451×10^{-3}	-6.28474×10^{-4}	2.99837×10^{-5}	-2.48576×10^{-5}
	μ_2	-1.30284×10^{-2}	8.29374×10^{-4}	6.99386×10^{-4}	3.95645×10^{-5}	-3.43927×10^{-5}
	σ_1	1.25344×10^{-2}	1.28374×10^{-4}	-5.98375×10^{-5}	-4.99380×10^{-5}	-5.99837×10^{-6}
	σ_2	1.27364×10^{-2}	-2.04634×10^{-3}	-1.47364×10^{-3}	2.98476×10^{-4}	6.97484×10^{-5}
	γ_1	-1.10264×10^{-2}	-1.02172×10^{-3}	-4.93846×10^{-5}	-6.10283×10^{-7}	7.98375×10^{-7}
	γ_2	-0.78725×10^{-2}	-0.90273×10^{-3}	5.97367×10^{-5}	5.83753×10^{-7}	9.01274×10^{-7}
	ν	2.40926×10^0	1.12027×10^{-1}	3.98365×10^{-2}	-2.48462×10^{-3}	-2.83765×10^{-4}
MSE	π_1	5.85644×10^{-3}	6.45364×10^{-4}	3.57464×10^{-5}	6.20183×10^{-5}	7.10993×10^{-5}
	μ_1	1.08744×10^{-1}	9.37436×10^{-1}	1.03764×10^{-2}	2.57463×10^{-4}	2.03464×10^{-4}
	μ_2	2.03937×10^{-1}	2.00713×10^{-1}	4.67483×10^{-2}	3.47367×10^{-3}	3.24536×10^{-3}
	σ_1	3.46357×10^{-1}	2.47464×10^{-1}	2.23433×10^{-1}	3.39283×10^{-3}	6.58475×10^{-4}
	σ_2	8.38474×10^{-1}	6.37364×10^{-1}	4.38475×10^{-1}	6.87364×10^{-2}	6.00836×10^{-2}
	γ_1	1.84746×10^{-2}	4.03847×10^{-4}	3.11972×10^{-4}	7.00374×10^{-5}	6.21002×10^{-5}
	γ_2	2.04464×10^{-2}	7.64533×10^{-3}	2.30283×10^{-4}	5.89472×10^{-5}	7.00353×10^{-5}
	ν	6.47465×10^1	5.87957×10^{-1}	1.93845×10^{-2}	1.74534×10^{-2}	1.48375×10^{-2}

Table 5. MC-Bias and MSE for ML estimates estimates in the strong separated components (SS) FM-TP-T model.

Measure	Parameter	Sample Size				
		$n = 150$	$n = 600$	$n = 1000$	$n = 2000$	$n = 4000$
MC-Bias	π_1	5.13426×10^{-3}	5.60483×10^{-3}	-1.40712×10^{-3}	-6.26826×10^{-4}	-6.95662×10^{-5}
	μ_1	-1.90273×10^{-2}	-1.24751×10^{-3}	1.20183×10^{-3}	3.45342×10^{-5}	2.74634×10^{-5}
	μ_2	1.15379×10^{-2}	-3.06344×10^{-3}	1.47333×10^{-3}	4.28144×10^{-5}	3.35242×10^{-5}
	σ_1	-1.03046×10^{-2}	-8.34452×10^{-5}	6.73645×10^{-5}	-4.87354×10^{-5}	-7.28374×10^{-6}
	σ_2	-1.38724×10^{-2}	-1.27844×10^{-3}	-1.00904×10^{-3}	3.49383×10^{-4}	9.88365×10^{-5}
	γ_1	-0.58746×10^{-2}	0.34452×10^{-3}	5.09847×10^{-5}	-5.46353×10^{-7}	8.73645×10^{-7}
	γ_2	-0.60273×10^{-2}	-0.24533×10^{-3}	6.08422×10^{-5}	5.27363×10^{-7}	8.69384×10^{-7}
	ν	2.37264×10^0	0.99775×10^{-1}	4.65635×10^{-2}	-1.74632×10^{-2}	-2.46354×10^{-3}
MSE	π_1	2.94256×10^{-3}	3.83748×10^{-4}	2.67464×10^{-4}	7.78374×10^{-5}	9.73646×10^{-5}
	μ_1	5.01324×10^{-3}	4.11293×10^{-4}	1.69837×10^{-4}	8.32847×10^{-6}	2.37462×10^{-6}
	μ_2	5.73648×10^{-3}	5.03744×10^{-4}	8.98474×10^{-5}	6.26353×10^{-6}	3.03733×10^{-6}
	σ_1	3.48373×10^{-2}	1.20385×10^{-3}	1.65464×10^{-4}	1.92736×10^{-6}	1.82037×10^{-6}
	σ_2	7.92746×10^{-2}	5.38474×10^{-3}	1.16354×10^{-3}	1.98274×10^{-5}	1.48263×10^{-5}
	γ_1	3.26354×10^{-3}	4.24846×10^{-4}	2.63846×10^{-4}	6.03947×10^{-5}	8.48375×10^{-5}
	γ_2	2.73145×10^{-3}	3.38475×10^{-4}	9.83746×10^{-5}	5.64544×10^{-5}	8.66555×10^{-5}
	ν	1.28736×10^2	3.01763×10^{-1}	9.02635×10^{-2}	7.27647×10^{-3}	8.83645×10^{-3}

