Communications for Statistical Applications and Methods
2020, Vol. 27, No. 2, 255–268

https://doi.org/10.29220/CSAM.2020.27.2.255
Print ISSN 2287-7843 / Online ISSN 2383-4757

# A fast approximate fitting for mixture of multivariate skew *t*-distribution via EM algorithm

Seung-Gu Kim[1,a]

[a]Department of Data and Information, Sangji University, Korea

## Abstract

A mixture of multivariate canonical fundamental skew $t$-distribution (CFUST) has been of interest in various fields. In particular, interest in the unsupervised learning society is noteworthy. However, fitting the model via EM algorithm suffers from significant processing time. The main cause is due to the calculation of many multivariate $t$-cdfs (cumulative distribution functions) in E-step. In this article, we provide an approximate, but fast calculation method for the in univariate fashion, which is the product of successively conditional univariate $t$-cdfs with Taylor's first order approximation. By replacing all multivariate $t$-cdfs in E-step with the proposed approximate versions, we obtain the admissible results of fitting the model, where it gives 85% reduction time for the 5 dimensional skewness case of the Australian Institution Sport data set. For this approach, discussions about rough properties, advantages and limits are also presented.

Keywords: mixture model, CFUST, approximate $t$-cdf, EM algorithm

## 1. Introduction

Multivariate skew $t$-distributions (MST) reflect data asymmetry and robustness to outliers. For this reason, there is an increasing demand for mixture models of MST in applications such as statistical machine learning. However, fitting this model with the multiple skewness factors by exact-EM (Lee and McLachlan, 2016a) suffers from impractical processing time. The main reason for this problem is that the computation for multivariate cumulative distribution function (cdf) corresponding to a dimension of skewness factors is very expensive in processing time, and many multivariate cdfs have to be calculated to fit the model. For this reason, application models using a single skewness factor are often proposed (Lin *et al.*, 2018). However, it is recommended to consider multiple skewness factors for better fit performance to asymmetric data on multidimensional space.

To cope with the fitting time problem, Lee and McLachlan (2016b) attempted parallel processing techniques with multi-processors for fitting this model, which is a non-statistical approach. They reported that their computing could reduce the processing time for model fit by about 60%.

For approximation of univariate $t$-cdf, Zogheib and Elsaheli (2015) provides $t$ distribution. For approximation of multivariate $t$-cdf, we find in Genz and Bretz (2009), which is implemented by function `pmvt` in R package `mvtnorm`, and it is used a base line compared with the approximation proposed in this article.

The calculation of univariate $t$-cdf based on author's experience is exceptionally accurate and very fast regardless of the calculation method. We now provide a method of approximating multivariate cdf

[1] Department of Data and Information, Sangji University, Sangjidae-gil 83 (Woo San-dong), Wonju-si, Gangwon-do 26339, Korea. E-mail: sgukim@sangji.ac.kr

Published 31 March 2020 / journal homepage: http://csam.or.kr

with the conditional univariate cdfs for faster fitting the model. The calculation of cdfs is somewhat inaccurate; however, the result of fitting should be acceptable. Kim (2016) showed encouraging results using this approach for fitting mixtures of multivariate skew normal distributions (MSN). In this paper, a similar approach is applied to fit the mixtures of MST.

In the Section 2, we introduce the fitting method for the mixture model of MST by expectation-maximization (EM) algorithm, and provide the computational problems. In Section 3, an approximation method for multivariate $t$-cdf is proposed. In Section 4, we fit the mixture model of MST based on the approximate cdf, and compare it with the exact fit. In Section 5, conclusions and discussions about some of the problems with the proposed method are provided.

## 2. Preliminaries

### 2.1. Mixture model of multivariate skew $t$-distribution

Let $\boldsymbol{Y}_j = (Y_{1j}, \ldots, Y_{pj})^T$ be a $p$-dimensional random vector observed at the $j^{th}$ $(j = 1, \ldots, n)$ observation, and $\boldsymbol{y}_j$ denotes a realization of $\boldsymbol{Y}_j$. Consider a finite mixture model of $\boldsymbol{Y}_j$ with $g$-components

$$f\left(\boldsymbol{y}_j; \boldsymbol{\Theta}\right) = \sum_{i=1}^{g} \pi_i f_i\left(\boldsymbol{y}_j; \boldsymbol{\theta}_i\right), \quad j = 1, \ldots, n, \tag{2.1}$$

where $\pi_i$ $(i = 1, \ldots, g)$ are the mixing proportions, $\boldsymbol{\Theta}$ denotes the vector of all parameters in the mixture model, $\boldsymbol{\theta}_i$ is the vector of parameters in the density $f_i(\cdot)$.

And we consider in (2.1)

$$f_i\left(\boldsymbol{y}_j; \boldsymbol{\theta}_i\right) = 2^q t_p\left(\boldsymbol{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Omega}_i, \nu_i\right) T_q\left(\tilde{\boldsymbol{x}}_{ij}\sqrt{\frac{\nu_i + p}{\nu_i + d_i\left(\boldsymbol{y}_j\right)}}; \boldsymbol{0}, \boldsymbol{\Psi}_i, \nu_i + p\right), \tag{2.2}$$

where $t_d(\cdot\,; \boldsymbol{m}, \boldsymbol{S}, \nu)$ and $T_d(\cdots; \boldsymbol{m}, \boldsymbol{S}, \nu)$ are the $d$-variate pdf (probability density function) and cdf of multivariate $t$-distribution with the location vector $\boldsymbol{m}$, the scale matrix $\boldsymbol{S}$ and the degree of freedom $\nu$, respectively, $\boldsymbol{\Omega}_i = \boldsymbol{\Sigma}_i + \boldsymbol{\Delta}_i\boldsymbol{\Delta}_i^T$, $\tilde{\boldsymbol{x}}_{ij} = \boldsymbol{\Delta}_i^T\boldsymbol{\Omega}_i^{-1}(\boldsymbol{y}_j - \boldsymbol{\mu}_i)$, $d_i(\boldsymbol{y}_j) = (\boldsymbol{y}_j - \boldsymbol{\mu}_i)^T\boldsymbol{\Omega}_i^{-1}(\boldsymbol{y}_j - \boldsymbol{\mu}_i)$, $\boldsymbol{\Psi}_i = \mathbf{I} - \boldsymbol{\Delta}_i^T\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Delta}_i$, and where $\boldsymbol{\mu}_i$ is a $p$-dimensional vector, $\boldsymbol{\Sigma}_i$ is a $p \times p$ positive definite matrix, and $\boldsymbol{\Delta}_j$ is the $p \times q$ matrix $(q \leq p)$ called a "$q$-ple" skewness parameter; the $q$ indicates the dimension of a latent truncated variable (i.e., fundamental variable) leading $q$-fold skewness in (2.2). When $q = 1$, the (2.2) refers to Pyne *et al.* (2009)'s model which tried to explain data asymmetry by an univariate latent variable. However, many applications, it requires $q > 1$ (up to $q = p$) in order to sufficiently accommodate to asymmetry folded toward multidimensional data space.

The model (2.2) refers to canonical fundamental skew $t$-distribution (CFUST) (Arellano-Valle and Genton, 2005) which is regarded as the standard form of MST recently. If $\nu \to \infty$, then it becomes canonical fundamental skew normal distribution (CFUSN), and additionally when $\boldsymbol{\Delta}_i = \boldsymbol{0}$, it reduces to a $p$-variate normal distribution $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$.

### 2.2. Fitting by exact-EM algorithm

Given $n$ observations $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_n^T)^T$, our aim is to maximize the log-likelihood

$$L(\boldsymbol{\Theta}) \propto \sum_{j=1}^{n} \log\left\{\sum_{i=1}^{g} \pi_i t_p\left(\boldsymbol{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Omega}_i, \nu_i\right) T_q\left(\tilde{\boldsymbol{x}}_{ij}\sqrt{\frac{\nu_i + p}{\nu_i + d_i(\boldsymbol{y}_j)}}; \boldsymbol{0}, \boldsymbol{\Psi}_i, \nu_i + p\right)\right\} \tag{2.3}$$

with respect to the $\boldsymbol{\Theta}$ for fitting the model. EM algorithm is used to estimate the parameters because the direct maximization of the $L(\boldsymbol{\Theta})$ is untractable.

A convenient hierarchical generative representation for the model (2.1) is given by

$$\begin{aligned}
\boldsymbol{Y}_j|\boldsymbol{x}_{ij}, u_{ij}, z_{ij} = 1 &\sim N_p\left(\boldsymbol{\mu}_i + \boldsymbol{\Delta}_i\boldsymbol{x}_{ij}, \frac{\boldsymbol{\Sigma}_i}{u_{ij}}\right),\\
\boldsymbol{X}_{ij}|u_{ij}, z_{ij} = 1 &\sim HN_q\left(\boldsymbol{0}, \frac{\mathbf{I}}{u_{ij}}\right),\\
U_{ij}|z_{ij} = 1 &\sim \Gamma\left(\frac{\nu_i}{2}, \frac{\nu_i}{2}\right),\\
\boldsymbol{Z}_j = \left(Z_{1j}, \ldots, Z_{gj}\right)^T &\sim \text{multinomial}(1, \pi_1, \ldots, \pi_g),
\end{aligned}$$

where multinomial$(1, \pi_1, \ldots, \pi_g)$ denotes the multinomial distribution with one draw and $g$ categories with probabilities $(\pi_1, \ldots, \pi_g)$, $\Gamma(\alpha,\beta)$ denotes the gamma distribution with the mean $\alpha/\beta$, and $HN_q(\boldsymbol{m}, \boldsymbol{S})$ denotes the half normal distribution with location vector $\boldsymbol{m}$ and scale matrix $\boldsymbol{S}$.

