



Università degli studi di Cagliari

**DOTTORATO DI RICERCA
IN MATEMATICA E CALCOLO SCIENTIFICO**

Ciclo XXIV

**TWO GENERALIZATIONS OF THE
SKEW-NORMAL DISTRIBUTION
AND
TWO VARIANTS OF MCCARTHY'S THEOREM**

Settore scientifico disciplinare di afferenza:

PROBABILITÀ E STATISTICA MATEMATICA - MAT/06

Presentata da	Valentina Mamei
Coordinatore Dottorato	Prof.ssa Giuseppina D'Ambra
Relatore	Dott.ssa Monica Musio

Esame finale anno accademico 2010 - 2011

Abstract

The thesis is structured into two main parts. The first and major part is concerned with the skew-normal distribution, introduced by Azzalini (1985) [6], while the second one is connected with the scoring rules. In part one the problem of finding confidence intervals for the skewness parameter of the skew-normal distribution is addressed. Two new five-parameter continuous distributions which generalize the skew-normal distribution as well as some other well-known distributions are proposed and studied. Some mathematical properties of both distributions are derived. Part two is focused on the extension of the theorem of characterization of scoring rules, due to McCarthy (1956) ([16] of part 2), in two directions: for countable infinite sample spaces, but with bounded score and for finite sample spaces, but with unbounded score.

Declaration

I declare that to the best of my knowledge the contents of this thesis are original and my work except where indicated otherwise.

Acknowledgements

Firstly, I want to thank my supervisor Monica Musio for her guidance during this work. I am very grateful for her permanent support and constant encouragement.

I wish to thank Professor Philip Dawid for accepting me as visiting student at the Statistical Laboratory of Cambridge University and for his valuable advice.

I would like to express my gratitude to Professor Maria Luisa Targhetta for introducing me to the Statistics.

I am also thankful to all members of the Department of Mathematics of University of Cagliari. Among them I would like to thank, in particular, Stefano Montaldo, Renzo I. Caddeo, Gianluca Bande, Paola Piu and Todor V. Gramtchev.

Next I would like to give a special thanks to Giorgia Tranquilli, my colleague and friend.

I especially want to thank my boyfriend Mauro for his permanent encouragement and for always believing in me.

Finally, but most importantly, I would like to thank my family for their constant support over all these years.

Introduction

This thesis focuses on two important topics of the mathematical statistics: the first one is related to the skew-normal distribution, the second one concerns the fundamental characterization of proper scoring rules, given by McCarthy (1956) [16].

The study of the first issue has led to the writing of two articles [45], [48] and two manuscripts [46], [47] on which the first part of this thesis is based. Part I (chapters from 1 to 4) deals with the skew-normal distribution (SN), introduced by Azzalini (1985) [6], which has been studied and generalized extensively. This model is a class of distributions that extends the Gaussian family by including a skewness parameter (λ). Despite its nice properties, this family presents some inferential problems linked to the estimation of the skewness parameter. In particular, its maximum likelihood estimate can be infinite especially for moderate sample sizes and is not clear how to calculate confidence intervals for this parameter. The objective of the first part of the thesis is twofold. Firstly, it aims to present how these inferential problems can be solved in a particular situation. More specifically, we are interested in the distribution of the maximum or minimum of two random variables which have a bivariate normal distribution. Order statistics of correlated normal variables appear in statistical applications. In a number of situations, especially in medical and the environmental contexts, even if observations are

taken in pairs, interest centres on the maximum or minimum value of the observations. Loperfido (2002) [43] shows that the minimum or maximum of two random variables with same mean and variance, whose distribution is jointly normal, is skew-normal with skewness parameter that can be expressed as a function of the correlation coefficient between the two initial variables. In this specific case we use the MLE of the correlation coefficient between the two initial variables to find the MLE of the corresponding skewness parameter of the skew-normal. Using the Fisher transformation ([37], [38]) we approximate the distribution of the skewness parameter λ and we are able to test hypotheses and to compute confidence intervals for λ . These theoretical intervals are then compared with parametric bootstrap intervals by means of a simulation study.

Secondly, it presents two new families of distributions which generalize the skew-normal one: the Beta skew-normal (*BSN*) and the Kumaraswamy skew-normal (*KwSN*) distributions. The *BSN*, which is flexible enough to support both unimodal and bimodal shape, arises naturally when we consider the distributions of order statistics of the skew-normal distribution. The Beta skew-normal can also be obtained as a special case of the Beta-generated distribution introduced by Jones (2004) [35]. The idea of Beta-generated family of distributions stemmed from the paper of Eugene et al. (2002) [23], wherein the Beta-normal distribution was introduced and its properties were studied. Some other Beta-generated families of distributions have also been discussed in the literature. For example, the Beta half-normal distribution has been defined and studied by Pescim et al. (2010) [52]. Kong et al. (2007) [39] presented results on the Beta-gamma distribution. All these works lead to some mathematical difficulties because the Beta distribution is not fairly tractable and, in particular, its cumulative distribution function involves the

incomplete Beta ratio. Following the idea of the class of Beta-generated distributions [35], Cordeiro and de Castro (2011) [17] proposed a new family of generalized distributions, called Kumaraswamy generalized family, by means of the Kumaraswamy distribution [40]. Some mathematical properties of the Kumaraswamy generalized family, derived by Cordeiro and de Castro (2011) [17], are usually much simpler than those properties of the Beta-generated class. In the same paper, they introduced some generalized distributions among these the Kumaraswamy-normal and the Kumaraswamy-gamma distributions. The Kumaraswamy generalized half-normal distribution has been defined and studied by Cordeiro et al. (2012) [19].

We use the cited generator approach suggested by Cordeiro and de Castro (2011) [17] to define a new model, called the Kumaraswamy skew-normal distribution, which extends the skew-normal one. We provide a comprehensive mathematical treatment of the new distribution and provide expansions for its distribution and density functions. Further we pay attention to three other generalizations of the skew-normal distribution: the Balakrishnan skew-normal (*SNB*) (Balakrishnan (2002) [10] as discussed by Arnold and Beaver (2002) [5], Gupta and Gupta (2004) [31], Sharafi and Behboodan (2008) [57]), the generalized Balakrishnan skew-normal (*GBSN*) (Yadegari et al. (2008) [59]) and a two-parameter generalization of the Balakrishnan skew-normal (*TBSN*) (Bahrami et al. (2009) [9]). The above three extensions are related to the Beta skew-normal and the Kumaraswamy skew-normal distributions for particular values of the parameters.

Given a random sample X_1, \dots, X_n from a distribution $F(x)$, in general the distribution of the related order statistics does not belong to the family of $F(x)$. In this thesis we show that the maximum between the X_i 's from a Balakrishnan skew-normal with parameters m and 1, denoted by $X_i \sim SNB_m(1)$,

is still a Balakrishnan skew-normal with parameters k and 1, where k is a function of m and n . Motivated by the well-known bounds for the moments [30], [33], and the variance of the order statistics [51], we study the problem of finding bounds for the moments and the variance of the Beta-generated family.

We obtain, inspired by the paper of Gupta and Nadarajah (2005) [32], general expressions for the moments of the Beta skew-normal and the Kumaraswamy skew-normal distributions. We introduce a bivariate Kumaraswamy skew-normal distribution (*BKwSN*) whose marginals are Kumaraswamy skew-normal distributions. This new distribution has been obtained from Frank's copula. The open-source software R is used extensively in implementing our results.

Part II (chapter 5) is dedicated to the scoring rules, which have been studied widely in statistical decision theory. They provide summary measures for the evaluation of probabilistic forecasts, by assigning a numerical score based on the forecast and on the event or value that materializes. More formally, a scoring rule $S(x, Q)$ is a loss function measuring the quality of a quoted distribution Q , for an uncertain quantity X , when the realized value of X is x . It is proper if it encourages honesty in the sense that the expected score $E_{X \sim P} S(x, Q)$, where X has distribution P , is minimized by the choice $Q = P$. McCarthy (1956) [16] fully characterizes the proper scoring rules, on finite sample spaces, as super-gradient of concave functions. This is formally proved in Hendrickson and Buehler (1971) [13].

Our main purpose is to generalize McCarthy's theorem for infinite countable spaces with bounded score and for finite spaces with infinite score. There are several other works generalizing, in one way or other, the classical McCarthy's theorem, as for instance Fang et al. (2010) [7], Hendrickson and

Buehler (1971) [13], Gneiting and Raftery (2007) [10].

The thesis is divided into six chapters and one appendix. The organization is as follows. In chapter 1 we remind several theoretical concepts used throughout the first part of the thesis. Note that only the most important definitions and properties are stated. In the first section of chapter 1 we describe briefly the skew-normal distribution and we list their most important properties. In the second one some generalizations of the skew-normal distribution are given. In the third one we remember the family of the Beta-generated distributions and also some particular cases of this family. In section 4 we present the Kumaraswamy generalized family and some special distributions of this class. Finally, the family of generalized Beta-generated distributions is illustrated and some examples are provided in the last section.

In chapter 2 the problem of finding the confidence intervals for the skewness parameter is addressed. In section 1 we utilize Fisher's transformation to construct test and confidence intervals for the skewness parameter. Section 2 is devoted to the construction of confidence intervals for the skewness parameter using the parametric bootstrap. In section 3 the results of a simulation study, which we conducted to compare confidence intervals obtained using both methods, are summarized. In section 4 we apply the proposed methodology to find approximate and bootstrap confidence intervals for the skewness parameter using data from a monitoring survey in Cagliari (Italy) and from a follow-up dataset of patients operated for a renal cancer in Strasbourg (France).

