CARLETON UNIVERSITY

SCHOOL OF
MATHEMATICS AND STATISTICS

HONOURS PROJECT

TITLE: ON SKEW-NORMAL MODELS: PROPERTIES AND ESTIMATION VIA EM ALGORITHM

AUTHOR: MICHELE BARANZIKA

SUPERVISOR: DR. SONG CAI

DATE: AUGUST 27, 2018

Contents

1.1 Introduction

This paper deals with several aspects of the skew-normal distribution with a particular focus on parameter estimation by means of an EM-algorithm that will be developed in the paper.

The normal distribution is used extensively to fit to datasets that display a certain level of symmetry. When we have datasets that display more asymmetry, the skew-normal distribution becomes more suitable for that type of data. It has also been shown that the tails of a skew-normal distribution are always similar or thinner than those of a normal distribution. A good example of the use of the skew-normal distribution is in the financial industry; the presence of skewness in financial assets, insurance risks and claims is a strong motivation for using the skew-normal distribution since the normal distribution does not accurately describe this type of data.

One of the benefits of using the EM algorithm as an optimisation method is that it guarantees an increase in the likelihood function due to its iterative nature. However, a major disadvantage of the EM algorithm is the slow convergence to the MLE.

2.1 The basic skew-normal distribution

The skew normal distribution is a generalisation of the normal distribution that allows for skewness. The probability density function of a random variable X that follows a skew normal distribution with skewness parameter λ is given by

$$f(x) = 2\varphi(x)\Phi(\lambda x)$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are respectively the probability density function and the cumulative distribution function of a standard normal distribution. We say that $X \sim SN(\lambda)$.

2.2 Properties of the skew-normal distribution

1. **For $\boldsymbol{X \sim SN(\lambda)}$, we say that $\boldsymbol{Y \sim SN(\mu, \sigma^2, \lambda)}$ if $\boldsymbol{Y = \sigma X + \mu}$.**

   Here we account for the scale and location parameters by performing a linear transformation on the random variable X.

   Proof:

   Let $Y = g(X) = \sigma X + \mu$, where $\sigma > 0$ and μ are the scale and location parameters respectively.
   Then $g^{-1}(Y) = \frac{Y-\mu}{\sigma}$ and so the Jacobian of the transformation is $J = \left|\frac{dg^{-1}}{dy}\right| = \left|\frac{1}{\sigma}\right| = \frac{1}{\sigma}$ since σ is always nonnegative.
   Hence

$$f_Y(y) = f_X\big(g^{-1}(y)\big) * J = 2\varphi\big(g^{-1}(y)\big)\Phi\big(\lambda g^{-1}(y)\big) * \frac{1}{\sigma} = \frac{2}{\sigma}\varphi(\frac{y-\mu}{\sigma})\Phi(\lambda\frac{y-\mu}{\sigma})$$

Note that in the basic formulation of the skew-normal distribution the scale and location parameters are merely 1 and 0 respectively.

2. **If $\boldsymbol{\lambda = 0}$, then $\boldsymbol{X \sim SN(0) \equiv N(0,1)}$**
3. **If $\boldsymbol{X \sim SN(\lambda)}$, then $\boldsymbol{-X \sim SN(-\lambda)}$**
4. **If $\boldsymbol{X \sim SN(\mu, \sigma^2, \lambda)}$, then $\boldsymbol{M_X(t) = 2exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\Phi(\sigma\delta t)}$ where $\boldsymbol{\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}}$**

From the moment generating function, we can determine the first 2 moments about the origin of the skew-normal distribution and subsequently obtain the mean and variance of the skew-normal distribution.

<u>First moment:</u>

$$E(X) = M_X'(t)|_{t=0}$$

$$= 2\left[(\mu + \sigma^2 t)\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\Phi(\sigma\delta t) + \sigma\delta\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\varphi(\sigma\delta t)\right]\Bigg|_{t=0}$$

$$= 2\left(\frac{\mu}{2} + \sigma\delta\frac{1}{\sqrt{2\pi}}\right)$$

$$= \mu + \sigma\delta\sqrt{\frac{2}{\pi}}$$

<u>Second moment:</u>

$$E(X^2) = M_X''(t)|_{t=0}$$

$$= 2\left[\sigma^2\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\Phi(\sigma\delta t) + (\mu + \sigma^2 t)^2\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\Phi(\sigma\delta t)\right.$$

