OKU Fen Bilimleri Enstitüsü Dergisi
6(1): 842-853, 2023

OKU Journal of The Institute of Science and Technology, 6(1): 842-853, 2023

Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi

Osmaniye Korkut Ata University Journal of The Institute of Science and Technology

A New Skew-Symmetric Gudermannian-Laplace Distribution with Properties and Application to Wind Speed Data

Abdullah YILMAZ[1*]

[1]Kırıkkale Üniversitesi Fen Edebiyat Fakültesi Aktüerya Bilimleri Bölümü, Kırıkkale

[1]https://orcid.org/ 0000-0002-1196-9541
*Sorumlu yazar: a.yilmaz@kku.edu.tr

Research Article

Article History:
Received: 08.01.2023
Accepted: 24.01.2023
Published online: 10. 03.2023

Keywords:
Skew-symmetric distribution
Skewness
Skew Normal distribution
Gudermanian distribution
Laplace distribution

ABSTRACT

We introduce a novel probability distribution that may be used to model both skewed and symmetric data. This new distribution, called the skew-symmetric Gudermanian-Laplace (SSGL) distribution, includes a shape parameter that allows it to change the asymmetry. Some fundamental statistical properties of the new distribution have been given explicit analytical expressions. The study also includes parameter estimations and simulation sections. We considered two datasets in the real-world data application. The first dataset is the "heights of 100 Australian athletes" data, which is discussed in many studies examining alternative skewed models. The second dataset contains the average wind speeds recorded by the İstanbul Çatalca meteorological observatory in January 2020. We showed that the SSGL distribution outperforms its well-known alternative, the Skew-Normal distribution, in both datasets. As a result of the study, it was concluded that the SSGL distribution is a suitable alternative for modeling skewed data.

Özellikleri ve Rüzgar Hızı Verilerine Uygulanması ile Yeni Çarpık-Simetrik Gudermannian-Laplace Dağılımı

Araştırma Makalesi

Makale Tarihçesi:
Geliş tarihi: 08.01.2023
Kabul tarihi:24.01.2023
Online Yayınlanma: 10.03.2023

Anahtar Kelimeler:
Çarpık-Simetrik dağılım
Çarpıklık
Çarpık Normal dağılım
Gudermanian dağılımı
Laplace dağılımı

ÖZ

Bu makalede hem çarpık hem de simetrik verileri modellemek için kullanılabilecek yeni bir olasılık dağılımı sunuyoruz. Çarpık simetrik Gudermanian-Laplace (SSGL) dağılımı olarak adlandırılan bu yeni dağılım, asimetriyi değiştirmesine izin veren bir şekil parametresi içerir. Yeni dağılımın bazı temel istatistiksel özelliklerine açık analitik ifadeler verilmiştir. Çalışma aynı zamanda parametre tahminleri ve simülasyon bölümlerini de içermektedir. Gerçek dünya veri uygulamasında ise iki veri kümesini ele aldık. İlk veri seti, alternatif çarpık olasılık modellerini inceleyen birçok çalışmada el alınan “100 Avustralyalı sporcunun boyları” verisidir. İkinci veri seti, İstanbul Çatalca meteoroloji gözlemevi tarafından Ocak 2020'de kaydedilen ortalama rüzgar hızlarıdır. SSGL dağılımının, her iki veri setinde de iyi bilinen alternatifi olan Çarpık-Normal dağılımından daha iyi performans gösterdiğini gösterdik. Çalışma sonucunda SSGL dağılımının çarpık verilerin modellenmesi için uygun bir alternatif olduğu sonucuna varılmıştır.

To Cite: Yılmaz A. A New Skew-Symmetric Gudermannian-Laplace Distribution with Properties and Application to Wind Speed Data. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2023; 6(1): 842-853.

1.Introduction

Data may have a longer tail on one side than the other, indicating that it is "skewed." Understanding skewed data is crucial for a data scientist or other professional who works with data because most real-world situations aren't symmetrical—real data sets are frequently skewed. Skewed data, on the other hand, can pose problems with statistical models because outliers, which frequently generate skew, can have a harmful effect on a model's performance. In this regard, the presence of a skewness parameter in a probability distribution improves modeling success. There are several methods of obtaining skew-adjustable probability distributions via a parameter. For a detailed review of these methods, we advise interested readers to check (Gupta & Kundu, 2009). In this study, we shall focus on one of these approaches, the family of skew-symmetric distributions, which was introduced by Azzalini (Azzalini, 1985).

The skew-symmetric distribution family is a wide family of probability density functions that include the skewness parameter(s). The following lemma defines the main frame of probability distributions in the family (Azzalini, 1985).