In the first section of chapter 3 we define the Beta skew-normal distribution, we present its properties and some special cases. In particular, the *BSN* contains the Beta half-normal distribution (Pescim et al. (2010) [52]) as limiting case. Besides, we investigate its shape properties. We give miscellaneous

results about bimodality properties of the Beta skew-normal distribution. We derive its moment generating function and we also compute numerically the first moment, the variance, the skewness and the kurtosis. We present two different methods which allow to simulate a BSN distribution. We show that the distributions of order statistics from the skew-normal distribution are Beta skew-normal and are log-concave. Furthermore, in the second section of this chapter we give some results concerning the SNB distribution. In particular, we derive the exact distributions of the largest order statistic from $SNB_m(1)$ and the shortest order statistic from $SNB_m(-1)$. Moreover, we explore the relationships between the BSN distribution and the other generalizations of the skew-normal. In section 4 we find bounds for the moments and the variance of the Beta-generated family. The special case of the Beta skew-normal distribution is treated in more detail. A maximum likelihood estimation is addressed in the last section.

Chapter 4 deals with the Kumaraswamy skew-normal distribution. In section 1 its properties and some special cases are presented. The first moment, the variance, the skewness and the kurtosis are numerically computed. Two different methods to simulate a $KwSN$ distribution are given. The second section is devoted to miscellaneous results on the Kumaraswamy skew-normal, among these an interesting theorem which relates the $KwSN$ and the normal distributions. The parameters of the new model are estimated by maximum likelihood in the third section. In the fourth section we considered the bivariate Kumaraswamy skew-normal distribution $BKwSN$ whose marginals are Kumaraswamy skew-normal distributions. We have established several properties of the proposed bivariate distribution using the properties of Frank's copula. In the last section we propose a generalization which nests both the Beta skew-normal and the Kumaraswamy skew-normal distributions.

The first section of chapter 5 is devoted to important definitions, properties and theorems issued from convex analysis, which are intensively used in the rest of the thesis. In the second section the most valuable and interesting features of scoring rules are mentioned. The concept of entropy related to a specific decision problem and the discrepancy function are introduced. Furthermore, we state McCarthy's theorem [16] as presented by Dawid et al. (2011) [4]. In the third section we remind two important characterization theorems, due to Grünwald and Dawid [11], the first one deals with bounded loss functions, while the second one is referred to unbounded loss. Finally, the last section focuses on two variants of McCarthy's theorem. In the final chapter our most important findings are summarized. Appendix contains a short description of the Lambert W function. The Bibliography appears at the end of each part.

Contents

Abstract	iii
Declaration	v
Acknowledgements	vii
Introduction	xv
List of Figures	xix
List of Tables	xxi
I	1
1 Literature Review	3
1.1 The skew-normal density	4
1.2 Skew-normal generalizations	9
1.3 The class of the Beta-generated distributions	15
1.3.1 Definition of the family	16
1.3.2 Expansion for the density function	17
1.3.3 Some special cases	18
1.4 The class of Kumaraswamy generalized distributions	24

1.4.1	Expansion of the density function	25
1.4.2	Some special cases	25
1.5	The class of the generalized Beta-generated distributions	29
1.5.1	Some special cases	30
2	Large sample confidence intervals for the skewness parameter	35
2.1	Approximate Confidence Intervals (ACI) for skewness parameter	36
2.2	Parametric Bootstrap Confidence Intervals (BCI)	38
2.3	Simulation study	39
2.4	Examples	45
2.4.1	PM ₁₀ concentrations	45
2.4.2	Creatinine clearance	47
3	The Beta skew-normal distribution	49
3.1	The Beta skew-normal	50
3.1.1	Definition and simple properties	50
3.1.2	Moment generating function and moments	54
3.1.3	Order statistics from the skew-normal distribution	59
3.1.4	Some interesting properties	60
3.1.5	Bimodal properties	62
3.1.6	Expansion for the density function	64
3.1.7	The $BSN(1, n, b)$	68
3.2	Further results	69
3.2.1	Skewing mechanism	73
3.3	Bounds of the moments and the variance of the Beta-generated distribution	74
3.3.1	Bounds of the moments	75

3.3.2	Bounds of the variance of the Beta-generated distribution	75
3.4	Maximum likelihood estimation	81
4	The Kumaraswamy skew-normal distribution	85
4.1	The Kumaraswamy skew-normal distribution	86
4.1.1	Definition and simple properties	86
4.1.2	Moment generating function and moments	88
4.1.3	Some interesting properties	91
4.1.4	Expansion for the density function	93
4.1.5	The $KwSN(1, n, b)$	97
4.2	Further results	98
4.2.1	An interesting theorem	100
4.3	Maximum likelihood estimation	106
4.4	Copulas	109
4.4.1	Definitions and basic properties	109
4.4.2	The bivariate Kumaraswamy skew-normal	113
4.5	The generalized Beta skew-normal distribution	119
4.5.1	Moment generating function and moments	122
4.5.2	New properties of the \mathcal{GBG} distribution	123
	Bibliography	125
II		133
5	Scoring rules	135
5.1	Basic concepts	135
5.1.1	Concave functions	136
5.1.2	Banach spaces	140

5.2	Scoring rules	141
5.2.1	Decision problems	141
5.2.2	Finite outcomes	142
5.2.3	Examples of proper scoring rules	143
5.2.4	Homogeneous scoring rules	145
5.3	Conjugacy	147
5.3.1	Finite Loss	150
5.3.2	Infinite Loss	150
5.4	New results	153
5.4.1	Infinite countable sample space	153
5.4.2	Finite sample space	155
	Bibliography	159
		163
6	Conclusions	163
A	The Lambert W function	167

List of Figures

1.1	<i>The $SN(\lambda)$ for different values of λ</i>	7
1.2	<i>The $SNB_n(\lambda)$ for different values of n and $\lambda = 1$</i>	11
1.3	<i>The $GBSN_{n,m}(\lambda)$ for different values of n, m and $\lambda = 1$</i>	13
1.4	<i>The $TBSN_{n,m}(\lambda_1, \lambda_2)$ for different values of $n, m, \lambda_1 = 1$ and $\lambda_2 = -1$</i>	15
1.5	<i>The $BN(a, b)$ for different values of a and b</i>	19
1.6	<i>The $BN(a, b)$ for different values of a and b</i>	20
1.7	<i>The $BHN(a, b)$ for different values of a and b</i>	22
1.8	<i>The $BG(a, b, \alpha, \beta)$ for different values of a, b, α and β</i>	23
1.9	<i>The $KwN(a, b)$ for different values of a and b</i>	26
1.10	<i>The $KwGa(a, b, \alpha, \beta)$ for different values of a, b, α and β</i>	27
1.11	<i>The $KwHN(a, b)$ for different values of a and b</i>	28
1.12	<i>The GBN density function for selected parameter values with $\mu = 0$ and $\sigma = 1$</i>	31
1.13	<i>The \mathcal{GBGa} density function for selected parameter values with $c = 2.5$ and $\beta = 1.5$</i>	32
1.14	<i>The $GBHN$ density function for selected parameter values with $c = 2.5$</i>	33
2.1	<i>Results of the simulation study</i>	43

3.1	Generated samples of the <i>BSN</i> distribution for some values of λ , a and b	53
3.2	The <i>BSN</i> (λ, a, b) for values of $a \geq 1$ and $b \geq 1$	56
3.3	The <i>BSN</i> (λ, a, b) for values of $a < 1$ and $b < 1$	57
4.1	The <i>KwSN</i> (λ, a, b) for different values of λ , a and b	88
4.2	The function $b \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right) \left(\frac{1}{2}\right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right)$	104
4.3	The function $W_0 \left(b \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right) \left(\frac{1}{2}\right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right)\right)$	104
4.4	The function $W_{-1} \left(b \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right) \left(\frac{1}{2}\right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right)\right)$	105
4.5	The univariate function of b obtained replacing (4.35)	105
4.6	The univariate function of b obtained replacing (4.36)	106
4.7	The <i>BKwSN</i> (1, 1, 1, 1, 1) density function	118
4.8	The <i>BKwSN</i> (-1, 1, 2, 4, 1, 2) density function	118
4.9	The <i>GBSN</i> density for different values of a , b , c and $\lambda = 1$	121
4.10	The <i>GBSN</i> density for different values of c , λ , $a < 1$ and $b < 1$	121
5.1	The Brier entropy	144
5.2	The Shannon entropy	145
A.1	The solid line shows W_0 and the dashed line W_{-1}	168

List of Tables

2.1	Results of the simulations with $\rho = 0.5$	44
2.2	Results of the simulations with $\rho = -0.5$	44
2.3	ACI and BCI of level 0.95 for λ using data from example 1 and 2	47
3.1	The first moment, the standard deviation, the skewness and the kurtosis of $BSN(\lambda, a, b)$ for different values of a, b and λ .	58
3.2	The variance of the $BSN(\lambda, a, b)$ and $\sigma_b^2(a)Var(Y)$ for different values of a, b and λ	80
4.1	The first moment, the standard deviation, the skewness and the kurtosis of $KwSN(\lambda, a, b)$ for different values of a, b and λ	90

Part I

Chapter 1

Literature Review

In this chapter we present some definitions and results known in the literature that will be useful later. In section 1 we introduce the definition and the main features of the skew-normal distribution. In section 2 we present some generalizations of the skew-normal distribution and list their key properties. The family of the Beta-generated distributions is introduced in section 3. Furthermore, in this section some models of this family are treated in detail. Section 4 presents the Kumaraswamy generalized family and provides some special models. Finally, in the last section we describe the class of the generalized Beta-generated distributions and include some examples.

Since most of the results are well-known in literature on this subject, we will not provide proofs. Throughout the thesis the notation \sim means “follows” or “has the distribution”, the notation $\stackrel{d}{=}$ indicates “equivalent in distribution”. Let us denote by pdf and cdf the density and the distribution function, respectively. The notation $N(0,1)$ is used to denote the normal distribution.

