Generating the Skew Normal random variable
Dariush Ghorbanzadeh, Philippe Durand, Luan Jaupi

To cite this version:

HAL Id: hal-02465107
https://hal-cnam.archives-ouvertes.fr/hal-02465107
Submitted on 5 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Generating the Skew Normal Random Variable

Dariush Ghorbanzadeh, Philippe Durand, Luan Jaupi

Abstract—In this paper, for the generating the Skew normal random variables, we propose a new method based on the combination of minimum and maximum of two independent normal random variables. The estimation of parameters using the maximum likelihood estimation and the methods of moments estimation method. A real data set has been considered to illustrate the practical utility of the paper.

keyword: Skew normal distribution, maximum likelihood estimation, methods of moments estimation.

I. INTRODUCTION

The univariate Skew Normal distribution has been studied by Azzalini [1], Azzalini [2], Henze [3] and Pewsey [4], has been studied and generalized extensively. The skew normal distribution family is well known for modeling and analyzing skewed data. It is the distribution family that extends the normal distribution family by adding a shape parameter to regulate the skewness, which has the higher flexibility in fitting a real data where some skewness is present.

The density function of the Skew Normal distribution with parameters \((\xi, \tau, \theta)\) is given by

\[ f_X(x) = \frac{2}{\tau} \varphi\left(\frac{x-\xi}{\tau}\right) \Phi\left(\theta \frac{x-\xi}{\tau}\right) \]

where \(\xi\) is location parameter, \(\tau > 0\) is scale parameters, \(\theta\) is asymmetry parameter and \(\varphi\) and \(\Phi\) are, respectively, the density and cumulative distribution functions of the standard normal distribution \(N(0,1)\). In this paper we use the notation \(SN(\xi, \tau, \theta)\) to denote this distribution.

For the simulation of \(SN(0,1,\theta)\) distribution, Henze [3], in his paper showed that, if \(U_1\) and \(U_2\) are identically and independently distributed \(N(0,1)\) random variables, then \(\theta U_1 + U_2 \sqrt{1+\theta^2}\) has the \(SN(0,1,\theta)\) distribution.

For the simulation of \(SN(0,1,\theta)\) distribution, recently Ghorbanzadeh et al. [5], have developed a method, called the Min-Max method, which consists in taking the combination of minimum and maximum of two independent random variables distributed \(N(0,1)\). In this paper, for the simulation of \(SN(\xi, \tau, \theta)\) distribution, we will use the Min-Max method.

II. GENERATING RANDOM VARIABLES BY MIN-MAX METHOD

Let \(m = \sqrt{\frac{1+\theta^2}{2}}\), \(U_1\) and \(U_2\) two independent and identically distributed \(N(m, \tau^2)\) random variables and \(U = \max(U_1, U_2)\) and \(V = \min(U_1, U_2)\). For simulation of the random variable \(X \sim SN(\xi, \tau, \theta)\), we take the combination of \(U\) and \(V\). First note that:

- If \(\theta = 0\), the density (1) becomes: \(\frac{1}{\tau} \varphi\left(\frac{x-\xi}{\tau}\right)\), simply simulate \(X \sim N(\xi, \tau^2)\).
- If \(\theta = 1\), the density (1) becomes:

\[ \frac{2}{\tau} \varphi\left(\frac{x-\xi}{\tau}\right) \left(1 - \Phi\left(\frac{x-\xi}{\tau}\right)\right) \]

we take \(X = \min(X_1, X_2)\).
- If \(\theta = -1\), the density (1) becomes:

\[ \frac{2}{\tau} \varphi\left(\frac{x-\xi}{\tau}\right) \Phi\left(\frac{x-\xi}{\tau}\right) \]

we take \(X = \max(X_1, X_2)\).

where \(X_1\) and \(X_2\) are identically and independently distributed \(N(\xi, \tau^2)\) random variables.