Lemma: *Let f be a density function symmetric about zero, and Ψ is a Lebesgue measurable function satisfying* (i) $0 \leq \Psi(x) \leq 1$, (ii) $\Psi(x) + \Psi(-x) = 1$ *for $x \in \mathbb{R}$. Then*

$$h(x) = 2f(x)\Psi\big(g(x)\big) \tag{1}$$

is a probability density function, where g is a function that is odd and continuous.

Proof: Let $t(x) = 2\Psi\big(g(x)\big) - 1$. It is easy to show that t is an odd function. Due to f is a pdf symmetric about zero it's an even function and $f(x)t(x)$ is an odd function. Thus

$$\int_{-\infty}^{\infty} h(x)dx - 1 = \int_{-\infty}^{\infty} 2f(x)\Psi\big(g(x)\big)dx - 1 = \int_{-\infty}^{\infty} f(x)t(x)dx = 0.$$

Many studies in the literature have focused on the skew-symmetric distributions produced by the various choices of base distribution f, skewing function Ψ, and odd function g. Azzalini (Azzalini, 1985) introduced the skew-normal distribution $SN(\lambda)$ in his pioneering work, and $h(x) = 2\varphi(x)\Phi(\lambda x)$ represents its density function, where φ and Φ are the probability density function (pdf) and cumulative density function (cdf) of a standard normal distribution, and $\lambda \in \mathbb{R}$ is the asymmetry parameter. Elal-Olivero (Elal-Olivero, 2010) introduced the alpha skewed-normal distribution ($ASN(\alpha)$) by taking the skewing function $\Psi(x) = ((1 - \alpha x)^2 + 1)(2 + \alpha^2)^{-1}$ and the base distribution φ. Arellano-Valle et al. (Arellano-Valle, Gómez, & Quintana, 2004) presented a skew-normal distribution generalization, where f and Ψ are the pdf and cdf of the normal distribution, respectively, and $g(x) = (\lambda_1 x)(1 + \lambda_2 x^2)^{-1/2}$, where $\lambda_1 \in \mathbb{R}$ and $\lambda_2 \geq 0$ are real constants. Ma and Genton (Ma & Genton, 2004) proposed the flexible-skew-normal distribution, where $g(x) = (\alpha x + \beta x^3)$. This is a particular case of flexible skew-generalized normal distribution (Nekoukhou, Alamatsaz, & Aghajani, 2013), with $(x) = (\lambda_1 x + \lambda_3 x^3)(1 + \lambda_2 x^2)^{-1/2}$. The purpose of this work is to introduce the new Skew-Symmetric Gudermannian-Laplace distribution.

The contents of the rest of this paper are as follows: The skew-symmetric Gudermannian-Laplace distribution is introduced in the next section. In addition, this section includes studies on skewness and kurtosis coefficients, entropy, and raw moments of distribution. The next section provides inference procedures for maximum likelihood estimation and simulation studies. The final two sections of the study are the application of real data section, which demonstrates the usefulness of the new distribution, and the conclusion section, which discusses some findings related to the proposed distribution.

2. Skew-Symmetric Gudermannian-Laplace Distribution

In this section, the probability density function of the skew-symmetric Gudermannian-Laplace distribution is presented with some basic properties. We use the base distribution in eq(1) as standardized generalized Gudermannian (GG) which is symmetric about zero. The pdf of standardized GG distribution is (Altun, 2019)

$$f(x) = \frac{e^{\frac{\pi x}{2}}}{e^{\pi x} + 1}, \quad x \in \mathbb{R}.$$

As skewing function Ψ, we use the cdf of the well-known Laplace distribution. Thus, by taking the odd function $g(x) = \lambda x$, our skewing function is obtained as

$$\Psi\big(g(x)\big) = \begin{cases} 1 - e^{-\lambda x}/2 & ,\lambda x \geq 0 \\ e^{\lambda x}/2 & ,\lambda x < 0 \end{cases}, \qquad \lambda, x \in \mathbb{R}.$$

Definition: A random variable X has the Skew-Symmetric Gudermannian-Laplace distribution with parameter λ, $X \sim SSGL(\lambda)$, if its pdf has the form

$$h(x;\lambda) = \frac{e^{\frac{\pi x}{2}}}{e^{\pi x} + 1}\left[\left(1 - e^{-|\lambda x|}\right)\mathrm{sgn}(\lambda x) + 1\right], \quad x \in \mathbb{R} \qquad (2)$$

where sgn is the signum function, $\lambda \in \mathbb{R}$ is a shape parameter and controlling the skewness.

