Geometric skew-Cauchy distribution as an alternative to the skew-normal and geometric skew-normal distributions

by

Mr. Xander Heaney 16118848

Contact details: xan97der@gmail.com

Submitted in partial fulfillment of the degree

Master of Science Advanced Data Analytics

In the Faculty of Natural & Agricultural Sciences

Supervisor: Dr. R. Ehlers

Co-supervisor: Prof. A. Bekker

Department of Statistics, University of Pretoria

Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

January 10, 2022

Declaration

I, *Xander Heaney*, declare that this dissertation, submitted in partial fulfillment of the degree *MSc Advanced Data Analytics* at the University of Pretoria, is my own work and has not been previously submitted at this or any other tertiary institution.

____

Mr. X Heaney

____

Dr. R. Ehlers

____

Prof. A. Bekker

____

January 10, 2022

Acknowledgements

This dissertation and work contained therein would not be possible without support. I would like to thank the following contributors:

- This research fall under ethics approval number NAS124/2019.
- This research was funded in part by the National Research Foundation (NRF) of South Africa, Reference: SRUG190308422768 grant No. 120839; Academic Statistics Funding Instrument grant No. 127946; SARChI Research Chair UID71199. The financial assistance of the NRF towards this research is acknowledged. Opinions expressed and conclusions arrived at in this dissertation are those of the authors and are not necessarily to be attributed to the NRF.
- A warm thank you to Prof. A. Jamalizadeh for his insights, guidance and suggestion to consider the geometric and Cauchy distributions.
- Prof A. Bekker, thank you for your involvement, thoughts, guidance and suggestions - specifically towards the end where your experience and knowledge helped finish the dissertation.
- To Dr R. Ehlers, thank you for being my supervisor. Thank you for all your constructive feedback, inquisitive mindset, reassurance and know that I admire your intellect. Most of all, thank you for your patience with me.
- To all my friends that supported me, thank you. I know that most of the time I said no to plans and your understanding thereof did not go by unnoticed. Thank you for dealing with me during difficult times.
- To my father, thank you for your prayers, support, love and for your reassurance that everything will work out in my favour.
- Gopika Ramkilawon, my best friend, colleague and study partner - thank you. The past 6 years has been an invaluable journey to share with you. Specifically, the last 2 years were very tough and I am blessed that you were by my side through it all. Thank you for every message of support, all your kind words, understanding and unconditional love. Thank you for teaching me the foundations of coding. Without your time, I will not be where I am today.
- I would like to dedicate this dissertation to all those who never had the opportunity. The opportunity to pursue tertiary education, the opportunity to learn and the opportunity to experience. This is also for those who believe that they are not good enough to accomplish, for those who are convinced that they are not able - for I know what that feels like.

- To my mother, Yvonne Heaney, thank you. You are my biggest supporter, best friend and greatest personal example of what hard work looks like. My emotional toughness and determination comes from you. Thank you for all your messages of support, all your prayers and all your love. There are many unspoken sacrifices that you made for me to be able to have the opportunity to be where I am in my life - physically, emotionally and spiritually. I love you. Always.

- Last, but certainly not least, to my heavenly Father. May I never forget how dependent I am on God, my Father who shows me undeserved mercy; Jesus, my saviour who never leaves my side and the Holy Spirit, my never-ending strength. May I never forget:

He, has never left me alone.

Abstract

The skew-normal distribution was popularised by Azzalini [4] to model skewed data. However, the skew-normal distribution is always unimodal. Kundu [24] recently presented the geometric skew-normal distribution by considering a geometric compounding sum of normal random variables. This distribution is more flexible than the skew-normal distribution since it can be multimodal.

In this dissertation we present a new distribution namely the geometric skew-Cauchy distribution. The idea follows a similar approach to that of Kundu's. The difference, however, is that we consider a geometric compounding sum of Cauchy random variables.

The inclusion of a simulation and application chapter demonstrates the practical use of this new distribution. It turns out that the geometric skew-Cauchy distribution is also more flexible than the skew-normal distribution.

It is concluded that this new distribution can be used as an alternative to the geometric skew-normal distribution since both distributions can be multimodal. The advantage over the geometric skew-normal distribution, is the ability of the geometric skew-Cauchy distribution to model fatter-tailed data.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

AIC - Akaike's Information Criterion

BIC - Bayesian Information Criterion

CDF - Cumulative distribution function

CGF - Cumulant generating function

EM - Expectation-Maximisation

i.i.d. - Independent and identically distributed

KS - Kolmogorov-Smirnov

MGF - Moment generating function

MLE - Maximum likelihood estimate

NR - Newton-Raphson

PDF - Probability density function

PMF - Probability mass function

Notations

$N(\cdot)$ - Normal distribution with parameters in brackets

$SN(\cdot)$ - Skew-normal distribution with parameters in brackets

$GE(\cdot)$ - Geometric distribution with parameter in brackets

$GSN(\cdot)$ - Geometric skew-normal distribution with parameters in brackets

$C(\cdot)$ - Cauchy distribution with parameters in brackets

$GSC(\cdot)$ - Geometric skew-Cauchy distribution with parameters in brackets

$SN_d(\cdot)$ - d-variate skew-normal distribution with parameters in brackets

$MVGSN_d(\cdot)$ - d-variate geometric skew-normal distribution with parameters in brackets

$MVGSC_d(\cdot)$ - d-variate geometric skew-Cauchy distribution with parameters in brackets

$arctan(\cdot)$ - inverse tangent

$\mathbb{R}$ - Real numbers

$\mathbb{R}^d$ - Real numbers in d dimensions

$L(\cdot)$ - Likelihood function

$l(\cdot)$ - Log-likelihood function

$l_c(\cdot)$ - Log-likelihood function based on the complete sample

$\phi(\cdot)$ - Standard normal distribution PDF

$\phi_d(\cdot)$ - d-variate standard normal distribution PDF

$\Phi(\cdot)$ - Standard normal distribution CDF

$M(t)$ - Moment generating function

$E(\cdot)$ - Expected value

$\text{var}(\cdot)$ - Variance

$b(\cdot)$ - Mill's ratio

$G(\cdot)$ - Cumulant generating function

γ_1 - Skewness

γ_2 - Kurtosis

$\gamma_{1,d}^M$ - Multivariate skewness

$\gamma_{2,d}^M$ - Multivariate kurtosis

$k_.$ - Cumulant

T - Transpose

$\stackrel{d}{=}$ - Equal in distribution

ϵ - Tolerance error

1 Chapter 1

1.1 Introduction and rationale

Developing an alternative distribution to the skew-normal and geometric skew-normal distributions serve as a clear motive for exploring this topic. The motive to delve into the theory is sustained by the fact that there is no literature that explicitly propose a distribution that entails the compounding sum of independent random variables that involve the geometric and Cauchy distributions. Establishing an extensive understanding of the theoretical construction of the skew-normal and geometric skew-normal distributions, form the basis for further investigation into an alternative consideration of a compounding sum of independent random variables. The rationale includes the fact that conducting research to develop an alternative distribution, will yield value in the financial and medical fields (among others). These include right-skewed income distributions of a population in a life assurance environment [13] as well as the right-skewed survival times of guinea pigs infected with tubercle bacilli [24].

This dissertation will follow a descriptive research approach with elements of experimental research present in the form of outcomes produced through simulation. Findings in the dissertation will be supported through implementation on real data.

1.2 Aims and objectives

The main goals of this dissertation that are expected to sustain the rationale are as follows:

- Discussing the skew-normal distribution.
- Discussing the geometric skew-normal distribution.
- Developing a new alternative distribution to the skew-normal and geometric skew-normal distributions.
- Investigating and developing the estimation of the alternative distribution.
- Providing an overview of the multivariate skew-normal distribution.
- Providing an overview of the multivariate geometric skew-normal distribution.
- Demonstrating a simulation study involving the new alternative distribution.
- Illustrating findings on real data.
- Providing consideration for future work that includes the multivariate extension of the new alternative distribution.

1.3 Literature review

The skew-normal distribution is presented by Azzalini [4]. The fundamental characteristics and other mathematical properties of Azzalini's skew-normal distribution received significant attention due to its flexibility [4]. This distribution includes the standard normal distribution as a special case. Moreover, the skew-normal distribution has a unimodal density function having both positive and negative skewness in attendance. A method to simulate random numbers from the skew-normal distribution is discussed by Henze [20]. Existing symmetric probability density functions (PDFs) can be skewed using the skewing methodology that is presented by Azzalini and Capitanio [6].

In contrast to the several attractive properties that the skew-normal distribution has, it presents a problem in expanding statistical inference procedures. The inferential issues regarding Azzalini's skew-normal distribution is highlighted by Pewsey [32], Gupta and Gupta [18], Yalcinkaya et al. [41] and Kundu [24]. In addition to the latter, Pewsey [32] also provided reasons to consider a more appropriate parameterization of the skew-normal distribution. Azzalini and Capitanio [6] noted that the skew-normal distribution cannot be used to model fat or moderate-tailed data and that the skew-normal distribution is well-known to have thin tails. Gupta and Gupta [18] noted that the maximum likelihood estimates (MLEs) for the unknown parameters may not exist for the skew-normal distribution in several cases.

A skew flexible-normal distribution is derived by Gomez et al. [15] as an extension to the skew-normal distribution which provides support for both unimodal and bimodal distributions.

To address the problem of modelling fat-tailed data, literature suggests the use of other skewed distributions frequently called skew-symmetric distributions. However, these skew-symmetric distributions do not accommodate for multimodality [18] [32]. Azzalini and Capitanio [6] provided and excellent piece of work with detailed discussions on different skew-symmetric distributions.

Kundu [24] proposed a new three-parameter skewed normal distribution to address the inferential issues regarding Azzalini's skew-normal distribution and the lack of accommodation for multimodality. The new distribution is called the geometric skew-normal distribution and is based on a construction of variables involving the geometric and normal distributions [24]. The distribution yields the MLEs of the unknown parameters quite conveniently using the expectation-maximization (EM) algorithm [24]. It is noted that the geometric skew-normal distribution is very flexible and the PDF can be unimodal or multimodal. Moreover, the geometric skew-normal distribution can be written as an infinite mixture of normal distributions [24].

No literature has been found up till now that explicitly propose a distribution that entails the construction of random variables involving the geometric and Cauchy distributions that accommodate for both multimodality as well as fatter tails in the data. It is of interest to utilize the Cauchy distribution since the Cauchy distribution is well-known to accommodate fatter tails than the normal distribution [40]. In fact, the Cauchy distribution accommodates fatter tails than the double-exponential distribution [40] and thus fatter tails in the data could be modelled more conveniently.

Extensions of the univariate distributions to multivariate distributions have been investigated. The multivariate extension of the univariate skew-normal distribution is presented by Azzalini and Valle [7]. Another extension of the univariate skew-normal distribution to the multivariate case is presented by Gupta and Chen [17]. The latter extension includes the multivariate extension of Azzalini and Valle [7] as a special case. Furthermore, Harrar and Gupta [19] derived a matrix variate skew-normal distribution.

Kundu [25] introduced a multivariate extension of the univariate geometric skew-normal distribution. The multivariate case addresses the inferential issues encountered in the multivariate skew-normal distribution by Azzalini and Valle [7]. Furthermore, the multivariate geometric skew-normal distribution compensates for the lack of skewed distributions available in higher dimensions [25]. The multivariate geometric skew-normal distribution yields a PDF that can be unimodal or multimodal and the MLEs of the unknown parameters can be obtained in a convenient manner in high dimensions [25].

It is of interest to consider future work on other multivariate extensions to address multimodality and fatter tails in the data. No literature has been found up till now that propose a multivariate extension to a distribution that entails the construction of random variables involving the geometric and Cauchy distributions. Providing consideration of the multivariate extension creates a basis for development of an additional skewed distribution in higher dimensions.

1.4 Dissertation outline

- In Chapter 2 the skew-normal distribution is revisited. The chapter is split into two parts, with the first part discussing the univariate case and the second part providing an overview of the multivariate case. The chapter proceeds to give theorems, a collection of some characteristics, algorithms on how to generate random numbers and outlines estimation theory.

- In Chapter 3 we revisit the univariate geometric skew-normal distribution that was presented by Kundu [24]. In addition, an overview of Kundu's [25] multivariate extension thereof is also provided. The chapter revisits theorems and derivations, discusses some characteristics, gives some conditional properties, outlines generation of random numbers and discusses estimation theory.

- An alternative distribution to the skew-normal and geometric skew-normal distributions is presented in Chapter 4. The chapter is devoted to developing a new distribution called the geometric skew-Cauchy distribution. The chapter proceeds to derive theorems, give a generation of random numbers algorithm, derives conditional properties and investigates the estimation of the new distribution.

Comparisons among the alternative distribution and the skew-normal and geometric skew-normal distributions are also given.

- Chapter 5 is devoted to evaluating the performance of the new alternative distribution. This is done by presenting a simulation study and also fitting the model to two real data sets. The models in Chapters 2 and 3 are also fitted for comparison. It is noted that the new alternative distribution is a competitive model against the skew-normal and geometric skew-normal models.
- The dissertation is concluded in Chapter 6, where a summary on what has been done is given as well as some conclusive remarks on the new distribution. Furthermore, consideration for future work is also outlined in this chapter.
- Two appendices (A and B respectively) are found at the end of this dissertation. Appendix A gives some fundamental definitions, lemmas and results for use throughout the dissertation. Appendix B gives some code that was used in the dissertation.

2 Chapter 2

Chapter 2 will discuss the skew-normal distribution. In Section 2.1, the univariate skew-normal distribution is revisited. An overview of the multivariate skew-normal distribution is presented in Section 2.2.

2.1 Univariate skew-normal distribution

This subsection will proceed to revisit the univariate skew-normal distribution. Figure 1 gives a summary of how Section 2.1 will proceed and what is expected from this subsection.

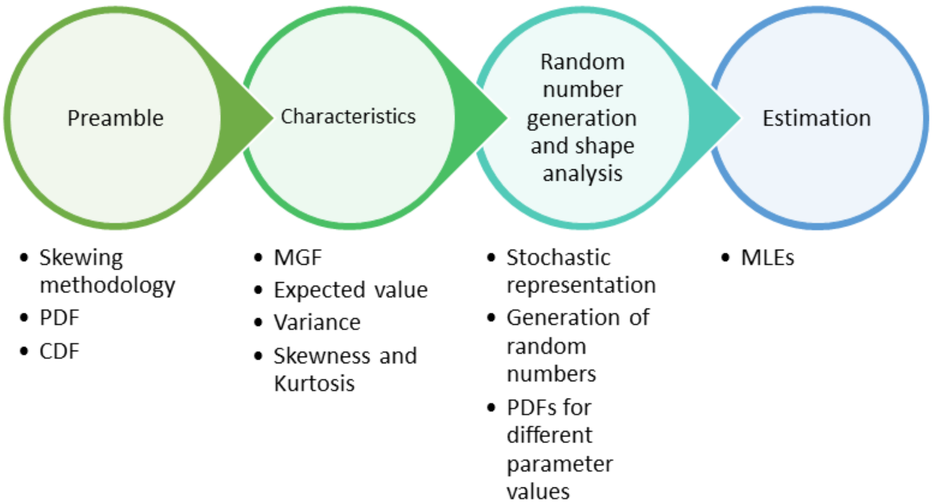


Figure 1: Summary of Section 2.1.

2.1.1 Preamble

The skew-normal distribution can be obtained by utilizing the skewing methodology that is presented by Azzalini and Capitanio [6]. A brief overview of the skewing methodology is stated in Proposition 1. This methodology is used to skew existing symmetric PDFs.

Proposition 1. *Denote by* $g_0(\cdot)$ *a PDF on* $\mathbb{R}^d$, *by* $F_0(\cdot)$ *a continuous cumulative distribution function (CDF) on* $\mathbb{R}$, *and by* $k(\cdot)$ *a real-valued function on* $\mathbb{R}^d$, *such that* $g_0(-x) = g_0(x), k(-x) = -k(x)$ *and* $F_0(-y) = 1 - F_0(y)$ *for all* $x \in \mathbb{R}^d, y \in \mathbb{R}$. *Then*

$$g_X(x) = 2g_0(x)F_0\{k(x)\} \tag{1}$$

is a PDF on $\mathbb{R}^d$ *[6].*

It should be noted that the symmetric base PDF is termed by $g_0(\cdot)$ and that the skewing mechanism is termed by $2F_0\{k(x)\}$. Consequently, the skewed version of the symmetric base PDF is termed by $g_X(\cdot)$.

In this subsection, using the notation that is stated in Proposition 1, the case where $g_0(\cdot) = \phi(\cdot)$, $F_0(\cdot) = \Phi(\cdot)$ (with $\phi(\cdot)$ and $\Phi(\cdot)$ representing the standard normal PDF and CDF) and $k(x) = \lambda x$ for $\lambda \in \mathbb{R}$, is used to obtain the corollaries below.

Theorem 1. *A random variable* X *has the skew-normal distribution if its PDF is given by the following:*

$$f_X(x) = 2\phi(x)\Phi(\lambda x) \tag{2}$$

where $\lambda \in \mathbb{R}$. *This is denoted by* $X \sim SN(\lambda)$ *[4].*

For use in applied work, location and scale parameters are introduced. The introduction is done via the transformation:

$$Y = \mu + \sigma X$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. The following corollary follows then immediately from (2).

Corollary 1. *If* $X \sim SN(\lambda)$ *and* $Y = \mu + \sigma X$ *then the random variable* Y *is said to have the skew-normal distribution with location parameter* μ *and scale parameter* σ*. Its PDF is given by the following:*

$$f_Y(y) = \frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\left(\frac{y-\mu}{\sigma}\right)\right) \tag{3}$$

where $-\infty < y < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$ *and* $\lambda \in \mathbb{R}$. *This is denoted by* $Y \sim SN(\mu, \sigma^2, \lambda)$.

Proof. Let $X \sim SN(\lambda)$ with the PDF as given in (2). Consider the random variable $Y = \mu + \sigma X$, where the location and scale parameters are denoted by $\mu \in \mathbb{R}$ and $\sigma > 0$ respectively. If $y = \mu + \sigma x$, then $u^{-1}(y) = \frac{y-\mu}{\sigma} = x$ with $\frac{d}{dy}u^{-1}(y) = \frac{1}{\sigma}$. From Bain and Engelhardt [8] and using (2), it follows that

$$\begin{aligned} f_Y(y) &= f_X\left(u^{-1}(y)\right)\left|\frac{d}{dy}u^{-1}(y)\right| \\ &= 2\phi\left(u^{-1}(y)\right)\Phi\left(\lambda(u^{-1}(y))\right)\left|\frac{d}{dy}\left(u^{-1}(y)\right)\right| \\ &= 2\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\left(\frac{y-\mu}{\sigma}\right)\right)\left|\frac{1}{\sigma}\right| \\ &= \frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\left(\frac{y-\mu}{\sigma}\right)\right). \end{aligned}$$

Thus, $Y \sim SN(\mu, \sigma^2, \lambda)$. □

Theorem 2. *A random variable X that follows the skew-normal distribution, that is $X \sim SN(\lambda)$ with PDF given in (2), has a CDF of the following form:*

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} 2\phi(t)\Phi(\lambda t)dt \\ &= \int_{-\infty}^{x} 2\phi(t)\int_{-\infty}^{\lambda t}\phi(u)dudt \\ &= 2\int_{-\infty}^{x}\int_{-\infty}^{\lambda t}\phi(t)\phi(u)dudt \\ &= \Phi(x) - T(x,\lambda) \end{aligned} \tag{4}$$

for $x, \lambda \in \mathbb{R}$, where $T(x, \lambda)$ is Owen's T-function [30] which is defined as

$$T(x,\lambda) = \frac{1}{\sqrt{2\pi}}\int_0^{\lambda}\frac{e^{-\frac{1}{2}x^2\left(1+t^2\right)}}{1+t^2}dt$$

for $x, \lambda \in \mathbb{R}$.

Proof. The full proof is provided in Azzalini [4]. □

2.1.2 Characteristics of the skew-normal distribution

The computation of the characteristics of $Y \sim SN(\mu, \sigma^2, \lambda)$ is done via the moment generating function (MGF) or, equivalently but somewhat more practical, via the cumulant generating function (CGF) (see (109) in Appendix A.1).

Moment generating function

Before presenting the MGF of a random variable with PDF given in (3), consider first the derivation of Lemma 1 given in (113) (see Appendix A.2).

Theorem 3. *If $X \sim SN(\lambda)$ then the MGF of the random variable $Y = \mu + \sigma X$ with PDF given in (3) is given by*

$$M_Y(t) = 2e^{t\mu+\frac{1}{2}t^2\sigma^2}\Phi(\delta\sigma t) \tag{5}$$

with $t \in \mathbb{R}$, $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ and $\Phi(\cdot)$ denoting the standard normal CDF [6].

Proof. Using (108) in Appendix A.1 and the PDF in (2), it follows that

$$\begin{aligned}
M_Y(t) &= E\left[e^{tY}\right] \\
&= E\left[e^{t(\mu+\sigma X)}\right] \\
&= \int_{\mathbb{R}} e^{t\mu+t\sigma x} 2\phi(x)\Phi(\lambda x)dx \\
&= 2e^{t\mu}\int_{\mathbb{R}} e^{t\sigma x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\Phi(\lambda x)dx \\
&= 2e^{t\mu}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(x^2-2t\sigma x\right)}\Phi(\lambda x)dx.
\end{aligned}$$

It can be observed that

$$\begin{aligned}
(x-t\sigma)^2 &= x^2 - 2t\sigma x + t^2\sigma^2 \\
\implies x^2 - 2t\sigma x &= (x-t\sigma)^2 - t^2\sigma^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
M_Y(t) &= 2e^{t\mu}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left((x-t\sigma)^2-t^2\sigma^2\right)}\Phi(\lambda x)dx \\
&= 2e^{t\mu+\frac{1}{2}t^2\sigma^2}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-t\sigma)^2}\Phi(\lambda x)dx.
\end{aligned}$$

Now, let $r = x - t\sigma$, then $x = r + t\sigma$ with $\frac{dx}{dr} = 1$. Let $\phi(\cdot)$ denote the standard normal PDF as before, then it is true that

$$\begin{aligned}
M_Y(t) &= 2e^{t\mu+\frac{1}{2}t^2\sigma^2}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}r^2}\Phi(\lambda(r+t\sigma))dr \\
&= 2e^{t\mu+\frac{1}{2}t^2\sigma^2}\int_{\mathbb{R}} \phi(r)\Phi(\lambda(r+t\sigma))dr \\
&= 2e^{t\mu+\frac{1}{2}t^2\sigma^2}E_R[\Phi(\lambda R + \lambda t\sigma)] \qquad (6)
\end{aligned}$$

where $R \sim N(0,1)$. Now, applying the result from Lemma 1 (see (113) in Appendix A.2), it follows from (6) that

$$\begin{aligned}
M_Y(t) &= 2e^{t\mu+\frac{1}{2}t^2\sigma^2}E_R[\Phi(\lambda R + \lambda t\sigma)] \\
&= 2e^{t\mu+\frac{1}{2}t^2\sigma^2}\Phi\left(\frac{\lambda t\sigma}{\sqrt{1+\lambda^2}}\right) \\
&= 2e^{t\mu+\frac{1}{2}t^2\sigma^2}\Phi\left(\delta\sigma t\right)
\end{aligned}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. □

Expected value, variance, skewness and kurtosis

The method used in this section follows in a similar fashion to that of what is described by Azzalini and Captanio [6]. The method computes the expected value, variance, third central moment and fourth central moment by using the cumulant generating function, $G_Y(t)$, as stated in (109) in Appendix A.1. The inverse Mills ratio, $b(\cdot)$, will be utilized in order to simplify the derivations of the central moments. The definition of the inverse Mills ratio is given by (111) in Appendix A.1.

Before proceeding with the derivations of the remaining characteristics, it is worthwhile to first consider some of the properties of the inverse Mills ratio. These properties, as well as their respective derivations are given in (118), (119) and (120) in Appendix A.3.

Expected value

Theorem 4. *Consider $Y \sim SN\left(\mu, \sigma^2, \lambda\right)$ with MGF given in (5), then the expected value is given by*

$$E[Y] = \mu + \delta\sigma\sqrt{\frac{2}{\pi}}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Proof. Using (109), (110), (111) in Appendix A.1, (116), (118) in Appendix A.3 and the fact that the first cumulant is the expected value [26], it follows from (5) that

$$\begin{aligned}
E[Y] &= \frac{d}{dt}G_Y(t)\Big|_{t=0} \\
&= \frac{d}{dt}\left(\log M_Y(t)\right)\Big|_{t=0} \\
&= \frac{d}{dt}\left(\log\left(2e^{t\mu+\frac{1}{2}t^2\sigma^2}\Phi(\delta\sigma t)\right)\right)\Big|_{t=0} \\
&= \frac{d}{dt}\left(t\mu + \frac{1}{2}t^2\sigma^2 + \log(2\Phi(\delta\sigma t))\right)\Big|_{t=0} \\
&= \left[\mu + \sigma^2 t + \frac{2\delta\sigma\phi(\delta\sigma t)}{2\Phi(\delta\sigma t)}\right]\Big|_{t=0} \qquad [1] \\
&= \left[\mu + \sigma^2 t + \delta\sigma b(\delta\sigma t)\right]\Big|_{t=0} \\
&= \mu + \delta\sigma b(0) \\
&= \mu + \delta\sigma\sqrt{\frac{2}{\pi}}
\end{aligned}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. □

The following is worth noting on the above proof:

[1] this step is labelled as (A^{**}) for later use.

Variance

Theorem 5. *Consider $Y \sim SN\left(\mu, \sigma^2, \lambda\right)$ with MGF given in (5), then the variance is given by*

$$\text{var}[Y] = \sigma^2\left(1 - \frac{2}{\pi}\delta^2\right)$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Proof. Using (109), (110), (111) in Appendix A.1, (118), (119) in Appendix A.3 and the fact that the second cumulant is the expected value [26], it follows from (5) that

$$\begin{aligned}
\text{var}[Y] &= \frac{d^2}{dt^2}G_Y(t)\bigg|_{t=0} \\
&= \frac{d}{dt}\left(\frac{d}{dt}\log M_Y(t)\right)\bigg|_{t=0} \\
&= \frac{d}{dt}\left(\mu + \sigma^2 t + \frac{\delta\sigma\phi(\delta\sigma t)}{\Phi(\delta\sigma t)}\right)\bigg|_{t=0} && [1] \\
&= \sigma^2 + \delta\sigma\frac{d}{dt}b(\delta\sigma t)\bigg|_{t=0} && [2] \\
&= \sigma^2 + \delta\sigma\left(-b(\delta\sigma t)[\delta\sigma t + b(\delta\sigma t)]\frac{d}{dt}(\delta\sigma t)\right)\bigg|_{t=0} \\
&= \sigma^2 + (\delta\sigma)^2(-b(\delta\sigma t)[\delta\sigma t + b(\delta\sigma t)])\big|_{t=0} \\
&= \sigma^2 + (\delta\sigma)^2\left(-(b(0))^2\right) \\
&= \sigma^2 + (\delta\sigma)^2\left(-\frac{2}{\pi}\right) \\
&= \sigma^2\left(1 - \frac{2}{\pi}\delta^2\right)
\end{aligned}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. □

The following is worth noting on the above proof:

[1] using (A^{**}),

[2] this step is labelled as (B^{**}) for later use.

Skewness

Standardisation of the third cumulant leads to the well-known measure of Fisher-Pearson moments skewness [31]. The Fisher-Pearson moment skewness, γ_1, is given by

$$\gamma_1 = \frac{E\left[(Y - E[Y])^3\right]}{(\text{var}[Y])^{\frac{3}{2}}}.$$

Theorem 6. *Consider $Y \sim SN\left(\mu, \sigma^2, \lambda\right)$ with MGF given in (5), then the third central moment is given by*

$$E\left[(Y - E[Y])^3\right] = \frac{1}{2}(4 - \pi)\left(\delta\sigma\sqrt{\frac{2}{\pi}}\right)^3$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Proof. Using (110) in Appendix A.1, (118), (120) in Appendix A.3 and the fact that the third cumulant is the third central moment [26], it follows from (5) that

$$\begin{aligned}
E\left[(Y - E[Y])^3\right] &= \frac{d^3}{dt^3}G_Y(t)\Big|_{t=0} \\
&= \frac{d}{dt}\left(\sigma^2 + \delta\sigma\frac{d}{dt}b(\delta\sigma t)\right)\Big|_{t=0} \qquad [1] \\
&= (\delta\sigma)^3 b''(\delta\sigma t)\big|_{t=0} \\
&= (\delta\sigma)^3\left(-b(\delta\sigma t) + (\delta\sigma t)^2 b(\delta\sigma t) + 3(\delta\sigma t)(b(\delta\sigma t))^2 + 2(b(\delta\sigma t))^3\right)\big|_{t=0} \\
&= (\delta\sigma)^3\left(-b(0) + 2(b(0))^3\right) \\
&= (\delta\sigma)^3\left(-\sqrt{\frac{2}{\pi}} + 2\left(\sqrt{\frac{2}{\pi}}\right)^3\right) \\
&= (\delta\sigma)^3\left(\frac{2}{\pi}\right)^{3/2}\left(2 - \frac{\pi}{2}\right) \\
&= \left(\delta\sigma\sqrt{\frac{2}{\pi}}\right)^3\frac{1}{2}(4 - \pi) \\
&= \frac{1}{2}(4 - \pi)\left(\delta\sigma\sqrt{\frac{2}{\pi}}\right)^3
\end{aligned}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. ☐

The following is worth noting on the above proof:

[1] using (B^{**}).

Utilizing the third central moment (i.e. the third cumulant), the Fisher-Pearson moment skewness is then yielded as

$$\gamma_1 = \frac{\frac{1}{2}(4-\pi)\left(\delta\sigma\sqrt{\frac{2}{\pi}}\right)^3}{\left(\sigma^2\left(1-\frac{2}{\pi}\delta^2\right)\right)^{\frac{3}{2}}}.$$

Kurtosis

The fourth central moment of the skew-normal distribution can be obtained by utilizing the following relationship

$$E\left[(Y-E[Y])^4\right] = k_4 + 3\left(E\left[(Y-E[Y])^2\right]\right)^2 \tag{7}$$

where $k_4 = \frac{d^4}{dt^4}G_Y(t)\Big|_{t=0}$ is derived in a similar fashion to that of Theorem 6 (presented below) and $E\left[(Y-E[Y])^2\right]$ is the variance of the distribution [26] [6].

Theorem 7. *Consider $Y \sim SN\left(\mu,\sigma^2,\lambda\right)$ with MGF given in (5), then the fourth central moment is given by*

$$E\left[(Y-E[Y])^4\right] = 2(\delta\sigma)^4\frac{4}{\pi^2}(\pi-3) + 3\sigma^4\left(1-\frac{2}{\pi}\delta^2\right)^2$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Proof. Using (110) given in Appendix A.1 as well as (119) and (120) given in Appendix A.3, it follows that

$$\begin{aligned}
k_4 &= \frac{d}{dt}\left[\frac{d^3}{dt^3}G_Y(t)\right]\Bigg|_{t=0} \\
&= \left[-(\delta\sigma)^4 b'(\sigma\delta t) - (\delta\sigma)^5 t b''(\delta\sigma t) - (\delta\sigma)^4 b'(\sigma\delta t) - (\delta\sigma)^4 2b'(\sigma\delta t)b'(\sigma\delta t) - (\delta\sigma)^4 2b''(\sigma\delta t)b(\sigma\delta t)\right]\Big|_{t=0} \\
&= -(\delta\sigma)^4 b'(0) - (\delta\sigma)^4 b'(0) - (\delta\sigma)^4 2[b'(0)]^2 - 2(\delta\sigma)^4 b''(0)b(0) \\
&= (\delta\sigma)^4\left[-2b'(0) - 2[b'(0)]^2 - 2b(0)b''(0)\right] \\
&= 2(\delta\sigma)^4\left[\frac{2}{\pi} - \frac{4}{\pi^2} - \sqrt{\frac{2}{\pi}}\left[2\left(\frac{2}{\pi}\right)^{\frac{3}{2}} - \sqrt{\frac{2}{\pi}}\right]\right] \\
&= 2(\delta\sigma)^4\left[\frac{2}{\pi} - \frac{4}{\pi^2} - 2\frac{4}{\pi^2} + \frac{2}{\pi}\right] \\
&= 2(\delta\sigma)^4\frac{4}{\pi}\left(1-\frac{3}{\pi}\right) \\
&= 2(\delta\sigma)^4\frac{4}{\pi^2}(\pi-3).
\end{aligned}$$

Now, using the relationship in (7) and the variance derived prior, it follows that

$$\begin{aligned} E\left[(Y-E[Y])^4\right] &= k_4 + 3\left(E\left[(Y-E[Y])^2\right]\right)^2 \\ &= 2(\delta\sigma)^4 \frac{4}{\pi^2}(\pi-3) + 3\sigma^4\left(1-\frac{2}{\pi}\delta^2\right)^2 \end{aligned}$$

□

Utilizing the fourth central moment produces the well-known measure of kurtosis, γ_2 [36], given by

$$\begin{aligned} \gamma_2 &= \frac{E\left[(Y-E[Y])^4\right]}{(\mathrm{var}[Y])^2} \\ &= \frac{2(\pi-3)\left(\delta\sigma\sqrt{\frac{2}{\pi}}\right)^4}{\left(\sigma^2\left(1-\frac{2}{\pi}\delta^2\right)\right)^2}. \end{aligned}$$

2.1.3 Generation of random numbers and illustration of PDF

In order to generate random numbers from the $SN\left(\mu,\sigma^2,\lambda\right)$ distribution with PDF given in (3), it is necessary to provide a stochastic representation of the distribution. This approach follows the suggestion by Henze [20].

Theorem 8. *If $W_1 \sim N(0,1)$ and $W_2 \sim N(0,1)$ are two independent normally distributed random variables, then*

$$X = \frac{\lambda|W_1| + W_2}{\sqrt{1+\lambda^2}} \sim SN(\lambda).$$

Proof. Let $W_1 \sim N(0,1)$ and $W_2 \sim N(0,1)$ be independent and let $m = \frac{\lambda}{\sqrt{1+\lambda^2}}$, $n = \frac{1}{\sqrt{1+\lambda^2}}$ and $X = \frac{\lambda|W_1|+W_2}{\sqrt{1+\lambda^2}} = m|W_1| + nW_2$. Then

$$\begin{aligned} P[X \le x] &= E_{W_1}\left[P\left[X \le x \mid W_1 = w_1\right]\right] \\ &= \int_{\mathbb{R}} P\left[m\left|w_1\right| + nW_2 \le x\right]\phi\left(w_1\right)dw_1 \\ &= \int_{-\infty}^{0} P\left[m\left|w_1\right| + nW_2 \le x\right]\phi\left(w_1\right)dw_1 + \int_{0}^{\infty} P\left[m\left|w_1\right| + nW_2 \le x\right]\phi\left(w_1\right)dw_1 \end{aligned} \tag{8}$$

noting that W_1 is symmetric about $w_1 = 0$, therefore

$$\begin{aligned} \int_{-\infty}^{0} P\left[m\left|w_1\right| + nW_2 \le x\right]\phi\left(w_1\right)dw_1 &= \int_{0}^{\infty} P\left[m\left|w_1\right| + nW_2 \le x\right]\phi\left(w_1\right)dw_1 \\ &= \int_{0}^{\infty} P\left[mw_1 + nW_2 \le x\right]\phi\left(w_1\right)dw_1. \end{aligned} \tag{9}$$

It can be observed from (8) and (9) that

$$
\begin{aligned}
P[X \leq x] &= 2 \int_0^{\infty} P\left[mw_1 + nW_2 \leq x\right] \phi\left(w_1\right) dw_1 \\
&= 2 \int_0^{\infty} P\left[W_2 \leq \frac{x - mw_1}{n}\right] \phi\left(w_1\right) dw_1 \\
&= 2 \int_0^{\infty} \Phi\left(\frac{x - mw_1}{n}\right) \phi\left(w_1\right) dw_1. \qquad (10)
\end{aligned}
$$

Using a well-known result from Bain and Engelhardt [8] (see (116) in Appendix A.3), it follows from (10) that

$$
\begin{aligned}
f_X(x) &= \frac{d}{dx} P[X \leq x] \\
&= 2 \int_0^{\infty} \frac{d}{dx} \Phi\left(\frac{x - mw_1}{n}\right) \phi\left(w_1\right) dw_1 \\
&= 2 \int_0^{\infty} \phi\left(\frac{x - mw_1}{n}\right) \frac{1}{n} \phi\left(w_1\right) dw_1 \\
&= 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(x^2 - 2xmw_1 + m^2 w_1^2\right)}{2n^2}} \frac{1}{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} w_1^2} dw_1 \\
&= 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2n^2}} \int_0^{\infty} \frac{1}{\sqrt{2\pi n^2}} e^{-\frac{w_1^2}{2} - \frac{\left(-2xmw_1 + m^2 w_1^2\right)}{2n^2}} dw_1 \\
&= 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2n^2}} \int_0^{\infty} \frac{1}{\sqrt{2\pi n^2}} e^{-\frac{\left(w_1^2 n^2 - 2xmw_1 + m^2 w_1^2\right)}{2n^2}} dw_1 \\
&= 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2n^2}} \int_0^{\infty} \frac{1}{\sqrt{2\pi n^2}} e^{-\frac{\left(w_1^2\left(m^2 + n^2\right) - 2xmw_1\right)}{2n^2}} dw_1 \\
&= 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2n^2}} \int_0^{\infty} \frac{1}{\sqrt{2\pi n^2}} e^{-\frac{\left(w_1^2 - 2xmw_1 + x^2 m^2\right)}{2n^2}} e^{\frac{x^2 m^2}{2n^2}} dw_1 \qquad \left(\text{since } m^2 + n^2 = 1 \text{ and induced } \pm \frac{x^2 m^2}{2n^2}\right) \\
&= 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2n^2} + \frac{x^2 m^2}{2n^2}} \int_0^{\infty} \frac{1}{\sqrt{2\pi n^2}} e^{-\frac{(w_1 - mx)^2}{2n^2}} dw_1 \\
&= 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \int_0^{\infty} \frac{1}{\sqrt{2\pi n^2}} e^{-\frac{(w_1 - mx)^2}{2n^2}} dw_1 \qquad \left(\text{since } -\frac{x^2}{2n^2} + \frac{x^2 m^2}{2n^2} = -\frac{x^2}{2}\right) \\
&= 2\phi(x) \int_0^{\infty} \frac{1}{\sqrt{2\pi n^2}} e^{-\frac{(w_1 - mx)^2}{2n^2}} dw_1. \qquad (11)
\end{aligned}
$$

Let $k = \frac{w_1 - mx}{n}$ then it is true that $u^{-1}(k) = kn + mx = w_1$ and $\frac{d}{dk} u^{-1}(k) = n$. It is important to note that there is a change in the bounds of the integral in (11). If $w_1 = 0$ then it implies that the lower bound becomes $k = \frac{-mx}{n}$. The upper bound does not change. Applying the transformation it follows from (11) that

$$
\begin{aligned}
f_X(x) &= 2\phi(x)\int_{\frac{-mx}{n}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{k^2}{2}}dk \\
&= 2\phi(x)\int_{\frac{-mx}{n}}^{\infty}\phi(k)dk \\
&= 2\phi(x)\left(\lim_{p\to\infty}\Phi(p) - \lim_{p\to\infty}\Phi\left(\frac{-mx}{n}\right)\right) \\
&= 2\phi(x)\left(1-\Phi\left(\frac{-mx}{n}\right)\right) \\
&= 2\phi(x)\Phi\left(\frac{mx}{n}\right). \qquad (12)
\end{aligned}
$$

Since $\frac{m}{n} = \frac{\frac{\lambda}{\sqrt{1+\lambda^2}}}{\frac{1}{\sqrt{1+\lambda^2}}} = \lambda$ it yields from (12) that

$$
f_X(x;\lambda) = 2\phi(x)\Phi(\lambda x).
$$

Hence, it is true that $X \sim SN(\lambda)$ with PDF as given in (2). □

Corollary 2. *If $W_1 \sim N(0,1)$ and $W_2 \sim N(0,1)$ are two independent normally distributed random variables, then*

$$
\begin{aligned}
Y &= \mu + \sigma X \\
&= \mu + \sigma\frac{\lambda|W_1| + W_2}{\sqrt{1+\lambda^2}} \sim SN\left(\mu, \sigma^2, \lambda\right) \qquad (13)
\end{aligned}
$$

with PDF as given in (3).