1.1 The skew-normal density

The skew-normal distribution refers to a parametric class of probability distributions which includes the standard normal as a special case. A random variable Z is said to be skew-normal with parameter λ , if its density function is

$$\phi(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad \text{with } \lambda, z \in \mathbb{R}, \quad (1.1)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution, respectively. We denote a random variable Z with the above density by $Z \sim SN(\lambda)$. The parameter λ controls skewness. The standard normal distribution is a skew-normal distribution with $\lambda = 0$.

We remind some properties of the SN distribution.

Properties of $SN(\lambda)$:

- a. As $\lambda \rightarrow \infty$, $\phi(z; \lambda)$ tends to the half-normal density.
- b. If Z is a $SN(\lambda)$ random variable, then $-Z$ is a $SN(-\lambda)$ random variable.
- c. If $Z \sim SN(\lambda)$, then $Z^2 \sim \chi^2(1)$.
- d. The density (1.1) is strongly unimodal, i.e. $\log \phi(z; \lambda)$ is a concave function of z .

The corresponding distribution function is

$$\Phi(z; \lambda) = 2 \int_{-\infty}^z \int_{-\infty}^{\lambda t} \phi(t)\phi(u)du dt = \Phi(z) - 2T(z; \lambda), \quad (1.2)$$

where $T(z; \lambda)$ is Owen's function. The properties of this function are:

1. $-T(z; \lambda) = T(z; -\lambda)$;
2. $T(-z; \lambda) = T(z; \lambda)$;

$$3. 2T(z; 1) = \Phi(z)\Phi(-z);$$

$$4. T(0; \lambda) = \frac{1}{2\pi} \arctan(\lambda).$$

Using the properties of Owen's function, we have immediately the following ones:

Property 1. $1 - \Phi(-z; \lambda) = \Phi(z; -\lambda).$

Property 2. $\Phi(z; 1) = \Phi(z)^2.$

Property 3. $\Phi(z; \lambda) + \Phi(z; -\lambda) = 2\Phi(z).$

Property 4. $\Phi(0; \lambda) = \frac{1}{2} - \frac{1}{\pi} \arctan(\lambda).$

The moment generating function of a random variable Z with skew-normal distribution is

$$M(t) = 2 \exp\left\{\frac{t^2}{2}\right\} \Phi(\delta t), \quad \text{where } \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}. \quad (1.3)$$

After some algebra, it is easy to obtain the following expressions:

$$\begin{aligned} E(Z) &= d\delta, & \text{var}(Z) &= 1 - (d\delta)^2, \\ \gamma_1(Z) &= \frac{1}{2}(4 - \pi) \text{sign}(\lambda) \left(\frac{E(Z)^2}{\text{var}(Z)}\right)^{\frac{3}{2}}, & \gamma_2(Z) &= 2(\pi - 3) \left(\frac{E(Z)^2}{\text{var}(Z)}\right)^2, \end{aligned}$$

where $d = \sqrt{\frac{2}{\pi}}$, and γ_1 , γ_2 denote the third and fourth standardized cumulants, respectively.

The maximum value of γ_1 is about 0.995, while for γ_2 it is 0.869.

The following proposition, due to Chiogna (1998) [14], generalizes the well known lemma:

Lemma 1. *If W is a normal random variable, then $E(\Phi(hW + k)) = \Phi\left(\frac{k}{\sqrt{1+h^2}}\right)$, for any real h, k .*

Proposition 1.

1. *If Z has distribution $SN(\lambda)$, then the random variable*

$Y = \Phi(hZ + k)$ has first moment:

$$E_Z(\Phi(hZ + k)) = \Phi\left(\frac{k}{\sqrt{1+h^2}}; \mu(h, \lambda)\right),$$

where $\mu(h, \lambda) = -\frac{h\lambda}{\sqrt{1+h^2+\lambda^2}}$.

2. *Analogously, if W has distribution $N(0, 1)$, then the random variable*

$Y = \Phi(hW + k; \lambda)$ has first moment:

$$E_W(\Phi(hW + k; \lambda)) = \Phi\left(\frac{k}{\sqrt{1+h^2}}; \mu(h, \lambda)\right),$$

where $\mu(h, \lambda) = \frac{\lambda}{\sqrt{1+h^2(1+\lambda^2)}}$.

Remark 1. *Lemma 1 can be used to prove that the skew-normal density function is a proper density and to derive the moment generating function (1.3).*

The class of skew-normal distributions can be generalized by the inclusion of the location and scale parameters which we identify as ξ and $\psi > 0$. Thus if $X \sim SN(\lambda)$, then $Y = \xi + \psi X$ is a skew-normal variable with parameters ξ , ψ and λ . We denote Y by $Y \sim SN(\xi, \psi, \lambda)$.

Plots of the skew-normal density function are illustrated in figure 1.1 for selected values of λ .

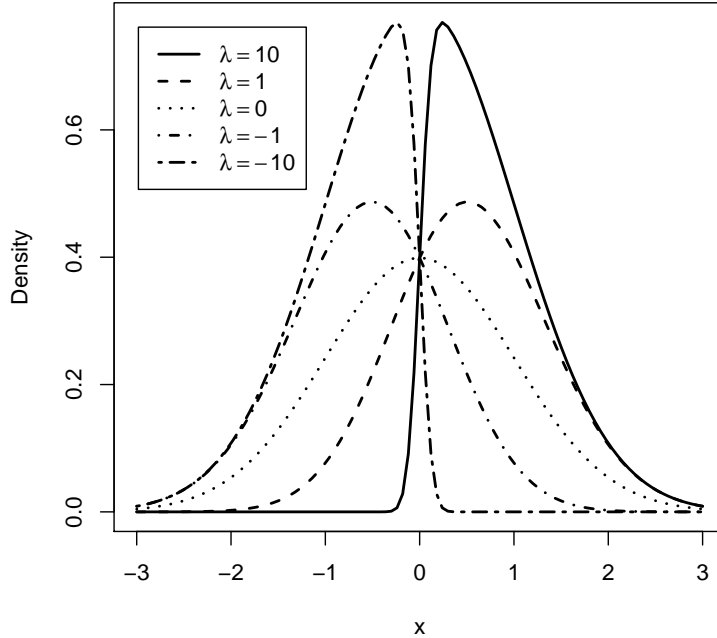


Figure 1.1: The $SN(\lambda)$ for different values of λ

Let us remind a result obtained by Loperfido (2008) [44], which will be useful in the second chapter. In [44] he has shown that any weighted average of the extremes of an exchangeable and bivariate normal random vector is skew-normal. More specifically, the following result holds:

Theorem 1. *Let X and Y be two random variables whose joint distribution is bivariate normal with $\mu_X = \mu_Y = \xi$, $\sigma_X^2 = \sigma_Y^2 = \psi^2$ and $Cov(X, Y) = \rho\psi^2$. Then for any two constants h and $k \neq -h$ the distribution of*

$$h \min(X, Y) + k \max(X, Y)$$

is

$$SN \left[\xi(h+k), \psi \sqrt{(h^2 + k^2 + 2\rho hk)}, \lambda = \frac{k-h}{|k+h|} \sqrt{\frac{1-\rho}{1+\rho}} \right].$$

In particular, for the choice $k = 0$ and $h = 1$, we find that the distribution of $\min(X, Y)$ is

$$SN \left[\xi, \psi, \lambda = -\sqrt{\frac{1-\rho}{1+\rho}} \right], \quad (1.4)$$

while, for $h = 0$ and $k = 1$, we have that the distribution of $\max(X, Y)$ is

$$SN \left[\xi, \psi, \lambda = \sqrt{\frac{1-\rho}{1+\rho}} \right]. \quad (1.5)$$

A proof of theorem 1 can be found in Loperfido (2002) [43] for $\xi = 0$ and $\psi = 1$, and in the more general case in Loperfido (2008) [44].

Despite the nice properties of the SN , inferential problems arise in the estimation of the skewness parameter. More specifically, its maximum likelihood estimator (MLE) can take infinite values with positive probability, especially for small or moderate sample sizes. In addition, it is not clear how to calculate confidence intervals for this parameter. Furthermore, the method of moments can give even worse results. Several solutions have been proposed to solve these problems, using numerical approximation methods, in both a classical and a Bayesian approach. Azzalini and Capitanio (1999) [7] suggested stopping the log-likelihood maximization procedure when the log-likelihood function reaches a value not significantly lower than the maximum. Another solution was proposed by Sartori (2005) [54], who developed a method based on a second-order modification of the likelihood equation that never produces boundary estimates. Liseo and Loperfido (2006) [42] use a Bayesian approach which modifies the likelihood function with a Jeffreys prior for the skewness parameter. They also prove that such prior is proper.

1.2 Skew-normal generalizations

Several modifications of the original skew-normal density have been developed in literature. Among these we mention the Balakrishnan skew-normal (*SNB*), the generalized Balakrishnan skew-normal (*GBSN*) and the two-parameter Balakrishnan skew-normal (*TBSN*). These generalizations are briefly presented in this section.

Balakrishnan (2002) [10] proposed a generalization of the standard skew-normal distribution as follows:

Definition 1. *A random variable X has Balakrishnan skew-normal distribution, denoted by $SNB_n(\lambda)$, if it has the following density function, with $n \in \mathbb{N}$,*

$$f_n(x; \lambda) = c_n(\lambda) \phi(x) \Phi(\lambda x)^n, \quad x \in \mathbb{R}, \lambda \in \mathbb{R}. \quad (1.6)$$

The coefficient $c_n(\lambda)$, which is a function of n and the parameter λ , is given by

$$c_n(\lambda) = \frac{1}{\int_{-\infty}^{\infty} \phi(x) \Phi(\lambda x)^n dx} = \frac{1}{E(\Phi(\lambda U)^n)}, \quad (1.7)$$

where $U \sim N(0, 1)$.

For $n = 0$ and $n = 1$, the above density reduces to the standard normal and the skew-normal distributions, respectively.

We denote a random variable X with density (1.6) in the special case $n = 2$ by $X \sim NSN(\lambda)$ (Sharafi and Behboodian (2006) [56]).

