On Some Properties and Estimation of a Skew Symmetric Density Function

M. Towhidi* and M. Shaghaghian

Shiraz University

Abstract. In this paper we consider a general setting of skew-symmetric distribution which was constructed by Azzalini (1985), and its properties are presented. A suitable empirical estimator for a skew-symmetric distribution is proposed. In data analysis, by comparing this empirical model with the estimated skew-normal distribution, we show that the proposed empirical model has a better fit in density estimation, via some simulations.

Keywords. Skew-distribution; Skew-normal; Skewness; Kernel estimation.

1 Introduction

A family of skew-symmetric distributions was introduced by Azzalini (1985), in the following form

$$\varphi(z; \lambda) = 2\phi(z)\Phi(\lambda z); \qquad -\infty < z < \infty \tag{1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and cumulative distribution function, respectively. He called this distribution "skew-normal distribution". The parameter λ varies in the real line and composes the skewness.

Some authors used other symmetric distributions like Student's t, Cauchy, Laplace, and Logistic for constructing some skew-symmetric distributions, (see e.g., Gupta et al., 2002).

---

* Corresponding author

A general form of skew distributions generated by the normal kernel was introduced by Nadarajah and Kotz (2003). This kind of skew distributions are in the form $2f(x)G(\lambda x)$, where $f(\cdot)$ is taken to be the standard normal density function while the cumulative distribution function $G(\cdot)$ is taken to come from Student's t, Uniform, Cauchy.

This paper is concerned with another general class of skew distributions of the form

$$f_X(x) = 2f(x)G(\lambda x); \qquad -\infty < x < \infty \qquad (2)$$

where $f(\cdot)$ is any symmetric density function and $G(\cdot)$ is a cumulative distribution function (cdf) of a symmetric random variable. The construction of these skew-symmetric distributions is due to Azzalini (1985). Obviously this class contains skew-normal distribution and the other normal kernel distributions.

These classes of skewed distributions are suitable for the analysis of data with unimodal empirical distribution functions having some skewness, which is often occurring in data analysis. Also these distributions can be used in studying robustness and Bayesian estimation (see e.g., Arnold et al., 1983, and Mukhopadhyay and Vidacovic, 1995).

In section 2, we present the structural and important properties of a skew-symmetric distribution with density (2). A method for constructing a suitable empirical skew density function in the special form of (2) is proposed in section 3. Then, we show that in data analysis, sometimes, the model (2) is more appropriate than (1) as a density function of the data. Also, this preference is shown via some simulations.

2   Some Properties of the General Skew-symmetric Distribution

A random variable $X_{f,G,\lambda}$ is said to have a skew distribution with functional parameters f, G and fixed parameter λ, if its probability density function (pdf) is given by

$$f_X(x) = 2f(x)G(\lambda x); \qquad -\infty < x < \infty$$

where f and g are two symmetric pdfs and F and G are their cdfs. The parameter λ varies in the real line and regulate the skewness. We use the notation $SD(f, G, \lambda)$ for this distribution and we can extend the above defi-

nition to a location-scale family.

Definition 1. The location-scale skew distribution is defined as that of $Z = \mu + \sigma X$, where $X \sim SD(f, G, \lambda)$, $\mu \in \mathbb{R}$ and $\sigma > 0$. Its density function is given by

$$f_Z(z) = \frac{2}{\sigma} f\left(\frac{z - \mu}{\sigma}\right) G\left(\lambda \frac{z - \mu}{\sigma}\right); \qquad \forall z \in \mathbb{R} \tag{3}$$

we denote this distribution by $SD(\mu, \sigma; f, G, \lambda)$.

In the following propositions, we present some structural properties of the $SD(f, G, \lambda)$.

Proposition 1. The following are two representations of a $SD(f, G, \lambda)$ random variable:

1. Let U be a random variable with pdf $f(\cdot)$ and conditionally on $U = u$, $S_U = 1$ with probability $G(\lambda u)$ and $S_U = -1$ with probability $1 - G(\lambda u)$, Then $X = S_U U \sim SD(f, G, \lambda)$.