Cumulative Distribution Function (Cdf) : It is clear that $h(x;0) = f(x)$, thus $H(x;0) = 2\pi^{-1} arctan[exp\{x\pi/2\}]$ (Altun, 2019). We give the cdf of the $X \sim SSGL(\lambda)$ random variable as two separate functions for $\lambda > 0$ and $\lambda < 0$ to save space. For $\lambda > 0$ cdf of X is

$$H(x;\lambda) = \begin{cases} sec(\lambda) + \dfrac{ie^{i\lambda} B_{-e^{-\pi x}}\left(\frac{1}{2} - \frac{\lambda}{\pi}, 0\right)}{\pi} & ,x < 0 \\ \dfrac{2\mathrm{gd}\left(\frac{\pi x}{2}\right) + \pi\, sec(\lambda) + ie^{i\lambda} B_{-e^{\pi x}}\left(\frac{1}{2} - \frac{\lambda}{\pi}, 0\right)}{\pi} & ,x \geq 0 \end{cases}, \qquad (3)$$

where $\mathrm{gd}(x) = 2arctan\big(tanh(x/2)\big)$ is the Gudermannian function and $B_z(a,b)$ is the incomplete beta function (Dutka, 1981). For $\lambda < 0$ we get

$$H(x;\lambda) = \begin{cases} -sec(\lambda) - \dfrac{ie^{-i\lambda}B_{-e^{-\pi x}}\left(\frac{\Lambda}{2\pi},0\right) - 4\,cot^{-1}\left(e^{\frac{\pi x}{2}}\right)}{\pi} + 2 & ,x < 0 \\ \dfrac{2e^{\frac{\Lambda x}{2}}\ {}_2F_1\left(1,\frac{\Lambda}{2\pi};\frac{\Lambda}{2\pi}+1;-e^{\pi x}\right)}{\Lambda} - \dfrac{1}{2}tan\left(\dfrac{\Lambda}{4}\right)\left(cot\left(\dfrac{\Lambda}{4}\right) - 1\right)^2 & ,x \geq 0 \end{cases}, \tag{4}$$

where ${}_2F_1$ is the hypergeometric function (Abramowitz & Stegun, 1964) and $\Lambda = 2\lambda + \pi$.

Raw Moments: Let $X \sim SSGL(\lambda)$, then even moments of X is given by

$$E\left(X^{2k}\right) = \left|\sum_{i=1}^{2k}\left(-\frac{1}{2}\right)^i \sum_{L=0}^{2i}(-1)^L\binom{2i}{L}(i-L)^{2k}\right|, \qquad k = 1,2,3,\dots \tag{5}$$

As can be seen, the even moments are unaffected by λ. Moreover, the values of even moments are $E(X^2) = 1$, $E(X^4) = 5$, $E(X^6) = 61, E(X^8) = 1385, \dots$ and are known as Euler numbers. Odd moments are calculated as

$$E\left(X^{2k-1}\right) = \frac{k!\left(-\zeta\left(k+1,\frac{2|\lambda|+\pi}{4\pi}\right) + \zeta\left(k+1,\frac{|\lambda|}{2\pi}+\frac{3}{4}\right) + \zeta\left(k+1,\frac{1}{4}\right) - \zeta\left(k+1,\frac{3}{4}\right)\right)}{\mathrm{sgn}(\lambda)2^k\pi^{k+1}}, \tag{6}$$

where ζ is the generalized Riemann zeta function (Edwards, 2001). By using raw moments, expected value and variance of X calculated as

$$E(X) = \frac{\mathrm{sgn}(\lambda)}{2\pi^2}\left(-\zeta\left(2,\frac{2|\lambda|+\pi}{4\pi}\right) + \zeta\left(2,\frac{|\lambda|}{2\pi}+\frac{3}{4}\right) + 16C\right), \tag{7}$$

$$Var(X) = 1 - \frac{1}{4\pi^4}\left(-\zeta\left(2,\frac{2|\lambda|+\pi}{4\pi}\right) + \zeta\left(2,\frac{|\lambda|}{2\pi}+\frac{3}{4}\right) + 16C\right)^2. \tag{8}$$

where $C \cong 0.915966$ and known as the Catalan number. Limiting case of an odd moment is

$$\lim_{|\lambda|\to\infty} E\left(X^{2k-1}\right) = \frac{(2k-1)\big(\zeta(k+1,1/k+1) - \zeta(k+1,3/k+1)\big)}{(k+1)\pi^4\mathrm{sgn}(\lambda)}.$$

Thus

$$\lim_{|\lambda|\to\infty} E(X) = 8C/\pi^2\mathrm{sgn}(\lambda) = -0.7425 \times\ sgn(\lambda),$$

and

$$\lim_{|\lambda|\to\infty} Var(X) = 1 - \frac{64C^2}{\pi^4} = 0.44876.$$

It is obvious that $Var(X) = 1$ for $\lambda = 0$.