For \(\theta \not\in \{-1,0,1\}\), note :

\[ \lambda_1 = \frac{1+\theta}{\sqrt{2(1+\theta^2)}} \quad \lambda_2 = \frac{1-\theta}{\sqrt{2(1+\theta^2)}} \]

We note that: \(\lambda_1^2 + \lambda_2^2 = 1\). For simulation of the random variable \(X \sim SN(\xi, \tau, \theta)\), we take the combination of \(U\) and \(V\) in the form:

\[ X = \lambda_1 U + \lambda_2 V \]

(3)

Proposition. The random variable \(X\) defined in the equation (3) has the Skew Normal distribution \(SN(\xi, \tau, \theta)\).

Proof: The pair \((U,V)\) has density:

\[ f_{U,V}(u,v) = \frac{2}{\tau^2} \varphi\left(\frac{u-m}{\tau}\right) \varphi\left(\frac{v-m}{\tau}\right) \mathbb{I}_{[u\leq v]}(u,v) \]

(4)

where \(\mathbb{I}\) is the indicator function.

Consider the transformation:

\[ x = \lambda_1 u + \lambda_2 v, \quad y = \lambda_1 u \].

The inverse transform is defined by:

\[ u = \frac{y}{\lambda_1} - \frac{x}{\lambda_2}, \quad v = \frac{x-y}{\lambda_2} \]

and the corresponding Jacobian is: \(J = \frac{1}{|\lambda_1 \lambda_2|}\). X density is defined by:

\[ f(x) = \frac{1}{|\lambda_1 \lambda_2|} \int_{\Delta} f_{U,V}\left(\frac{y}{\lambda_1}, \frac{x-y}{\lambda_2}\right) dy \]

\[ = \frac{2}{\tau^2 |\lambda_1 \lambda_2|} \int_{\Delta} \varphi\left(\frac{y-m_1}{\lambda_1 \tau}\right) \varphi\left(\frac{y-x+m \lambda_2}{\lambda_2 \tau}\right) dy \]

(5)

where \(\Delta = \left\{(x-y) \leq \frac{y}{\lambda_1}\right\}\).

For the density of the standard normal distribution \(N(0,1)\),

---

Dariush Ghorbanzadeh, Philippe Durand, Luan Jaupi are in the département Mathématiques-Statistiques, Conservatoire National des Arts et Métiers, 292 rue Saint martin, 75141 Paris FRANCE e-mail: dariush.ghorbanzadeh@cnam.fr, philippe.durand@cnam.fr, leonjaupi@cnam.fr
we have the following classical property.
\[
\varphi\left(\frac{x - \mu_1}{\sigma_1}\right) \varphi\left(\frac{x - \mu_2}{\sigma_2}\right) = \\
\varphi\left(\frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} \left(\frac{x - \mu_1 \sigma_2^2 + \mu_2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)\right) \varphi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)
\]

By using relation (6), the relation (5) becomes
\[
f(x) = \frac{2}{\tau^2 |\lambda_1 \lambda_2|} \varphi\left(\frac{x - m(\lambda_1 + \lambda_2)}{\tau}\right) \\
\times \int_{\Delta} \varphi\left(\frac{1}{\lambda_1 \lambda_2 \tau} \left(y - x \lambda_1^2 - m \lambda_1 \lambda_2 (\lambda_2 - \lambda_1)\right)\right) dy
\]

Taking account of \(m(\lambda_1 + \lambda_2) = \xi\), we obtain,
\[
f(x) = \frac{2}{\tau^2 |\lambda_1 \lambda_2|} \varphi\left(\frac{x - \xi}{\tau}\right) \\
\times \int_{\Delta} \varphi\left(\frac{1}{\lambda_1 \lambda_2 \tau} \left(y - x \lambda_1^2 - m \lambda_1 \lambda_2 (\lambda_2 - \lambda_1)\right)\right) dy
\]

