Computational methods applied to a skewed generalised normal family

A. Bekker[a], J.T. Ferreira[a,1], M. Arashi[b], & B.W. Rowland[a]

[a]Department of Statistics, University of Pretoria, Pretoria, South Africa

[b]Department of Statistics, School of Mathematical Sciences, Shahrood University of Technology, Shahrood, Iran

ABSTRACT

Some characteristics of the normal distribution may not be ideal to model in many applications. We develop a skew generalised normal ($\mathcal{SGN}$) distribution by applying a skewing method to a generalised normal distribution, and study some meaningful statistical characteristics. Computational methods to approximate, and a well-constructed efficient computational approach to estimate these characteristics, are presented. A stochastic representation of this distribution is derived and numerically implemented. The skewing method is extended to the elliptical class resulting in a more general framework for skewing symmetric distributions. The proposed distribution is applied in a fitting context and to approximate particular binomial distributions.

Keywords and phrases: Approximating binomial distribution; Skew-normal; Skew generalised normal; Stochastic representation.

2000 *Mathematics Subject Classification*: 62F03; 62F10; 62E99.

1 Introduction

Azzalini (Azzalini 1985) introduced the skew-normal ($\mathcal{SN}$) distribution, which includes the standard normal distribution as a special case, and studied the basic mathematical properties thereof. The skewing methodology that is used to skew existing symmetric probability density functions (PDFs) is stated:

Proposition 1 *(Azzalini 1985) Denote by $f_0(\cdot)$ a probability density function (PDF) on $\mathbb{R}^d$, by $G_0(\cdot)$ a continuous cumulative distribution function (CDF) on $\mathbb{R}$, and by $w(\cdot)$ a real-valued function on $\mathbb{R}^d$, such that $f_0(-x) = f_0(x)$, $w(-x) = -w(x)$ and $G_0(-y) = 1 - G_0(y)$ for all $x \in \mathbb{R}^d$, $y \in \mathbb{R}$. Then*

$$f_X(x) = 2f_0(x)\, G_0\{w(x)\} \tag{1}$$

is a PDF on $\mathbb{R}^d$.

Note that f_0 is termed the *symmetric base PDF*, $2G_0\{w(x)\}$ is termed the *skewing mechanism* and f_X is termed the skewed version of the symmetric base PDF.
Azzalini's innovation of coupling the symmetric component with a skewing mechanism generates distributions with flexible tail behaviour. This methodology provides a platform for considering other candidates rather than the normal distribution as the symmetric component.

(Azzalini 2013) remarked that the $\mathcal{SN}$ distribution has short tails making it unsuitable for use when there is a need for a model to have heavier tails than the normal distribution. One method to solve this problem, as suggested in (Azzalini 2013), is to use a *symmetric base* PDF f_0, (see Proposition 1) with heavier tails than the normal distribution. Following this motivation, this paper focuses on the use of Subbotin's generalised normal distribution (Subbotin 1923) as f_0, which provides enough flexibility to allow for tails heavier than that of the normal distribution. This idea has been briefly discussed in (Salinas, Arellano-Valle, and Gómez 2007; Azzalini 2013) and is used as a point of departure. The purpose of this paper is to enrich and develop theory in this area and provide novel methods of estimating the characteristics of this distribution. Literature on Azzalini's model and related models are widely presented in literature (Azzalini and Regoli 2011; Arellano-Valle, Del Pino, and San Martın 2002; Bahrami and Qasemi 2015; Arnold, Beaver, Azzalini, Balakrishnan, Bhaumik, Dey, Cuadras, and Sarabia 2002; Pourahmadi 2007; Yadegari, Gerami, and Khaledi 2008).

In Section 2 the generalised normal distribution (Subbotin 1923) is given and a stochastic representation is presented. Thereafter, the use of the generalised normal ($\mathcal{GN}$) distribution as a symmetric base (see Proposition 1) gives rise to a skew generalised normal distribution termed the $\mathcal{SGN}$ distribution. Methods to calculate basic

[1]johan.ferreira@up.ac.za

statistical properties of the $\mathcal{SGN}$ and a corresponding stochastic representation is presented in Section 3. In Section 4 a generalised skewing mechanism is coupled with the generalised normal distribution. In Section 5 the $\mathcal{SGN}$ is firstly applied in a distribution fitting context. Thereafter, the $\mathcal{SGN}$ distribution is considered as a candidate to approximate the binomial distribution. Final remarks are made in Section 6.

2 Skew generalised normal distribution

In this section the $\mathcal{GN}$ distribution is revisited and a sampling scheme to generate random variates from this distribution is proposed. Thereafter, the $\mathcal{SGN}$ distribution is focussed on. Three methods of evaluating the characteristics (i.e. expected value, variance, skewness and kurtosis) of this distribution are presented. Finally, a stochastic representation of the $\mathcal{SGN}$ distribution is developed.

Definition 2 *(Subbotin 1923) A random variable X has the generalised normal distribution with location parameter μ and scale parameter α if its PDF is given by*

$$\phi^*(x;\mu,\alpha,\beta) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)}e^{-\left|\frac{x-\mu}{\alpha}\right|^\beta},\ x\in\mathbb{R} \tag{2}$$

where $\Gamma(\cdot)$ denotes the gamma function, $\mu\in\mathbb{R}$ and $\alpha,\beta\in\mathbb{R}^+$. This is denoted by $X\sim\mathcal{GN}(\mu,\alpha,\beta)$.

Following a similar approach in (Azzalini 2013), a stochastic representation of the $\mathcal{GN}(\mu,\alpha,\beta)$ distribution is proposed. This provides a method to generate random numbers from $X\sim\mathcal{GN}(\mu,\alpha,\beta)$.

Theorem 3 *If $X\sim\mathcal{GN}(\mu,\alpha,\beta)$ then,*

$$X=\begin{cases}\mu+\alpha P^{\frac{1}{\beta}} & \text{, with probability } \frac{1}{2}\\ \mu-\alpha P^{\frac{1}{\beta}} & \text{, with probability } \frac{1}{2}\end{cases}$$

where $P\sim Gamma\left(\frac{1}{\beta},1\right)$.