    This result is due to Azzalini (1986).

2. Let U and V be two independent random variables with symmetric pdfs f, g, and cdfs F, G, respectively. Then the conditional distribution of U given $V < \lambda U$ is $SD(f, G, \lambda)$.

Proof. We only prove part (2) of the proposition. Under the assumptions of part (2), we can find the conditional distribution of U given $V < \lambda U$,

$$\begin{aligned}
f_{U|V<\lambda U}(u) &= \frac{\int_{-\infty}^{\lambda u} f(u)g(v)dv}{\mathrm{P}(V < \lambda U)} \\
&= \frac{f(u)G(\lambda u)}{\mathrm{P}(V < \lambda U)} \\
&= \frac{f(u)G(\lambda u)}{\int_{-\infty}^{+\infty} \mathrm{P}(V < \lambda u)f(u)du} \\
&= \frac{f(u)G(\lambda u)}{\int_{-\infty}^{+\infty} G(\lambda u)f(u)du} \\
&= 2f(u)G(\lambda u).
\end{aligned}$$

It can be easily found the following properties of $SD(f, G, \lambda)$:

Proposition 2. Let U and V be two independent random variables with

symmetric pdfs f, g, and cdfs F, G, respectively and $X_{f,G,\lambda}$ has a $SD(f, G, \lambda)$. We can conclude the following results:

(a) $X_{f,G,0} \overset{d}{=} U$

(b) $-X_{f,G,\lambda} \overset{d}{=} X_{f,G,-\lambda}$

(c) $|X_{f,G,\lambda}| \overset{d}{=} |U|$

(d)   If λ tends to $+\infty$, then $X_{f,G,\lambda}$ converges to $|U|$ in distribution. Also, if λ tends to $-\infty$, then $X_{f,G,\lambda}$ converges to $-|U|$ in distribution.

(e)   The distribution of $T_1 = \min(U, V)$ is a mixture of the distributions $SD(f, G, -1)$ and $SD(g, F, -1)$ with probability $\frac{1}{2}$.

(f)   The distribution of $T_2 = \max(U, V)$ is a mixture of the distributions $SD(f, G, 1)$ and $SD(g, F, 1)$ with probability $\frac{1}{2}$.

It is useful to present a new representation of the pdf (2) such that some common measures of skewness exactly appear in the visual form of the density function.

Remark 1.   In the pdf (2), λ is a measure of skewness and can be for example Pearson's skewness coefficients or AG measure (Arnold and Groeneveld, 1995) defined as:
$$AG = 2P(X > \text{mode}) - 1.$$

In this paper, for simplicity, we define a new measure for calculating the skewness of the distribution of a random variable X as

$$\begin{aligned}
\lambda' &= P(X - \mu > 0) - P(X - \mu < 0) \\
&= 2P(X - \mu > 0) - 1
\end{aligned} \tag{4}$$

where $\mu = E(X)$. λ' takes values in $[-1, 1,]$ and has an intuitive interpretation for unimodal distributions. λ' takes a negative (positive) values for left (right) skewed distributions and is equal to zero for symmetric distributions.

According to Remark 1, we can state $f_X(x)$ in (2) on the basis of λ', in the form:
$$f_X(x) = 2f(x)G_1(\lambda' x); \qquad \forall x. \tag{5}$$

where $G_1(\cdot)$ is a cdf which is defined as:

$$G_1(x) = G(\frac{\lambda}{\lambda'} x).$$

3 Empirical Skewed Version of a Symmetric Distribution

This section is allocated to present a method for constructing a suitable empirical skew density function to analyze the data. An appropriate estimate for the pdf $f_X(x)$ in (4) (with $G_1 = F$, the cdf of $f(\cdot)$) can be found by introducing logical estimates $\hat{f}_n(\cdot)$, $\hat{\lambda}'$,

$$\hat{f}_X(x) = 2\hat{f}_n(x)\hat{F}_n(\hat{\lambda}'x). \tag{6}$$

Let $X_1, \ldots, X_n$ be a random sample from pdf $f_X(x)$, then the skewness parameter λ', which is defined in (4), can be estimated from the data by:

$$\hat{\lambda}' = \frac{2}{n} \sum_{i=1}^{n} I_{(0,+\infty)}(X_i - \bar{X}) - 1 \tag{7}$$

where $I(\cdot)$ is the indicator function of the set A.