Theorem 8 and Corollary 2 provide a representation that can be used to conveniently generate values randomly from the skew-normal distribution. A short algorithm is provided that summarises the generation from the $SN\left(\mu,\sigma^2,\lambda\right)$ distribution.

Algorithm 1 Generating a random sample of size n from the $SN\left(\mu,\sigma^2,\lambda\right)$ distribution

---

1: ***Required:***
   - *Define the value of μ for $\mu \in \mathbb{R}$.*
   - *Define the value of σ for $\sigma > 0$.*
   - *Define the value of λ for $\lambda \in \mathbb{R}$.*

2: *Generate the independent values W_1 and W_2 from the $N(0,1)$ distribution.*

3: *Recall the stochastic representation that is given in (13). Define the variable Y accordingly, using steps 1 and 2.*

4: *Repeat steps 2 and 3, n times to obtain the n values from the distribution of Y.*

Using the PDF representation that is given in (3), Figure 2 gives a visual representation of the PDFs for the $SN\left(\mu, \sigma^2, \lambda\right)$ distribution. These PDFs are plotted with $\mu = 0$ and $\sigma^2 = 1$ kept constant throughout, varying the values of λ.

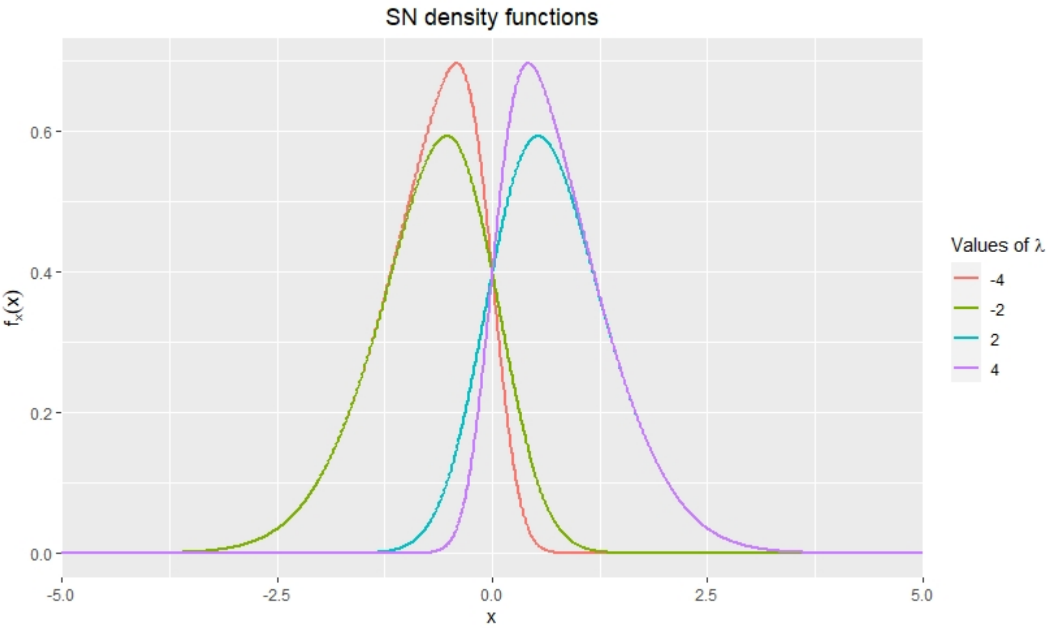


Figure 2: PDFs of the distribution with PDF in (3) for different values of λ.

It can be observed from Figure 2 that the skew-normal distribution exhibits a positive skewness for $\lambda > 0$ and that the shape of the distribution becomes more skewed as the value of λ increases. It can also be seen from Figure 2 that the skew-normal distribution exhibits a negative skewness for $\lambda < 0$. In addition to the aforementioned, it can be observed that the skew-normal distribution exhibits a strictly *unimodal PDF* [6].

The characteristics that include the expected value, variance, skewness and kurtosis are given in Table 1 for different values of λ with $\mu = 0$ and $\sigma = 1$ kept constant throughout.

Table 1: Values of some characteristics of the skew-normal distribution for different values of λ

λ	$E[Y]$	$var[Y]$	γ_1	γ_2
-4	-0.774	0.400	-0.784	0.632
-2	-0.714	0.491	-0.454	0.305
2	0.714	0.491	0.454	0.305
4	0.744	0.400	0.784	0.632

It can be observed from Table 1 that the expected value increases as the value of λ increases. The same observation holds for the skewness. However, as $|\lambda|$ decreases, the kurtosis decreases. It is observed that the variance increases as $|\lambda|$ decreases.

The characteristics that include the expected value, variance, skewness and kurtosis are given in Table

2 for different values of μ with $\lambda = 2$ and $\sigma = 1$ kept constant throughout.

Table 2: Values of some characteristics of the skew-normal distribution for different values of μ

μ	$E[Y]$	$var[Y]$	γ_1	γ_2
-4	-3.286	0.491	0.454	0.305
-2	-1.286	0.491	0.454	0.305
2	2.714	0.491	0.454	0.305
4	4.714	0.491	0.454	0.305

It can be observed from Table 2 that the expected value increases as the value of μ increases. The remaining characteristics remain unchanged as the value of μ changes. This is due to the fact that the expected value, skewness and kurtosis are not functions of μ.

The characteristics that include the expected value, variance, skewness and kurtosis are given in Table 3 for different values of σ with $\mu = 2$ and $\lambda = 2$ kept constant throughout.

Table 3: Values of some characteristics of the skew-normal distribution for different values of σ

σ	$E[Y]$	$var[Y]$	γ_1	γ_2
0.1	2.071	0.005	0.454	0.305
0.5	2.357	0.123	0.454	0.305
0.9	2.642	0.397	0.454	0.305
1.3	2.928	0.829	0.454	0.305

It can be observed from Table 3 that the expected value increases as the value of σ increases. Furthermore, the variance increases as the value of σ increases. The remaining characteristics remain unchanged as the value of σ changes.

2.1.4 Estimation

The points which maximize the log-likelihood function in the parameter space of the distribution is known as the maximum likelihood estimators [41]. Various literature note the difficulty that is encountered with maximum likelihood estimates (MLEs) of the skew-normal distribution; see Gupta and Gupta [18], Yalcinkaya et al. [41], Pewsey [32] and Kundu [24]. This section will proceed to give the log-likelihood function as well as the normal equations of the skew-normal distribution.

Maximum likelihood estimators

Theorem 9. *Let $\{y_1, y_2, ..., y_n\}$ be a random sample of size n from $Y \sim SN\left(\mu, \sigma^2, \lambda\right)$ with pdf as given in (3). The log-likelihood function is then given by*

$$l\left(\mu, \sigma, \lambda\right) = n\ln 2 - n\ln\sigma - \frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2 + \sum_{i=1}^{n}\ln\left[\Phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)\right]. \tag{14}$$

Proof. From Bain and Engelhardt [8] and using (3), it follows that

$$\begin{aligned}
l(\mu,\sigma,\lambda) &= \ln\left[\prod_{i=1}^{n} f_{Y_i}(y_i)\right] \\
&= \sum_{i=1}^{n} \ln\left[f_{Y_i}(y_i)\right] \\
&= \sum_{i=1}^{n} \ln\left[\frac{2}{\sigma}\phi\left(\frac{y_i-\mu}{\sigma}\right)\Phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)\right] \\
&= n\ln 2 - n\ln\sigma + \sum_{i=1}^{n}\ln\left[\phi\left(\frac{y_i-\mu}{\sigma}\right)\right] + \sum_{i=1}^{n}\ln\left[\Phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)\right] \\
&= n\ln 2 - n\ln\sigma + \sum_{i=1}^{n}\left[\ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2}\right]\right] + \sum_{i=1}^{n}\ln\left[\Phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)\right] \\
&= n\ln 2 - n\ln\sigma + \sum_{i=1}^{n}\left[\frac{-1}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right] + \sum_{i=1}^{n}\ln\left[\Phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)\right] \\
&= n\ln 2 - n\ln\sigma - \frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2 + \sum_{i=1}^{n}\ln\left[\Phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)\right].
\end{aligned}$$

□

It is proceeded to obtain the normal equations by taking the partial derivatives of the log-likelihood function in (14) with respect to the unknown parameters and equating them to 0. The normal equations are subsequently given in (15), (16) and (17).

$$\frac{\partial l(\mu,\sigma,\lambda)}{\partial\mu} = \frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right) - \frac{\lambda}{\sigma}\sum_{i=1}^{n}\frac{\phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)}{\Phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)} = 0, \tag{15}$$

$$\frac{\partial l(\mu,\sigma,\lambda)}{\partial\sigma} = -\frac{2n}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n}(y_i-\mu)^2 + \lambda\sum_{i=1}^{n}\frac{\phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)}{\Phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)}(y_i-\mu) = 0, \tag{16}$$

$$\frac{\partial l(\mu,\sigma,\lambda)}{\partial\lambda} = \sum_{i=1}^{n}\frac{\phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)}{\Phi\left(\lambda\left(\frac{y_i-\mu}{\sigma}\right)\right)}\left(\frac{y_i-\mu}{\sigma}\right) = 0. \tag{17}$$

The solutions to the normal equations are called the maximum likelihood estimators of the unknown parameters. Since the normal equations contain non-linear functions defined by $\omega(z_i) = \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)}$ where $z_i = \frac{y_i-\mu}{\sigma}$, the exact solutions of the normal equations cannot be obtained [41]. This briefly outlines the inferential problem that is evident with the skew-normal distribution. Furthermore, Azzalini and Capitanio [5] noted that at $\lambda = 0$ of the profile log-likelihood, there is always an inflection point. In addition to the latter, the expected Fisher information matrix becomes singular at $\lambda = 0$ [5]. Yalcinkaya et al. [41] suggest the use of iterative methods to try and solve the normal equations. In addition to the work from Yalcinkaya et al. [41], other iterative methods to deal with the non-linearity of the normal equations can be viewed in the excellent monograph by Kantar and Senoglu [22]. An example, is to resort to the use of the quasi-Newton method by Dennis and More [12] in order to solve the normal equations.

However, Kundu [24] developed the more flexible geometric skew-normal distributions to circumvent the inferential issues encountered. Moreover, Kundu [24] developed the geometric skew-normal distribution to address the fact that the PDF of the skew-normal distribution is strictly unimodal.

2.2 Multivariate skew-normal distribution

This subsection will proceed to provide an overview of the multivariate skew-normal distribution. A brief summary of proceedings in Section 2.2 is provided in Figure 3.

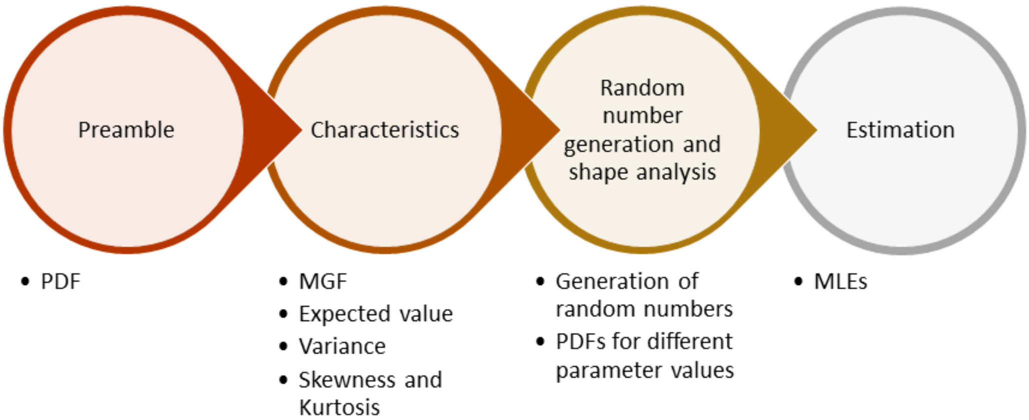


Figure 3: Summary of Section 2.2.

2.2.1 Preamble

The multivariate extension of (2) was momentarily discussed by Azzalini [4]. The formal extensions of (2) and (3) was discussed by Azzalini and Capitanio [6]. The formal extension of the univariate skew-normal distribution to the multivariate skew-normal distribution constitutes the most convenient option that involves a modulation factor of Gaussian type. This modulation factor utilizes a multivariate normal base density [6]. The multivariate extensions presented in the theorems below are still of the type presented in Proposition 1.

Theorem 10. *Consider the d-dimensional random variable* $\mathbf{Z} = (Z_1, Z_2, ..., Z_d)^T$. *Then* $\mathbf{Z}$ *has the multivariate skew-normal distribution if its PDF is given by the following:*

$$f_d(\mathbf{z}) = 2\phi_d(\mathbf{z}; \mathbf{\Omega})\, \Phi\left(\boldsymbol{\alpha}^T \mathbf{z}\right) \tag{18}$$

where

$$
\begin{aligned}
\boldsymbol{\alpha}^T &= \frac{\boldsymbol{\lambda}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Delta}^{-1}}{\left(1+\boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{\lambda}\right)^{\frac{1}{2}}} \\
\boldsymbol{\Delta} &= \operatorname{diag}\left(\left(1-\delta_1^2\right)^{\frac{1}{2}}, \ldots,\left(1-\delta_d^2\right)^{\frac{1}{2}}\right) \\
\boldsymbol{\lambda} &= \left(\lambda\left(\delta_1\right), \ldots, \lambda\left(\delta_d\right)\right)^{\mathrm{T}} \\
\boldsymbol{\Omega} &= \boldsymbol{\Delta}\left(\boldsymbol{\Psi}+\boldsymbol{\lambda} \boldsymbol{\lambda}^T\right) \boldsymbol{\Delta} \\
\delta_i &= \frac{\lambda_i}{\sqrt{1+\lambda_i^2}} \text{ for } i=1,2, \ldots, d \text{ and } \lambda_i \in \mathbb{R} \\
\boldsymbol{\Psi}&: d \times d \text{ correlation matrix}
\end{aligned}
$$

and noting that $\phi_d(\mathbf{z};\boldsymbol{\Omega})$ represents the PDF of the d-variate normal distribution with standardised marginals and the positive definite $d \times d$ covariance matrix $\boldsymbol{\Omega}$. This is denoted by $\boldsymbol{Z} \sim SN_d(\boldsymbol{\Omega},\boldsymbol{\alpha})$ [6].

It is proceeded to introduce location and scale parameters for the use in applied work. The introduction is done via the transformation:

$$
\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\omega}\boldsymbol{Z}
$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\omega}$ =diag$(\omega_1, \omega_2, ..., \omega_d)$ with ω_i positive, leading to the general form of $SN_d(\cdot)$ variables. The next corollary is immediate from the transformation [7].

Corollary 3. *The d-dimensional random variable Y has the multivariate skew-normal distribution with location parameter $\boldsymbol{\mu}$ and scale parameter $\boldsymbol{\omega}$ if its PDF is given by the following:*

$$
f_d(\boldsymbol{y}) = 2\phi_d\left(\boldsymbol{y}-\boldsymbol{\mu};\boldsymbol{\Omega}^*\right)\Phi\left(\boldsymbol{\alpha}^T\left(\frac{\boldsymbol{y}-\boldsymbol{\mu}}{\boldsymbol{\omega}}\right)\right) \tag{19}
$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\omega}$ =diag$(\omega_1, \omega_2, ..., \omega_d)$ and $\boldsymbol{\Omega^*} = \boldsymbol{\omega}\boldsymbol{\Omega}\boldsymbol{\omega}$. The latter implying that the covariance structure is not affected by the introduction of skewness since $\boldsymbol{\omega}$ is assumed to be diagonal. This is denoted by $\boldsymbol{Y} \sim SN_d(\boldsymbol{\mu},\boldsymbol{\Omega^*},\boldsymbol{\alpha})$.

Proof. Let $\boldsymbol{Z} \sim SN_d(\boldsymbol{\Omega},\boldsymbol{\alpha})$ with PDF as given in (18). Consider the d-dimensional random variable $\boldsymbol{Y} = \boldsymbol{\mu}+\boldsymbol{\omega}\boldsymbol{Z}$, where the location and scale parameters are denoted by $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\omega}$ =diag$(\omega_1, \omega_2, ..., \omega_d)$ respectively. If $\boldsymbol{y} = \boldsymbol{\mu}+\boldsymbol{\omega}\boldsymbol{z}$, then $\boldsymbol{u^{-1}}(\boldsymbol{y}) = \frac{\boldsymbol{y}-\boldsymbol{\mu}}{\boldsymbol{\omega}} = \boldsymbol{z}$ with $\frac{d}{d\boldsymbol{y}}\boldsymbol{u^{-1}}(\boldsymbol{y}) = \frac{\mathbf{1}}{\boldsymbol{\omega}}$. Also, if $\boldsymbol{\omega}\boldsymbol{u^{-1}}(\boldsymbol{y}) = \boldsymbol{y}-\boldsymbol{\mu} = \boldsymbol{\omega}\boldsymbol{z}$ with $\frac{d}{d\boldsymbol{y}}\boldsymbol{u^{-1}}(\boldsymbol{y}) = 1$, then it is true that $\boldsymbol{\omega}\boldsymbol{z} \sim SN_d(\boldsymbol{\Omega^*},\boldsymbol{\alpha})$ where $\boldsymbol{\Omega^*} = \boldsymbol{\omega}\boldsymbol{\Omega}\boldsymbol{\omega}$. Thus, from Bain and

Engelhardt [8] and using (18), it follows that

$$
\begin{aligned}
f_d(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Omega^*},\boldsymbol{\alpha}) &= f_d(\boldsymbol{z};\boldsymbol{\Omega},\boldsymbol{\alpha})\left|\frac{d}{d\boldsymbol{y}}\boldsymbol{u^{-1}}(\boldsymbol{y})\right| \\
&= f_d(\boldsymbol{u^{-1}}(\boldsymbol{y});\boldsymbol{\Omega},\boldsymbol{\alpha})\left|\frac{d}{d\boldsymbol{y}}\boldsymbol{u^{-1}}(\boldsymbol{y})\right| \\
&= 2\phi_d\left(\boldsymbol{u^{-1}}(\boldsymbol{y});\boldsymbol{\Omega}\right)\Phi\left(\boldsymbol{\alpha}^T\boldsymbol{u^{-1}}(\boldsymbol{y})\right)\left|\frac{1}{\boldsymbol{\omega}}\right| \\
&= 2|\boldsymbol{\omega}|^{-1}\phi_d\left(\frac{\boldsymbol{y}-\boldsymbol{\mu}}{\boldsymbol{\omega}};\boldsymbol{\Omega}\right)\Phi\left(\boldsymbol{\alpha}^T\left(\frac{\boldsymbol{y}-\boldsymbol{\mu}}{\boldsymbol{\omega}}\right)\right) \\
&= 2\phi_d\left(\boldsymbol{y}-\boldsymbol{\mu};\boldsymbol{\Omega}^*\right)\Phi\left(\boldsymbol{\alpha}^T\left(\frac{\boldsymbol{y}-\boldsymbol{\mu}}{\boldsymbol{\omega}}\right)\right)
\end{aligned}
$$

where the last step follows since $\boldsymbol{\omega z} \sim SN_d(\boldsymbol{\Omega^*},\boldsymbol{\alpha})$ with $\boldsymbol{\Omega^*} = \boldsymbol{\omega\Omega\omega}$. Thus, $\boldsymbol{Y} \sim SN_d(\boldsymbol{\mu},\boldsymbol{\Omega^*},\boldsymbol{\alpha})$. □

2.2.2 Characteristics of the multivariate skew-normal distribution

This section will proceed by revisiting select characteristics of the multivariate skew-normal distribution. These characteristics include the likes of the expected value and variance (among others). Before revisiting some of the characteristics, firstly the MGF of the random variable with PDF given in (19) will be derived.

Moment generating function

Before deriving the MGF, it is worthwhile to consider the theory presented in Lemmas 2 and 3 that are given in (114) and (115) respectively (see Appendix A.2). These Lemmas present convenient theory that ease the derivations of some of the remaining characteristics of the multivariate skew-normal distribution that will be discussed. Lemma 2 is an instantaneous extension of Lemma 1.

Theorem 11. *The MGF of a d-dimensional random variable $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\omega Z}$ with PDF as given in (19) is given by*

$$
M_{\boldsymbol{Y}}(\boldsymbol{t}) = 2e^{\left\{\boldsymbol{t}^T\boldsymbol{\mu}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Omega}^*\boldsymbol{t}\right\}}\Phi\left(\boldsymbol{\delta}^T\boldsymbol{\omega t}\right) \tag{20}
$$

with $\boldsymbol{t} \in \mathbb{R}^d, \boldsymbol{Z} \sim SN_d(\boldsymbol{\Omega},\boldsymbol{\alpha}), \boldsymbol{\delta} = \frac{\boldsymbol{\Omega\alpha}}{\sqrt{1+\boldsymbol{\alpha^T\Omega\alpha}}}$ and $\Phi(\cdot)$ denoting the standard normal CDF.

Proof. Consider the d-dimensional random variable $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\omega Z}$ where $\boldsymbol{Y} \sim SN_d(\boldsymbol{\mu},\boldsymbol{\Omega^*},\boldsymbol{\alpha})$ and $\boldsymbol{Z} \sim SN_d(\boldsymbol{\Omega},\boldsymbol{\alpha})$. Then from Johnson and Wichern [21], using (19) and (115) in Appendix A.2 and the fact that $\boldsymbol{\omega z} = \boldsymbol{y} - \boldsymbol{\mu}$, it follows that

$$\begin{aligned}
M_{\boldsymbol{Y}}(\boldsymbol{t}) &= E\left[e^{\boldsymbol{t}^T\boldsymbol{Y}}\right] \\
&= E\left[e^{\boldsymbol{t}^T(\boldsymbol{\mu}+\boldsymbol{\omega Z})}\right] \\
&= \int_{\mathbb{R}^d} e^{\boldsymbol{t}^T\boldsymbol{\mu}+\boldsymbol{t}^T\boldsymbol{\omega z}} 2\phi_d(\boldsymbol{z};\boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}^T\boldsymbol{z})d\boldsymbol{z} \\
&= e^{\boldsymbol{t}^T\boldsymbol{\mu}} \int_{\mathbb{R}^d} 2e^{\boldsymbol{t}^T\boldsymbol{\omega z}} 2\phi_d(\boldsymbol{z};\boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}^T\boldsymbol{z})d\boldsymbol{z} \\
&= e^{\boldsymbol{t}^T\boldsymbol{\mu}} \int_{\mathbb{R}^d} 2e^{\boldsymbol{t}^T(\boldsymbol{y}-\boldsymbol{\mu})} 2\phi_d(\boldsymbol{y}-\boldsymbol{\mu};\boldsymbol{\Omega}^*)\Phi(\boldsymbol{\alpha}^T\boldsymbol{\omega}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}))d(\boldsymbol{y}-\boldsymbol{\mu}) \\
&= 2e^{\boldsymbol{t}^T\boldsymbol{\mu}} \times e^{\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Omega}^*\boldsymbol{t}} \times \Phi\left(\boldsymbol{\delta}^T\boldsymbol{\omega t}\right) \\
&= 2e^{\left\{\boldsymbol{t}^T\boldsymbol{\mu}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Omega}^*\boldsymbol{t}\right\}}\Phi\left(\boldsymbol{\delta}^T\boldsymbol{\omega t}\right)
\end{aligned}$$

where $\boldsymbol{\delta} = \frac{\boldsymbol{\Omega\alpha}}{\sqrt{1+\boldsymbol{\alpha^T\Omega\alpha}}}$ and $\boldsymbol{\Omega}^* = \boldsymbol{\omega\Omega\omega}$. □

Expected value, variance, skewness and kurtosis

The method that will be used to derive the characteristics follows in a similar fashion to that of what is described by Azzalini and Capitanio [6]. As before, the inverse Mills ratio will be utilized in order to simplify the derivations. The definition can be viewed in (111) (see Appendix A.1).

Expected value

Theorem 12. *Consider the d-dimensional random variable $\boldsymbol{Y} \sim SN_d(\boldsymbol{\mu}, \boldsymbol{\Omega}^*, \boldsymbol{\alpha})$ with MGF given in (20), then the expected value is given by*

$$E[\boldsymbol{Y}] = \boldsymbol{\mu^T} + \boldsymbol{\delta^T\omega}\sqrt{\frac{2}{\pi}}$$

where $\boldsymbol{\delta} = \frac{\boldsymbol{\Omega\alpha}}{\sqrt{1+\boldsymbol{\alpha^T\Omega\alpha}}}$.

Proof. Using (109), (111) in Appendix A.1, (116), (118) in Appendix A.3, it follows from (20) that

$$
\begin{aligned}
E[\boldsymbol{Y}] &= \frac{d}{d\boldsymbol{t}} G_{\boldsymbol{Y}}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}} \\
&= \frac{d}{d\boldsymbol{t}} log M_{\boldsymbol{Y}}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}} \\
&= \frac{d}{d\boldsymbol{t}} \boldsymbol{t^T \mu} + \frac{1}{2}\boldsymbol{t^T \Omega^* t} + log\left(2\Phi\left(\boldsymbol{\delta^T \omega t}\right)\right)\Big|_{\boldsymbol{t}=\boldsymbol{0}} \\
&= \boldsymbol{\mu^T} \frac{d}{d\boldsymbol{t}}\left[\left(\frac{1}{2}\boldsymbol{t^T \Omega^* t}\right)\right] + \frac{d}{d\boldsymbol{t}}\left[log\left(2\Phi\left(\boldsymbol{\delta^T \omega t}\right)\right)\right]\Big|_{\boldsymbol{t}=\boldsymbol{0}} \\
&= \boldsymbol{\mu^T} + \frac{d}{d\boldsymbol{t}}\left[\left(\frac{1}{2}\boldsymbol{\Omega^* t}\right)^T \boldsymbol{t}\right] + \frac{d}{d\boldsymbol{t}}\left[\left(\frac{1}{2}\boldsymbol{t^T \Omega^*}\right)\boldsymbol{t}\right] + \frac{d}{d\boldsymbol{t}}\left[log\left(2\Phi\left(\boldsymbol{\delta^T \omega t}\right)\right)\right]\Big|_{\boldsymbol{t}=\boldsymbol{0}} \\
&= \boldsymbol{\mu^T} + \frac{1}{2}\boldsymbol{t^T \Omega^{*T}} + \frac{1}{2}\boldsymbol{t^T \Omega^*} + \frac{\frac{d}{d\boldsymbol{t}} 2\Phi\left(\boldsymbol{\delta^T \omega t}\right)}{2\Phi\left(\boldsymbol{\delta^T \omega t}\right)}\Bigg|_{\boldsymbol{t}=\boldsymbol{0}} \\
&= \boldsymbol{\mu^T} + \frac{1}{2}\boldsymbol{t^T}\left(\boldsymbol{\Omega^{*T}} + \boldsymbol{\Omega^*}\right) + \frac{2\boldsymbol{\delta^T \omega}\phi\left(\boldsymbol{\delta^T \omega t}\right)}{2\Phi\left(\boldsymbol{\delta^T \omega t}\right)}\Bigg|_{\boldsymbol{t}=\boldsymbol{0}} \\
&= \boldsymbol{\mu^T} + \frac{1}{2}\boldsymbol{t^T}\left(\boldsymbol{\Omega^{*T}} + \boldsymbol{\Omega^*}\right) + \boldsymbol{\delta^T \omega} b\left(\boldsymbol{\delta^T \omega t}\right)\Big|_{\boldsymbol{t}=\boldsymbol{0}} \\
&= \boldsymbol{\mu^T} + \frac{1}{2}\boldsymbol{t^T}\left(\boldsymbol{2\Omega^*}\right) + \boldsymbol{\delta^T \omega} b\left(\boldsymbol{\delta^T \omega t}\right)\Big|_{\boldsymbol{t}=\boldsymbol{0}} \qquad [1] \\
&= \boldsymbol{\mu^T} + 0 + \boldsymbol{\delta^T \omega} b(0) \\
&= \boldsymbol{\mu^T} + \boldsymbol{\delta^T \omega}\sqrt{\frac{2}{\pi}}
\end{aligned}
$$

where $\boldsymbol{\delta} = \frac{\boldsymbol{\Omega\alpha}}{\sqrt{1+\boldsymbol{\alpha^T \Omega \alpha}}}$. □

The following is worth noting on the above proof:

[1] $\boldsymbol{\Omega^*}$ is a positive definite matrix.

Theorem 13. *Consider the d-dimensional random variable* $\boldsymbol{Y} \sim SN_d(\boldsymbol{\mu}, \boldsymbol{\Omega^*}, \boldsymbol{\alpha})$ *with MGF given in (20), then the variance is given by*

$$var[\boldsymbol{Y}] = \boldsymbol{\Omega^*} - \boldsymbol{\omega\delta\delta^T\omega}\left(\frac{2}{\pi}\right)$$

where $\boldsymbol{\delta} = \frac{\boldsymbol{\Omega\alpha}}{\sqrt{1+\boldsymbol{\alpha^T \Omega \alpha}}}$.

Proof. Using (109) in Appendix A.1, (118), (119) in Appendix A.3, it follows from (20) that

$$\begin{aligned}
\text{var}[\boldsymbol{Y}] &= \frac{d}{d\boldsymbol{t}d\boldsymbol{t^T}} G_{\boldsymbol{Y}}(\boldsymbol{t})\bigg|_{\boldsymbol{t=0}} \\
&= \frac{d}{d\boldsymbol{t^T}}\left(\boldsymbol{\mu^T} + \frac{1}{2}\left(\boldsymbol{\Omega^{*T}} + \boldsymbol{\Omega^*}\right)^T \boldsymbol{t} + \boldsymbol{\delta^T\omega} b\left(\boldsymbol{\delta^T\omega t}\right)\right)\bigg|_{\boldsymbol{t=0}} \\
&= \frac{d}{d\boldsymbol{t^T}}\left(\boldsymbol{t^T}\frac{1}{2}\left(\boldsymbol{\Omega^{*T}} + \boldsymbol{\Omega^*}\right)\right) + \boldsymbol{\delta^T\omega}\frac{d}{d\boldsymbol{t^T}} b\left(\boldsymbol{\delta^T\omega t}\right)\bigg|_{\boldsymbol{t=0}} \\
&= \frac{1}{2}\left(\boldsymbol{\Omega^{*T}} + \boldsymbol{\Omega^*}\right) + \boldsymbol{\omega^T\delta}\frac{d}{d\boldsymbol{t^T}} b\left(\boldsymbol{t^T}\left(\boldsymbol{\omega^T\delta}\right)^{\boldsymbol{T}}\right)\bigg|_{\boldsymbol{t=0}} \\
&= \frac{1}{2}(2\boldsymbol{\Omega^*}) + \boldsymbol{\omega^T\delta}\left[-b\left(\boldsymbol{t^T}\left(\boldsymbol{\omega^T\delta}\right)^{\boldsymbol{T}}\right)\left\{\boldsymbol{t^T}\left(\boldsymbol{\omega^T\delta}\right)^{\boldsymbol{T}} + b\left(\boldsymbol{t^T}\left(\boldsymbol{\omega^T\delta}\right)^{\boldsymbol{T}}\right)\right\}\right] \\
&\quad \times \frac{d}{d\boldsymbol{t^T}}\left(\boldsymbol{t^T}\left(\boldsymbol{\omega^T\delta}\right)^{\boldsymbol{T}}\right)\bigg|_{\boldsymbol{t=0}} \\
&= \boldsymbol{\Omega^*} + \boldsymbol{\omega^T\delta\delta^T\omega}\left[-b(0)\left\{0 + b(0)\right\}\right] \\
&= \boldsymbol{\Omega^*} + \boldsymbol{\omega^T\delta\delta^T\omega}\left[(-b(0))^2\right] \\
&= \boldsymbol{\Omega^*} + \boldsymbol{\omega^T\delta\delta^T\omega}\left(\frac{2}{\pi}\right) \\
&= \boldsymbol{\Omega^*} + \boldsymbol{\omega\delta\delta^T\omega}\left(\frac{2}{\pi}\right) && [1]
\end{aligned}$$

where $\boldsymbol{\delta} = \frac{\boldsymbol{\Omega\alpha}}{\sqrt{1+\boldsymbol{\alpha^T\Omega\alpha}}}$. □

The following is worth noting on the above proof:

[1] $\boldsymbol{\omega}$ is a diagonal matrix and so $\boldsymbol{\omega} = \boldsymbol{\omega^T}$.

The variance can also be written in another form:

$$\begin{aligned}
\text{var}[\boldsymbol{Y}] &= \boldsymbol{\Omega^*} - \boldsymbol{\omega\delta\delta^T\omega}\left(\frac{2}{\pi}\right) \\
&= \boldsymbol{\Omega^*} - \boldsymbol{\omega\delta} b(0)b(0)\boldsymbol{\delta^T\omega} \\
&= \boldsymbol{\omega\Omega\omega} - \boldsymbol{\omega\delta}[b(0)]^2\boldsymbol{\delta^T\omega} \\
&= \boldsymbol{\omega\Sigma_z\omega}
\end{aligned}$$

where $\boldsymbol{\Sigma_z} = \boldsymbol{\Omega} - \boldsymbol{\delta}\left[b(0)\right]^2\boldsymbol{\delta^T} = \text{var}[\boldsymbol{Z}]$ with $\boldsymbol{Z} \sim SN_d(\boldsymbol{\Omega}, \boldsymbol{\alpha})$ [6].

Skewness and kurtosis

The higher order derivatives of $G_{\boldsymbol{Y}}(\boldsymbol{t})$ that are used to obtain the remaining cumulants (third and fourth) are given by Azzalini and Capitanio [6]. The k^{th} order derivative of $G_{\boldsymbol{Y}}(\boldsymbol{t})$ for $k > 2$ takes the form [6]:

$$\frac{\mathrm{d}^k}{\mathrm{d}t_i\ \mathrm{d}t_j ... \mathrm{d}t_h} G_{\boldsymbol{Y}}(\boldsymbol{t}) = b^k\left(\boldsymbol{\delta}^T\boldsymbol{\omega t}\right)\omega_i\omega_j...\omega_h\delta_i\delta_j...\delta_h \qquad \text{for } h = 3, 4, ... \tag{21}$$

where the expressions of $b^k(\cdot)$ for $k \leq 4$ is given by the inverse Mills ratio Properties 1 till 4 respectively. The multivariate skewness and kurtosis measures that were introduced by Mardia [27] are obtained by the evaluation at $\boldsymbol{t} = \mathbf{0}$ of the derivatives obtained using (21). Particularly, the evaluation of (21) at $\boldsymbol{t} = \mathbf{0}$ for $k = 3$ yields the third cumulant and therefore the third central moment. Using the third central moment, Mardia's [27] multivariate skewness measure is given by

$$\gamma_{1,d}^{M} = \left(\frac{4-\pi}{2}\right)^2 \left(\boldsymbol{\mu_z^T} \boldsymbol{\Sigma_z^{-1}} \boldsymbol{\mu_z}\right)^3$$

where $\boldsymbol{\mu_z} = \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}$. In addition to the latter, the evaluation of (21) at $\boldsymbol{t} = \mathbf{0}$ for $k = 4$ yields the fourth cumulant. Using the relationship in (7), the fourth central moment is obtained. Mardia's [28] multivariate kurtosis measure is then given by

$$\gamma_{2,d}^{M} = 2\left(\pi - 3\right) \left(\boldsymbol{\mu_z^T} \boldsymbol{\Sigma_z^{-1}} \boldsymbol{\mu_z}\right)^2$$

where $\boldsymbol{\mu_z} = \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}$.

2.2.3 Generation of random numbers and illustration of PDF

It is proceeded to provide an algorithm that describes how to generate from the multivariate skew-normal distribution. In this case, the algorithm is specifically set to be implemented in R, although the idea can be extended to be used in general. The steps that are given consider the particular case of the bivariate skew-normal distribution, that is, the $SN_2(\boldsymbol{\mu}, \boldsymbol{\Omega^*}, \boldsymbol{\alpha})$ distribution.

