Symmetric Geometric Skew Normal Regression Model

Debasis Kundu [*]and Deepak Prajapati [†]

Abstract

Recently, Kundu (2014, Sankhya, Ser. B, 167–189, 2014) proposed a geometric skew normal distribution as an alternative to Azzalini's skew normal distribution. The geometric skew normal distribution can be a skewed distribution, it can be heavy tailed as well as multimodal also, unlike Azzalini's skew normal distribution. It can be easily extended to the multivariate case also. The multivariate geometric skew normal distribution also has several desirable properties. In this paper we have proposed a symmetric geometric skew normal distribution as an alternative to a symmetric distribution like normal distribution, log Birnbaum-Saunders distribution, Student-t distribution etc. It is a very flexible class of distributions, of which normal distribution is a special case. The proposed model has three unknown parameters, and it is observed that the maximum likelihood estimators of the unknown parameters cannot be obtained in explicit forms. In this paper we have proposed a very efficient expectation maximization (EM) algorithm, and it is observed that the proposed EM algorithm works very well. We have further considered a location shift symmetric geometric skew normal regression model. It is a more flexible than the standard Gaussian regression model. The maximum likelihood estimators of the unknown parameters are obtained based on expectation maximization algorithm. Extensive simulation experiments and the analyses of two data sets have been presented to show the effectiveness of the proposed model and the estimation techniques.

Key Words and Phrases: Absolute continuous distribution; singular distribution; Fisher information matrix; EM algorithm; joint probability distribution function; joint probability density function.

[*]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: kundu@iitk.ac.in, Phone no. 91-512-2597141, Fax no. 91-512-2597500, Corresponding author.

[†]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: kundu@iitk.ac.in, Phone no. 91-512-2597141, Fax no. 91-512-2597500.

1 Introduction

Azzalini [3] introduced a skew normal distribution which has received considerable attention in the last three decades. It has three parameters, and it is a skewed distribution of which normal distribution is a special case. Form now on we call it as the Azzalini's skew normal (ASN) distribution. A three-parameter ASN distribution has the following probability density function (PDF):

$$f(x) = 2\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\frac{\lambda(x-\mu)}{\sigma}\right).$$

Here $\phi(\cdot)$ is the PDF of a standard normal distribution and the $\Phi(\cdot)$ is the corresponding cumulative distribution function (CDF), $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter and $-\infty < \lambda < \infty$ is the skewness or tilt parameter. Note that when $\lambda = 0$, it becomes a normal PDF with mean μ and standard deviation σ. It has some interesting physical interpretation also as a hidden truncation model, see for example Arnold et al. [2] and Arnold and Beaver [1]. It is a very flexible class of distribution functions, and due to which it has been used to analyze various skewed data sets. Although, it has several desirable properties, it has been observed that the maximum likelihood estimators of the unknown parameters may not always exist, see for example Gupta and Gupta [10]. It can be shown that for any sample size if the data come from a ASN distribution, there is a positive probability that the MLEs do not exist. The problem becomes more severe for higher dimension. Moreover, the probability density function (PDF) of a ASN distribution is always unimodal and thin tailed. Due to these limitations, ASN distribution cannot be used for analyzing for a large class of skewed data sets.

To overcome the problem of the ASN distribution, Gupta and Gupta [10] proposed the power normal distribution which has the following CDF

$$F(x) = \left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha}.$$

Here also, $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter and $\alpha > 0$ is the skewness parameter. When $\alpha = 1$, it becomes a normal distribution function with mean μ and standard deviation σ. Therefore, in this case also the normal distribution can be obtained as a special case. It has been shown by Gupta and Gupta [10] that for $\alpha > 1$ it is positively skewed and for $\alpha < 1$, it is negatively skewed. They have obtained several other interesting properties of the power normal distribution in the same paper. Kundu and Gupta [18] provided an efficient estimation procedure and also defined bivariate power normal distribution. Although, the power normal distribution as proposed by Gupta and Gupta [10] has several desirable properties, and the maximum likelihood estimators also always exist, it is always unimodal and thin tailed similar to the ASN distribution. Therefore, if the data indicate that the observations are coming from a heavy tailed or multimodal distribution, it can be used to analyze that data set.

Recently, Kundu [14] introduced a three-parameter geometric skew normal (GSN) distribution as an alternative to the popular Azzalini's [3] skew normal (ASN) distribution or the power normal distribution of Gupta and Gupta [10]. The GSN distribution can be defined as follows. Suppose $X_1, X_2, \ldots,$ are independent and identically distributed (i.i.d.) normal random variables with mean μ, variance σ^2 and N is a geometric random variable with parameter p. Here, a geometric random variable with parameter p will be denoted by GE(p), and it has the following probability mass function:

$$P(N = n) = p(1-p)^{n-1}; \qquad n = 1, 2, \ldots. \tag{1}$$

It is assumed that N and X_i's are independently distributed. The random variable

$$X \stackrel{d}{=} \sum_{i=1}^{N} X_i,$$

is said to have a GSN distribution with the parameters μ, σ and p. Form now on it will be denoted by GSN(μ, σ, p).

The GSN distribution also has three parameters similar to the ASN or power normal distribution. But the main advantage of the GSN distribution over the ASN or the power normal distribution is that the GSN distribution is more flexible than them in the sense its PDF can have more variety of shapes compared to the ASN distribution. The PDF of the GSN distribution can be symmetric, skewed, unimodal, bimodal, multimodal shaped also. Moreover, the GSN distribution can be heavy tailed also depending on the parameter values. In case of an ASN distribution, it is observed that the maximum likelihood (ML) estimators may not always exist. In fact it can be shown that the ML estimates will not exist if the all the data points are of the same sign. But in case of the GSN distribution, the ML estimates of the unknown parameters exist, if the sample size is greater than three. If $\mu = 0$, it becomes a symmetric distribution and it is called a symmetric GSN (SGSN) distribution. From now on a GSN(μ, σ, p) with $\mu = 0$ will be denoted by SGSN(σ, p).

As GSN has been introduced, along the same line multivariate geometric skew normal (MGSN) distribution has been introduced by Kundu [17] as an alternative to Azzalini's multivariate skew normal distribution. In this case the marginals are GSN distributions, and the joint PDF can be unimodal, bimodal and multimodal also. Multivariate normal distribution can be obtained as a special case. It has also several interesting properties, and the moments and product moments can be obtained quite conveniently from the joint characteristic function, which can be expressed in explicit form. It has several characterization properties similar to the multivariate normal distribution. A brief review of the MGSN distribution will be provided in Section 3.

In recent time Birnbaum-Saunders (BS) distribution has received a considerable amount of attention in the statistical and some related literature. The BS distribution was originally derived by Birnbaum and Saunders [7] by showing that the fatigue failure is caused by the development and growth of cracks from the cyclic loading. The BS distribution can be defined as follows. Suppose T is a non-negative random variable, T is said to have a BS distribution with shape parameter α and scale parameter β, it will be denoted by BS(α, β), if the cumulative distribution function (CDF) of T is given by

$$F_{BS}(t;\alpha,\beta) = \Phi\left\{\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right\}, \qquad t > 0, \tag{2}$$

and zero, otherwise. Here $\Phi(\cdot)$ is the CDF of a standard normal distribution. If T follows ($\sim$), BS(α, β), then $Y = \ln T$ is said to have a log-BS (LBS) distribution. The CDF of Y can be written

$$F_{LBS}(y;\alpha,\beta) = \Phi\left\{\frac{2}{\alpha}\sinh\left(\frac{y - \ln\beta}{2}\right)\right\}, \qquad -\infty < y < \infty, \tag{3}$$

and it will be denoted by LBS(α, β). The LBS distribution was originally proposed by Rieck [25], see also Rieck and Nedelman [26] and Kundu [15, 16] in this respect. It has been observed by Rieck [25] that the LBS distribution is symmetric, it is strongly unimodal for $\alpha < 2$ and for $\alpha \geq 2$, it is bimodal. It may be mentioned that SGSN distribution is more flexible than the LBS distribution in the sense, SGSN distribution can be heavy tailed also, where as LBS distribution can never be a heavy tailed. For a comprehensive review on BS distribution one is referred to the review article by Balakrishnan and Kundu [6]. Although, LBS distribution can be bimodal and the maximum likelihood estimators of the unknown parameters always exist, it cannot be multimodal or heavy tailed.

In this paper we have introduced a location-shift SGSN (LS-SGSN) distribution. It is observed that due to presence of three parameters it is more flexible than a two-parameter normal distribution. Moreover, the normal distribution can be obtained as a special case of the LS-SGSN distribution. We have provided several properties of the LS-SGSN distribution, obtain the characteristic function and different moments. It is interesting to observe that although the PDF of a SGSN distribution can be obtained as an infinite series, the characteristic function can be obtained in explicit form. The LS-SGSN distribution has three parameters. The ML estimators cannot be obtained in explicit forms, they have to be obtained by solving a three dimensional optimization problem. Some standard numerical methods like Newton-Raphson or Gauss-Newton may be used to compute the ML estimates. But it involves providing efficient initial guesses, otherwise, the algorithm may not converge, and even if it converges it may converge to a local minimum rather than a global minimum. To avoid that we have treated this problem as a missing value problem, and we have used the expectation maximization (EM) algorithm to compute the ML estimators. It is observed that at each 'E'-step, the corresponding 'M'-step can be performed explicitly. Hence, no optimization problem needs to be solved numerically at the 'M'-step and they are unique. Moreover, at the last step of the EM algorithm one can easily obtain the observed Fisher information matrix based on the method of Louis [22]. Hence, the confidence intervals of the unknown parameters also can be obtained quite conveniently. We have performed extensive simulation experiments and the analysis of one data set has been presented for illustrative purposes.

We have further considered the multiple linear regression model in presence of additive SGSN errors. Note that the analysis of multiple linear regression model in presence of additive ASN distribution has been well studied in the statistical literature both from the classical and Bayesian view points, see for example Sahu et al. [27], Lachos et al. [20] and the references cited therein. Similarly, log-linear BS regression model also has been well studied in the literature, see for example Zhang et al. [30] and Balakrishnan and Kundu [6]. It is expected that the proposed multiple regression model will be more flexible than the ASN and LBS multiple regression models.