Sharafi et al. (2008) [57] give the following theorems.

Theorem 2. *If U, U_1, U_2, \dots, U_n are i.i.d. $N(0, 1)$, then*

$$U | (\max(U_1, U_2, \dots, U_n) \leq \lambda U) \sim SNB_n(\lambda). \quad (1.8)$$

Theorem 3. *If $Y \sim N(0, 1)$ and $W \sim SNB_{n-1}(\lambda)$ are independent, then conditionally $W | (\lambda W > Y) \sim SNB_n(\lambda)$.*

Theorem 4. *The moment generating function of $X \sim SNB_n(\lambda)$ is*

$$M_X(t) = c_n(\lambda) e^{\frac{t^2}{2}} a_n(t, \lambda), \quad (1.9)$$

where

$$a_n(t, \lambda) = E(\Phi^n(\lambda V)), \quad V \sim N(t, 1). \quad (1.10)$$

From the previous theorem it is easy to find the following recursion formula:

$$E(X^k) = (k-1)E(X^{k-2}) + \frac{n}{\sqrt{2\pi}} \frac{\lambda}{(1+\lambda^2)^{\frac{k}{2}}} \frac{c_n(\lambda)}{c_{n-1}\left(\frac{\lambda}{\sqrt{1+\lambda^2}}\right)} E(W^{k-1}), \quad (1.11)$$

where $W \sim SNB_{n-1}\left(\frac{\lambda}{\sqrt{1+\lambda^2}}\right)$.

The following formulae are useful in the sequel:

- $\frac{1}{c_1(\lambda)} = \frac{1}{2}$;
- $\frac{1}{c_2(\lambda)} = \frac{1}{4} + \frac{1}{2\pi} \arcsin(\rho) = \frac{1}{\pi} \arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\right)$;
- $\frac{1}{c_3(\lambda)} = \frac{1}{8} + \frac{3}{4\pi} \arcsin(\rho)$;

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

For $n \geq 4$, there is no closed form for $c_n(\lambda)$.

The class of Balakrishnan skew-normal distributions can be generalized by the inclusion of the location and scale parameters which we identify as μ and $\sigma > 0$. Thus if $X \sim SNB_n(\lambda)$, then $Y = \mu + \sigma X$ is a Balakrishnan skew-normal variable with parameters μ , σ and λ . We denote Y by $Y \sim SNB_n(\mu, \sigma, \lambda)$.

Remark 2. *Sharafi and Behboodian (2008) [57] have shown that for $\lambda = 1$, (1.6) is the density function of the $(n+1)$ -th order statistic $X_{(n+1)}$ in a sample of size $n+1$ from $N(0, 1)$. Moreover, for $\lambda = -1$, (1.6) is the density function of the first order statistic $X_{(1)}$ in a sample of size $n+1$ from $N(0, 1)$.*

Graphical illustrations of (1.6) are shown in figure 1.2.

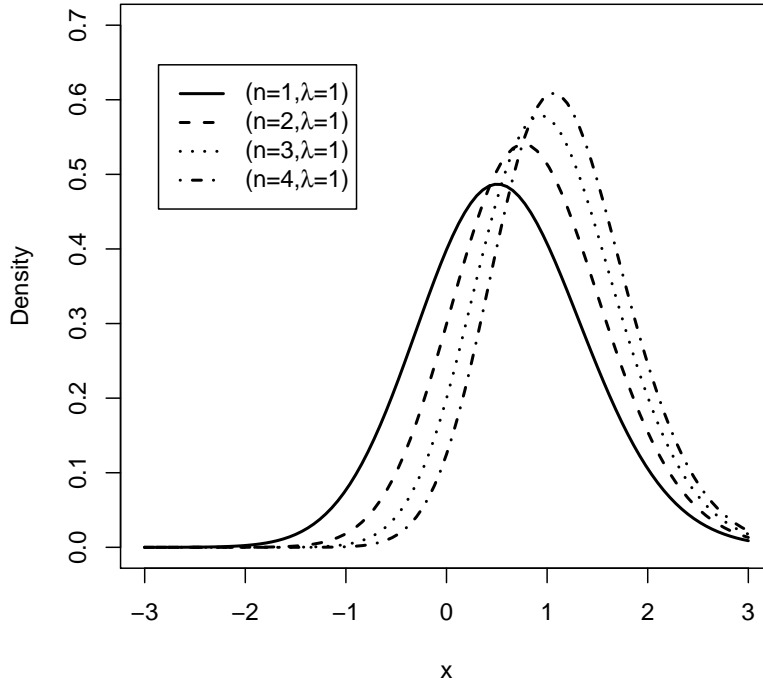


Figure 1.2: *The $SNB_n(\lambda)$ for different values of n and $\lambda = 1$*

Recently, Yadegari et al. (2008) [59] introduced the following generalization of the Balakrishnan skew-normal distribution and explained some important properties of this distribution.

Definition 2. *A random variable X is said to have a generalized Balakrishnan skew-normal distribution, denoted by $GBSN_{n,m}(\lambda)$, with parameters $n, m \in \mathbb{N}$ and $\lambda \in \mathbb{R}$, if its density function has the following form:*

$$f_{n,m}(x; \lambda) = \frac{1}{C_{n,m}(\lambda)} \phi(x) \Phi(\lambda x)^n (1 - \Phi(\lambda x))^m, \quad x \in \mathbb{R}, \quad (1.12)$$

where

$$C_{n,m}(\lambda) = \sum_{i=0}^m \binom{m}{i} (-1)^i \int_{-\infty}^{\infty} \phi(x) \Phi(\lambda x)^{n+i} dx. \quad (1.13)$$

For $m = 0$, this density reduces to the Balakrishnan skew-normal.

Remark 3. Let X_1, \dots, X_n be a random sample from a $N(0, 1)$. Then the j -th order statistic is a $GBSN_{j-1, n-j}(1)$, with $j = 1, \dots, n$. In this case we have that

$$C_{j-1, n-j}(1) = \sum_{i=0}^{n-j} \binom{n-j}{i} (-1)^i \int_{-\infty}^{\infty} \phi(x) \Phi(x)^{j-1+i} dx = \frac{n!}{(j-1)!(n-j)!}. \quad (1.14)$$

We recall some properties of this distribution, which have been studied by Yadegari et al. (2008) [59].

Property 5. $GBSN_{n,m}(\lambda) \stackrel{d}{=} GBSN_{m,n}(-\lambda)$ and $GBSN_{n,m}(-\lambda) \stackrel{d}{=} GBSN_{m,n}(\lambda)$.

Property 6. If $X \sim GBSN_{n,m}(\lambda)$, then $-X \sim GBSN_{n,m}(-\lambda)$. Moreover, for any $\lambda \neq 0$, $X \stackrel{d}{=} -X$ if and only if $n = m$.

Property 7. Let $X \sim GBSN_{n,m}(\lambda)$ be independent of a random sample Y_1, Y_2, \dots, Y_k from a normal distribution. Then $(Y_{(k)} \leq \lambda X) \sim GBSN_{n+k,m}(\lambda)$ and $X | (Y_{(1)} \geq X) \sim GBSN_{n,m+k}(\lambda)$, where $Y_{(k)}$ and $Y_{(1)}$ are the largest and the smallest order statistics, respectively.

Several generalized Balakrishnan skew-normal densities are illustrated in figure 1.3.

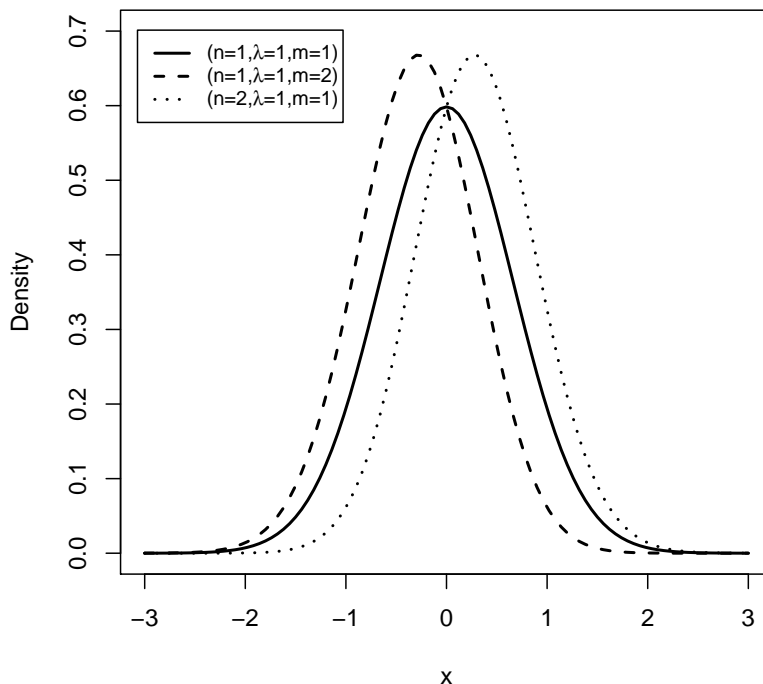


Figure 1.3: The $GBSN_{n,m}(\lambda)$ for different values of n , m and $\lambda = 1$

Bahrami et al. (2009) [9] discussed a two-parameter generalized skew-normal distribution which includes the skew-normal, the Balakrishnan skew-normal and the generalized Balakrishnan skew-normal as special cases.

Definition 3. A random variable Z has a two-parameter Balakrishnan skew-normal distribution with parameters $\lambda_1, \lambda_2 \in \mathbb{R}$, denoted by $T_{n,m}(\lambda_1, \lambda_2)$, if its pdf is

$$f_{n,m}(z; \lambda_1, \lambda_2) = \frac{1}{c_{n,m}(\lambda_1, \lambda_2)} \phi(z) \Phi(\lambda_1 z)^n \Phi(\lambda_2 z)^m, \quad z \in \mathbb{R}, \quad (1.15)$$

and n, m are non-negative integer numbers. The coefficient $c_{n,m}(\lambda_1, \lambda_2)$, which

is a function of the parameters n , m , λ_1 and λ_2 , is given by

$$c_{n,m}(\lambda_1, \lambda_2) = E(\Phi(\lambda_1 X)^n \Phi(\lambda_2 X)^m), \text{ where } X \sim N(0, 1). \quad (1.16)$$

The following properties are direct consequences of definition 3.