Proof. Let $P=\left|\frac{X-\mu}{\alpha}\right|^\beta$. The property, $\mathbb{P}[X-\mu>0]=\mathbb{P}[X-\mu<0]=\frac{1}{2}$, follows from the symmetry of $X\sim\mathcal{GN}(\mu,\alpha,\beta)$. It follows from (Bain and Engelhardt 1992) that

$$f_P(p)=\sum_{j=1}^{2}f_X\left(u_j^{-1}(p)\right)\left|\frac{d}{dp}u_j^{-1}(p)\right|=\frac{1}{\Gamma\left(\frac{1}{\beta}\right)}e^{-p}z^{\frac{1}{\beta}-1}$$

for $p\in\mathbb{R}^+$, and the result follows that $P\sim Gamma\left(\frac{1}{\beta},1\right)$. ■

Since software that can generate gamma distributed random numbers is readily available, Theorem 3 provides a representation to easily generate random numbers from a generalised normal distribution.

Using the same notation defined in Proposition 1, the case where $f_0=\phi^*$, $G_0=\Phi$ (with ϕ^* denoting the PDF defined in (2), $\Phi(\cdot)$ representing the standard normal CDF) and where $w(x)=\sqrt{2}\lambda x$ for $\lambda\in\mathbb{R}$ yields the following definition of the $\mathcal{SGN}$ distribution.

Definition 4 *A random variable X has the $\mathcal{SGN}$ distribution with location parameter μ and scale parameter α if its PDF is given by*

$$f_X(x;\mu,\alpha,\beta,\lambda) = \frac{2}{\alpha}\phi^*\left(\frac{x-\mu}{\alpha};\beta\right)\Phi\left(\sqrt{2}\lambda\left(\frac{x-\mu}{\alpha}\right)\right),\ x\in\mathbb{R} \tag{3}$$

where $\mu\in\mathbb{R}$, $\alpha,\beta\in\mathbb{R}^+$ and $\lambda\in\mathbb{R}$. This is denoted by $X\sim\mathcal{SGN}(\mu,\alpha,\beta,\lambda)$.

When $\mu=0$, $\alpha=\sqrt{2}$ and $\beta=2$ the $\mathcal{SGN}$ distribution with PDF as in (3), the PDF collapses to that of Azzalini's $\mathcal{SN}$ (Azzalini 1985).

3 Statistical Properties

In this section, three methods are proposed to evaluate some statistical characteristics (expected value, standard deviation, skewness and kurtosis) of the $\mathcal{SGN}$ distribution. Four different parameter structures of the $\mathcal{SGN}$ are considered as candidates and the results are tabulated. Finally, a stochastic representation of the $\mathcal{SGN}$ distribution is proposed.

3.1 Method 1

The acceptance-rejection (AR) method (Lange 2010) is used to generate random skew generalised normal variates from $\mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ distribution with PDF given by (3). Fig. 1 shows the results of the AR algorithm for the first parameter structure considered. The green and red points indicate the random numbers that are respectively accepted and rejected as candidates from the $\mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ distribution with PDF given by (3).

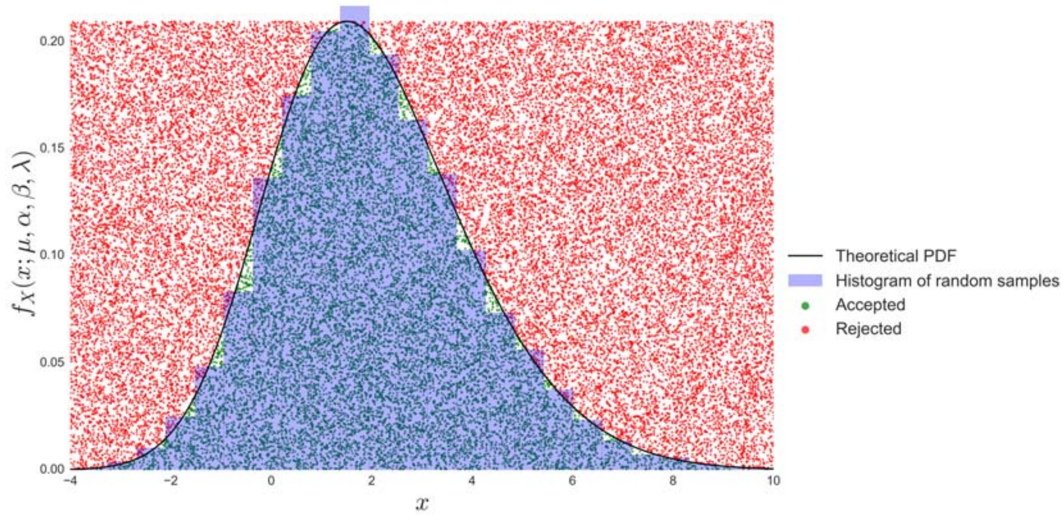

Figure 1. Histogram of the realised random variates drawn from $\mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ with theoretical PDF given by (3) overlaid for the first parameter structure i.e. $(\mu = 0, \alpha = 4, \beta = 2, \lambda = 2)$.

Estimates of the characteristics of the $\mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ distribution can be obtained by performing calculations on the realised random variates. The drawback of this method is that the AR algorithm may perform poorly or fail outright for certain parameter structures.

3.2 Method 2

This method uses the stochastic representation of the $\mathcal{GN}$ distribution as derived in Section 2.

Theorem 5 *If $X \sim \mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ with PDF as defined in* (3) *then*

$$\mathbb{E}[X^r] = \mathbb{E}_{X_*}\left[2X_*^r \Phi\left(\sqrt{2}\lambda\left(\frac{X_* - \mu}{\alpha}\right)\right)\right]$$

where $X_* \sim \mathcal{GN}(\mu, \alpha, \beta)$ has PDF as in (2).