For the domain \(\Delta\), we have the following three cases:
case 1 \(\theta \in (-1, 0) \cup (0, 1)\), we have:
\[
|\lambda_1 \lambda_2| = \left|\frac{1 - \theta^2}{2(1 + \theta^2)}\right| \text{ and } \Delta = \{y \geq \frac{\lambda_1 x}{\lambda_1 + \lambda_2}\}
\]
case 2 \(\theta < -1\), we have: \(|\lambda_1 \lambda_2| = \frac{\theta^2 - 1}{2(1 + \theta^2)}\) and
\[
\Delta = \{y \geq \frac{\lambda_1 x}{\lambda_1 + \lambda_2}\}
\]
case 3 \(\theta > 1\), we have: \(|\lambda_1 \lambda_2| = \frac{\theta^2 - 1}{2(1 + \theta^2)}\) and
\[
\Delta = \{y \leq \frac{\lambda_1 x}{\lambda_1 + \lambda_2}\}
\]
Using equation (8) and the three cases above, we get the result.

III. Inference

Let \(X_1, \ldots, X_n\) be a sample of size \(n\) from a \(\mathcal{SN}(\xi, \tau, \theta)\) distribution.

A. Maximum likelihood Estimation

The log-likelihood is given by
\[
\ell(\xi, \tau, \theta) = n \log 2 - n \log \tau + \sum_{i=1}^{n} \log \left(\varphi\left(\frac{x_i - \xi}{\tau}\right)\right) + \\
\sum_{i=1}^{n} \log \left(\Phi\left(\frac{x_i - \xi}{\tau}\right)\right)
\]
The maximum likelihood estimators (MLE) of \((\xi, \tau, \theta)\), denoted \((\hat{\xi}, \hat{\tau}, \hat{\theta})\), are the numerical solution of the system of equations :
\[
\begin{align*}
\tau^2 & = \frac{1}{n} \sum_{i=1}^{n} (x_i - \xi)^2 \\
\tau \theta & = \frac{n}{\sum_{i=1}^{n} \varphi\left(\frac{x_i - \xi}{\tau}\right)} = \sum_{i=1}^{n} \varphi\left(\frac{x_i - \xi}{\tau}\right) \\
\sum_{i=1}^{n} \frac{(x_i - \xi) \varphi\left(\frac{x_i - \xi}{\tau}\right)}{\Phi\left(\frac{x_i - \xi}{\tau}\right)} & = 0
\end{align*}
\]

B. Methods of Moments Estimation

Let \(m_k\) be the centered moment of order \(k\) of data defined by
\[
m_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^k
\]
where \(\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i\), and denote by \(\gamma_1\) the skewness coefficient defined by \(\gamma_1 = \frac{m_4}{m_2^{3/2}}\).

The method of moment estimators (MME) of \((\xi, \tau, \theta)\), denoted \((\hat{\xi}, \hat{\tau}, \hat{\theta})\), are obtained by solving the set of three equations:
\[
\begin{align*}
\xi + a \delta & = \bar{x}_n \\
\tau^2 (1 - a^2 \delta^2) & = m_2 \\
\frac{b \delta^3}{(1 - a^2 \delta^2)^{3/2}} & = \gamma_1
\end{align*}
\]
where \(a = \sqrt{\frac{2}{\pi}}, b = \frac{4 - \pi}{2}\) and \(\delta = \frac{\theta}{\sqrt{1 + \theta^2}}\).

We note that the sign of \(\delta\) is the same as the sign of \(\gamma_1\). The equation (10c) admits the solution
\[
\delta = \text{sign}(\gamma_1) \sqrt{\frac{1}{\alpha^2 (b^2/3 + |\gamma_1|^{2/3})}}
\]
where \( \text{sign}(\gamma_1) \) is the sign of \( \gamma_1 \). Consequently,

\[
\begin{align*}
\tilde{\theta} &= \frac{\tilde{\delta}}{1 - \tilde{\delta}^2} \\
\tilde{\tau} &= \sqrt{\frac{m_2}{1 - a^2\tilde{\delta}^2}} \\
\tilde{\xi} &= \bar{x} - a\tilde{\tau}\tilde{\delta}
\end{align*}
\]

(12)

IV. SIMULATION RESULTS

In order to study the performance of the method, we simulated 100-samples of sizes \( n = 5000 \). For each sample, we calculated the parameter estimators, MLE and MME, the Table I summarizes the results obtained for different values of \( (\xi, \tau, \theta) \).