Properties of $T_{n,m}(\lambda_1, \lambda_2)$:

1. $TBSN_{1,1}(\lambda_1, 0) = SN(\lambda_1)$ and $TBSN_{1,1}(0, \lambda_2) = SN(\lambda_2)$;
2. $TBSN_{n,m}(\lambda, \lambda) = SNB_{n+m}(\lambda)$;
3. $TBSN_{n,m}(\lambda_1, 0) = SNB_n(\lambda_1)$ and $TBSN_{n,m}(0, \lambda_2) = SNB_m(\lambda_2)$;
4. $TBSN_{n,m}(\lambda, -\lambda) = GBSN_{n,m}(\lambda)$ and $TBSN_{n,m}(-\lambda, \lambda) = GBSN_{m,n}(\lambda)$;
5. $TBSN_{n,m}(0, 0) = TBSN_{0,0}(\lambda_1, \lambda_2) = N(0, 1)$.

Remark 4. Let Z_1, \dots, Z_n be i.i.d. $N(0, 1)$ and $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$ be the corresponding order statistics, then $Z_{(i)} \sim TBSN_{i-1, n-i}(1, -1)$.

Theorem 5. (Representation theorem) If $X, V_1, \dots, V_n, U_1, \dots, U_m$ are i.i.d. $N(0, 1)$ and let $V_{(n)} = \max(V_1, \dots, V_n)$ and $U_{(m)} = \max(U_1, \dots, U_m)$, then

$$X | (V_{(n)} < \lambda_1 X, U_{(m)} < \lambda_2 X) \sim TBSN_{n,m}(\lambda_1, \lambda_2).$$

Theorem 6. If $Z \sim TBSN_{n-1, m-1}(\lambda_1, \lambda_2)$ and Y_1, Y_2 i.i.d. $N(0, 1)$ are independent, then $Z | (Y_1 < \lambda_1 Z, Y_2 < \lambda_2 Z) \sim TBSN_{n,m}(\lambda_1, \lambda_2)$.

The location-scale two-parameter Balakrishnan skew-normal distribution is defined as the distribution of $Y = \mu + \sigma X$, where $X \sim TBSN_{n,m}(\lambda_1, \lambda_2)$. Hence, $\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameters, respectively. We denote Y by $Y \sim TBSN_{n,m}(\mu, \sigma, \lambda_1, \lambda_2)$.

Figure 1.4 illustrates some of the possible shapes of $TBSN_{n,m}(\lambda_1, \lambda_2)$ density function.

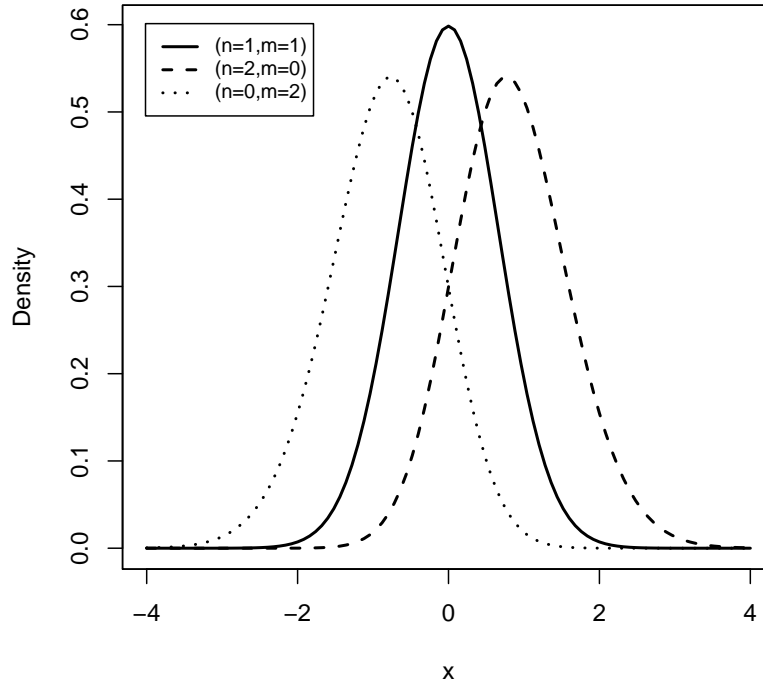


Figure 1.4: The $TBSN_{n,m}(\lambda_1, \lambda_2)$ for different values of n , m , $\lambda_1 = 1$ and $\lambda_2 = -1$

In the rest of the thesis we denote by $\phi(\cdot; \lambda)$ and $\Phi(\cdot; \lambda)$ the density and the distribution functions of the $SN(\lambda)$ distribution, respectively. The density function of $TBSN_{n,m}(\lambda_1, \lambda_2)$ will be indicated by $f_{n,m}(\cdot; \lambda_1, \lambda_2)$.

1.3 The class of the Beta-generated distributions

In this section the class of the Beta-generated distributions is described.

1.3.1 Definition of the family

The Beta distribution in its standard form ($Beta(a, b)$) is specified by its density function

$$f(x; a, b) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)}, \quad \text{for } 0 < x < 1, \quad a > 0 \text{ and } b > 0. \quad (1.17)$$

Starting from the Beta distribution, Jones (2004) [35] defined a new class of probability distributions, called Beta-generated family. Following the notation of Jones, the class of the Beta-generated distributions is defined as follows.

Definition 4. Let $F(\cdot)$ be a continuous distribution function with density function $f(\cdot)$. The univariate family of distributions generated by $F(\cdot)$, called Beta-generated family (Beta-F), with parameters $a > 0$ and $b > 0$, has pdf

$$g_{F(x)}^B(x; a, b) = \frac{1}{B(a, b)} (F(x))^{a-1} (1-F(x))^{b-1} f(x), \quad x \in \mathbb{R}, \quad (1.18)$$

where $B(a, b)$ is the complete Beta function.

Thus, this family of distributions has distribution function given by:

$$G_{F(x)}^B(x; a, b) = I_{F(x)}(a, b), \quad a, b > 0, \quad (1.19)$$

where the function $I_{F(x)}$ denotes the incomplete Beta ratio defined by

$$I_y(a, b) = \frac{B_y(a, b)}{B(a, b)}, \quad (1.20)$$

where

$$B_y(a, b) = \int_0^y z^{a-1} (1-z)^{b-1} dz, \quad 0 < y \leq 1, \quad (1.21)$$

is the incomplete Beta function. Replacing (1.20) and (1.21) in (1.19), we get that this family of distributions has distribution function

$$G_{F(x)}^B(x; a, b) = \frac{1}{B(a, b)} \int_0^{F(x)} z^{a-1} (1-z)^{b-1} dz. \quad (1.22)$$

Remark 5. Let $f(\cdot)$ be unimodal and continuously differentiable, if $a \geq 1$ and $b \geq 1$ then $g_{F(\cdot)}^B(\cdot; a, b)$ is also unimodal. The strong unimodality, i.e. log-concavity, of $f(\cdot)$ implies strong unimodality of $g_{F(\cdot)}^B(\cdot; a, b)$.

The density $g_{F(\cdot)}^B(\cdot; a, b)$ will be more tractable when both functions $F(\cdot)$ and $f(\cdot)$ have simple analytic expressions.

1.3.2 Expansion for the density function

Cordeiro and Lemonte (2011) [18] derived some properties of the *Beta*–*F* family using an important expansion for the density (1.18).

For $b > 0$ real non-integer,

$$(1-z)^{b-1} = \sum_{i=0}^{\infty} \frac{(-1)^i \Gamma(b)}{\Gamma(b-i) i!} z^i \quad (1.23)$$

is defined for $|z| < 1$. Replacing the above expansion into equation (1.22) if b is real non-integer, we have

$$\begin{aligned} G_F^B(x; a, b) &= \sum_{i=0}^{\infty} \frac{\Gamma(b)}{B(a, b)} \int_0^{F(x)} \frac{(-1)^i}{i! \Gamma(b-i)} z^{a+i-1} dz = \\ &= \sum_{i=0}^{\infty} \frac{\Gamma(b)}{B(a, b)} \frac{(-1)^i}{i! \Gamma(b-i)} \frac{F(x)^{a+i}}{a+i} = \\ &= \sum_{i=0}^{\infty} w_i(a, b) F(x)^{a+i}, \end{aligned} \quad (1.24)$$

where $w_i(a, b) = \frac{\Gamma(b)}{B(a, b)} \frac{(-1)^i}{i! \Gamma(b-i)} \frac{1}{a+i}$.

If b is an integer, the index i in the previous sum stops at $b-1$. If a is a real

non-integer, using (1.23) twice, $F(x)^{a+i}$ can be expressed as

$$\begin{aligned}
F(x)^{a+i} &= (1 - (1 - F(x)))^{a+i} = \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(a+i)}{\Gamma(a+i-k)k!} (1 - F(x))^k = \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(a+i)}{\Gamma(a+i-k)k!} \sum_{j=0}^k (-1)^j \binom{k}{j} F(x)^j = \\
&= \sum_{j=0}^{\infty} \sum_{k=j}^{\infty} \frac{(-1)^{k+j} \Gamma(a+i)}{\Gamma(a+i-k)k!} \binom{k}{j} F(x)^j = \\
&= \sum_{j=0}^{\infty} s_j(a+i) F(x)^j, \tag{1.25}
\end{aligned}$$

where $s_j(a+i) = \sum_{k=j}^{\infty} \frac{(-1)^{k+j} \Gamma(a+i)}{\Gamma(a+i-k)k!} \binom{k}{j}$. Consequently, the distribution function (1.24) becomes

$$\begin{aligned}
G_{F(x)}^B(x; a, b) &= \sum_{i=0}^{\infty} w_i(a, b) \sum_{j=0}^{\infty} s_j(a+i) F(x)^j = \\
&= \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} w_i(a, b) s_j(a+i) F(x)^j = \sum_{j=0}^{\infty} t_j(a, b) F(x)^j, \tag{1.26}
\end{aligned}$$

where $t_j(a, b) = \sum_{i=0}^{\infty} w_i(a, b) s_j(a+i)$.