$$\left. + 2\sigma\delta(\mu + \sigma^2 t)\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\varphi(\sigma\delta t) - 2(\sigma\delta)^2 t\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\varphi(\sigma\delta t)\right]\Bigg|_{t=0}$$

$$= 2\left(\frac{\sigma^2}{2} + \frac{\mu^2}{2} + \frac{2\mu\sigma\delta}{\sqrt{2\pi}}\right)$$

$$= \sigma^2 + \mu^2 + 2\mu\sigma\delta\sqrt{\frac{2}{\pi}}$$

From the above derivations, we can see that

$$Mean = E(X) = \mu + \sigma\delta\sqrt{\frac{2}{\pi}}$$

and

$$Variance = E(X^2) - [E(X)]^2 = \sigma^2 + \mu^2 + 2\mu\sigma\delta\sqrt{\frac{2}{\pi}} - \left(\mu + \sigma\delta\sqrt{\frac{2}{\pi}}\right)^2$$

$$\therefore Variance = \sigma^2\left(1 - \frac{2\delta^2}{\pi}\right)$$

2.3 Stochastic representation of the skew-normal distribution

Consider 2 independent standard normal variables V and U. Now introduce a new variable W, which is correlated to the variable V, defined by the relationship $W = \frac{\lambda V - U}{\sqrt{1+\lambda^2}}$ such that the correlation between V and W is a function δ of λ, i.e, $cor(V, W) = \delta(\lambda)$.

Truncating W below 0 gives us that $(V|W > 0) \sim SN(\lambda)$. So, by property 1 defined above, if we define a random variable X such that $X = \sigma V + \mu$, then $(X|W > 0) \sim SN(\mu, \sigma^2, \lambda)$.

A bi-product of this is the bivariate normal variable (V, W) with standardised marginals and correlation δ, whose pdf is given by

$$f(w, v) = 2\varphi_B(w, v; \delta) = \frac{1}{\pi(1-\delta^2)} \exp\left[-\frac{1}{2(1-\delta^2)}(w^2 - 2\delta wv + v^2)\right]$$

This representation of the skew-normal distribution will allow us to develop an EM algorithm to estimate the parameters μ, σ^2, and λ.

Another representation of the skew-normal distribution is based on the two independent variables V and U, as introduced above. If we derive a new random variable Z such that

$$Z = \sqrt{1-\delta^2}U + \delta|V|$$

Then we say that $Z \sim SN(\lambda)$, where δ is an arbitrary value between -1 and 1, and is a function of λ.

3.1 EM algorithm for parameter estimation

The expectation maximisation (EM) algorithm is an iterative method of finding the maximum likelihood estimates of parameters that depends on unobserved variables. In this method, iterations are performed until convergence is achieved.

The EM algorithm is oftentimes used because the pdf of the distribution whose parameters are to be estimated is very intricate and without a simple expression.
A common application of the EM algorithm is in the estimation of mixture models.

The algorithm

As the name suggests, the EM algorithm is performed in two main steps namely the expectation step (E-step) and the maximisation step (M-step) which are performed during each iteration.

The E-step consists of computing the expected value of the complete data log-likelihood conditional on the observed data, with respect to the current parameter estimate. This expectation is referred to as the Q-function.

The M-step in turn consists of maximising the above Q-function with respect to the parameter we are estimating, in order to get the MLE.

Once the M-step is done, we use a convergence rule of choice to determine whether to continue the iterative process or to stop and accept the most recent MLE.

Setup:

Let $X = \{X_1, X_2, \dots, X_n\}$ represent the observed data and $Y = \{Y_1, Y_2, \dots, Y_m\}$ represent the unobserved data so that the pair (X, Y) represents the complete data that depends on a parameter vector θ, i.e, $f(x, y; \theta)$ is the joint pdf of the complete data. Our goal is to find the MLE $\hat{\theta}$ of θ.
Let the conditional pdf of the unobserved data Y given the observed data $X = x$ be given by $g(y|x; \theta)$.

The complete likelihood function is given by

$$L^C(\theta) = \prod_i f(x_i, y_i; \theta)$$

Therefore, the complete log-likelihood function is given by

$$\log L^C(\theta) = \sum_i \log f(x_i, y_i; \theta)$$

E-step: Given a previous estimate $\hat{\theta}^{(k)}$ of θ, the Q-function is determined as

$$Q_{\hat{\theta}^{(k)}}(\theta) = E_{\hat{\theta}^{(k)}}[\log L^C(\theta)|X] = \int \log L^C(\theta)\, g(y|x; \hat{\theta}^{(k)})dy$$

Note that it may be useful to determine the conditional first and second moment about the origin of the unobserved data Y in order to get a simplified expression for the Q-function.