Proof. By definition

$$\begin{aligned} \mathbb{E}[X^r] &= \int_{\mathbb{R}} x^r \frac{\beta}{\alpha\Gamma\left(\frac{1}{\beta}\right)} e^{-\left|\frac{x-\mu}{\alpha}\right|^{\beta}} \Phi\left(\sqrt{2}\lambda\left(\frac{x-\mu}{\alpha}\right)\right) dx \\ &= \mathbb{E}_{X_*}\left[2X_*^r \Phi\left(\sqrt{2}\lambda\left(\frac{X_* - \mu}{\alpha}\right)\right)\right] \end{aligned}$$

where $X_* \sim \mathcal{GN}(\mu, \alpha, \beta)$ has PDF as in (2). ■

Since Theorem 3 can be used to generate random $\mathcal{GN}$ variates, Theorem 5 provides a method to calculate the moments of $X \sim \mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ with PDF given by (3), which are then used to calculate the characteristics of the respective distribution.

3.3 Method 3

A novel expression for the moments of $X \sim \mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ with PDF given by (3), is presented.

Theorem 6 *If $X \sim \mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ with PDF given by (3) then*

$$\mathbb{E}[X^r] = \sum_{k=0}^{r} \binom{r}{k} \mu^{r-k} \alpha^k \mathbb{E}[Z^k] \tag{4}$$

where

$$\mathbb{E}[Z^k] = \begin{cases} \frac{\Gamma\left(\frac{k+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)} & \text{, for } k \text{ even} \\ \frac{\Gamma\left(\frac{k+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)} \left\{2\mathbb{E}_Q\left[\Phi\left(\sqrt{2}\lambda Q^{\frac{1}{\beta}}\right)\right] - 1\right\} & \text{, for } k \text{ odd} \end{cases}$$

where $Q \sim Gamma\left(\frac{k+1}{\beta}, 1\right)$.

Proof. Consider a random variable $Z \sim \mathcal{SGN}(0, 1, \beta, \lambda)$ with PDF as defined in (3) then

$$\begin{aligned} \mathbb{E}[Z^k] &= \int_0^\infty z^k \frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)} e^{-z^\beta} \Phi\left(\sqrt{2}\lambda z\right) dz + (-1)^r \int_0^\infty z^k \frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)} e^{-z^\beta} \Phi\left(-\sqrt{2}\lambda z\right) dz \\ &= I_1 + I_2. \end{aligned}$$

Let $q = z^\beta$ then it follows that

$$I_I = \frac{\Gamma\left(\frac{k+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)} \mathbb{E}_Q\left[\Phi\left(\sqrt{2}\lambda Q^{\frac{1}{\beta}}\right)\right].$$

Similarly,

$$I_2 = \frac{(-1)^k \Gamma\left(\frac{k+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)} \mathbb{E}_Q\left[1 - \Phi\left(\sqrt{2}\lambda Q^{\frac{1}{\beta}}\right)\right].$$

Therefore result (**??**) follows. To extend this approach for a random variable $X = \mu + \alpha Z$, so that $X \sim \mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ with PDF given by (3), it follows from the binomial theorem that

$$\mathbb{E}[X^r] = \sum_{k=0}^{r} \binom{r}{k} \mu^{r-k} \alpha^k \mathbb{E}[Z^k].$$

■

3.4 Simulation results

Four different parameter structures of the $\mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ distribution are considered. The three methods presented are used to estimate the characteristics of the distribution for a certain parameter structure. The time taken by each of the methods to obtain a result is also recorded. The results of the simulation are summarised in Table 1.

Remark 7 *It is clear that Method 2 performs comparatively poorly. Method 1 may outperform Method 3, however, the latter will work for any valid parameter structure which is not the case with Method 1.*

3.5 Stochastic representation

After noting that the AR algorithm is unable to draw appropriate samples from $\mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ for certain parameter structures, it was undertaken to investigate a more stable sampling scheme that generates random variates from $X \sim \mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ with PDF given by (3). A stochastic representation is developed that is useful for generating random numbers from a $\mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ distribution. Contrary to the AR method, this stochastic representation is able to draw random samples from $X \sim \mathcal{SGN}(\mu, \alpha, \beta, \lambda)$ with PDF given by (3) for any valid parameter structure.

		Simulation 1	Simulation 2	Simulation 3	Simulation 4
4*Parameter structure	μ	0	0	0	0
	α^2	16	16	25	4
	β	2	5	3	2
	λ	2	2	-2	25
3*Time taken (seconds)	Method 1	1.09	**0.094**	**0.110**	**0.111**
	Method 2	33.362	22.545	20.018	10.545
	Method 3	**0.094**	3.017	4.094	0.219

Table 1: Time taken by each method to obtain estimates of the characteristics of $X \sim \mathcal{SGN}\left(\mu, \alpha^2, \beta, \lambda\right)$ with PDF given by (3)

Lemma 8 *Let X and Y be a random variables with respective PDFs $f_X(x)$ and $f_Y(y)$ both symmetric about zero. If $W = -\lambda Y$ then*

$$\mathbb{P}[X + W < 0] = \mathbb{P}[X < \lambda Y] = \frac{1}{2}$$

Proof. Consider the random variable $W = -\lambda Y$ with PDF $f_W(w)$.
Case 1: $\lambda \neq 0$. Since Y is a symmetric random variable it follows that

$$\mathbb{P}[W \geq w] = \mathbb{P}\left[Y \geq \frac{w}{\lambda}\right] = \mathbb{P}[W \leq -w].$$

Symmetry of W follows since $\mathbb{P}[W \geq w] = \mathbb{P}[W \leq -w]$.
Let $Z = X + W$. Using the convolution of marginal PDFs of symmetric random variables X and W to obtain

$$\begin{aligned} f_Z(z) &= f_{X+W}(z) \\ &= \int_{\mathbb{R}} f_X(-z + w) f_W(-w)\, dw && (5) \\ &= f_{X+W}(-z) && (6) \\ &= f_Z(-z). && (7) \end{aligned}$$

Therefore $Z = X + W = X - \lambda Y$ is a symmetric random variable around zero and it follows that

$$\mathbb{P}[Z < 0] = \mathbb{P}[X < \lambda Y] = \frac{1}{2}.$$

Case 2: $\lambda = 0$

$$\mathbb{P}[Z < 0] = \mathbb{P}[X < 0] = \frac{1}{2}$$

since X is symmetric random variables around zero. ■

Theorem 9 *Let $U \sim \mathcal{GN}(\beta)$ with PDF $\phi^*(\cdot; 0, 1, \beta)$ given by (2), and $U_1 \sim \mathcal{N}(0, 1)$ with U and U_1 independent. If*

$$Z = U \quad \textit{whenever} \quad U_1 \leq \sqrt{2}\lambda U$$

then $Z \sim \mathcal{SGN}(0, 1, \beta, \lambda)$ with PDF $f_Z(z; 0, 1\beta, \lambda)$ as in (3).