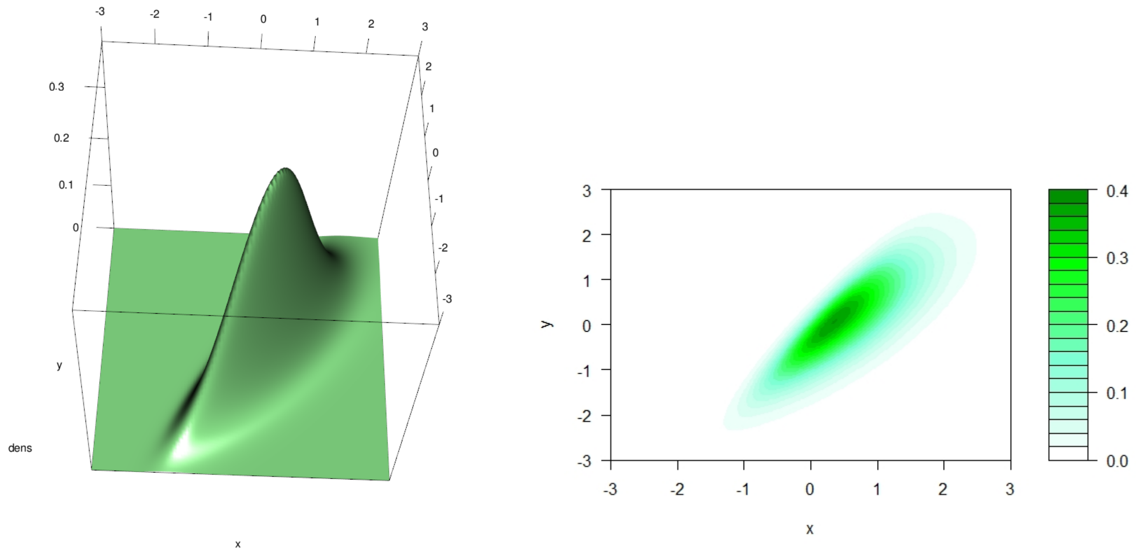
Algorithm 2 Generation from the $SN_d(\boldsymbol{\mu}, \boldsymbol{\Omega^*}, \boldsymbol{\alpha})$ distribution in R. In this case specifically where $d = 2$

1: ***Required:*** *Define the sample size to generate say N.*
2: ***Required:*** *Define two sequences. That is, define a sequence, say x, of length N that starts at -3 and ends at 3 (start and end values are arbitrary values). Then define the second sequence, say y, in a similar manner.*
3: ***Required:*** *Combine the defined sequences into a grid. In R, this translates to a 2D grid.*
4: ***Required:*** *Define the parameters of the distribution. That is*

- *define $\boldsymbol{\mu}$,*
- *define $\boldsymbol{\Omega^*}$ and*
- *define $\boldsymbol{\alpha}$.*

5: *Generate associated PDF values from the bivariate skew-normal distribution by using the rmsn-function, specifying the aforementioned parameter values.*
6: *Compress the numbers into a list object. This will yield the PDF values associated with the required sample (grid coordinates).*

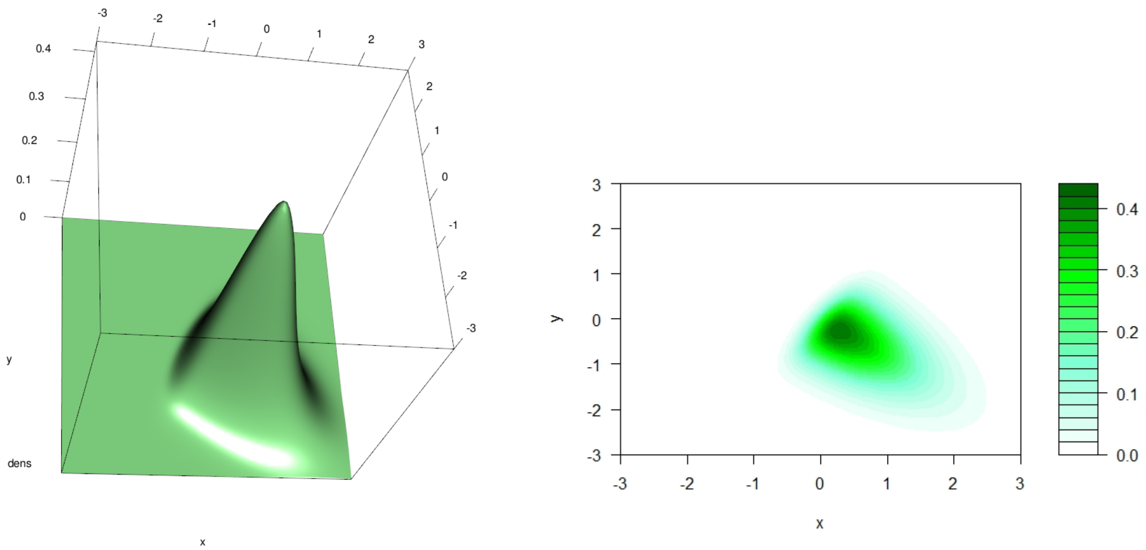
Using the PDF representation that is given in (19), PDFs of the bivariate $SN_2(\boldsymbol{\mu}, \boldsymbol{\Omega}^*, \boldsymbol{\alpha})$ distribution are provided in Figures 4 and 5 for different parameter values. These PDFs are plotted with $\boldsymbol{\mu} = (0,0)^T$ and $\boldsymbol{\alpha} = (5,-3)^T$ kept constant throughout, varying the values of Ω_{ii} for $i = 1,2$. It can be observed from Figures 4 and 5 that the multivariate skew-normal distribution exhibits a strictly unimodal PDF [7] [25].



(a) 3D plot

(b) Contour plot

Figure 4: Bivariate skew-normal PDF with $\Omega^*_{11} = \Omega^*_{22} = 1$ with $\Omega^*_{12} = \Omega^*_{21} = 0.7$.



(a) 3D plot

(b) Contour plot

Figure 5: Bivariate skew-normal PDF with $\Omega^*_{11} = \Omega^*_{22} = 1$ with $\Omega^*_{12} = \Omega^*_{21} = -0.7$.

2.2.4 Estimation

The estimation section considers the approach where the location parameters have been assumed to be different. The approach considers a direct regression environment where the location parameter $\boldsymbol{\mu_i}$ is related to a set of p explanatory variables $\boldsymbol{x_i}$ [5] [6]. That is, a regression environment where the i^{th} component $\boldsymbol{y_i} \in \mathbb{R}^d$ of $\boldsymbol{y} = (\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y_n})^T$ is sampled from $\boldsymbol{Y_i} \sim SN_d(\boldsymbol{\mu}, \boldsymbol{\Omega^*}, \boldsymbol{\alpha})$. There is independence among the $\boldsymbol{Y_i}$'s; assuming $\boldsymbol{\mu_i}$ has the following relation to $\boldsymbol{x_i}$:

$$\boldsymbol{\mu_i} = \boldsymbol{x_i^T}\boldsymbol{\beta}, \quad i = 1, 2, ..., n \tag{22}$$

for some $p \times d$ matrix $\boldsymbol{\beta}$ of unknown parameters, where the covariates vector $\boldsymbol{x_i}$ has 1 in the first position [6]. The covariate vectors $\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n}$ are then arranged in a $n \times p$ matrix $\boldsymbol{X}$ (with $n > p$) having rank p [6].

Theorem 14. *Let $\boldsymbol{y} = (\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y_n})^T$ be a random sample from the $SN_d(\boldsymbol{\mu_i}, \boldsymbol{\Omega^*}, \boldsymbol{\alpha})$ distribution for $i = 1, 2, ..., n$ with PDF in (19). Assuming the relationship in (22), the log-likelihood function is then given by*

$$l\left(\boldsymbol{X\beta}, \boldsymbol{\Omega^*}, \boldsymbol{\alpha}\right) = -\frac{1}{2}n \ln |\boldsymbol{\Omega^*}| - \frac{1}{2}n \,\mathrm{tr}\left(\boldsymbol{\Omega^{*-1}V}\right) + \sum_{i=1}^{n} b\left(\boldsymbol{\alpha}^T \boldsymbol{\omega^{-1}}\left(\boldsymbol{y} - \boldsymbol{X\beta}\right)\right) \tag{23}$$

where $b(\cdot) = \ln(2\Phi(\cdot))$ and $\boldsymbol{V} = \sum_{i=1}^{n}(\boldsymbol{y} - \boldsymbol{X\beta})(\boldsymbol{y} - \boldsymbol{X\beta})^T$.

Proof. From Bain and Engelhardt [8], using (19) and the fact that scalar$(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A})$ [21], the likelihood is derived as

$$\begin{aligned}
L\left(\boldsymbol{X\beta}, \boldsymbol{\Omega^*}, \boldsymbol{\alpha}\right) &= \prod_{i=1}^{n} f_d(\boldsymbol{y}) \\
&= \prod_{i=1}^{n} 2|\boldsymbol{\omega}|^{-1}\left(2\pi\right)^{-\frac{d}{2}} |\boldsymbol{\Omega^*}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})^{\boldsymbol{T}}\boldsymbol{\Omega^{*-1}}(\boldsymbol{y}-\boldsymbol{X\beta})} \Phi\left(\boldsymbol{\alpha^T}\boldsymbol{\omega^{-1}}\left(\boldsymbol{y} - \boldsymbol{X\beta}\right)\right) \\
&= 2|\boldsymbol{\omega}|^{-n}\left(2\pi\right)^{-\frac{nd}{2}} |\boldsymbol{\Omega^*}|^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}-\boldsymbol{X\beta})^{\boldsymbol{T}}\boldsymbol{\Omega^{*-1}}(\boldsymbol{y}-\boldsymbol{X\beta})} \sum_{i=1}^{n} \Phi\left(\boldsymbol{\alpha^T}\boldsymbol{\omega^{-1}}\left(\boldsymbol{y} - \boldsymbol{X\beta}\right)\right) \\
&= 2|\boldsymbol{\omega}|^{-n}\left(2\pi\right)^{-\frac{nd}{2}} |\boldsymbol{\Omega^*}|^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\left[(\boldsymbol{y}-\boldsymbol{X\beta})^{\boldsymbol{T}}\boldsymbol{\Omega^{*-1}}(\boldsymbol{y}-\boldsymbol{X\beta})\right]} \sum_{i=1}^{n} \Phi\left(\boldsymbol{\alpha^T}\boldsymbol{\omega^{-1}}\left(\boldsymbol{y} - \boldsymbol{X\beta}\right)\right) \\
&= 2|\boldsymbol{\omega}|^{-n}\left(2\pi\right)^{-\frac{nd}{2}} |\boldsymbol{\Omega^*}|^{-\frac{n}{2}} e^{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Omega^{*-1}}\sum_{i=1}^{n}(\boldsymbol{y}-\boldsymbol{X\beta})(\boldsymbol{y}-\boldsymbol{X\beta})^{\boldsymbol{T}}\right]} \sum_{i=1}^{n} \Phi\left(\boldsymbol{\alpha^T}\boldsymbol{\omega^{-1}}\left(\boldsymbol{y} - \boldsymbol{X\beta}\right)\right) \\
&= 2|\boldsymbol{\omega}|^{-n}\left(2\pi\right)^{-\frac{nd}{2}} |\boldsymbol{\Omega^*}|^{-\frac{n}{2}} e^{-\frac{1}{2}n\mathrm{tr}\left[\boldsymbol{\Omega^{*-1}}\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}-\boldsymbol{X\beta})(\boldsymbol{y}-\boldsymbol{X\beta})^{\boldsymbol{T}}\right]} \sum_{i=1}^{n} \Phi\left(\boldsymbol{\alpha^T}\boldsymbol{\omega^{-1}}\left(\boldsymbol{y} - \boldsymbol{X\beta}\right)\right)
\end{aligned}$$

and hence the log-likelihood is then obtained as

$$
\begin{aligned}
l\left(\boldsymbol{X\beta},\boldsymbol{\Omega^*},\boldsymbol{\alpha}\right) &= \ln\left[L\left(\boldsymbol{X\beta},\boldsymbol{\Omega^*},\boldsymbol{\alpha}\right)\right] \\
&= -n\ln|\boldsymbol{\omega}| - \frac{1}{2}nd\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Omega^*}| - \frac{1}{2}n\mathrm{tr}\left[\boldsymbol{\Omega^{*-1}}\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)^{\boldsymbol{T}}\right] \\
&+ \sum_{i=1}^{n}\ln\left(2\Phi\left(\boldsymbol{\alpha^T}\boldsymbol{\omega^{-1}}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)\right)\right) \\
&\propto -\frac{1}{2}\ln|\boldsymbol{\Omega^*}| - \frac{1}{2}n\mathrm{tr}\left(\boldsymbol{\Omega^{*-1}V}\right) + \sum_{i=1}^{n}b\left(\boldsymbol{\alpha}^T\boldsymbol{\omega^{-1}}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)\right)
\end{aligned}
$$

where $b(\cdot) = \ln(2\Phi(\cdot))$ and $\boldsymbol{V} = \sum_{i=1}^{n}(\boldsymbol{y}-\boldsymbol{X\beta})(\boldsymbol{y}-\boldsymbol{X\beta})^T$. □

Maximization of (23) should be conducted numerically over a parameter space with dimension size of $pd + \frac{d(d+3)}{2}$. It is proceeded to follow a provisional reparametrization that will ease the maximization of the log-likelihood [5].

Let $\boldsymbol{\eta} = \boldsymbol{\omega^{-1}\alpha}$ as a parameter replacing $\boldsymbol{\alpha}$ in the final term of (23). Now (23) without the final summation is the same as the Gaussian log-likelihood [6] [21] . Therefore, the maximization of (23) with respect to $\boldsymbol{\Omega^*}$ is equivalent to maximizing the Gaussian log-likelihood for fixed $\boldsymbol{\beta}$ [5]. The well-known solution is given as

$$
\begin{aligned}
\hat{\boldsymbol{\Omega}}^*\left(\boldsymbol{\beta}\right) &= V\left(\boldsymbol{\beta}\right) \\
&= n^{-1}\boldsymbol{u^T u}
\end{aligned}
\tag{24}
$$

where $\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{X\beta}$ [5]. Substituting (24) into (23) yields the profile log-likelihood as

$$
l^*\left(\boldsymbol{\beta},\boldsymbol{\eta}\right) = -\frac{1}{2}\ln|V(\boldsymbol{\beta})| - \frac{1}{2}nd + \boldsymbol{1_n^T}b(\boldsymbol{u\eta}) \tag{25}
$$

which will also be maximized numerically, however with lower dimensions. The parameter space has been reduced to $d(p+1)$ parameter components [6]. Numerical maximization of (25) is accelerated by supplying partial derivatives to the quasi-Newton algorithm [5]. The quasi-Newton algorithm is fully explained by Dennis and More [12]. The partial derivatives of (25) are given as [5]

$$
\begin{aligned}
\frac{\partial l^*\left(\boldsymbol{\beta},\boldsymbol{\eta}\right)}{\partial\boldsymbol{\beta}} &= \boldsymbol{X^T u}V(\boldsymbol{\beta})^{-1} - \boldsymbol{X^T}b'(\boldsymbol{u\eta})\boldsymbol{\eta^T} \\
\frac{\partial l^*\left(\boldsymbol{\beta},\boldsymbol{\eta}\right)}{\partial\boldsymbol{\eta}} &= \boldsymbol{u^T}b'(\boldsymbol{u\eta}).
\end{aligned}
$$

Once the value for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\eta}}$ are obtained using the quasi-Newton algorithm, the MLE of $\boldsymbol{\Omega^*}$ is obtained as

$\hat{\boldsymbol{\Omega}}^{*} = n^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)^{T}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)$. From here, it is proceeded to obtain $\hat{\boldsymbol{\omega}}$ following convenient notation, with the MLE of $\boldsymbol{\alpha}$ then obtained as $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\omega}}\hat{\boldsymbol{\eta}}$ using the equivariance property of MLEs [5].

3 Chapter 3

Chapter 3 will discuss the geometric skew-normal distribution. In Section 3.1, the univariate geometric skew-normal distribution is revisited. An overview of the multivariate geometric skew-normal distribution is presented in Section 3.2.

3.1 Univariate geometric skew-normal distribution

This subsection will proceed to revisit the univariate geometric skew-normal distribution. The geometric skew-normal distribution was presented by Kundu [24]. This distribution is conveniently more flexible than the skew-normal distribution as its PDF can be multimodal and the MLEs can be obtained in explicit forms [24]. Figure 6 gives a summary of how Section 3.1 will proceed.

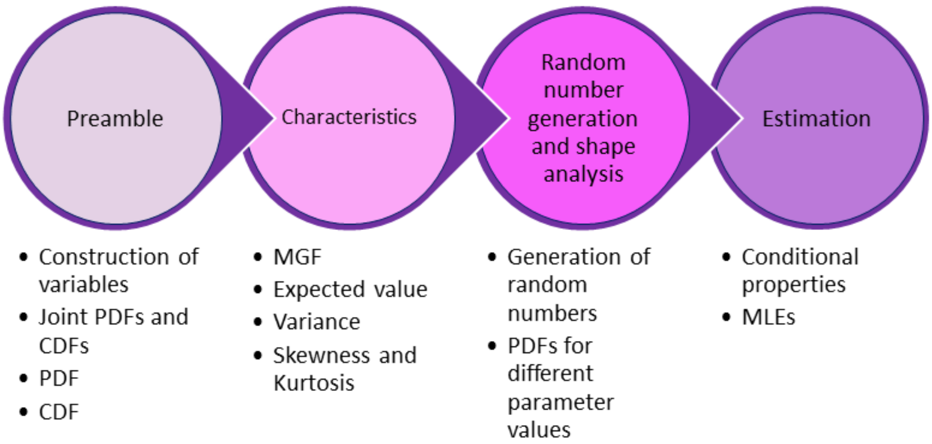


Figure 6: Summary of Section 3.1.

3.1.1 Preamble

Throughout this section a normal random variable with mean μ and variance σ^2 will be denoted by $N\left(\mu, \sigma^2\right)$. A geometric random variable with parameter p will be denoted by $GE\left(p\right)$. The functions $\phi(\cdot)$ and $\Phi(\cdot)$ represent the standard normal distribution PDF and CDF respectively.

Theorem 15. *A random variable Y that follows the geometric distribution with parameter p, that is $Y \sim GE\left(p\right)$, has a probability mass function (PMF) of the following form [8]:*

$$f_Y\left(y\right) = p\left(1-p\right)^{y-1} \tag{26}$$

for $y = 1, 2, ...$ and $0 < p \leq 1$.

Theorem 16. *Let $\{X_i : i = 1, 2, ...\}$ be independent and identically distributed (i.i.d.) $N\left(\mu, \sigma^2\right)$ random variables and suppose that $N \sim GE(p)$, with N and $X_i's$ independently distributed. Then define*

$$X \stackrel{d}{=} \sum_{i=1}^{N} X_i \tag{27}$$

where $\stackrel{d}{=}$ indicates equal in distribution. It is then observed that X is a geometric skew-normal random variable with parameters μ, σ and p [24]. This will be denoted by $GSN(\mu, \sigma, p)$.

Theorem 17. *Let $X \sim GSN(\mu, \sigma, p)$ given in (27) where $N \sim GE(p)$. The joint PDF of the variable (X, N) is given by*

$$f_{X,N}(x,n) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi n}} e^{-\frac{1}{2n\sigma^2}(x-n\mu)^2} p(1-p)^{n-1} & \text{if } 0 < p < 1 \\ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} & \text{if } p = 1 \end{cases} \tag{28}$$

using the convention that $0^0 = 1$ when $p = 1$ and noting that $-\infty < x < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$ and n is any positive integer [24].

Proof. Let $N \sim GE(p)$. The proof will be given for the case where $0 < p < 1$. If $p = 1$, then $P(N = n) = 1$ since it is assumed that $0^0 = 1$. The result for $p = 1$ is then immediate using the $N(\mu, \sigma^2)$ distribution PDF that is stated and proved in Bain and Engelhardt [8].

Recall that if $X \sim N(\mu, \sigma^2)$, then $X|(N = n) = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$ [8]. Hence, from the latter, Bain and Engelhardt [8] and (26), it follows that

$$\begin{aligned} f_{X,N}(x,n) &= f_{X|N}(x|n) \times f_N(n) \\ &= \frac{1}{\sigma\sqrt{2\pi n}} e^{-\frac{1}{2n\sigma^2}(x-n\mu)^2} p(1-p)^{n-1} \end{aligned}$$

noting that $-\infty < x < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$ and n is any positive integer. □

Theorem 18. *Let $X \sim GSN(\mu, \sigma, p)$ given in (27) where $N \sim GE(p)$. The joint CDF of the variable (X, N) is given by*

$$F_{X,N}(x,n) = p \sum_{k=1}^{n} \Phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1} \tag{29}$$

for $-\infty < x < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$, $0 < p < 1$ and n is any positive integer [24].

Proof. Making use of conditional probability (112) in Appendix A.1 and the fact that $X|(N = k) \sim$

$N(k\mu, k\sigma^2)$ [8], it follows that

$$\begin{aligned}
F_{X,N}(x,n) &= P\left(X \leq x, N \leq n\right) \\
&= P\left(X \leq x, N = 1\right) + P\left(X \leq x, N = 2\right) + \\
&\quad \cdots + P\left(X \leq x, N = n\right) \\
&= \sum_{k=1}^{n} P\left(X \leq x, N = k\right) \\
&= \sum_{k=1}^{n} P\left(X \leq x | N = k\right) P\left(N = k\right) \\
&= p \sum_{k=1}^{n} \Phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1}.
\end{aligned}$$

□

Naturally, the question arises of what would happen to the CDF if $p = 1$? If $p = 1$, then $P(N = 1) = 1$ and $F_{X,N}(x,n) = \Phi\left(\frac{x-\mu}{\sigma}\right)$.

Theorem 19. *Let $X \sim GSN\left(\mu, \sigma, p\right)$ given in (27) where $N \sim GE(p)$. The CDF of the random variable X is given by*

$$F_X(x) = p \sum_{k=1}^{\infty} \Phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1} \tag{30}$$

for $-\infty < x < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$ and $0 < p < 1$ [24].

Proof. From Stewart [37], it follows that

$$\begin{aligned}
F_X(x) &= \lim_{n\to\infty} F_{X,N}(x,n) \\
&= \lim_{n\to\infty} p \sum_{k=1}^{n} \Phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1} \\
&= p \sum_{k=1}^{\infty} \Phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1}.
\end{aligned}$$

□

As, before the question arises of what would happen to the CDF if $p = 1$? If $p = 1$, then $P(N = 1) = 1$ and $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$.

Theorem 20. *Let $X \sim GSN\left(\mu, \sigma, p\right)$ given in (27) where $N \sim GE(p)$. The PDF of the random variable*

X is given by

$$f_X(x) = \sum_{k=1}^{\infty} \frac{p}{\sigma\sqrt{k}} \phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1} \tag{31}$$

for $-\infty < x < \infty$*,* $\mu \in \mathbb{R}$*,* $\sigma > 0$ *and* $0 < p < 1$ *[24].*

Proof. Using (116) in Appendix A.3, it follows that

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \frac{d}{dx} p \sum_{k=1}^{\infty} \Phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1} \\ &= \sum_{k=1}^{\infty} \frac{p}{\sigma\sqrt{k}} \phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1}. \end{aligned}$$

□

If $p = 1$, then $X \sim N(\mu, \sigma^2)$ [24]. When $\mu = 0$ and $\sigma = 1$, then the distribution of X is known as the standard geometric skew-normal distribution denoted by $GSN(0, 1, p)$, with the pdf in (31) becoming

$$f_X(x) = p \sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \phi\left(\frac{x}{\sqrt{k}}\right) (1-p)^{k-1}. \tag{32}$$

The $GSN(0, 1, p)$ distribution is a symmetric distribution around 0 for all values of p [24].

3.1.2 Characteristics of the geometric skew-normal distribution

The computation of the characteristics of $X \sim GSN(\mu, \sigma, p)$ is done via the MGF or, equivalently but somewhat more practical, via the CGF (see (109) in Appendix A.1).

Moment generating function

The MGF of a random variable with PDF as given in (31) will be derived. Before deriving the MGF of the geometric skew-normal distribution it is worthwhile deriving the MGF of the geometric distribution first.

Theorem 21. *The MGF of the random variable* $Y \sim GE(p)$ *with PMF given in (26) is given by*

$$M_Y(t) = \frac{pe^t}{1 - qe^t} \tag{33}$$

with $t \in \mathbb{R}$ *and* $q = 1 - p$*.*

Proof. Using (108) in Appendix A.1 and the sum of geometric series, it follows that

$$\begin{aligned}M_Y(t) &= E\left[e^{tY}\right]\\ &= \sum_{y=1}^{\infty} e^{ty} p q^{y-1}\\ &= \frac{p}{q}\sum_{y=1}^{\infty}\left(qe^t\right)^y\\ &= pe^t\sum_{y=0}^{\infty}\left(qe^t\right)^y\\ &= \frac{pe^t}{1-qe^t}.\end{aligned}$$

□

Theorem 22. *The MGF of a random variable $X \sim GSN(\mu, \sigma, p)$ with PDF given in (31) is given by*

$$M_X(t) = \frac{pe^{\mu t+\frac{1}{2}\sigma^2 t^2}}{1-(1-p)e^{\mu t+\frac{1}{2}\sigma^2 t^2}} \tag{34}$$

where $t \in \mathbb{R}$ [24].

Proof. From Bain and Engelhardt [8]

$$\begin{aligned}M_X(t) &= E\left[e^{tX}\right]\\ &= E_N\left[E\left(e^{tX}|N\right)\right]\end{aligned} \tag{35}$$

where $N \sim GE(p)$. Since the $X_i's$ are i.i.d. $N(\mu, \sigma^2)$ random variables, it follows that

$$\begin{aligned}E\left(e^{tX}|N=n\right) &= E\left[e^{t\sum_{i=1}^{n} X_i}\right]\\ &= \prod_{i=1}^{n} E\left[e^{tX_i}\right]\\ &= \prod_{i=1}^{n} M_{X_i}(t)\\ &= \left[M_Y(t)\right]^n\\ &= \left[e^{\mu t+\frac{1}{2}\sigma^2 t^2}\right]^n.\end{aligned} \tag{36}$$

Substituting (36) into (35), using the fact that if $Y \sim N(\mu, \sigma^2)$, then $\ln(M_Y(t)) = \mu t + \frac{1}{2}\sigma^2 t^2$ and using

(33), it follows that

$$\begin{aligned} M_X(t) &= E_N\left[e^{N\mu t+\frac{1}{2}N\sigma^2t^2}\right] \\ &= E_N\left[e^{N\ln(M_Y(t))}\right] \\ &= M_N\left(\ln(M_Y(t))\right) \\ &= \frac{pe^{\mu t+\frac{1}{2}\sigma^2t^2}}{1-(1-p)e^{\mu t+\frac{1}{2}\sigma^2t^2}} \end{aligned}$$

where $t \in \mathbb{R}$. □

Expected value, variance, skewness and kurtosis

This section will proceed by discussing the expected value and variance of the geometric skew-normal distribution. The third- and fourth central moments are stated without proof, however a result is provided that can be utilized to obtain them in a convenient matter. Consequently, the skewness and kurtosis are also provided. The method that will be used for derivations in this section follows in a similar fashion to that of what is described by Azzalini and Captanio [6]. The method derives the expected value and variance by using the cumulant generating function, $G_Y(t)$, as stated in (109) (see Appendix A.1).

Expected value

Theorem 23. *Consider $X \sim GSN(\mu, \sigma, p)$ with MGF as given in (34), then the expected value is given in Kundu [24] by*

$$E[X] = \frac{\mu}{p}.$$

Proof. Using (109) and (110) in Appendix A.1, it follows that

$$\begin{aligned}
E[X] &= \frac{d}{dt} G_X(t)\bigg|_{t=0} \\
&= \frac{d}{dt} \ln M_X(t)\bigg|_{t=0} \\
&= \frac{d}{dt} \ln \left[\frac{pe^{\mu t+\frac{1}{2}\sigma^2 t^2}}{1-qe^{\mu t+\frac{1}{2}\sigma^2 t^2}} \right]\Bigg|_{t=0} \\
&= \frac{d}{dt} \left[\ln(pe^{\mu t+\frac{1}{2}\sigma^2 t^2}) - \ln\left(1-qe^{\mu t+\frac{1}{2}\sigma^2 t^2}\right) \right]\bigg|_{t=0} \\
&= \frac{d}{dt} \ln\left(pe^{\mu t+\frac{1}{2}\sigma^2 t^2}\right) - \frac{d}{dt} \ln\left(1-qe^{\mu t+\frac{1}{2}\sigma^2 t^2}\right)\bigg|_{t=0} \\
&= \frac{pe^{\mu t+\frac{1}{2}\sigma^2 t^2}\left(\mu+\sigma^2 t\right)}{pe^{\mu t+\frac{1}{2}\sigma^2 t^2}} - \frac{-qe^{\mu t+\frac{1}{2}\sigma^2 t^2}\left(\mu+\sigma^2 t\right)}{1-qe^{\mu t+\frac{1}{2}\sigma^2 t^2}}\Bigg|_{t=0} \\
&= \mu + \frac{q(\mu)}{1-q} \\
&= \frac{\mu}{p}.
\end{aligned}$$

□

Variance

Theorem 24. *Consider $X \sim GSN(\mu, \sigma, p)$ with MGF as given in (34), then the variance is given in Kundu [24] by*

$$var[X] = \frac{\sigma^2 p + \mu^2(1-p)}{p^2}.$$

Proof. Using (109) and (110) in Appendix A.1, it follows that

$$
\begin{aligned}
\text{var}[X] &= \frac{d^2}{dt^2} G_X(t)\bigg|_{t=0} \\
&= \frac{d}{dt}\left(\frac{d}{dt}\ln M_X(t)\right)\bigg|_{t=0} \\
&= \frac{d}{dt}\left(\mu+\sigma^2 t+\frac{qe^{\mu t+\frac{1}{2}\sigma^2t^2}(\mu+\sigma^2 t)}{1-qe^{\mu t+\frac{1}{2}\sigma^2t^2}}\right)\Bigg|_{t=0} \\
&= \frac{d}{dt}\left([\mu+\sigma^2 t]\left[1+\frac{qe^{\mu t+\frac{1}{2}\sigma^2t^2}}{1-qe^{\mu t+\frac{1}{2}\sigma^2t^2}}\right]\right)\Bigg|_{t=0} \\
&= \sigma^2\left[1+\frac{qe^{\mu t+\frac{1}{2}\sigma^2t^2}}{1-qe^{\mu t+\frac{1}{2}\sigma^2t^2}}\right]+[\mu+\sigma^2 t]\left[qe^{\mu t+\frac{1}{2}\sigma^2t^2}(\mu+\sigma^2 t)\left(1-qe^{\mu t+\frac{1}{2}\sigma^2t^2}\right)\right] \\
&\quad\times\left[\frac{-qe^{\mu t+\frac{1}{2}\sigma^2t^2}\left(-qe^{\mu t+\frac{1}{2}\sigma^2t^2}\right)(\mu+\sigma^2 t)}{\left[1-qe^{\mu t+\frac{1}{2}\sigma^2t^2}\right]^2}\right]\Bigg|_{t=0} \qquad [1] \\
&= \sigma^2\left[1+\frac{q}{1-q}\right]+[\mu]\left[\frac{q(\mu)(1-q)-q(-q)(\mu)}{[1-q]^2}\right] \\
&= \sigma^2\left[\frac{p+q}{p}\right]+\mu\left[\frac{qp\mu+q^2\mu}{p^2}\right] \\
&= \frac{\sigma^2p^2+\sigma^2pq+\mu^2pq+\mu^2q^2}{p^2} \\
&= \frac{\sigma^2p\,(p+q)+\mu^2q\,(p+q)}{p^2} \\
&= \frac{[p+(1-p)]\left[\sigma^2p+\mu^2(1-p)\right]}{p^2} \\
&= \frac{\sigma^2p+\mu^2(1-p)}{p^2}.
\end{aligned}
$$

□

The following is worth noting on the above proof:

[1] labelled as (A^{**}) for later use.

Skewness

Standardisation of the third cumulant produces the well-known measure of Fisher-Pearson moment skewness [31]. The Fisher-Pearson moment skewness, γ_1, is given by

$$\gamma_1 = \frac{E\left[(X-E[X])^3\right]}{(\text{var}[X])^{\frac{3}{2}}}.$$

Theorem 25. *Consider $X \sim GSN\,(\mu,\sigma,p)$ with MGF as given in (34), then the third central moment* *is given in Kundu [24] by*

$$E\left[(X-E(X))^3\right]=\frac{1-p}{p^3}\left(\mu^3(2p^2-p+2)+2\mu^2p^2+\mu\sigma^2(3-p)p\right). \tag{37}$$

The third central moment (which in turn is the third cumulant) can be obtained in a similar fashion to that of the variance (second cumulant) by obtaining the derivative of (A^{**}). The third central moment as given in (37) can otherwise be directly obtained using (28) in terms of an infinite series as described in Kundu [24]. The infinite series is used to obtain higher order moments and it is given as

$$E\left(X^m\right)=p\sum_{n=1}^{\infty}(1-p)^{n-1}u_m\left(n\mu,n\sigma^2\right) \tag{38}$$

where $u_m(n\mu,n\sigma^2)=E(Y^m)$ with $Y\sim N(n\mu,n\sigma^2)$ [24]. Utilizing the third central moment, the Fisher-Pearson moment skewness is then given in Kundu [24] as

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

Kurtosis

Corollary 4. *Consider $X\sim GSN\left(\mu,\sigma,p\right)$ with MGF as given in (34), then the fourth central moment is given in Kundu [24] by*

$$E\left[(X-E(X))^4\right]=\mu^4(1-p)\left(p^2-6p+6\right)-2\mu^2\sigma^2p(1-p)\left(p^2+3p-6\right)+3\sigma^4p^2. \tag{39}$$

As before, the fourth central moment given in (39) can be directly obtained using the infinite series given in (38). Utilization of the fourth central moment produces the well-known measure of kurtosis, γ_2 [36]. The kurtosis is given in Kundu [24] by

$$\begin{aligned}\gamma_2&=\frac{E\left[(X-E[X])^4\right]}{(\mathrm{var}[X])^2}\\&=\frac{\mu^4(1-p)\left(p^2-6p+6\right)-2\mu^2\sigma^2p(1-p)\left(p^2+3p-6\right)+3\sigma^4p^2}{\left(\sigma^2p+\mu^2(1-p)\right)^2}.\end{aligned}$$

3.1.3 Generation of random numbers and illustration of PDF

In order to generate from the $GSN\left(\mu,\sigma,p\right)$ distribution with PDF given in (31) a short algorithm is provided by Kundu [24]. In the algorithm the distributions as given in (27) are used as well as the fact that $X|(N=k)\sim N(k\mu,k\sigma^2)$.

Algorithm 3 Generation from the $GSN(\mu, \sigma, p)$ distribution

1: ***Required:***
- *Define the value of* μ *for* $\mu \in \mathbb{R}$.
- *Define the value of* σ *for* $\sigma > 0$.
- *Define the value of* p *for* $0 < p \leq 1$.

2: *Generate the value* k *from the* $GE(p)$ *distribution.*
3: *Generate the value* x *from* $N(k\mu, k\sigma^2)$, *where* k *is from step 2.*
4: *To obtain a required sample of size* n *repeat step 3,* n *times using the same* k *from step 2.*

Using the PDF given in (31), graphs of the PDFs of the $GSN(\mu, \sigma, p)$ distribution are provided in Figure 7. These PDFs are plotted with $p = 0.5$ and $\sigma = 1$ kept constant throughout, varying the values of μ.

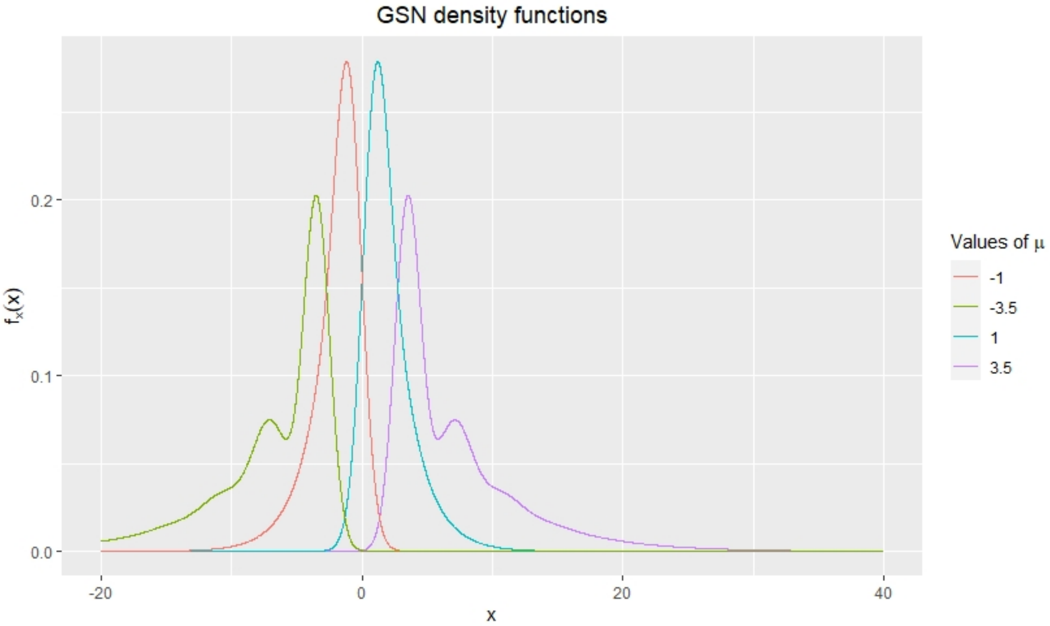


Figure 7: PDFs of the $GSN(\mu, \sigma, p)$ distribution for different values of μ.

It can be seen from Figure 7 that the PDFs of the $GSN(\mu, \sigma, p)$ distribution can take on various shapes depending on the parameter values for μ. The PDFs are positively skewed when $\mu > 0$ and negatively skewed when $\mu < 0$. It is also observed that the PDFs can be unimodal or multimodal [24]. The green PDF ($\mu = -3.5$) and the purple PDF ($\mu = 3.5$) in Figure 7 depict multimodal PDFs, whereas the remaining densities depict unimodal PDFs. The former observation of a multimodal PDF is different from the skew-normal distribution which is always unimodal. It appears as if the $GSN(\mu, \sigma, p)$ distribution is more flexible than the skew-normal distribution [24].

The characteristics that include the expected value, variance, skewness and kurtosis are given in Table 4 for different values of μ, with $p = 0.5$ and $\sigma = 1$ kept constant throughout.

Table 4: Values of some characteristics of the geometric skew-normal distribution for different values of μ

μ	$E[X]$	$var[X]$	γ_1	γ_2
-4	-7	26.5	-2.463	6.166
-2	-2	4	-1.375	4.5
2	2	4	1.875	4.5
4	7	26.5	2.822	6.166

It can be observed from Table 4 that the expected value increases as the value of μ increases. The same observation is noted for the skewness. It is also observed that as the $|\mu|$ decreases, the variance and the kurtosis decreases.

The characteristics that include the expected value, variance, skewness and kurtosis are given in Table 5 for different values of p, with $\mu = 2$ and $\sigma = 1$ kept constant throughout.