Skewness and Kurtosis: The third standardized moment is the skewness of a random variable X, which is defined as

$$SK_X = E\left(\frac{X-\mu_X}{\sigma_X}\right)^3 = \frac{1}{\sigma_X^3}\left(E(X^3) - 3\mu\sigma_X^2 - \mu_X^3\right)$$

where $\mu_X = E(X)$ and $\sigma_X^2 = Var(X)$. The skewness coefficient of random variable $X \sim SSGL(\lambda)$ can be calculated using eq(8), eq(9), and eq(10) as

$$SK_X(\lambda) = \frac{\begin{array}{c}\mathrm{sgn}(\lambda)^3\left(-\zeta(2,\Lambda_1) + \zeta(2,\Lambda_2) + 16C\right)^3 \\ -3\pi^2\mathrm{sgn}(\lambda)\begin{pmatrix} 2\pi^2\left(\zeta(2,\Lambda_2) - \zeta(2,\Lambda_1)\right) \\ +\zeta(4,\Lambda_1) - \zeta(4,\Lambda_2) \\ +32\pi^2 C - \zeta\left(4,\frac{1}{4}\right) + \zeta\left(4,\frac{3}{4}\right)\end{pmatrix}\end{array}}{4\pi^6\left(1 - \frac{\mathrm{sgn}(\lambda)^2\left(-\zeta(2,\Lambda_1) + \zeta(2,\Lambda_2) + 16C\right)^2}{4\pi^4}\right)^{3/2}},$$

where $\Lambda_1 = \frac{2|\lambda|+\pi}{4\pi}$ and $\Lambda_2 = \frac{|\lambda|}{2\pi} + \frac{3}{4}$. Considering the limit case,

$$SK_X(\lambda) \xrightarrow[|\lambda|\to\infty]{} \frac{-96\pi^4 C + 4096C^3 + 3\pi^2\left(\zeta\left(4,\frac{1}{4}\right) - \zeta\left(4,\frac{3}{4}\right)\right)}{\mathrm{sgn}(\lambda)4(\pi^4 - 64C^2)^{3/2}} = \mathrm{sgn}(\lambda) \times 1.7978$$

is obtained for $|\lambda| \to \infty$. As can be seen, the skewness of the distribution follows the same sign as the parameter λ. It is obvious that $SK_X(0) = 0$ and denotes the symmetric case.

The kurtosis coefficient of a random variable is defined as its 4th central moment and is expressed as

$$KR_X = E\left(\frac{X-\mu_X}{\sigma_X}\right)^4 = \frac{1}{\sigma_X^4}\left(E(X^4) - 4\mu_X E(X^3) + 6\mu_X^2 E(X^2) - 3\mu_X^4\right).$$

The skewness coefficient of random variable $X \sim SSGL(\lambda)$ can be calculated using eq(7), eq(8), and eq(10) as

$$KR_X(\lambda)=\frac{-\frac{3\mathrm{sgn}(\lambda)^4\begin{pmatrix}-\zeta(2,\Lambda_1)\\+\zeta(2,\Lambda_2)+16C\end{pmatrix}^4}{16\pi^8}+\frac{3\mathrm{sgn}(\lambda)^2\begin{pmatrix}-\zeta(2,\Lambda_1)\\+\zeta(2,\Lambda_2)+16C\end{pmatrix}\begin{pmatrix}\pi^2\left(\zeta(2,\Lambda_2)-\zeta(2,\Lambda_1)\right)\\+\zeta(4,\Lambda_1)-\zeta(4,\Lambda_2)\\+16\pi^2C-\zeta\left(4,\frac{1}{4}\right)+\zeta\left(4,\frac{3}{4}\right)\end{pmatrix}}{2\pi^6}+5}{\left(1-\frac{\mathrm{sgn}(\lambda)^2(-\zeta(2,\Lambda_1)+\zeta(2,\Lambda_2)+16C)^2}{4\pi^4}\right)^2}$$

It is easy the calculate $KR_X(0)=5$. The limiting case of kurtosis is

$$KR_X(\lambda)\xrightarrow[|\lambda|\to\infty]{}\frac{-12288C^4+384\pi^4C^2+4\pi^2C\left(\psi^{(3)}\left(\frac{3}{4}\right)-\psi^{(3)}\left(\frac{1}{4}\right)\right)+5\pi^8}{(\pi^4-64C^2)^2}=7.978.$$

Using numerical approach, we observe that the kurtosis coefficient is minimum at λ=1/2 and equal to 4.8981.