Proof. Let $Z = U | \left\{U_1 \leq \sqrt{2}\lambda U\right\}$. Then

$$\mathbb{P}[Z \leq z] = \frac{\mathbb{P}\left[U \leq z, U_1 \leq \sqrt{2}\lambda U\right]}{\mathbb{P}\left[U_1 \leq \sqrt{2}\lambda U\right]}. \tag{8}$$

Since U and U_1 are independent it follows that

$$\mathbb{P}\left[U \leq z, U_1 \leq \sqrt{2}\lambda U\right] = \int_{-\infty}^{x} \phi^*(u; 0, 1, \beta)\, \Phi\left(\sqrt{2}\lambda u\right) du. \tag{9}$$

Since $f_U(u)$ and $f_{U_1}(u_1)$ are both symmetric about zero, applying Proposition 1 it follows that

$$\mathbb{P}\left[U_1 \le \sqrt{2}\lambda U\right] = \frac{1}{2}. \tag{10}$$

Using (9) and (10) in (8) it follows that

$$\mathbb{P}[Z \le z] = \int_{-\infty}^{x} 2\phi^*(u;0,1,\beta)\,\Phi\left(\sqrt{2}\lambda u\right) du$$

with PDF $2\phi^*(z;0,1,\beta)\,\Phi\left(\sqrt{2}\lambda z\right)$ which is the PDF given by (3). ■

Corollary 10 *Let $U \sim \mathcal{GN}(\beta)$ with PDF $\phi^*(\cdot;0,1,\beta)$ given by (2) and $U_1 \sim \mathcal{N}(0,1)$ with U and U_1 independent. If*

$$X = \mu + \alpha U \quad \text{whenever } U_1 \le \sqrt{2}\lambda U$$

then $X \sim \mathcal{SGN}(\mu,\alpha,\beta,\lambda)$ with PDF given by (3).

Since there is readily available software that can generate normal distributed random numbers and the sampling scheme in Section 2 can be used to generate generalised normal random numbers. Theorem 9 and Corollary 10 provide a representation to easily generate random numbers from a $\mathcal{SGN}$ distribution for any valid parameter structure. Figure 2 shows histograms of the random samples taken from $X \sim \mathcal{SGN}(\mu,\alpha,\beta,\lambda)$ using the stochastic representation in Corollary 10 with the corresponding theoretical PDF (3) overlaid.

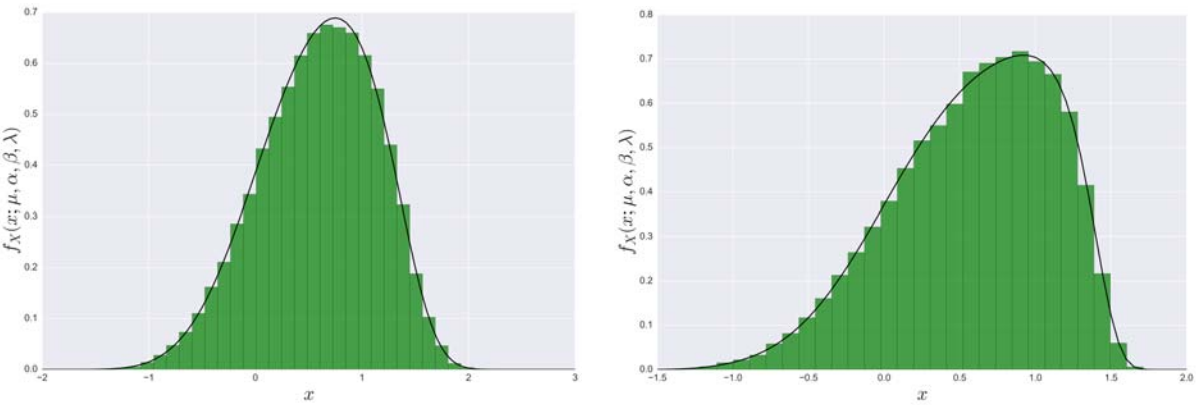

Figure 2. Histograms of realised random samples of size 10 000 taken from $X \sim \mathcal{SGN}_I(\mu,\alpha,\beta,\lambda)$ with the corresponding theoretical PDF (3), overlaid for different values of μ,α,β and λ.

4 Generalisation of the skewing mechanism

In this section a generalisation of (3) is proposed by assuming the skewing mechanism has the elliptically contoured distribution (Arashi and Nadarajah 2017).

Definition 11 *A random variable Y has the skew elliptical generalised-normal distribution with location parameter μ, scale parameter α and shape parameter β if its PDF is given by*

$$f_Y(y;\mu,\alpha\beta,\lambda) = \frac{2}{\alpha}\phi^*\left(\frac{y-\mu}{\alpha};\beta\right)\Phi^E\left(\sqrt{2}\lambda\frac{y-\mu}{\alpha}\right) \tag{11}$$

where $\alpha,\beta \in \mathbb{R}^+$, $\lambda \in \mathbb{R}$ and $\Phi^E(\cdot)$ is the CDF of an elliptically contoured distribution (Arashi and Nadarajah 2017).