4.2. Applications

In this section, we apply the FM-TP-SMN models on some various real data sets to show the performance of the proposed models and estimates in applications.

BMI Data

We considered the body mass index (BMI) data set collected for men aged between 18 and 80 years. The BMI data set was gathered with the National Health and Nutrition Examination Survey in the US National Center for Health Statistics (NCHS) of the Center for Disease Control (CDC). A strong relationship between the obesity problem and many chronic diseases has attracted attention in recent years, that is, most people with an obesity problem will have chronic diseases. The ratio of body weight in kilograms and height in squared meters (BMI) is a measure to determine the rate of relationship between overweight and obesity. In this way, a person with BMI > 25 is considered overweighted, while BMI > 30 is considered obese.

This dataset had 4579 participants with BMI records, but for modeling with finite mixture models, participants with weights within 39.50–70.00 kg and 95.01–196.80 kg with 1069 and 1054 participants were considered in the first and second subgroups respectively. Lin et al. [7] were first analyzed this dataset by considering the reports in 1999–2000 and 2001–2002, and were fitted the FM-normal, FM-T, FM-SN and FM-ST, always with two components, and then [5,13] fitted the FM-SMSN models to this dataset. The results, obtained by [13], were general and involved the results by [5,7]. So we fitted the proposed FM-TP-SMN models to this dataset and compared obtained results in the [13].

Table 6 contains the ML estimates of the FM-TP-SMN models with two components, and the Log-likelihood, AIC and BIC criterions of the proposed FM-TP-SMN models and FM-SMSN taken from Table 1 due to [13] appear in Table 7.

Table 6. ML estimation results for fitting FM-TP-SMN models to the body mass index (BMI) data.

Parameter	Fitted Model				
	FM-Normal	FM-TP-N	FM-TP-T	FM-TP-SL	FM-TP-CN
π_1	0.391	0.5400	0.4600	0.5374	0.5282
μ_1	21.412	20.7520	20.744	20.7982	21.0629
μ_2	32.548	30.1324	30.155	30.0169	29.9851
σ_1	2.0176	5.0343	5.0303	4.2785	1.5215
σ_2	6.4180	7.4040	7.4038	6.3213	2.4601
γ_1	–	0.6798	0.6806	0.6709	0.6302
γ_2	–	0.8695	0.8673	0.8771	0.8785
ν	–	–	7.9939	2.1950	(0.971, 0.081)

Table 7. Model selection criteria for fitting FM-TP-SMN and FM-scale mixtures of the skew normal (SMSN) models to the BMI data. The best values are marked in bold.