M-step: We maximise the Q-function by solving

$$\frac{\partial Q_{\hat{\theta}^{(k)}}(\theta)}{\partial \theta} = 0$$

$$\Rightarrow \qquad \hat{\theta}^{(k+1)} = \underset{\theta}{\operatorname{argmax}}\, Q_{\hat{\theta}^{(k)}}(\theta)$$

<u>Convergence rule:</u> For illustration purposes, we will use a relative reduction in $\|\theta\|$ to determine convergence. We choose a small positive real number ξ, say $\xi = 10^{-10}$. We stop the iterative process when the following test is satisfied

$$\frac{\left\|\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}\right\|}{\left\|\hat{\theta}^{(k)}\right\|} < \xi$$

where $\|\cdot\|$ is the Euclidean norm.

3.2 EM algorithm for the skew-normal case

We will use the stochastic representation developed in section 2.3 to derive an EM algorithm for the skew-normal distribution to estimate $\theta = \{\mu, \sigma^2, \lambda\}$

We will let $X = \sigma V + \mu$ be the observed data and $Y = \sigma W$ be the unobserved data such that the pair (X, Y) forms the complete data.

Recall that the pair (V, W) forms a bivariate standard normal distribution truncated below $W = 0$, and since the pair (X, Y) is a linear transformation of (V, W), we can easily compute the complete likelihood function.

First, we need to find the Jacobian of the transformation

$X = \sigma V + \mu$

$\Longrightarrow V = \dfrac{X - \mu}{\sigma}$

$Y = \sigma W$

$$\Longrightarrow W = \frac{Y}{\sigma}$$

$$J = \begin{vmatrix} \frac{\partial V}{\partial X} & \frac{\partial V}{\partial Y} \\ \frac{\partial W}{\partial X} & \frac{\partial W}{\partial Y} \end{vmatrix} = \begin{vmatrix} \frac{1}{\sigma} & 0 \\ 0 & \frac{1}{\sigma} \end{vmatrix} = \left|\frac{1}{\sigma^2}\right| = \frac{1}{\sigma^2}$$

So, the joint distribution of the complete data is given by

$$f_{X,Y}(x, y; \theta) = \frac{1}{\sigma^2} f_{W,V}\left(\frac{x-\mu}{\sigma}, \frac{y}{\sigma}; \delta\right) = \frac{2}{\sigma^2} \varphi_B\left(\frac{x-\mu}{\sigma}, \frac{y}{\sigma}; \delta\right)$$

$$= \frac{1}{\sigma^2 \pi (1-\delta^2)} \exp\left[-\frac{1}{2(1-\delta^2)}\left(\frac{y^2}{\sigma^2} - \frac{2\delta y(x-\mu)}{\sigma^2} + \frac{(x-\mu)^2}{\sigma^2}\right)\right]$$

Hence the complete log-likelihood function is given by

$$\log L^C(\theta) = \log\left[\prod_i f(x_i, y_i; \theta)\right] = \sum_i \log f(x_i, y_i; \theta)$$

$$\therefore \log L^C(\theta) = -2\log\sigma - \log\pi - \log(1-\delta^2) - \sum_i \left[\frac{{y_i}^2 - 2\delta y_i(x_i-\mu) + (x_i-\mu)^2}{2(1-\delta^2)\sigma^2}\right]$$

It can be shown that the conditional pdf of W given V is

$$f(w|v) = \frac{f(w, v)}{\varphi(v)\Phi(\lambda v)}, \qquad w > 0$$

So that

$$E[W|v] = \delta v + \frac{\varphi(\lambda v)}{\Phi(\lambda v)}\sqrt{1-\delta^2}$$

$$\Longrightarrow\ E[Y|x] = \ E[Y = \sigma W|x = \sigma v + \mu] = \sigma E\left[W|v = \frac{x-\mu}{\sigma}\right]$$

$$= \sigma\left[\delta\frac{x-\mu}{\sigma} + \frac{\varphi\left(\lambda\frac{x-\mu}{\sigma}\right)}{\Phi\left(\lambda\frac{x-\mu}{\sigma}\right)}\sqrt{1-\delta^2}\right] = \delta(x-\mu) + \sigma\frac{\varphi\left(\lambda\frac{x-\mu}{\sigma}\right)}{\Phi\left(\lambda\frac{x-\mu}{\sigma}\right)}\sqrt{1-\delta^2}$$