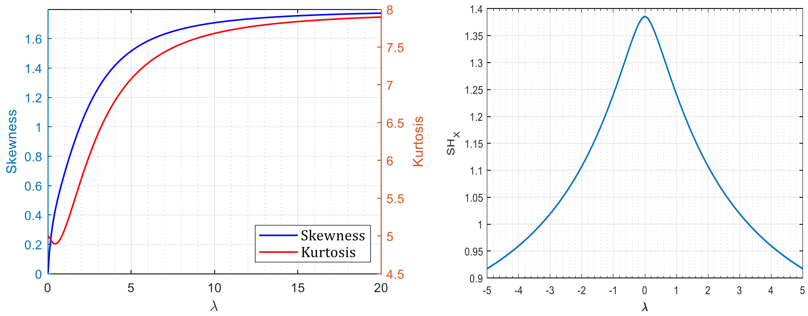

Figure 1. Skewness and kurtosis (left) and Shannon entropy (right) of $SSGL(\lambda)$ distribution according to λ.

Note that, $SK_X(-\lambda)=-SK_X(\lambda)$ and $KR_X(-\lambda)=KR_X(\lambda)$. The left panel of Figure 1 shows the effect of the λ parameter on the skewness and kurtosis coefficients in the range $\lambda\in[0,20]$, and can be interpreted for $\lambda\in[-20,20]$.

Shannon Entropy: Entropy is a measure of the variation or uncertainty of a random variable. Shannon entropy, defined as $SH_X=E(-\ln f_X(X))$, is the most well-known measure of entropy. The right panel of Figure 1 shows the Shannon entropy graph for the random variable $X\sim SSGL(\lambda)$ in the range of λ values between [-5,5]. The highest entropy value has been numerically observed to be 1.386 at $\lambda=0$. Given that the variance in eq(8) reaches its highest value of 1 at $\lambda=0$, we may argue that the uncertainty in the distribution reaches its maximum in the symmetric case. On the other hand, the sign of the λ parameter has no effect on the entropy value.

Location-Scale Extension: Location and scale parameters, μ and σ respectively, can be introduced by means of $X=\sigma Z+\mu$, where Z is a random variable with density eq(2). Thus, the pdf of X is obtained as

$$h(x;\mu,\sigma,\lambda)=\frac{e^{\frac{\pi}{2\sigma}(x-\mu)}}{\sigma\left(e^{\frac{\pi}{\sigma}(x-\mu)}+1\right)}\left[\operatorname{sgn}\big(\lambda(x-\mu)\big)\left(1-e^{-\left|\frac{\lambda(x-\mu)}{\sigma}\right|}\right)+1\right], \tag{9}$$

where $\mu\in\mathbb{R}$, $\sigma>0$, and $\lambda\in\mathbb{R}$. We use the notation $SSGL(\mu,\sigma,\lambda)$ for X. Figure 2 illustrates the pdf of the location scale extended $SSGL$ distribution with different parameter values.

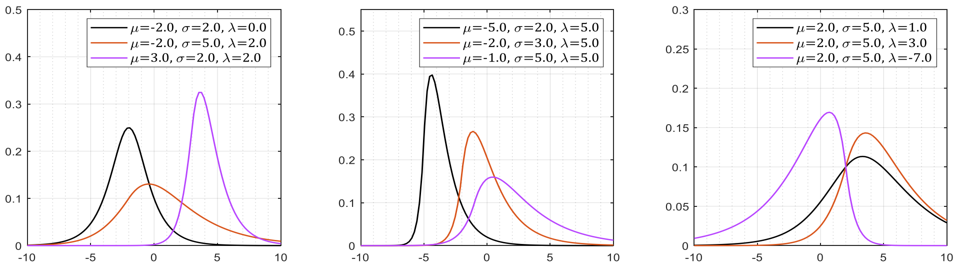

Figure 2. Plots of SSGL(μ,σ,λ) pdf.

3.Estimation of Parameters and Simulation

In this section, we will study the maximum likelihood estimators of the parameters of the $SSGL(\mu,\sigma,\lambda)$ distribution. Let $X_1, X_2, \ldots, X_n$ be a random sample from this distribution. The logarithmic likelihood function can be expressed as

$$L(\mu,\sigma,\lambda;\underline{X})=\sum_{i=1}^{n}\log e^{\frac{\pi(X_i-\mu)}{2\sigma}}\left(\frac{1-e^{-\left|\frac{\lambda(X_i-\mu)}{\sigma}\right|}}{\operatorname{sgn}\big(\lambda(X_i-\mu)\big)}+1\right)-\sum_{i=1}^{n}\log\sigma\left(e^{\frac{\pi(X_i-\mu)}{\sigma}}+1\right)$$

directly from eq(9). We obtain the following normal equations by taking the first derivatives of L with respect to μ,σ and λ, and setting them to zero:

$$\sum\log\left(\frac{e^{\frac{\pi\Delta}{2}}}{2\sigma^3(e^{\pi\Delta}+1)^2}\left(\begin{array}{c}\dfrac{-2(e^{\pi\Delta}+1)\left(\lambda\Delta\sigma\mathrm{Abs}'(\lambda\Delta)+\sigma^2\operatorname{sgn}'(\sigma)\left(e^{|\lambda\Delta|}-1\right)\right)}{\operatorname{sgn}(\lambda\Delta)e^{|\lambda\Delta|}}\\ +2\pi\Delta\sigma e^{\pi\Delta}\left(\dfrac{1-e^{-|\lambda\Delta|}}{\operatorname{sgn}(\lambda\Delta)}+1\right)-\pi\Delta\sigma(e^{\pi\Delta}+1)\left(\dfrac{1-e^{-|\lambda\Delta|}}{\operatorname{sgn}(\lambda\Delta)}+1\right)\\ -2\sigma(e^{\pi\Delta}+1)\left(\dfrac{1-e^{-|\lambda\Delta|}}{\operatorname{sgn}(\lambda\Delta)}+1\right)\end{array}\right)\right)=0$$

$$\sum\log\left(\frac{\Delta e^{\frac{\pi\Delta}{2}-|\lambda\Delta|}\left(\operatorname{sgn}(\lambda\Delta)\mathrm{Abs}'(\lambda\Delta)+\sigma\left(e^{|\lambda\Delta|}-1\right)\operatorname{sgn}'(\lambda\Delta)\right)}{\sigma(e^{\pi\Delta}+1)}\right)=0$$

$$\sum\log\left(\frac{e^{\frac{\pi\Delta}{2}-|\lambda\Delta|}}{2\sigma^2(e^{\pi\Delta}+1)^2}\left(\begin{array}{c}\dfrac{\pi(e^{\pi\Delta}-1)\left(e^{|\lambda\Delta|}-1\right)-2\lambda(e^{\pi\Delta}+1)\mathrm{Abs}'(\lambda\Delta)}{\operatorname{sgn}(\lambda\Delta)}\\ -\dfrac{2\lambda\sigma(e^{\pi\Delta}+1)\left(e^{|\lambda\Delta|}-1\right)}{\operatorname{sgn}'(\lambda\Delta)}+\dfrac{\pi(e^{\pi\Delta}-1)}{e^{-|\lambda\Delta|}}\end{array}\right)\right)=0$$

where $\Delta = (X_i - \mu)/\sigma$, respectively. Thus, the maximum likelihood (ML) estimates of the parameters μ, σ, λ, say $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\lambda}$, can be obtained by simultaneously and numerically solving these equations.

Monte-Carlo Simulation: We performed Monte Carlo simulation studies to illustrate the estimation performance of the obtained ML estimators. Since the quantile function of the distribution cannot be obtained analytically, the following algorithm can be used to generate random variables from the distribution.

Step 1. Set parameter values (μ, σ, λ)
Step 2. Generate $U \sim U(0,1)$
Step 3. Solve $H(Z;\lambda) - U = 0$ with respect to Z, where $H(Z;\lambda)$ is the cdf of $SSGL(\lambda)$
Step 4. Calculate $X = \sigma Z + \mu$

Different parameter values are used in Monte Carlo simulations. Table 1 shows the mean absolute bias (Bias) and mean squared error (MSE) values obtained from simulations repeated 1000 times for different sample sizes n=30, 50, 100 and 1000. The formula used for computing Bias and MSE are given in follows

$$Bias = \frac{1}{1000}\sum_{i=1}^{1000} \left|\theta - \hat{\theta}\right|,$$
$$MSE = \frac{1}{1000}\sum_{i=1}^{1000} \left(\theta - \hat{\theta}\right)^2,$$

where θ represents the real parameter value and $\hat{\theta}$ is the maximum likelihood estimate of θ.

Table 1. Monte-Carlo simulation results.

		$\hat{\mu}$		$\hat{\sigma}$		$\hat{\lambda}$	
Parameter values	n	Bias	MSE	Bias	MSE	Bias	MSE
$\mu = -2, \sigma = 2, \lambda = 1.5$	30	0.6046	0.6875	0.4017	0.2425	4.0394	124.85
	50	0.4872	0.4061	0.3508	0.1842	2.0410	29.356
	100	0.3175	0.1584	0.2419	0.0905	0.8769	1.3217
	1000	0.0920	0.0137	0.0795	0.0103	0.2414	0.0961
$\mu = -2, \sigma = 2, \lambda = 7.5$	30	0.2375	0.1482	0.3691	0.2056	15.434	544.14
	50	0.1593	0.0640	0.2754	0.1258	9.4193	290.77
	100	0.1052	0.0187	0.1953	0.0579	5.3101	107.08
	1000	0.0261	0.0011	0.0562	0.0046	0.7638	0.9050
$\mu = -4, \sigma = 6, \lambda = -2.5$	30	1.4335	4.1693	1.1915	2.1655	7.2708	248.88
	50	0.9649	1.6478	1.0101	1.5355	3.3004	76.773
	100	0.7528	0.9349	0.7843	0.9005	1.3555	7.4157
	1000	0.1963	0.0663	0.2164	0.0752	0.3088	0.1555
$\mu = -3, \sigma = 4, \lambda = -7.5$	30	0.4714	0.4785	0.6439	0.6683	16.324	601.18
	50	0.2830	0.1332	0.5064	0.4064	9.2901	267.51
	100	0.1966	0.0643	0.3789	0.2089	4.6045	90.125
	1000	0.0573	0.0051	0.1225	0.0233	0.7168	0.8628