With the missing data $\{\boldsymbol{X}_{ij}, U_{ij}, Z_{ij}\}$ and the observed data $\{\boldsymbol{Y}_j\}$, the log-likelihood for complete data is given by

$$\begin{aligned}
L_c(\boldsymbol{\Theta}) \propto &\sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\left[\log \pi_i - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}u_{ij}\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i - \boldsymbol{\Delta}_i\boldsymbol{x}_{ij}\right)^T \boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i - \boldsymbol{\Delta}_i\boldsymbol{x}_{ij}\right)\right]\\
&+ \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\left[\frac{\nu_i}{2}\log\left(\frac{\nu_i}{2}\right) - \log\Gamma\left(\frac{\nu_i}{2}\right) + \left(\frac{\nu_i}{2}\right)\left(\log u_{ij} - u_{ij}\right)\right]. \qquad (2.4)
\end{aligned}$$

The EM algorithm, at the $(k + 1)$ iteration, maximizes $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(k)}) = E[L_c(\boldsymbol{\Theta})|\boldsymbol{y}, \boldsymbol{\Theta}^{(k)}]$ with respect to $\boldsymbol{\Theta}$, which is as follows.

In E-step, Calculate the conditional expectations of missing data on $\boldsymbol{y}_j$: First, for $i = 1, \ldots, g$; $j = 1, \ldots, n$, the conditional expectations of $Z_{ij}$

$$\tau_{ij}^{(k+1)} = E\left[Z_{ij}|\boldsymbol{y}_j, \boldsymbol{\Theta}^{(k)}\right] = \frac{\pi_i^{(k)} f_i\left(\boldsymbol{y}_j; \boldsymbol{\theta}_i^{(k)}\right)}{\sum_{h=1}^{g} \pi_h^{(k)} f_h\left(\boldsymbol{y}_j; \boldsymbol{\theta}_h^{(k)}\right)},$$

where

$$f_i\left(\boldsymbol{y}_j; \boldsymbol{\theta}^{(k)}\right) = 2^q t_p\left(\boldsymbol{y}_j; \boldsymbol{\mu}_i^{(k)}, \boldsymbol{\Omega}_i^{(t)}, \nu_i^{(k)}\right) T_q\left(\tilde{\boldsymbol{x}}_{ij}^{(k)}\sqrt{\frac{\nu_i^{(t)} + p}{\nu_i^{(k)} + d_i^{(k)}\left(\boldsymbol{y}_j\right)}}; \boldsymbol{0}, \boldsymbol{\Psi}_i^{(k)}, \nu_i^{(k)} + p\right). \qquad (2.5)$$

Second, the conditional expectations of $U_{ij}$

$$\begin{aligned}
u_{ij}^{(k+1)} &= E\left[U_{ij}|\boldsymbol{y}_j, Z_{ij} = 1, \boldsymbol{\Theta}^{(k)}\right]\\
&= \left(\frac{\nu_i^{(k)} + p}{\nu_i^{(t)} + d_i^{(k)}\left(\boldsymbol{y}_j\right)}\right)\frac{T_q\left(\tilde{\boldsymbol{x}}_{ij}^{(k)}\sqrt{\frac{\nu_i^{(k)}+p+2}{\nu_i^{(k)}+d_i^{(k)}(\boldsymbol{y}_j)}}; \boldsymbol{0},\ \boldsymbol{\Psi}_i^{(k)},\ \nu_i^{(k)} + p + 2\right)}{T_q\left(\tilde{\boldsymbol{x}}_{ij}^{(k)}\sqrt{\frac{\nu_i^{(k)}+p}{\nu_i^{(k)}+d_i^{(k)}(\boldsymbol{y}_j)}}; \boldsymbol{0},\ \boldsymbol{\Psi}_i^{(k)},\ \nu_i^{(k)} + p\right)}. \qquad (2.6)
\end{aligned}$$

Third, a couple of $q$-dimensional moment

$$\begin{aligned} \boldsymbol{m}_{ij}^{(k+1)} &= E\left[\boldsymbol{X}_{ij}|\boldsymbol{y}_j, Z_{ij} = 1, \boldsymbol{\Theta}^{(k)}\right], \\ \boldsymbol{M}_{ij}^{(k+1)} &= E\left[\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T|\boldsymbol{y}_j, Z_{ij} = 1, \boldsymbol{\Theta}^{(k)}\right], \end{aligned} \tag{2.7}$$

which are first and second moment of the $q$-variate truncated $t$ distribution with the support $(0,\infty)^q$

$$Tt_q\left(\tilde{\boldsymbol{x}}_{ij}^{(k)}\sqrt{\frac{\nu_i^{(k)}+p+2}{\nu_i^{(k)}+d_i^{(k)}\left(\boldsymbol{y}_j\right)}},\ \boldsymbol{\Psi}_i^{(t)},\ \nu_i^{(k)}+p+2;(0,\infty)^q\right),$$

respectively.

Lastly, the conditional expectations of $\log U_{ij}$

$$\begin{aligned} \log u_{ij}^{(k+1)} &= E\left[\log U_{ij}|\boldsymbol{y}_j, Z_{ij} = 1, \boldsymbol{\Theta}^{(k)}\right], \\ &= u_{ij}^{(k+1)} - \left(\frac{\nu_i^{(k)}+p}{\nu_i^{(k)}+d_i^{(k)}\left(\boldsymbol{y}_j\right)}\right) - \log\left(\frac{\nu_i^{(k)}+p}{2}\right) + \varphi\left(\frac{\nu_i^{(k)}+d_i^{(k)}\left(\boldsymbol{y}_j\right)}{2}\right) \\ &\quad + 2\left[\frac{d}{d\nu}\log T_q\left(\tilde{\boldsymbol{x}}_{ij}^{(k)}\sqrt{\frac{\nu_i^{(k)}+p}{\nu_i^{(k)}+d_i^{(k)}\left(\boldsymbol{y}_j\right)}};\boldsymbol{0},\ \boldsymbol{\Psi}_i^{(t)},\ \nu_i^{(k)}+p\right)\right]_{\nu_i=\nu_i^{(k)}}, \end{aligned} \tag{2.8}$$

where $\varphi(\cdot)$ denotes the digamma function.

And in M-step, for $i = 1,\ldots,g$,

$$\begin{aligned} \pi_i^{(k+1)} &= \frac{1}{n}\sum_{j=1}^n \tau_{ij}^{(k+1)}, \\ \boldsymbol{\mu}_i^{(k+1)} &= \frac{\sum_{j=1}^n w_{ij}^{(k+1)}\left(\boldsymbol{y}_j - \boldsymbol{\Delta}_i^{(t)}\boldsymbol{m}_{ij}^{(k+1)}\right)}{\sum_{j=1}^n w_{ij}^{(k+1)}}, \\ \boldsymbol{\Delta}_i^{(k+1)} &= \left[\sum_{j=1}^n w_{ij}^{(k+1)}\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i^{(t+1)}\boldsymbol{m}_{ij}^{(k+1)}\right)\right]\left[\sum_{j=1}^n w_{ij}^{(k+1)}\boldsymbol{M}_{ij}^{(k+1)}\right]^{-1}, \\ \boldsymbol{\Sigma}_i^{(k+1)} &= \frac{1}{\sum_{j=1}^n \tau_{ij}^{(k+1)}}\left[\sum_{j=1}^n w_{ij}^{(k+1)}\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i^{(k+1)}\right)\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i^{(k+1)}\right)^T - \boldsymbol{\Delta}_i^{(k+1)}\sum_{j=1}^n w_{ij}^{(k+1)}\boldsymbol{m}_{ij}^{(k+1)}\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i^{(k+1)T}\right)\right. \\ &\quad \left. - \sum_{j=1}^n w_{ij}^{(k+1)}\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i^{(k+1)}\right)\boldsymbol{m}_{ij}^{(k+1)T}\boldsymbol{\Delta}_i^{(k+1)T} + \boldsymbol{\Delta}_i^{(k+1)}\left(\sum_{j=1}^n w_{ij}^{(k+1)}\boldsymbol{M}_{ij}^{(k+1)}\right)\boldsymbol{\Delta}_i^{(k+1)T}\right], \end{aligned}$$

where $w_{ij}^{(k+1)} = \tau_{ij}^{(k+1)}u_{ij}^{(k+1)}$. And $\nu_i^{(k+1)}$ is obtained by solving the following equation

$$0 = \log\left(\nu_i^{\frac{k+1}{2}}\right) - \varphi\left(\nu_i^{\frac{k+1}{2}}\right) + 1 - \bar{e}_i^{(k+1)},$$

where $\bar{e}_i^{(k+1)} = \sum_{j=1}^n \tau_{ij}^{(k+1)}(\log u_{ij}^{(k+1)} - u_{ij}^{(k+1)})/\sum_{j=1}^n \tau_{ij}^{(k+1)}$.

For the initial estimates $\mathbf{\Theta}^{(0)}$, we follow Lin (2010)'s method: First, $g$ clusters $G_1, \ldots, G_g$ are roughly determined by a clustering method like the Kmeans. And then, for each data matrix $\boldsymbol{Y}_{G_i}$, compute $\pi_i^{(0)} = \#\{G_i\}/n$, for a constant $a \in (0,1)$ (e.g., $a = 0.9$), $\mathbf{\Sigma}_i^{(0)} = \boldsymbol{S}_i - (1-a)\mathrm{Diag}(\boldsymbol{S}_i)$, $\mathbf{\Delta}_i^{(0)} = \mathrm{diag}(K\sqrt{(1-a)/(1-2/\pi)})$, and $\boldsymbol{\mu}_i^{(0)} = \boldsymbol{m}_i - \sqrt{2/\pi}\mathbf{\Delta}^{(0)}$, where $K = \mathrm{sign}(\mathrm{skewness}(\boldsymbol{Y}_{G_i}))$, and $\boldsymbol{m}_i$ and $\boldsymbol{S}_i$ are the sample mean vector and the sample covariance matrix for $\boldsymbol{Y}_{G_i}$.