Table 5: Values of some characteristics of the geometric skew-normal distribution for different values of p

p	$E[X]$	$var[X]$	γ_1	γ_2
0.1	20	370	2.026	5.992
0.3	6.667	34.444	2.229	5.904
0.7	2.857	3.878	2.907	5.039
0.9	2.222	1.605	2.160	3.834

It can be observed from Table 5 that the expected value decreases as the value of p increases. The same observation is noted for the variance and the kurtosis. It is also observed that as the p increases, the skewness increases.

The characteristics that include the expected value, variance, skewness and kurtosis are given in Table 6 for different values of σ, with $\mu = 2$ and $p = 0.5$ kept constant throughout.

Table 6: Values of some characteristics of the geometric skew-normal distribution for different values of σ

σ	$E[X]$	$var[X]$	γ_1	γ_2
0.1	4	8.02	3.174	6.489
0.5	4	8.50	3.006	6.239
0.9	4	9.62	2.685	5.771
1.1	4	10.42	2.500	5.509

It can be observed from Table 6 that the expected value remains unchanged as the value of σ changes. The latter is as a result of the fact that the expected value is not a function of σ. It is observed that the variance increases as the value of σ increases. In contrast, both the skewness and the kurtosis decreases as the value of σ increases.

3.1.4 Estimation

Conditional properties

Before continuing with the estimation section, it is worthwhile to first furnish different conditional properties as given in Kundu [24]. These conditional properties provide further detail into the distribution and contribute knowledge on specific conditions enclosed on the distribution. Some, but not all of the properties will be utilized in the derivation of the estimation theory.

Theorem 26. *Consider (X, N) which has the joint PDF as given by (28), and let $m \leq n$ be positive integers. The conditional CDF of (X, N) given $N \leq n$ is given by*

$$P\left(X \leq x, N \leq m | N \leq n\right) = \frac{p}{1-(1-p)^n} \sum_{k=1}^{m} \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1} \tag{40}$$

for $-\infty < x < \infty$ and $0 < p < 1$ [24].

Proof. From (112), (29) and the CDF of a $GE(p)$ distribution, it follows that

$$\begin{aligned}
P\left(X \leq x, N \leq m | N \leq n\right) &= \frac{P\left(X \leq x, N \leq m\right)}{P\left(N \leq n\right)} \\
&= \frac{p \sum_{k=1}^{m} \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right)(1-p)^{k-1}}{P\left(N \leq n\right)} \\
&= \frac{p \sum_{k=1}^{m} \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right)(1-p)^{k-1}}{1-(1-p)^n} \\
&= \frac{p}{1-(1-p)^n} \sum_{k=1}^{m} \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1}.
\end{aligned}$$

□

It follows directly from (40), that for $m = n$ and using (29), it is true that

$$\begin{aligned}
P\left(X \leq x, N \leq n | N \leq n\right) &= P\left(X \leq x | N \leq n\right) \\
&= \frac{P\left(X \leq x, N \leq n\right)}{P\left(N \leq n\right)} \\
&= \frac{p}{1-(1-p)^n} \sum_{k=1}^{n} \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1}.
\end{aligned} \tag{41}$$

Theorem 27. *Consider (X, N) which has the joint PDF as given by (28). Suppose that $0 \leq x \leq y$ and*

n is a positive integer. The conditional CDF of (X, N) given $X \leq y$ is given in Kundu [24] by

$$P(X \leq x, N \leq n | X \leq y) = \frac{\sum_{k=1}^{n} (1-p)^k \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right)}{\sum_{k=1}^{\infty} (1-p)^k \Phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)}. \tag{42}$$

Proof. Using (112) in Appendix A.1, (29) and since $x \leq y$, it follows that

$$\begin{aligned}
P(X \leq x, N \leq n | X \leq y) &= \frac{P(X \leq x, N \leq n)}{P(X \leq y)} \\
&= \frac{p \sum_{k=1}^{n} \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right)(1-p)^{k-1}}{p \sum_{k=1}^{\infty} \Phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)(1-p)^{k-1}} \\
&= \frac{\sum_{k=1}^{n} (1-p)^k \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right)}{\sum_{k=1}^{\infty} (1-p)^k \Phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)}.
\end{aligned}$$

□

It follows directly from (42) that

$$\begin{aligned}
P(N \leq n, X \leq y | X \leq y) &= P(N \leq n | X \leq y) \\
&= \frac{P(N \leq n, X \leq y)}{P(X \leq y)} \\
&= \frac{\sum_{k=1}^{n} (1-p)^k \Phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)}{\sum_{k=1}^{\infty} (1-p)^k \Phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)}
\end{aligned} \tag{43}$$

for n as any positive integer [24]. From Kundu [24], using (28) and (31), the conditional PMF of N given $X = x$ is

$$\begin{aligned}
f_{N|X}(n|x) &= \frac{f_{N,X}(n,x)}{f_X(x)} \\
&= \frac{\frac{1}{\sigma\sqrt{n}} \phi\left(\frac{x-n\mu}{\sigma\sqrt{n}}\right) p(1-p)^{n-1}}{\sum_{k=1}^{\infty} \frac{1}{\sigma\sqrt{k}} \phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) p(1-p)^{k-1}} \\
&= \frac{(1-p)^{n-1} e^{-\frac{1}{2\sigma^2 n}(x-n\mu)^2} / \sqrt{n}}{\sum_{k=1}^{\infty} (1-p)^{k-1} e^{-\frac{1}{2\sigma^2 k}(x-k\mu)^2} / \sqrt{k}}
\end{aligned} \tag{44}$$

and in turn then the conditional expectation of N given $X = x$ becomes

$$\begin{aligned}
E(N \mid X = x) &= \sum_{n=1}^{\infty} n f_{N|X}(n|x) \\
&= \frac{\sum_{n=1}^{\infty} (1-p)^{n-1} e^{-\frac{1}{2\sigma^2 n}(x-n\mu)^2} \sqrt{n}}{\sum_{k=1}^{\infty} (1-p)^{k-1} e^{-\frac{1}{2\sigma^2 k}(x-k\mu)^2} / \sqrt{k}}
\end{aligned} \tag{45}$$

with

$$
\begin{aligned}
E\left(N^{-1} \mid X = x\right) &= \sum_{n=1}^{\infty} n^{-1} f_{N|X}(n|x) \\
&= \frac{\sum_{n=1}^{\infty}(1-p)^{n-1} e^{-\frac{1}{2\sigma^2 n}(x-n\mu)^2} / n^{3/2}}{\sum_{k=1}^{\infty}(1-p)^{k-1} e^{-\frac{1}{2\sigma^2 k}(x-k\mu)^2} / \sqrt{k}}. \qquad (46)
\end{aligned}
$$

Maximum likelihood estimators

Theorem 28. *Suppose that $\{x_1, x_2, ..., x_n\}$ is a sample of size n from the GSN(μ, σ, p) distribution with PDF as given in (31). The log-likelihood function is then given in Kundu [24] by*

$$
l(\mu, \sigma, p) = \sum_{i=1}^{n} \ln \left[\sum_{k=1}^{\infty} \frac{p}{\sigma\sqrt{k}} \phi\left(\frac{x_i - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1} \right]. \qquad (47)
$$

Proof. From Bain and Engelhardt [8] and using (31), it follows that

$$
\begin{aligned}
l(\mu, \sigma, p) &= \ln\left(\prod_{i=1}^{n} f_{X_i}(x_i)\right) \\
&= \sum_{i=1}^{n} ln\left[f_{X_i}(x_i)\right] \\
&= \sum_{i=1}^{n} \ln \left[\sum_{k=1}^{\infty} \frac{p}{\sigma\sqrt{k}} \phi\left(\frac{x_i - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1} \right]
\end{aligned}
$$

□

It is proceeded to obtain the normal equations by taking the partial derivatives of the log-likelihood function given in (47) and equating them to 0. The partial derivatives are with respect to μ, σ, p. However, clearly the MLEs cannot be obtained in explicit form as it is required to solve three non-linear equations simultaneously [24] [25]. In order to circumvent this problem, Kundu [24] proposes the use of the EM-algorithm to compute the MLEs.

Theorem 29. *Let $\{(x_1, m_1), (x_2, m_2), ..., (x_n, m_n)\}$ be a random sample of size n from the joint distribution of (X, N). The log-likelihood function based on the complete sample is given in Kundu [24] by*

$$
l_c(\mu, \sigma, p) \propto -n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{(x_i - m_i\mu)^2}{m_i} + n \ln p + \ln(1-p) \sum_{i=1}^{n} (m_i - 1) \qquad (48)
$$

without the additive constant and the MLEs of the unknown parameters are obtained by Kundu [24] as

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{\sum_{k=1}^{n} m_k}, \quad \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - m_i\hat{\mu})^2}{m_i}, \quad \hat{p} = \frac{n}{\sum_{i=1}^{n} m_i}. \tag{49}$$

Proof. Using (28), it follows that

$$\begin{aligned}
l_c(\mu,\sigma,p) &= \ln\left(\prod_{i=1}^{n} f_{X,N}(x_i,m_i)\right) \\
&= \sum_{i=1}^{n} \ln\left[f_{X,N}(x_i,m_i)\right] \\
&= \sum_{i=1}^{n} \ln\left[\frac{p}{\sigma\sqrt{2\pi m_i}} e^{\frac{-1}{2}\left(\frac{x_i - m_i\mu}{\sigma\sqrt{m_i}}\right)^2}(1-p)^{m_i - 1}\right] \\
&\propto \sum_{i=1}^{n}\left[\frac{-1}{2m_i\sigma^2}(x_i - m_i\mu)^2 + (m_i - 1)\ln(1-p) + \ln\left(\frac{p}{\sigma\sqrt{2\pi}}\right)\right] \\
&= \frac{-1}{2\sigma^2}\sum_{i=1}^{n}\frac{(x_i - m_i\mu)^2}{m_i} + \ln(1-p)\sum_{i=1}^{n}(m_i - 1) + n\ln(p) - n\ln\left(\sigma\sqrt{2\pi}\right) \\
&\propto -n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{(x_i - m_i\mu)^2}{m_i} + n\ln p + \ln(1-p)\sum_{i=1}^{n}(m_i - 1)\,.
\end{aligned}$$

It is then proceeded to obtain the normal equations based on the complete sample. This is done by taking the partial derivatives of the complete log-likelihood and setting them equal to 0, followed by solving for the unknown parameters. Thus, the MLE for μ is obtained as

$$\begin{aligned}
\frac{\partial l_c(\mu,\sigma,p)}{\partial\mu} &= 0 \\
\frac{-1}{\sigma^2}\left(\sum_{i=1}^{n}\frac{(x_i - m_i\hat{\mu})}{m_i}\right)(-m_i) &= 0 \\
\frac{-1}{\sigma^2}\sum_{i=1}^{n}(x_i - m_i\hat{\mu}) &= 0 \\
\sum_{i=1}^{n}(x_i - m_i\hat{\mu}) &= 0 \\
\hat{\mu} &= \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} m_i}
\end{aligned}$$

with the MLE for σ^2 obtained as

$$\begin{aligned}
\frac{\partial l_c(\mu, \sigma, p)}{\partial \sigma} &= 0 \\
\frac{-n}{\hat{\sigma}} + \frac{2}{2\hat{\sigma}^3} \sum_{i=1}^{n} \frac{(x_i - m_i \mu)^2}{m_i} &= 0 \\
\frac{n}{\hat{\sigma}} &= \frac{1}{\hat{\sigma}^3} \sum_{i=1}^{n} \frac{(x_i - m_i \mu)^2}{m_i} \\
\hat{\sigma^2} &= \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - m_i \hat{\mu})^2}{m_i}
\end{aligned}$$

and lastly the MLE for p being obtained as

$$\begin{aligned}
\frac{\partial l_c(\mu, \sigma, p)}{\partial p} &= 0 \\
\frac{-\sum_{i=1}^{n}(m_i - 1)}{1 - \hat{p}} + \frac{n}{\hat{p}} &= 0 \\
-\hat{p} \sum_{i=1}^{n} m_i + n\hat{p} + n - n\hat{p} &= 0 \\
-\hat{p} \sum_{i=1}^{n} m_i &= -n \\
\hat{p} &= \frac{n}{\sum_{i=1}^{n} m_i}.
\end{aligned}$$

□

The complete log-likelihood in (48) directly results that the MLEs of the unknown parameters can be obtained in explicit forms based on the complete samples [24]. The EM-algorithm by Kundu [24] can subsequently be implemented and the algorithm is summarised in Algorithm 4.

Algorithm 4 EM-algorithm to obtain the MLEs of the $GSN(\mu, \sigma, p)$ distribution [24].

1: ***Required:*** *Denote* $\mu^{(k)}, \sigma^{(k)}$ *and* $p^{(k)}$ *as the estimates of* μ, σ *and* p *at the* k^{th} *stage of the algorithm. Initial guesses for* $\mu^{(k)}$ *and* $\sigma^{(k)}$ *can be taken as the sample mean and sample covariance, with* $p^{(k)}$ *as random adhering to the parameter constraint.*

2: ***'E'-step:*** *Obtain the pseudo log-likelihood function at the* k^{th} *stage by replacing the missing values of the complete log-likelihood function in (48) with their expectations.*

3: *The pseudo log-likelihood is given as follows:*

$$l_s^{(k)}(\mu, \sigma, p) = -\,n \ln \sigma - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 c_i^{(k)} - 2\mu \sum_{i=1}^{n} x_i + \mu^2 \sum_{i=1}^{n} d_i^{(k)}\right) + n \ln p + \ln(1-p) \sum_{i=1}^{n} \left(d_i^{(k)} - 1\right) \tag{50}$$

where $c_i^{(k)} = \frac{1}{m_i}$ *and* $d_i^{(k)} = m_i$ *and can be obtained using (46) and (45) respectively, by replacing* x, μ, σ, p *with their* k^{th} *stage estimates [24].*

4: ***'M'-step:*** *Maximise the pseudo log-likelihood as given in (50) with respect to the unknown parameters.*

5: *The 'M'-step yields the following:*

$$\mu^{(k+1)} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} d_i^{(k)}},$$

$$\sigma^{(k+1)} = \frac{1}{\sqrt{n}} \times \sqrt{\sum_{i=1}^{n} x_i^2 c_i^{(k)} - 2\mu^{(k+1)} \sum_{i=1}^{n} x_i + \left(\mu^{(k+1)}\right)^2 \sum_{i=1}^{n} d_i^{(k)}},$$

$$p^{(k+1)} = \frac{n}{\sum_{i=1}^{n} d_i^{(k)}}$$

noting specifically where (k) *and* $(k+1)$ *is used [24].*

6: *Repeat steps 2 - 5 until convergence.*

3.2 Multivariate geometric skew-normal distribution

This subsection will proceed to provide an overview of the multivariate extension of the geometric skew-normal distribution. This extension is presented by Kundu [25]. Figure 8 gives a summary of how Section 3.2 will proceed.

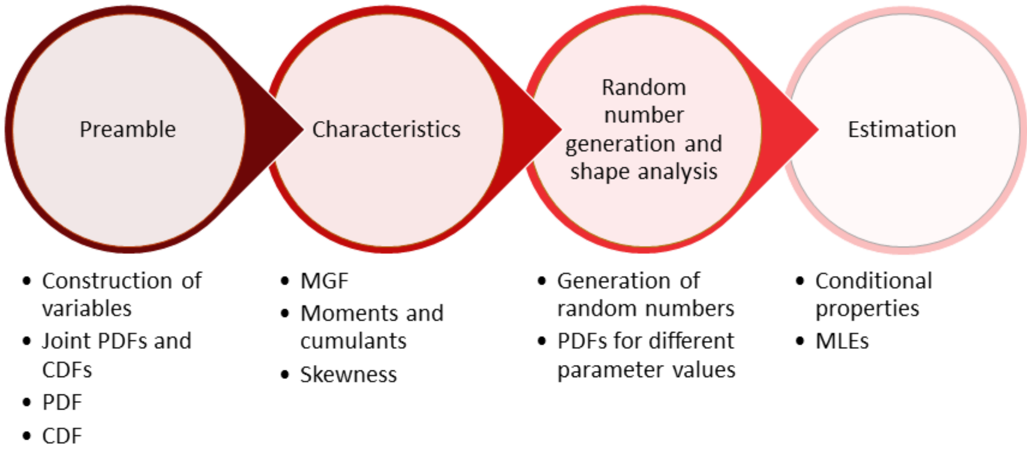


Figure 8: Summary of Section 3.2.

3.2.1 Preamble

Throughout this subsection a geometric random variable with parameter p will be denoted by $GE(p)$. A d-variate normal random variable with covariance matrix $\boldsymbol{\Sigma}$ and mean vector $\boldsymbol{\mu}$ will be denoted by $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The corresponding CDF and PDF will be denoted by $\Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ respectively. Furthermore, the convention that $0^0 = 1$ will also be used [25].

Theorem 30. *Let $\{\boldsymbol{X_i} : i = 1, 2, ...\}$ be i.i.d $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random variables and suppose that $N \sim GE(p)$, with N and $\boldsymbol{X_i'}s$ independently distributed. Then define*

$$\boldsymbol{X} \overset{dist.}{=} \sum_{i=1}^{N} \boldsymbol{X_i} \tag{51}$$

where $\overset{dist.}{=}$ indicates equal in distribution. It is then observed that $\boldsymbol{X}$ is a multivariate geometric skew-normal random variable with parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ and p [25]. This will be denoted by $MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$, where $d = 1, 2, ...$ indicates the number of variables.

Theorem 31. *Let $\boldsymbol{X} \sim MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ and $N \sim GE(p)$ with composition given by (51), then the joint PDF of the variable $(\boldsymbol{X}, N)$ is given by*

$$f_{\boldsymbol{X},N}(\boldsymbol{x}, n) = \begin{cases} \frac{p(1-p)^{n-1}}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}} n^{\frac{d}{2}}} e^{\frac{-1}{2n}(\boldsymbol{x}-n\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-n\boldsymbol{\mu})} & \text{if } 0 < p < 1 \\ \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{\frac{-1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} & \text{if } p = 1 \end{cases} \tag{52}$$

using the convention that $0^0 = 1$ when $p = 1$ and noting that $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$: $d \times d$ positive definite *covariance matrix and n is any positive integer [25].*

Theorem 32. *Let $\boldsymbol{X} \sim MVGSN_d\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p\right)$ and $N \sim GE(p)$ with composition given by (51), then the joint CDF of the variable $(\boldsymbol{X}, N)$ is given by*

$$F_{\boldsymbol{X},N}(\boldsymbol{x}, n) = \sum_{k=1}^{n} p\left(1-p\right)^{k-1} \Phi_d\left(k\boldsymbol{\mu}, k\boldsymbol{\Sigma}\right) \tag{53}$$

for $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$: $d \times d$ positive definite covariance matrix and $0 < p < 1$ [25].

Proof. Using (112) in Appendix A.1 and the fact that $\boldsymbol{X}|(N = k) \sim N_d(k\boldsymbol{\mu}, k\boldsymbol{\Sigma})$, it follows that

$$\begin{aligned}
F_{\boldsymbol{X},N}(\boldsymbol{x}, n) &= P\left(\boldsymbol{X} \leq \boldsymbol{x}, N \leq n\right) \\
&= P\left(\boldsymbol{X} \leq \boldsymbol{x}, N = 1\right) + P\left(\boldsymbol{X} \leq \boldsymbol{x}, N = 2\right) + \cdots \\
&\quad + P\left(\boldsymbol{X} \leq \boldsymbol{x}, N = n\right) \\
&= \sum_{k=1}^{n} P\left(\boldsymbol{X} \leq \boldsymbol{x}, N = k\right) \\
&= \sum_{k=1}^{n} P\left(\boldsymbol{X} \leq \boldsymbol{x} | N = k\right) P\left(N = k\right) \\
&= \sum_{k=1}^{n} p\left(1-p\right)^{k-1} \Phi_d\left(k\boldsymbol{\mu}, k\boldsymbol{\Sigma}\right)
\end{aligned}$$

□

Naturally, the question arises of what would happen to the CDF if $p = 1$? If $p = 1$, then $P(N = 1) = 1$ and $F_{\boldsymbol{X},N}(\boldsymbol{x}, n) = \Phi_d\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$.

Theorem 33. *Let $\boldsymbol{X} \sim MVGSN_d\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p\right)$ and $N \sim GE(p)$ with composition given by (51), then the CDF of the random variable $\boldsymbol{X}$ is given by*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{k=1}^{\infty} p\left(1-p\right)^{k-1} \Phi_d\left(k\boldsymbol{\mu}, k\boldsymbol{\Sigma}\right) \tag{54}$$

for $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$: $d \times d$ positive definite covariance matrix and $0 < p < 1$ [25].

Proof. From Stewart [37], it follows that

$$\begin{aligned}
F_{\boldsymbol{X}}(\boldsymbol{x}) &= \lim_{n\to\infty} F_{\boldsymbol{X},N}(\boldsymbol{x}, n) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} p\left(1-p\right)^{k-1} \Phi_d\left(k\boldsymbol{\mu}, k\boldsymbol{\Sigma}\right) \\
&= \sum_{k=1}^{\infty} p\left(1-p\right)^{k-1} \Phi_d\left(k\boldsymbol{\mu}, k\boldsymbol{\Sigma}\right).
\end{aligned}$$

□

As, before the question arises of what would happen to the CDF if $p = 1$? If $p = 1$, then $P(N = 1) = 1$ and $F_{\boldsymbol{X},N}(\boldsymbol{x}, n) = \Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Theorem 34. *Let* $\boldsymbol{X} \sim MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ *and* $N \sim GE(p)$ *with composition given by (51), then the PDF of the random variable* $\boldsymbol{X}$ *is given by*

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}} k^{\frac{d}{2}}} e^{\frac{-1}{2k}(\boldsymbol{x}-k\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})} \tag{55}$$

for $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$*:* $d \times d$ *positive definite covariance matrix and* $0 < p < 1$ *[25].*

Proof. Using (116) in Appendix A.3, it follows that

$$\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}) &= \frac{d}{d\boldsymbol{x}} F_{\boldsymbol{X}}(\boldsymbol{x}) \\
&= \frac{d}{d\boldsymbol{x}} \sum_{k=1}^{\infty} p(1-p)^{k-1} \Phi_d(k\boldsymbol{\mu}, k\boldsymbol{\Sigma}) \\
&= \sum_{k=1}^{\infty} p(1-p)^{k-1} \phi_d(k\boldsymbol{\mu}, k\boldsymbol{\Sigma}) \\
&= \sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}} k^{\frac{d}{2}}} \mathrm{e}^{\frac{-1}{2k}(\boldsymbol{x}-k\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})}
\end{aligned}$$

□

If $p = 1$, then $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. When $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$, then the distribution of $\boldsymbol{X}$ becomes $MVGSN_d(p)$ and it is known as the standard multivariate geometric skew-normal distribution [25]. The $MVGSN_d(p)$ distribution is a symmetric and unimodal distribution for all values of p and d [25].

3.2.2 Characteristics of the multivariate geometric skew-normal distribution

The characteristics that will be provided in this section include the moments and cumulants as well as the skewness. The MGF will also be given in this section.

Moment generating function

Theorem 35. *The MGF of a random variable* $\boldsymbol{X} \sim MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ *with PDF given in (55) is given in Kundu [25] by*

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \frac{p\mathrm{e}^{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}+(1/2)\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}}}{1-(1-p)\mathrm{e}^{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}+(1/2)\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}}} \tag{56}$$

where $\boldsymbol{t} \in A_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ *and*

$$\begin{aligned} A_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p) &= \left\{\boldsymbol{t}; \boldsymbol{t} \in \mathbb{R}^d, (1-p)\mathrm{e}^{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}+(1/2)\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}} < 1\right\} \\ &= \left\{\boldsymbol{t}; \boldsymbol{t} \in \mathbb{R}^d, \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t} + \ln(1-p) < 0\right\}. \end{aligned}$$

Proof. From Bain and Engelhardt [8]

$$\begin{aligned} M_{\boldsymbol{X}}(\boldsymbol{t}) &= E\left[e^{\boldsymbol{X}^T\boldsymbol{t}}\right] \\ &= E_N\left[E\left(e^{\boldsymbol{X}^T\boldsymbol{t}}|N\right)\right] \end{aligned} \tag{57}$$

where $N \sim GE(p)$. Since the $\boldsymbol{X_i}$'s are i.i.d. $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random variables, it follows that

$$\begin{aligned} E\left(e^{\boldsymbol{X}^T\boldsymbol{t}}|N=n\right) &= E\left[e^{\left(\sum_{i=1}^n \boldsymbol{X}_i\right)^T\boldsymbol{t}}\right] \\ &= \prod_{i=1}^n E\left[e^{\boldsymbol{X}_i^T\boldsymbol{t}}\right] \\ &= \prod_{i=1}^n M_{\boldsymbol{X}_i}(\boldsymbol{t}) \\ &= [M_{\boldsymbol{Y}}(\boldsymbol{t})]^n \\ &= \left[\mathrm{e}^{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}+(1/2)\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}}\right]^n. \end{aligned} \tag{58}$$

Substituting (58) into (57), using the fact that if $\boldsymbol{Y} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then $\ln(M_{\boldsymbol{Y}}(\boldsymbol{t})) = \mathrm{e}^{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}+(1/2)\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}}$ and using (33), it follows that

$$\begin{aligned} M_{\boldsymbol{X}}(\boldsymbol{t}) &= E_N\left[\mathrm{e}^{N\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}+(1/2)N\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}}\right] \\ &= E_N\left[e^{N\ln(M_{\boldsymbol{Y}}(\boldsymbol{t}))}\right] \\ &= M_N\left(\ln(M_{\boldsymbol{Y}}(\boldsymbol{t}))\right) \\ &= \frac{p\mathrm{e}^{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}+(1/2)\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}}}{1-(1-p)\mathrm{e}^{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}+(1/2)\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}}}. \end{aligned}$$

□

Moments, cumulants and skewness

The cumulants and moments of $\boldsymbol{X}$ can be obtained from the MGF in (56) for $i, j = 1, 2, ..., d$. Suppose $\boldsymbol{X} = (X_1, X_2, ..., X_d)^T \sim MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ and denote $\boldsymbol{\mu}^T = (\mu_1, \mu_2, ..., \mu_d)$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$ for $i, j =$

$1, 2, ..., d$. Then the moments and cumulants are given in Kundu [25] by

$$E(X_i) = \frac{\partial}{\partial t_i} M_{\mathbf{X}}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} = \frac{\mu_i}{p} \tag{59}$$

$$E(X_i X_j) = \frac{\partial^2}{\partial t_i \partial t_j} M_{\mathbf{X}}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} = \frac{p\sigma_{ij} + \mu_i\mu_j(2-p)}{p^2}. \tag{60}$$

Hence, Kundu [25] presented

$$\operatorname{Var}(X_i) = \frac{p\sigma_{ii} + \mu_i^2(1-p)}{p^2} \tag{61}$$

$$\operatorname{Cov}(X_i, X_j) = \frac{p\sigma_{ij} + \mu_i\mu_j(1-p)}{p^2} \tag{62}$$

and

$$\operatorname{Corr}(X_i, X_j) = \frac{p\sigma_{ij} + \mu_i\mu_j(1-p)}{\sqrt{p\sigma_{ii} + \mu_i^2(1-p)}\sqrt{p\sigma_{jj} + \mu_j^2(1-p)}}. \tag{63}$$

When considering the correlation between X_i and X_j for $i \neq j$, it is observed from (63) that there is dependence on μ_i and μ_j and not only on σ_{ij} [25]. If $\mu_j, \mu_i \to \infty$ for fixed p and σ_{ij}, then $\operatorname{Corr}(X_i, X_j) \to 1$. If $\mu_j, \mu_i \to \infty$, then $\operatorname{Corr}(X_i, X_j) \to -1$ [25]. It can therefore be concluded that there is dependence between X_i and X_j in this case, although they may be uncorrelated (in the case of a standard $MVGSN_d(\cdot)$ distribution) [25].

It is also of interest to provide the multivariate skewness indices of the $MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ distribution. Various multivariate skewness measures have been introduced in literature on multivariate distributions. The most common one is the skewness index of Mardia [27] [28]. In order to proceed, it is necessary to present the following notations of a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$:

$$\mu_{i_1,\ldots,i_s}^{(r_1,\ldots,r_s)} = E\left[\prod_{k=1}^{s} (X_{r_k} - \mu_{r_k})^{i_k}\right] \tag{64}$$

where $\mu_{r_k} = E(X_{r_k}), k = 1, \ldots, s$. It then follows that Mardia [27] defined the multivariate skewness index as

$$\gamma_1 = \sum_{r,s,t=1}^{d} \sum_{r',s',t'=1}^{d} \sigma^{rr'}\sigma^{ss'}\sigma^{tt'}\mu_{111}^{rst}\mu_{111}^{r's't'}. \tag{65}$$

Here σ^{jk} for $j, k = 1, \ldots, d$ denotes the $(j,k)^{th}$ element of the inverse of the covariance matrix of the random vector $\boldsymbol{X}$ [25]. In the case of the $MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ distribution [25],

$$\mu_{111}^{(qwv)} = \frac{1}{p^4}\left\{p(1-p)(2-p)\mu_q\mu_w\mu_v + p^2(1-p)\left(\mu_v\sigma^{wq} + \mu_q\sigma^{wv} + \mu_w\sigma^{qv}\right)\right\}. \tag{66}$$

It is observed from (66) that if $p = 1$ then $\gamma_1 = 0$. In addition to the latter, if $\mu_j = 0$ for all $j = 1, \ldots, d$, then $\gamma_1 = 0$. Furthermore, Kundu [25] also notes that the skewness index γ_1 may diverge to either ∞ or $-\infty$ as $p \to 0$, when $\mu_j \neq 0$ for some $j = 1, \ldots, d$.

3.2.3 Generation of random numbers and illustration of PDF

In order to generate from the $MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ distribution with PDF given in (55) a short algorithm is provided by Kundu [25].

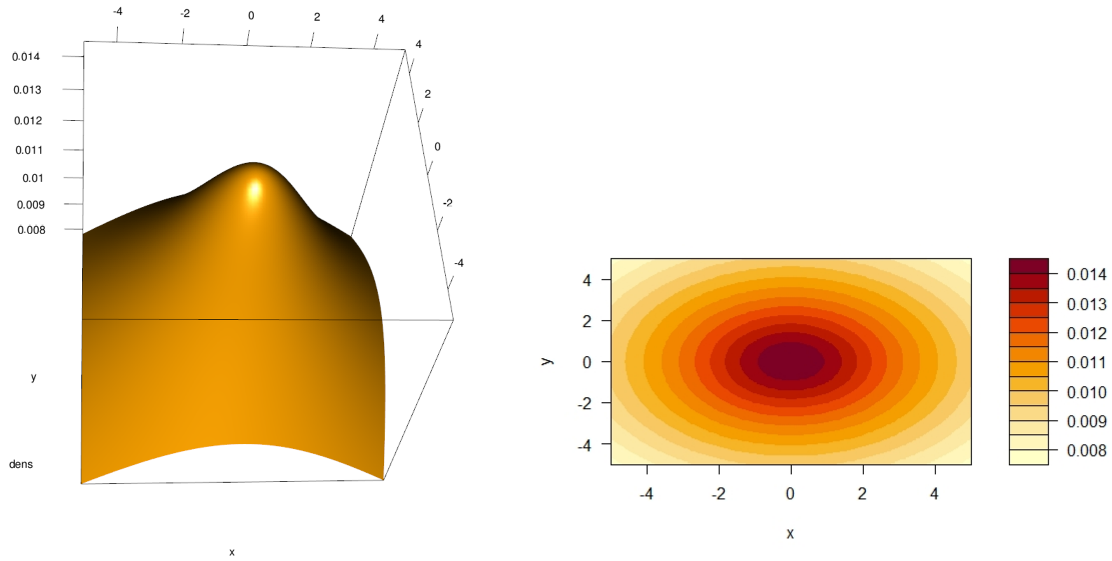
Algorithm 5 Generation from the $MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ distribution

1: ***Required:***
- *Define the values of* $\boldsymbol{\mu}$ *for* $\boldsymbol{\mu} \in \mathbb{R}^d$.
- *Define the values of* $\boldsymbol{\Sigma}$ *for* $\boldsymbol{\Sigma} : d \times d$ *positive definite covariance matrix.*
- *Define the value of* p *for* $0 < p \leq 1$.

2: *Generate the value* k *from the* $GE(p)$ *distribution with* p *defined as before.*

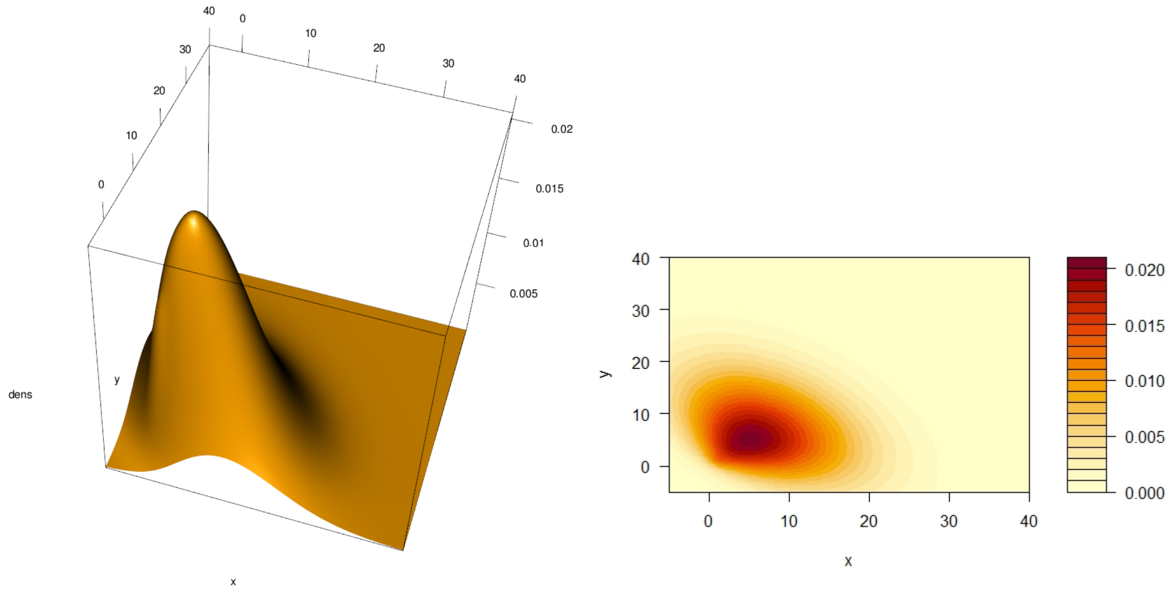
3: *Recall that* $\boldsymbol{X}|(N = k) \sim N_d(k\boldsymbol{\mu}, k\boldsymbol{\Sigma})$ *with* $N \sim GE(p)$. *Generate the sample* $\boldsymbol{X}$ *from the* $N_d(k\boldsymbol{\mu}, k\boldsymbol{\Sigma})$ *distribution, where* k *is from step 2. This* $\boldsymbol{X}$ *is then the required sample.*

Using the PDF given in (55), graphs of the PDFs of the $MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ distribution (for $d = 2$) are provided in Figures 9, 10 and 11. That is, the latter figures plot the PDFs of the bivariate geometric skew-normal distribution, given specific parameter values.
Figure 9 is plotted for $\boldsymbol{\mu} = (0,0)^T$, $\sigma_{11} = \sigma_{22} = 2$, $\sigma_{12} = \sigma_{21} = 0$ and $p = 0.75$. The PDFs in Figures 10 and 11 are plotted with $\sigma_{11} = \sigma_{22} = 1$, $\sigma_{12} = \sigma_{21} = -0.5$ and $p = 0.5$ kept constant throughout, varying the values of $\boldsymbol{\mu}$. It is observed from Figures 9, 10 and 11 that the PDF of the $MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ distribution can be unimodal or multimodal depending on the parameter values. The latter observation of multimodality as seen in Figure 11, is different from the multivariate skew-normal distribution that is always unimodal [25].



(a) 3D plot (b) Contour plot

Figure 9: Bivariate geometric skew-normal PDF.

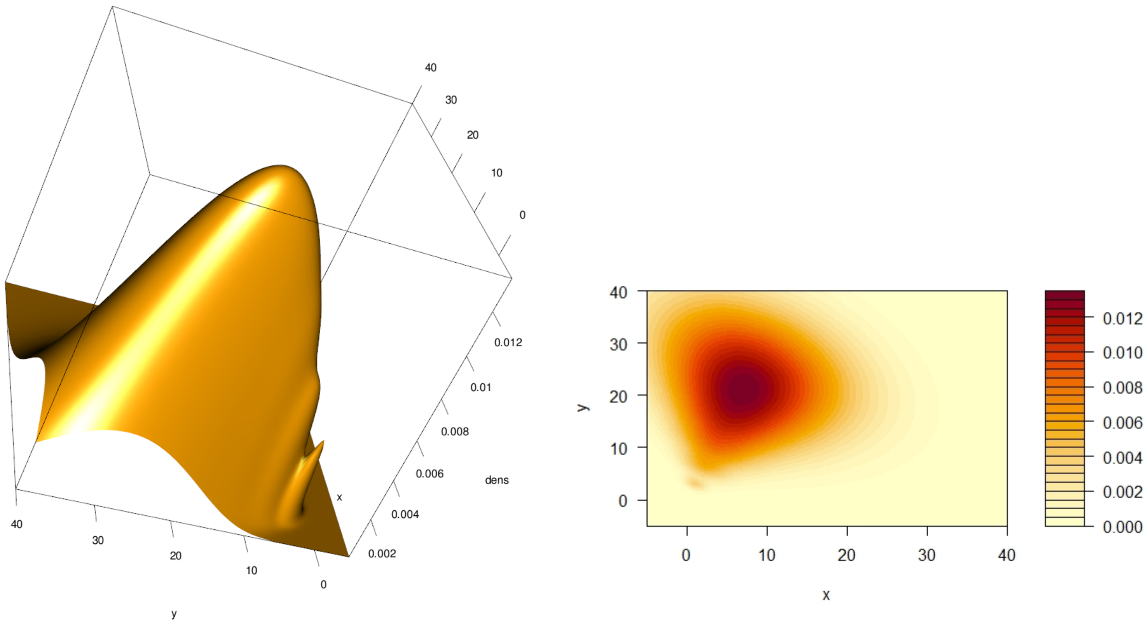


(a) 3D plot (b) Contour plot

Figure 10: Bivariate geometric skew-normal PDF with $\boldsymbol{\mu} = (1, 1)^T$.

3.2.4 Estimation

Before continuing with the estimation section, it is worthwhile to first furnish a couple of conditional properties presented by Kundu [25] that will be utilized in the derivation of the estimation theory.