As seen in Table 1, we conducted our simulations with high and low skewness and sigma values. When we examine bias and MSE values in this table as the sample size increases, both the Bias and MSE values decrease for all parameter values. This shows that the estimations are precise and accurate, implying that they are consistent and unbiased. Because the ML estimators are asymptotically unbiased, this is an expected result.

4.Application to Real Data

The purpose of this section is to demonstrate the usefulness of the $SSGL$ distribution by using two real-world data sets.

Australian Athletes Data: The first set is the heights (in centimeters) of 100 Australian athletes data (Telford & Cunningham, 1991), which is a popular data set in the literature, especially in studying skewed distributions. We employed Azzalini's skew Normal (SN) distribution and the well-known Normal (N) distribution to compare the modeling success of the SSGL distribution. The results are presented in Table 2 along with the maximum likelihood estimates, log-likelihood value (LH), Akaike information criterion (AIC) and Bayes information criterion (BIC) values, and Kolmogorov-Smirnov statistics with associated p-value (KS). In the same table, the observed values of some statistics and the theoretical values calculated by parameter estimates of these statistics are also presented.

Table 2. Summary of fits for Australian athletes data set.

	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	-LH	KS (p)	AIC	BIC
SSGL	177.920	8.762	-0.614	348.064	0.046 (0.98)	702.127	709.943
SN	182.269	11.232	-1.718	350.303	0.075 (0.60)	706.607	714.422
N	174.590	8.242	-	352.321	0.090 (0.37)	708.641	713.852
	Mean	Std.Dev.	Skewness	Kurtosis	Q1	Median	Q3
SSGL	174.636	8.122	-0.521	4.919	170.270	175.237	179.610
SN	174.524	8.135	-0.370	3.233	168.934	174.323	179.662
Data	174.594	8.242	-0.560	4.197	170.950	175.000	179.700

According to the KS values in Table 2, the goodness of fit of all three models could not be rejected. However, compared to the other two distributions based on LH, AIC, and BIC values, the SSGL distribution fits better. Considering the values reported in studies (Hasanalipour & Sharafi, 2012) and (Jamalizadeh, Behboodian, & Balakrishnan, 2008), the SSGL distribution is more successful than the alternatives mentioned in these studies.

Wind Speed Data: This dataset contains the average wind speeds recorded by the İstanbul Çatalca meteorological observatory (41°10'04.9"N, 28°29'27.1"E) in January 2020 at 2-hour intervals.

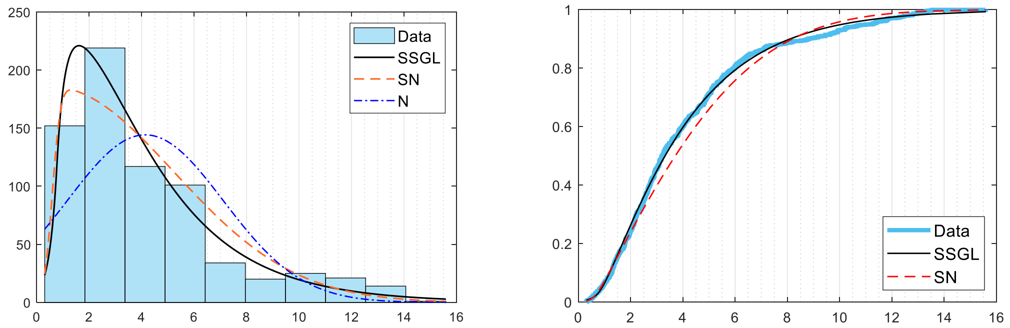

Figure 3. Histogram of wind speed data with fitted densities (left), empirical cdf and fitted cdf (right).

Examining Table 3, we find that the KS test does not accept the goodness of fit of the SN and N distributions, but the SSGL distribution is reasonable. AIC and BIC values also show that the SSGL distribution provides a better fit.

Table 3. Summary of fits for wind speed data set.