Also, a natural estimator for the pdf $f(\cdot)$ in (5), can be derived by a kernel method (Silverman, 1986), with the normal density as a kernel function.

According to part (c) of proposition (2), the pdf of the random variable $Z = |X|$ is:

$$f_Z(z) = 2f(z); \qquad z \geqslant 0$$

and a kernel estimate of the density function $f_Z(\cdot)$ according to the random sample $Z_1 = |X_1|, \ldots, Z_n = |X_n|$ is:

$$\hat{f}_Z(z) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{z - z_i}{h}\right)$$

where $K(\cdot)$ is a normal kernel function and h is an smoothing parameter.

Hence a kernel estimate of $f(\cdot)$ is as follows:

$$\hat{f}_n(x) = \begin{cases} \frac{1}{2}\, \hat{f}_Z(x) & \text{if } x > 0, \\[2mm] \frac{1}{2}\, \hat{f}_Z(-x) & \text{if } x < 0. \end{cases} \tag{8}$$

Finally an appropriate estimate of $F(\cdot)$, the cdf of $f(\cdot)$, is:

$$\hat{F}_n(x) = \int_{-\infty}^{x} \hat{f}_n(t)\, dt. \tag{9}$$

www.SID.ir

In two examples below, we compare two estimated skewed version of symmetric distributions:

$$\begin{cases} \hat{f}_X^1(x) = 2\phi(x)\,\Phi(\hat{\lambda}x), \\ \hat{f}_X^2(x) = 2\hat{f}_n(x)\hat{F}_n(\hat{\lambda}'x). \end{cases}$$

where $\hat{\lambda}$ is the MLE of λ in the pdf(1), and $\hat{\lambda}'$, $\hat{f}_n$, $\hat{F}_n$, are defined in (7), (8), (9), respectively.

From these examples, we can conclude that in some situations, the model $\hat{f}_X^2(x)$ is preferred over $\hat{f}_X^1(x)$ for estimating the density function of X .

Example 1. Arellano-Valle et al. (2004) analyze the data concerning the height of 100 Australian athletes. They claimed fitting a $SN(\lambda_1)$ model to the data would be inadequate, and introduced a new class of generalized skew-normal distribution by defining a parameter $\lambda_2 > 0$. Now, we want to compare the two fitted density functions $\hat{f}_X^1(x)$, $\hat{f}_X^2(x)$, with each other on the basis of the data. Note that the MLEs of parameters μ, σ, λ are:

$$\hat{\mu} = 174.58, \qquad \hat{\sigma} = 8.201, \qquad \hat{\lambda} = 0.0016,$$

and $\hat{\lambda}'$ can be calculated from the relation (7), $\hat{\lambda}' = 0.02$.

Figure 1 shows plots of the relative frequency curve, $f_{s-\text{plus}}(x)$, and the estimated pdfs $\hat{f}_X^1(x)$ and $\hat{f}_X^2(x)$. Also, for the sake of comparison, we have used two criteria

$$\sup_{x_i} |f_{s-\text{plus}}(x_i) - \hat{f}_X(x_i)| \qquad \text{and} \qquad \frac{1}{n}\sum_{i=1}^n (f_{s-\text{plus}}(x_i) - \hat{f}_X(x_i))^2.$$

$$\sup_{x_i} |f_{s-\text{plus}}(x_i) - \hat{f}_X^1(x_i)| = 0.009350,$$
$$\sup_{x_i} |f_{s-\text{plus}}(x_i) - \hat{f}_X^2(x_i)| = 0.0066180.$$

$$\frac{1}{n}\left[\sum_{i=1}^n \left\{f_{s-\text{plus}}(x_i) - \hat{f}_X^1(x_i)\right\}^2\right] = 0.00001637776,$$
$$\frac{1}{n}\left[\sum_{i=1}^n \left\{f_{s-\text{plus}}(x_i) - \hat{f}_X^2(x_i)\right\}^2\right] = 8.1719 \times 10^{-6}.$$