Expansions for the Beta-generated density function can be obtained by simple differentiation of (1.24) for a integer and of (1.26) for a real non-integer.

1.3.3 Some special cases

We now present some special cases of the class of the Beta-generated distributions.

The Beta-normal distribution

When in (1.18) $F(x)$ is the normal distribution function with parameters μ and σ , we have the Beta-normal family, introduced by Eugene et al. [23], whose distribution function is given by

$$G_{\Phi\left(\frac{x-\mu}{\sigma}\right)}^B(x; a, b, \mu, \sigma) = \frac{1}{B(a, b)} \int_0^{\Phi\left(\frac{x-\mu}{\sigma}\right)} z^{a-1} (1-z)^{b-1} dz, \quad x \in \mathbb{R}, \tag{1.27}$$

and the corresponding probability density function is

$$g_{\Phi\left(\frac{x-\mu}{\sigma}\right)}^B(x; a, b, \mu, \sigma) = \frac{1}{B(a, b)} \left(\Phi\left(\frac{x-\mu}{\sigma}\right) \right)^{a-1} \left(1 - \Phi\left(\frac{x-\mu}{\sigma}\right) \right)^{b-1} \sigma^{-1} \phi\left(\frac{x-\mu}{\sigma}\right), \quad (1.28)$$

where $\sigma^{-1}\phi\left(\frac{x-\mu}{\sigma}\right)$ and $\Phi\left(\frac{x-\mu}{\sigma}\right)$ are the normal density and distribution with parameters μ and σ , respectively.

A random variable X with Beta-normal distribution with vector of parameters $\xi = (0, 1, a, b)$ is denoted by $X \sim BN(a, b)$.

The following figures plot the density function of the BN distribution for some values of a and b .

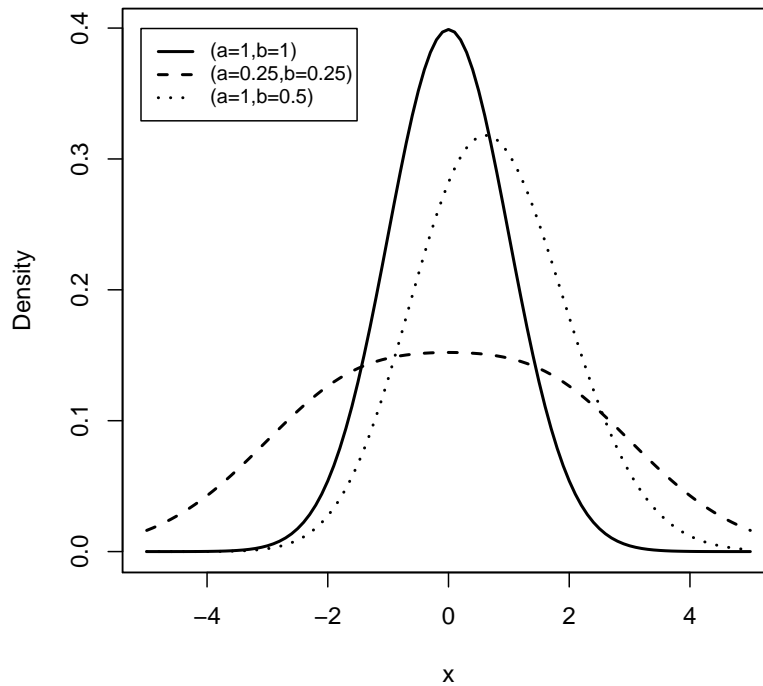


Figure 1.5: The $BN(a, b)$ for different values of a and b

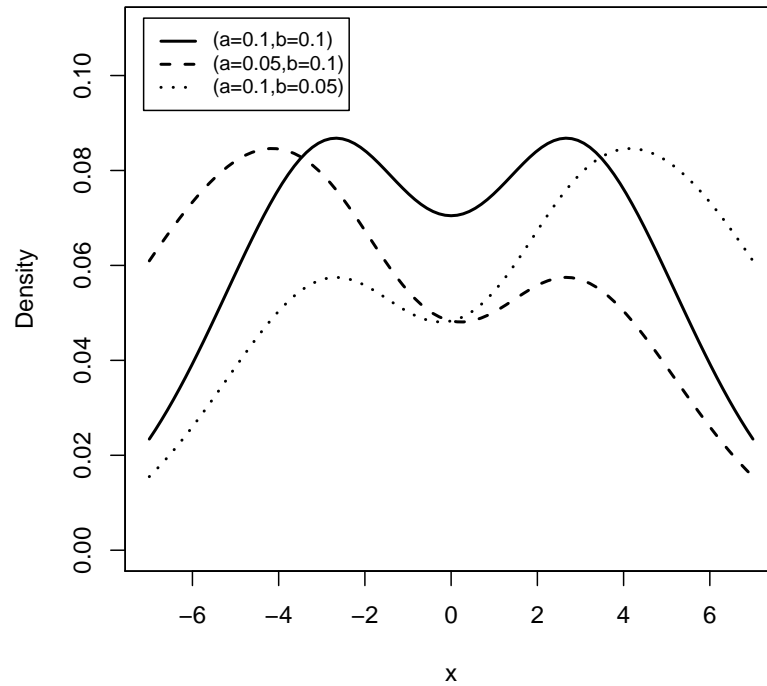


Figure 1.6: *The $BN(a,b)$ for different values of a and b*

Eugene et al. [23] showed that the Beta-normal distribution is symmetric about μ when $a = b$. Furthermore, they noted that this distribution can model both unimodal and bimodal data. Numerically, they found that when a and b are less than 0.214 the Beta-normal distribution is bimodal.

However if a and b are both larger than 0.214 the distribution is always unimodal.

Famoye et al. (2004) [24] studied the bimodality properties of the Beta-normal distribution.

In particular, they proved the following results.

Proposition 2. *A mode of the $BN(a, b, \mu, \sigma)$ is any point $x_0 = x_0(a, b)$ that satisfies*

$$x_0 = \sigma \left\{ (a-1) \frac{\phi\left(\frac{x_0-\mu}{\sigma}\right)}{\Phi\left(\frac{x_0-\mu}{\sigma}\right)} - (b-1) \frac{\phi\left(\frac{x_0-\mu}{\sigma}\right)}{1-\Phi\left(\frac{x_0-\mu}{\sigma}\right)} \right\} + \mu. \quad (1.29)$$

Corollary 1. *If $a = b$ and one mode of $BN(a, b, \mu, \sigma)$ is at x_0 , then the other mode is at the point $2\mu - x_0$.*

Corollary 2. *If $BN(a, b, \mu, \sigma)$ has a mode at x_0 , then $BN(b, a, \mu, \sigma)$ has a mode at $2\mu - x_0$.*

Corollary 3. *The modal point $x_0(a, b)$ is an increasing function of a and a decreasing function of b .*

Corollary 4. *The bimodal property of $BN(a, b, \mu, \sigma)$ is independent of the parameters μ and σ .*

The Beta half-normal distribution

The cdf of the half-normal distribution is $F(x) = 2\Phi(x) - 1$, with $x > 0$. By inserting $F(x)$ in (1.18), we obtain the Beta half-normal density function

$$g_{2\Phi(x)-1}^B(x; a, b) = \frac{2^b}{B(a, b)} (2\Phi(x) - 1)^{a-1} (1 - \Phi(x))^{b-1} \phi(x), \quad x > 0, \quad (1.30)$$

and the relative distribution function

$$G_{2\Phi(x)-1}^B(x; a, b) = \int_0^{2\Phi(x)-1} \frac{1}{B(a, b)} t^{a-1} (1-t)^{b-1} dt, \quad x > 0. \quad (1.31)$$

When X is a random variable following the BHN distribution, it is denoted by $X \sim BHN(a, b)$.

Remark 6. *The Beta half-normal distribution arises as a special case of the Beta generalized half-normal one studied by Pescim et al. (2010) [52].*

Figure 1.7 plots some shapes of the BHN distribution for some values of a and b .

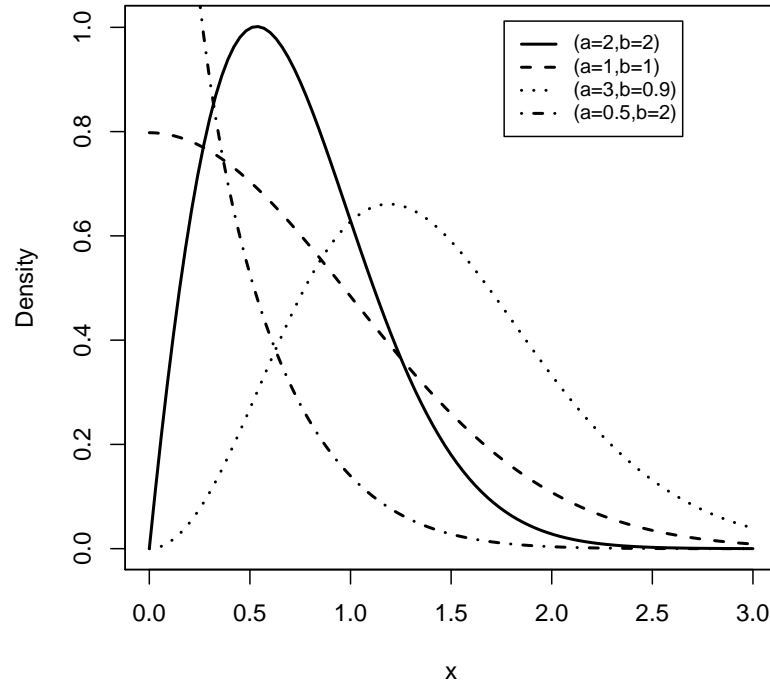


Figure 1.7: *The $BHN(a,b)$ for different values of a and b*

The Beta-gamma distribution

Let X be a gamma random variable with distribution function

$$F(x; \alpha, \beta) = \frac{\gamma_{\beta x}(\alpha)}{\Gamma(\alpha)}, \quad x > 0, \quad \alpha, \beta > 0, \quad (1.32)$$

where $\gamma_z(s) = \int_0^z t^{s-1} e^{-t} dt$ is the incomplete gamma function and $\Gamma(\cdot)$ is the gamma function.