(a) 3D plot

(b) Contour plot

Figure 11: Bivariate geometric skew-normal PDF with $\boldsymbol{\mu} = (1, 3)^T$.

Kundu [25] noted that using (52) and (55), the conditional PMF of N given $\boldsymbol{X} = \boldsymbol{x}$ is given by

$$
\begin{aligned}
P(N = n \mid \boldsymbol{X} = \boldsymbol{x}) &= \frac{P\left(N = n, \boldsymbol{X} = \boldsymbol{x}\right)}{P\left(\boldsymbol{X} = \boldsymbol{x}\right)} \\
&= \frac{\frac{p(1-p)^{n-1}}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}} n^{\frac{d}{2}}} \mathrm{e}^{\frac{-1}{2n}(\boldsymbol{x}-n\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-n\boldsymbol{\mu})}}{\sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}} k^{\frac{d}{2}}} \mathrm{e}^{\frac{-1}{2k}(\boldsymbol{x}-k\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})}} \\
&= \frac{(1-p)^{n-1} \mathrm{e}^{-(1/2n)(\boldsymbol{x}-n\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-n\boldsymbol{\mu})} n^{-d/2}}{\sum_{k=1}^{\infty} (1-p)^{k-1} \mathrm{e}^{-(1/2k)(\boldsymbol{x}-k\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})} k^{-d/2}}
\end{aligned}
$$

and in turn the conditional expectation of N given $\boldsymbol{X} = \boldsymbol{x}$ becomes

$$
E(N \mid \boldsymbol{X} = \boldsymbol{x}) = \frac{\sum_{n=1}^{\infty} (1-p)^{n-1} \mathrm{e}^{-(1/2n)(\boldsymbol{x}-n\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-n\boldsymbol{\mu})} n^{-d/2+1}}{\sum_{k=1}^{\infty} (1-p)^{k-1} \mathrm{e}^{-(1/2k)(\boldsymbol{x}-k\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})} k^{-d/2}} \tag{67}
$$

with

$$
E\left(N^{-1} \mid \boldsymbol{X} = \boldsymbol{x}\right) = \frac{\sum_{n=1}^{\infty} (1-p)^{n-1} \mathrm{e}^{-(1/2n)(\boldsymbol{x}-n\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-n\boldsymbol{\mu})} n^{-d/2-1}}{\sum_{k=1}^{\infty} (1-p)^{k-1} \mathrm{e}^{-(1/2k)(\boldsymbol{x}-k\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})} k^{-d/2}}. \tag{68}
$$

Maximum likelihood estimators

In order to obtain the MLEs of the unknown parameters it is necessary to maximize the log-likelihood function with respect to the unknown parameters.

Theorem 36. *Suppose $\mathcal{D} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n}\}$ is a random sample of size n from the $MVGSN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$, distribution with PDF as given in (55). In Kundu [25], the log-likelihood function is then given by*

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p) = \sum_{i=1}^{n} \ln \left[\sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2} k^{d/2}} \mathrm{e}^{-(1/2k)(\boldsymbol{x}_i - k\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - k\boldsymbol{\mu})} \right]. \tag{69}$$

Proof. From Bain and Engelhardt [8] and using (55), it follows that

$$\begin{aligned} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p) &= \ln \left[\prod_{i=1}^{n} f_{\boldsymbol{X_i}}(\boldsymbol{x_i}) \right] \\ &= \sum_{i=1}^{n} \ln f_{\boldsymbol{X_i}}(\boldsymbol{x_i}) \\ &= \sum_{i=1}^{n} \ln \left[\sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2} k^{d/2}} \mathrm{e}^{-(1/2k)(\boldsymbol{x}_i - k\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - k\boldsymbol{\mu})} \right] \end{aligned}$$

□

It is proceeded to obtain the MLEs by maximizing (69) with respect to the unknown parameters. That is, obtaining the normal equations by taking partial derivatives of (69) and equating them to 0. These partial derivatives are with respect to $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ and p. The latter translates to solving an optimization problem of $\left[d + 1 + \frac{d(d+1)}{2}\right]$ dimensions yielding a complicated issue for large d [25]. To circumvent this problem, it is assumed that p is known and the MLEs will be estimated using the EM-algorithm [25]. In essence it will be proceeded to maximize $l\left(\hat{\boldsymbol{\mu}}(p), \hat{\boldsymbol{\Sigma}}(p), p\right)$ to compute the MLE of p, that is $\hat{p}$. Then the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ will be obtained as $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}(p)$ and $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(p)$ respectively. The EM-algorithm will be implemented for a known p, where the sample mean and sample covariance will be used as initial guesses of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for all values of p [25].

Theorem 37. *Let $\{(\boldsymbol{x_1}, m_1), (\boldsymbol{x_2}, m_2), ..., (\boldsymbol{x_n}, m_n)\}$ be a random sample of size n from the random variable $(\boldsymbol{X}, N)$. In Kundu [25], the log-likelihood function based on the complete sample is given by*

$$l_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{n}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^{n} \frac{1}{m_i} \left(\boldsymbol{x_i} - m_i \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{x_i} - m_i \boldsymbol{\mu}\right) \tag{70}$$

without the additive constant and the MLEs of the unknown parameters are obtained as

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{n} \boldsymbol{x_i}}{\sum_{i=1}^{n} m_i} \tag{71}$$

and

$$
\begin{aligned}
\hat{\boldsymbol{\Sigma}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - m_i\hat{\boldsymbol{\mu}}\right)\left(\boldsymbol{x_i} - m_i\hat{\boldsymbol{\mu}}\right)^T \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}\frac{1}{m_i}\boldsymbol{x_i}\boldsymbol{x_i}^T - \sum_{i=1}^{n}\left(\hat{\boldsymbol{\mu}}\boldsymbol{x_i}^T + \boldsymbol{x_i}\hat{\boldsymbol{\mu}}^T\right) + \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^T\sum_{i=1}^{n} m_i\right].
\end{aligned}
\tag{72}
$$

Proof.

$$
\begin{aligned}
l_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \left[\prod_{i=1}^{n} f_{\boldsymbol{X},N}(\boldsymbol{x}, n)\right] \\
&= \sum_{i=1}^{n}\ln\left[f_{\boldsymbol{X},N}(\boldsymbol{x}, n)\right] \\
&= \sum_{i=1}^{n}\ln\left[\frac{p(1-p)^{m_i-1}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}m_i^{d/2}}\mathrm{e}^{-(\frac{1}{2m_i})(\boldsymbol{x}_i - m_i\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - m_i\boldsymbol{\mu})}\right] \\
&= \sum_{i=1}^{n}\left[-\left(\frac{1}{2m_i}\right)(\boldsymbol{x}_i - m_i\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - m_i\boldsymbol{\mu}) + (m_i - 1)\ln(1-p)\right] \\
&\quad + \sum_{i=1}^{n}\left[\ln\left(\frac{p}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}m_i^{d/2}}\right)\right] \\
&= -\frac{1}{2}\sum_{i=1}^{n}\frac{1}{m_i}(\boldsymbol{x}_i - m_i\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - m_i\boldsymbol{\mu}) + \ln(1-p)\sum_{i=1}^{n}(m_i - 1) + n\ln(p) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| \\
&\quad - \frac{nd}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n} ln(m_i) \\
&\propto -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right).
\end{aligned}
$$

In order to obtain the MLE for $\boldsymbol{\mu}$ it is necessary to obtain the partial derivative of (70) and set it equal to 0, followed by solving for the unknown $\boldsymbol{\mu}$. Thus, the MLE for $\boldsymbol{\mu}$ is given by

$$
\begin{aligned}
\frac{\partial l_c(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial\boldsymbol{\mu}} &= \mathbf{0} \\
-\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - m_i\hat{\boldsymbol{\mu}}\right)(m_i) &= \mathbf{0} \\
\sum_{i=1}^{n}\left(\boldsymbol{x_i} - m_i\hat{\boldsymbol{\mu}}\right) &= \mathbf{0} \\
\sum_{i=1}^{n}\boldsymbol{x_i} &= \sum_{i=1}^{n} m_i\hat{\boldsymbol{\mu}} \\
\hat{\boldsymbol{\mu}} &= \frac{\sum_{i=1}^{n}\boldsymbol{x_i}}{\sum_{i=1}^{n} m_i}.
\end{aligned}
$$

Obtaining the MLE for $\boldsymbol{\Sigma}$ involves using the matrix property where trace($\boldsymbol{CB}$) = trace($\boldsymbol{BC}$) and the fact that the trace of a scalar is a scalar [33]. In addition to the latter, another property will be used that was developed by Anderson [2] (see (121) in Appendix A.3). Now recall from (70) that

$$\begin{aligned} l_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right) \\ &= -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right)\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right)^T\right] \end{aligned} \tag{73}$$

and since

$$\begin{aligned} &\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right)\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right)^T \\ &= \sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - n\bar{\boldsymbol{x}} + n\bar{\boldsymbol{x}} - m_i\boldsymbol{\mu}\right)\left(\boldsymbol{x_i} - n\bar{\boldsymbol{x}} + n\bar{\boldsymbol{x}} - m_i\boldsymbol{\mu}\right)^T \\ &= \sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - n\bar{\boldsymbol{x}}\right)\left(\boldsymbol{x_i} - n\bar{\boldsymbol{x}}\right)^T + \sum_{i=1}^{n}\frac{1}{m_i}\left(n\bar{\boldsymbol{x}} - m_i\boldsymbol{\mu}\right)\left(n\bar{\boldsymbol{x}} - m_i\boldsymbol{\mu}\right)^T \\ &= \boldsymbol{A} + \sum_{i=1}^{n}\frac{1}{m_i}\left(n\bar{\boldsymbol{x}} - m_i\boldsymbol{\mu}\right)\left(n\bar{\boldsymbol{x}} - m_i\boldsymbol{\mu}\right)^T \end{aligned}$$

it follows that (73) becomes

$$l_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{A} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right). \tag{74}$$

Since $\boldsymbol{\Sigma}$ is a positive definite matrix, $\boldsymbol{\Sigma}^{-1}$ is also positive definite. It is also noted that $\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x_i} - m_i\boldsymbol{\mu}\right) \geq \mathbf{0}$ and is $\mathbf{0}$ if $\sum_{i=1}^{n} m_i\boldsymbol{\mu} = \sum_{i=1}^{n}\boldsymbol{x_i} = n\bar{\boldsymbol{x}}$. In order to maximize the first term and the second term of (74), the result given in (121) is used and the MLE of $\boldsymbol{\Sigma}$ is found to be

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \frac{1}{n}\boldsymbol{A} \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - n\bar{\boldsymbol{x}}\right)\left(\boldsymbol{x_i} - n\bar{\boldsymbol{x}}\right)^T \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\left(\boldsymbol{x_i} - m_i\hat{\boldsymbol{\mu}}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x_i} - m_i\hat{\boldsymbol{\mu}}\right) \\ &= \frac{1}{n}\left[\sum_{i=1}^{n}\frac{1}{m_i}\boldsymbol{x_i}\boldsymbol{x_i}^T - \sum_{i=1}^{n}\left(\hat{\boldsymbol{\mu}}\boldsymbol{x_i}^T + \boldsymbol{x_i}\hat{\boldsymbol{\mu}}^T\right) + \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^T\sum_{i=1}^{n} m_i\right]. \end{aligned}$$

□

The log-likelihood function in (70) directly results that the MLEs of the unknown parameters can be obtained in explicit forms based on the complete samples. The EM-algorithm presented by Kundu [25] can subsequently be implemented for a given p. The main idea of the EM-algorithm follows along the line of utilizing the complete log-likelihood by maximizing the conditional expectation thereof, all based on the observed data at hand and the current value of $\boldsymbol{\delta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, say $\tilde{\boldsymbol{\delta}}$. The following notation will be used in the proceeding parts:

$$h_i = E\left(N|\boldsymbol{X} = \boldsymbol{x_i}, \tilde{\boldsymbol{\delta}}\right)$$
$$g_i = E\left(N^{-1}|\boldsymbol{X} = \boldsymbol{x_i}, \tilde{\boldsymbol{\delta}}\right)$$

where h_i and g_i can be obtained from (67) and (68) respectively. The EM-algorithm is summarised in Algorithm 6.

Algorithm 6 EM-algorithm to obtain the MLEs of the $MVGSN(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p)$ distribution [25].

1: ***Required:*** *Denote* $\tilde{\boldsymbol{\delta}}$ *as the current value of* $\boldsymbol{\delta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Define an initial guess for* $\tilde{\boldsymbol{\delta}}$, *say* $\tilde{\boldsymbol{\delta}}^{(0)}$. *The sample mean vector and sample covariance matrix as initial guesses will suffice.*
2: ***'E'-step:*** *Obtain the conditional expectation denoted by* $Q\left(\boldsymbol{\delta}|\tilde{\boldsymbol{\delta}}\right)$ *with* $\tilde{\boldsymbol{\delta}}$ *being the current value.*
3: *The conditional expectation is given as follows:*

$$\begin{aligned} Q\left(\boldsymbol{\delta}|\tilde{\boldsymbol{\delta}}\right) &= E\left[l_c\left(\boldsymbol{\delta}|D\left(\tilde{\boldsymbol{\delta}}\right)\right)\right] \\ &= -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\left(\sum_{i=1}^{n} g_i \boldsymbol{x_i}\boldsymbol{x_i}^{\mathrm{T}} - \sum_{i=1}^{n}\left(\boldsymbol{x_i}\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\mu}\boldsymbol{x_i}^{\mathrm{T}}\right) + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\sum_{i=1}^{n} h_i\right)\right\} \end{aligned} \tag{75}$$

where h_i *and* g_i *and can be obtained using (67) and (68) respectively [25].*
4: ***'M'-step:*** *Maximise* $Q(\boldsymbol{\delta}|\tilde{\boldsymbol{\delta}})$ *with respect to* $\boldsymbol{\delta}$ *to obtain* $\boldsymbol{\delta}^{(1)}$. *That is, obtain* $\boldsymbol{\delta}^{(1)} = \arg\max_{\boldsymbol{\delta}} Q(\boldsymbol{\delta} \mid \tilde{\boldsymbol{\delta}})$, *where* $\arg\max_{\boldsymbol{\delta}} Q(\boldsymbol{\delta} \mid \tilde{\boldsymbol{\delta}})$ *indicates the value for which* $Q(\boldsymbol{\delta}|\tilde{\boldsymbol{\delta}})$ *is a maximum.*
5: *The 'M'-step yields the following:*

$$\begin{aligned} \boldsymbol{\mu}^{(k)} &= \frac{\sum_{i=1}^{n} \boldsymbol{x_i}}{\sum_{j=1}^{n} h_j}, \\ \boldsymbol{\Sigma}^{(k)} &= \frac{1}{n}\left[\sum_{i=1}^{n} g_i \boldsymbol{x_i}\boldsymbol{x_i}^{\mathrm{T}} - \sum_{i=1}^{n}\left(\boldsymbol{x_i}\boldsymbol{\mu}^{(k)\mathrm{T}} + \boldsymbol{\mu}^{(k)}\boldsymbol{x_i}^{\mathrm{T}}\right) + \boldsymbol{\mu}^{(k)}\boldsymbol{\mu}^{(k)\mathrm{T}}\sum_{i=1}^{n} h_i\right] \end{aligned}$$

where $\boldsymbol{\mu}^{(k)}$ *and* $\boldsymbol{\Sigma}^{(k)}$ *are the estimates of* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *at the* k^{th} *stage of the algorithm [25].*
6: *Repeat steps 2 - 5 until convergence.*

4 Chapter 4

Chapter 4 will develop the geometric skew-Cauchy distribution. In Section 4.1, the new alternative distribution called the geometric skew-Cauchy distribution is presented.

4.1 Univariate geometric skew-Cauchy distribution

This section will proceed to introduce a new approach to model skewed data. The new distribution will be presented in a similar fashion to the univariate geometric skew-normal distribution. In this section, the geometric skew-Cauchy distribution is introduced. The geometric skew-Cauchy distribution can be used as an alternative to the skew-normal and the geometric skew-normal distributions. Figure 12 gives a summary of how Section 4.1 will proceed.

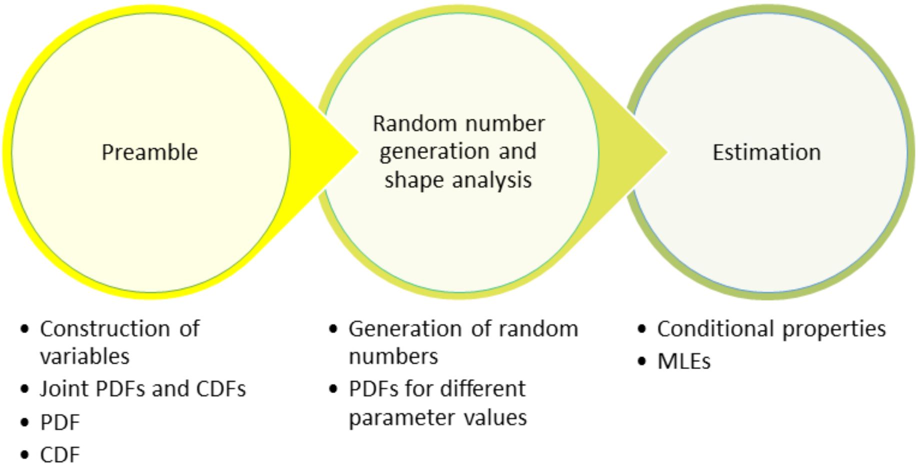


Figure 12: Summary of Section 4.1.

4.1.1 Preamble

Throughout this section a Cauchy random variable with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$ will be denoted by $C(\mu, \sigma)$. Let $Z \sim C(0, 1)$ be a standard Cauchy random variable, then $Y = \mu + \sigma Z \sim C(\mu, \sigma)$ [40].

Theorem 38. *A random variable Y that follows the Cauchy distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, that is $Y \sim C(\mu, \sigma)$, has a PDF of the following form:*

$$f_Y(y) = \frac{1}{\pi\sigma}\left[\frac{\sigma^2}{(y-\mu)^2 + \sigma^2}\right] \tag{76}$$

for $-\infty < y < \infty$ [40].

Proof. The full proof is provided in Walck [40] □

Theorem 39. *A random variable Y that follows the Cauchy distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, that is $Y \sim C(\mu, \sigma)$, has a CDF of the following form:*

$$F_Y(y) = \frac{1}{\pi} \arctan\left[\frac{y-\mu}{\sigma}\right] + \frac{1}{2} \tag{77}$$

for $-\infty < y < \infty$ [40].

Proof. From Stewart [37] and using (76), it follows that

$$\begin{aligned}
F_Y(y) &= \int_{-\infty}^{y} f_G(g)dg \\
&= \int_{-\infty}^{y} \frac{\frac{\sigma}{\pi}}{(g-\mu)^2+\sigma^2} dg \\
&= \frac{\sigma}{\pi}\left[\frac{1}{\sigma}\arctan\left(\frac{g-\mu}{\sigma}\right)\right]\Big|_{-\infty}^{y} \\
&= \frac{1}{\pi}\left[\arctan\left(\frac{y-\mu}{\sigma}\right) - \arctan(-\infty)\right] \\
&= \frac{1}{\pi}\left[\frac{\pi}{2} + \arctan\left(\frac{y-\mu}{\sigma}\right)\right] \\
&= \frac{1}{\pi}\arctan\left(\frac{y-\mu}{\sigma}\right) + \frac{1}{2}.
\end{aligned}$$

□

Theorem 40. *Let $Y_i \sim C(\mu, \sigma)$ with $\mu \in \mathbb{R}, \sigma > 0$ and $n \in \{1, 2, ...\}$. Then for the Y_i's independent, it is true that:*

$$\sum_{i=1}^{n} Y_i \sim C(n\mu, n\sigma). \tag{78}$$

Proof. The full proof is provided in Walck [40]. □

Theorem 41. *Let $\{Y_i : i = 1, 2, ...\}$ be i.i.d. $C(\mu, \sigma)$ random variables and suppose that $N \sim GE(p)$, with N and Y_i's independently distributed. Then define*

$$X \stackrel{d}{=} \sum_{i=1}^{N} Y_i \tag{79}$$

where $\stackrel{d}{=}$ indicates equal in distribution. It is then said that X is a geometric skew-Cauchy random variable with parameters $\mu \in \mathbb{R}, \sigma > 0$ and $0 < p \leq 1$. This will be denoted by $GSC(\mu, \sigma, p)$.

Theorem 42. *Let $X \sim GSC(\mu, \sigma, p)$ and $N \sim GE(p)$ with composition given by (79), then the joint PDF of the variable (X, N) is given by*

$$f_{X,N}(x,n) = \begin{cases} \frac{1}{\pi\sigma n}\left[\frac{1}{1+\left(\frac{x-n\mu}{n\sigma}\right)^2}\right] p(1-p)^{n-1} & \text{if } 0 < p < 1 \\ \frac{1}{\pi\sigma}\left[\frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2}\right] & \text{if } p = 1 \end{cases} \tag{80}$$

using the convention that $0^0 = 1$ when $p = 1$ and noting that $-\infty < x < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$ and n is any positive integer.

Proof. Let $N \sim GE(p)$. The proof will be given for the case where $0 < p < 1$. When $p = 1$, then $P(N = n) = 1$. The result for $p = 1$ is then immediate using the Cauchy distribution PDF in (76).

Recall that if $Y \sim C(\mu, \sigma)$, then $\sum_{i=1}^{n} Y_i \sim C(n\mu, n\sigma)$ as given in (78). Then it follows that $X|(N = n) = \sum_{i=1}^{n} Y_i \sim C(n\mu, n\sigma)$ [8]. Hence, from the latter, Bain and Engelhardt [8], (76) and (26) it follows that

$$\begin{aligned} f_{X,N}(x,n) &= f_{X|N}(x|n) \times f_N(n) \\ &= \frac{1}{\pi n\sigma}\left[\frac{n^2\sigma^2}{(x-n\mu)^2 + n^2\sigma^2}\right] p(1-p)^{n-1} \\ &= \frac{1}{\pi\sigma n}\left[\frac{1}{1+\left(\frac{x-n\mu}{n\sigma}\right)^2}\right] p(1-p)^{n-1} \end{aligned}$$

noting that $-\infty < x < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$ and n is any positive integer. □

Theorem 43. *Let $X \sim GSC(\mu, \sigma, p)$ and $N \sim GE(p)$ with composition given by (79), then the joint CDF of the variable (X, N) is given by*

$$F_{X,N}(x,n) = p\sum_{k=1}^{n}\left[\frac{1}{\pi}\arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2}\right](1-p)^{k-1} \tag{81}$$

for $-\infty < x < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$ and $0 < p < 1$.

Proof. Making use of (112) in Appendix A.1, (77), (78) and the fact that $X|(N = k) \sim C(k\mu, k\sigma)$, it follows that

$$
\begin{aligned}
F_{X,N}(x,n) &= P\left(X \leq x, N \leq n\right) \\
&= P\left(X \leq x, N = 1\right) + P\left(X \leq x, N = 2\right) + \\
&\quad \cdots + P\left(X \leq x, N = n\right) \\
&= \sum_{k=1}^{n} P\left(X \leq x, N = k\right) \\
&= \sum_{k=1}^{n} P\left(X \leq x | N = k\right) P\left(N = k\right) \\
&= p \sum_{k=1}^{n} \left[\frac{1}{\pi} \arctan\left(\frac{x - k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}.
\end{aligned}
$$

□

Naturally, the question arises of what would happen to the CDF if $p = 1$? If $p = 1$, then $P(N = 1) = 1$ and $F_{X,N}(x,n) = \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sigma}\right) + \frac{1}{2}$, hence only the Cauchy distribution CDF.

Theorem 44. *Let $X \sim GSC\left(\mu, \sigma, p\right)$ and $N \sim GE(p)$ with composition given by (79), then the CDF of the random variable X is given by*

$$
F_X(x) = p \sum_{k=1}^{\infty} \left[\frac{1}{\pi} \arctan\left(\frac{x - k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1} \tag{82}
$$

for $-\infty < x < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$ and $0 < p < 1$.

Proof. From Stewart [37], it follows that

$$
\begin{aligned}
F_X(x) &= \lim_{n\to\infty} F_{X,N}(x,n) \\
&= \lim_{n\to\infty} p \sum_{k=1}^{n} \left[\frac{1}{\pi} \arctan\left(\frac{x - k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1} \\
&= p \sum_{k=1}^{\infty} \left[\frac{1}{\pi} \arctan\left(\frac{x - k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}.
\end{aligned}
$$

□

As before, the question arises of what would happen to the CDF if $p = 1$? If $p = 1$, then $P(N = 1) = 1$ and $F_X(x) = \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sigma}\right) + \frac{1}{2}$.

Theorem 45. *Let $X \sim GSC(\mu, \sigma, p)$ and $N \sim GE(p)$ with composition given by (79), then the PDF of the random variable X is given by*

$$f_X(x) = \frac{p}{\pi\sigma} \sum_{k=1}^{\infty} \frac{1}{k} \left[\frac{1}{1 + \left(\frac{x-k\mu}{k\sigma}\right)^2} \right] (1-p)^{k-1} \tag{83}$$

for $-\infty < x < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$ and $0 < p < 1$.

Proof. Using (116) in Appendix A.3, it follows that

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \frac{d}{dx} p \sum_{k=1}^{\infty} \left[\frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2} \right] (1-p)^{k-1} \\ &= p \sum_{k=1}^{\infty} \left[\frac{d}{dx} \left\{ \frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) (1-p)^{k-1} \right\} \right] \\ &= \frac{p}{\pi\sigma} \sum_{k=1}^{\infty} \frac{1}{k} \left[\frac{1}{1 + \left(\frac{x-k\mu}{k\sigma}\right)^2} \right] (1-p)^{k-1}. \end{aligned}$$

□

If $p = 1$, then $X \sim C(\mu, \sigma)$. When $\mu = 0$ and $\sigma = 1$, then the distribution of X is known as the standard geometric skew-Cauchy distribution denoted by $GSC(0, 1, p)$, with the pdf in (83) becoming

$$f_X(x) = \frac{p}{\pi} \sum_{k=1}^{\infty} \frac{1}{k} \left[\frac{1}{1 + \left(\frac{x}{k}\right)^2} \right] (1-p)^{k-1}. \tag{84}$$

The $GSC(0, 1, p)$ distribution is a symmetric distribution around 0 for all values of p. The symmetry follows from the fact that $f_X(x) = f_X(-x)$ for all x in (84). The latter can be observed in Figure 13 that illustrate symmetric PDFs around 0 for different values of p (with $\mu = 0$ and $\sigma = 1$).

It should be noted that no characteristics (expected value and variance) for the $GSC(\mu, \sigma, p)$ distribution will be presented in this dissertation. Since the MGF does not exist for the Cauchy distribution [40], proceeding to derive the MGF for a random variable Y with PDF given in (83) is not feasible. The latter results that some characteristics such as the expected value and variance do not exist since computation thereof require the MGF.

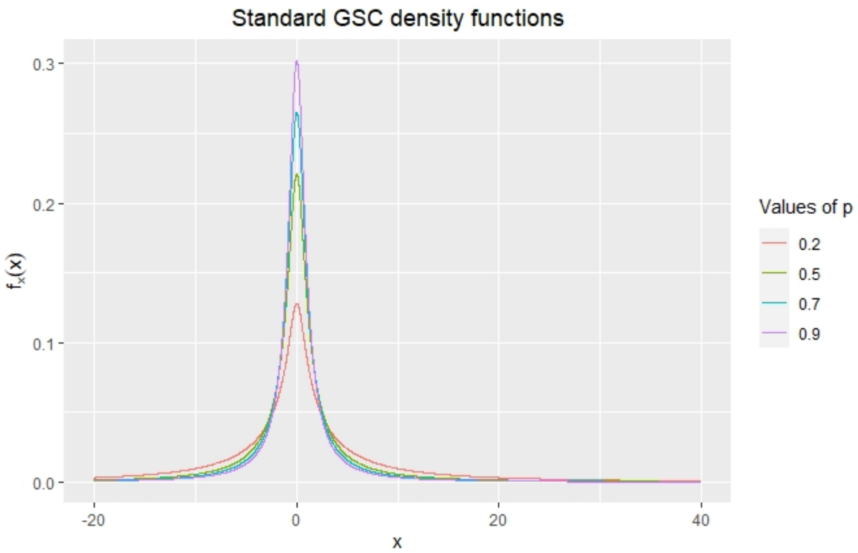


Figure 13: PDFs of the $GSC(0, 1, p)$ distribution with PDF in (84) for different values of p.

4.1.2 Generation of random numbers and illustration of PDF

In order to generate from the $GSC(\mu, \sigma, p)$ distribution with PDF given in (83) a short algorithm is provided.

Algorithm 7 Generation from the $GSC(\mu, \sigma, p)$ distribution

---

1: ***Required:***
- *Define the value of* μ *for* $\mu \in \mathbb{R}$.
- *Define the value of* σ *for* $\sigma > 0$.
- *Define the value of* p *for* $0 < p \leq 1$.

2: *Generate the value* n *from the* $GE(p)$ *distribution.*
3: *Generate the value* x *from* $C(n\mu, n\sigma)$*, where* n *is from step 2.*
4: *To obtain a required sample of size* j *repeat step 3,* j *times using the same* n *from step 2.*

Using the PDF given in (83), graphs of the PDFs of the $GSC(\mu, \sigma, p)$ distribution are provided in Figures 14, 15 and 16. These PDFs are plotted with $p = 0.5$ and $\sigma = 1$ kept constant throughout, varying the values of μ.

It can be seen from Figures 14, 15 and 16 that the PDFs of the $GSC(\mu, \sigma, p)$ distribution can take on various shapes depending on the parameter values for μ. The PDFs are positively skewed when $\mu > 0$ and negatively skewed when $\mu < 0$. It is also observed that the PDFs can be unimodal or multimodal, as the PDF in Figure 15 depicts a bimodal PDF when the value of $\mu = 4.5$ and the PDF in Figure 16 depicts a multimodal PDF when the value of $\mu = 8$. The unimodal PDFs are observed for the values of $\mu = -1$ and $\mu = 1$ in Figure 14. It is therefore determined that as $|\mu|$ increases, a multimodal PDF seems evident (keeping p and σ constant in this case).

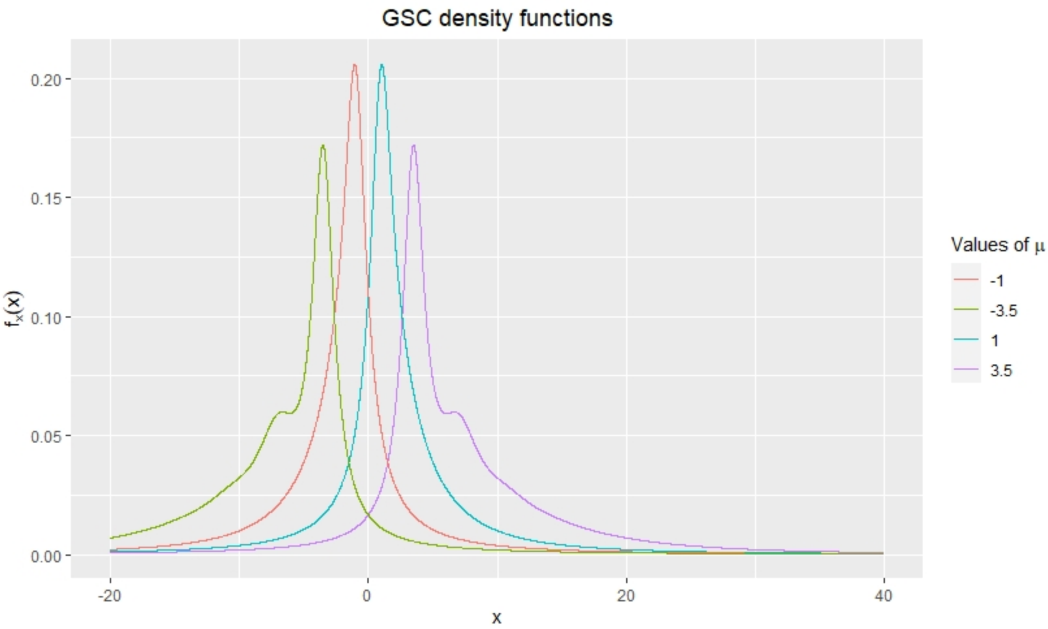


Figure 14: PDFs of the distribution with PDF in (83) for different values of μ.

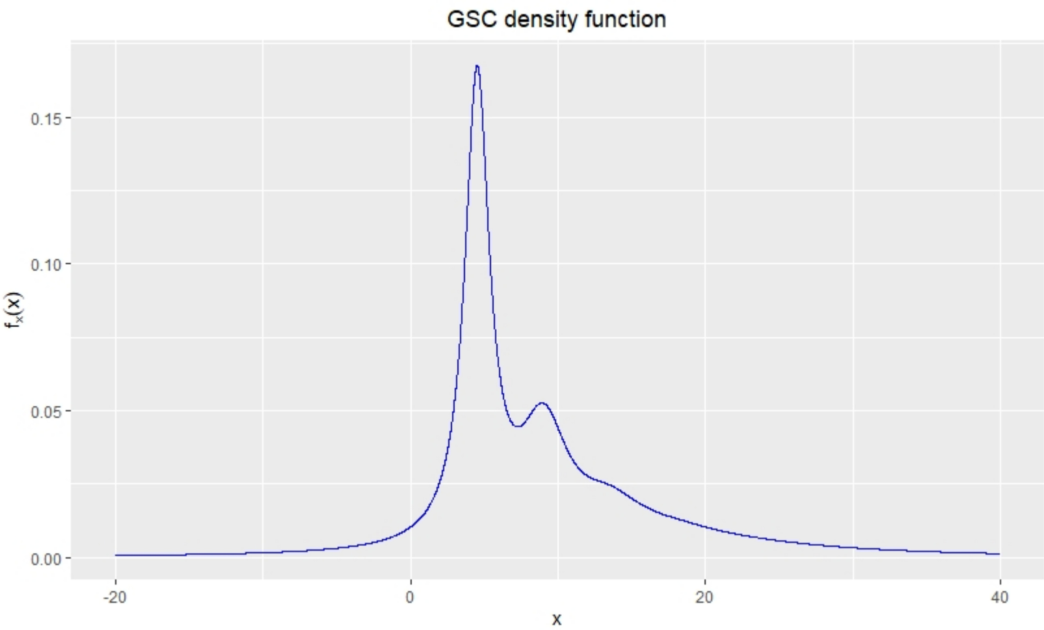


Figure 15: PDF of the distribution with PDF in (83) for the value of $\mu = 4.5$.

The observation of multimodality is different from the skew-normal distribution which is always unimodal. Therefore, the $GSC(\mu, \sigma, p)$ distribution is more flexible than the $SN\left(\mu, \sigma^2, \lambda\right)$ distribution and it can be used as an alternative to model skewed data.

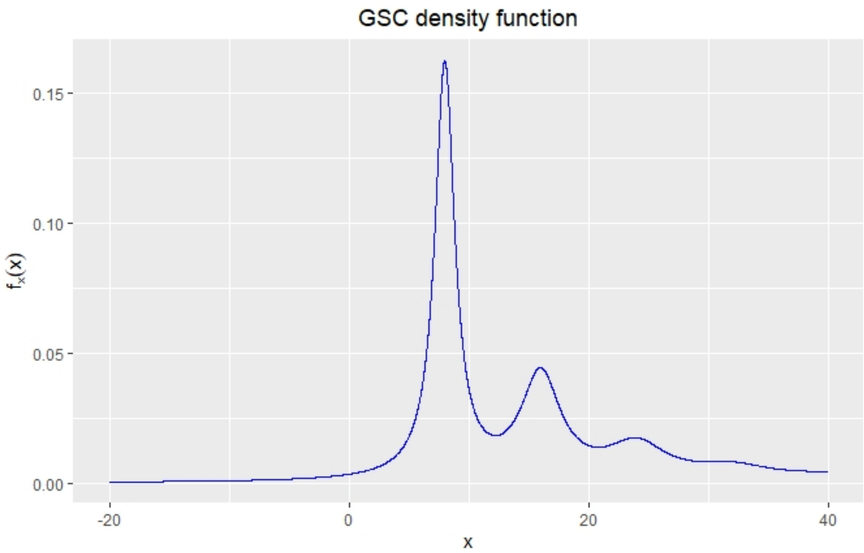


Figure 16: PDF of the distribution with PDF in (83) for the value of $\mu = 8$.

The comparison of the PDFs of the $GSC(\mu, \sigma, p)$ distribution against the the $GSN(\mu, \sigma, p)$ distribution is given in Figure 17. These PDFs are plotted with $p = 0.5$, $\sigma = 1$ and $\mu = 3.5$. The multimodality property coincides with the same property as observed in the $GSN(\mu, \sigma, p)$ distribution. However, as can be seen in Figure 18, it is evident that the $GSC(\mu, \sigma, p)$ distribution has fatter tails in both the upper- and lower-tails of the distribution when compared to the $GSN(\mu, \sigma, p)$ distribution.

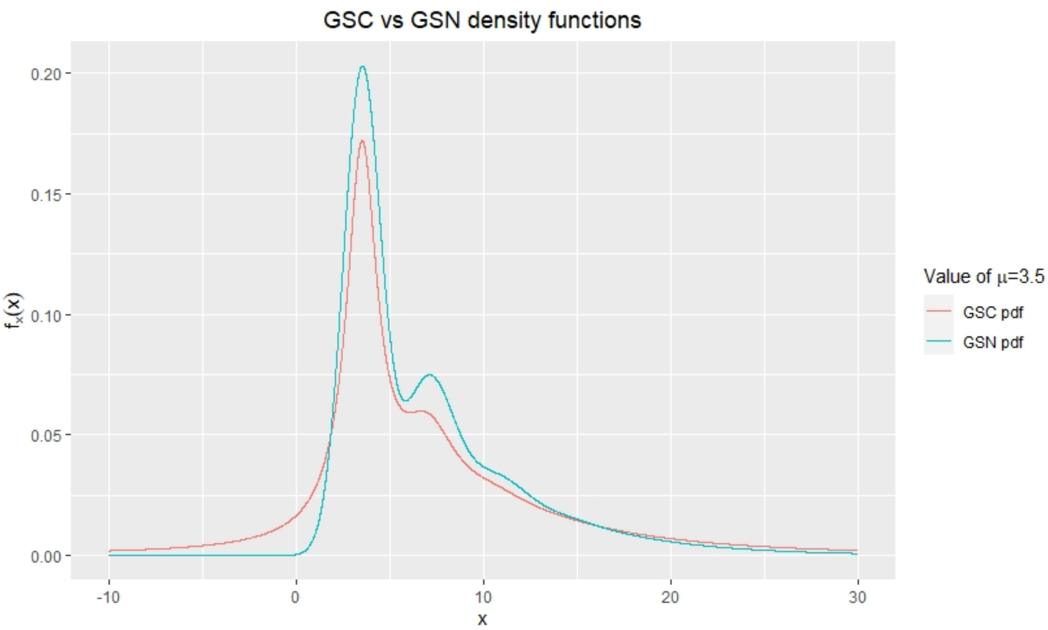


Figure 17: PDFs of (83) against (31) for the value of $\mu = 3.5$.