The approach in (Chu 1973) is used to represent (11) as

$$f_Y(y;\mu,\alpha\beta,\lambda)=\frac{2}{\alpha}\phi^*\left(\frac{y-\mu}{\alpha};\beta\right)\int_0^\infty \mathcal{W}(t)\,\Phi_{N(0,t^{-1})}\left(\sqrt{2}\lambda\left(\frac{y-\mu}{\alpha}\right)\right)dt \tag{12}$$

where $\beta\in\mathbb{R}^+$, $\lambda\in\mathbb{R}$ and $\Phi_{N(0,t^{-1})}$ is the CDF of a $N(0,t^{-1})$ distribution and where $\mathcal{W}(\cdot)$ is a weighting function on $(0,\infty)$. If the weighting function used in (12) is $\mathcal{W}(t)=\delta(t-1)$, where $\delta(\cdot)$ is the dirac delta function, (11) simplifies to (3).

Assuming now that $\mathcal{W}(t)=\frac{\nu\left(\frac{\nu t}{2}\right)^{\frac{\nu}{2}-1}}{2\Gamma\left(\frac{\nu}{2}\right)e^{\frac{\nu t}{2}}}$ is used in (12), it results in a new contribution as stated in the following definition.

Definition 12 *A random variable Y has the t-skewed generalised-normal distribution with location parameter μ and scale parameter α if its PDF is given by*

$$f_Y(y;\mu,\alpha,\nu,\beta,\lambda) = \tfrac{2}{\alpha}\phi^*\left(\tfrac{y-\mu}{\alpha};\beta\right)\mathcal{T}_\nu\left(\sqrt{2}\lambda\tfrac{y-\mu}{\alpha}\right),\ y\in\mathbb{R} \tag{13}$$

where $\mu\in\mathbb{R}$, $\nu\in\mathbb{Z}^+$ and $\alpha,\beta\in\mathbb{R}^+$ and $\lambda\in\mathbb{R}$ and where $\mathcal{T}_\nu(\cdot)$ represents the CDF of a t distribution with degrees of freedom ν This is denoted by $Y\sim\mathcal{SGN}_{\mathcal{T}_\nu}(\mu,\alpha,\beta,\lambda)$.

A stochastic sampling scheme similar to Corollary 10 is developed to draw samples from (13).

Corollary 13 *If $U\sim\mathcal{GN}(\beta)$ and $U_1\sim t(\nu)$, where $t(\nu)$ denotes t distribution with ν degrees of freedom, with U and U_1 independent then conditionally*

$$Y=\mu+\alpha U \quad whenever \quad U_1\leq\sqrt{2}\lambda U$$

then $Y\sim\mathcal{SGN}_{\mathcal{T}_\nu}(\mu,\alpha,\beta,\lambda)$ with PDF $f_Y(y;\mu,\alpha,\nu,\beta,\lambda)$ as in (13).

Corollary 13 provides a representation to easily generate random numbers from a SGN_{T_ν} distribution for any valid parameter structure. Figure 3 shows histograms of the random samples taken from $X\sim\mathcal{SGN}_{\mathcal{T}_\nu}(\mu,\alpha,\beta,\lambda)$ using the stochastic representation in Corollary 13 with the corresponding theoretical PDF (13) overlaid.

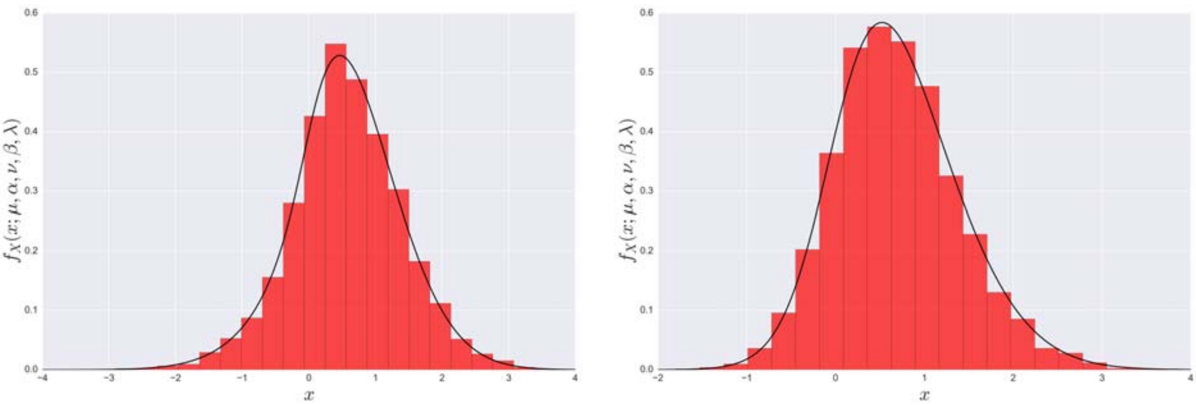

Figure 3. Histograms of realised random samples of size 10 000 taken from $X\sim\mathcal{SGN}_{\mathcal{T}_\nu}(\mu,\alpha,\beta,\lambda)$ with the corresponding theoretical PDF (13), overlaid for different values of μ,α,ν,β and λ.

5 Applications

5.1 Fitting to data

An Australian athletes data set containing various measurements on athletes specialising in different sports is used (Telford and Cunningham 1991). The variable of interest is the caliper measurement obtained from each athlete, which provides an indication of body-fat percentage. It is vital to perform a test on whether a distribution is a candidate for fitting to a particular data set. The Kolmogorov-Smirnov (KS) goodness-of-fit test is performed to assess the suitability of fitting the $\mathcal{SGN}$ distribution to this data set. The standard critical values of the $K-S$ statistic do not apply when any parameters of the candidate distribution ($\mathcal{SGN}$ in this case) are estimated from the data. A Monte Carlo approach must be used instead to construct an appropriate test and is outlined as follows:

Level of significance, α	Critical value, $q_{1-\alpha}$
0.01	0.138527
0.05	0.145027
0.1	0.1562062

Table 2: Simulated critical values of the $K-S$ goodness of fit test at various levels of significance

