Econometrics and Statistics
Volume 19, July 2021, Pages 151-168

EM algorithm using overparameterization for the multivariate skew-normal distribution

Toshihiro Abe [a], Hironori Fujisawa [b d], Takayuki Kawashima [c d], Christophe Ley [e]

https://doi.org/10.1016/j.ecosta.2021.03.003

Abstract

A stochastic representation with a latent variable often enables us to make an EM algorithm to obtain the maximum likelihood estimate. The skew-normal distribution has such a simple stochastic representation with a latent variable, and consequently one expects to have a convenient EM algorithm. However, even for the univariate skew-normal distribution, existing EM algorithms constructed using a stochastic representation require a solution of a complicated estimating equation for the skewness parameter, making it difficult to extend such an idea to the multivariate skew-normal distribution. A stochastic representation with overparameterization is proposed, which has not been discussed yet. The approach allows the construction of an efficient EM algorithm in a closed form, which can be extended to a mixture of multivariate skew-normal distributions. The proposed EM algorithm can be regarded as an accelerated version with momentum (which is known as an acceleration technique of the algorithm in optimization) of a recently proposed EM algorithm. The novel EM algorithm is applied to real data and compared with the command msn.mle from the **R** package sn.

Introduction

The most popular skew distribution on the real line is the skew-normal of Azzalini, 1985, Azzalini, 1986. Consequently its multivariate extensions have been investigated in the literature: Azzalini and Dalla Valle (1996) introduced the multivariate skew-normal distribution, emphasizing specifically on the bivariate case, while Azzalini and Capitanio (1999) further investigated properties of multivariate skew-normal distributions and briefly defined multivariate skew-elliptical distributions. A similar class of skew-elliptical distributions is due to Branco and Dey (2001). Genton and Loperfido (2005) extended the latter into the so-called generalized skew-elliptical distributions, and Azzalini and Capitanio (2003) and Wang et al. (2004) independently to multivariate skew-symmetric distributions. Besides its many nice stochastic properties, the skew-normal distribution is also (in)famous for inferential problems, such as a singular Fisher information matrix in the vicinity of symmetry (Hallin and Ley, 2014). For an extensive overview of the univariate and multivariate skew-normal distributions, extensions and properties, we refer the reader to the review paper Azzalini (2005), the monograph Azzalini and Capitanio (2014) and Section3.1 of Ley (2015) (where this type of distributions is considered in the general context of asymmetric distributions).

In this paper, our goal is to improve maximum likelihood estimation for the multivariate skew-normal distribution by devising a more efficient EM algorithm (Dempster et al., 1977). Several stochastic representations have been proposed for multivariate skew-normal distributions. A stochastic representation with a latent variable enables a convenient design of an EM algorithm (McLachlan and Krishnan, 2007), however the M-step does not always produce simple expressions. For instance, Lin et al. (2007) constructed an EM algorithm for a mixture of univariate skew-normal distributions, but the EM algorithm obtained in their work demands a solution of a complicated estimating equation for the skewness parameter. The same problem occurs even in a usual univariate skew-normal distribution and, therefore, it is difficult to extend the idea to the multivariate case. Lin (2009) considered an EM algorithm for a mixture of multivariate skew-normal distributions, but their multivariate skew-normal distribution was based on Sahu et al. (2003) and was thus different from the “original one” by Azzalini and Capitanio (1999). The EM algorithm in Lin (2009) demands multi-dimensional numerical integrations, which are difficult to compute in a high-dimensional case. Ferreira et al. (2016) addressed the inference problem for multivariate skew scale mixtures of normal (SSMN) distributions, but the resulting algorithm demands a numerical optimization for the hyperparameter. More recently, Maleki et al. (2020) provided the Expectation/Conditional Maximization Either (ECME) algorithm (an extension of the EM-algorithm) for multivariate scale mixture of skew-normal (SMSN) distributions in vector autoregressive (VAR) models, but with the same numerical difficulties.

A recent more appealing EM algorithm has been designed by Chen et al. (2014). They devised an expression of the skew-normal probability density function which implied an EM algorithm in a closed form for the multivariate skew-normal distribution, although they considered regularized multivariate regression models with skew-t error distributions. In this paper we incorporate an overparameter into a conventional stochastic representation and then obtain the EM algorithm in a closed form. Our EM algorithm is similar to that in Chen et al. (2014), but our proposal can be regarded as an accelerated algorithm with momentum, which is known as an acceleration technique of the algorithm in optimization (Nesterov, 2004). In numerical studies, the proposed EM algorithm increases the likelihood more rapidly than that of Chen et al. (2014).