In this paper we provide the likelihood inference of the unknown parameters of the proposed SGSN regression model. It is observed that the ML estimators of the unknown parameters cannot be obtained in closed forms. In this case also we have used a very efficient

EM algorithm to avoid solving non-linear optimization problem. The implementation of the proposed EM algorithm is quite simple in practice. Moreover, using the method of Louis [22] at the last step of the EM algorithm the observed Fisher information matrix also can be obtained in a standard manner. Hence, the confidence intervals of the unknown parameters based on the observed Fisher information matrix also can be constructed. Simulation experimental results and analysis of a data set have been presented.

The rest of the paper is organized as follows. In Section 2 and Section 3 we provide a brief review of the GSN distribution and MGSN distribution, respectively. The SGSN, LS-SGSN distributions have been introduced and the EM algorithm have been discussed in Section 4. In Section 5, we have formulated the LS-SGSN regression model and discussed the corresponding EM algorithm. Simulation results have been presented in Section 6, and the analysis of two data sets have been presented in Section 7. Finally we conclude the paper in Section 8.

2 Geometric Skew Normal Distribution

In the previous section we have provided the definition of a GSN distribution. In this section we provide a brief review and some properties of a GSN distribution. All the details can be obtained in Kundu [14]. If $X \sim \text{GSN}(\mu, \sigma, p)$, then the CDF and PDF of X become

$$F_{GSN}(x; \mu, \sigma, p) = p \sum_{k=1}^{\infty} \Phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right)(1-p)^{k-1}, \tag{4}$$

and

$$f_{GSN}(x; \mu, \sigma, p) = \sum_{k=1}^{\infty} \frac{p}{\sigma\sqrt{k}} \phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right)(1-p)^{k-1}, \tag{5}$$

respectively, for $x \in \mathbb{R}$. Here $\phi(\cdot)$ denotes the PDF of a standard normal distribution function.

The PDF of the GSN distribution (5) can take variety of shapes. It can be symmetric, positively skewed, negatively skewed, unimodal, bimodal, multimodal and heavy tailed also. For different shapes of the PDF of a GSN distribution see Figure 1 and also Kundu [14]. It is clear that it can be positively skewed, negatively skewed, unimodal, multimodal, heavy tailed also depending on the parameter values.

If $X \sim \text{GSN}(\mu, \sigma, p)$, then the characteristic function of X is

$$\phi_X(t) = E(e^{itX}) = \frac{pe^{\left(i\mu t - \frac{\sigma^2 t^2}{2}\right)}}{1-(1-p)e^{\left(i\mu t - \frac{\sigma^2 t^2}{2}\right)}}, \qquad t \in \mathbb{R}.$$

The mean and variance of X are

$$E(X) = \frac{\mu}{p} \qquad \text{and} \qquad V(X) = \frac{\mu^2(1-p)+\sigma^2 p}{p^2}.$$

Any higher order moment can be obtained as infinite series as follows

$$E(X^m) = p\sum_{n=1}^{\infty}(1-p)^{n-1}c_m(n\mu, n\sigma^2).$$

Here, $c_m(n\mu, n\sigma^2) = E(Y^m)$, where $Y \sim N(n\mu, n\sigma^2)$. It may be mentioned that c_m can be obtained using confluent hyper geometric function, see for example Johnson, Kotz and Balakrishnan [13]. The skewness of a $\text{GSN}(\mu, \sigma, p)$ can be written as

$$\gamma_1 = \frac{(1-p)(\mu^3(2p^2-p+1)+2\mu^2p^2+\mu\sigma^2(3-p)p)}{(\sigma^2 p + \mu^2(1-p))^{3/2}}.$$

Hence, unlike normal distribution, the skewness depends on all the three parameters.

If $\mu = 0$, it becomes symmetric, and the CDF and PDF of a SGSN distribution become

$$F_{SGSN}(x; \sigma, p) = p\sum_{k=1}^{\infty}\Phi\left(\frac{x}{\sigma\sqrt{k}}\right)(1-p)^{k-1}, \tag{6}$$

and

$$f_{SGSN}(x; \sigma, p) = \sum_{k=1}^{\infty}\frac{p}{\sigma\sqrt{k}}\phi\left(\frac{x}{\sigma\sqrt{k}}\right)(1-p)^{k-1}, \tag{7}$$

respectively, for $x \in \mathbb{R}$. The PDF of SGSN are provided in Figure 2 for different σ and p values.

It is clear from Figure 2 that the tail probabilities of a SGSN distribution increase as p decreases. It behaves like a heavy tailed distribution as p tends to zero. As p tends to one, it behaves like a normal distribution.

If $X \sim \text{SGSN}(\sigma, p)$, then the characteristic function of X can be obtained as

$$\phi_X(t) = E(e^{itX}) = \frac{pe^{\left(-\frac{\sigma^2 t^2}{2}\right)}}{1-(1-p)e^{\left(-\frac{\sigma^2 t^2}{2}\right)}}, \qquad t \in \mathbb{R}.$$

Using the characteristic function or otherwise, the mean and variance of X can be expressed as follows:

$$E(X) = 0 \qquad \text{and} \qquad V(X) = \frac{\sigma^2}{p}.$$

For higher moments, it can be easily observed from the characteristic function that if $X \sim \text{SGSN}(1, p)$, then

$$E(X^m) = pd_m \sum_{n=1}^{\infty} (1-p)^{n-1} n^{m/2},$$

where $d_m = E(Z^m)$, $Z \sim \text{N}(0, 1)$, and

$$d_m = \begin{cases} 0 & \text{if} \quad \text{m is odd} \\ \frac{2^{m/2}\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi}} & \text{if} \quad \text{m is even.} \end{cases}$$

In Figure 3 we have plotted the characteristic functions of SGSN for different parameter values. Just to show how different it can be from a normal distribution we have plotted the characteristic function of the corresponding normal distribution with mean zero, and variance σ^2/p. It is clear that the characteristic functions can be quite different particularly for small values of p. But as p approaches one, they become close to each other.

It can be easily seen that the variance diverges as p approaches zero. The GSN distribution is known to be infinitely divisible and geometrically stable. For different other properties and estimation methods, interested readers are referred to the original article of Kundu [14].

The following derivations are needed for further development. These results will be used to develop EM algorithm for SGSN distribution. Suppose $X \sim \text{SGSN}(\sigma, p)$ and N is the associated $\text{GE}(p)$ random variable, then for $0 < p < 1$, and $-\infty < x < \infty$, $n = 1, 2, \ldots$,

$$P(X \leq x, N = n) = P(X \leq x | N = n) P(N = n) = p(1-p)^{n-1} \Phi\left(\frac{x}{\sigma\sqrt{n}}\right), \qquad (8)$$

for $p = 1$,

$$P(X \leq x, N = n) = \begin{cases} \Phi\left(\frac{x}{\sigma}\right) & \text{if } \ n = 1 \\ 0 & \text{if } \ n > 1. \end{cases} \qquad (9)$$

Hence, the joint PDF, $f_{X,N}(x,n)$, of (X, N) for $0 < p < 1$, $-\infty < x < \infty$, $n = 1, 2, \ldots$, becomes

$$f_{X,N}(x,n) = p(1-p)^{n-1} \frac{1}{\sigma\sqrt{2\pi n}} e^{-\frac{x^2}{2n\sigma^2}},$$

and for $p = 1$,

$$f_{X,N}(x,n) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} & \text{if } \ n = 1 \\ 0 & \text{if } \ n > 1. \end{cases} \qquad (10)$$

Therefore,

$$P(N = n | X = x) = \frac{(1-p)^{n-1} e^{-\frac{x^2}{2n\sigma^2}} / \sqrt{n}}{\sum_{j=1}^{\infty} (1-p)^{j-1} e^{-\frac{x^2}{2j\sigma^2}} / \sqrt{j}}. \qquad (11)$$

It can be easily shown that (11) is an unimodal function in n. Hence, there exits a unique n_0 such that $P(N = n_0 | X = x) > P(N = n | X = x)$, for any $n \neq n_0$. In fact n_0 can be obtained as the minimum value of $n \geq 1$, such that

$$\frac{P(N = n+1 | X = x)}{P(N = n | X = x)} = \frac{\sqrt{n}(1-p)}{\sqrt{n+1}} e^{\frac{x^2}{2\sigma^2 n(n+1)}} < 1. \qquad (12)$$

Moreover, from (11) one can easily obtain

$$E(N | X = x) = \frac{\sum_{n=1}^{\infty} \sqrt{n}(1-p)^{n-1} e^{-\frac{x^2}{2n\sigma^2}}}{\sum_{j=1}^{\infty} (1-p)^{j-1} e^{-\frac{x^2}{2j\sigma^2}} / \sqrt{j}}, \qquad (13)$$

and

$$E(1/N | X = x) = \frac{\sum_{n=1}^{\infty} n^{-3/2}(1-p)^{n-1} e^{-\frac{x^2}{2n\sigma^2}}}{\sum_{j=1}^{\infty} (1-p)^{j-1} e^{-\frac{x^2}{2j\sigma^2}} / \sqrt{j}}. \qquad (14)$$

3 Multivariate Geometric Skew Normal Distribution

Azzalini and Dalla Valle [5] introduced a multivariate distribution with ASN marginals. We call it as multivariate ASN (MASN) distribution and it can be defined as follows: A random vector $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$ is a d-variate MASN distribution, if it has the following PDF

$$g(\mathbf{z}) = 2\phi_d(\mathbf{z}, \boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}^\top \mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^d,$$

here $\phi_d(\mathbf{z}, \boldsymbol{\Omega})$ denotes the PDF of a d-variate multivariate normal distribution with standardized marginals and correlation matrix $\boldsymbol{\Omega}$. Here the parameter vector $\boldsymbol{\alpha}$ is known as the shape vector, and depending on the shape vector, the PDF of a MASN distribution can take variety of shapes. The PDF is always unimodal and when $\boldsymbol{\alpha} = \mathbf{0}$, then $\mathbf{Z}$ has the multivariate normal distribution with mean vector $\mathbf{0}$ and correlation matrix $\boldsymbol{\Omega}$. Although, MASN is a very flexible distribution, but if the marginals are heavy tailed or multimodal it cannot be used, see for example the excellent recent monograph by Azzalini and Capitanio [4]. Moreover, if the data come from a MASN distribution it can be shown that the MLEs do not exist with a positive probability. Due to these reasons several kernels instead of normal kernel have been used, but they have their own problems.