The Exact-EM algorithm is presented roughly. In M-step, there is no expensive part in computation time. However, all conditional expectations from (2.5) to (2.8) in E-Step contain $q$-dimensional cdfs $T_q(\cdot)$. Therefore, at every iteration, many time-consuming multidimensional cdfs must be calculated for each observation $j$ and component $i$ when $q > 1$. Especially many multidimensional cdf calculations are required to obtain $\boldsymbol{m}_{ij}^{(k+1)}$ and $\boldsymbol{M}_{ij}^{(k+1)}$ in (2.7) by Ho *et al.* (2012)'s routine. This is the main reason why the mixture model of CFUST is impractical for "Big data".

It is well known that the calculation of univariate $t$-cdf is exceptionally accurate and very fast regardless of the calculation method. Thus we present an approximate but faster way to compute multivariate $t$-cdf in univariate fashion in the following section.

## 3. Approximation of *t*-cdf

Consider Taylor's first order approximation of the continuous function $h(X)$ of a random variable $X$ at $\eta = E(X)$

$$E(h(X)) \approx h(\eta) + h'(\eta)E(X-\eta) = h(\eta). \tag{3.1}$$

For convenience of notation, write $t$-cdf $T_q(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{\Sigma}, \nu)$ as $T_q(\boldsymbol{x}; \nu)$, and $t$-pdf $t_q(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{\Sigma}, \nu)$ as $t_q(\boldsymbol{x}; \nu)$. And for $\mathbf{\Sigma} = \{\sigma_{ij}\}$, write the $\sigma_{ii}$ as $\sigma_i^2$ .

Based on (3.1), the $t$-cdf is approximated at the first stage by

$$\begin{aligned}
T_q(\boldsymbol{x}; \nu) &= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_q} t_q\left(\ell_1, \ell_2, \ldots, \ell_q; \nu\right) d\ell_q \cdots d\ell_2 d\ell_1 \\
&= \int_{-\infty}^{x_1} t_1(\ell_1; \nu)\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_q} t_{q-1}\left(\ell_2, \ldots, \ell_q|\ell_1; \nu+1\right) d\ell_q \cdots d\ell_2 d\ell_1 \\
&= T_1(x_1; \nu)\int_{-\infty}^{x_1}\frac{t_1(\ell_1;\nu)}{T_1(x_1;\nu)}\left[\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_q} t_{q-1}\left(\ell_2, \ldots, \ell_q|\ell_1; \nu+1\right) d\ell_q \cdots d\ell_2\right] d\ell_1 \\
&= T_1(x_1; \nu)E\left[\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_q} t_{q-1}\left(\ell_2, \ldots, \ell_q|X_1; \nu+1\right) d\ell_q \cdots d\ell_2\middle| X_1 \le x_1\right] \\
&\approx T_1(x_1; \nu)\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_q} t_{q-1}\left(\ell_2, \ldots, \ell_q|\eta_1; \nu+1\right) d\ell_q \cdots d\ell_2 \\
&= T_1(x_1; \nu)T_{q-1}\left(\boldsymbol{x}_{(1)}|\eta_1; \nu+1\right),
\end{aligned} \tag{3.2}$$

where $\boldsymbol{x}_{(1,2,\ldots)}$ denotes the $\boldsymbol{x}$ with $(x_1, x_2, \ldots)$ deleted,

$$\eta_1 = E[X_1|X_1 \le x_1] = \mu_1 - \sigma_1\left(\frac{\nu+z_1^2}{\nu-1}\right)\frac{t_1(z_1;\nu)}{T_1(z_1;\nu)}, \tag{3.3}$$

which is the 1st moment of univariate left-truncated $t$-distribution $t_1(\mu_1, \sigma_1, \nu; (-\infty, x_1))$, and $z_1 = (x_1-\mu_1)/\sigma_1$. Note in the second equation of (3.2) that the degree of freedom of the conditional $t$ distribution increases by the dimension of the condition that is 1 here.

By similarly way for $T_{q-1}(\boldsymbol{x}_{(1)}|\eta_1;\nu+1)$ in (3.2), we obtain at the second stage

$$T_{q-1}\left(\boldsymbol{x}_{(1)}|\eta_1;\nu+1\right) \approx T_1\left(x_2|\eta_1;\nu+1\right)T_{q-2}\left(\boldsymbol{x}_{(1,2)}|\eta_1,\eta_2;\nu+2\right).$$

Thus $T_q(\boldsymbol{x};\nu) \approx T_1(x_1;\nu)T_1(x_2|\eta_1;\nu+1)T_{q-2}(\boldsymbol{x}_{(1,2)}|\eta_1,\eta_2;\nu+2)$, where

$$\eta_2 = E\left[X_2|X_1=\eta_1, X_2 \le x_2\right] = \mu_{2|1} - \sigma_{2|1}\left[\frac{(\nu+1)+z_2^2}{(\nu+1)-1}\right]\frac{t_1(z_2;\nu+1)}{T_1(z_2;\nu+1)}, \tag{3.4}$$

$z_2 = (x_2-\mu_{2|1})/\sigma_{2|1}$, $\mu_{2|1} = \mu_2+\sigma_{21}(\eta_1-\mu_1)^2/\sigma_1^2$, $\sigma_{2|1} = \alpha_1(\sigma_2^2-\sigma_{21}^2/\sigma_1^2)$, and $\alpha_1 = \{\nu+(\eta_1-\mu_1)^2/\sigma_1^2\}/(\nu+1)$.

Subsequently, after $q$ stage, we obtain

$$\begin{aligned}
T_q(\boldsymbol{x};\nu) &\approx T_1(x_1;\nu)T_1\left(x_2|\eta_1;\nu+1\right)T_1\left(x_3|\eta_1,\eta_2;\nu+2\right)\times\cdots\times T_1\left(x_q|\eta_1,\ldots,\eta_{q-1};\nu+q-1\right)\\
&= T_1(x_1;\nu)\prod_{i=2}^{q} T_1\left(x_i|\boldsymbol{\eta}_{i_-};\nu+i-1\right), \qquad (3.5)
\end{aligned}$$

where $i_- = \{1,2,\ldots,i-1\}$, thus $\boldsymbol{\eta}_{i_-} = (\eta_1,\ldots,\eta_{i-1})^T$, and

$$\begin{aligned}
\eta_i &= E\left[X_i|X_1=\eta_1,\ldots,X_{i-1}=\eta_{i-1}|X_i \le x_i\right]\\
&= \mu_{i|i_-} - \sigma_{i|i_-}\left[\frac{(\nu+i-1)+z_i^2}{(\nu+i-1)-1}\right]\frac{t_1(z_i;\nu+i-1)}{T_1(z_i;\nu+i-1)}, \qquad (3.6)\\
\mu_{i|i_-} &= E\left[X_i|X_1=\eta_1,\ldots,X_{i-1}=\eta_{i-1}\right]\\
&= \mu_i + \boldsymbol{\Sigma}_{i,i_-}\boldsymbol{\Sigma}_{i_-,i_-}^{-1}\left(\boldsymbol{\eta}_{i_-}-\boldsymbol{\mu}_{i_-}\right), \qquad (3.7)\\
\sigma_{i|i_-} &= \operatorname{var}\left[X_i|X_1=\eta_1,\ldots,X_{i-1}=\eta_{i-1}\right]\\
&= \alpha_{i-1}\left[\sigma_i^2 - \boldsymbol{\Sigma}_{i,i_-}\boldsymbol{\Sigma}_{i_-,i_-}^{-1}\boldsymbol{\Sigma}_{i_-,i}\right], \qquad (3.8)\\
\alpha_{i-1} &= \frac{(\nu+i-2)+\left(\boldsymbol{\eta}_{i_-}-\boldsymbol{\mu}_{i_-}\right)^T\boldsymbol{\Sigma}_{i_-,i_-}^{-1}\left(\boldsymbol{\eta}_{i_-}-\boldsymbol{\mu}_{i_-}\right)}{\nu+i-1}, \qquad (3.9)\\
z_i &= \frac{x_i-\mu_{i|i_-}}{\sigma_{i|i_-}}, \qquad (3.10)
\end{aligned}$$

specifying $\alpha_0 = 1$, and $\boldsymbol{X}_{i_-} = (X_1,\ldots,X_{i-1})^T$, $\boldsymbol{\mu}_{i_-} = E(\boldsymbol{X}_{i_-})$, $\boldsymbol{\Sigma}_{i_-,i_-} = \operatorname{var}(\boldsymbol{X}_{i_-})$, and $\boldsymbol{\Sigma}_{i,i_-} = \boldsymbol{\Sigma}_{i_-,i}^T = \operatorname{cov}(X_i,\boldsymbol{X}_{i_-})$, respectively.

Note that the approximate result of (3.5) is the product of all univariate $t$-cdfs, therefore calculated quickly. It is also easy to see that it becomes the approximate multivariate normal cdf when $\nu \to \infty$.