This paper is organized as follows. In Section2, we introduce a stochastic representation for the multivariate skew-normal distribution with overparameterzation, and provide some basic results. In Section3, we construct the related EM algorithm in a closed form, and in the final algorithm, we can eliminate the overparameter τ. In Section4, using the moment estimator, we propose an initial value for the EM algorithm. Next in Section5 we show that the proposed EM algorithm is an accelerated algorithm with momentum, compared to the EM algorithm of Chen et al. (2014). In Section6, we consider a mixture of multivariate skew-normal distributions, and in Section7 we present some real data analyses using AIS data. Numerical performances are evaluated by means of simulations in Section8. A conclusion is given in Section9 while proofs are provided in the Appendix.

Section snippets

Multivariate skew-normal distributions derived from a new stochastic representation with overparameterization

First, we review the stochastic representation of the multivariate skew-normal distribution as given by Azzalini and Capitanio (1999). Suppose $\begin{pmatrix} \boldsymbol{Y} \\ Y_0 \end{pmatrix} \sim N_{p+1}\left(\mathbf{0}, \Omega^*\right), \quad \Omega^* = \begin{pmatrix} \Omega & \boldsymbol{\delta}_0 \\ \boldsymbol{\delta}_0^\top & 1 \end{pmatrix}$, where $\boldsymbol{\delta}_0 \in \mathbb{R}^p$ and Ω is a $p \times p$ symmetric positive definite matrix. Then $\boldsymbol{U} = \operatorname{sgn}(Y_0)\boldsymbol{Y}$ has as density

$f(\boldsymbol{u}) = 2\Phi\left(\boldsymbol{\alpha}^\top \boldsymbol{u}\right)\phi_p\left(\boldsymbol{u}; \mathbf{0}, \Omega\right), \quad \boldsymbol{u} \in \mathbb{R}^p$,where $\boldsymbol{\alpha} = \Omega^{-1}\boldsymbol{\delta}_0 / \left(1 - \boldsymbol{\delta}_0^\top \Omega^{-1}\boldsymbol{\delta}_0\right)^{1/2}$, $\Phi(z)$ is the cumulative distribution function of the standard scalar normal distribution, and $\phi_p\left(\boldsymbol{u}; \mathbf{0}, \Omega\right)$ is the p-variate normal density function with mean vector $\mathbf{0}$ and...

EM algorithm for the multivariate skew-normal distribution

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be independent observations drawn from $\mathrm{SN}_p^*\left(\boldsymbol{\mu}, \Omega, \boldsymbol{\lambda}\right)$. Let $\boldsymbol{\theta}$ be the parameter vector that consists of $\boldsymbol{\mu}$, Ω, and $\boldsymbol{\delta}$ (or $\boldsymbol{\lambda}$). In this section, we consider the maximum likelihood estimation of $\boldsymbol{\theta}$ based on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, and provide an EM algorithm in a closed form.

Let $\left(\boldsymbol{y}_1^\top, y_{01}\right)^\top, \ldots, \left(\boldsymbol{y}_n^\top, y_{0n}\right)^\top$ be independent observations drawn from a multivariate normal distribution $N_{p+1}\left(\mathbf{0}, \Sigma\right)$. Let $\boldsymbol{x}_i = \boldsymbol{\mu} + \operatorname{sgn}\left(y_{0i}\right)\boldsymbol{y}_i$ for $i = 1, \ldots, n$. Then, it can be noted that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are the observations drawn from $\mathrm{SN}_p^*\left(\boldsymbol{\mu}, \Omega, \boldsymbol{\lambda}\right)$, as...

Initial value setting using moment estimator

The update algorithm requires an initial value, $\widehat{\boldsymbol{\theta}}^{(0)}$. In this section, we present a method for setting an initial value using a moment estimator.