Kundu [17] proposed multivariate GSN (MGSN) distribution along the same line as the GSN distribution. The MGSN distribution has been defined as follows. Let us use the following notations. A d-variate normal random variable with mean vector $\boldsymbol{\mu}$ and the disersion matrix $\boldsymbol{\Sigma}$ will be denoted by $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The corresponding PDF and CDF will be denoted $\phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, respectively. Now a d-variate MGSN distribution can be defined as the following. Suppose $\boldsymbol{X}_i$ for $i = 1, 2, 3, \ldots$ are i.i.d. $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $N \sim \mathrm{GE}(p)$ and they are independently distributed. Then

$$\boldsymbol{X} = \sum_{i=1}^{N} \boldsymbol{X}_i$$

is said to have MGSN distribution with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and p and will be denoted by $\text{MGSN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$.

MGSN is a very flexible multivariate distribution, and its joint PDF can take variety of shapes. It can be unimodal, bimodal as well as multimodal also. Since the marginals are GSN distributions, therefore the marginals can be positively and negatively skewed, and it can be heavy tailed also. If $\boldsymbol{X} \sim \text{MGSN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$, then the PDF and CDF of $\boldsymbol{X}$ can be written as

$$f_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, p) = \sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}k^{d/2}} e^{-\frac{1}{2k}(\boldsymbol{x}-k\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})}$$

and

$$F_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, p) = \sum_{k=1}^{\infty} p(1-p)^{k-1} \Phi_d(\boldsymbol{x}; k\boldsymbol{\mu}, k\boldsymbol{\Sigma}),$$

respectively. When $\boldsymbol{\mu} = \boldsymbol{0}$, the PDF and CDF of $\boldsymbol{X}$ can be written as

$$f_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{0}, \boldsymbol{\Sigma}, p) = \sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}k^{d/2}} e^{-\frac{1}{2k}\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{x}}$$

and

$$F_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, p) = \sum_{k=1}^{\infty} p(1-p)^{k-1} \Phi_d(\boldsymbol{x}; \boldsymbol{0}, k\boldsymbol{\Sigma}),$$

respectively.

Similar to the GSN distribution, when $p = 1$, MGSN distribution becomes multivariate normal distribution. The PDF of a MGSN distribution can take variety of shapes. Contour plots of the PDF of a MGSN distribution for different parameter values when $d = 2$ are provided in Figure 4. It is clear that it can be positively and negative skewed, unimodal, bimodal, multimodal, heavy tailed depending on the parameter values. When $\boldsymbol{\mu} = \boldsymbol{0}$, then it becomes a symmetric distribution.

If $\boldsymbol{X} \sim \text{MGSN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$, then the characteristic function of $\boldsymbol{X}$ is

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \frac{p e^{i\boldsymbol{\mu}^\top \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma}\boldsymbol{t}}}{1-(1-p)e^{i\boldsymbol{\mu}^\top \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma}\boldsymbol{t}}}, \qquad \boldsymbol{t} \in \mathbb{R}^d.$$

Since the characteristic function is in a compact form, many properties can be derived quite conveniently. Different moments, product moments, multivariate skewness can be obtained using the characteristic function, see Kundu [17] for details.

MGSN distribution has several properties like multivariate normal distribution. For example if $\boldsymbol{X}$ is a d-variate MGSN distribution then if we partition $\boldsymbol{X}$ as

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix},$$

where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are of the orders d_1 and $n-d_1$, respectively. Then $\boldsymbol{X}_1$ is d_1-variate MGSN distribution and $\boldsymbol{X}_2$ is a $n-d_1$ variate MGSN distribution. Similarly, if $\boldsymbol{X}$ is d-variate MGSN distribution and $\boldsymbol{D}$ is a $s \times d$ matrix, with rank $s \leq d$, then $\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{X}$ is a s-variate MGSN distribution. It has been shown that $\boldsymbol{X}$ is d-variate MGSN distribution if and only if $Y = \boldsymbol{c}^\top \boldsymbol{X}$ is a GSN distribution, for all $\boldsymbol{c} \in \mathbb{R}^d \neq \boldsymbol{0}$. Several other interesting properties including canonical correlation, majorization and characterization have been provided in Kundu [17]. Estimation of the unknown parameters is an important problem. A d-variate MGSN distribution has $1 + d + d(d+1)/2$ unknown parameters. The usual maximum likelihood computation involves solving a $1 + d + d(d+1)/2$ variate optimization problem. Therefore, even for $d = 3$, it involves solving a 10 dimensional optimization problem, and clearly it is a non-trivial problem. A very efficient EM algorithm has been proposed by Kundu [17], which does not require solving any optimization problem, i.e. at each 'E'-step, the corresponding 'M'-step can be performed explicitly. Due to this reason, MGSN can be used quite efficiently in practice even for large d.

4 Location Shift SGSN Distribution

4.1 Model Description

Now we will introduce location shift SGSN distribution, which can be used quite effectively for analyzing symmetric data as an alternative to any symmetric distribution such as normal, log-BS, Student-t distributions etc. A random variable Y is called a location shift SGSN (LS-SGSN) distribution if

$$Y = \theta + X, \tag{15}$$

where $\theta \in \mathbb{R}$ and $X \sim \text{SGSN}(\sigma, p)$. If the random variable Y has the form (15), then it will be denoted by LS-SGSN(θ, σ, p). If $Y \sim$ LS-SGSN(θ, σ, p), then the CDF and PDF of Y become

$$F_Y(y; \theta, \sigma, p) = p \sum_{k=1}^{\infty} \Phi\left(\frac{y-\theta}{\sigma\sqrt{k}}\right)(1-p)^{k-1}, \tag{16}$$

and

$$f_Y(y; \theta, \sigma, p) = \sum_{k=1}^{\infty} \frac{p}{\sigma\sqrt{k}} \phi\left(\frac{y-\theta}{\sigma\sqrt{k}}\right)(1-p)^{k-1}, \tag{17}$$

respectively. Clearly, LS-SGSN distribution will be a more flexible than a normal distribution due to the presence of an extra parameter. Moreover, normal distribution can be obtained as a special case of the LS-SGSN distribution. We need the following derivations for further development, when $0 < p < 1$, $y \in \mathbb{R}$ and $n = 1, 2, \ldots$.

$$P(Y \leq y, N = n) = P(X \leq y - \theta, N = n) = p(1-p)^{n-1}\Phi\left(\frac{x-\theta}{\sigma\sqrt{n}}\right).$$

and

$$f_{Y,N}(y, n) = p(1-p)^{n-1}\frac{1}{\sigma\sqrt{2\pi n}}e^{-\frac{(x-\theta)^2}{2\sigma^2 n}}.$$

If $Y \sim$ LS-SGSN(θ, σ, p), then the characteristic function of Y can be obtained as

$$\phi_Y(t) = E(e^{itY}) = e^{it\theta}\frac{pe^{\left(-\frac{\sigma^2 t^2}{2}\right)}}{1-(1-p)e^{\left(-\frac{\sigma^2 t^2}{2}\right)}}, \qquad t \in \mathbb{R}.$$

The mean and variance of Y are

$$E(Y) = \theta \qquad \text{and} \qquad V(Y) = \frac{\sigma^2}{p},$$

respectively.

4.2 Maximum Likelihood Estimators

The proposed LS-SGSN distribution has three parameters. Now we will discuss about the estimation of the unknown parameters of the LS-SGSN distribution and the associated confidence intervals.

Suppose we have a random sample of size n, say,

$$\mathcal{D}_1 = \{x_1, \ldots, x_n\}$$

from LS-SGSN(θ, σ, p). The log-likelihood function based on the observation $\mathcal{D}_1$ can be written as

$$l(\theta, \sigma, p) = n \ln p - n \ln \sigma + \sum_{i=1}^{n} \ln \left\{ \sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \phi\left(\frac{x_i - \theta}{\sigma\sqrt{k}}\right) (1-p)^{k-1} \right\}. \tag{18}$$

Therefore, the ML estimators of the unknown parameters can be obtained by maximizing (18) with respect to the unknown parameters. The normal equations can be obtained as

$$\dot{l}_\theta = \frac{\partial l(\theta, \sigma, p)}{\partial \theta} = 0, \quad \dot{l}_\sigma = \frac{\partial l(\theta, \sigma, p)}{\partial \sigma} = 0, \quad \dot{l}_p = \frac{\partial l(\theta, \sigma, p)}{\partial p} = 0. \tag{19}$$

Hence, the ML estimates can be obtained by solving all the three normal equations in (19) simultaneously. Clearly, the ML estimates cannot be obtained in explicit forms, and one needs to use some iterative methods to solve (19). One may use Newton-Raphson type algorithm to solve these non-linear equations, but it has its own problem of local convergence and the choice of initial values. To avoid that we have proposed to use expectation maximization

(EM) type algorithm to compute the ML estimates. It is observed that at each E-step the corresponding M-step can be obtained explicitly. Hence, it can be implemented very easily.