The latter can be also be observed in Tables 7 and 8, where the $GSC(\mu, \sigma, p)$ distribution has greater lower-tail probabilities and smaller upper-tail probabilities than the $GSN(\mu, \sigma, p)$ distribution. This is

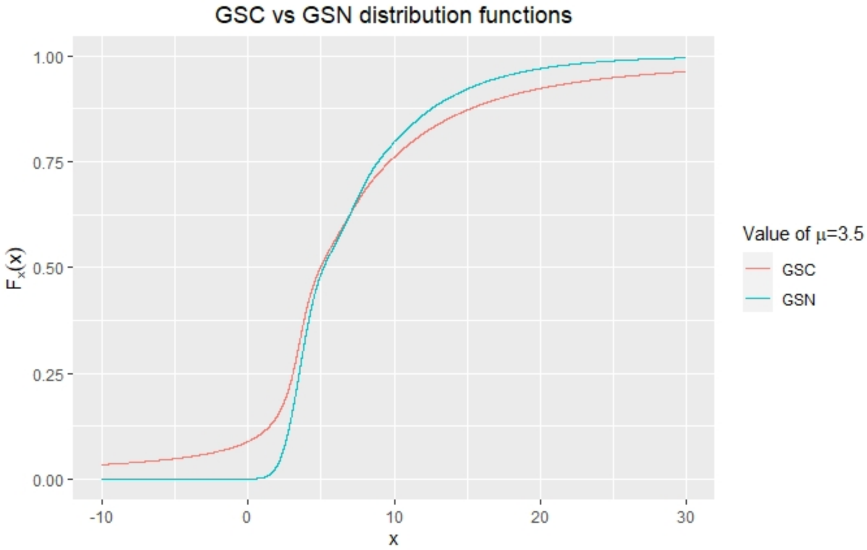


Figure 18: CDFs of (83) against (31) for the value of $\mu = 3.5$.

indicative of the ability of the $GSC(\mu, \sigma, p)$ distribution to model fatter tails. The probabilities in Tables 7 and 8 were calculated with the parameters set to $p = 0.5$, $\sigma = 1$ and $\mu = 3.5$.

Table 7: Lower tail probabilities of the $GSN(\mu, \sigma, p)$ vs. $GSC(\mu, \sigma, p)$ distribution

	$GSN(\mu, \sigma, p)$	$GSC(\mu, \sigma, p)$
$x = -3$	< 0.001	0.058
$x = -2.5$	< 0.001	0.062
$x = -2$	< 0.001	0.065
$x = -1.5$	< 0.001	0.070
$x = -1$	< 0.001	0.075
$x = -0.5$	< 0.001	0.081
$x = 0$	< 0.001	0.089
$x = 0.5$	< 0.001	0.098
$x = 1$	0.003	0.110
$x = 1.5$	0.011	0.126

Table 8: Upper tail probabilities of the $GSN(\mu, \sigma, p)$ vs. $GSC(\mu, \sigma, p)$ distribution

	$GSN(\mu, \sigma, p)$	$GSC(\mu, \sigma, p)$
$x = 20$	0.970	0.923
$x = 21$	0.975	0.930
$x = 22$	0.980	0.936
$x = 23$	0.983	0.941
$x = 24$	0.986	0.945
$x = 25$	0.989	0.950
$x = 26$	0.991	0.952
$x = 27$	0.992	0.956
$x = 28$	0.994	0.958
$x = 29$	0.995	0.961

It is also observed that as the value of p increases the distribution exhibits less fatter tails (keeping the values of μ and σ constant). The latter can be observed in Figure 19.

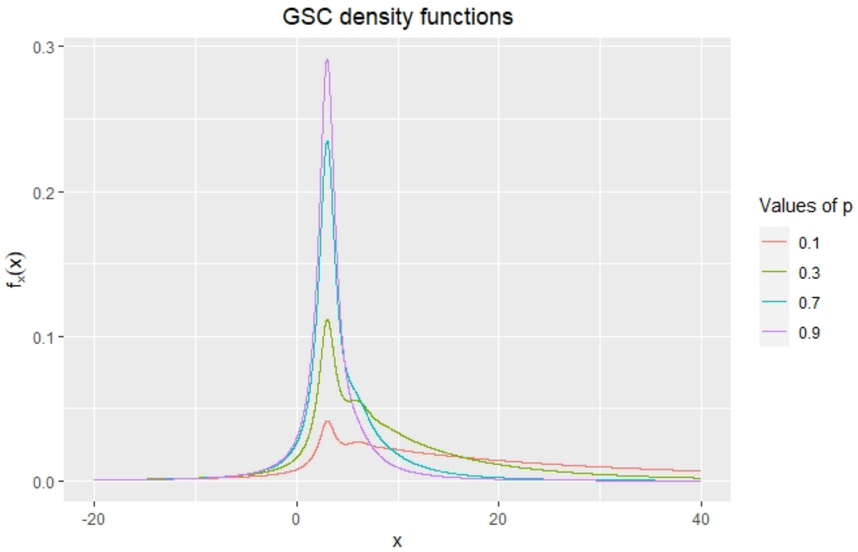


Figure 19: PDF of the distribution with PDF in (83) for different values of p.

It is therefore concluded that the $GSC(\mu, \sigma, p)$ distribution can be used as an alternative to the $GSN(\mu, \sigma, p)$ distribution to model skewed data. It seems that the $GSC(\mu, \sigma, p)$ distribution is a reasonably more flexible distribution than the $GSN(\mu, \sigma, p)$ distribution when it comes to modelling fat tails. Hence, the $GSC(\mu, \sigma, p)$ distribution can model data that is *skewed, multimodal and exhibits fatter tails.*

4.1.3 Estimation

Conditional properties

Before continuing with the estimation section, it is worthwhile to first furnish different conditional properties that will be utilized in the estimation theory. These conditional properties provide further detail into the distribution and contribute knowledge on specific conditions enclosed on the distribution.

Theorem 46. *Consider (X, N) which has the joint PDF as given by (80), and let $m \leq n$ be positive integers. The conditional CDF of (X, N) given $N \leq n$ is given by*

$$P\left(X \leq x, N \leq m | N \leq n\right) = \frac{p}{1-(1-p)^{n-1}} \sum_{k=1}^{m} \left[\frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1} \tag{85}$$

for $-\infty < x < \infty$ and $0 < p < 1$.

Proof. From (112), (81), the CDF of a $GE(p)$ distribution and the fact that $m \leq n$, it follows that

$$\begin{aligned} P(X \leq x, N \leq m | N \leq n) &= \frac{P(X \leq x, N \leq m)}{P(N \leq n)} \\ &= \frac{p \sum_{k=1}^{m} \left[\frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}}{P(N \leq n)} \\ &= \frac{p \sum_{k=1}^{m} \left[\frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}}{1-(1-p)^n} \\ &= \frac{p}{1-(1-p)^n} \sum_{k=1}^{m} \left[\frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}. \end{aligned}$$

□

It follows directly from (85), that for $m = n$ and using (81), it is true that

$$\begin{aligned} P(X \leq x, N \leq n | N \leq n) &= \frac{P(X \leq x, N \leq n)}{P(N \leq n)} \\ &= \frac{p}{1-(1-p)^n} \sum_{k=1}^{n} \left[\frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}. \end{aligned} \tag{86}$$

Theorem 47. *Consider (X, N) which has the joint PDF as given by (80). Suppose that $0 \leq x \leq y$ and n is any positive integer. The conditional CDF of (X, N) given $X \leq y$ is given by*

$$P(X \leq x, N \leq n | X \leq y) = \frac{\sum_{k=1}^{n} \left[\frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}}{\sum_{k=1}^{\infty} \left[\frac{1}{\pi} \arctan\left(\frac{y-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}}. \tag{87}$$

Proof. Using (85), (112) in Appendix A.1 and since $x \leq y$, it follows that

$$\begin{aligned} P(X \leq x, N \leq n | X \leq y) &= \frac{P(X \leq x, N \leq n)}{P(X \leq y)} \\ &= \frac{p \sum_{k=1}^{n} \left[\frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}}{p \sum_{k=1}^{\infty} \left[\frac{1}{\pi} \arctan\left(\frac{y-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}} \\ &= \frac{\sum_{k=1}^{n} \left[\frac{1}{\pi} \arctan\left(\frac{x-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}}{\sum_{k=1}^{\infty} \left[\frac{1}{\pi} \arctan\left(\frac{y-k\mu}{k\sigma}\right) + \frac{1}{2}\right] (1-p)^{k-1}}. \end{aligned}$$

□

It follows directly from (87) that

$$
\begin{aligned}
P(N \le n, X \le y | X \le y) &= P(N \le n | X \le y) \\
&= \frac{P(N \le n, X \le y)}{P(X \le y)} \\
&= \frac{\sum_{k=1}^{n} \left[\frac{1}{\pi}\arctan\left(\frac{y-k\mu}{k\sigma}\right) + \frac{1}{2}\right](1-p)^{k-1}}{\sum_{k=1}^{\infty} \left[\frac{1}{\pi}\arctan\left(\frac{y-k\mu}{k\sigma}\right) + \frac{1}{2}\right](1-p)^{k-1}}
\end{aligned} \tag{88}
$$

for n as any positive integer. Using (80) and (83), the conditional PMF of N given $X = x$ is

$$
\begin{aligned}
f_{N|X}(n|x) &= \frac{f_{N,X}(n,x)}{f_X(x)} \\
&= \frac{\frac{1}{\pi n\sigma}\left[\frac{1}{1+\left(\frac{x-n\mu}{n\sigma}\right)^2}\right](1-p)^{n-1}}{\frac{1}{\pi\sigma}\sum_{k=1}^{\infty}\frac{1}{k}\left[\frac{1}{1+\left(\frac{x-k\mu}{k\sigma}\right)^2}\right](1-p)^{k-1}} \\
&= \frac{(1-p)^{n-1}\left[\frac{1}{1+\left(\frac{x-n\mu}{n\sigma}\right)^2}\right]/n}{\sum_{k=1}^{\infty}(1-p)^{k-1}\left[\frac{1}{1+\left(\frac{x-k\mu}{k\sigma}\right)^2}\right]/k}
\end{aligned} \tag{89}
$$

and in turn the conditional expectation of N given $X = x$ becomes

$$
\begin{aligned}
E(N|X=x) &= \sum_{n=1}^{\infty} nP(N=n|X=x) \\
&= \frac{\sum_{n=1}^{\infty}(1-p)^{n-1}\left[\frac{1}{1+\left(\frac{x-n\mu}{n\sigma}\right)^2}\right]}{\sum_{k=1}^{\infty}(1-p)^{k-1}\left[\frac{1}{1+\left(\frac{x-k\mu}{k\sigma}\right)^2}\right]/k}
\end{aligned} \tag{90}
$$

with

$$
\begin{aligned}
E\left(N^{-1}|X=x\right) &= \sum_{n=1}^{\infty} n^{-1}P(N=n|X=x) \\
&= \frac{\sum_{n=1}^{\infty}(1-p)^{n-1}\left[\frac{1}{1+\left(\frac{x-n\mu}{n\sigma}\right)^2}\right]/n^2}{\sum_{k=1}^{\infty}(1-p)^{k-1}\left[\frac{1}{1+\left(\frac{x-k\mu}{k\sigma}\right)^2}\right]/k}.
\end{aligned} \tag{91}
$$

Maximum likelihood estimators

In order to obtain the MLEs of the unknown parameters it is necessary to maximize the log-likelihood function with respect to the unknown parameters.

Theorem 48. *Suppose that $\{x_1, x_2, ..., x_n\}$ is a sample of size n from the $GSC(\mu, \sigma, p)$ distribution with PDF as given in (83). The log-likelihood function is given by*

$$l(\mu, \sigma, p) = \sum_{i=1}^{n} \ln \left[\sum_{k=1}^{\infty} \frac{p}{\pi \sigma k} \left[\frac{1}{1 + \left(\frac{x - k\mu}{k\sigma}\right)^2} \right] (1-p)^{k-1} \right]. \tag{92}$$

Proof. From Bain and Engelhardt [8] and using (83), it follows that

$$\begin{aligned} l(\mu, \sigma, p) &= \ln \left(\prod_{i=1}^{n} f_{X_i}(x_i) \right) \\ &= \sum_{i=1}^{n} ln \left[f_{X_i}(x_i) \right] \\ &= \sum_{i=1}^{n} \ln \left[\sum_{k=1}^{\infty} \frac{p}{\pi \sigma k} \left[\frac{1}{1 + \left(\frac{x - k\mu}{k\sigma}\right)^2} \right] (1-p)^{k-1} \right]. \end{aligned}$$

□

It is proceeded to obtain the normal equations by taking the partial derivatives of the log-likelihood function given in (92) and equating them to 0. The partial derivatives are with respect to μ, σ and p. In order to circumvent the problem of having to solve three non-linear equations, an iterative method will be used to find the MLEs. For this purpose, the log-likelihood based on the complete sample will be used, as an explicit expression for the MLE of p can be obtained. For the remaining parameters, the non-linear equations need to be solved by using the Newton-Raphson (NR) method.

Theorem 49. *Let $\{(x_1, m_1), (x_2, m_2), ..., (x_n, m_n)\}$ be a random sample of size n from the joint distribution of (X, N) in (80). The log-likelihood based on the complete sample is given by*

$$l_c(\mu, \sigma, p) \propto n \ln p - n \ln \sigma + \ln(1-p) \sum_{i=1}^{n} (m_i - 1) - \sum_{i=1}^{n} \ln \left(1 + \left(\frac{x_i - m_i \mu}{m_i \sigma} \right)^2 \right) \tag{93}$$

without the additive constant and the MLE of p is obtained as

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} m_i}. \tag{94}$$

Proof. Using (80), it follows that

$$\begin{aligned}
l_c(\mu,\sigma,p) &= \ln\left(\prod_{i=1}^{n} f_{X,N}(x_i,m_i)\right)\\
&= \sum_{i=1}^{n} \ln\left[f_{X,N}(x_i,m_i)\right]\\
&= \sum_{i=1}^{n} \ln\left\{\frac{p}{\pi\sigma m_i}\left[\frac{1}{1+\left(\frac{x_i-m_i\mu}{m_i\sigma}\right)^2}\right](1-p)^{m_i-1}\right\}\\
&= \sum_{i=1}^{n}\left\{\ln\left(\frac{p}{\pi\sigma m_i}\right)+(m_i-1)\ln(1-p)+\ln(1)-\ln\left[1+\left(\frac{x_i-m_i\mu}{m_i\sigma}\right)^2\right]\right\}\\
&\propto n\ln p - n\ln\sigma + \ln(1-p)\sum_{i=1}^{n}(m_i-1) - \sum_{i=1}^{n}\ln\left[1+\left(\frac{x_i-m_i\mu}{m_i\sigma}\right)^2\right].
\end{aligned}$$

It is then proceeded to obtain the normal equation for p based on the complete sample. This is done by taking the partial derivative of the complete log-likelihood with respect to p and setting it equal to 0, followed by solving for the unknown parameter. Thus, the MLE for p is obtained as

$$\begin{aligned}
\frac{\partial l_c(\mu,\sigma,p)}{\partial p} &= 0\\
\frac{-\sum_{i=1}^{n}(m_i-1)}{1-\hat{p}}+\frac{n}{\hat{p}} &= 0\\
-\hat{p}\sum_{i=1}^{n} m_i + n\hat{p} + n - n\hat{p} &= 0\\
-\hat{p}\sum_{i=1}^{n} m_i &= -n\\
\hat{p} &= \frac{n}{\sum_{i=1}^{n} m_i}.
\end{aligned}$$

□

The log-likelihood in (93) directly results that the MLE of p can be obtained in an explicit form. Using (93), the normal equations of the remaining unknown parameters can be written as

$$\begin{aligned}
\frac{\partial l_c(\mu,\sigma,p)}{\partial \mu} &= 0\\
\sum_{i=1}^{n}\left\{\left[\frac{-1}{1+\left(\frac{x_i-m_i\hat{\mu}}{m_i\sigma}\right)^2}\right]\left[2\left(\frac{x_i-m_i\hat{\mu}}{m_i\sigma}\right)\right]\left[\frac{-1}{\sigma}\right]\right\} &= 0\\
\sum_{i=1}^{n}\left\{\frac{\frac{2}{\sigma}\left(\frac{x_i-m_i\hat{\mu}}{m_i\sigma}\right)}{1+\left(\frac{x_i-m_i\hat{\mu}}{m_i\sigma}\right)^2}\right\} &= 0 \qquad (95)
\end{aligned}$$

and

$$\frac{\partial l_c(\mu,\sigma,p)}{\partial\sigma}=0$$

$$\frac{-n}{\hat{\sigma}}-\sum_{i=1}^{n}\left\{\frac{1}{1+\left(\frac{x_i-m_i\hat{\mu}}{m_i\hat{\sigma}}\right)^2}\left[2\left(\frac{x_i-m_i\hat{\mu}}{m_i\hat{\sigma}}\right)\right]\left[-\frac{x_i-m_i\hat{\mu}}{m_i\hat{\sigma^2}}\right]\right\}=0$$

$$\frac{-n}{\hat{\sigma}}-\sum_{i=1}^{n}\left\{\frac{-2\left(\frac{x_i-m_i\hat{\mu}}{m_i\hat{\sigma}}\right)^2}{\hat{\sigma}\left[1+\left(\frac{x_i-m_i\hat{\mu}}{m_i\hat{\sigma}}\right)^2\right]}\right\}=0 \tag{96}$$

respectively. From (95) and (96) is it observed that explicit expressions for $\hat{\mu}$ and $\hat{\sigma}$ are not possible, hence the necessity to solve these using the NR method. Furthermore, from (95) and (96) it is observed that these expressions contain the term m_i. Since the explicit expression of $\hat{p}$ contains the term m_i as well, it yields that the estimate for p will update with each iterative step that is run within the estimation algorithm to obtain the MLEs for μ and σ. Before providing the algorithm to estimate the unknown parameters, it is proceeded to first discuss the NR algorithm specifically for use in the estimation of parameters of the geometric skew-Cauchy distribution. A more general discussion on the NR method and its use to solve a non-linear system of equations can be viewed in the excellent monograph by Press and Vetterling [34].

Newton-Raphson (NR) algorithm

Derivatives of the complete log-likelihood sample given in (93) will be used to obtain the MLEs for μ and σ. These derivatives can be constituted in a vector of length two, $\mathbf{F}(\mu,\sigma)$, with the components of the vector being functions of μ and σ:

$$\mathbf{F}(\mu,\sigma)=\begin{bmatrix}\frac{\partial l_c(\mu,\sigma,p)}{\partial\mu}\\ \frac{\partial l_c(\mu,\sigma,p)}{\partial\sigma}\end{bmatrix}$$

$$=\begin{bmatrix}\sum_{i=1}^{n}\left\{\frac{2\left(\frac{x_i-m_i\mu}{m_i\sigma}\right)}{\sigma\left[1+\left(\frac{x_i-m_i\mu}{m_i\sigma}\right)^2\right]}\right\}\\ \frac{-n}{\sigma}-\sum_{i=1}^{n}\left\{\frac{\frac{-2}{\sigma}\left(\frac{x_i-m_i\mu}{m_i\sigma}\right)^2}{1+\left(\frac{x_i-m_i\mu}{m_i\sigma}\right)^2}\right\}\end{bmatrix}. \tag{97}$$

It is then necessary to solve the system of equations $\mathbf{F}(\mu,\sigma) = \mathbf{0}$. Consider the Jacobian matrix $\mathbf{J}(\mu,\sigma)$ of the vector equation $\mathbf{F}$:

$$\mathbf{J}(\mu,\sigma) = \begin{bmatrix} \frac{\partial^2 l_c(\mu,\sigma,p)}{\partial\mu^2} & \frac{\partial^2 l_c(\mu,\sigma,p)}{\partial\mu\partial\sigma} \\ \frac{\partial^2 l_c(\mu,\sigma,p)}{\partial\sigma\partial\mu} & \frac{\partial^2 l_c(\mu,\sigma,p)}{\partial\sigma^2} \end{bmatrix}. \tag{98}$$

After conducting some algebraic calculations, the elements of the Jacobian matrix are obtained as:

$$\frac{\partial^2 l_c(\mu,\sigma,p)}{\partial\mu^2} = \sum_{i=1}^{n} \left\{ \frac{-\frac{2}{\sigma^2} + \frac{2}{\sigma^2}\left(\frac{x_i - m_i\mu}{m_i\sigma}\right)^2}{\left[1 + \left(\frac{x_i - m_i\sigma}{m_i\sigma}\right)^2\right]^2} \right\}, \tag{99}$$

$$\frac{\partial^2 l_c(\mu,\sigma,p)}{\partial\mu\partial\sigma} = \sum_{i=1}^{n} \left\{ \frac{-\frac{4}{\sigma^2}\left(\frac{x_i - m_i\mu}{m_i\sigma}\right)}{\left[1 + \left(\frac{x_i - m_i\sigma}{m_i\sigma}\right)^2\right]^2} \right\}, \tag{100}$$

$$\frac{\partial^2 l_c(\mu,\sigma,p)}{\partial\sigma\partial\mu} = -\sum_{i=1}^{n} \left\{ \frac{\frac{4}{\sigma^2}\left(\frac{x_i - m_i\mu}{m_i\sigma}\right)}{\left[1 + \left(\frac{x_i - m_i\sigma}{m_i\sigma}\right)^2\right]^2} \right\}, \tag{101}$$

$$\frac{\partial^2 l_c(\mu,\sigma,p)}{\partial\sigma^2} = \frac{n}{\sigma^2} - \sum_{i=1}^{n} \left\{ \frac{\frac{6}{\sigma^2}\left(\frac{x_i - m_i\mu}{m_i\sigma}\right)^2 + \frac{2}{\sigma^2}\left(\frac{x_i - m_i\mu}{m_i\sigma}\right)^4}{\left[1 + \left(\frac{x_i - m_i\sigma}{m_i\sigma}\right)^2\right]^2} \right\}. \tag{102}$$

The following iterative procedure represents the NR method for non-linear systems [9]:

$$\begin{bmatrix} \mu^{(n)} \\ \sigma^{(n)} \end{bmatrix} = \begin{bmatrix} \mu^{(n-1)} \\ \sigma^{(n-1)} \end{bmatrix} - \mathbf{J}^{-1} \begin{bmatrix} \mu^{(n-1)} \\ \sigma^{(n-1)} \end{bmatrix} \mathbf{F} \begin{bmatrix} \mu^{(n-1)} \\ \sigma^{(n-1)} \end{bmatrix} \tag{103}$$

for $n \geq 1$, where the initial values for the parameters μ^0 and σ^0 will be given from the values obtained via a grid search for parameter optimization, and $\mathbf{J}^{-1}$ is the inverse of the Jacobian matrix in (98). The iterative method will continue until a given tolerance error, say ϵ, is reached between the n^{th} and $(n+1)^{th}$ iterative values.

The grid search for parameter optimization is summarised in Algorithm 8.

Algorithm 8 Simple grid search for parameter optimization

1: ***Required:***
   - *Define the step size.*
   - *Define the lower and upper bound for μ. The minimum and maximum value of the sample at hand will suffice.*
   - *Define a sequence of values for μ using the step size and the predefined minimum and maximum values.*
   - *Define a sequence of values for σ using the step size and a minimum value of greater than 0, with a maximum value taken as the sample standard deviation.*
   - *Define a sequence of values for p using the step size and a minimum value of greater than 0, with a maximum less than or equal to 1.*

2: *For a given p, calculate a matrix than contains the log-likelihood values using all possible combinations of μ and σ.*
3: *Find the maximum log-likelihood value in the matrix in step 2 and save it into a new matrix. This will yield the μ and σ combination that gives the maximum log-likelihood for a given p.*
4: *Repeat steps 2 and 3, for all possible values of p.*
5: *Find the maximum of the matrix in step 4. This will yield the optimimum combination of parameters.*

The grid search is constructed in such a manner that it conducts an exhaustive search across the entire grid, rather than searching until it finds a local maximum. The values obtained from the grid search are then used as the starting values for the NR method and in turn the estimation algorithm.

The algorithm that yields the estimates for μ, σ and p can subsequently be implemented and is summarised in Algorithm 9.

Algorithm 9 Algorithm to obtain the MLEs of the $GSC(\mu, \sigma, p)$ distribution.

1: ***Required:*** *Run the grid search in algorithm (8) to obtain initial values for μ, σ and p.*
2: *Obtain the conditional expectations $c_i = \frac{1}{m_i}$ and $d_i = m_i$ by using (91) and (90) respectively as well as the parameter estimates at the current stage of the algorithm.*
3: *Estimate the value of p by using (94) as well as d_i from step 2.*
4: *Obtain the elements of the vector* $\mathbf{F}$ *in (97) and the Jacobian matrix* $\mathbf{J}$ *in (98) by replacing the missing values with their expectations.*
5: *Step 4 yields the following:*

$$\mathbf{F}(\mu,\sigma)=\begin{bmatrix} \sum_{i=1}^{n}\left\{\frac{\frac{2}{\sigma}(x_ic_i-\mu)}{1+\frac{1}{\sigma^2}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)}\right\} \\ \frac{-n}{\sigma}-\sum_{i=1}^{n}\left\{\frac{\frac{-2}{\sigma^3}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)}{1+\frac{1}{\sigma^2}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)}\right\}\end{bmatrix}$$

and

$$\mathbf{J}\left(\mu,\sigma\right)=\begin{bmatrix} \sum_{i=1}^{n}\left\{\frac{-\frac{2}{\sigma^2}+\frac{2}{\sigma^4}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)}{\left[1+\frac{1}{\sigma^2}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)\right]^2}\right\} & \sum_{i=1}^{n}\left\{\frac{-\frac{4}{\sigma^3}(x_ic_i-\mu)}{\left[1+\frac{1}{\sigma^2}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)\right]^2}\right\} \\ -\sum_{i=1}^{n}\left\{\frac{\frac{4}{\sigma^3}(x_ic_i-\mu)}{\left[1+\frac{1}{\sigma^2}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)\right]^2}\right\} & \frac{n}{\sigma^2}-\sum_{i=1}^{n}\left\{\frac{\frac{6}{\sigma^4}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)+\frac{2}{\sigma^2}\left[\frac{1}{\sigma^2}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)\right]^2}{\left[1+\frac{1}{\sigma^2}\left(x_i^2c_i^2-2x_ic_i\mu+\mu^2\right)\right]^2}\right\}\end{bmatrix}$$

where $c_i = \frac{1}{m_i}$ from step 2.
6: *Run the iterative procedure in (103) to obtain the estimates for μ and σ.*
7: *Repeat steps 2 - 6 until a tolerance error ϵ is obtained.*

5 Chapter 5

This chapter will proceed to present a simulation study and conduct analysis on two real data sets. This will be done to assess the efficacy of the newly proposed geometric skew-Cauchy distribution as an alternative to the skew-normal and geometric skew-normal distributions.

The first data set will be analysed using the different estimation algorithms presented in previous chapters for the $GSN(\mu, \sigma, p)$ model and the $GSC(\mu, \sigma, p)$ model. The *sn* package in R will be used to estimate the $SN(\mu, \sigma^2, \lambda)$ model. The second data set will be analysed using the *DEoptim* package in R for all three models considered in previous chapters [1]. It is noted that the estimation algorithms as well as the use of the *DEoptim* package yield the same results for parameter estimates.

Throughout this chapter, in order to asses which model is a better fit, the log-likelihood (higher is better), the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the Kolmogorov-Smirnov (KS) test statistic values will be calculated.

The KS test statistic is the distance between the empirical and fitted CDFs, for which a lower value is better [16]. The KS test statistic, T_1, is suggested by Kolmogorov as:

$$T_1 = \max_{-\infty < x < +\infty} \mid S(x) - F(x) \mid$$

where $S(x)$ is the empirical CDF and $F(x)$ is the fitted CDF [11]. The KS test statistic is used in the Kolmogorov-Smirnov test to determine whether the sample that yields the CDF, $S(x)$, emanates from the population that yields the CDF, $F(x)$. In essence, a smaller T_1 indicates that $S(x)$ and $F(x)$ are closer to each other, which in turn indicates that the sample values are more likely to emanate from the population values. Further details on the Kolmogorov-Smirnov test can be viewed in the excellent chapter on non-parametric modelling by Guidici [16].

The AIC approximates the quantity of information lost by a given model that was fitted to the data. Hence, a lower AIC value is preferred [1]. The AIC value can be used as a method for model selection and it is calculated as follows:

$$AIC = 2k - 2\ln(L)$$

where k is the number of estimated parameters in the model and L is the maximum value of the likelihood function of the model [1]. In essence, the AIC approximates the quality of each model in relation to the quality of other models, given a multitude of models for the data at hand [1]. The AIC handles the trade-off between the simplicity of the model and the goodness-of-fit of the model by introducing a penalty

[1] Full explanation on the *DEoptim* package can be viewed on https://cran.r-project.org/web/packages/DEoptim/DEoptim.pdf.

term for the number of parameters in the model [1]. On the other hand, the BIC also yields a method for model selection, where a model is preferred with a lower BIC value [35]. The BIC value is closely connected to the AIC value since it also handles the trade-off between simplicity and goodness-of-fit of the model. The difference, however, is the fact that the BIC introduces a larger penalty term for the number of parameters in the model [38]. In essence, the BIC penalizes an increase in parameters more severely than the AIC. The BIC value is calculated as follows:

$$BIC = k\ln(n) - 2\ln(L)$$

where k is the number of estimated parameters in the model, n is the sample size and L is the maximum value of the likelihood function of the model [35].

5.1 Simulation

This subsection will proceed to simulate a data set from the $GSC(\mu, \sigma, p)$ distribution with PDF in (83). It has been decided to simulate the data set with a sample size of $j = 5000$ using Algorithm 7 as suggested in Section 4.1.2. The simulation was conducted with the following specification:

$$\mu = 2, \quad \sigma = 1, \quad p = 0.7.$$

After running the grid search in Algorithm 8, the starting values for the estimation algorithm were obtained as:

$$\mu = 2.15, \quad \sigma = 0.84, \quad p = 0.8$$

and proceeding to run the estimation algorithm given in Algorithm 9, the MLEs of the unknown parameters were obtained as:

$$\hat{\mu} = 2.012, \quad \hat{\sigma} = 1.021, \quad \hat{p} = 0.980$$

with the associated log-likelihood value becoming -12930.05. The fitted PDF is plotted against the histogram of the simulated data in Figure 20.

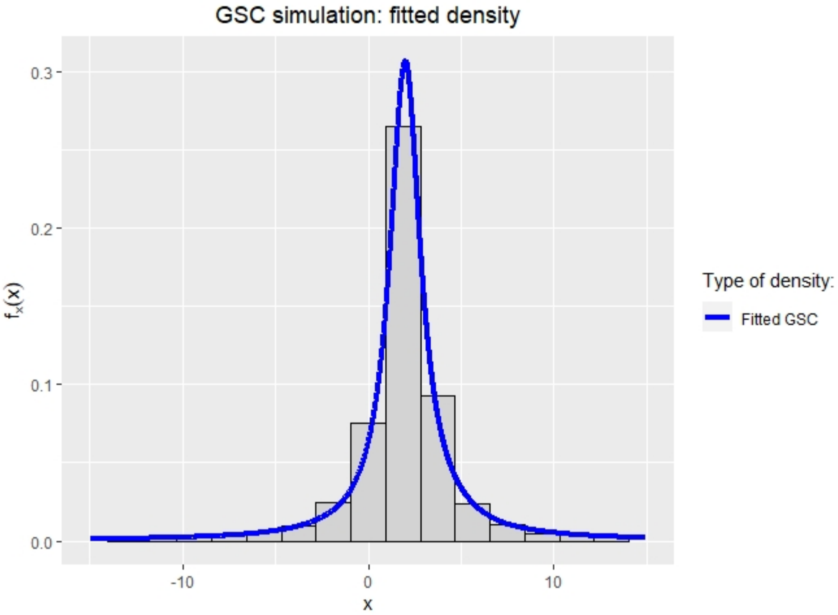


Figure 20: Histogram of simulated data with fitted $GSC(\mu, \sigma, p)$ PDF.

The estimates are given against the true parameters in Table 9 for comparison.

Table 9: Simulated data estimates comparison

Parameter	True parameters	$GSC(\mu, \sigma, p)$ estimates
μ	2	2.012
σ	1	1.021
p	0.7	0.980

It is observed from Table 9 that the estimated parameters for μ and σ are close to the true parameters used in the simulation. Using all the estimated parameters to plot the fitted PDF, it is observed from Figure 20 that the fitted PDF fits the data well indicating that the estimation algorithm performs satisfactory on a sample size of 5000 observations. It is noted, that other sample sizes were also considered. These include sample sizes of $j = 100, 200, 500, 1000, 10000$ respectively. The estimation algorithm also yielded satisfactory performance on the other sample sizes.

However, it is noted from Table 9 that the estimate for p is not close to the true parameter. The fact that the parameter estimate of p does not converge to the true parameter prompts further investigation into why this estimated result was obtained for simulated data. The following was done to investigate the behaviour of p:

- The impact of p on the log-likelihood was determined. This can be seen in Figure 21. It is clear that the log-likelihood value increases as the value of p increases. This was conducted for $\mu = 2$ and $\sigma = 1$ kept constant, using a sample size of $j = 5000$ as in the above simulation.

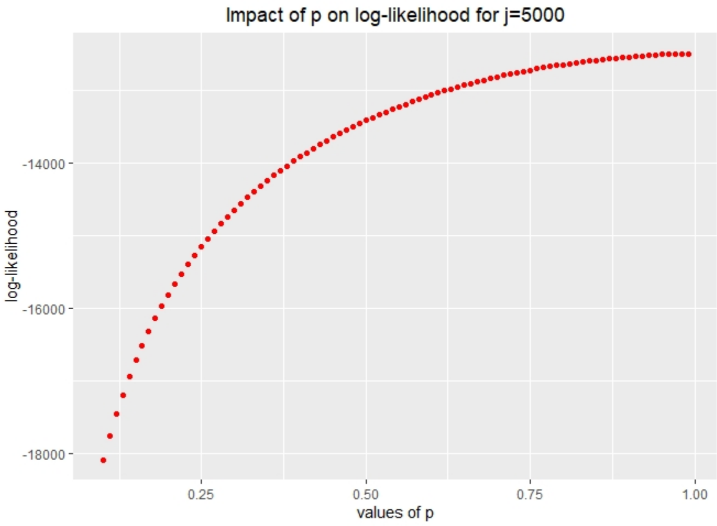
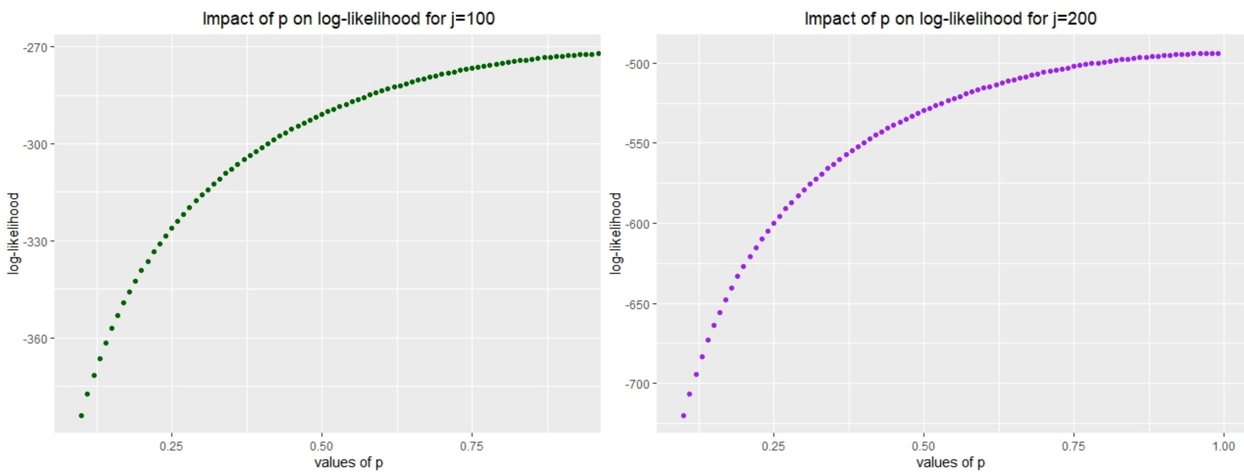


Figure 21: Change in the log-likelihood in (92) for different p for $j = 5000$.

- The process was repeated for other sample sizes as well, keeping $\mu = 2$ and $\sigma = 1$ constant. The sample sizes of $j = 100, 200, 500, 1000$ were used respectively. The impact on the log-likelihood can be seen in Figures 22 and 23. It is noted that a smaller sample size yields a higher log-likelihood value, given a higher p. Furthermore, the process was also repeated for other values of μ and σ kept constant, whilst varying the value of p. The same results depicted in Figures 22 and 23 are observed.

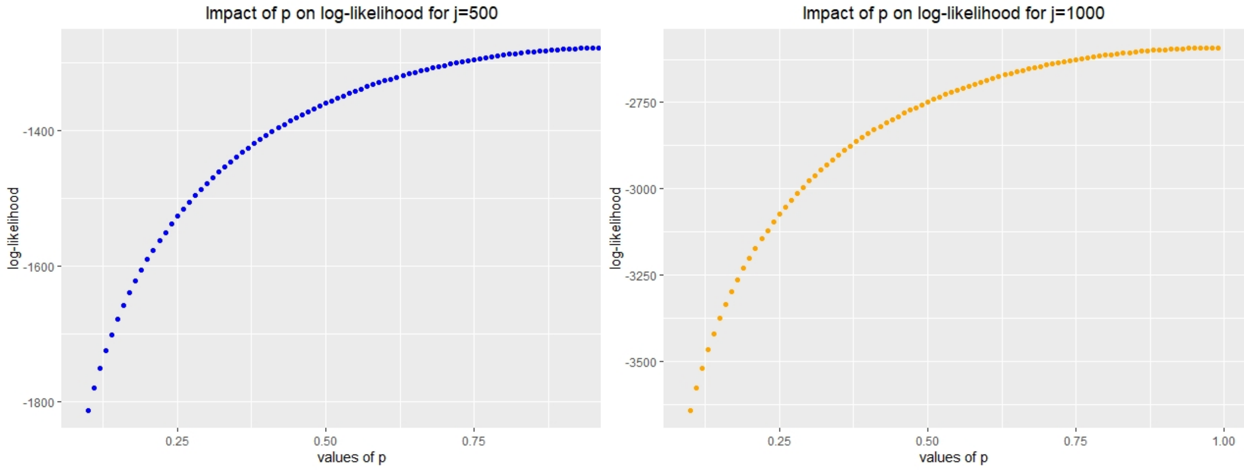


(a) Sample size $j = 100$.