The Beta-gamma cumulative distribution function is defined by substituting (1.32) into equation (1.19). Hence, the associated density function with four

positive parameters a , b , α and β has the form

$$g_{F(x;\alpha,\beta)}^B(x;a,b,\alpha,\beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{B(a,b)\Gamma(\alpha)^{a+b-1}} \gamma_{\beta x}(\alpha)^{a-1} (\Gamma(\alpha) - \gamma_{\beta x}(\alpha))^{b-1}, \quad (1.33)$$

with $x > 0$. A random variable Y with density function (1.33) is denoted by $Y \sim BG(a, b, \alpha, \beta)$.

Some properties of the Beta-gamma distribution are discussed in Kong et al. (2007) [39]. For $\alpha = \frac{\nu}{2}$ and $\beta = \frac{1}{2}$, the random variable Y has a Beta chi-square distribution that we will indicate by $B\chi^2(a, b, \nu)$.

Figure 1.8 illustrates several of the possible shapes obtained from (1.33) under various choices of a , b , α and β .

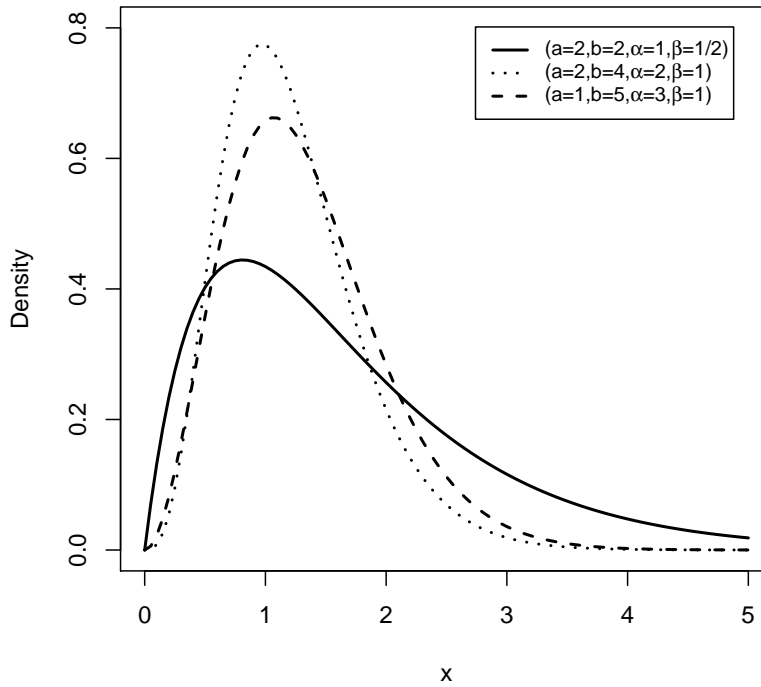


Figure 1.8: *The $BG(a, b, \alpha, \beta)$ for different values of a , b , α and β*

1.4 The class of Kumaraswamy generalized distributions

The Kumaraswamy distribution is not very common among statisticians and has been little studied in the literature. However, in a very recent paper, Jones (2009) [36] explored it, and he highlighted several advantages of this distribution over the Beta one. We can remember among them that the Kumaraswamy distribution ($Kw(a, b)$) has very easy cdf and quantile function, which do not involve any special functions, and imply a simple formula for random variate generation. Its cumulative distribution function is defined by

$$F(x; a, b) = 1 - (1 - x^a)^b, \quad 0 < x < 1, \quad (1.34)$$

where $a, b > 0$ are two additional parameters whose role is to introduce asymmetry and produce distributions with heavier tails. The probability density function is

$$f(x; a, b) = abx^{a-1}(1 - x^a)^{b-1}. \quad (1.35)$$

Cordeiro and de Castro (2011) [17] combined the works of Eugene et al. (2002) [23] and Jones (2004) [35] to construct a new class of distributions, called the Kumaraswamy generalized family ($Kw - F$).

From an arbitrary distribution function $F(x)$, the cdf of the $Kw - F$ is defined by

$$G_{F(x)}^K(x; a, b) = 1 - (1 - F(x)^a)^b, \quad (1.36)$$

where $a, b > 0$ are two additional parameters. Correspondingly, the density function of this family of distributions has a very simple form

$$g_{F(x)}^K(x; a, b) = abf(x)F(x)^{a-1}(1 - F(x)^a)^{b-1}. \quad (1.37)$$

Some structural properties of the $Kw - F$ distribution derived by Cordeiro and de Castro (2011) [17] are usually simpler than the corresponding properties of

the *Beta* – *F* distribution. They introduced some of these generalized forms but not discussed them in detail.

Remark 7. *If $f(\cdot)$ is a symmetric density function around 0, then $g_{F(\cdot)}^K(\cdot; a, b)$ will not be a symmetric even when $a = b$.*

1.4.1 Expansion of the density function

Using the expansion (1.23), the density function $g_{F(x)}^K(x; a, b)$, for $b > 0$ real non-integer, can be expanded as

$$g_{F(x)}^K(x; a, b) = f(x) \sum_{i=0}^{\infty} (-1)^i ab \binom{b-1}{i} F(x)^{a(i+1)-1} = f(x) \sum_{i=0}^{\infty} w_i(a, b) F(x)^{a(i+1)-1}, \quad (1.38)$$

where $w_i(a, b) = (-1)^i ab \binom{b-1}{i}$. If b is an integer, the index i in the previous sum stops at $b - 1$. If a is real non-integer, $F(x)^{a(i+1)-1}$ can be expanded as follows

$$F(x)^{a(i+1)-1} = [1 - (1 - F(x))]^{a(i+1)-1} = \sum_{j=0}^{\infty} (-1)^j \binom{a(i+1)-1}{j} (1 - F(x))^j, \quad (1.39)$$

and then

$$F(x)^{a(i+1)-1} = \sum_{j=0}^{\infty} \sum_{r=0}^j (-1)^{j+r} \binom{a(i+1)-1}{j} \binom{j}{r} F(x)^r. \quad (1.40)$$

Hence, the density $g_{F(x)}^K(x; a, b)$ can be rewritten in the form

$$g_{F(x)}^K(x; a, b) = f(x) \sum_{i,j=0}^{\infty} \sum_{r=0}^j (-1)^{i+j+r} ab \binom{b-1}{i} \binom{a(i+1)-1}{j} \binom{j}{r} F(x)^r. \quad (1.41)$$

1.4.2 Some special cases

Two examples, which are highlighted in [17], are the Kumaraswamy-normal and the Kumaraswamy-gamma distributions.

The Kumaraswamy-normal distribution

Replacing $f(\cdot)$ and $F(\cdot)$ into (1.37) by the pdf and the cdf of the normal distribution with parameters μ and σ , we obtain the Kumaraswamy-normal distribution (KwN) with density function given by

$$g_{\Phi\left(\frac{x-\mu}{\sigma}\right)}^K(x; a, b, \mu, \sigma) = ab\phi\left(\frac{x-\mu}{\sigma}\right)\left(\Phi\left(\frac{x-\mu}{\sigma}\right)\right)^{a-1}\left(1-\Phi\left(\frac{x-\mu}{\sigma}\right)\right)^{b-1},$$

where $x \in \mathbb{R}$, $a, b, \sigma > 0$, $\mu \in \mathbb{R}$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and the distribution functions of the standard normal distribution, respectively. For later reference, we denote a random variable X with the above pdf by $KwN(a, b, \mu, \sigma)$.

The KwN distribution with $a = 2$ and $b = 1$ reduces to the skew-normal distribution with shape parameter equal to one.

Different densities with $\mu = 0$ and $\sigma = 1$ are plotted in figure 1.9.

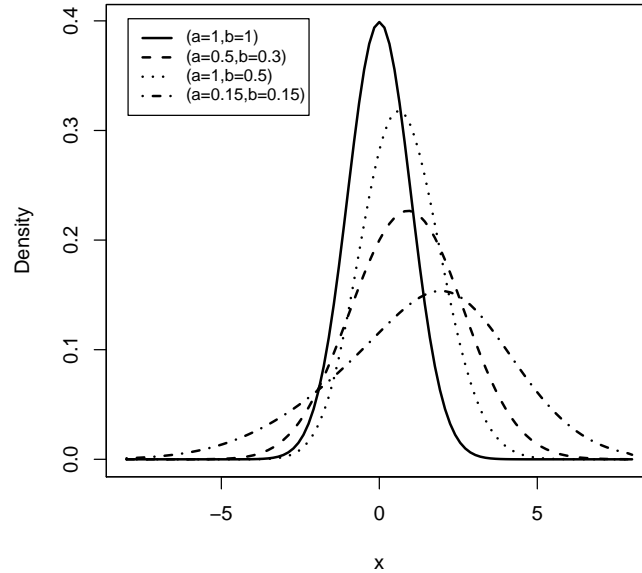


Figure 1.9: *The $KwN(a, b)$ for different values of a and b*

The Kumaraswamy-gamma distribution

Let $F(x)$ be the gamma distribution function with parameters α and β . The general form for the density of a random variable X following a Kumaraswamy-gamma distribution, say $X \sim KwGa(a, b, \alpha, \beta)$, can be expressed as

$$g_{F(x)}^K(x; a, b, \alpha, \beta) = ab \left(\frac{\Gamma_{\beta x}(\alpha)}{\Gamma(\alpha)} \right)^{a-1} \left(1 - \left(\frac{\Gamma_{\beta x}(\alpha)}{\Gamma(\alpha)} \right)^a \right)^{b-1} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0.$$

The Kumaraswamy-exponential distribution is obtained setting $\alpha = 1$. The $KwGa(1, b, 1, \beta)$ simplifies to the exponential distribution with parameter $b\beta$. We will denote the $KwGa(a, b, \frac{\nu}{2}, \frac{1}{2})$ by $Kw\chi^2(a, b, \nu)$.