V. APPLICATION

In this section, we illustrate the use of the estimation procedures described in the previous section. The variable to be considered is the average length of stay for patients who are in hospital for acute care because of problems, hepatobiliary system and pancreas, and die for this cause. The sample, under study, corresponds to 1082 hospitals in 10 states of the United States. For more information see columns 4 in http://lib.stat.cmu.edu/data-expo/1997/ascii/p07.dat. The data were used by Elal-Olivero [6] by modeling with a model "alpha-Skew-Normal distribution". Table II summarizes the results obtained by the two methods of estimation.

TABLE I: Estimators of \( (\xi, \tau, \theta) \) by the MLE and MME methods.

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>MLE</th>
<th>MME</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi )</td>
<td>4.1192</td>
<td>3.8856</td>
</tr>
<tr>
<td>( \tau )</td>
<td>2.3006</td>
<td>2.4731</td>
</tr>
<tr>
<td>( \theta )</td>
<td>2.0608</td>
<td>2.6286</td>
</tr>
</tbody>
</table>

By the MLE method, we get the same results as Elal-Olivero [6]. The results obtained by the MME method are similar as the results obtained by the MLE method.

VI. CONCLUSION

The simulation method proposed in this paper is simpler to use than the three methods, the inverse transform method, the composition method and the acceptance-rejection method. The results of estimation of the parameters, by the two estimation methods, from the simulations are very satisfactory. For the application, we obtain the same estimation values as other author using the same data.

REFERENCES

### TABLE II: Statistics of estimators of $(\xi, \tau, \theta)$ for size $n = 5000$.

<table>
<thead>
<tr>
<th></th>
<th>$\xi = -6, \tau = 8, \theta = -7$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>MME</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean of $\xi$</td>
<td>mean of $\xi$</td>
<td>mean of $\tau$</td>
<td>mean of $\theta$</td>
<td>mean of $\theta$</td>
</tr>
<tr>
<td></td>
<td>std of $\xi$</td>
<td>std of $\tau$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
</tr>
<tr>
<td>MLE</td>
<td>-6.0017</td>
<td>0.2109</td>
<td>8.0045</td>
<td>0.2930</td>
<td>-7.2612</td>
</tr>
<tr>
<td></td>
<td>std of $\xi$</td>
<td>std of $\tau$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
</tr>
<tr>
<td>MME</td>
<td>-6.1489</td>
<td>0.8805</td>
<td>7.9178</td>
<td>0.4288</td>
<td>-7.5284</td>
</tr>
<tr>
<td></td>
<td>std of $\xi$</td>
<td>std of $\tau$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
</tr>
<tr>
<td></td>
<td>$\xi = 2, \tau = 5, \theta = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>MME</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean of $\xi$</td>
<td>mean of $\xi$</td>
<td>mean of $\tau$</td>
<td>mean of $\theta$</td>
<td>mean of $\theta$</td>
</tr>
<tr>
<td></td>
<td>std of $\xi$</td>
<td>std of $\tau$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
</tr>
<tr>
<td>MLE</td>
<td>2.0025</td>
<td>0.2396</td>
<td>4.9891</td>
<td>0.2382</td>
<td>3.0745</td>
</tr>
<tr>
<td></td>
<td>std of $\xi$</td>
<td>std of $\tau$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
</tr>
<tr>
<td>MME</td>
<td>2.0054</td>
<td>0.0788</td>
<td>5.0546</td>
<td>0.0795</td>
<td>3.2742</td>
</tr>
<tr>
<td></td>
<td>std of $\xi$</td>
<td>std of $\tau$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
<td>std of $\theta$</td>
</tr>
</tbody>
</table>