1. The null hypothesis, H_0, is that the data are from a $\mathcal{SGN}$ distribution;
2. Fit the $\mathcal{SGN}$ distribution to the original data using maximum likelihood estimation;
3. Calculate the KS distance, d^*, between the data and the fitted $\mathcal{SGN}$ distribution;
4. Bootstrap the original data and fit the $\mathcal{SGN}$ distribution again;
5. Calculate KS distance, d_i, between the bootstrapped data and the fitted $\mathcal{SGN}$ distribution;
6. Repeat steps $4-5$ M times to obtain the set $d=\{d_1,d_2,\cdots,d_M\}$;
7. Calculate the $(1-\alpha)^{th}$ sample quantiles $q_{1-\alpha}$ of d_M at levels of significance: $\alpha=0.01,\ 0.05,\ 0.1$;
8. If $d^*<q_{1-\alpha}$ the null hypothesis cannot be rejected at α level of significance and there is not enough evidence to suggest that the data are not from a $\mathcal{SGN}$ distribution. If this is the case, the KS test indicated that the $\mathcal{SGN}$ distribution is a suitable candidate to fit to the data;
9. An approximate $p-value$ can also be obtained as the proportion of time the elements in d_M are greater than d^*.

The test is performed with $M=5000$. The test statistic is calculated as $d^*=0.1174142$. The critical values at different levels of significance are presented in Table 2.

Since $d^*<q_{1-\alpha}$ for $\alpha=0.01,\ 0.05,\ 0.1$ and the approximate $p-value$ is calculated as 0.545109, the null hypothesis that the data are from a $\mathcal{SGN}$ distribution cannot be rejected. Therefore it is concluded that the $\mathcal{SGN}$ distribution is a suitable candidate to fit to this data set.

In what follows, the $\mathcal{SGN}$ model, along with relevant alternative models which include the generalised Balakrishnan skew-normal ($\mathcal{GBSN}$) model (see Definition 14) is fitted to the data set.

Definition 14 *A random variable X has the generalised Balakrishnan skew-normal distribution with location parameter μ and scale parameter σ if its PDF is given by*

$$f_X(x;\mu,\sigma,n,\lambda_1,\lambda_2) = \frac{c_n(\mu,\sigma,\lambda_1,\lambda_2)}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi^n\left(\frac{\lambda_1(x-\mu)}{\sqrt{\sigma^2+\lambda_2(y-\mu)^2}}\right) \quad , x\in\mathbb{R} \tag{14}$$

where $\mu\in\mathbb{R}$, $\sigma\in\mathbb{R}^+$, $n\in\mathbb{R}^+$, $\lambda_1\in\mathbb{R}$, $\lambda_2\in\mathbb{R}^+$ and

$$c_n(\mu,\sigma,\lambda_1,\lambda_2) = \frac{1}{\mathbb{E}_B\left[\Phi^n\left(\frac{\lambda_1(B-\mu)}{\sqrt{\sigma^2+\lambda_2(B-\mu)^2}}\right)\right]}$$

with $B\sim\mathcal{N}(\mu,\sigma^2)$. This is denoted by $Y\sim\mathcal{GBSN}(\mu,\sigma^2,n,\lambda_1,\lambda_2)$ (Hasanalipour and Sharafi 2012).

The normal, $\mathcal{SN}$ (Azzalini 1985), $\mathcal{SGN}$ and $\mathcal{GBSN}$ distributions are fitted to the caliper measurement obtained from each athlete. All distributions were fitted using the method of maximum likelihood. In particular, to fit the $\mathcal{SGN}$ distribution, let $x_1,\cdots,x_m$ be a random sample of size m from $\mathcal{SGN}(\mu,\alpha,\beta,\lambda)$ and maximise the likelihood function given by

$$\mathcal{L}(\mu,\alpha,\beta,\lambda)=\left(\frac{2}{\alpha}\right)^m\prod_{i=1}^{m}\phi^*\left(\frac{x_i-\mu}{\alpha};\beta\right)\Phi\left(\sqrt{2}\lambda\left(\frac{x_i-\mu}{\alpha}\right)\right)$$

using well-known optimisation techniques. The adequacy of the fit of the four distributions is assessed by the BIC and AIC information criteria.

	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\alpha}^2$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{n}$	AIC	BIC
$\mathcal{N}$	22.96	2.86^2	-	-	-	-	-	-	1001.336	997.336
$\mathcal{SN}$	19.97	4.13^2	-	-	2.313	-	-	-	986.199	980.199
$\mathcal{SGN}$	20.80	-	3.842^2	1.381	1.074	-	-	-	**984.349**	**976.349**
$\mathcal{GBSN}$	19.76	4.288^2	-	-	-	2.535	1.092	1.1613677	989.820	979.820

Table 3: AIC and BIC criteria obtained for each of the fitted distributions

As a visual assessment of goodness of fit, the estimated PDFs of the three distributions and the empirical histogram are plotted in the Figure 4. The results in Table 3 identify the $\mathcal{SGN}(\mu, \alpha, \lambda)$ distribution with PDF $f_X(x; \mu, \alpha, \lambda)$ given by (3) as the best fit for the given data.

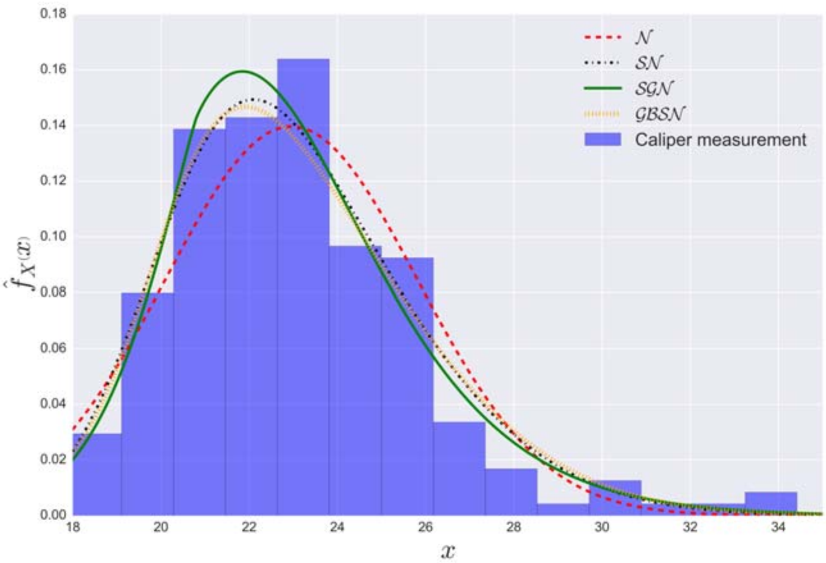

Figure 4. Empirical histogram of data with overlaid fitted PDFs.