The basic idea of the proposed EM type algorithm is to treat this problem as a missing value problem as follows: Suppose with each x_i we also observe the corresponding value of N, say m_i. Therefore, the complete observation becomes

$$\mathcal{D}_{1c} = \{(x_1, m_1), \ldots, (x_n, m_n)\}. \tag{20}$$

Based on the complete data, the log-likelihood function becomes

$$l(\theta, \sigma, p) = n \ln p + (\sum_{i=1}^{n} m_i - n) \ln(1-p) - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{(x_i - \theta)^2}{m_i}. \tag{21}$$

Hence, the ML estimates of θ, σ and p based on the complete observation $\mathcal{D}_{1c}$ can be obtained by maximizing (21). If we denote them as $\widehat{\theta}_c$, $\widehat{\sigma}_c$ and $\widehat{p}_c$, respectively, then

$$\widehat{\theta}_c = \frac{1}{\sum_{i=1}^{n} \frac{1}{m_i}} \sum_{i=1}^{n} \frac{x_i}{m_i}, \quad \widehat{\sigma}_c^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \widehat{\theta}_c)^2}{m_i}, \quad \widehat{p}_c = \frac{n}{\sum_{i=1}^{n} m_i}. \tag{22}$$

Hence, in this case the EM algorithm takes the following form. Suppose at the k-th stage of the EM algorithm the estimates of θ, σ and p are $\theta^{(k)}$, $\sigma^{(k)}$ and $p^{(k)}$, respectively. Then at the 'E'-step of the EM algorithm, the 'pseudo' log-likelihood function can be written as

$$l_s(\theta, \sigma, p | \theta^{(k)}, \sigma^{(k)}, p^{(k)}) = n \ln p + (\sum_{i=1}^{n} a_i^{(k)} - n) \ln(1-p) - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} b_i^{(k)} (x_i - \theta)^2. \tag{23}$$

Here,

$$a_i^{(k)} = E(N | Y = x_i, \theta^{(k)}, \sigma^{(k)}, p^{(k)}) = \frac{\sum_{n=1}^{\infty} \sqrt{n} (1 - p^{(k)})^{n-1} e^{-\frac{(x_i - \theta^{(k)})^2}{2n(\sigma^{(k)})^2}}}{\sum_{j=1}^{\infty} (1 - p^{(k)})^{j-1} e^{-\frac{(x_i - \theta^{(k)})^2}{2j(\sigma^{(k)})^2}} / \sqrt{j}} \tag{24}$$

and

$$b_i^{(k)} = E(1/N | Y = x_i, \theta^{(k)}, \sigma^{(k)}, p^{(k)}) = \frac{\sum_{n=1}^{\infty} n^{-3/2} (1 - p^{(k)})^{n-1} e^{-\frac{(x_i - \theta^{(k)})^2}{2n(\sigma^{(k)})^2}}}{\sum_{j=1}^{\infty} (1 - p^{(k)})^{j-1} e^{-\frac{(x_i - \theta^{(k)})^2}{2j(\sigma^{(k)})^2}} / \sqrt{j}}. \tag{25}$$

Therefore, at the 'M'-th step the maximization of the 'pseudo' log-likelihood functions provides:

$$\theta^{(k+1)} = \frac{\sum_{i=1}^{n} b_i^{(k)} x_i}{\sum_{j=1}^{n} b_j^{(k)}}, \quad \sigma^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} b_i^{(k)} (x_i - \theta^{(k+1)})^2, \quad p^{(k+1)} = \frac{n}{\sum_{i=1}^{n} a_i^{(k)}}. \tag{26}$$

Note that to start the EM algorithm, we need some initial estimates of θ, σ and p. We suggest the following. Use the sample mean and the sample standard deviation as the initial estimates of θ and σ, respectively. We can start the EM algorithm for different initial values of p from $(0, 1)$. The EM algorithm can be explicitly written as follows:

EM Algorithm

Step 1: Choose some initial estimates of θ, σ and p, say $\theta^{(1)}$, $\sigma^{(1)}$ and $p^{(1)}$ as described above. Put $k = 1$.

Step 2: Compute $a_i^{(k)}$ and $b_i^{(k)}$ as in (24) and (25), respectively.

Step 3: Compute $\theta^{(k+1)}$, $\sigma^{(k+1)}$ and $p^{(k+1)}$ as in (26).

Step 4: Check the convergence. If the convergence criterion is satisfied then stop, otherwise put $k = k + 1$, and go to Step 2.

The observed Fisher information matrix also can be easily obtained as the last step of the corresponding algorithm. The observed Fisher information matrix can be written as $\mathbf{I} = \mathbf{B} - \mathbf{S}\mathbf{S}^\top$. Here, $\mathbf{B}$ is a 3×3 Hessian matrix of the 'pseudo' log-likelihood function (23) and $\mathbf{S}$ is the corresponding gradient vector. If we denote $\mathbf{B} = ((b_{ij}))$ and $\mathbf{S} = (s_i)$, then at the k-th stage of the EM algorithm

$$\begin{aligned}
b_{11} &= \frac{1}{(\sigma^{(k)})^2}\sum_{i=1}^n b_i^{(k)}, b_{12} = b_{21} = \frac{2}{(\sigma^{(k)})^3}\sum_{i=1}^n b_i^{(k)}(x_i - \theta^{(k)}), \\
b_{22} &= -\frac{n}{(\sigma^{(k)})^2} + \frac{3}{(\sigma^{(k)})^4}\sum_{i=1}^n b_i^{(k)}(x_i - \theta^{(k)})^2 \\
b_{33} &= \frac{n}{(p^{(k)})^2} + \frac{\sum_{i=1}^n a_i^{(k)} - n}{(1-p^{(k)})^2}, b_{13} = b_{31} = b_{23} = b_{32} = 0.
\end{aligned}$$

$$s_1 = \frac{1}{(\sigma^{(k)})^2}\sum_{i=1}^n b_i^{(k)}(x_i - \theta^{(k)}), s_2 = -\frac{n}{\sigma^{(k)}} + \frac{1}{(\sigma^{(k)})^3}\sum_{i=1}^n b_i^{(k)}(x_i - \theta^{(k)})^2, s_3 = \frac{n}{p^{(k)}} - \frac{\sum_{i=1}^n a_i^{(k)}}{1-p^{(k)}}.$$

Therefore, if $\mathbf{I}^{-1} = ((fij))$, then $100(1-\alpha)\%$ confidence intervals of θ, σ and p can be obtained as

$$(\widehat{\theta} - z_{\alpha/2} f^{11}, \widehat{\theta} + z_{\alpha/2} f^{11}), \quad (\widehat{\sigma} - z_{\alpha/2} f^{22}, \widehat{\sigma} + z_{\alpha/2} f^{22}), \quad (\widehat{p} - z_{\alpha/2} f^{33}, \widehat{p} + z_{\alpha/2} f^{33}).$$

respectively. Here z_α denotes the α-th percentile point of a standard normal distribution.

4.3 Testing of Hypothesis

In this section we discuss some testing of hypotheses problems which have some practical importance.

Problem 1: We want to test

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0. \tag{27}$$

The problem is of interest as it tests whether the distribution has a specific mean or not. We propose to use the likelihood ratio test (LRT) for this purpose. To compute the LRT statistic, we need to compute the ML estimates of σ and p, when $\theta = \theta_0$. In this case also we can use the same EM algorithm with the obvious modification that at each stage $\theta^{(k)}$ is replaced by θ_0. Therefore, if $\widehat{\theta}$, $\widehat{\sigma}$ and $\widehat{p}$ denote the ML estimates of θ, σ and p, respectively, without any restriction, and $\widetilde{\sigma}$ and $\widetilde{p}$ denote the ML estimators of σ and p, respectively, under H_0, then

$$2(l(\widehat{\theta}, \widehat{\sigma}, \widehat{p}) - l(\theta_0, \widetilde{\sigma}, \widetilde{p})) \longrightarrow \chi^2_1.$$

Problem 2: We want to test

$$H_0 : p = 1 \quad \text{vs.} \quad H_1 : p < 1. \tag{28}$$

The problem is of interest as it tests whether the distribution is normal or not. In this case under H_0, the ML estimates of θ and σ can be obtained as

$$\widetilde{\theta} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{and} \quad \widetilde{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \widetilde{\theta})^2}{n}}.$$

In this case p is in the boundary under H_0, hence, the standard results do not hold. But using Theorem 3 of Self and Liang [28], it follows that

$$2(l(\widehat{\theta}, \widehat{\sigma}, \widehat{p}) - l(\widetilde{\theta}, \widetilde{\sigma}, 1) \longrightarrow \frac{1}{2} + \frac{1}{2}\chi_1^2.$$

5 SGSN Regression Model

In this section we introduce the following regression model:

$$\mathbf{Y} = \mathbf{X}^\top \boldsymbol{\beta} + \epsilon, \tag{29}$$

here $\mathbf{Y}$ is a $n \times 1$ observed vector, $\mathbf{X} = (X_1, X_2, \ldots, X_k)^\top$ is a k-dimensional covariate vector, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)^\top$ is a k-vector of regression coefficients, $\mathbf{a}^\top$ denotes the transpose of an arbitrary vector $\mathbf{a}$ and $\epsilon \sim \text{SGSN}(\sigma, p)$. Note that it is a standard multiple linear regression model when ϵ has a normal distribution with mean zero and finite variance. Since, the normal distribution is a special case of the proposed SGSN distribution, therefore, the model (29) will be more flexible than the standard multiple linear regression model.

The main aim of this section is to discuss the estimation of the unknown parameters of the model (29). Let

$$\mathcal{D} = \{(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \ldots, (y_n, \mathbf{x}_n)\},$$

be n independent observations from the model (29), the problem is to estimate the unknown parameters $\beta_1, \ldots, \beta_k, \sigma, p$. We use the following notation $\mathbf{x}_i = (x_{1i}, \ldots, x_{ki})^\top$, $i = 1, \ldots, n$,

Let us try to compute the ML estimates of the unknown parameters. The log-likelihood function of the observed data $\mathcal{D}$ based on the model (29) can be obtained as

$$l(\boldsymbol{\theta}) = n \ln p - n \ln(1-p) - n \ln \sigma + \sum_{i=1}^{n} \ln \left(\sum_{j=1}^{\infty} \frac{(1-p)^j}{\sqrt{j}} \phi\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma\sqrt{j}} \right) \right), \tag{30}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma, p)^\top$. Therefore, the ML estimates of the unknown parameters can be obtained by maximizing the the log-likelihood function (30) with respect to the unknown parameters. Taking derivatives with respect to the unknown parameters, the normal equations become:

$$\dot{l}_{\beta_1} = \frac{\partial}{\partial \beta_1} l(\boldsymbol{\theta}) = 0, \ldots, \dot{l}_{\beta_k} = \frac{\partial}{\partial \beta_k} l(\boldsymbol{\theta}) = 0, \quad \dot{l}_{\sigma} = \frac{\partial}{\partial \sigma} l(\boldsymbol{\theta}) = 0, \quad \dot{l}_{p} = \frac{\partial}{\partial p} l(\boldsymbol{\theta}) = 0. \tag{31}$$

Clearly, they cannot be solved explicitly. One needs to solve $k+2$ non-linear equations simultaneously to compute the ML estimates of the unknown parameters. To avoid that problem, we propose to use this problem as a missing value problem, and provide an efficient EM algorithm to compute the ML estimates of the unknown parameters. The main idea is as follows. Suppose along with $(y_i, \mathbf{x}_i)$ we also observe m_i, where m_i denotes the value of N in this case, for $i = 1, 2, \ldots, n$. Therefore, the complete data will be of the form:

$$\mathcal{D}^c = \{(y_1, \mathbf{x}_1, m_1), \ldots, (y_n, \mathbf{x}_n, m_n)\}. \tag{32}$$