$$E[W^2|v] = 1 - \delta^2 + (\delta v)^2 + \frac{\varphi(\lambda v)}{\Phi(\lambda v)}\delta\sqrt{1-\delta^2}v$$

$$\Longrightarrow\ E[Y^2|x] = E[Y^2 = E[\sigma^2 W^2|x = \sigma v + \mu] = \sigma^2\ E\left[W^2|v = \frac{x-\mu}{\sigma}\right]$$

$$= \sigma^2\ \left[1 - \delta^2 + \left(\delta\frac{x-\mu}{\sigma}\right)^2 + \frac{\varphi\left(\lambda\frac{x-\mu}{\sigma}\right)}{\Phi\left(\lambda\frac{x-\mu}{\sigma}\right)}\delta\sqrt{1-\delta^2}\frac{x-\mu}{\sigma}\right]$$

For simplicity, let $m_1 = \ E[Y|x]$ and $m_2 = \ E[Y^2|x]$

<u>E-step:</u>

For a predetermined initial value of the parameter we are estimating, say $\hat{\theta}^{(k)}$, the Q-function is given by

$$Q_{\hat{\theta}^{(k)}}(\theta) = \int \log L^C(\theta\,)\, f\left(y_i\middle|x_i;\hat{\theta}^{(k)}\right)dy$$

$$= \int\left\{-2\log\sigma - \log\pi - \log(1-\delta^2) - \sum_i\left[\frac{{y_i}^2 - 2\delta y_i(x_i-\mu) + (x_i-\mu)^2}{2(1-\delta^2)\sigma^2}\right]\right\} f\left(y_i\middle|x_i;\hat{\theta}^{(k)}\right)dy$$

$$= n[-2\log\sigma - \log\pi - \log(1-\delta^2)] - \frac{1}{2(1-\delta^2)\sigma^2}\sum_i\left[\widehat{m}_2^{(k)} - 2\delta(x_i-\mu)\widehat{m}_1^{(k)} + (x_i-\mu)^2\right]$$

Where

$$\widehat{m}_1^{(k)} = \hat{\delta}^{(k)}\left(x_i - \hat{\mu}^{(k)}\right) + \hat{\sigma}^{(k)} \frac{\varphi\left(\hat{\lambda}^{(k)} \frac{x_i - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right)}{\Phi\left(\hat{\lambda}^{(k)} \frac{x_i - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right)} \sqrt{1 - \hat{\delta}^{(k)^2}}$$

$$\widehat{m}_2^{(k)} = \hat{\sigma}^{(k)^2}\left[1 - \hat{\delta}^{(k)^2} + \left[\hat{\delta}^{(k)} \frac{x_i - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right]^2 + \frac{\varphi\left(\hat{\lambda}^{(k)} \frac{x_i - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right)}{\Phi\left(\hat{\lambda}^{(k)} \frac{x_i - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right)} \hat{\delta}^{(k)} \sqrt{1 - \hat{\delta}^{(k)^2}} \left(\frac{x_i - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right)\right]$$

<u>M-step:</u>

We now maximise $Q_{\hat{\theta}^{(k)}}(\theta)$ with respect to θ to get the MLE $\hat{\theta}^{(k+1)}$.

By solving

$$\frac{\partial Q_{\hat{\theta}^{(k)}}(\theta)}{\partial \theta} = 0$$

We obtain the following MLEs for each component of the parameter vector θ

$$\hat{\mu}^{(k+1)} = \frac{1}{n} \sum_i x_i - \frac{\hat{\delta}^{(k+1)}}{n} \sum_i \widehat{m}_1^{(k)}$$

$$\hat{\sigma}^{(k+1)^2} = \frac{1}{2n\left(1 - \hat{\delta}^{(k+1)^2}\right)} \sum_i \left[\left(x_i - \hat{\mu}^{(k+1)}\right)^2 - 2\hat{\delta}^{(k+1)}\left(x_i - \hat{\mu}^{(k+1)}\right)\widehat{m}_1^{(k)} + \widehat{m}_2^{(k)}\right]$$

$$\hat{\lambda}^{(k+1)} = \frac{\hat{\delta}^{(k+1)}}{\sqrt{1 - \hat{\delta}^{(k+1)^2}}}$$