We investigate exact cdfs and approximate cdfs with $\boldsymbol{\mu} = \boldsymbol{0}$ for various realizations $\boldsymbol{x}$, scale matrices $\boldsymbol{\Sigma}$, and degree of freedom $\nu$: $\boldsymbol{x} = c\boldsymbol{1}$ for $c \in \{-2,-1,0,1,2\}$, $\nu \in \{5,10,20,100,\infty\}$, and the scale matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & r & r & r\\ r & 1 & r & r\\ \vdots & \vdots & \ddots & \vdots\\ r & r & \cdots & 1 \end{bmatrix},$$

Table 1: Comparison between exact cdf $F$ and approximates $F^*$ when $q = 3$

| $\boldsymbol{x}$ | $\nu$ | $r$ |  |  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  | 0.0 |  |  | 0.1 |  |  | 0.5 |  |  | 0.9 |  |  |
|  |  | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ |
| $-(2, 2, 2)$ | 5 | 0.00094 | 0.00091 | 0.00002 | 0.00165 | 0.00162 | 0.00003 | 0.00829 | 0.00850 | −0.00020 | 0.02842 | 0.03588 | −0.00746 |
|  | 10 | 0.00024 | 0.00023 | 0.00001 | 0.00052 | 0.00050 | 0.00002 | 0.00423 | 0.00402 | 0.00020 | 0.01885 | 0.02199 | −0.00314 |
|  | 20 | 0.00008 | 0.00008 | 0.00000 | 0.00021 | 0.00021 | 0.00001 | 0.00264 | 0.00243 | 0.00021 | 0.01429 | 0.01591 | −0.00162 |
|  | 100 | 0.00002 | 0.00002 | 0.00000 | 0.00007 | 0.00007 | 0.00000 | 0.00160 | 0.00146 | 0.00014 | 0.01094 | 0.01163 | −0.00070 |
|  | ∞ | 0.00001 | 0.00001 | 0.00000 | 0.00005 | 0.00005 | 0.00000 | 0.00138 | 0.00126 | 0.00012 | 0.01013 | 0.01066 | −0.00052 |
| $-(1, 1, 1)$ | 5 | 0.00926 | 0.00942 | −0.00016 | 0.01432 | 0.01460 | −0.00028 | 0.04713 | 0.04995 | −0.00282 | 0.11696 | 0.14536 | −0.02840 |
|  | 10 | 0.00649 | 0.00654 | −0.00005 | 0.01075 | 0.01074 | 0.00001 | 0.04059 | 0.04058 | 0.00001 | 0.10741 | 0.12866 | −0.02125 |
|  | 20 | 0.00520 | 0.00523 | −0.00003 | 0.00902 | 0.00898 | 0.00003 | 0.03720 | 0.03639 | 0.00081 | 0.10239 | 0.12016 | −0.01777 |
|  | 100 | 0.00423 | 0.00424 | −0.00001 | 0.00768 | 0.00764 | 0.00004 | 0.03446 | 0.03324 | 0.00122 | 0.09835 | 0.11340 | −0.01504 |
|  | ∞ | 0.00399 | 0.00399 | 0.00000 | 0.00736 | 0.00731 | 0.00005 | 0.03380 | 0.03248 | 0.00132 | 0.09733 | 0.11172 | −0.01438 |
| $(0, 0, 0)$ | 5 | 0.12500 | 0.12500 | 0.00000 | 0.14891 | 0.15123 | −0.00232 | 0.25001 | 0.26843 | −0.01842 | 0.39232 | 0.46324 | −0.07092 |
|  | 10 | 0.12500 | 0.12500 | 0.00000 | 0.14891 | 0.14988 | −0.00096 | 0.25001 | 0.26085 | −0.01085 | 0.39233 | 0.46208 | −0.06975 |
|  | 20 | 0.12500 | 0.12500 | 0.00000 | 0.14891 | 0.14927 | −0.00036 | 0.24999 | 0.25741 | −0.00741 | 0.39232 | 0.46198 | −0.06965 |
|  | 100 | 0.12500 | 0.12500 | 0.00000 | 0.14891 | 0.14882 | 0.00009 | 0.25000 | 0.25482 | −0.00482 | 0.39233 | 0.46224 | −0.06991 |
|  | ∞ | 0.12500 | 0.12500 | 0.00000 | 0.14891 | 0.14872 | 0.00020 | 0.25000 | 0.25420 | −0.00420 | 0.39232 | 0.46236 | −0.07004 |
| $(1, 1, 1)$ | 5 | 0.56196 | 0.57466 | −0.01270 | 0.57823 | 0.59323 | −0.01501 | 0.64867 | 0.68886 | −0.04019 | 0.74750 | 0.80950 | −0.06201 |
|  | 10 | 0.57804 | 0.58512 | −0.00708 | 0.59378 | 0.60237 | −0.00859 | 0.66276 | 0.69759 | −0.03482 | 0.75995 | 0.82386 | −0.06391 |
|  | 20 | 0.58661 | 0.59029 | −0.00368 | 0.60204 | 0.60703 | −0.00500 | 0.67014 | 0.70268 | −0.03254 | 0.76648 | 0.83151 | −0.06504 |
|  | 100 | 0.59373 | 0.59449 | −0.00076 | 0.60888 | 0.61089 | −0.00201 | 0.67621 | 0.70722 | −0.03101 | 0.77181 | 0.83773 | −0.06593 |
|  | ∞ | 0.59556 | 0.59556 | 0.00000 | 0.61063 | 0.61188 | −0.00125 | 0.67777 | 0.70843 | −0.03066 | 0.77316 | 0.83928 | −0.06612 |
| $(2, 2, 2)$ | 5 | 0.86338 | 0.88390 | −0.02052 | 0.86743 | 0.88721 | −0.01978 | 0.88809 | 0.91731 | −0.02923 | 0.92295 | 0.94796 | −0.02500 |
|  | 10 | 0.89707 | 0.90694 | −0.00987 | 0.89960 | 0.90969 | −0.01008 | 0.91442 | 0.93831 | −0.02389 | 0.94210 | 0.96307 | −0.02097 |
|  | 20 | 0.91501 | 0.91963 | −0.00461 | 0.91672 | 0.92212 | −0.00540 | 0.92828 | 0.94974 | −0.02145 | 0.95204 | 0.97032 | −0.01828 |
|  | 100 | 0.92960 | 0.93047 | −0.00087 | 0.93076 | 0.93276 | −0.00200 | 0.93971 | 0.95929 | −0.01958 | 0.95971 | 0.97589 | −0.01618 |
|  | ∞ | 0.93329 | 0.93329 | 0.00000 | 0.93432 | 0.93553 | −0.00121 | 0.94253 | 0.96172 | −0.01919 | 0.96175 | 0.97725 | −0.01550 |

for $r \in \{0, 0.1, 0.5, 0.9\}$. For the exact cdf, we exploited the function `pmvt` in R package `mvtnorm`, where it is the average of 10 replications.

In Table 1, exact cdfs denoted by $F$ and approximate cdfs denoted by$F^*$ are presented for $q = 3$. Results for $q = 2, 4, 5, 10$ are given in the Appendix A. It is shown that the approximations are as a whole satisfactory, except in extreme situations where $\boldsymbol{x} = \boldsymbol{0}$ and $r = 0.9$. As expected, the greater the degree of freedom, the better the approximation. Therefore the error is the smallest at normal distribution ($\nu = \infty$). The larger the dimension, the worse the approximation. For computational speed in time, the approximation technique is about 1, 7, 12, 61 times faster than exact calculation for $q = 2, 3, 5, 10$ respectively.

There is one thing to note here. As Kim (2016) points out in the case of a normal distribution that if the elements of $\boldsymbol{x} = (x_1, \ldots, x_q)^T$ are different, then they must be sorted in ascending order before the approximate cdf is calculated, otherwise it gives a worse approximations. This property is confirmed to be applied to the $t$-distribution as well, though its result is not included in this paper. Therefore, it is necessary to sort it in ascending order such as $x_1 \geq x_2 \geq \cdots \geq x_q$ before calculating the approximate cdf, where each $x_i$ is standardized.

Our main concern is to see how the approximation technique differs from the exact fit in the mixture of CFUST fit. We fit the model by replacing all cdfs in E-step with approximate cdf proposed in this section. In the next section presents the performance of the mixture of CFUST based on the approximate cdf.

## 4. Experiments

In this section, the mixture of CFUST model fits the Australian Institution Sport (AIS) data set provided in Cook and Weisberg (1994) that contains 11 variables measured for 100 female and 102 male athletes. This data set is well known to be asymmetrical and contains outliers: As athletes are selected

Table 2: Estimates by Exact-EM and Approx-EM (maximum degree of freedom is specified by 300)

| | | Estimates | | | | |
|---|---|---|---|---|---|---|
| | | $\hat{\pi}_i$ | $\hat{\boldsymbol{\mu}}_i$ | $\hat{\boldsymbol{\Sigma}}_i$ | $\hat{\boldsymbol{\Delta}}_i$ | $\hat{\nu}_i$ |
| Exact-EM | Component1 | 0.47 | $(-0.27, -0.16, 1.23)$ | $\begin{bmatrix} 0.24 & 0.31 & 0.02 \\ 0.31 & 0.40 & 0.02 \\ 0.02 & 0.02 & 0.00 \end{bmatrix}$ | $\begin{bmatrix} 0.61 & 0.20 & 0.44 \\ 0.02 & 0.70 & 0.15 \\ 0.01 & -0.02 & 0.55 \end{bmatrix}$ | 24.44 |
| | Component2 | 0.53 | $(-0.70, -0.11, 0.17)$ | $\begin{bmatrix} 0.02 & 0.08 & 0.01 \\ 0.08 & 0.38 & -0.02 \\ 0.01 & -0.02 & 0.13 \end{bmatrix}$ | $\begin{bmatrix} -0.63 & 0.74 & -0.06 \\ -0.98 & 0.37 & 0.16 \\ -0.61 & 0.14 & 1.11 \end{bmatrix}$ | 300 |
| Approx-EM | Component1 | 0.47 | $(-0.11, 0.02, -1.21)$ | $\begin{bmatrix} 0.27 & 0.34 & 0.02 \\ 0.34 & 0.46 & 0.03 \\ 0.02 & 0.03 & 0.00 \end{bmatrix}$ | $\begin{bmatrix} 0.56 & 0.09 & 0.41 \\ -0.02 & 0.54 & 0.12 \\ 0.00 & -0.03 & 0.54 \end{bmatrix}$ | 19.40 |
| | Component2 | 0.53 | $(-0.72, -0.31, 0.31)$ | $\begin{bmatrix} 0.02 & 0.07 & 0.00 \\ 0.07 & 0.36 & -0.08 \\ 0.00 & -0.08 & 0.15 \end{bmatrix}$ | $\begin{bmatrix} -0.63 & 0.74 & -0.04 \\ -0.94 & 0.43 & 0.29 \\ -0.68 & 0.11 & 1.04 \end{bmatrix}$ | 300 |