First we will show that if m_i's are known, then the ML estimates of $\boldsymbol{\beta}$, σ and p can be obtained in explicit forms. Based on the complete data (32), the log-likelihood function without the additive constant becomes

$$l_c(\boldsymbol{\beta}, \sigma, p) = n \ln p + (\sum_{i=1}^{n} m_i - n) \ln(1-p) - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{1}{m_i} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right)^2. \tag{33}$$

It can be easily seen that $l_c(\boldsymbol{\beta}, \sigma, p)$ as given in (33) is an unimodal function. If $\widehat{\boldsymbol{\beta}}_c$, $\widehat{\sigma}_c$ and $\widehat{p}$ maximize $l_c(\boldsymbol{\beta}, \sigma, p)$, then they can be obtained as

$$\widehat{\boldsymbol{\beta}}_c = \left[\sum_{i=1}^{n} \frac{1}{m_i} \mathbf{x}_i \mathbf{x}_i^\top\right]^{-1}, \quad \widehat{\sigma}_c^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \left(y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_c\right)^2, \quad \widehat{p} = \frac{n}{\sum_{i=1}^{n} m_i}. \tag{34}$$

Therefore, it is clear that if m_i's are known then the ML estimates of $\boldsymbol{\beta}$, σ and p can be obtained quite conveniently, and one does not need to solve any non-linear equation. This is the main motivation of the proposed EM algorithm. Now we are ready to provide the EM algorithm for this problem. We will show how to move from the r-th step to the $(r+1)$-th step of the EM algorithm.

We will use the following notations. Let us denote $\boldsymbol{\beta}^{(r)}$, $\sigma^{(r)}$ and $p^{(r)}$ as the estimates of $\boldsymbol{\beta}$, σ and p, respectively, for $i = 1, 2, \ldots, n$, at the r-th iteration of the EM algorithm. Then at the 'E'-step of the EM algorithm, the pseudo log-likelihood function becomes:

$$l_s(\boldsymbol{\beta}, \sigma, p | \boldsymbol{\beta}^{(r)}, \sigma^{(r)}, p^{(r)}) = n \ln p + (\sum_{i=1}^{n} c_i^{(r)} - n) \ln(1-p) - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} d_i^{(r)} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right)^2, \tag{35}$$

where

$$c_i^{(r)} = E(N | y_i, \mathbf{x}_i, \boldsymbol{\beta}^{(r)}, \sigma^{(r)}, p^{(r)}) = \frac{\sum_{n=1}^{\infty} \sqrt{n} (1-p^{(r)})^{n-1} e^{-\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(r)})^2}{2n(\sigma^{(r)})^2}}}{\sum_{j=1}^{\infty} (1-p^{(r)})^{j-1} e^{-\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(r)})^2}{2j(\sigma^{(r)})^2}} / \sqrt{j}} \tag{36}$$

and

$$d_i^{(r)} = E(1/N | y_i, \mathbf{x}_i, \boldsymbol{\beta}^{(r)}, \sigma^{(r)}, p^{(r)}) = \frac{\sum_{n=1}^{\infty} n^{-3/2} (1-p^{(k)})^{n-1} e^{-\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(r)})^2}{2n(\sigma^{(k)})^2}}}{\sum_{j=1}^{\infty} (1-p^{(k)})^{j-1} e^{-\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(r)})^2}{2j(\sigma^{(k)})^2}} / \sqrt{j}}. \tag{37}$$

Therefore, at the 'M'-th step, the maximization of the 'pseudo' log-likelihood function provides

$$\boldsymbol{\beta}^{(r+1)} = \left[\sum_{i=1}^{n} d_i^{(r)} \mathbf{x}_i \mathbf{x}_i^\top\right]^{-1}, \quad \sigma^{(r+1)} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} d_i^{(r)} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(r+1)}\right)^2}, \quad \widehat{p} = \frac{n}{\sum_{i=1}^{n} c_i^{(r)}}, \tag{38}$$

moreover, these are unique solutions.

Now to start the EM algorithm, we suggest the following initial guesses. Use ordinary least squares estimates of $\boldsymbol{\beta}$ as the initial estimate of $\boldsymbol{\beta}$ and the corresponding square root of the residual sums of squares as the initial estimate of σ. As before, we can start the EM algorithm with different initial values of $p \in (0, 1)$. The EM algorithm can be written as follows.

EM Algorithm

Step 1: Choose some initial estimates of $\boldsymbol{\beta}$, σ and p, say $\boldsymbol{\beta}^{(1)}$, $\sigma^{(1)}$ and $p^{(1)}$ as described above. Put $r = 1$.

Step 2: Compute $c_i^{(r)}$ and $c_i^{(r)}$ as in (36) and (37), respectively.

Step 3: Compute $\boldsymbol{\beta}^{(r+1)}$, $\sigma^{(r+1)}$ and $p^{(r+1)}$ as in (38).

Step 4: Check the convergence. If the convergence criterion is satisfied then stop, otherwise put $r = r+1$, and go to Step 2.

In this case also the observed Fisher information matrix also can be obtained as before, in the last step of the EM algorithm. The observed Fisher information matrix can be written as $\widetilde{\mathbf{I}} = \widetilde{\mathbf{B}} - \widetilde{\mathbf{S}}\widetilde{\mathbf{S}}^\top$. Here, $\widetilde{\mathbf{B}}$ is a $(k+3)\times(k+3)$ Hessian matrix of the 'pseudo' log-likelihood function (35) and $\widetilde{\mathbf{S}}$ is the corresponding gradient vector. If we denote $\widetilde{\mathbf{B}} = ((\widetilde{b}_{ij}))$ and $\widetilde{\mathbf{S}} = (\widetilde{s}_i)$, then at the r-th stage of the EM algorithm, for $j, l = 1, \ldots, k$ and for $j \neq l$,

$$\widetilde{b}_{jj} = \frac{1}{(\sigma^{(r)})^2}\sum_{i=1}^n d_i^{(r)} x_{ji}^2, \qquad \widetilde{b}_{jl} = \widetilde{b}_{lj} = \frac{1}{(\sigma^{(r)})^3}\sum_{i=1}^n d_i^{(r)} x_{ji} x_{li},$$

$$\begin{aligned}\widetilde{b}_{(k+1)(k+1)} &= -\frac{n}{(\sigma^{(r)})^2} + \frac{3}{(\sigma^{(r)})^4}\sum_{i=1}^n d_i^{(k)}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(r)})^2,\\ \widetilde{b}_{j(k+1)} &= \widetilde{b}_{(k+1)j} = \frac{2}{(\sigma^{(r)})^3}\sum_{i=1}^n d_i^{(k)}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(r)})x_{ji},\end{aligned}$$

$$\widetilde{b}_{(k+2)(k+2)} = \frac{n}{(p^{(r)})^2} + \frac{\sum_{i=1}^n c_i^{(r)} - n}{(1-p^{(r)})^2}, \qquad b_{j(k+2)} = 0; \quad j = 1, \ldots, k+1.$$

$$s_j = \frac{1}{(\sigma^{(r)})^2}\sum_{i=1}^n d_i^{(r)}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(r)})\mathbf{x}_{ji}; \quad j = 1, \ldots k,$$

$$s_{(k+1)} = -\frac{n}{\sigma^{(r)}} + \frac{1}{(\sigma^{(r)})^3}\sum_{i=1}^n d_i^{(r)}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(r)})^2, \quad s_{(k+2)} = \frac{n}{p^{(r)}} - \frac{\sum_{i=1}^n c_i^{(r)} - n}{1-p^{(r)}}.$$

Therefore, if $\widetilde{\mathbf{I}}^{-1} = ((\widetilde{f}ij))$, then $100(1-\alpha)\%$ confidence intervals of β_j, σ and p can be obtained as

$$(\widehat{\beta}_j - z_{\alpha/2}\widetilde{f}^{jj}, \widehat{\beta}_j + z_{\alpha/2}\widetilde{f}^{jj}); \quad j = 1, \ldots, k,$$

$$(\widehat{\sigma} - z_{\alpha/2}\widetilde{f}^{(k+1)(k+1)}, \widehat{\sigma} + z_{\alpha/2}\widetilde{f}^{(k+1)(k+1)}), \quad (\widehat{p} - z_{\alpha/2}\widetilde{f}^{(k+2)(k+2)}, \widehat{p} + z_{\alpha/2}\widetilde{f}^{(k+2)(k+2)}),$$

respectively. Here z_α denotes the α-th percentile point of a standard normal distribution.

6 Simulation Results

In this section we present some simulation results both for the LS-SGSN and SGSN regression models. The main idea is to see how the proposed EM algorithms behave in these cases. All the computations are performed using R software, and it can be obtained from the corresponding author on request.

6.1 LS-SGSN Model

In this section we have generated samples from a LS-SGSN distribution for different sample sizes n and different p values. We have kept the values of θ and σ to be same, namely $\theta = 0$ and $\sigma = 1$. We have taken $n = 25$, 50, 75 and 100, and $p = 0.2$, 0.4, 0.6 and 0.8. We have used the EM algorithm as it has been described in Section 3. In all the cases, we have used the true value as the initial guesses. We stop the iteration when the absolute difference of the consecutive estimates is less that 10^{-6}, for all the unknown parameters. We have reported the average ML estimates of θ, σ, p, and the associated mean squared errors (MSEs) based on 1000 replications. The results are reported in Table 1.

Some of the points are quite clear from the above experimental results. It is observed that ML estimates of θ, σ and p provide unbiased estimates of the corresponding parameters. As sample size increases, in all the cases considered, the MSEs decrease. It indicates consistency property of the ML estimates. The EM algorithm converges in all the cases. Hence, the EM algorithm works well in this case.