Table 1. The relative results of Example 2

	Degrees of Freedom	
	3	5
$\sup\limits_{x_i}\|f_{\mathrm{s-plus}}(x_i) - \hat{f}_X^1(x_i)\|$	0.034707	0.0239784
$\sup\limits_{x_i}\|f_{\mathrm{s-plus}}(x_i) - \hat{f}_X^2(x_i)\|$	0.021089	0.017344
$\frac{1}{n}[\sum\limits_{i=1}^{n}\{f_{\mathrm{s-plus}}(x_i) - \hat{f}_X^1(x_i)\}^2]$	8.53326×10^{-6}	5.7552678×10^{-6}
$\frac{1}{n}[\sum\limits_{i=1}^{n}\{f_{\mathrm{s-plus}}(x_i) - \hat{f}_X^2(x_i)\}^2]$	7.390103×10^{-6}	4.7391424×10^{-6}

Example 2.   We have generated the data sets from two chi-square distributions with distinct degrees of freedom. Figure 2 and 3 show the graphs of $f_{\mathrm{s-plus}}(x)$ and the estimated pdfs $\hat{f}_X^1(x)$ and $\hat{f}_X^2(x)$.

Table 1 presents the results of comparing the two fitted density functions with the relative frequency function, $f_{\mathrm{s-plus}}(x)$, via some reasonable criteria.
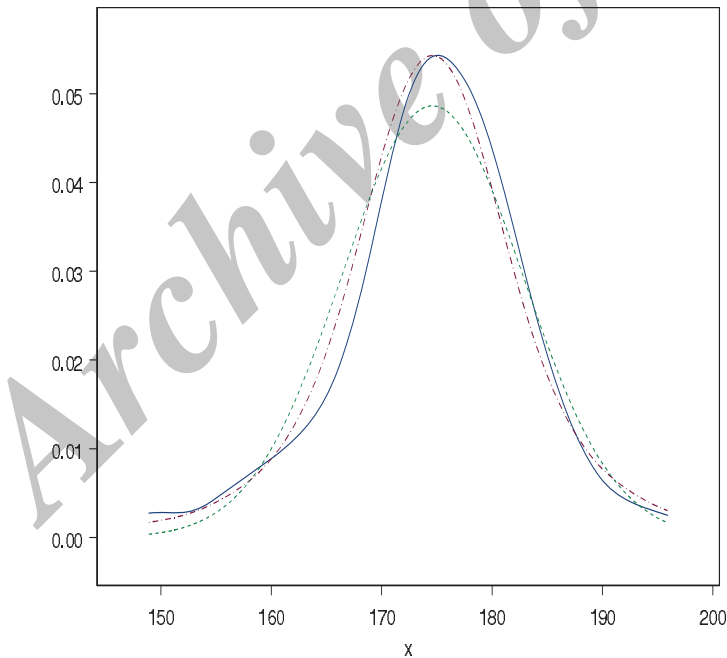