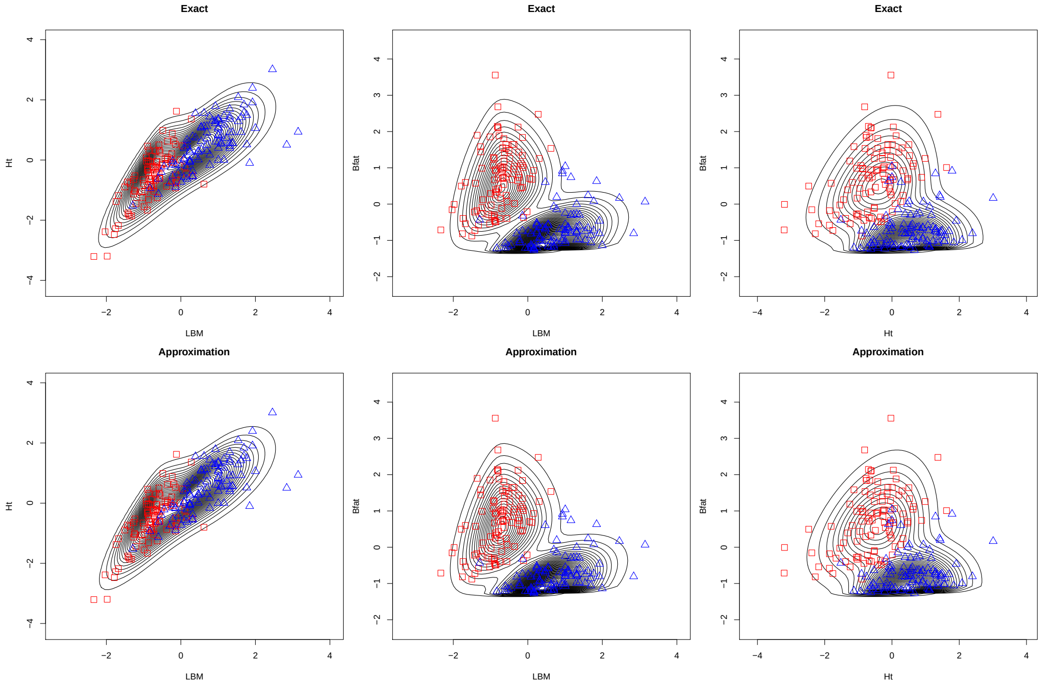

Figure 1: *Results of fit for* (LBM, Ht, Bfat). *(a) 1st row: By Exact-EM, (b) 2nd row: By Approx-EM. Symbols:* □ *and* △ *indicate female and male, respectively.* $a = 0.9$ *for the initial estimates. Note: Fit contours are calculated with exact cdfs given their estimates.*

from the average person (i.e., they are truncated from the normal), their various physical characteristics are expected to be skewed.

First, we experiment for 3 variables (LBM, Ht, Bfat) among them, those are standardized before fit. Approx-EM is the fitting technique based on the proposed approximate cdf.

In Table 2, the estimates of parameter fitted by Exact-EM and Approx-EM with $q = 3$ under the same initial estimates for $\pi_i$, $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$, and $\boldsymbol{\Delta}_i$ $(i = 1, 2)$. They shows quite slightly difference. In Figure 1, the fitted contours corresponding to their estimates are displayed (1st row: Exact-EM, 2nd row:

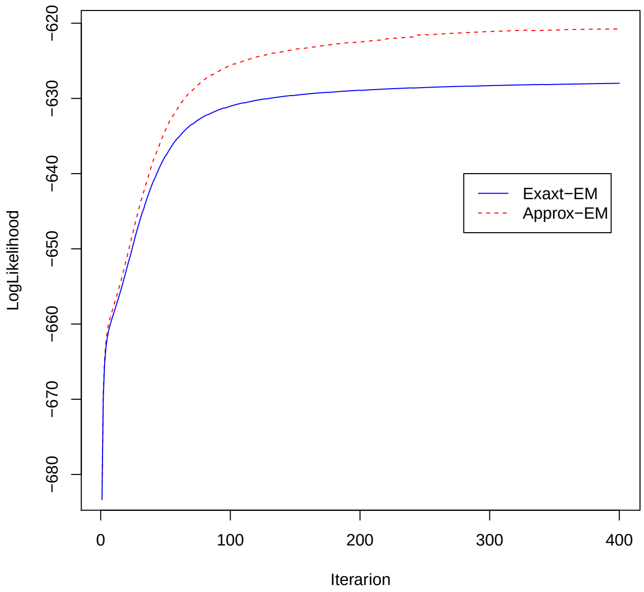

Figure 2: *Log-likelihood plots over iteration. solid line (blue): Exact-EM, dot line (red): Approx-EM.*

Table 3: Log-Likelihoods and execution time (PC used: Intel Core i7-8700 3.2GHz)

| Skewness dimension | $q = 2$ | | $q = 3$ | | $q = 4$ | | $q = 5$ | |
|---|---|---|---|---|---|---|---|---|
| method | Exact-EM | Approx-EM | Exact-EM | Approx-EM | Exact-EM | Approx-EM | Exact-EM | Approx-EM |
| Data Set | (BMI, Bfat) | | (LBM, Ht ,Bfat) | | (BMI, Bfat, LBM, ferr) | | (BMI, Bfat, LBM, ferr, hg) | |
| # of iterations | 400 | | 400 | | 300 | | 300 | |
| Log-Likelihood | −487.54 | −487.72 | −627.99 | −628.35 | −808.97 | −809.71 | −992.0 | −994.23 |
| Etime (minutes) | 5.99 | 5.07 | 38.40 | 9.58 | 61.41 | 12.76 | 131.02 | 19.14 |
| Time reduction (%) | 15 | | 75 | | 80 | | 85 | |

Approx-EM), which are so similar that it is difficult to visually distinguish the differences.

Figure 2 illustrates the log-likelihood plot over the iteration of the two methods. Near the convergence point, Approx-EM's log likelihood calculated by approximate cdfs is over-evaluated than the Exact-EM. This phenomenon is presumed to be due to the approximate cdf that has a generally positive bias. However, the convergence path patterns of the two algorithms look very similar.

Table 3 shows the log-likelihoods and the fitting times when $q = 2, 3, 4, 5$, where the log likelihood of Approx-EM is calculated by exact cdfs under its estimates for a fair comparison. The log likelihoods for Exact-EM and Approx-EM are nearly the same. Approx-EM shows a slightly lower fitness as dimension of the skewness increases. The results of fit contour for two data set (BMI, Bfat, LBM, ferr) with $q = 4$ and (BMI, Bfat, LBM, ferr, hg) with $q = 5$ are presented in the Appendix B.

For the processing time for fitting, Approx-EM is less than Exact-EM and becomes clearer as $q$ increases. It reduces the processing time by 85% when $q = 5$, which is better than Lee and McLachlan (2016b)'s results with a 60% reduction time.

All experiments in this section was implemented purely in R, and the PC used was Intel Core i7-8700 (3.2GHz).

## 5. Conclusion and discussions

The fit by the EM algorithm of the mixture model of CFUST with multidimensional skewness factor suffers from processing time. The main reason is that many multidimensional cdfs have to be calculated in E-step. In this paper, we propose an approximated EM algorithm that fits the model by approximating the multidimensional cdf in univariate fashion. Various experiments have shown the effectiveness of the proposed approximation method.

For applicability of approximation with (3.1), (as an anonymous reviewer noted) if $h$ is convex, then $E[h(X)] \geq h[E(X)]$, otherwise $E[h(X)] \leq h[E(X)]$ by Jensen's inequality. But the cdf in (3.2) is not strictly convex nor concave, which probably leads (3.1) to be applicable because each data point could be placed arbitrarily.

In case of $q \leq 5$, the performance of the proposed method is valid, but it is not confirmed how it is for further dimension. The approximate performance is deteriorated if $q$ is very large. For a more accurate approximation, one can consider the second order Taylor's expansion in (3.1)

$$E(h(X)) \approx h(\eta) + h'(\eta)E(X-\eta) + \frac{1}{2}h''(\eta)E(X-\eta)^2. \tag{5.1}$$

Unfortunately, however, the author found later that $h''(\eta)$ contains $(q-2)$-dimensional cdfs, so that the second order approximations might have no gain unless $q$ is less than and equal to 4.

Approx-EM is somewhat sensitive to initial estimates; in addition, the monotonicity of the algorithm is sometimes slightly broken near the convergence point depending on the data set and the initial estimates. These problems should be solved in future studies.