6.2 SGSN Regression Model

In this section we have performed some simulation experiments for SGSN Regression model. Here, we have taken $k = 2$, $\beta_1 = 1$, $\beta_2 = 2$, $\sigma = 1$. We have varied $n = 25$, 50, 75 and 100, and

n		$p = 0.2$	$p = 0.4$	$p = 0.6$	$p = 0.8$
25	θ	0.0003	0.0062	0.0007	-0.0026
		(0.1230)	(0.0738)	(0.0621)	(0.0483)
	σ	0.9251	0.9177	0.9293	0.9314
		(0.0471)	(0.0419)	(0.0369)	(0.0374)
	p	0.1883	0.3703	0.5699	0.7546
		(0.0224)	(0.0116)	(0.0103)	(0.0081)
50	θ	0.0094	-0.0004	-0.0030	0.0067
		(0.0651)	(0.0398)	(0.0320)	(0.0243)
	σ	0.9515	0.9427	0.9519	0.9569
		(0.0396)	(0.0326)	(0.0294)	(0.0211)
	p	0.1890	0.3792	0.5768	0.7748
		(0.0121)	(0.0099)	(0.0089)	(0.0075)
75	θ	0.0104	0.0036	0.0060	0.0028
		(0.0432)	(0.0271)	(0.0189)	(0.0149)
	σ	0.9606	0.9747	0.9720	0.9683
		(0.0273)	(0.0250)	(0.0231)	(0.0200)
	p	0.1891	0.3835	0.5775	0.7787
		(0.0099)	(0.0079)	(0.0056)	(0.0047)
100	θ	-0.0004	-0.0018	-0.0023	0.0016
		(0.0318)	(0.0196)	(0.0144)	(0.0115)
	σ	0.9805	0.9767	0.9889	0.9839
		(0.0165)	(0.0234)	(0.0193)	(0.0165)
	p	0.1948	0.3927	0.5860	0.7880
		(0.0071)	(0.0056)	(0.0032)	(0.0027)

Table 1: ML estimates of θ, σ, p and the associated MSEs (reported below within brackets) for different sample sizes and for different p values.

$p = 0.2$, 0.4, 0.6 and 0.8. We have generated the entries of the $\mathbf{X}$ matrix from i.i.d. normal random variables, with mean zero and variance one. For a fixed, n, the generated $\mathbf{X}$ matrix remains fixed. The error random variables are generated from a SGSN(σ, p) distribution. For each n, σ and p, we generate Y. We compute the ML estimates of β_1, β_2, σ and p, based on the EM algorithm as described in Section 4. In all the cases we report the average ML estimates and the associated MSEs bases on 1000 replications. The results are reported in Table 2.

Some of the points are quite clear from both the tables. It is observed that even for the SGSN regression model the proposed EM algorithm works quite well in all the cases considered. The ML estimates provide consistent estimates of the corresponding unknown parameters, and the EM algorithm also converges in each case.

7 Real Data Analysis

7.1 LS-SGSN Model

In this section we present the analysis of one data set based on LS-SGSN model mainly for illustrative purposes. The data set has been obtained from Lawless [21] (page 228). It arose from test on the endurance of deep groove ball bearings. It represents the number of million revolutions before failure for each of the 23 ball bearings in the life test and they are as follows:

17.88 28.92 33.00 41.52 42.12 45.60 48.40 51.84 51.96 54.12 55.56 67.80 68.64 68.64 68.88 84.12 93.12 98.64 105.12 105.84 127.92 128.04 173.40.

The mean and standard deviation of the data points are 72.2 and 36.7, respectively. We have used the EM algorithm with the initial estimates of θ, σ and p as 72.2, 36.7 and 0.5, respectively. We stop the iteration when the difference between the consecutive estimates are less that 10^{-4} for all the three parameters. The maximum likelihood estimates and the associated log-likelihood (ll) value are as follows:

$$\widehat{\theta} = 68.443, \quad \widehat{\sigma} = 26.088 \quad \widehat{p} = 0.554, \quad ll = -127.4590.$$

The corresponding 95% confidence intervals of θ, σ and p are obtained as:

$$(56.513, 80.373), \quad (18.857, 33.320), \quad (0.409, 0.699),$$

respectively. To check the effect of the initial guess to the final estimates, we have used different other initial guesses, mainly the p values ranging from 0.1 to 0.9, but in all the cases it converge to the same estimate. It confirms that the initial estimates do not affect the convergence of the EM algorithm.

7.2 SGSN Regression Model

In this section we have analyzed one Rocket Propellant Data set from Montgomery et al. [23] (Chapter 2). The data represent the shear strength (Y) and the age of the propellant (X). Here the shear strength is in psi and the age is in weeks. The observations are given below in the format (Y, X).

(2158.70, 15.50) (1678.15, 23.75) (2316.00, 8.00) (2061.30, 17.00) (2207.50, 5.50) (1708.30, 19.00) (1784.70, 24.00) (2575.00, 2.50) (2357.90, 7.50) (2256.70, 11.00) (2165.20, 13.00) (2399.55, 3.75) (1779.80, 25.00) (2336.75, 9.75) (1765.30, 22.00) (2053.50, 18.00) (2414.40, 6.00) (2200.50, 12.50) (2654.20, 2.00) (1753.70, 21.50).

We have used the following model to analyze the data:

$$Y = \beta_1 + \beta_2 X + \epsilon.$$

Here ϵ is assumed to follow a GSSN distribution with mean zero and finite variance. Assuming that ϵ follows normal distribution, we obtain the least squares estimates of β_1, β_2 and σ as 2627.82, -37.15 and 91.17, respectively. We have used these estimates as the initial estimates and we have used the initial estimate of p as 0.5. Using the EM algorithm we obtain the ML estimates of the unknown parameters and the corresponding log-likelihood value are as follows:

$$\widehat{\beta}_1 = 2649.58, \quad \widehat{\beta}_2 = -37.58, \quad \widehat{\sigma} = 53.42, \quad \widehat{p} = 0.34, \quad ll = -88.97.$$

The associated 95% confidence intervals are as follows:

$$(2584.95, 2714.18), \quad (-41.83, -33.34), \quad (36.87, 69.98), \quad (0.22, 0.46).$$

8 Conclusions

In this paper we have considered LS-SGSN distribution. It is symmetric, and it has three parameters. Hence, it is more flexible than the normal distribution. The normal distribution can be obtained as as special case of the LS-SGSN distribution. We have proposed a very efficient EM algorithm, and it is observed that the EM algorithm works quite well. We have further considered SGSN regression model, which is more flexible than the standard Gaussian regression model. In this case also we have proposed a very efficient EM algorithm, and the performance of the proposed EM algorithm is quite satisfactory.

References

[1] Arnold, B.C. and Beaver, R.J. (2000), "Hidden truncation models", *Sankhya*, vol. 62, 23- 35.

[2] Arnold, B.C. and Beaver, R.J., Groeneveld, R.A. and Meeker , W.Q. (1993), "The non-truncated marginal of a truncated bivariate normal distribution", *Psychometrika*, vol. 58, 471 - 488.

[3] Azzalini, A.A. (1985), "A class of distributions which include the normal", *Scandinavian Journal of Statistics*, vol. 12, 171 – 178.

[4] Azzalini, A.A. and Capitanio, A. (2014), *The skew-normal and related families*, Cambridge University Press, Cambridge, United Kingdom.

[5] Azzalini, A.A. and Dalla Valle, A. (1996), "The multivariate skew normal distribution", *Biometrika*, vol. 83, 715 - 726

[6] Balakrishnan, N. and Kundu, D. (2019), "Birnbaum-Saunders Distribution: A Review of Models, Analysis and Applications (with discussions)", *Applied Stochastic Models in Business and Industry*, vol. 35, no. 1, 4 – 132.

[7] Birnbaum, Z. and Saunders, S.C. (1969), "A new family of life distributions", *Journal of Applied Probability*, vol. 6, 319 – 327.

[8] Cook, R.D. (1986), "Assessment of local influence", *Journal of the Royal Statistical Society*, Ser. B, vol. 48, 133 – 169.

[9] Fachini, J.B., Ortega, E.M.M., Louzada-Neto, F. (2008), "Influence diagnostics for polynomial hazards in presence of covariates", *Statistical Methods and its Applications*, vol. 17, 413 – 433.

[10] Gupta, R.C. and Gupta, R.D. (2004), "Generalized skew normal model", *TEST*, vol. 13, 1- 24.

[11] Hashimoto, E.M., Ortega, E.M.M., Cancho, V.G., Cordeiro, G.M. (2010), "The log-exponentiated Weibull regression model for interval censored data", *Computational Statistics and Data Analysis*, vol. 54, 1017 – 1035.

[12] Hashimoto, E.M., Ortega, E.M.M., Cordeiro, G.M., Cancho, V.G. (2014), "The Poisson Birnbaum-Saunders model with long-term survivors", *Statistics*, vol. 48, no. 6, 1394 – 1413.

[13] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995), *Continuous Univariate Distribution*, Volume 1, John Wiley and Sons, New York, USA.

[14] Kundu, D. (2014), “Geometric skewed normal distribution”, *Sankhya*, Ser. B., vol. 76, Part 2, 167 – 189.

[15] Kundu, D. (2015a), “Bivariate sinh-normal distribution and a related model”, *Brazilian Journal of Probability and Statistics*, vol. 29, no. 3, 590 - 607.

[16] Kundu, D. (2015b), “Bivariate log Birnbaum-Saunders Distribution”, *Statistics*, vol. 49, no. 4, 900 - 917, 2015.

[17] Kundu, D. (2017), “Multivariate geometric skew normal distribution”, *Statistics*, vol. 51, no. 6, 1377 - 1397, 2017.

[18] Kundu, D. and Gupta, R.D., “Power Normal Distribution”, *Statistics*, vol. 47, no. 1, 110 - 125, 2013.

[19] Kundu, D. and Nekoukhou, V. (2018), “Univariate and bivariate geometric discrete generalized exponential distribution”, *Journal of Statistical Theory and Practice*, vol. 12, no. 3, 595 – 614.

[20] Lachos, V.H., Bolfarine, H., Arellano-Valle, R.B., Montenegro, L.C. (2007), “Likelihood-Based Inference for Multivariate Skew-Normal Regression Models”, *Communications in Statistics - Theory and Methods*, vol. , 36, 1769 – 1786.

[21] Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, New York, Wiley.

[22] Louis, T. A. (1982) “Finding the observed information matrix when using the EM algorithm”, *Journal of the Royal Statistical Society*, Series B, vol. 44, no. 2, 226–233.

[23] Montgomery, D.C., Peck, E.A., Vining, G.G. (2001), *Introduction to Linear Regression Analysis*, Third edition, Wiley, New York.