Where

$$\hat{\delta}^{(k+1)} = \frac{\sqrt{1+4c^2} - 1}{2c}$$

And

$$c = 2\left\{\sum_i \left[\left(x_i - \hat{\mu}^{(k+1)}\right)^2 - 2\hat{\delta}^{(k+1)}\left(x_i - \hat{\mu}^{(k+1)}\right)\widehat{m}_1^{(k)} + \widehat{m}_2^{(k)}\right]\right\}^{-1} \sum_i \left(x_i - \hat{\mu}^{(k+1)}\right)\widehat{m}_1^{(k)}$$

These estimates represent $\hat{\theta}^{(k+1)} = \left\{\hat{\mu}^{(k+1)}, \hat{\sigma}^{(k+1)^2}, \hat{\lambda}^{(k+1)}\right\}$. Note that the estimate $\hat{\theta}^{(k+1)}$ has to be found by solving the above equations. Fortunately, this can be done easily with a fixed-point iteration nested within each EM iteration.

Using this MLE $\hat{\theta}^{(k+1)}$ together with an initial value of $\hat{\theta}^{(0)}$, the EM iterations are performed until a convergence is observed based on the chosen convergence rule.

3.3 EM algorithm pseudo code

- Randomly generate a sample of size n from a $SN\sim(\mu_r, \sigma_r^2, \lambda_r)$ distribution.
- Set the initial value $\hat{\theta}^{(0)} = \left\{\hat{\mu}^{(0)}, \hat{\sigma}^{(0)}, \hat{\lambda}^{(0)}\right\}$ to start the iteration and ξ to perform the convergence test. Set iteration counter $k = 0$.
- Compute $\hat{\theta}^{(k+1)} = \left\{\hat{\mu}^{(k+1)}, \hat{\sigma}^{(k+1)}, \hat{\lambda}^{(k+1)}\right\}$ by solving the equations in Section 3.2 using fixed-point iteration.

- Compute $t = \frac{\|\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}\|}{\|\hat{\theta}^{(k+1)}\|} = \frac{\sqrt{\left(\hat{\mu}^{(k+1)} - \hat{\mu}^{(k)}\right)^2 + \left(\hat{\sigma}^{(k+1)} - \hat{\sigma}^{(k)}\right)^2 + \left(\hat{\lambda}^{(k+1)} - \hat{\lambda}^{(k)}\right)^2}}{\sqrt{\hat{\mu}^{(k+1)^2} + \hat{\sigma}^{(k+1)^2} + \hat{\lambda}^{(k+1)^2}}}$
- If $t > \xi$, continue iteration.
- Otherwise, stop iteration and chose $\hat{\theta}^{(k+1)}$ as the MLE of the skew-normal distribution.

4.1 Simulation study

We will conduct our simulation in R, using the “sn” package. With a sample size of n = 200, we will generate numbers from a $SN \sim (0,4,4)$ and run the simulation N = 2000 times.
Once we have estimated N MLEs, we will then find the bias and mean squared error (MSE) of each parameter in order to measure their “quality”.

A summary of the simulation can be found below:

Simulation i	$\hat{\mu}_i$	$\hat{\sigma}_i$	$\hat{\lambda}_i$
1	0.10706	2.0323	4.2062
2	-0.016431	1.9709	3.9808
3	-0.029879	1.9991	4.6441
…	…	…	…
999	0.22855	1.8734	3.3605
1000	0.060500	1.8546	3.6052
1001	-0.063491	1.9391	4.0931

…	…	…	…
1998	-0.022415	1.9174	4.0675
1999	0.28566	1.8229	2.6940
2000	0.053638	1.9443	3.1390

Next, we summarise the bias and mean squared error of each estimated parameter

$\hat{\theta}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$
$E(\hat{\theta})$	0.0089270	1.9924	4.3134
$var(\hat{\theta})$	0.016335	0.019633	1.7910
$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$	0.0089270	-0.0076023	0.31344
$MSE(\hat{\theta}) = var(\hat{\theta}) + [Bias(\hat{\theta})]^2$	0.016415	0.019691	1.8893

<u>Conclusion:</u>

Based on the summarised results above, it is safe to conclude that the MLEs are unbiased and have low MSEs. The estimation accuracy for parameter λ is relatively low but still reasonable. This makes them good candidates to estimate the parameters of the skew-normal distribution.