## Appendix A: Tables of approximate cdf

Table A.1: Comparison between true cdf $F$ and approximates $F^*$ when $q = 2$

| $\boldsymbol{x}$ | $\nu$ | $r$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.0 | | | 0.1 | | | 0.5 | | | 0.9 | | |
| | | $F$ | $F^*$ | $F-F^*$ | $F$ | $F^*$ | $F-F^*$ | $F$ | $F^*$ | $F-F^*$ | $F$ | $F^*$ | $F-F^*$ |
| | 5 | 0.00575 | 0.00579 | −0.00004 | 0.00732 | 0.00741 | −0.00009 | 0.01643 | 0.01725 | −0.00082 | 0.03477 | 0.04076 | −0.00599 |
| | 10 | 0.00248 | 0.00244 | 0.00004 | 0.00342 | 0.00337 | 0.00005 | 0.00960 | 0.00960 | 0.00000 | 0.02367 | 0.02664 | −0.00297 |
| $-2\mathbf{1}$ | 20 | 0.00131 | 0.00129 | 0.00002 | 0.00194 | 0.00192 | 0.00003 | 0.00662 | 0.00651 | 0.00011 | 0.01836 | 0.02010 | −0.00174 |
| | 100 | 0.00065 | 0.00064 | 0.00000 | 0.00105 | 0.00105 | 0.00001 | 0.00453 | 0.00442 | 0.00011 | 0.01433 | 0.01529 | −0.00096 |
| | ∞ | 0.00052 | 0.00052 | 0.00000 | 0.00087 | 0.00087 | 0.00000 | 0.00405 | 0.00395 | 0.00010 | 0.01336 | 0.01416 | −0.00080 |
| | 5 | 0.03868 | 0.03883 | −0.00015 | 0.04579 | 0.04633 | −0.00054 | 0.08023 | 0.08448 | −0.00425 | 0.13643 | 0.15774 | −0.02131 |
| | 10 | 0.03196 | 0.03175 | 0.00021 | 0.03862 | 0.03851 | 0.00011 | 0.07155 | 0.07288 | −0.00134 | 0.12622 | 0.14304 | −0.01682 |
| $-1\mathbf{1}$ | 20 | 0.02856 | 0.02841 | 0.00016 | 0.03498 | 0.03483 | 0.00015 | 0.06707 | 0.06743 | −0.00036 | 0.12092 | 0.13535 | −0.01443 |
| | 100 | 0.02585 | 0.02581 | 0.00004 | 0.03205 | 0.03198 | 0.00007 | 0.06343 | 0.06322 | 0.00021 | 0.11659 | 0.12909 | −0.01250 |
| | ∞ | 0.02517 | 0.02517 | 0.00000 | 0.03132 | 0.03128 | 0.00004 | 0.06251 | 0.06219 | 0.00032 | 0.11549 | 0.12751 | −0.01202 |
| | 5 | 0.25000 | 0.25000 | 0.00000 | 0.26594 | 0.26837 | −0.00243 | 0.33333 | 0.34986 | −0.01653 | 0.42822 | 0.47611 | −0.04789 |
| | 10 | 0.25000 | 0.25000 | 0.00000 | 0.26594 | 0.26712 | −0.00118 | 0.33333 | 0.34413 | −0.01079 | 0.42822 | 0.47543 | −0.04721 |
| $0\mathbf{1}$ | 20 | 0.25000 | 0.25000 | 0.00000 | 0.26594 | 0.26653 | −0.00059 | 0.33333 | 0.34139 | −0.00805 | 0.42822 | 0.47522 | −0.04700 |
| | 100 | 0.25000 | 0.25000 | 0.00000 | 0.26594 | 0.26609 | −0.00015 | 0.33333 | 0.33926 | −0.00593 | 0.42822 | 0.47514 | −0.04692 |
| | ∞ | 0.25000 | 0.25000 | 0.00000 | 0.26594 | 0.26598 | −0.00004 | 0.33333 | 0.33874 | −0.00541 | 0.42822 | 0.47513 | −0.04692 |
| | 5 | 0.67546 | 0.68646 | −0.01100 | 0.68258 | 0.69452 | −0.01194 | 0.71701 | 0.74252 | −0.02551 | 0.77321 | 0.81223 | −0.03902 |
| | 10 | 0.69106 | 0.69707 | −0.00601 | 0.69773 | 0.70441 | −0.00668 | 0.73065 | 0.75215 | −0.02150 | 0.78533 | 0.82553 | −0.04020 |
| $1\mathbf{1}$ | 20 | 0.69931 | 0.70242 | −0.00311 | 0.70572 | 0.70947 | −0.00374 | 0.73781 | 0.75735 | −0.01953 | 0.79166 | 0.83255 | −0.04089 |
| | 100 | 0.70612 | 0.70676 | −0.00064 | 0.71233 | 0.71360 | −0.00128 | 0.74371 | 0.76172 | −0.01802 | 0.79686 | 0.83828 | −0.04141 |
| | ∞ | 0.70786 | 0.70786 | 0.00000 | 0.71401 | 0.71465 | −0.00064 | 0.74520 | 0.76285 | −0.01765 | 0.79818 | 0.83971 | −0.04153 |
| | 5 | 0.90381 | 0.91510 | −0.01129 | 0.90539 | 0.91639 | −0.01100 | 0.91449 | 0.93028 | −0.01578 | 0.93283 | 0.94821 | −0.01539 |
| | 10 | 0.92909 | 0.93443 | −0.00534 | 0.93003 | 0.93546 | −0.00543 | 0.93621 | 0.94860 | −0.01239 | 0.95028 | 0.96312 | −0.01284 |
| $2\mathbf{1}$ | 20 | 0.94205 | 0.94458 | −0.00253 | 0.94268 | 0.94550 | −0.00282 | 0.94735 | 0.95816 | −0.01081 | 0.95910 | 0.97033 | −0.01123 |
| | 100 | 0.95243 | 0.95291 | −0.00048 | 0.95284 | 0.95374 | −0.00090 | 0.95631 | 0.96592 | −0.00960 | 0.96612 | 0.97589 | −0.00977 |
| | ∞ | 0.95502 | 0.95502 | 0.00000 | 0.95537 | 0.95583 | −0.00046 | 0.95855 | 0.96786 | −0.00931 | 0.96786 | 0.97725 | −0.00939 |

Table A.2: Comparison between true cdf $F$ and approximates $F^*$ when $q = 4$

| | | $r$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{x}$ | $\nu$ | 0.0 | | | 0.1 | | | 0.5 | | | 0.9 | | |
| | | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ |
| **−21** | 5 | 0.00019 | 0.00017 | 0.00002 | 0.00050 | 0.00044 | 0.00006 | 0.00504 | 0.00494 | 0.00010 | 0.02497 | 0.03276 | −0.00780 |
| | 10 | 0.00003 | 0.00003 | 0.00000 | 0.00012 | 0.00010 | 0.00001 | 0.00236 | 0.00207 | 0.00029 | 0.01618 | 0.01903 | −0.00285 |
| | 20 | 0.00001 | 0.00001 | 0.00000 | 0.00003 | 0.00003 | 0.00000 | 0.00136 | 0.00115 | 0.00020 | 0.01214 | 0.01332 | −0.00118 |
| | 100 | 0.00000 | 0.00000 | 0.00000 | 0.00001 | 0.00001 | 0.00000 | 0.00076 | 0.00063 | 0.00013 | 0.00914 | 0.00945 | −0.00030 |
| | ∞ | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00063 | 0.00053 | 0.00010 | 0.00844 | 0.00859 | −0.00015 |
| **−11** | 5 | 0.00242 | 0.00246 | −0.00004 | 0.00523 | 0.00526 | −0.00003 | 0.03186 | 0.03306 | −0.00120 | 0.10542 | 0.13695 | −0.03153 |
| | 10 | 0.00141 | 0.00145 | −0.00003 | 0.00349 | 0.00346 | 0.00003 | 0.02666 | 0.02560 | 0.00106 | 0.09631 | 0.11865 | −0.02233 |
| | 20 | 0.00099 | 0.00102 | −0.00003 | 0.00270 | 0.00268 | 0.00002 | 0.02403 | 0.02240 | 0.00163 | 0.09157 | 0.10961 | −0.01803 |
| | 100 | 0.00070 | 0.00071 | −0.00001 | 0.00213 | 0.00210 | 0.00003 | 0.02194 | 0.02006 | 0.00188 | 0.08772 | 0.10261 | −0.01488 |
| | ∞ | 0.00063 | 0.00063 | 0.00000 | 0.00199 | 0.00196 | 0.00003 | 0.02142 | 0.01950 | 0.00192 | 0.08676 | 0.10090 | −0.01414 |
| **01** | 5 | 0.06250 | 0.06250 | 0.00000 | 0.08709 | 0.08864 | −0.00155 | 0.19998 | 0.21586 | −0.01588 | 0.36930 | 0.45423 | −0.08493 |
| | 10 | 0.06250 | 0.06250 | 0.00000 | 0.08709 | 0.08750 | −0.00042 | 0.20001 | 0.20760 | −0.00759 | 0.36930 | 0.45239 | −0.08308 |
| | 20 | 0.06250 | 0.06250 | 0.00000 | 0.08709 | 0.08703 | 0.00006 | 0.20000 | 0.20403 | −0.00404 | 0.36935 | 0.45234 | −0.08300 |
| | 100 | 0.06250 | 0.06250 | 0.00000 | 0.08709 | 0.08670 | 0.00038 | 0.20002 | 0.20147 | −0.00145 | 0.36933 | 0.45295 | −0.08362 |
| | ∞ | 0.06250 | 0.06250 | 0.00000 | 0.08709 | 0.08663 | 0.00046 | 0.20000 | 0.20088 | −0.00087 | 0.36929 | 0.45322 | −0.08393 |
| **11** | 5 | 0.47103 | 0.48039 | −0.00936 | 0.49624 | 0.50976 | −0.01352 | 0.59825 | 0.64662 | −0.04837 | 0.72962 | 0.80787 | −0.07825 |
| | 10 | 0.48541 | 0.49070 | −0.00529 | 0.51045 | 0.51827 | −0.00782 | 0.61196 | 0.65445 | −0.04249 | 0.74239 | 0.82283 | −0.08044 |
| | 20 | 0.49306 | 0.49581 | −0.00275 | 0.51800 | 0.52269 | −0.00469 | 0.61923 | 0.65950 | −0.04027 | 0.74898 | 0.83089 | −0.08191 |
| | 100 | 0.49944 | 0.50000 | −0.00056 | 0.52426 | 0.52644 | −0.00218 | 0.62519 | 0.66433 | −0.03914 | 0.75433 | 0.83744 | −0.08311 |
| | ∞ | 0.50107 | 0.50107 | 0.00000 | 0.52586 | 0.52742 | −0.00156 | 0.62668 | 0.66568 | −0.03900 | 0.75572 | 0.83906 | −0.08334 |
| **21** | 5 | 0.82706 | 0.85490 | −0.02784 | 0.83391 | 0.86075 | −0.02684 | 0.86678 | 0.90731 | −0.04054 | 0.91587 | 0.94785 | −0.03198 |
| | 10 | 0.86714 | 0.88068 | −0.01355 | 0.87173 | 0.88566 | −0.01393 | 0.89623 | 0.93013 | −0.03391 | 0.93606 | 0.96305 | −0.02699 |
| | 20 | 0.88905 | 0.89547 | −0.00642 | 0.89231 | 0.90003 | −0.00772 | 0.91194 | 0.94302 | −0.03108 | 0.94650 | 0.97032 | −0.02381 |
| | 100 | 0.90737 | 0.90857 | −0.00120 | 0.90961 | 0.91280 | −0.00319 | 0.92509 | 0.95409 | −0.02900 | 0.95502 | 0.97589 | −0.02087 |
| | ∞ | 0.91206 | 0.91206 | 0.00000 | 0.91403 | 0.91621 | −0.00217 | 0.92845 | 0.95696 | −0.02850 | 0.95713 | 0.97725 | −0.02012 |