(b) Sample size $j = 200$.

Figure 22: Change in the log-likelihood in (92) for different p for $j = 100, 200$.

- The observation of an increasing log-likelihood for different p is also noted from investigation of the simulation of the $GSN(\mu, \sigma, p)$ distribution. Thus, the log-likelihood also increases for an increasing p (keeping μ and σ constant).



(a) Sample size $j = 500$.

(b) Sample size $j = 1000$.

Figure 23: Change in the log-likelihood in (92) for different p for $j = 500, 1000$.

- Recall that if $N \sim GE(p)$ and $X \sim GSC(\mu, \sigma, p)$ with composition in (79), then $X|(N = n) = \sum_{i=1}^{n} Y_i \sim C(n\mu, n\sigma)$ using (78). Naturally, then $X \stackrel{d}{=} n\mu + n\sigma Z$, where $Z \sim C(0, 1)$. The latter stochastic representation was also used to simulate random numbers from the $GSC(\mu, \sigma, p)$ distribution, with $\mu = 2, \sigma = 1$ and $p = 0.7$. Proceeding with estimation, the estimate obtained for p again differs significantly from the true parameter, with p converging to a value close to 1. The impact on the log-likelihood was investigated and it increases for an increasing p.

- Following the approach suggested in Walck [40] to simulate random numbers from the Cauchy distribution, the stochastic representation of $X \stackrel{d}{=} n\mu + n\sigma(\frac{u}{v})$, where $u \sim N(0, 1), v \sim N(0, 1)$ and $(N = n) \sim GE(p)$, was also used to simulate random numbers for the $GSC(\mu, \sigma, p)$ distribution. The parameters were kept at $\mu = 2, \sigma = 1$ and $p = 0.7$. Proceeding with estimation, the estimate obtained for p again differs significantly from the true parameter, with p converging to a value close to 1. The impact on the log-likelihood was investigated and it increases for an increasing p.

- It is noted that the same observation with regards to the estimate of p is also evident when the *DEoptim* package was used to conduct the aforementioned investigation.

It is worthwhile to conduct further research on the behaviour of p and reasons why this peculiarity is evident for simulated data.

5.2 Guinea pig data

The first data set considers the survival times of guinea pigs that were obtained freely from Bjerkedal [10]. The guinea pigs were injected with different doses of tubercle bacilli due to their particular high vulnerability to tuberculosis [24] [10].

The total number of observations in the data is 72. A histogram of the data is plotted in Figure 24. The coefficient of skewness of the sample is calculated as 1.759. It can be observed from the histogram as well as the value of the sample skewness that the data is right skewed and hence suffices the need to fit distributions that can model skewed data.

Before analysing the data, it has been decided to divide all the observations in the data by 50 for computational purposes as this will not affect the inference that will be conducted.

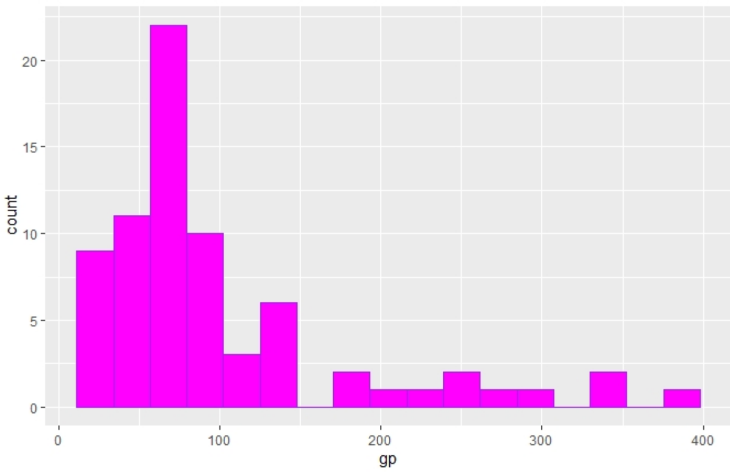


Figure 24: Histogram of guinea pig data.

The first model to be fitted will be the skew-normal model. That is, the $SN(\mu, \sigma^2, \lambda)$ distribution with PDF in (3). The MLEs of the unknown parameters are given in Table 10. The associated log-likelihood value becomes -115.252. The KS test statistic is 0.185. In addition to the latter, the AIC and BIC values are given by 236.504 and 243.334 respectively. The fitted PDF along with the histogram of the data is given in Figure 29.

The second model to be fitted will be the geometric skew-normal model. That is, the $GSN(\mu, \sigma, p)$ distribution with PDF in (31). The MLEs of the unknown parameters are given in Table 10. The fit statistics of the model is given in Table 11. The fitted PDF along with the histogram of the data is given in Figure 29. The empirical survival function and the fitted survival function is given in Figure 25. Furthermore, the empirical CDF and the fitted CDF is given in Figure 26.

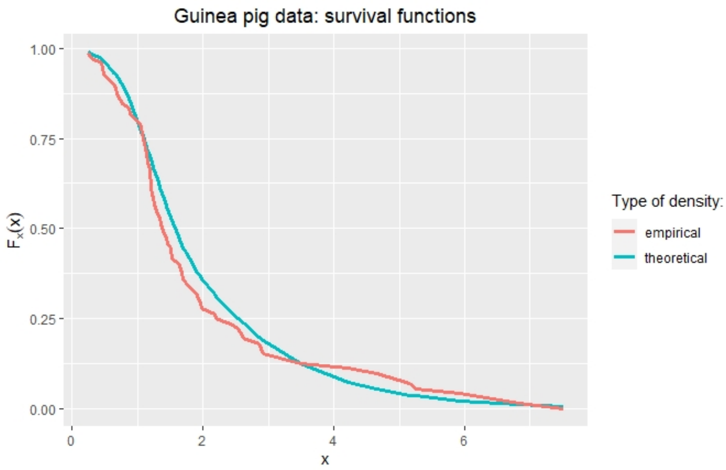


Figure 25: Survival functions of fitted $GSN(\mu, \sigma, p)$ model.

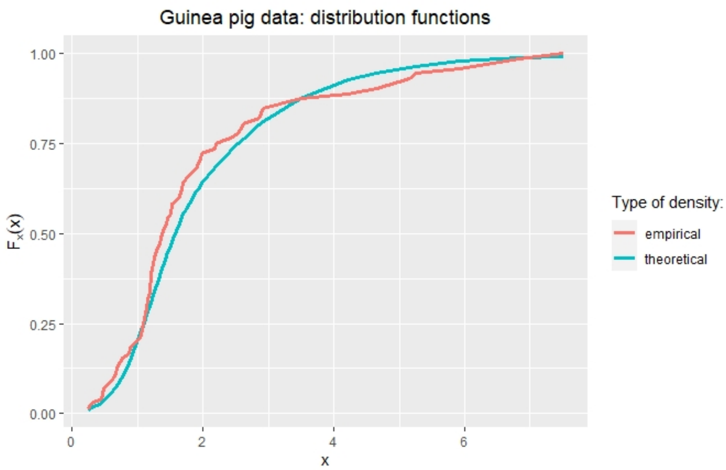


Figure 26: Distribution functions of fitted $GSN(\mu, \sigma, p)$ model.

The third model to be fitted will be the geometric skew-Cauchy model. That is, the $GSC(\mu, \sigma, p)$ distribution with PDF in (83). The MLEs of the unknown parameters are given in Table 10. The fit statistics of the model is given in Table 11. The fitted PDF along with the histogram of the data is given in Figure 29. The empirical survival function and the fitted survival function is given in Figure 27. Furthermore, the empirical CDF and the fitted CDF is given in Figure 28.

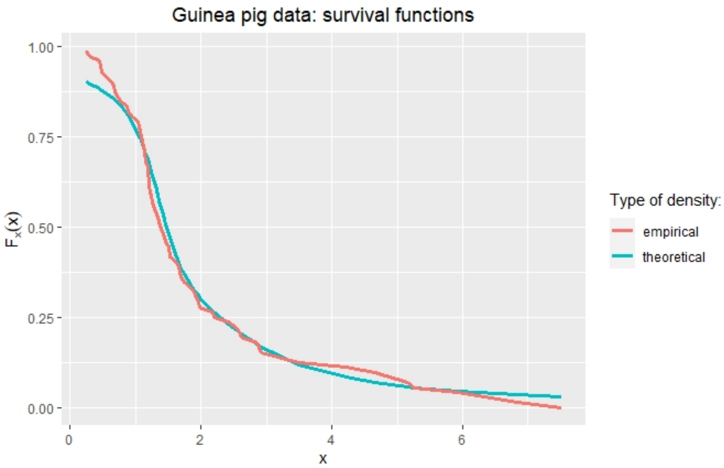


Figure 27: Survival functions of fitted $GSC(\mu, \sigma, p)$ model.

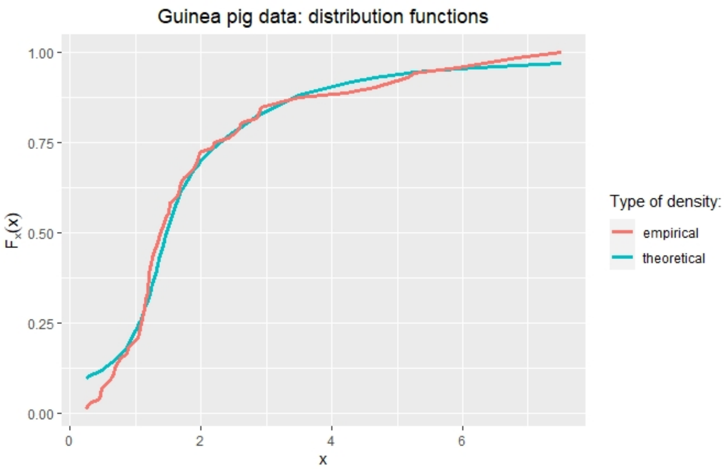


Figure 28: Distribution functions of fitted $GSC(\mu, \sigma, p)$ model.

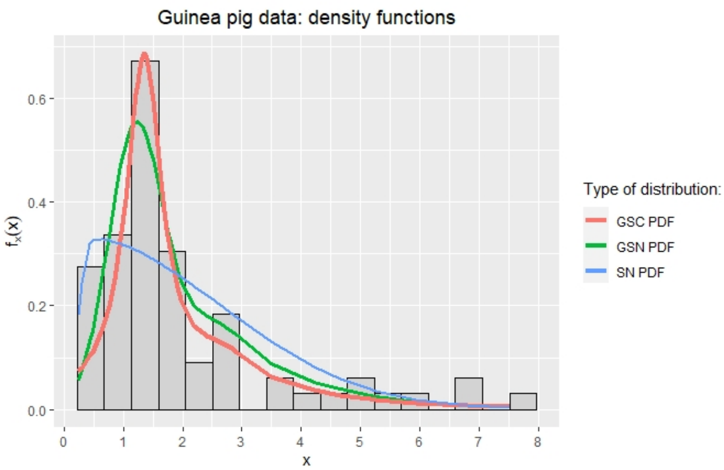


Figure 29: All three fitted models for the guinea pig data.

Table 10: Parameter estimates of three models for the guinea pig data

$SN(\mu, \sigma^2, \lambda)$	$GSN(\mu, \sigma, p)$	$GSC(\mu, \sigma, p)$
$\hat{\mu} = 0.227$	$\hat{\mu} = 1.202$	$\hat{\mu} = 1.351$
$\hat{\sigma} = 2.388$	$\hat{\sigma} = 0.457$	$\hat{\sigma} = 0.369$
$\hat{\lambda} = 183.446$	$\hat{p} = 0.602$	$\hat{p} = 0.773$

Table 11: Three models for the guinea pig data

	$SN(\mu, \sigma^2, \lambda)$	$GSN(\mu, \sigma, p)$	$GSC(\mu, \sigma, p)$
Log-likelihood	-115.252	-107.431	-108.213
KS test statistic	0.185	0.107	0.086
AIC	236.504	220.862	222.426
BIC	243.334	227.692	229.256

It can be observed from Table 11 that the new $GSC(\mu, \sigma, p)$ model provides a better fit than the $SN(\mu, \sigma^2, \lambda)$ model to this data, considering the log-likelihood, the KS test statistic, AIC and the BIC values.

It is observed that the $GSC(\mu, \sigma, p)$ model provides a better fit to the data than the $GSN(\mu, \sigma, p)$ model based on the KS test statistic. This can also be observed from Figures 26 and 28 where the empirical CDF of the $GSC(\mu, \sigma, p)$ model is closer to the fitted CDF, as opposed to the $GSN(\mu, \sigma, p)$ model.

However, the $GSN(\mu, \sigma, p)$ model provides a better fit than both the $GSC(\mu, \sigma, p)$ model and the $SN(\mu, \sigma^2, \lambda)$ model when considering the log-likelihood, AIC and the BIC.

The fact that the $GSN(\mu, \sigma, p)$ model performs better than the $GSC(\mu, \sigma, p)$ model could be due to the fact that the data does not exhibit fat tails (for which the $GSC(\mu, \sigma, p)$ model is appropriate). Figure 30 depicts the fitted CDFs of the $GSC(\mu, \sigma, p)$ and $GSN(\mu, \sigma, p)$ models. It can be observed from Figure 30 that the $GSC(\mu, \sigma, p)$ model accommodates fatter lower- and upper-tails than the $GSN(\mu, \sigma, p)$ model. The latter can also be observed in Tables 12 and 13, where the $GSC(\mu, \sigma, p)$ model has greater lower-tail probabilities and smaller upper-tail probabilities, indicative of the ability to model fatter tails than the $GSN(\mu, \sigma, p)$ model.

Table 12: Lower tail probabilities of $GSN(\mu, \sigma, p)$ vs. $GSC(\mu, \sigma, p)$ model

	$GSN(\mu, \sigma, p)$	$GSC(\mu, \sigma, p)$
$x = 0.4$	0.004	0.113
$x = 0.5$	0.008	0.124
$x = 0.6$	0.016	0.136
$x = 0.7$	0.030	0.152
$x = 0.8$	0.052	0.171

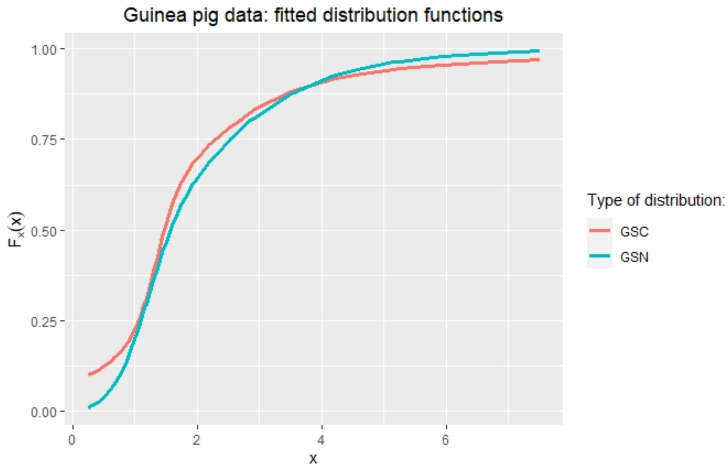


Figure 30: Distribution functions of fitted $GSC(\mu, \sigma, p)$ and $GSN(\mu, \sigma, p)$ models.

Table 13: Upper tail probabilities of $GSN(\mu, \sigma, p)$ vs. $GSC(\mu, \sigma, p)$ model

	$GSN(\mu, \sigma, p)$	$GSC(\mu, \sigma, p)$
$x = 4$	0.966	0.908
$x = 4.5$	0.980	0.926
$x = 5$	0.988	0.940
$x = 5.5$	0.993	0.949
$x = 6$	0.996	0.957

Considering Figures 26 and 28, it can also be observed that the $GSC(\mu, \sigma, p)$ model is a better fit in the middle part of the data, whereas the $GSN(\mu, \sigma, p)$ model is a better fit at the tails. The latter supports the fact that the data does not exhibit fat tails and hence the $GSC(\mu, \sigma, p)$ model not being a better fit than the $GSN(\mu, \sigma, p)$ model for the tail data.

5.3 Danish fire loss data

The second data set that will be analysed considers data that comprises Danish fire losses. The data contains individual losses above 1 million Danish Krones. The data were analysed by McNeil [29] and were collected by a reinsurance company in Denmark [14]. The data set spans over the period from 3 January 1980 till 31 December 1990 [14].

The total number of observations in the data is 2167. A histogram of the data is plotted in Figure 31. The coefficient of skewness is calculated as 18.737. It can be observed from the histogram as well as the value of the sample skewness that the data is right skewed and hence suffices the need to fit distributions that can model skewed data.

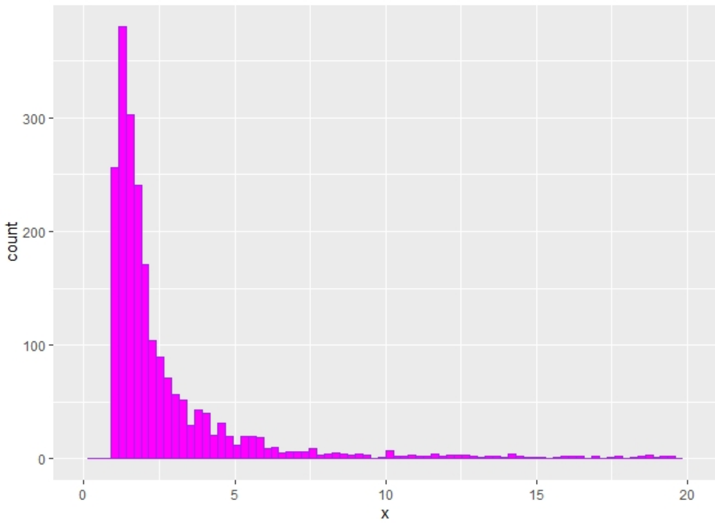


Figure 31: Histogram of Danish fire loss data.

The first model to be fitted will be the skew-normal model. That is, the $SN(\mu, \sigma^2, \lambda)$ distribution with PDF in (3). The MLEs of the unknown parameters are given in Table 14. The associated log-likelihood value becomes -6301.178. The KS test statistic is 0.988. In addition to the latter, the AIC and BIC values are given by 12608.36 and 12625.4 respectively.

The second model to be fitted will be the geometric skew-normal model. That is, the $GSN(\mu, \sigma, p)$ distribution with PDF in (31). The MLEs of the unknown parameters are given in Table 14. The fit statistics of the model is given in Table 15. The fitted PDF along with the histogram of the data is given in Figure 36. The empirical survival function and the fitted survival function is given in Figure 32. Furthermore, the empirical CDF and the fitted CDF is given in Figure 33.

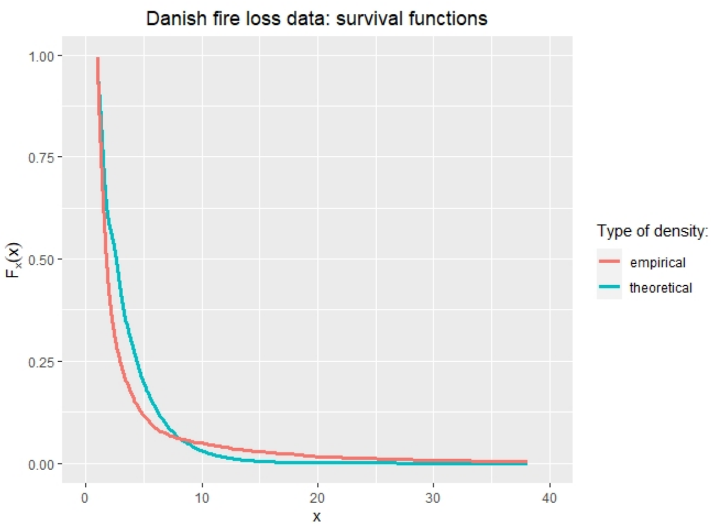


Figure 32: Survival functions of fitted $GSN(\mu, \sigma, p)$ model.

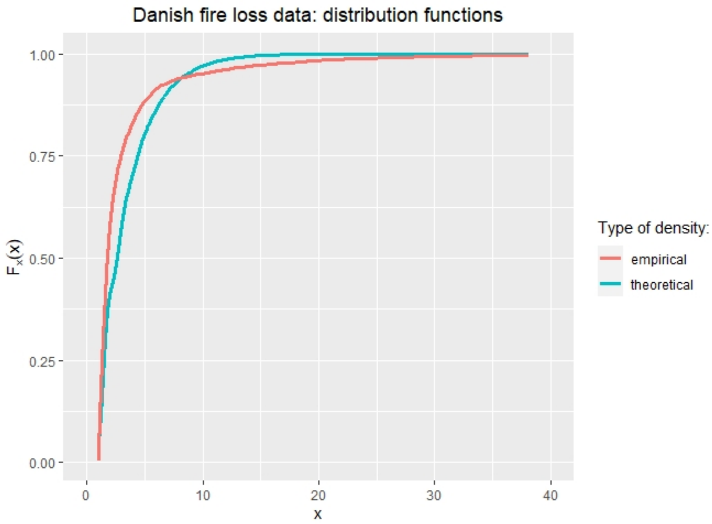


Figure 33: Distribution functions of fitted $GSN(\mu, \sigma, p)$ model.

The third model to be fitted will be the geometric skew-Cauchy model. That is, the $GSC(\mu, \sigma, p)$ distribution with PDF in (83). The MLEs of the unknown parameters are given in Table 14. The fit statistics of the model is given in Table 15. The fitted PDF along with the histogram of the data is given in Figure 36. The empirical survival function and the fitted survival function is given in Figure 34. Furthermore, the empirical CDF and the fitted CDF is given in Figure 35.

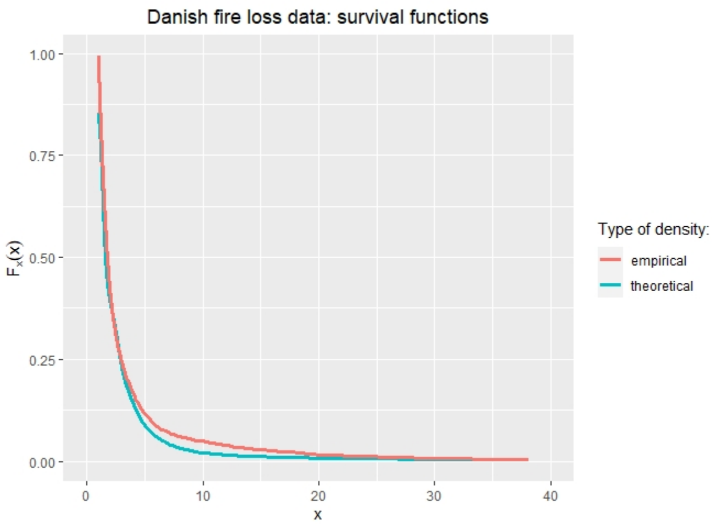


Figure 34: Survival functions of fitted $GSC(\mu, \sigma, p)$ model.

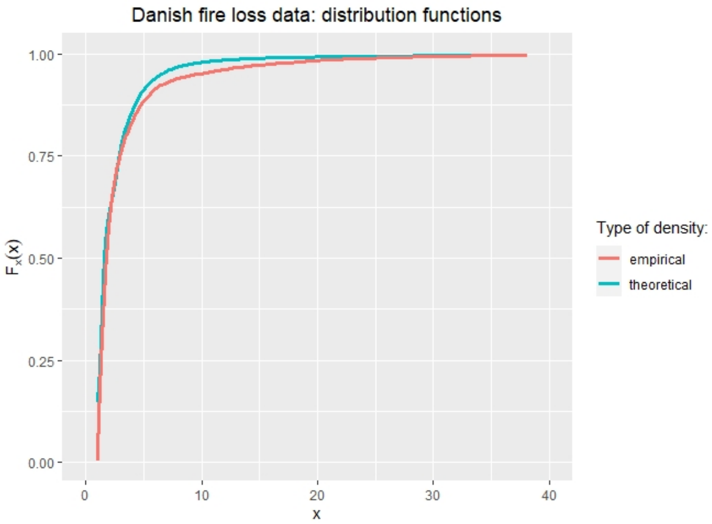


Figure 35: Distribution functions of fitted $GSC(\mu, \sigma, p)$ model.

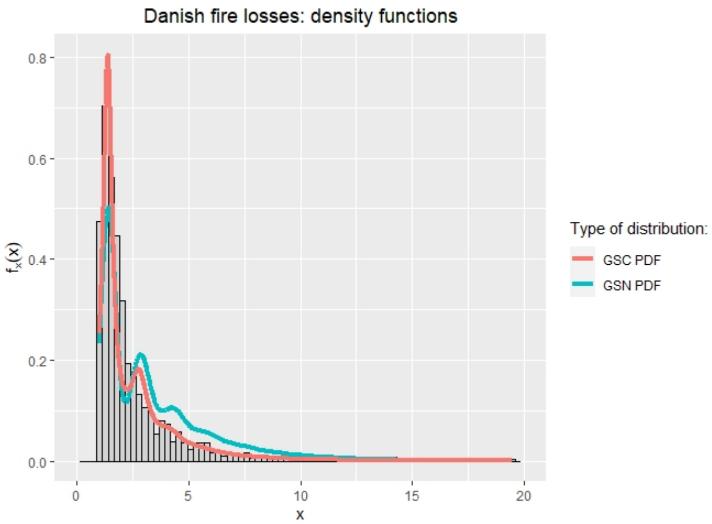


Figure 36: All three fitted models for the Danish fire loss data.

Table 14: Parameter estimates of three models for the Danish fire loss data

$SN(\mu, \sigma^2, \lambda)$	$GSN(\mu, \sigma, p)$	$GSC(\mu, \sigma, p)$
$\hat{\mu} = 0.909$	$\hat{\mu} = 1.414$	$\hat{\mu} = 1.383$
$\hat{\sigma} = 8.844$	$\hat{\sigma} = 0.333$	$\hat{\sigma} = 0.254$
$\hat{\lambda} = 183.446$	$\hat{p} = 0.418$	$\hat{p} = 0.625$

Table 15: Three models for the Danish fire loss data

	$SN(\mu, \lambda, \sigma^2)$	$GSN(\mu, \sigma, p)$	$GSC(\mu, \sigma, p)$
Log-likelihood	-6301.178	-4111.879	-3852.932
KS test statistic	0.988	0.992	0.987
AIC	12608.36	8229.757	7711.863
BIC	12625.4	8246.801	7728.907

It can be observed from Table 15 that the new $GSC(\mu, \sigma, p)$ model provides a better fit than both the $SN(\mu, \sigma^2, \lambda)$ model as well as the $GSN(\mu, \sigma, p)$ model to this data, considering the log-likelihood, the KS test statistic, AIC and the BIC values. The latter can also be observed from Figures 33 and 35 where the empirical CDF of the $GSC(\mu, \sigma, p)$ model is closer to the fitted CDF, as opposed to the $GSN(\mu, \sigma, p)$ model. It seems that the $GSC(\mu, \sigma, p)$ model is a better fit to both the middle part and the tails of the data. Furthermore, it is observed from Figure 36 that the $GSC(\mu, \sigma, p)$ model provides a closer fit to the data as compared to the $GSN(\mu, \sigma, p)$ model. In fact, the $GSC(\mu, \sigma, p)$ model is fitted close to the tail data of the histogram, whereas the $GSN(\mu, \sigma, p)$ model is relatively far off from the tail data.

The above results from the analysis of the Danish fire loss data show that the $GSC(\mu, \sigma, p)$ model is a competitive model. Thus, it can be used as an alternative to the skew-normal and geometric skew-normal distributions.

6 Chapter 6

The dissertation is concluded by providing a summary of what has been done and outlining considerations for future work.

6.1 Summary

In this dissertation the geometric skew-Cauchy distribution was introduced. The latter distribution utilized the geometric and Cauchy distributions in a compounding sum of i.i.d. random variables. This was in addition to the existing utilization of the geometric and normal distributions in a compounding sum of i.i.d. random variables.

The new model has the same number of parameters as the skew-normal distribution, and is more flexible since it can be unimodal or multimodal. In addition, although the geometric skew-normal distribution can also be unimodal or multimodal, the new model is shown to be reasonably more flexible since it can accommodate for fatter tails in the data.

The new alternative distribution is presented through theoretical development, descriptive research and experimental research in the form of simulation. An algorithm is presented to find the MLEs of the unknown parameters of the new model. The findings are supported through implementation on two real data sets. The aims and objectives have been met to sustain the rationale of conducting research on an alternative distribution. It is concluded that the newly proposed three-parameter distribution can be used as an alternative to the skew-normal and geometric skew-normal distributions. Furthermore, it appears overall as if the geometric skew-Cauchy distribution is a competitive model to consider as an alternative to the other models revisited in the dissertation.

Consideration is presented in the following subsection on generalizing the new model to the multivariate case. It will be worthwhile to continue developing the multivariate case as an alternative consideration to other multivariate distributions such as the multivariate skew-normal and geometric skew-normal distributions.

6.2 Future work

Subbotin introduced the generalised normal distribution that is more flexible than the normal distribution to allow for fatter tails [39]. Arellano-Valle et. al. [3] introduced a skew-generalised normal distribution which contains the skew-normal distribution as a special case. It would be possible to consider utilizing the geometric distribution and either the generalised normal or the skew-generalised normal distributions in a compounding sum of independent random variables.

Another possible idea for future work is to extend the univariate geometric skew-Cauchy distribution to the multivariate case. The motivation to explore the multivariate case is supported by the fact that there is a lack of skewed distributions in dimensions greater than one [25]. For this reason, the proceeding generalization of the univariate geometric skew-Cauchy is outlined.

The multivariate Cauchy distribution is a special case of the multivariate student's t-distribution with $\nu = 1$ degree of freedom [23].

Theorem 50. *A d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)^T$ is said to have the d-variate t distribution if its joint PDF is given by*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{(\pi\nu)^{\frac{d}{2}}\Gamma\left(\frac{\nu}{2}\right)|\mathbf{\Sigma}|^{\frac{1}{2}}}\left[1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^{-\frac{(\nu+d)}{2}} \tag{104}$$

with degrees of freedom $\nu > 0$, $d \times d$ mean vector $\boldsymbol{\mu}$, and $d \times d$ positive definite covariance matrix $\mathbf{\Sigma}$ (and with $\mathbf{R}$ denoting the corresponding $d \times d$ correlation matrix) [23].

If $d = 1$, $\boldsymbol{\mu} = 1$ and $\mathbf{\Sigma} = 1$, then (104) becomes the PDF of the univariate student's t-distribution [23]. Since the multivariate Cauchy distribution is a special case of (104), the following corollary follows immediately.

Corollary 5. *A d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)^T$ is said to have the d-variate Cauchy distribution if its joint PDF is given by*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma\left(\frac{1+d}{2}\right)}{(\pi)^{\frac{d}{2}}\Gamma\left(\frac{1}{2}\right)|\mathbf{\Sigma}|^{\frac{1}{2}}}\left[1 + (\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^{-\frac{(1+d)}{2}} \tag{105}$$

with $d \times d$ mean vector $\boldsymbol{\mu}$ and $d \times d$ positive definite covariance matrix $\mathbf{\Sigma}$ (with $\mathbf{R}$ denoting the corresponding $d \times d$ correlation matrix).

Theorem 51. *If $\mathbf{X}$ has the d-variate t distribution with degrees of freedom $\nu > 0$, $d \times d$ mean vector $\boldsymbol{\mu}$, and $d \times d$ positive definite covariance matrix $\mathbf{\Sigma}$, then for any nonsingular scalar matrix $\mathbf{C}$ and for any $\mathbf{a}, \mathbf{CX}+\mathbf{a}$ has the d-variate t distribution with degrees of freedom ν, mean vector $\mathbf{C}\boldsymbol{\mu}+\mathbf{a}$, and covariance matrix $\mathbf{C\Sigma C}^T$ [23].*

Naturally, the latter result can be applied to the multivariate Cauchy distribution. This yields that the distribution of a linear combination of d-variate Cauchy random variables, will also follow the multivariate Cauchy distribution [23]. A d-variate Cauchy random variable with location parameter $\boldsymbol{\mu}$ and scale parameter $\mathbf{\Sigma}$, will be denoted by $C_d(\boldsymbol{\mu}, \mathbf{\Sigma})$.

Theorem 52. *Let $\{\mathbf{Y}_i : i = 1, 2, ...\}$ be i.i.d $C_d(\boldsymbol{\mu}, \mathbf{\Sigma})$ random variables and suppose that $N \sim GE(p)$,* *with N and $\boldsymbol{Y_i'}s$ independently distributed. Then define*

$$\boldsymbol{X} \overset{dist.}{=} \sum_{i=1}^{N} \boldsymbol{Y_i} \tag{106}$$

where $\overset{dist.}{=}$ indicates equal in distribution. It is then observed that $\boldsymbol{X}$ is a multivariate geometric skew-Cauchy random variable with parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ and p. This will be denoted by $MVGSC_d\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p\right)$, where $d = 1, 2, ...$ indicates the number of variables.

Future work can include the derivations of the joint PDF as well as the joint CDF of the variable $(\boldsymbol{X}, N)$. In turn, this can be used to derive the CDF of the random variable $\boldsymbol{X}$ in (106). A possible representation of the PDF of the random variable $\boldsymbol{X}$ in (106) is given by the proceeding theorem for $0 < p < 1$. If $p = 1$, then $\boldsymbol{X} \sim C_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Theorem 53. *Let $\boldsymbol{X} \sim MVGSC_d\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p\right)$ and $N \sim GE(p)$ with composition given by (106), then the PDF of the random variable $\boldsymbol{X}$ is given by*

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{k=1}^{\infty} p(1-p)^{k-1} \Psi_d(k\boldsymbol{\mu}, k\boldsymbol{\Sigma}) \tag{107}$$

where

$$\Psi_d(k\boldsymbol{\mu}, k\boldsymbol{\Sigma}) = \frac{\Gamma\left(\frac{1+d}{2}\right)}{(k\pi)^{\frac{d}{2}} \left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right)} \left[1 + \left(\boldsymbol{x} - k\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{x} - k\boldsymbol{\mu}\right)\right]$$

for $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$: $d \times d$ positive definite covariance matrix and $0 < p < 1$.

Bibliography

[1] H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

[2] T. Anderson. *An introduction to multivariate statistical analysis.* J. Wiley and Sons, 2003.

[3] R. Arellano-Valle, H. Gomez, and F. Quintana. A new class of skew-normal distributions. *Communications in statistics - theory and methods*, 33(7):1465–1480, 2004.

[4] A. Azzalini. A class of distributions which includes the normal one. *Scandinavian Journal of Statistics*, 12(2):171–178, 1985.

[5] A. Azzalini and A. Capitanio. Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, 61(3):579–602, 1999.

[6] A. Azzalini and A. Capitanio. *The skew-normal and related families.* Cambridge University Press, 2013.

[7] A. Azzalini and D. Valle. The mulitvariate skew-normal distribution. *Biometrika*, 83(4):715–726, 1996.

[8] L. Bain and M. Engelhardt. *Introduction to probability and mathematical statistics.* Thomson Learning, 1992.

[9] N. Balakrishnan, Y. Ye, and R. Al-Jarallah. Likelihood inference for Type II bivariate Polya-Aeppli distribution. *Communications in statistics - simulation and computation*, 46(8):6463–6453, 2017.

[10] T. Bjerkedal. Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *American Journal of Hygiene*, 72:130–148, 1960.

[11] I. Chakravarti, R. Laha, and J. Roy. *Handbook of methods of applied statistics.* J. Wiley and Sons, 1967.

[12] J. Dennis and J. More. Quasi-Newton methods, motivation and theory. *Society for Inudstrial and Applied Mathematics*, 19(1):46–89, 1977.

[13] A. Eberl. and B. Klar. Asymptotic distributions and performance of empirical skewness measures. *Computational Statistics & Data Analysis*, 146:106939, 2020.

[14] M. Eling. Fitting insurance claims to skewed distributions: are the skew-normal and skew-student good models? *Working papers on risk management and insurance*, 98(1):1–31, 2011.

[15] H. Gomez, D. Elal-Olivero, H. Salinas, and H. Bolfarine. Bimodal extension based on the skew-normal distribution with application to pollen data. *Environmetrics*, 22:50–62, 2011.

[16] P. Guidici and S. Figini. *Applied data mining for business and industry*, chapter 4, pages 109–112. J. Wiley and Sons, 2009.

[17] A. Gupta and J. Chen. A class of multivariate skew-normal models. *Annals of the Institute of Statistical Mathematics*, 56(2):305–315, 2004.

[18] R. Gupta and R. Gupta. Generalized skew-normal model. *Spanish Society of Statistics and Operations Research Test*, 13(2):501–524, 2004.

[19] S. Harrar and A. Gupta. On matrix variate skew-normal distributions. *Statistics*, 42(2):179–194, 2008.

[20] N. Henze. A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, 13(4):271–275, 1986.

[21] R. A. Johnson and D. W. Wichern. *Applied multivariate statistical analysis*. Prentice-Hall Inc., 2002.

[22] Y. M. Kanatar and B. Senoglu. A comparative study for the location and scale parameters of the Weibull distribution with given shape parameter. *Computers and Geosciences*, 34:1900–1909, 2008.

[23] S. Kotz and S. Nadarajah. *Multivariate t distributions and their applications*. Cambridge University Press, 2004.

[24] D. Kundu. Geometric skew-normal distribution. *The Indian Journal of Statistics*, 76(2):167–189, 2014.

[25] D. Kundu. Multivariate geometric skew-normal distribution. *Journal of Theoretical and Applied Statistics*, 51(6):1377–1397, 2017.

[26] John D. MacArthur. Cumulants. Stat306, 2013.

[27] K. Mardia. Measure of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530, 1970.

[28] K. Mardia and P. Zemroch. Measures of multivariate skewness and kurtosis. *Journal of the Royal Statistical Society.*, 24(2):262–265, 1975.

[29] A. McNeil. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27(1):117–137, 1997.

[30] D. B. Owen. Tables for computing bivariate normal probabilities. *The Annals of Mathematical Statistics*, 27(4):1075–1090, 1956.

[31] K. Pearson. Contributions to the mathematical theory of evolution: skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London*, 186:343–414, 1895.

[32] A. Pewsey. Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27(7):859–870, 2000.