Criterion	Fitted Model								
	FM-Normal	FM-TP-N	FM-SN	FM-TP-T	FM-ST	FM-TP-SL	FM-SSL	FM-TP-CN	FM-SCN
Log-likelihood	−6911.76	−6870.30	−6979.47	−6856.65	−6869.03	−6857.14	−6867.99	−6871.65	−6865.25
AIC	13833.35	13754.72	13972.95	13729.41	13754.06	13730.41	13751.98	13812.31	13748.51
BIC	13961.61	13763.79	13821.22	13737.61	13782.33	13735.85	13780.24	13775.89	13776.77

As noted by Lin et al. [4] and Prates et al. [13], the criteria values in Table 7 indicate that the heavy tailed FM-SMSN models (FM-ST, FM-SCN and FM-SSL) had a better fit than the ordinary FM-NOR and FM-SN models, and also the FM-SSL and FM-ST were the best fitted models. Such results are for the FM-TP-SMN (with corresponding FM-SMSN counterparts) models, while the FM-TP-SMN models were more reasonable than FM-TP-SMN models. However, the FM-TP-SL and FM-TP-T were the best models. In Figure 2, we plot the fitted FM-TP-T and FM-ST densities curved on the histogram of BMI data.

Figure 2. Histogram of body mass index (BMI) data with fitted FM-TP-true (T) (left) and FM-ST (right) models with two components.

4.3. UScrime Data

As a further application of the FM-TP-SMN models and proposed methodology, we consider the effect of punishment regimes on crime rates [33,34], which is of high interest to criminologists. This has been studied using aggregate data of 47 US states for 1960 given in this data frame, and we

consider the 13th column of this data frame which is due to income inequality. The data are available under the UScrime function in the MASS R package.

Table 8 contains the ML estimates of the FM-TP-SMN models with two components, and the Log-likelihood, AIC and BIC criteria of the proposed FM-TP-SMN and FM-SMSN models.

The log-likelihood values in Table 8 indicate that the FM-TP-N and FM-TP-T are the best models within the FM-TP-SMN models, while the FM-SCN is the best model in the class of the FM-SMSN models. The AIC and BIC criteria chose the FM-TP-N model (asymmetric components) within the FM-TP-SMN models, while they chose the FM-NOR model (symmetric components) in the class of the FM-SMSN models, which has symmetrical components. Among all competitors, the criteria chose the FM-TP-N model belonging to the proposed FM-TP-SMN models, which is more reasonable. In Figure 3 is plotted the histogram of US crime data with the curves of FM-TP-SMN and FM-SMSN models. These graphical visualizations show the suitability of the asymmetrical components and proposed FM-TP-SMN models.

Table 8. ML estimation results and model selection criteria for fitting FM-TP-SMN models to the UScrime data. The best values are marked in bold.

Criterion	Fitted Model								
	FM-Normal	FM-TP-N	FM-SN	FM-TP-T	FM-ST	FM-TP-SL	FM-SSL	FM-TP-CN	FM-SCN
π_1	0.615	0.463	0.607	0.462	0.607	0.463	0.607	0.509	0.582
μ_1	166.674	173.927	175.448	173.924	174.558	173.931	175.350	170.931	175.390
μ_2	237.632	246.160	227.951	246.060	228.018	245.998	227.888	204.645	224.584
σ_1	18.283	20.399	20.170	20.263	19.597	20.294	19.835	18.626	16.362
σ_2	20.455	59.464	22.778	59.268	22.634	58.813	22.478	42.728	20.201
$\gamma_1(\lambda_1)$	—	0.071	−0.687	0.071	−0.613	0.071	−0.679	0.176	−0.872
$\gamma_2(\lambda_2)$	—	0.257	0.555	0.258	0.548	0.259	0.560	0.782	0.579
ν	—	—	—	39.934	100	39.726	37.590	0.745	0.672
τ	—	—	—	—	—	—	—	0.999	0.990
Log-likelihood	−232.226	−228.215	−232.274	−228.865	−231.275	−228.211	−231.265	−229.864	−230.281
AIC	474.452	470.437	478.548	473.734	478.549	472.426	478.531	477.748	478.562
BIC	483.703	483.386	491.499	488.535	491.500	487.228	491.482	494.399	491.513

Figure 3. Histogram of UScrime data with fitted FM-TP-SMN and FM-SMSN models with two components.