Table A.3: Comparison between true cdf $F$ and approximates $F^*$ when $q = 5$

| | | $r$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{x}$ | $\nu$ | 0.0 | | | 0.1 | | | 0.5 | | | 0.9 | | |
| | | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ |
| **−21** | 5 | 0.00004 | 0.00004 | 0.00001 | 0.00017 | 0.00014 | 0.00004 | 0.00344 | 0.00316 | 0.00028 | 0.02257 | 0.03051 | −0.00795 |
| | 10 | 0.00000 | 0.00000 | 0.00000 | 0.00003 | 0.00002 | 0.00001 | 0.00151 | 0.00120 | 0.00030 | 0.01446 | 0.01690 | −0.00243 |
| | 20 | 0.00000 | 0.00000 | 0.00000 | 0.00001 | 0.00001 | 0.00000 | 0.00082 | 0.00063 | 0.00020 | 0.01076 | 0.01148 | −0.00072 |
| | 100 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00043 | 0.00032 | 0.00011 | 0.00802 | 0.00794 | 0.00007 |
| | ∞ | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00035 | 0.00026 | 0.00009 | 0.00738 | 0.00717 | 0.00020 |
| **−11** | 5 | 0.00068 | 0.00068 | 0.00000 | 0.00214 | 0.00209 | 0.00006 | 0.02341 | 0.02340 | 0.00001 | 0.09758 | 0.13058 | −0.03300 |
| | 10 | 0.00033 | 0.00034 | −0.00001 | 0.00128 | 0.00124 | 0.00003 | 0.01913 | 0.01738 | 0.00175 | 0.08878 | 0.11086 | −0.02208 |
| | 20 | 0.00020 | 0.00021 | −0.00001 | 0.00091 | 0.00090 | 0.00002 | 0.01703 | 0.01490 | 0.00212 | 0.08425 | 0.10140 | −0.01715 |
| | 100 | 0.00012 | 0.00012 | 0.00000 | 0.00066 | 0.00065 | 0.00002 | 0.01535 | 0.01313 | 0.00222 | 0.08053 | 0.09427 | −0.01373 |
| | ∞ | 0.00010 | 0.00010 | 0.00000 | 0.00061 | 0.00059 | 0.00002 | 0.01494 | 0.01272 | 0.00222 | 0.07962 | 0.09257 | −0.01295 |
| **01** | 5 | 0.03125 | 0.03125 | 0.00000 | 0.05286 | 0.05369 | −0.00083 | 0.16667 | 0.17871 | −0.01204 | 0.35273 | 0.44723 | −0.09450 |
| | 10 | 0.03125 | 0.03125 | 0.00000 | 0.05287 | 0.05280 | 0.00006 | 0.16667 | 0.17022 | −0.00355 | 0.35274 | 0.44454 | −0.09180 |
| | 20 | 0.03125 | 0.03125 | 0.00000 | 0.05286 | 0.05247 | 0.00040 | 0.16667 | 0.16672 | −0.00006 | 0.35275 | 0.44445 | −0.09170 |
| | 100 | 0.03125 | 0.03125 | 0.00000 | 0.05286 | 0.05226 | 0.00060 | 0.16667 | 0.16434 | 0.00233 | 0.35276 | 0.44537 | −0.09261 |
| | ∞ | 0.03125 | 0.03125 | 0.00000 | 0.05286 | 0.05222 | 0.00064 | 0.16665 | 0.16381 | 0.00284 | 0.35273 | 0.44578 | −0.09305 |
| **11** | 5 | 0.39761 | 0.40117 | −0.00357 | 0.43060 | 0.44042 | −0.00982 | 0.55874 | 0.61147 | −0.05274 | 0.71630 | 0.80676 | −0.09046 |
| | 10 | 0.40915 | 0.41121 | −0.00206 | 0.44271 | 0.44830 | −0.00559 | 0.57194 | 0.61832 | −0.04638 | 0.72904 | 0.82209 | −0.09305 |
| | 20 | 0.41525 | 0.41625 | −0.00100 | 0.44910 | 0.45255 | −0.00345 | 0.57891 | 0.62330 | −0.04439 | 0.73558 | 0.83044 | −0.09486 |
| | 100 | 0.42029 | 0.42047 | −0.00018 | 0.45436 | 0.45626 | −0.00190 | 0.58462 | 0.62843 | −0.04381 | 0.74105 | 0.83724 | −0.09619 |
| | ∞ | 0.42157 | 0.42157 | 0.00000 | 0.45570 | 0.45726 | −0.00156 | 0.58607 | 0.62992 | −0.04386 | 0.74239 | 0.83891 | −0.09652 |
| **21** | 5 | 0.79414 | 0.82771 | −0.03357 | 0.80404 | 0.83650 | −0.03246 | 0.84876 | 0.89913 | −0.05037 | 0.91024 | 0.94779 | −0.03755 |
| | 10 | 0.83903 | 0.85555 | −0.01651 | 0.84596 | 0.86314 | −0.01718 | 0.88054 | 0.92323 | −0.04268 | 0.93132 | 0.96304 | −0.03172 |
| | 20 | 0.86418 | 0.87206 | −0.00788 | 0.86931 | 0.87908 | −0.00977 | 0.89767 | 0.93728 | −0.03961 | 0.94233 | 0.97031 | −0.02799 |
| | 100 | 0.88573 | 0.88719 | −0.00145 | 0.88931 | 0.89373 | −0.00442 | 0.91216 | 0.94967 | −0.03751 | 0.95114 | 0.97589 | −0.02475 |
| | ∞ | 0.89131 | 0.89131 | 0.00000 | 0.89449 | 0.89773 | −0.00324 | 0.91585 | 0.95293 | −0.03708 | 0.95343 | 0.97725 | −0.02382 |