5.2 Approximating the binomial distribution

Consider a random variable $X \sim Binomial\,(n, p)$. A normal distribution with expected value np and variance $np(1-p)$ is usually used to approximate the binomial distribution when n is large or when p is close to 0.5 (in which case the binomial distribution PDF is approximately symmetric). However, it is well known that the PDF is not symmetric for $p \neq 0.5$ and exhibits a non-negligible degree of skewness for both large and small values of p. It is therefore of interest to consider approximating the binomial distribution using the $\mathcal{SN}$ and $\mathcal{SGN}$ distributions (as Chang et. al. 2008 considered) in order to account for the skewness present.

Methodology

The adopted methodology is different compared to the approach of Chang et. al. (2008).

1. Let $X \sim Binomial\,(n, p)$;
2. Let the classic normal approximation to the binomial distribution be $A \sim \mathcal{N}\left(\hat{\mu} = np, \hat{\sigma}^2 = np\,(1-p)\right)$ with PDF $f_A(a; \hat{\mu}, \hat{\sigma})$ and calculate $d = \max_{0 \leq i \leq n} |f_X(i; n, p) - f_A(i; \hat{\mu}, \hat{\sigma})|$;
3. Let the $\mathcal{SN}$ approximation be $B \sim \mathcal{SN}\left(\hat{\mu}, \hat{\sigma}^2, \hat{\lambda}\right)$ with PDF $f_B\left(b; \hat{\mu}, \hat{\sigma}, \hat{\lambda}\right)$ as in (Azzalini 1985) with estimated parameters numerically minimising the maximum distance between $f_X(x; n, p)$ and $f_B\left(b; \hat{\mu}, \hat{\sigma}, \hat{\lambda}\right)$ (i.e minimising $d = \max_{0 \leq i \leq n} \left|f_X(i; n, p) - f_B\left(i; \hat{\mu}, \hat{\sigma}, \hat{\lambda}\right)\right|$);

	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\alpha}^2$	$\hat{\beta}$	$\hat{\lambda}$	d
$\mathcal{N}$	28.5	0.1937^2	-	-	-	0.06289
$\mathcal{SN}$	28.624	1.232^2	-	-	0.140	0.01944
$\mathcal{SGN}$	29.672	-	2.083^2	1.696	-1.702	**0.01165**

Table 4: d obtained for each of the distributions approximating the $Binomial\,(30, 0.95)$ distribution

4. Let the $\mathcal{SGN}$ approximation be $C \sim \mathcal{SGN}\left(\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\lambda}\right)$ with PDF $f_C\left(c; \hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\lambda}\right)$ as in (3) with estimated parameters numerically minimising the maximum distance between $f_X\,(x; n, p)$ and $f_C\left(c; \hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\lambda}\right)$ (i.e minimising $d = \max_{0 \leq i \leq n} \left| f_X\,(i; n, p) - f_C\left(i; \hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\lambda}\right)\right|$);
5. The initial values in the optimisation algorithm used to estimate the parameters of the $\mathcal{SN}$ approximation are $\{\mu, \sigma, \lambda\} = \left\{np, \sqrt{np\,(1-p)}, 0\right\}$. λ is set zero so that the optimisation algorithm begins with a symmetric distribution;
6. The initial values in the optimisation algorithm used to estimate the parameters of the $\mathcal{SGN}$ approximations $\{\mu, \alpha, \beta, \lambda\} = \left\{np, \sqrt{np(1-p}, 2, 0\right\}$. β is set to two and λ is set to zero so that the optimisation algorithm begins with a symmetric distribution with normal tail behavior;
7. In Step 5 and Step 6, the parameters are set in this way to ensure that the algorithm used does not favour one distribution over another;
8. The approximation error is calculated for $i = 0, \cdots, n$ as
   (a) $f_X\,(i; n, p) - f_A\,(i; \hat{\mu}, \hat{\sigma})$ for the $\mathcal{N}$ distribution;
   (b) $f_X\,(i; n, p) - f_B\left(i; \hat{\mu}, \hat{\sigma}, \hat{\lambda}\right)$ for the $\mathcal{SN}$ distribution and
   (c) $f_X\,(i; n, p) - f_C\left(i; \hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\lambda}\right)$ for the $\mathcal{SGN}$ distribution.

Case 1: $n = 30$, $p = 0.95$
Table 4 and Figure 5 below summarise the results obtain when the $\mathcal{N}$, $\mathcal{SN}$ and $\mathcal{SGN}$ are used to approximate $X \sim Binomial\,(30, 0.95)$:

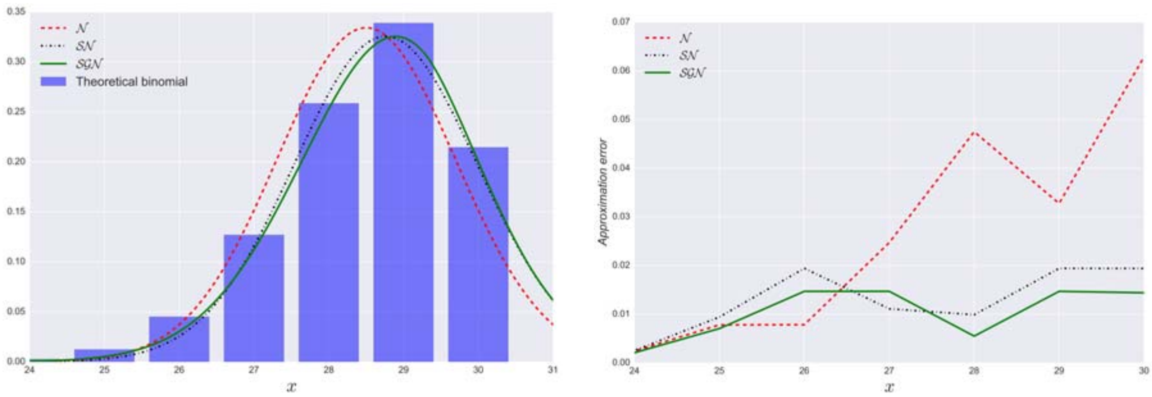

Figure 5. $\mathcal{N}$, $\mathcal{SN}$ (see (Azzalini 1985)) and $\mathcal{SGN}$ (see (3)) approximations to $Binomial\,(30, 0.95)$ distribution.