Let the cumulant-generating function be denoted by $K(\boldsymbol{t}) = \log M(\boldsymbol{t})$. It follows from Lemma3 that

$$\frac{\partial K(\boldsymbol{t})}{\partial \boldsymbol{t}} = \boldsymbol{\mu} + \Omega\boldsymbol{t} + \frac{\phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)}{\Phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)}\boldsymbol{\gamma},$$

where $\boldsymbol{\gamma} = \Omega^{1/2}\boldsymbol{\delta} = \left(\gamma_1, \ldots, \gamma_p\right)^\top$, and

$$\frac{\partial^2 K(\boldsymbol{t})}{\partial \boldsymbol{t}\partial \boldsymbol{t}^\top} = \Omega + \frac{\phi'\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)\Phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right) - \phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)^2}{\Phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)^2}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top,$$

$$\mathrm{E}[\boldsymbol{X}] = \left.\frac{\partial K(\boldsymbol{t})}{\partial \boldsymbol{t}}\right|_{\boldsymbol{t}=\mathbf{0}} = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\gamma},$$

$$V[\boldsymbol{X}] = \left.\frac{\partial^2 K(\boldsymbol{t})}{\partial \boldsymbol{t}\partial \boldsymbol{t}^\top}\right|_{\boldsymbol{t}=\mathbf{0}} = \Omega - \frac{2}{\pi}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top.$$

We moreover get

$$\frac{\partial^3 K(\boldsymbol{t})}{\partial t_k^3} = \left(\frac{\phi''\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)}{\Phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)} - 3\frac{\phi'\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)\phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)}{\Phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)^2} + 2\frac{\phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)^3}{\Phi\left(\boldsymbol{\gamma}^\top\boldsymbol{t}\right)^3}\right)\gamma_k^3,$$

...

$$s_k = \mathrm{E}\left[\left(X_k - \mathrm{E}\left[X_k\right]\right)^3\right] = \frac{\partial^3 K(\boldsymbol{t})}{}$$

Comparison with the EM algorithm of Chen etal. (2014)

Typesetting math: 79%

The EM algorithm of Chen et al. (2014) can be expressed as follows:

$$\begin{aligned}\widehat{\boldsymbol{\mu}}^{(k+1)} &= \overline{\boldsymbol{x}} - c_{\boldsymbol{\lambda}^{(k)}} \left\{\widehat{\Omega}^{(k)}\right\}^{1/2} \hat{\boldsymbol{\delta}}^{(k)} \frac{1}{n} \sum_{i=1}^{n} \rho_1 \left(\hat{v}_i^{(k)}\right), \\ \widehat{\Omega}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}^{(k)}\right) \left(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}^{(k)}\right)^{\top}, \\ \widehat{\boldsymbol{\lambda}}^{(k+1)} &= \left\{\widehat{\Omega}^{(k)}\right\}^{-1/2} \widehat{\boldsymbol{\psi}}_n^{(k)} \left(\widehat{\boldsymbol{\mu}}^{(k)}\right),\end{aligned}$$

where $\widehat{\boldsymbol{\psi}}_n^{(k)} (\boldsymbol{\mu}) = \frac{1}{n} \sum_{i=1}^{n} \rho_1 \left(\hat{v}_i^{(k)}\right) (\boldsymbol{x}_i - \boldsymbol{\mu})$ .Our EM algorithm is given by (8), (9), and (12). To easily compare our algorithm with their algorithm, we provide the updated formula of $\boldsymbol{\lambda} = \boldsymbol{\delta}/\sqrt{1 - \boldsymbol{\delta}^{\top} \boldsymbol{\delta}}$ instead of $\boldsymbol{\delta}$ in (12):$\widehat{\boldsymbol{\lambda}}^{(k+1)} = a_n^{(k)} \left\{\Omega^{(k+1)}\right\}^{-1/2} \widehat{\boldsymbol{\psi}}_n^{(k)} \left(\widehat{\boldsymbol{\mu}}^{(k+1)}\right)$,

where $a_n^{(k)} = \frac{1}{\sqrt{1+\left\{\widehat{\boldsymbol{\lambda}}^{(k)}\right\}^{\top}\left\{\widehat{\Omega}^{(k)}\right\}^{-1/2} \widehat{\boldsymbol{\psi}}_n^{(k)}\left(\widehat{\boldsymbol{\mu}}^{(k)}\right)-\left\{\widehat{\boldsymbol{\psi}}_n^{(k)}\left(\widehat{\boldsymbol{\mu}}^{(k+1)}\right)\right\}^{\top}\left\{\right.}}$...