[24] Ortega, E.M.M., Bolfarine, H., Paula, G.A. (2003), "Influence diagnostics in generalized log-gamma regression model", *Computational Statistics and Data Analysis*, vol. 42, 165 – 186.

[25] Rieck, J.R. (1989), "Statistical analysis for the Birnbaum-Saunders fatigue life distribution", Ph.D. thesis, Clemson University, Department of Mathematical Sciences, Canada.

[26] Rieck, J.R. and Nedelman, J.R. (1991), "A log-linear model for the Birnbaum-Saunders distribution", *Technometrics*, vol. 33, 51 – 60.

[27] Sahu, S. K., Dey, D. K., Branco, M. D. (2003), "A new class of multivariate skew distributions with applications to Bayesian regression models", *Canadian Journal of Statistics*, vol. 31, no. 2, 129 – 150.

[28] Self, S.G. and Liang, K-L. (1987), "Asymptotic properties of the maximum likelihood estimators and likelihood ratio test under non-standard conditions", *Journal of the American Statistical Association*, vol. 82, 605 – 610.

[29] Silva, G.O., Ortega, E.M.M., Cancho, V.G., Barreto, M.L. (2008), "Log-Burr XII regression model with censored data", *Computational Statistics and Data Analysis*, vol. 52, 3820 – 3842.

[30] Zhang, Y., Lu, X., Desmond, A.F. (2016), "Variable selection in a log-linear Birnbaum-Saunders regression model for high-dimensional survival data via elastic-net and stochastic EM", *Technometrics*, vol. 58, no. 3, 383 – 392.

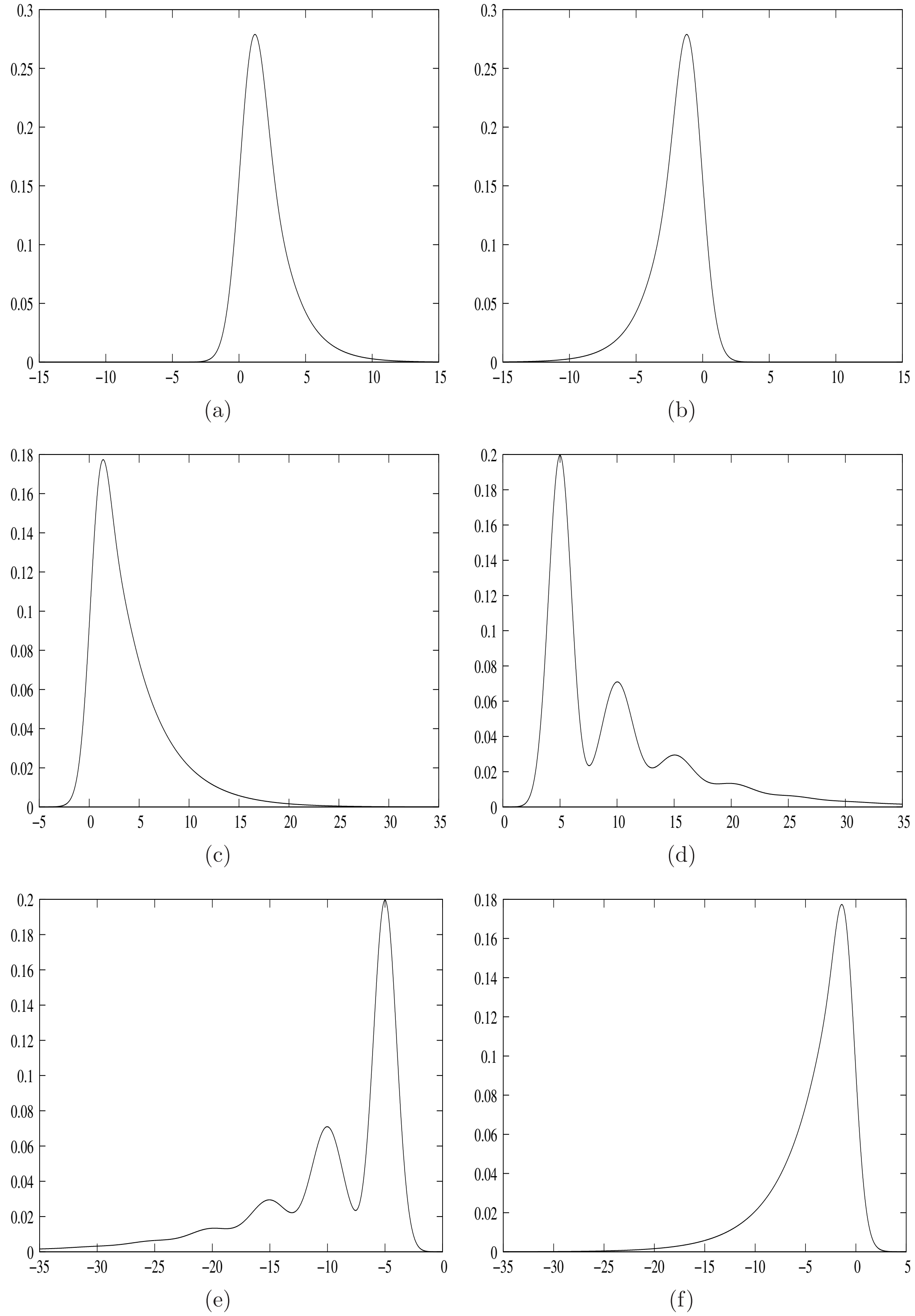

Figure 1: PDF plots of GSN(μ, σ, p) distribution for different (μ, σ, p) values: (a) (1.0,1,0.5) (b) (-1.0,1.0,0.5) (c) (1.0,1.0,0.25) (d) (5.0,1.0,0.5) (e) (-5.0,1.0,0.5) (-1.0,1.0,0.25).

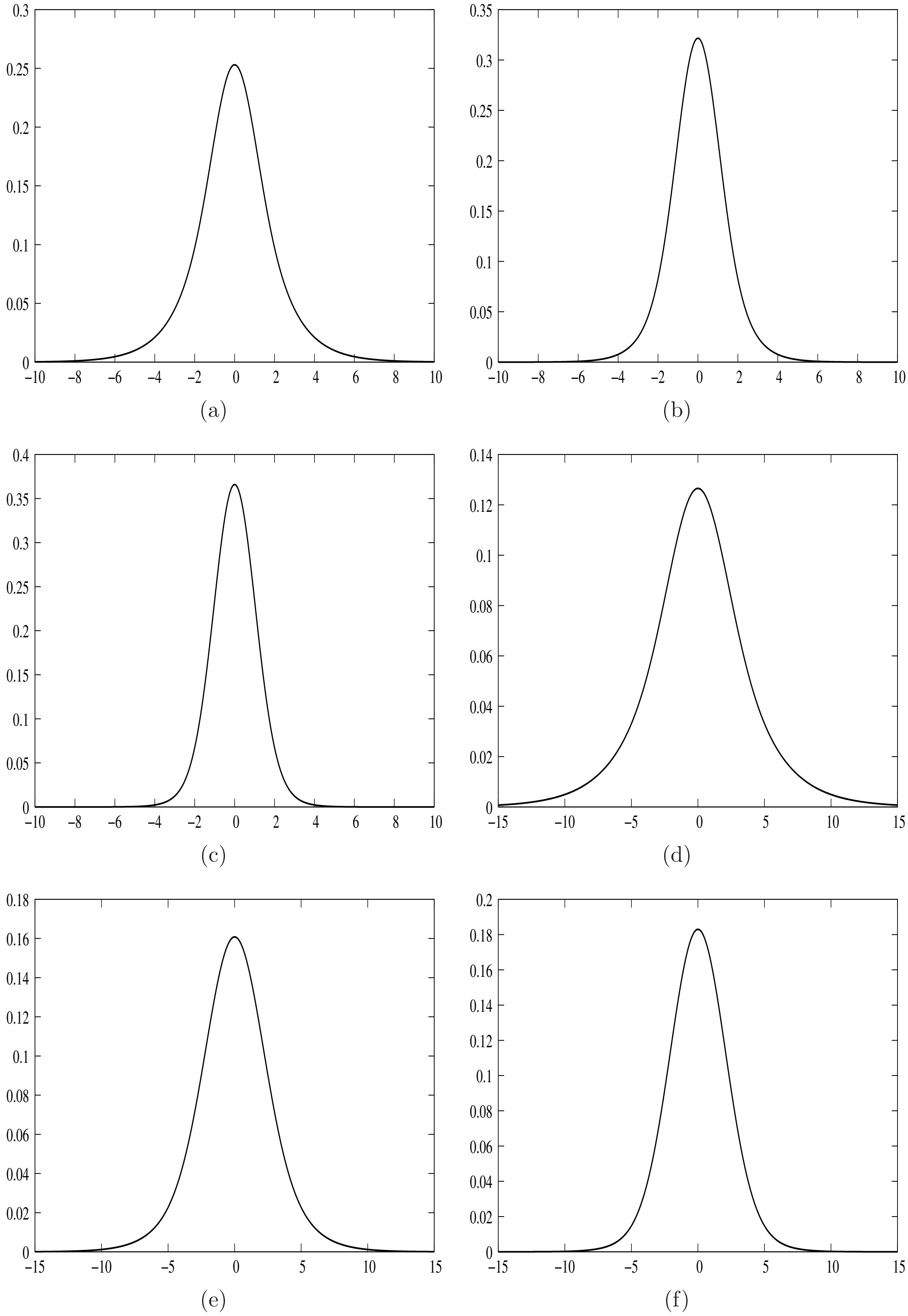

Figure 2: PDF plots of SGSN(σ, p) distribution for different (σ, p) values: (a) (1.0,0.25) (b) (1.0,0.50) (c) (1.0,0.75) (d) (2.0,0.25) (e) (2.0,0.50) (f) (2.0,0.75).