5. Conclusions

We have proposed a flexible family of TP-SMN distributions for application in clustering problems. The TP-SMN family is capable of representing distributions of symmetric/asymmetric and light/heavy tailed forms, which contains the well-known symmetric SMN family and is a reasonable competitor for asymmetric SMSN family as a special case. Estimation of the finite mixtures of the FM-TP-SMN parameters is relatively straightforward via the ECME algorithm with a fast convergence (a few iterations loop). Considering a Bayesian approach and the flexible TP-SMN family in the Autoregressive and ARMA processes from [35,36] can be further topics of research.

Author Contributions: M.M., J.E.C.-R. and M.R.M. wrote the paper and contributed the reagents/analysis/materials tools; M.M. conceived, designed, and performed the experiments, and analyzed the data. All authors read and approved the final manuscript. All authors have read and approved the final manuscript.

Acknowledgments: The authors thank the editor and three anonymous referees for their helpful comments and suggestions.

Conflicts of Interest: The authors declare that there is no conflict of interest in the publication of this paper.

References

1. McLachlan, G.; Peel, D. *Finite Mixture Models*; John Wiley and Sons: New York, NY, USA, 2000.
2. Contreras-Reyes, J.E.; Cortés, D.D. Bounds on Rényi and Shannon Entropies for Finite Mixtures of Multivariate Skew-Normal Distributions: Application to Swordfish (*Xiphias gladius Linnaeus*). *Entropy* **2016**, *18*, 382. [[CrossRef](#)]
3. Frühwirth-Schnatter, S.; Pyne, S. Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics* **2010**, *11*, 317–336. [[CrossRef](#)] [[PubMed](#)]
4. Lin, T.I.; Lee, J.C.; Yen, S.Y. Finite mixture modelling using the skew normal distribution. *Stat. Sin.* **2007**, *17*, 909–927.
5. Basso, R.M.; Lachos, V.H.; Cabral, C.R.B.; Ghosh, P. Robust mixture modeling based on scale mixtures of skew-normal distributions. *Comput. Stat. Data Anal.* **2010**, *54*, 2926–2941. [[CrossRef](#)]
6. Lin, T.I. Robust mixture modeling using multivariate skew t distributions. *Stat. Comput.* **2009**, *20*, 343–356. [[CrossRef](#)]
7. Lin, T.I.; Lee, J.C.; Hsieh, W.J. Robust Mixture Modelling Using the Skew t Distribution. *Stat. Comput.* **2007**, *17*, 81–92. [[CrossRef](#)]
8. Contreras-Reyes, J.E.; López Quintero, F.O.; Yáñez, A.A. Towards Age Determination of Southern King Crab (*Lithodes santolla*) Off Southern Chile Using Flexible Mixture Modeling. *J. Mar. Sci. Eng.* **2018**, *6*, 157. [[CrossRef](#)]
9. Maleki, M.; Arellano-Valle, R.B. Maximum a-posteriori estimation of autoregressive processes based on finite mixtures of scale-mixtures of skew-normal distributions. *J. Stat. Comput. Sim.* **2017**, *87*, 1061–1083. [[CrossRef](#)]
10. Maleki, M.; Arellano-Valle, R.B.; Dey, D.K.; Mahmoudi, M.R.; Jalali, S.M.J. A Bayesian Approach to Robust Skewed Autoregressive Processes. *Calcutta Stat. Assoc. Bull.* **2017**, *69*, 165–182. [[CrossRef](#)]
11. Maleki, M.; Wraith, D.; Arellano-Valle, R.B. Robust finite mixture modeling of multivariate unrestricted skew-normal generalized hyperbolic distributions. *Stat. Comput.* **2018**, in press. [[CrossRef](#)]
12. Maleki, M.; Wraith, D.; Arellano-Valle, R.B. A flexible class of parametric distributions for Bayesian linear mixed models. *Test* **2018**, in press. [[CrossRef](#)]
13. Prates, M.O.; Lachos, V.H.; Cabral, C. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *J. Stat. Soft.* **2000**, *54*, 1–20. [[CrossRef](#)]
14. Hajrajab, A.; Maleki, M. Nonlinear semiparametric autoregressive model with finite mixtures of scale mixtures of skew normal innovations. *J. Appl. Stat.* **2019**, in press. [[CrossRef](#)]
15. Maleki, M.; Wraith, D. Mixtures of multivariate restricted skew-normal factor analyzer models in a Bayesian framework. *Comput. Stat.* **2019**, in press. [[CrossRef](#)]
16. Andrews, D.R.; Mallows, C.L. Scale mixture of normal distribution. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 99–102. [[CrossRef](#)]