Table A.4: Comparison between true cdf $F$ and approximates $F^*$ when $q = 10$

| | | *r* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{x}$ | $\nu$ | 0.0 | | | 0.1 | | | 0.5 | | | 0.9 | | |
| | | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ | $F$ | $F^*$ | $F - F^*$ |
| | 5 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00110 | 0.00068 | 0.00042 | 0.01705 | 0.02444 | −0.00739 |
| | 10 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00039 | 0.00018 | 0.00021 | 0.01051 | 0.01104 | −0.00052 |
| **−21** | 20 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00018 | 0.00007 | 0.00010 | 0.00759 | 0.00661 | 0.00098 |
| | 100 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00007 | 0.00003 | 0.00004 | 0.00550 | 0.00415 | 0.00135 |
| | ∞ | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00006 | 0.00002 | 0.00003 | 0.00501 | 0.00367 | 0.00135 |
| | 5 | 0.00000 | 0.00000 | 0.00000 | 0.00007 | 0.00005 | 0.00002 | 0.00885 | 0.00677 | 0.00208 | 0.07783 | 0.11142 | −0.03359 |
| | 10 | 0.00000 | 0.00000 | 0.00000 | 0.00003 | 0.00002 | 0.00001 | 0.00676 | 0.00424 | 0.00251 | 0.07005 | 0.08620 | −0.01615 |
| **−11** | 20 | 0.00000 | 0.00000 | 0.00000 | 0.00001 | 0.00001 | 0.00000 | 0.00576 | 0.00336 | 0.00240 | 0.06603 | 0.07525 | −0.00922 |
| | 100 | 0.00000 | 0.00000 | 0.00000 | 0.00001 | 0.00001 | 0.00000 | 0.00498 | 0.00280 | 0.00218 | 0.06278 | 0.06793 | −0.00515 |
| | ∞ | 0.00000 | 0.00000 | 0.00000 | 0.00001 | 0.00000 | 0.00000 | 0.00479 | 0.00268 | 0.00210 | 0.06196 | 0.06638 | −0.00442 |
| | 5 | 0.00098 | 0.00098 | 0.00000 | 0.00659 | 0.00625 | 0.00034 | 0.09091 | 0.08662 | 0.00429 | 0.30749 | 0.42494 | −0.11745 |
| | 10 | 0.00098 | 0.00098 | 0.00000 | 0.00659 | 0.00602 | 0.00057 | 0.09092 | 0.07879 | 0.01213 | 0.30748 | 0.41704 | −0.10956 |
| **01** | 20 | 0.00098 | 0.00098 | 0.00000 | 0.00659 | 0.00597 | 0.00062 | 0.09092 | 0.07601 | 0.01490 | 0.30747 | 0.41577 | −0.10830 |
| | 100 | 0.00098 | 0.00098 | 0.00000 | 0.00659 | 0.00598 | 0.00061 | 0.09090 | 0.07458 | 0.01633 | 0.30744 | 0.41749 | −0.11005 |
| | ∞ | 0.00098 | 0.00098 | 0.00000 | 0.00659 | 0.00600 | 0.00059 | 0.09090 | 0.07438 | 0.01652 | 0.30748 | 0.41854 | −0.11107 |
| | 5 | 0.18661 | 0.16142 | 0.02519 | 0.23905 | 0.22630 | 0.01275 | 0.43970 | 0.48947 | −0.04977 | 0.67626 | 0.80413 | −0.12787 |
| | 10 | 0.18345 | 0.16850 | 0.01495 | 0.24035 | 0.23108 | 0.00927 | 0.44986 | 0.49085 | −0.04099 | 0.68897 | 0.82002 | −0.13106 |
| **11** | 20 | 0.18100 | 0.17259 | 0.00840 | 0.24057 | 0.23433 | 0.00624 | 0.45516 | 0.49465 | −0.03949 | 0.69566 | 0.82909 | −0.13343 |
| | 100 | 0.17845 | 0.17656 | 0.00189 | 0.24048 | 0.23789 | 0.00260 | 0.45948 | 0.50060 | −0.04112 | 0.70111 | 0.83663 | −0.13553 |
| | ∞ | 0.17772 | 0.17772 | 0.00000 | 0.24041 | 0.23900 | 0.00141 | 0.46056 | 0.50275 | −0.04219 | 0.70247 | 0.83849 | −0.13603 |
| | 5 | 0.66710 | 0.71159 | −0.04450 | 0.69146 | 0.73768 | −.04623 | 0.78613 | 0.87133 | −0.08521 | 0.89271 | 0.94771 | −0.05500 |
| | 10 | 0.72086 | 0.74375 | −0.02289 | 0.74117 | 0.76733 | −.02617 | 0.82329 | 0.89791 | −0.07463 | 0.91632 | 0.96302 | −0.04670 |
| **21** | 20 | 0.75440 | 0.76524 | −0.01083 | 0.77128 | 0.78754 | −.01626 | 0.84416 | 0.91538 | −0.07121 | 0.92849 | 0.97031 | −0.04182 |
| | 100 | 0.78581 | 0.78766 | −0.00185 | 0.79894 | 0.80885 | −.00991 | 0.86215 | 0.93260 | −0.07045 | 0.93863 | 0.97589 | −0.03726 |
| | ∞ | 0.79443 | 0.79443 | 0.00000 | 0.80646 | 0.81532 | −.00886 | 0.86691 | 0.93750 | −0.07059 | 0.94118 | 0.97725 | −0.03607 |

## Appendix B: Results of fit

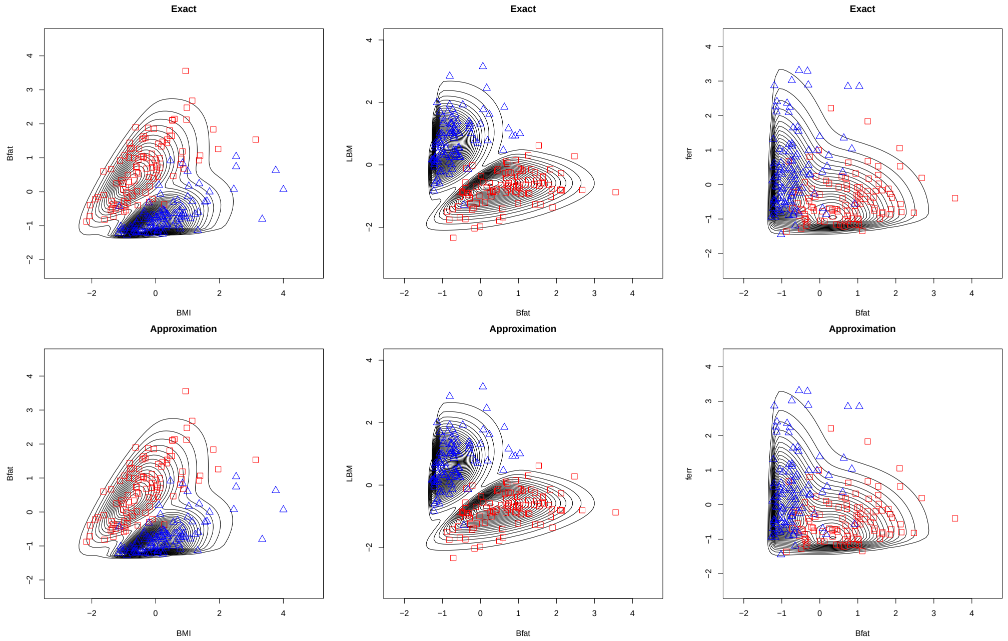

Figure B.1: *Results of fit for* (BMI, Bfat, LBM, ferr). *(a) 1st row: By Exact-EM, (b) 2nd row: By Approx-EM. Symbols:* □ *and* △ *indicate female and male, respectively.* $a = 0.9$ *for the initial estimates.*

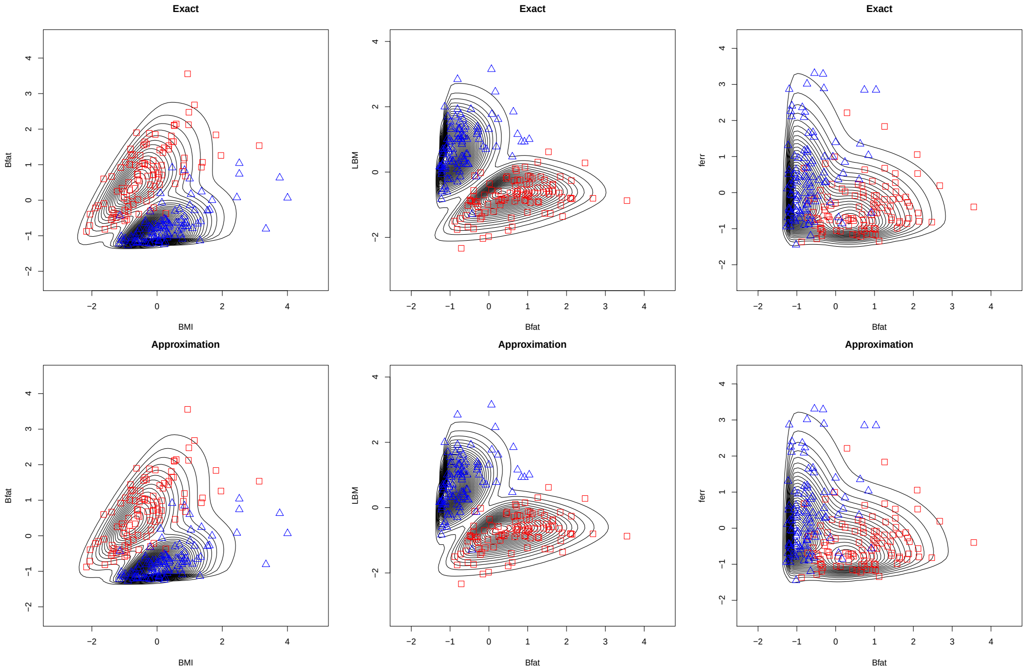

Figure B.2: *Results of fit for* (BMI, Bfat, LBM, ferr, hg). *(a) 1st row: By Exact-EM, (b) 2nd row: By Approx-EM. Symbols:* □ *and* △ *indicate female and male, respectively.* $a = 0.9$ *for the initial estimates.*

## References

Arellano-Valle RB and Genton MG (2005). On fundamental skew distributions, *Journal of Multivariate Analysis*, **96**, 93–116

Cook RD and Weisberg S (1994). *An Introduction to Regression Graphics*, John Wiley & Sons, New York.

Genz A and Bretz F (2009). Computation of multivariate normal and *t* probabilities, *Lecture Notes in Statistics*, **195**, Springer-Verlag, Heidelberg.

Ho HJ, Lin TI, Chen HY, and Wang WL (2012). Some results on the truncated multivariate *t* distribution, *Journal of Multivariate Analysis*, **96**, 93–116.

Kim SG (2016). An approximation fitting for mixture of multivariate skew normal distribution via EM algorithm, *Korean Journal of Applied Statistics*, **29**, 513–523.

Lee SX and McLachlan GJ (2016a). Finite mixtures of canonical fundamental skew *t*-distributions: The unification of the unrestricted and restricted skew *t*-mixture models, *Statistics and Computing*, **26**, 573–586.

Lee SX and McLachlan GJ (2016b). A simple parallel EM algorithm for statistical learning via mixture models. arXiv:1606.02054 [stat.CO] 7 Jun 2016.

Lin TI (2010). Robust mixture modelling using multivariate skew *t* distribution, *Statistics and Computing*, **20**, 343–356.

Lin TI, Wang WL, McLachlan GJ, and Lee SX (2018). Robust mixtures of factor analysis models using the restricted multivariate skew-t distribution, *Statistical Modelling*, **18**, 50–72.

Pyne S, Hu X, Wang K, *et al.* (2009). Automated high-dimensional flow cytometric data analysis. In

*Proceedings of the National Academy of Sciences of the United States of America*, **106**, 8519–8524.

Zogheib B and Elsaheli A (2015). Approximations of the *t* distribution, *Applied Mathematical Sciences*, **9**, 2445–2449.

*Received January 13, 2020; Revised February 4, 2020; Accepted February 12, 2020*