Figure 1. The relative frequency curve for the height of 100 Australian athletes (solid line) and plots of the fitted densities $\hat{f}_X^1(x)$ (dotted line), $\hat{f}_X^2(x)$ (dashed line)
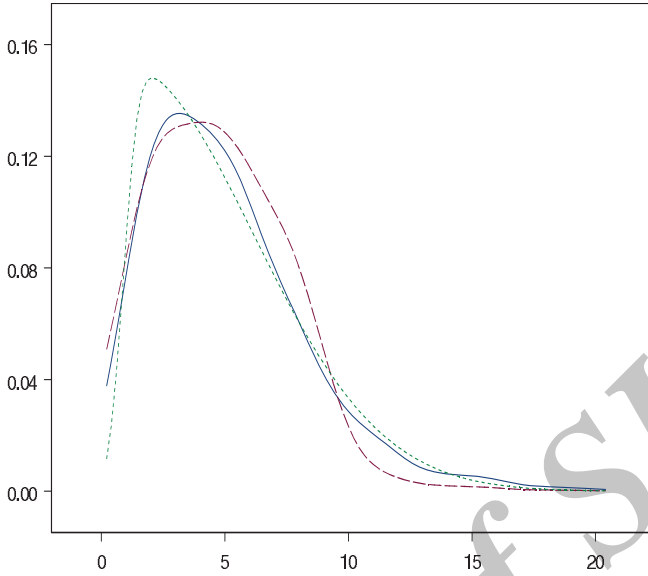
Figure 2.   The graphs of $f_{\mathrm{s-plus}}(x)$ (solid line), $\hat{f}_X^1(x)$ (dotted line), and $\hat{f}_X^2(x)$ (dashed line) for generated data set of chi-square distribution with 3 degrees of freedom
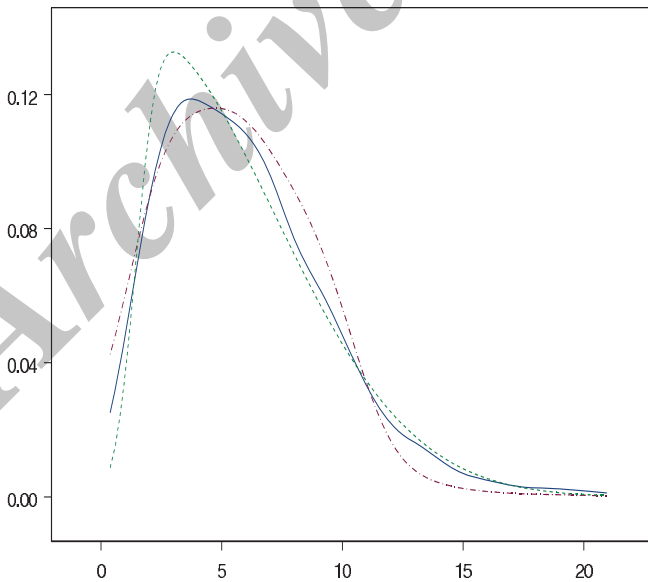

Figure 3. The graphs of $f_{\mathrm{s-plus}}(x)$ (solid line), $\hat{f}_X^1(x)$ (dotted line), and $\hat{f}_X^2(x)$ (dashed line) for generated data set of chi-square distribution with 5 degrees of freedom

Acknowledgements

References

Arellano-Valle, R.B.; Gomez, H.W. and Quintana, F.A. (2004). A new class of skew-normal distributions. *Commun. Statist. Theory Meth.* **33**, 1465-1480.

Arnold, B.C.; Beaver, R.J.; Groeneveld, R.A. and Meeker, W.Q. (1983). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika* **58**, 471-488.

Arnold, B.C. and Groeneveld, R.A. (1995). Measuring skewness with respect to the mode. *The American Statistican* **49**, 34-38.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171-178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones.*Statistica.* **46**, 199-208.

Gupta, A.K.; Chang, F.C. and Huang, W.J. (2002). Some skew-symmetric models. Random Operators. *Stochastic Equations.* **10**, 133-140.

Mukhopadhyay, S. and Vidakovic, B. (1995). Efficiency of linear Bayes rules for a normal mean: skewed priors class. *The Statistician* **44**, 389-397.

Nadarajah, S. and Kotz, S. (2003). Skewed distributions generated by the normal kernel. *Statist. Prob. Letters* **65**, 269-277.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.

M. Towhidi
Department of statistics,
Shiraz University,
Shiraz, Iran.
Po. Box 71454.
e-mail: *mtowhidi@susc.ac.*

M. Shaghaghian
Department of statistics,
Shiraz University,
Shiraz, Iran.
Po. Box 71454.
e-mail: *Shaghaghian.m@gmail.com*