It is observed that using the $\mathcal{SGN}$ distribution to approximate a binomial distribution with p either large or small results in a overall more accurate approximation compared to both the $\mathcal{N}$ and $\mathcal{SN}$ distributions. This is due to the $\mathcal{SGN}$ having two parameters, i.e. β and λ, adding flexibility in accounting for skewness of the binomial distribution exhibited when p is large or small. In both cases above, the $\mathcal{SGN}$ resulted in the minimum d, and by this measure it is conclude that the $\mathcal{SGN}$ distribution outperforms both the $\mathcal{N}$ and $\mathcal{SN}$ distributions in approximating a binomial distribution with p either large or small.

6 Final remarks

A stochastic representation of the $\mathcal{GN}$ distribution is derived and the characteristics of the $\mathcal{SGN}$ distribution are computationally investigated. An acceptance-rejection algorithm is employed to sample from the $\mathcal{SGN}$ distribution and shortfalls of this approach are noted. Thereafter, two further methods which approximate the characteristics of the $\mathcal{SGN}$ distribution are derived and compared. Given the shortfalls of the acceptance-rejection algorithm, it was then undertaken to derive a stochastic representation for the $\mathcal{SGN}$ distribution. Combining the generalised normal distribution with a skewing mechanism belonging to the elliptical class provides a new, more general framework for skew-symmetric distributions. A special case where the weighting function in the new framework resulted in the CDF of the t-distribution to act as the skewing mechanism is illustrated and a stochastic representation of this case was developed. A distribution fitting application is presented and it is found that the $\mathcal{SGN}$ distribution was the best fit for the given data, outperforming the $\mathcal{GBSN}$ distribution. A second application which involves the approximating the binomial distribution using the $\mathcal{N}$, $\mathcal{SN}$ and $\mathcal{SGN}$ distributions is also presented. It is determined the the $\mathcal{SGN}$ distribution outperforms the $\mathcal{N}$ and $\mathcal{SN}$ distributions in approximating particular binomial distributions. Further possible extensions include using the $\mathcal{GN}$ distribution in the skewing mechanism and using the $\mathcal{SGN}$ distribution in the form of a beta generated distribution defined in (Mameli and Musio 2013).

Acknowledgements

This work is based upon research supported by the National Research Foundation, South Africa (Grant number: 91497; 109214; 102640). Opinions expressed an conclusions arrived at in this study are those of the authors and are not necessarily attributed to the NRF.

References

Arashi, M. and S. Nadarajah (2017). Generalized elliptical distributions. *Communications in Statistics - Theory and Methods 46*(6412-6432), 13.

Arellano-Valle, R., G. Del Pino, and E. San Martın (2002). Definition and probabilistic properties of skew-distributions. *Statistics & Probability Letters 58*(2), 111–121.

Arnold, B. C., R. J. Beaver, A. Azzalini, N. Balakrishnan, A. Bhaumik, D. K. Dey, C. M. Cuadras, and J. M. Sarabia (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test 11*(1), 7–54.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12*(2), 171–178.

Azzalini, A. (2013). *The skew-normal and related families*, Volume 3. Cambridge University Press.

Azzalini, A. and G. Regoli (2011). Some properties of skew-symmetric distributions. *Annals of the Institute of Statistical Mathematics 64*, 857–879.

Bahrami, W. and E. Qasemi (2015). A flexible skew-generalised normal distribution. *Journal of Statistical Research of Iran 11*(2), 131–145.

Bain, L. J. and M. Engelhardt (1992). *Introduction to probability and mathematical statistics*. Number 202. Thomson Learning.

Chang, C., J. Lin, N. Pal, and M. Chiang. (2008). A note on improved approximation of the binomial distribution by the skew-normal distribution. *The American Statistician 62*(2), 167–170.

Chu, K. C. (1973). Estimation and decision for linear system with elliptical random processes. *IEEE Transactions on Automatic Control 18*, 499–505.

Hasanalipour, P. and M. Sharafi (2012). A new generaslised Balakishnan skew-normal distribution. *Statistical Papers 53*, 219–228.

Lange, K. (2010). *Numerical Analysis for Statisticians* (2 ed.). Springer.

Mameli, V. and M. Musio (2013). A generalization of the skew-normal distribution: the beta skew-normal. *Communications in Statistics - Theory and Methods 42*(12), 2229–2244.

Pourahmadi, M. (2007). Construction of skew-normal random variables: Are they linear combinations of normal and half-normal? *Journal of Statistical Theory and Application 3*, 314–328.

Salinas, H. S., R. B. Arellano-Valle, and H. W. Gómez (2007). The extended skew-exponential power distribution and its derivation. *Communications in Statistics - Theory and Methods 36*(9), 1673–1689.

Subbotin, M. (1923). On the law of frequency of error. *Matematicheskii Sbornik 31*(2), 296–301.

Telford, R. D. and R. Cunningham (1991). Sex, sport and body-size dependency of hematology in highly trained athletes. *Medicine and Science in Sports and Exercise 23*, 788–794.

Yadegari, I., A. Gerami, and M. J. Khaledi (2008). A generalization of the Balakrishnan skew-normal distribution. *Statistics and Probability Letters 78*(10), 1165–1167.