References

1. Adcock, C.J., Eling, M. and Loperfido, N. (2012) *Skewed distributions in finance and actuarial science*, Working papers on risk management and insurance no.122

2. Azzalini, A. and Capitanio, A. (2013) *The Skew-Normal and related families*, Cambridge University Press, Cambridge, UK

3. Azzalini, A. and Dalla Valle, A. (1996) *The multivariate skew-normal distribution*, Biometrika, London, UK

Appendix

R code:

```
library("sn")
n <- 200  # sample size
kSim <- 2000  # number of simulation repetitions

# relative error precision for testing convergence of EM iteration
err_prec <- 1e-10

# relative error precision for testing convergence of inner
# fixed point iteration (within each EM iteration).
err_prec_inner <- 1e-12

# true parameter values
mu0 <- 0
sig0 <- 2
lam0 <- 4
#del0 <- lam0/sqrt(1+lam0^2)

mu_sim <- numeric(kSim)
sig_sim <- numeric(kSim)
lam_sim <- numeric(kSim)
nstep_sim <- numeric(kSim)  # number of EM iterations for each simulation iteration
```

```
# Function for fixed point iteration inside EM iteration
inner_f <- function(par, x, m1, m2) {
  mu <- par[1]
  sig <- par[2]
  del <- par[3]

  mu1 <- mean(x) - mean(m1) * del
  v_tmp <- mean( (x-mu)^2 - 2*del*(x-mu)*m1 + m2 )
  sig1 <- sqrt( v_tmp / 2 / (1-del^2) )
  c_tmp <- 2 * mean((x-mu) * m1)/v_tmp
  del1 <- ( sqrt(1+4*c_tmp^2) - 1 )/(2*c_tmp)

  return(c(mu1, sig1, del1))
}

for (iSim in 1:kSim) { # main loop
  x <- rsn(n, xi=mu0, omega=sig0, alpha=lam0)

  # Initial values of EM: taken to be around true parameter
  # values for convenience
  mu_pre <- mu0 + rnorm(1, 0, 0.1)
  sig_pre <- sig0 + 0.1
  lam_pre <- lam0 + rnorm(1, 0, 0.1)
```

```
  del_pre <- lam_pre/sqrt(1+lam_pre^2)

  nstep <- 0  # initial value of EM iteration counter

  rel_err <- 10  # initialize value of relative error; take this intitial value to be greater than
err_prec.
  while(rel_err > err_prec) {

    z_pre <- (x-mu_pre)/sig_pre
    m1_pre <- del_pre * (x-mu_pre) + sig_pre *
      sqrt(1-del_pre^2) * dnorm(lam_pre*z_pre, 0, 1) /
      pnorm(lam_pre*z_pre, 0, 1)
    m2_pre <- sig_pre^2 * ( 1 - del_pre^2 + del_pre^2*z_pre^2
      + del_pre * sqrt(1-del_pre^2) * dnorm(lam_pre*z_pre, 0, 1)
      / pnorm(lam_pre*z_pre, 0, 1) * z_pre )

    par0 <- c(mu_pre, sig_pre, del_pre)
    rel_err_inner <- 10  # initialize value of relative error for inner loop; take this intitial value to
be greater than err_prec_inner.
    while (rel_err_inner > err_prec_inner) {
      par1 <- inner_f(par0, x, m1_pre, m2_pre)

      rel_err_inner <- norm(cbind(par1 - par0),
           type="F")/norm(cbind(par1), type="F")
```

```
    par0 <- par1

  }

  par_pre <- c(mu_pre, sig_pre, lam_pre)
  par_cur <- par1
  #lam_cur <-  del_cur / sqrt(1-del_cur^2)
  par_cur[3] <- par1[3] / sqrt(1-par1[3]^2)

  rel_err <- norm(cbind(par_cur - par_pre),
      type="F")/norm(cbind(par_cur), type="F")

  # count the number of EM iterations
  nstep <- nstep + 1

  # refresh the stored value at m step.
  mu_pre <- par1[1]
  sig_pre <- par1[2]
  del_pre <- par1[3]
  lam_pre <- par_cur[3]

}
```

```
  # Store results
  mu_sim[iSim] <- par_cur[1]
  sig_sim[iSim] <- par_cur[2]
  lam_sim[iSim] <- par_cur[3]
  nstep_sim[iSim] <- nstep  ## number of EM iterations for each simulation iteration

}

save(mu_sim, sig_sim, lam_sim, nstep_sim, file="results.RData")

# bias
mean(mu_sim) - mu0
mean(sig_sim) - sig0
mean(lam_sim) - lam0

# variance
mean((mu_sim-mean(mu_sim))^2)
mean((sig_sim-mean(sig_sim))^2)
mean((lam_sim-mean(lam_sim))^2)

# MSE
mean((mu_sim-mu0)^2)
mean((sig_sim-sig0)^2)
mean((sig_sim-sig0)^2)
```