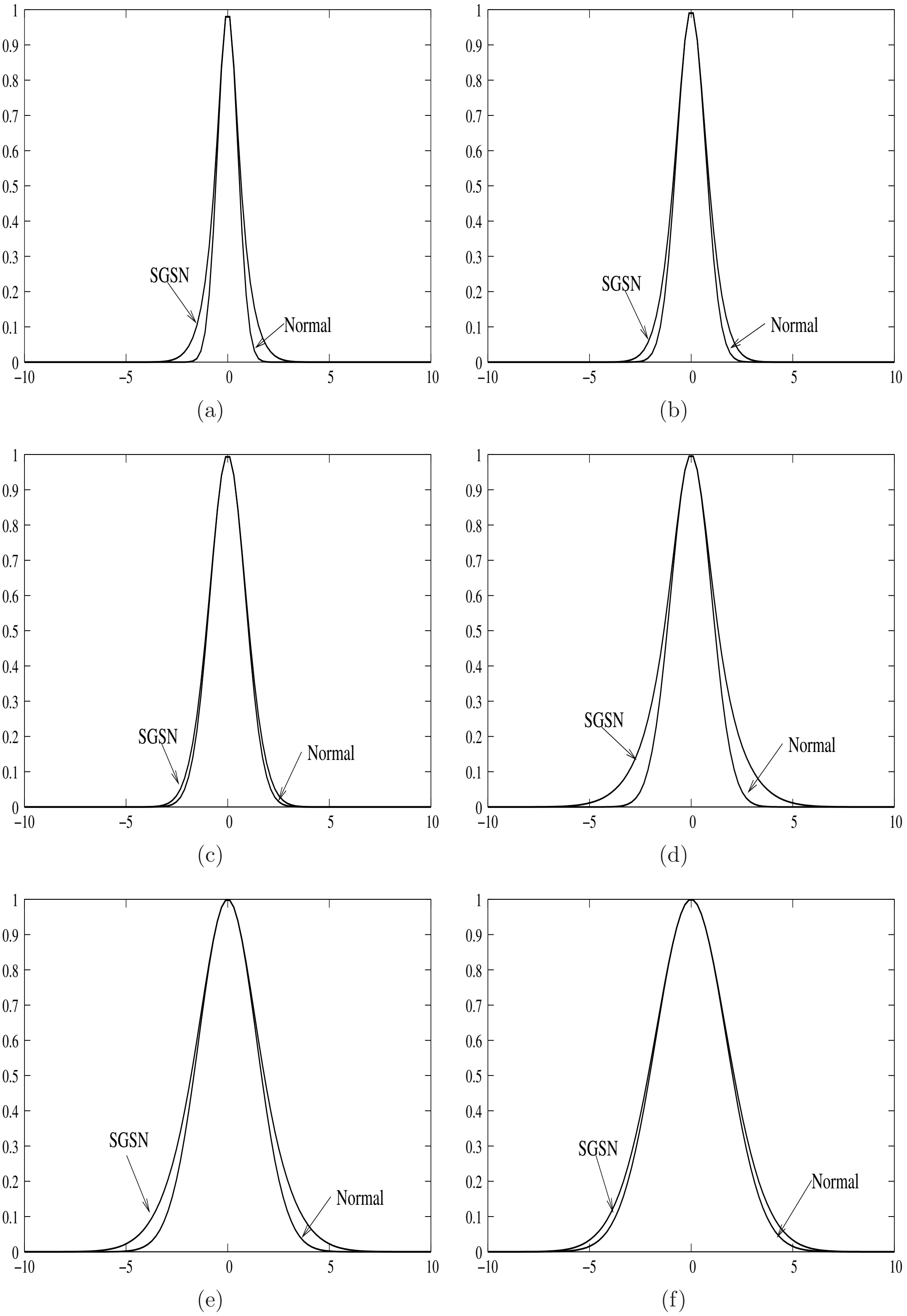

Figure 3: CHF plots of SGSN(σ, p) distribution for different (σ, p) values: (a) (1.0,0.25) (b) (1.0,0.50) (c) (1.0,0.75) (d) (0.5,0.25) (e) (0.5,0.50) (f) (0.5,0.75).

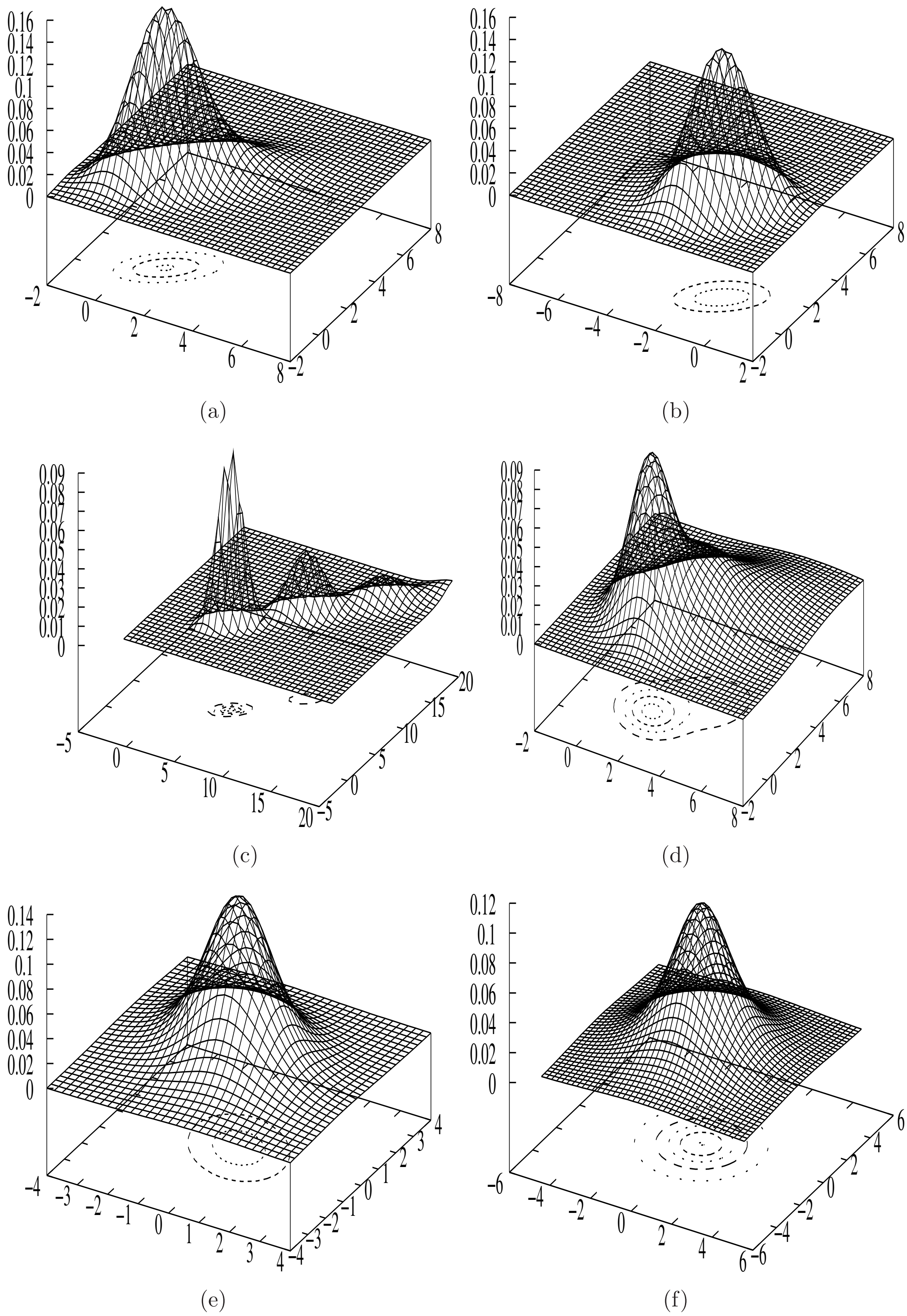

Figure 4: PDF plots of $\mathrm{MGSN}_2(\boldsymbol{\mu}, \Sigma, p)$ distribution for different $(\mu_1, \mu_1, \sigma_{11}, \sigma_{22}, \sigma_{12}, p)$ values: (a) (1.0,1.0,1.0,0.5,0.75) (b) (-1.0,1.0,1.0,1.0,0.5,0.75) (c) (5,5,1.0,1.0,0.5,0.5) (d) (1.5,1.5,1.0,1.0,-0.25,0.5) (e) (0.0,0.0,1.0,1.0,-0.25,0.5) (f) (0.0, 0.0, 1.0, 1.0, 0.0, 0.25).

n		$p = 0.2$	$p = 0.4$	$p = 0.6$	$p = 0.8$
25	β_1	1.0063	1.0044	0.9998	0.9979
		(0.1348)	(0.0824)	(0.0612)	(0.0487)
	β_2	2.0006	1.9964	2.0010	2.0062
		(0.0952)	(0.0568)	(0.0485)	(0.0400)
	σ	0.9274	0.9286	0.9356	0.9398
		(0.0264)	(0.0243)	(0.0238)	(0.0229)
	p	0.1966	0.3878	0.5729	0.7714
		(0.0014)	(0.0061)	(0.0173)	(0.0207)
50	β_1	1.0043	1.0008	0.9987	1.0013
		(0.0502)	(0.0316)	(0.0227)	(0.0197)
	β_2	2.0011	1.9995	2.0001	2.0058
		(0.0703)	(0.0427)	(0.0331)	(0.0269)
	σ	0.9508	0.9323	0.9389	0.9428
		(0.0173)	(0.0211)	(0.0212)	(0.0199)
	p	0.1987	0.3889	0.5787	0.7899
		(0.0008)	(0.0048)	(0.0143)	(0.0178)
75	β_1	0.9917	1.0001	0.9997	1.0001
		(0.0422)	(0.0217)	(0.0208)	(0.0146)
	β_2	2.0005	2.0005	2.0003	2.0041
		(0.0367)	(0.0226)	(0.0193)	(0.0149)
	σ	0.9615	0.9576	0.9545	0.9758
		(0.0165)	(0.0142)	(0.0128)	(0.0111)
	p	0.1998	0.3976	0.5889	0.7987
		(0.0004)	(0.0028)	(0.0013)	(0.0011)
100	β_1	0.9993	1.0021	1.0016	0.9997
		(0.0242)	(0.0163)	(0.0127)	(0.0095)
	β_2	1.9992	1.9987	1.9970	2.0002
		(0.0261)	(0.0147)	(0.0126)	(0.0100)
	σ	0.9893	0.9892	0.9888	0.9917
		(0.0056)	(0.0058)	(0.0059)	(0.0056)
	p	0.2001	0.3999	0.5998	0.7999
		(0.0001)	(0.0016)	(0.0008)	(0.0005)

Table 2: ML estimates of β_1, β_2, σ and the associated MSEs (reported below within brackets) for different sample sizes and for different p values.