[33] D. Poole. *Linear algebra: a modern introduction*. Cengage learning, 2015.

[34] W. Press and W. Vetterling. *Numerical recipes in Pascal: the art of scientific computing*. London, England: Cambridge University Press, 1989.

[35] G. Schwwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[36] R. Staudte. Inference for quantile measures of kurtosis, peakedness and tail weight. *Communication in statistics- theory and methods*, 46(7):3148–3163, 2014.

[37] J. Stewart. *Calculus: early transcendentals*. Cengage Learning, 2015.

[38] P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE signal processing magazine*, 21(4):36–47, 2004.

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

[40] C. Walck. *Hand-book on statistical distributions for experimentalists*. Particle Physics Group Fysikum, 2007.

[41] A. Yalcinkaya, B. Senoglu, and U. Yolcu. Maximum likelihood estimation for the parameters of skew-normal distribution using genetic algorithm. *Swarm and Evolutionary Computation*, 38:127–138, 2018.

Appendix A: definitions, lemmas and results

A.1 Definitions

Definition 1. *If X is a random variable then the MGF of X is given by the following expected value formulation [8]:*

$$M_X(t) = E\left(e^{tX}\right). \tag{108}$$

This is called the MGF of X if the expected value exists for all values of t in some interval of the form $-k < t < k$ for some $k > 0$.

Definition 2. *The cumulant generating function (CGF) of a random variable Y, that is the function $G_Y(t)$, is defined as the logarithm of the moment generating function of the random variable Y [6]. Thus,*

$$G_Y(t) = \log(M_Y(t)). \tag{109}$$

Definition 3. *The n^{th} cumulant, k_n, of a random variable Y can be obtained via the n^{th} derivative of the CGF of the random variable for $n = 1, 2, 3, 4$. Thus,*

$$k_n = \frac{d^n}{dt^n} G_Y(t) \tag{110}$$

for $n = 1, 2, 3, 4$ [6].

Definition 4. *The inverse Mills ratio is defined as the ratio between the PDF and the CDF of the standard normal distribution [6]. That is, the inverse Mills ratio is defined as:*

$$b(y) = \frac{\phi(y)}{\Phi(y)} \tag{111}$$

for $y \in \mathbb{R}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the PDF and CDF of the standard normal distribution respectively.

Definition 5. *The conditional probability of event A, given event B, is defined as [8]:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{112}$$

if $P(B) \neq 0$.

A.2 Lemmas

Lemma 1. *If $V \sim N(0,1)$ then*

$$E_V[\Phi(hV+m)] = \Phi\left(\frac{m}{\sqrt{1+h^2}}\right) \tag{113}$$

for $h, m \in \mathbb{R}$ [6], with $\Phi(\cdot)$ denoting the standard normal CDF.

Proof. If $V \sim N(0,1)$ then it is true that $Y = hV + m \sim N(m, h^2)$ [8]. If $Z \sim N(0,1)$ then it is true that $Z - Y \sim N(-m, 1+h^2)$ [8]. Now let $X = \Phi(Y)$. Then it follows that

$$\begin{aligned} E(X) &= E_Y\left[\Phi(Y)\right] \\ &= \int_{-\infty}^{\infty}\int_0^y e^{-\frac{z^2}{2}} dz f(y) dy \\ &= E_Y\left[P_z\left(Z \leq y|y\right)\right] \\ &= P_{Y,Z}\left(Z \leq Y\right) \\ &= P_{Y,Z}\left(Z - Y \leq 0\right). \end{aligned}$$

Thus,

$$E_V[\Phi(hV+m)] = \Phi\left(\frac{m}{\sqrt{1+h^2}}\right).$$

□

Lemma 2. *If $\boldsymbol{V} \sim N_d(\boldsymbol{0}, \boldsymbol{\Sigma})$ then*

$$E\left[\Phi(\boldsymbol{h^T V} + \boldsymbol{m})\right] = \Phi\left(\frac{m}{\sqrt{1+\boldsymbol{h^T \Sigma h}}}\right) \tag{114}$$

for $\boldsymbol{h} \in \mathbb{R}^d$ and $m \in \mathbb{R}$ [6], with $\Phi(\cdot)$ denoting the standard normal CDF.

Proof. The proof follows in a similar fashion to that of the proof of (113), by noting that $\boldsymbol{h^T V} \sim N(\boldsymbol{0}, \boldsymbol{h^T \Sigma h})$ when $\boldsymbol{V} \sim N_d(\boldsymbol{0}, \boldsymbol{\Sigma})$. □

Lemma 3. *If $\boldsymbol{B}$ is a symmetric positive definite $d \times d$ matrix, $\boldsymbol{a}$ and $\boldsymbol{k}$ are d-vectors and k_0 is a scalar, then*

$$\begin{aligned} I &= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}|\boldsymbol{B}|^{1/2}} e^{\left\{-\frac{1}{2}\left(\boldsymbol{x}^\top \boldsymbol{B}^{-1}\boldsymbol{x} - 2\boldsymbol{a}^\top \boldsymbol{x}\right)\right\}} \Phi\left(k_0 + \boldsymbol{k}^\top \boldsymbol{x}\right) \mathrm{d}\boldsymbol{x} \\ &= e^{\left(\frac{1}{2}\boldsymbol{a}^\top \boldsymbol{B}\boldsymbol{a}\right)} \Phi\left(\frac{k_0 + \boldsymbol{k}^\top \boldsymbol{B}\boldsymbol{a}}{\sqrt{1+\boldsymbol{k}^\top \boldsymbol{A}\boldsymbol{k}}}\right). \end{aligned} \tag{115}$$

Proof. Using (114), and the fact that $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{\mu}$ and $\boldsymbol{\mu} = \boldsymbol{B}\boldsymbol{a}$, it follows that

$$
\begin{aligned}
I &= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}|\boldsymbol{B}|^{1/2}} e^{\left\{-\frac{1}{2}\left(\boldsymbol{x}^\top \boldsymbol{B}^{-1}\boldsymbol{x} - 2\boldsymbol{a}^\top \boldsymbol{x}\right)\right\}} \Phi\left(k_0 + \boldsymbol{k}^\top \boldsymbol{x}\right) \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}|\boldsymbol{B}|^{1/2}} e^{\left\{-\frac{1}{2}\left((\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \boldsymbol{\mu}^T \boldsymbol{B}^{-1}\boldsymbol{\mu}\right)\right\}} \Phi\left(k_0 + \boldsymbol{k}^\top \boldsymbol{x}\right) \mathrm{d}\boldsymbol{x} \\
&= e^{\left\{\frac{1}{2}\boldsymbol{a}^T \boldsymbol{B}\boldsymbol{a}\right\}} \int_{\mathbb{R}^d} \phi\left(\boldsymbol{y}; \boldsymbol{B}\right) \Phi\left(k_0 + \boldsymbol{k}^T\left(\boldsymbol{y} + \boldsymbol{\mu}\right)\right) d\boldsymbol{y} \\
&= e^{\left(\frac{1}{2}\boldsymbol{a}^\top \boldsymbol{B}\boldsymbol{a}\right)} \Phi\left(\frac{k_0 + \boldsymbol{k}^\top \boldsymbol{B}\boldsymbol{a}}{\sqrt{1 + \boldsymbol{k}^\top \boldsymbol{A}\boldsymbol{k}}}\right).
\end{aligned}
$$

□

A.3 Results

Theorem 54. *Consider the random variable Y. Let $f(y)$ and $F(y)$ denote the PDF and CDF of Y respectively. Then,*

$$
f(y) = \frac{d}{dy} F(y) \tag{116}
$$

where the derivative exists [8].

Theorem 55. *The CDF of a standard normal random variable Y can be represented as:*

$$
\Phi(y) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{2}\right)^k y^{2k+1}}{k!(2k+1)}. \tag{117}
$$

Proof. From Bain and Engelhardt [8] and Stewart [37], it follows that

$$
\begin{aligned}
\Phi(y) &= \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{y}{\sqrt{2}}\right) \\
&= \frac{1}{2} + \frac{1}{2}\frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{y}{\sqrt{2}}\right)^{2k+1}}{k!(2k+1)} \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{2}\right)^k y^{2k+1}}{k!(2k+1)}.
\end{aligned}
$$

□

Theorem 56. *Let $b(\cdot)$ represent the inverse Mills ratio as given in (111). Let $b'(\cdot)$ and $b''(\cdot)$ denote the first and second derivatives of $b(\cdot)$ respectively. The following properties of the inverse Mills hold:*

1. ***Property 1.***

$$b(0) = \sqrt{\frac{2}{\pi}} \tag{118}$$

2. ***Property 2.***

$$b'(y) = \frac{d}{dy}b(y) = -b(y)[y + b(y)] \tag{119}$$

3. ***Property 3.***

$$b''(y) = \frac{d^2}{dy^2}b(y) = -b(y) + y^2 b(y) + 3y(b(y))^2 + 2(b(y))^3 \tag{120}$$

Proof. The derivations are done freely utilizing the product, quotient and chain rules as given in Stewart [37]. For Property 1, from Bain and Engelhard [8] by definition

$$\begin{aligned} b(0) &= \frac{\phi(0)}{\Phi(0)} \\ &= \frac{\frac{1}{\sqrt{2\pi}}}{\frac{1}{2}} \\ &= \sqrt{\frac{2}{\pi}}. \end{aligned}$$

For Property 2, we make use of the fact that

$$\begin{aligned} \phi'(y) &= \frac{d}{dy}\phi(y) \\ &= \frac{d}{dy}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}\right) \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}(-y) \\ &= -y\phi(y). \end{aligned}$$

Then, using (116) and (111), it follows that

$$\begin{aligned}
b'(y) &= \frac{d}{dy}b(y) \\
&= \frac{d}{dy}\left(\frac{\phi(y)}{\Phi(y)}\right) \\
&= \frac{\left(\frac{d}{dy}\phi(y)\right)\Phi(y) - \phi(y)\left(\frac{d}{dy}\Phi(y)\right)}{(\Phi(y))^2} \\
&= \frac{-y\phi(y)\Phi(y) - \phi(y)\phi(y)}{(\Phi(y))^2} \\
&= -\left(\frac{y\phi(y)}{\Phi(y)} + \left(\frac{\phi(y)}{\Phi(y)}\right)^2\right) \\
&= -\frac{\phi(y)}{\Phi(y)}\left[y + \frac{\phi(y)}{\Phi(y)}\right] \\
&= -b(y)[y + b(y)].
\end{aligned}$$

For Property 3, using Property 2, it follows that

$$\begin{aligned}
b''(y) &= \frac{d^2}{dy^2}b(y) \\
&= \frac{d}{dy}\left(\frac{d}{dy}b(y)\right) \\
&= \frac{d}{dy}(-b(y)[y + b(y)]) \\
&= -\frac{d}{dy}\left(yb(y) + (b(y))^2\right) \\
&= -\left(\frac{d}{dy}(yb(y)) + \frac{d}{dy}(b(y))^2\right) \\
&= -\left(\left(\frac{d}{dy}y\right)b(y) + y\left(\frac{d}{dy}b(y)\right) + 2b(y)\left(\frac{d}{dy}b(y)\right)\right) \\
&= -(b(y) + y(-b(y)[y + b(y)]) + 2b(y)(-b(y)[y + b(y)])) \\
&= -\left(b(y) - y^2b(y) - y(b(y))^2 - 2y(b(y))^2 - 2(b(y))^3\right) \\
&= -\left(b(y) - y^2b(y) - 3y(b(y))^2 - 2(b(y))^3\right) \\
&= -b(y) + y^2b(y) + 3y(b(y))^2 + 2(b(y))^3.
\end{aligned}$$

□

Corollary 6. *Property 3 yields that all $b^r(y)$ for $r > 1$ can be written as functions of $b(y)$ and powers of y [6]. That is,*

$$b^r(y) = \frac{d^r}{dy^r}b_0(y) \text{ for } r = 1, 2, ...$$

where

$$b_0(y) = log\left[2\Phi(y)\right].$$

Theorem 57. *If D is positive definite matrix of order p, the maximum of*

$$f(G) = -N\log|G| - \operatorname{tr} G^{-1}D$$

with respect to positive definite matrices G exists, occurs at

$$G = (1/N)D \tag{121}$$

and has the value

$$f[(1/N)D] = pN\log N - N\log|D| - pN[2].$$

Proof. The full proof is provided in Anderson [2]. □

Appendix B: code

Data analysis

```
#####################################################################
#Data analysis of the guinea pig data set
gp_nodiv <- c(12,15,22,24,24,32,32,33,34,
38,38,43,44,48,52,53,54,54,
55,56,57,58,58,59,60,60,60,
60,61,62,63,65,65,67,68,70,
70,72,73,75,76,76,81,83,84,
85,87,91,95,96,98,99,109,110,
121,127,129,131,143,146,146,175,175,
211,233,258,258,263,297,341,341,376)
gp <- gp_nodiv/50

#Plot of the data
library(ggplot2)
ggplot(data.frame(gp), aes(x=gp)) +
geom_histogram(aes(y=..density..),color='darkblue', fill='magenta', bins=17) +
scale_x_continuous(breaks = seq(0, 8, by = 1))

#####################################################################
#Skew-normal work
#Trying to obtain the skew-normal parameter
library(sn)
cp_space_est <- sn.mple(y=gp)$cp
dp_space_est <- cp2dp(cp_space_est,family="SN") #Conversion of CP parameters to DP
    ↪ parameters
dp_space_est
mean <- dp_space_est[1]
sd <- dp_space_est[2]
gamma1 <- dp_space_est[3]

#For the likelihood value
```

```
testdata <- data.frame(gp)
X <- gp
mod <- selm(X ~ 1, data=testdata)
summary(mod)

#Fitting to obtain the theoretical PDF of the skew-normal
sn_dens <- dsn(x=gp, xi=mean, omega=sd, alpha=gamma1)

#Trying to obtain the KS TS using a function
ks.test(gp, 'psn', xi=mean, omega=sd, alpha=gamma1)

#Obtaining the KS using first principles
#Empirical quantiles
sorted_dat <- sort(gp, decreasing = FALSE)
EQ <- cbind(seq(1,length(gp),1)/length(gp),sorted_dat)
EQ[length(gp),1] <- 0.99 #Assigning 0.99 as the 99th quantile

#Theoretical quantiles
quants <- EQ[,1]
theoSNq <- psn(x=gp, xi=mean, omega=sd, alpha=gamma1)
t_quant <- cbind(theoSNq, sorted_dat)

#Statistic Guidici page 110.
T1 = max(abs(EQ[,1]-t_quant[,1]))
T1

#AIC for the SN model
aic_sn <- 2*3 - 2*(-115.252)
aic_sn

#BIC for the SN model
bic_sn <- 3*log(length(gp)) - 2*(-115.252)
bic_sn
```

```
######################################################################
#Plotting the theoretical PDF
ggplot(data.frame(gp)) +
geom_histogram(aes(x = gp, y = ..density..), color='black', fill='lightgrey', bins=17)
    ↪ +
scale_x_continuous(breaks = seq(0, 8, by = 1)) +
#geom_line(aes(gp, gsn_pdf, color = 'GSN PDF'), size = 1.2) +
geom_line(aes(gp, dsn(gp, xi=mean, omega=sd, alpha=19.7001), color = 'SN PDF'), size
    ↪ =1.2) +
labs(x="x",y=parse(text="f[x](x)"),
title="Guinea pig data: density function") +
scale_color_discrete(name=expression("Type of distribution:")) +
theme(plot.title = element_text(hjust = 0.5))

######################################################################
#GSN work
#Parameters given in the document that were obtained by the EM
p <- 0.6022818
mu <- 1.202389
sigma <- 0.4565002
lower <- 1
upper <- 100

######################################################################
#Plotting the theoretical PDF
ggplot(data.frame(gp,gsn_pdf)) +
geom_histogram(aes(x = gp, y = ..density..), color='black', fill='lightgrey', bins=17)
    ↪ +
scale_x_continuous(breaks = seq(0, 8, by = 1)) +
geom_line(aes(gp, gsn_pdf, color = 'GSN PDF'), size = 1.2) +
#geom_line(aes(gp, dsn(gp, xi=mean, omega=sd, alpha=19.7001), color = 'SN PDF'), size
    ↪ =0.8) +
```

```
labs(x="x",y=parse(text="f[x](x)"),
title="Guinea pig data: density function") +
scale_color_discrete(name=expression("Type of distribution:")) +
theme(plot.title = element_text(hjust = 0.5))

######################################################################
#Plotting the theoretical CDF
ggplot(data.frame(gp,gsn_cdf)) +
geom_line(aes(gp, gsn_cdf, color = 'GSN CDF'), size = 1.2) +
labs(x="x",y=parse(text="F[x](x)"),
title="Guinea pig data: distribution function") +
scale_color_discrete(name=expression("Type of distribution:")) +
theme(plot.title = element_text(hjust = 0.5))

######################################################################
#Obtaining the log-likelihood based on the estimates
pdf_mat <- as.matrix(gsn_pdf)
log_pdf <- log(pdf_mat)
ll_gsn <- sum(log_pdf)
ll_gsn

#Obtaining the empirical CDF
cdf_emp_gsnf <- ecdf(x = gp) # is a function
cdf_emp_gsn <- cdf_emp_gsnf(gp) # values

#Plotting the EMP vs THEO CDF:
plot(cdf_emp_gsnf, main = "Empirical and Theoretical CDF", xlab = "y", ylab = "F(y)")
    ↪ # empirical cdf
lines(x = gp, y = gsn_cdf, col = "red", type = "l") # theoretical cdf

ggplot(data.frame(gp,gsn_cdf)) +
geom_line(aes(gp, gsn_cdf, color = 'theoretical'), size = 1.2) +
geom_line(aes(gp, cdf_emp_gsn, color = 'empirical'), size = 1.2) +
```

```
labs(x="x",y=parse(text="F[x](x)"),
title="Guinea pig data: distribution functions") +
scale_color_discrete(name=expression("Type of density:")) +
theme(plot.title = element_text(hjust = 0.5))

#Obtaining the theoretical survival function
gsn_surv <- 1 - gsn_cdf

#Obtaining the empirical survival function
gsn_surv_emp <- 1 - cdf_emp_gsn

#Implementation check, S(x) = 1 - F(x) where F(.) is the cdf:
#Checking if the theoretical survival function is close to the empirical survival
    ↪ function
all.equal(gsn_surv, 1 - cdf_emp_gsn)

#Plotting the EMP vs THEO survival function:
plot(stepfun(x = gp, y = c(1, gsn_surv_emp)), main = "Empirical and Theoretical
    ↪ Survival Function", xlab = "y", ylab = "S(y)") # empirical S
lines(x = gp, y = gsn_surv, col = "red", type = "l") # theoretical S

ggplot(data.frame(gp,gsn_surv)) +
geom_line(aes(gp, gsn_surv, color = 'theoretical'), size = 1.2) +
geom_line(aes(gp, gsn_surv_emp, color = 'empirical'), size = 1.2) +
labs(x="x",y=parse(text="F[x](x)"),
title="Guinea pig data: survival functions") +
scale_color_discrete(name=expression("Type of density:")) +
theme(plot.title = element_text(hjust = 0.5))

######################################################################
#Obtaining the KS using first principles
#Empirical quantiles
sorted_dat <- sort(gp, decreasing = FALSE)
```

```
EQ <- cbind(seq(1,length(gp),1)/length(gp),sorted_dat)
EQ[length(gp),1] <- 0.99 #Assigning 0.99 as the 99th quantile

#Theoretical quantiles
quants <- EQ[,1]
theoGSNq <- gsn_cdf
t_quant <- cbind(theoGSNq, sorted_dat)

#Statistic Guidici page 110.
T1 = max(abs(EQ[,1]-t_quant[,1]))
T1

######################################################################
#Trying to obtain the AIC and the BIC
aic_gsn <- 2*3 - 2*(ll_gsn)
aic_gsn

bic_gsn <- 3*log(length(gp)) - 2*(ll_gsn)
bic_gsn


######################################################################
#GSC work
#Parameters as obtained via the NR algorithm
p <- p
mu <- est_old[1]
sigma <- est_old[2]
lower <- 1
upper <- 100

######################################################################
#Plotting the theoretical PDF
ggplot(data.frame(gp,gsc_pdf)) +
```

```
geom_histogram(aes(x = gp, y = ..density..), color='black', fill='lightgrey', bins=17)
    ↪ +
scale_x_continuous(breaks = seq(0, 8, by = 1)) +
geom_line(aes(gp, gsc_pdf, color = 'GSC PDF'), size = 1.2) +
labs(x="x",y=parse(text="f[x](x)"),
title="Guinea pig data: density function") +
scale_color_discrete(name=expression("Type of distribution:")) +
theme(plot.title = element_text(hjust = 0.5))

######################################################################
#Plotting the theoretical CDF
ggplot(data.frame(gp,gsc_cdf)) +
geom_line(aes(gp, gsc_cdf, color = 'GSC CDF'), size = 1.2) +
labs(x="x",y=parse(text="F[x](x)"),
title="Guinea pig data: distribution function") +
scale_color_discrete(name=expression("Type of distribution:")) +
theme(plot.title = element_text(hjust = 0.5))

######################################################################
#Obtaining the log-likelihood based on the estimates
pdf_mat <- as.matrix(gsc_pdf)
log_pdf <- log(pdf_mat)
ll_gsc <- sum(log_pdf)
ll_gsc

#Obtaining the empirical CDF
cdf_emp_gscf <- ecdf(x = gp) # is a function
cdf_emp_gsc <- cdf_emp_gscf(gp) # values

#Plotting the EMP vs THEO CDF:
plot(cdf_emp_gscf, main = "Empirical and Theoretical CDF", xlab = "y", ylab = "F(y)")
    ↪ # empirical cdf
lines(x = gp, y = gsc_cdf, col = "red", type = "l") # theoretical cdf
```

```
#lines(x = gp, y = gsn_cdf, col = "blue", type = "l")

ggplot(data.frame(gp,gsc_cdf)) +
geom_line(aes(gp, gsc_cdf, color = 'theoretical'), size = 1.2) +
geom_line(aes(gp, cdf_emp_gsc, color = 'empirical'), size = 1.2) +
labs(x="x",y=parse(text="F[x](x)"),
title="Guinea pig data: distribution functions") +
scale_color_discrete(name=expression("Type of density:")) +
theme(plot.title = element_text(hjust = 0.5))

#Obtaining the theoretical survival function
gsc_surv <- 1 - gsc_cdf

#Obtaining the empirical survival function
gsc_surv_emp <- 1 - cdf_emp_gsc

#Implementation check, S(x) = 1 - F(x) where F(.) is the cdf:
#Checking if the theoretical survival function is close to the empirical survival
    ↪ function
all.equal(gsc_surv, 1 - cdf_emp_gsc)

#Plotting the EMP vs THEO survival function:
plot(stepfun(x = gp, y = c(1, gsc_surv_emp)), main = "Empirical and Theoretical
    ↪ Survival Function", xlab = "y", ylab = "S(y)") # empirical S
lines(x = gp, y = gsc_surv, col = "red", type = "l") # theoretical S
#lines(x = gp, y = gsn_surv, col = "blue", type = "l")

ggplot(data.frame(gp,gsc_surv)) +
geom_line(aes(gp, gsc_surv, color = 'theoretical'), size = 1.2) +
geom_line(aes(gp, gsc_surv_emp, color = 'empirical'), size = 1.2) +
labs(x="x",y=parse(text="F[x](x)"),
title="Guinea pig data: survival functions") +
scale_color_discrete(name=expression("Type of density:")) +
```

```
theme(plot.title = element_text(hjust = 0.5))

######################################################################
#Obtaining the KS using first principles
#Empirical quantiles
sorted_dat <- sort(gp, decreasing = FALSE)
EQ <- cbind(seq(1,length(gp),1)/length(gp),sorted_dat)
EQ[length(gp),1] <- 0.99 #Assigning 0.99 as the 99th quantile

#Theoretical quantiles
quants <- EQ[,1]
theoGSCq <- gsc_cdf
t_quant <- cbind(theoGSCq, sorted_dat)

#Statistic Guidici page 110.
T1 = max(abs(EQ[,1]-t_quant[,1]))
T1

######################################################################
#Trying to obtain the AIC and the BIC
ll_gsc <- -108.213
aic_gsc <- 2*3 - 2*(ll_gsc)
aic_gsc

bic_gsc <- 3*log(x_iter) - 2*(ll_gsc)
bic_gsc

#tail probability checks
#Checking an upper range of values say from x=20 till x=30
xrange_upper<-seq(4,7,0.5)
gsn2<-sapply(xrange_upper, function(i) sum(gsn_cdf_nosum(seq(lower, upper, 1), i=i)))
gsc2<-sapply(xrange_upper, function(i) sum(gsc_cdf_nosum(seq(lower, upper, 1), i=i)))
```

```
#Checking a lower range of values say from x=-3 till x=2
xrange_lower<-seq(0.4,0.8,0.1)
gsn3<-sapply(xrange_lower, function(i) sum(gsn_cdf_nosum(seq(lower, upper, 1), i=i)))
gsc3<-sapply(xrange_lower, function(i) sum(gsc_cdf_nosum(seq(lower, upper, 1), i=i)))

#####################################################################
#Parms for SN
mean <- 0.2270935
sd <- 2.388109
lams <- 19.7001

#Parms for GSN
p <- 0.6022818
mu <- 1.202389
sigma <- 0.4565002

lower <- 1
upper <- 100

#Parms for GSC
p <- 0.773
mu <- 1.351
sigma <- 0.369
lower <- 1
upper <- 100

#####################################################################
#Plotting the theoretical PDF
ggplot(data.frame(gp,gsn_pdf)) +
geom_histogram(aes(x = gp, y = ..density..), color='black', fill='lightgrey', bins=17)
    ↪ +
scale_x_continuous(breaks = seq(0, 8, by = 1)) +
geom_line(aes(gp, gsn_pdf, color = 'GSN PDF'), size = 1.3) +
```

```
geom_line(aes(gp, gsc_pdf, color = 'GSC PDF'), size = 1.6) +
geom_line(aes(gp, dsn(gp, xi=mean, omega=sd, alpha=19.7001), color = 'SN PDF'), size
    ↪ =1) +
labs(x="x",y=parse(text="f[x](x)"),
title="Guinea pig data: density functions") +
scale_color_discrete(name=expression("Type of distribution:")) +
theme(plot.title = element_text(hjust = 0.5))

######################################################################
#Data analysis for the Dansih fire loss data
library(fExtremes)
danishClaims
Danish<- danishClaims[,2]



#Histogram only
library(ggplot2)
ggplot(data.frame(Danish), aes(x=Danish)) +
geom_histogram(color='purple', fill='magenta', bins=80) +
scale_x_continuous(name="x", limits = c(0, 20))

########################################################
#Skew-normal work

gp <- Danish

#Trying to obtain the skew-normal parameters
library(sn)
cp_space_est <- sn.mple(y=gp)$cp
dp_space_est <- cp2dp(cp_space_est,family="SN") #Conversion of CP parameters to DP
    ↪ parameters
dp_space_est
mean <- dp_space_est[1]
```

```
sd <- dp_space_est[2]
gamma1 <- dp_space_est[3]

#For the likelihood value
testdata <- data.frame(gp)
X <- gp
mod <- selm(X ~ 1, data=testdata)
summary(mod)

#Fitting to obtain the theoretical PDF of the skew-normal
sn_dens <- dsn(x=gp, xi=mean, omega=sd, alpha=gamma1)

#Trying to obtain the KS TS using a function
ks.test(gp, 'psn', xi=mean, omega=sd, alpha=gamma1)

#Obtaining the KS using first principles
#Empirical quantiles
sorted_dat <- sort(gp, decreasing = FALSE)
EQ <- cbind(seq(1,length(gp),1)/length(gp),sorted_dat)
EQ[length(gp),1] <- 0.99 #Assigning 0.99 as the 99th quantile

#Theoretical quantiles
quants <- EQ[,1]
theoSNq <- psn(x=gp, xi=mean, omega=sd, alpha=gamma1)
t_quant <- cbind(theoSNq, sorted_dat)

#Statistic Guidici page 110.
T1 = max(abs(EQ[,1]-t_quant[,1]))
T1

#AIC for the SN model
aic_sn <- 2*3 - 2*(-6301.178)
aic_sn
```

```
#BIC for the SN model
bic_sn <- 3*log(length(gp)) - 2*(-6301.178)
bic_sn

#####################################################
#The geometric skew-Cauchy model

x <- Danish

library(DEoptim)

LGSC=function(theta){
p=theta[1]
mu=theta[2]
sigma=theta[3]
t=0
for(k in 1:1000){
t1=(1/k)*((1-p)^(k-1))*(1+((x-k*mu)/(k*sigma))^2)^(-1)
t=t+t1}

l=sum(log((p/(pi*sigma))*t))
return(-l)
}

h=DEoptim(LGSC,lower=c(0,-10,0),
upper=c(1,10,10),control = DEoptim.control(trace = TRUE))
par.hat=h$optim$bestmem
names(par.hat)=NULL
phat=par.hat[1]
muhat=par.hat[2]
sigmahat=par.hat[3]
value=-h$optim$bestval
```

```
AIC=6-2*value

gsc_parms <- par.hat
gsc_ll <- value
gsc_aic <- AIC

gsc_parms
gsc_ll
gsc_aic

p <- 0.6246752
mu <- 1.3827333
sigma <- 0.2535249
lower <- 1
upper <- 1000

#Plotting the theoretical PDF
ggplot(data.frame(Danish,gsc_pdf)) +
geom_histogram(aes(x = Danish, y = ..density..), color='purple', fill='magenta', bins
    ↪ =80) +
scale_x_continuous(name="x", limits = c(0, 20)) +
geom_line(aes(Danish, gsc_pdf, color = 'GSC PDF'), size = 1.2, color='black') +
labs(x="x",y=parse(text="f[x](x)"),
title="Danish fire loss data: density function") +
scale_color_discrete(name=expression("Type of distribution:")) +
theme(plot.title = element_text(hjust = 0.5))

#Obtaining the KS using first principles
#Empirical quantiles
sorted_dat <- sort(Danish, decreasing = FALSE)
EQ <- cbind(seq(1,length(Danish),1)/length(Danish),sorted_dat)
EQ[length(Danish),1] <- 0.99 #Assigning 0.99 as the 99th quantile
```

```
#Theoretical quantiles
quants <- EQ[,1]
theoGSCq <- gsc_cdf
t_quant <- cbind(theoGSCq, sorted_dat)

#Statistic Guidici page 110.
T1 = max(abs(EQ[,1]-t_quant[,1]))
T1

#Trying to obtain the AIC and the BIC
ll_gsc <- gsc_ll
aic_gsc <- 2*3 - 2*(ll_gsc)
aic_gsc

bic_gsc <- 3*log(length(Danish)) - 2*(ll_gsc)
bic_gsc

#Obtaining the empirical CDF
cdf_emp_gscf <- ecdf(x = Danish) # is a function
cdf_emp_gsc <- cdf_emp_gscf(Danish) # values

#Plotting the EMP vs THEO CDF:
plot(cdf_emp_gscf, main = "Empirical and Theoretical CDF", xlab = "y", ylab = "F(y)")
    ↪ # empirical cdf
lines(x = Danish, y = gsc_cdf, col = "red", type = "l") # theoretical cdf
#lines(x = Danish, y = gsn_cdf, col = "blue", type = "l")

ggplot(data.frame(Danish,gsc_cdf)) +
geom_line(aes(Danish, gsc_cdf, color = 'theoretical'), size = 1.2) +
geom_line(aes(Danish, cdf_emp_gsc, color = 'empirical'), size = 1.2) +
scale_x_continuous(name="x", limits = c(0, 40)) +
labs(x="x",y=parse(text="F[x](x)"),
title="Danish fire loss data: distribution functions") +
```

```
scale_color_discrete(name=expression("Type of density:")) +
theme(plot.title = element_text(hjust = 0.5))

#Obtaining the theoretical survival function
gsc_surv <- 1 - gsc_cdf

#Obtaining the empirical survival function
gsc_surv_emp <- 1 - cdf_emp_gsc

#Implementation check, S(x) = 1 - F(x) where F(.) is the cdf:
#Checking if the theoretical survival function is close to the empirical survival
    ↪ function
all.equal(gsc_surv, 1 - cdf_emp_gsc)

#Plotting the EMP vs THEO survival function:
plot(stepfun(x = Danish, y = c(1, gsc_surv_emp)), main = "Empirical and Theoretical
    ↪ Survival Function", xlab = "y", ylab = "S(y)") # empirical S
lines(x = Danish, y = gsc_surv, col = "red", type = "l") # theoretical S
#lines(x = Danish, y = gsn_surv, col = "blue", type = "l")

ggplot(data.frame(Danish,gsc_surv)) +
geom_line(aes(Danish, gsc_surv, color = 'theoretical'), size = 1.2) +
geom_line(aes(Danish, gsc_surv_emp, color = 'empirical'), size = 1.2) +
scale_x_continuous(name="x", limits = c(0, 40)) +
labs(x="x",y=parse(text="F[x](x)"),
title="Danish fire loss data: survival functions") +
scale_color_discrete(name=expression("Type of density:")) +
theme(plot.title = element_text(hjust = 0.5))



#######################################################
#The geometric skew-normal model
```

```
x <- Danish

library(DEoptim)

LGSN=function(theta){
p=theta[1]
mu=theta[2]
sigma=theta[3]
t=0
for(k in 1:1000){
t1=(1/sqrt(k))*((1-p)^(k-1))*(dnorm(((x-k*mu)/(sigma*sqrt(k))), mean=0, sd=1))
t=t+t1}

l=sum(log((p/(sigma))*t))
return(-l)
}

h=DEoptim(LGSN,lower=c(0,-10,0),
upper=c(1,10,10),control = DEoptim.control(trace = TRUE))
par.hat=h$optim$bestmem
names(par.hat)=NULL
phat=par.hat[1]
muhat=par.hat[2]
sigmahat=par.hat[3]
value=-h$optim$bestval
AIC=6-2*value

gsn_parms <- par.hat
gsn_ll <- value
gsn_aic <- AIC

gsn_parms
gsn_ll
```

```
gsn_aic


p <- 0.417621
mu <- 1.413684
sigma <- 0.333302
lower <- 1
upper <- 1000


#Plotting the theoretical PDF
ggplot(data.frame(Danish,gsn_pdf)) +
geom_histogram(aes(x = Danish, y = ..density..), color='purple', fill='magenta', bins
    ↪ =80) +
scale_x_continuous(name="x", limits = c(0, 20)) +
geom_line(aes(Danish, gsn_pdf, color = 'GSC PDF'), size = 1.2, color='black') +
labs(x="x",y=parse(text="f[x](x)"),
title="Guinea pig data: density function") +
scale_color_discrete(name=expression("Type of distribution:")) +
theme(plot.title = element_text(hjust = 0.5))


#Obtaining the KS using first principles
#Empirical quantiles
sorted_dat <- sort(Danish, decreasing = FALSE)
EQ <- cbind(seq(1,length(Danish),1)/length(Danish),sorted_dat)
EQ[length(Danish),1] <- 0.99 #Assigning 0.99 as the 99th quantile


#Theoretical quantiles
quants <- EQ[,1]
theoGSNq <- gsn_cdf
t_quant <- cbind(theoGSNq, sorted_dat)


#Statistic Guidici page 110.
T1 = max(abs(EQ[,1]-t_quant[,1]))
T1
```

```
#Trying to obtain the AIC and the BIC
ll_gsn <- gsn_ll
aic_gsn <- 2*3 - 2*(ll_gsn)
aic_gsn

bic_gsn <- 3*log(length(Danish)) - 2*(ll_gsn)
bic_gsn

#Obtaining the empirical CDF
cdf_emp_gsnf <- ecdf(x = Danish) # is a function
cdf_emp_gsn <- cdf_emp_gsnf(Danish) # values

#Plotting the EMP vs THEO CDF:
plot(cdf_emp_gsnf, main = "Empirical and Theoretical CDF", xlab = "y", ylab = "F(y)")
    ↪ # empirical cdf
lines(x = Danish, y = gsn_cdf, col = "red", type = "l") # theoretical cdf

ggplot(data.frame(Danish,gsn_cdf)) +
geom_line(aes(Danish, gsn_cdf, color = 'theoretical'), size = 1.2) +
geom_line(aes(Danish, cdf_emp_gsn, color = 'empirical'), size = 1.2) +
scale_x_continuous(name="x", limits = c(0, 40)) +
labs(x="x",y=parse(text="F[x](x)"),
title="Danish fire loss data: distribution functions") +
scale_color_discrete(name=expression("Type of density:")) +
theme(plot.title = element_text(hjust = 0.5))

#Obtaining the theoretical survival function
gsn_surv <- 1 - gsn_cdf

#Obtaining the empirical survival function
gsn_surv_emp <- 1 - cdf_emp_gsn
```

```
#Implementation check, S(x) = 1 - F(x) where F(.) is the cdf:
#Checking if the theoretical survival function is close to the empirical survival
    ↪ function
all.equal(gsn_surv, 1 - cdf_emp_gsn)

#Plotting the EMP vs THEO survival function:
plot(stepfun(x = Danish, y = c(1, gsn_surv_emp)), main = "Empirical and Theoretical
    ↪ Survival Function", xlab = "y", ylab = "S(y)") # empirical S
lines(x = Danish, y = gsn_surv, col = "red", type = "l") # theoretical S

ggplot(data.frame(Danish,gsn_surv)) +
geom_line(aes(Danish, gsn_surv, color = 'theoretical'), size = 1.2) +
geom_line(aes(Danish, gsn_surv_emp, color = 'empirical'), size = 1.2) +
scale_x_continuous(name="x", limits = c(0, 40)) +
labs(x="x",y=parse(text="F[x](x)"),
title="Danish fire loss data: survival functions") +
scale_color_discrete(name=expression("Type of density:")) +
theme(plot.title = element_text(hjust = 0.5))

######################################################################
#Plotting the theoretical PDF
ggplot(data.frame(Danish,gsn_pdf)) +
geom_histogram(aes(x = Danish, y = ..density..), color='black', fill='lightgrey', bins
    ↪ =80) +
scale_x_continuous(name="x", limits = c(0, 20)) +
geom_line(aes(Danish, gsn_pdf, color = 'GSN PDF'), size = 1.6) +
geom_line(aes(Danish, gsc_pdf, color = 'GSC PDF'), size = 1.6) +
#geom_line(aes(gp, dsn(gp, xi=mean, omega=sd, alpha=gamma1), color = 'SN PDF'), size
    ↪ =1) +
labs(x="x",y=parse(text="f[x](x)"),
title="Danish fire losses: density functions") +
scale_color_discrete(name=expression("Type of distribution:")) +
theme(plot.title = element_text(hjust = 0.5))
```