	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	-LH	KS (p)	AIC	BIC
SSGL	0.770	4.547	15.720	1607.526	0.0274 (0.65)	3221.052	3234.723
SN	0.591	4.631	18.140	1616.539	0.0768 (0.00)	3239.078	3252.748
N	4.138	2.979	-	1766.838	0.1299 (0.00)	3537.676	3546.789
	Mean	Std.Dev.	Skewness	Kurtosis	Q1	Median	Q3
SSGL	4.128	3.066	1.759	7.849	1.934	3.321	5.444
SN	4.280	2.799	0.983	3.855	2.219	3.836	6.000
Data	4.138	2.979	1.340	4.363	2.034	3.225	5.338

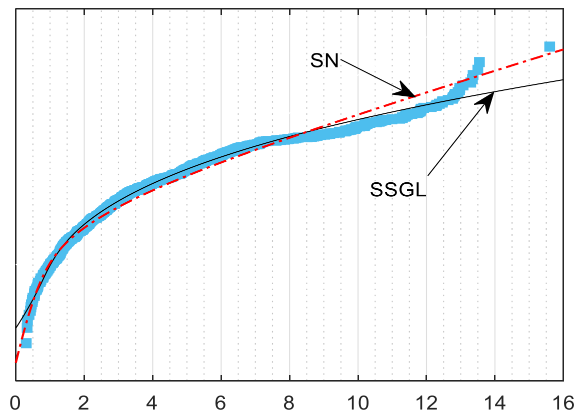

Figure 4. Probability plot for wind speed data.

Table 3 also includes some statistics of wind speed data. If the theoretical values of these statistics calculated with the estimated parameters for the SSGL and SN distributions are examined, one sees that the mean and standard deviation values of wind speed are more accurately estimated by SSGL. The same may be said for the first quartile (Q1), median, and third quartile (Q3). When we examine Figure 3 and Figure 4, it is seen that SSGL fits better than SN in most of the empirical distribution. At the end of the right tail, the SN distribution provides a better fit than SSGL distribution. This explains why the skewness of the data is better predicted by SN.

5.Conclusion

In this study, we derived a new skew-symmetric model called SSGL to model skewed data. The closed-form pdf and cdf of the resulting distribution were obtained in the study. Furthermore, statistically significant features of the distribution, such as raw moments, skewness and kurtosis coefficients, and Shannon entropy, have been investigated. In addition, maximum likelihood estimators for unknown parameters of the new distribution were studied. The performance of these estimators in the study was also compared to a series of Monte Carlo simulation studies that had been performed. Given the information obtained from the simulation study, it can be said that all obtained estimators of the SSGL parameters are asymptotically consistent and unbiased. Finally, the usability of the derived distribution has been exemplified by applications performed on two real-world datasets. In both samples, the SSGL distribution provides a better fit than Azzalini's SN distribution. As a result of this study, it was concluded that the SSGL distribution is a suitable alternative for modeling skewed data, especially with the help of computer programs.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The author declares that there is no conflict of interest.

Author's Contributions

The contribution of the author is 100%.

References

Abramowitz, M., Stegun, I. A. Handbook of mathematical functions with formulas, graphs, and mathematical tables. US Government printing office, New York 1964.

Altun E. The generalized Gudermannian distribution: inference and volatility modelling. Statistics 2019; 53(2): 364–386.

Arellano-Valle RB., Gómez HW., Quintana FA. A new class of skew-normal distributions. Communications in statistics-Theory and Methods 2004; 33(7): 1465–1480.

Azzalini A. A class of distributions which includes the normal ones. Scandinavian journal of statistics 1985; 171–178.

Dutka, J. (1981). The incomplete Beta function-a historical profile. Archive for history of exact sciences 1981; 24(1): 11–29.

Edwards, H. M. Riemann's zeta function. Courier Corporation, New York 2001.

Elal-Olivero D. Alpha-skew-normal distribution. Proyecciones Antofagasta 2010; 29(3): 224–240.

Gupta R., Kundu D. Introduction of shape/skewness parameter(s) in a probability distribution. J Probab Stat Sci 2009; 7.

Hasanalipour P., Sharafi M. A new generalized Balakrishnan skew-normal distribution. Statistical Papers 2012; 53(1): 219–228.

Jamalizadeh A., Behboodian J., Balakrishnan N. A two-parameter generalized skew-normal distribution. Statistics & probability letters 2008; 78(13): 1722–1726.

Ma Y., Genton MG. Flexible class of skew-symmetric distributions. Scandinavian Journal of Statistics 2004; 31(3): 459–468.

Nekoukhou V., Alamatsaz MH., Aghajani AH. A flexible skew-generalized normal distribution. Communications in Statistics-Theory and Methods 2013; 42(13): 2324–2334.

Telford, R. D., Cunningham, R. B. Sex, sport, and body-size dependency of hematology in highly trained athletes. Medicine and science in sports and exercise 1991; 23: 788–794.