EM algorithm for a mixture of multivariate skew-normal distributions

Suppose that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are drawn from a mixture of multivariate skew-normal distributions with probability density function $f_{\mathrm{mix}} (\boldsymbol{x}) = \sum_{j=1}^{g} w_j f_{\mathrm{SN}} \left(\boldsymbol{x}; \boldsymbol{\xi}_j\right)$,where $w_j > 0, \sum_{j=1}^{g} w_j = 1, \boldsymbol{\xi}_j$ is a parameter vector that consists of $\boldsymbol{\mu}_j, \Omega_j, \boldsymbol{\lambda}_j$, and $f_{\mathrm{SN}} \left(\boldsymbol{x}; \boldsymbol{\xi}_j\right)$ is the probability density function of $\mathrm{SN}_p^* \left(\boldsymbol{\mu}_j, \Omega_j, \boldsymbol{\lambda}_j\right)$. In this section, we construct an EM algorithm for this distribution. The basic idea is based on a combination of Section3 and the EM algorithm for a normal mixture model.

Let Σ_j be the matrix Σ in which Ω, δ,...

Analysis of Australian Institute of Sport data

The EM algorithm formulated in this paper is applied to the Australian Institute of Sport (AIS) data (Cook and Weisberg, 1994) with $n = 202$ observations and $p = 6$ variables, namely BMI (body mass index), ssf (sum of skin folds), pcBfat (body fat percentage), LBM (lean body mass), Height and Weight. The EM algorithm is compared with the classical MLE command msn.mle in the R package sn, which is a standard package for skew-normal distributions. First, we adopt a univariate skew-normal distribution...

Simulation Study

We conducted a Monte Carlo simulation study with 5,000 simulations for sample sizes $n = 200, 500, 1000$ for 2-dimensional skew-normal distributions with $\mu_1 = \mu_2 = 0,\ \Omega_{11} = \Omega_{22} = 1,\ \Omega_{12} = 0.5$ and $(\lambda_1, \lambda_2)^{\top} = (0.5, 1)^{\top}, (1, 2)^{\top}$, and then we calculated the biases and mean squared errors (MSEs) of the estimates obtained via the proposed EM algorithm. The results are listed in Table4, which shows that, as expected, the biases and MSEs decrease with increasing n.

We also have investigated the performance of ML estimation ...

Conclusion

The multivariate skew-normal distribution of Azzalini and Capitanio (1999) was considered and a different stochastic representation with an overparameter was proposed. It led to an explicit expression of the EM

Typesetting math: 79%

orithm for a mixture of multivariate skew-normal distributions was also considered, and numerical experiments showed that the proposed EM algorithm was effective. Our approach further paves the way for a fast estimation of the shape and degrees of freedom parameters…

Declaration of Competing Interest

Toshihiro Abe was supported in part by JSPS KAKENHI [grant numbers 19K11869 and 19KK0287]; and Nanzan University Pache Research Subsidy I-A-2 for the 2020 academic year.

Hironori Fujisawa was supported in part by JSPS KAKENHI [grant number 17K00065].

Takayuki Kawashima was supported in part by JSPS KAKENHI [grant numbers 19K24340 and 19H04071].

Christophe Ley is supported by the FWO Krediet aan Navorsers grant with reference number 1510391N.…

Acknowledgements

Toshihiro Abe was supported in part by JSPS KAKENHI [grant numbers 19K11869 and 19KK0287]. Hironori Fujisawa was supported in part by JSPS KAKENHI [grant number 17K00065]. Takayuki Kawashima was supported in part by JSPS KAKENHI [grant numbers 19K24340 and 19H04071]. Christophe Ley is supported by the FWO Krediet aan Navorsers grant with reference number 1510391N.…

References (22)

M.D. Branco *et al.*
A general class of multivariate skew-elliptical distributions
Journal of Multivariate Analysis (2001)

L. Chen *et al.*
Regularized multivariate regression models with skew-t error distributions
Journal of Statistical Planning and Inference (2014)

T.I. Lin
Maximum likelihood estimation for multivariate skew normal mixture models
Journal of Multivariate Analysis (2009)

A. Azzalini
A class of distributions which includes the normal ones
Scandinavian Journal of Statistics (1985)

A. Azzalini
Further results on a class of distributions which includes the normal ones
Statistica (1986)

A. Azzalini
…istribution and related multivariate families

Scandinavian Journal of Statistics (2005)

A. Azzalini *et al.*
Statistical applications of the multivariate skew normal distribution
Journal of the Royal Statistical Society: Series B (Statistical Methodology) (1999)

A. Azzalini *et al.*
Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution
Journal of the Royal Statistical Society: Series B (Statistical Methodology) (2003)

A. Azzalini *et al.*
The Skew-Normal and Related Families. vol. 3
(2014)

A. Azzalini *et al.*
The multivariate skew-normal distribution
Biometrika (1996)

View more references

Cited by (4)

View full text