17. Lange, K.L.; Little, R.; Taylor, J. Robust statistical modeling using t distribution. *J. Am. Stat. Assoc.* **1989**, *84*, 881–896. [CrossRef]
18. Lange, K.L.; Sinsheimer, J.S. Normal/independent distributions and their applications in robust regression. *J. Comput. Graph. Stat.* **1993**, *2*, 175–198.
19. Maleki, M.; Nematollahi, A.R. Autoregressive Models with Mixture of Scale Mixtures of Gaussian innovations. *Iranian J. Sci. Technol. Trans. A* **2017**, *41*, 1099–1107. [CrossRef]
20. Arellano-Valle, R.B.; Gómez, H.; Quintana, F.A. Statistical inference for a general class of asymmetric distributions. *J. Stat. Plan. Inf.* **2005**, *128*, 427–443. [CrossRef]
21. Hoseinzadeh, A.; Maleki, M.; Khodadadi, Z.; Contreras-Reyes, J.E. The Skew-Reflected-Gompertz distribution for analyzing symmetric and asymmetric data. *J. Comput. Appl. Math.* **2019**, *349*, 132–141. [CrossRef]
22. Moravveji, B.; Khodadai, Z.; Maleki, M. A Bayesian Analysis of Two-Piece distributions based on the Scale Mixtures of Normal Family. *Iranian J. Sci. Technol. Trans. A* **2018**, in press. [CrossRef]
23. Maleki, M.; Mahmoudi, M.R. Two-Piece Location-Scale Distributions based on Scale Mixtures of Normal family. *Commun. Stat. Theor. Meth.* **2017**, *46*, 12356–12369. [CrossRef]
24. Rubio, F.J.; Steel, M.F.G. Inference in Two-Piece Location-Scale Models with Jeffreys Priors. *Bayesian Anal.* **2014**, *9*, 1–22. [CrossRef]
25. Branco, M.D.; Dey, D.K. A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113. [CrossRef]
26. Mudholkar, G.S.; Hutson, A.D. The epsilon-skew-normal distribution for analyzing near-normal data. *J. Stat. Plan. Inf.* **2000**, *83*, 291–309. [CrossRef]
27. Meng, X.; Rubin, D.B. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **2017**, *80*, 267–278. [CrossRef]
28. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–38. [CrossRef]
29. Liu, C.; Rubin, D.B. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* **1994**, *81*, 633–648. [CrossRef]
30. R Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018; ISBN 3-900051-07-0. Available online: <http://www.R-project.org> (accessed on 12 December 2018).
31. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
32. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
33. Ehrlich, I. Participation in illegitimate activities: A theoretical and empirical investigation. *J. Political Econ.* **1973**, *81*, 521–565. [CrossRef]
34. Vandaele, W. Participation in illegitimate activities: Ehrlich revisited. In *Deterrence and Incapacitation*; Blumstein, A., Cohen, J., Nagin, D., Eds.; US National Academy of Sciences: Washington, DC, USA, 1978; pp. 270–335.
35. Ghasami, S.; Khodadadi, Z.; Maleki, M. Autoregressive Processes with Generalized Hyperbolic Innovations. *Commun. Stat. Comput. Sim.* **2019**, in press. [CrossRef]
36. Zarrin, P.; Maleki, M.; Khodadadi, Z.; Arellano-Valle, R.B. Time series process based on the unrestricted skew normal process. *J. Stat. Comput. Sim.* **2019**, *89*, 38–51. [CrossRef]