Figure 1.10 illustrates some of the possible shapes of the $KwGa(a, b, \alpha, \beta)$ density for selected values of a , b , α and β .

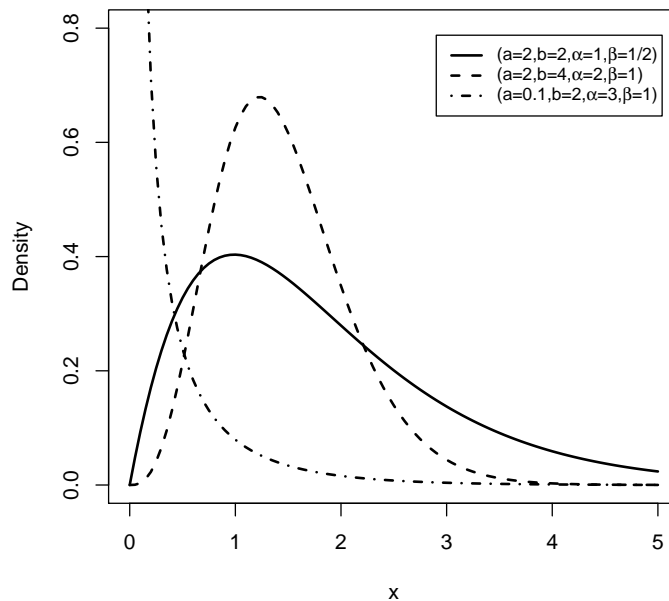


Figure 1.10: *The $KwGa(a, b, \alpha, \beta)$ for different values of a , b , α and β*

The Kumaraswamy half-normal distribution

Let $F(x)$ be the standard half-normal distribution function. The density of a random variable X having a Kumaraswamy half-normal distribution with parameters a and b , denoted by $KwHN(a,b)$, is

$$g_{F(x)}^k(x;a,b) = 2ab\phi(x)(2\Phi(x)-1)^{a-1}(1-(2\Phi(x)-1)^a)^{b-1}, \quad x > 0,$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ represent the density and distribution functions of the standard normal distribution, respectively.

The half-normal distribution arises as the particular case $a = b = 1$.

Figure 1.11 plots some densities of the $KwHN(a,b,\mu,\sigma)$ distribution with $\mu = 0$ and $\sigma = 1$.

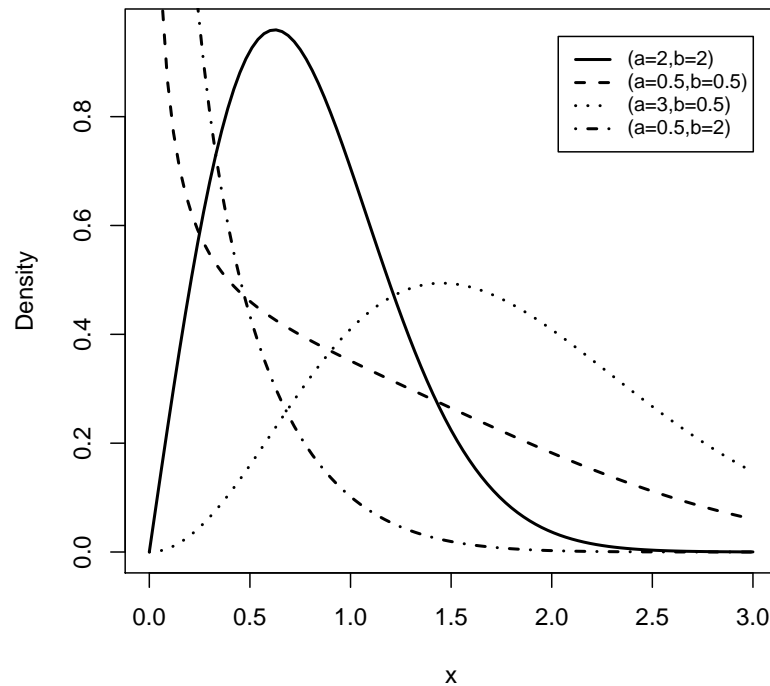


Figure 1.11: *The $KwHN(a,b)$ for different values of a and b*

Remark 8. We remind that the Kumaraswamy half-normal distribution is a special case of the Kumaraswamy generalized half-normal distribution, defined by Cordeiro et al. (2012) [19].

1.5 The class of the generalized Beta-generated distributions

The generalized Beta-generated distribution of the first kind was introduced by McDonald (1984) [49].

Definition 5. A variable X is said to have a generalized Beta-generated distribution of the first kind with positive parameters a , b and c , say $GB(a, b, c)$, if its density is given by

$$g(x; a, b, c) = \frac{cx^{ac-1}(1-x^c)^{b-1}}{B(a, b)}, \quad \text{with } 0 < x < 1. \quad (1.42)$$

If $c = 1$ the variable X is a Beta of the first kind with parameters a and b .

In the special case $a = 1$, equation (1.42) reduces to the Kumaraswamy distribution with parameters b and c .

Recently, Alexander et al. (2011) [3] proposed the class of the generalized Beta-generated distributions which is defined as follows.

For a continuous distribution function $F(x)$ with density $f(x)$, the family of the generalized Beta-generated distributions ($\mathcal{GBG} - F$) is characterized by its density:

$$g_{F(x)}^{\mathcal{GBG}}(x; a, b, c) = \frac{c}{B(a, b)} f(x) [F(x)^{ac-1} (1 - F(x)^c)^{b-1}]. \quad (1.43)$$

Two important special cases are the Beta-generated distribution ($c = 1$), and the Kumaraswamy generalized distribution ($a = 1$).

Remark 9. The \mathcal{GBG} distribution obtained from $F(x)$ is a standard Beta-generated distribution generated by $F(x)^c$.

1.5.1 Some special cases

We include two examples of the class of the generalized Beta-generated distributions given by Alexander et al. (2011) [3]: the generalized Beta-normal and the generalized Beta-gamma distributions.

Furthermore, we define a new distribution of this family useful for our purposes, called the generalized Beta half-normal distribution.

Generalized Beta-normal distribution

Replacing $F(\cdot)$ and $f(\cdot)$ into (1.43) by the cdf and the pdf of the normal distribution with parameters μ and σ , we obtain the generalized Beta-normal distribution (GBN) with density function given by

$$g_{\Phi\left(\frac{x-\mu}{\sigma}\right)}^{\mathcal{GBG}}(x; a, b, c, \mu, \sigma) = \frac{c}{B(a, b)} \phi\left(\frac{x-\mu}{\sigma}\right) \left(\Phi\left(\frac{x-\mu}{\sigma}\right)\right)^{ac-1} \left(1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right)^{b-1}, \quad (1.44)$$

where $x \in \mathbb{R}$, $a, b, c, \sigma > 0$, $\mu \in \mathbb{R}$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and the cdf of the standard normal distribution, respectively.

A random variable X following the GBN distribution is denoted by

$$X \sim GBN(a, b, c, \mu, \sigma).$$

Setting $c = 1$, (1.44) reduces to the Beta-normal distribution proposed by Eugene et al. (2002) [23].

When $a = 1$ the Kumaraswamy-normal is obtained. The GBN distribution with $\mu = 0$, $\sigma = 1$, $b = 1$ and $ac = 2$ coincides with the skew-normal distribution with shape parameter equal to one.

Plots of the *GBN* density function for selected parameter values are given in figure 1.12.

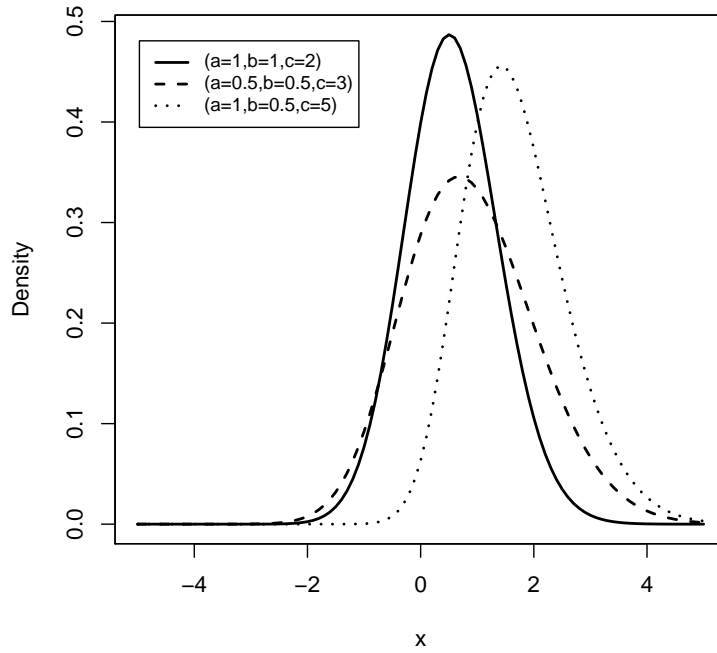


Figure 1.12: The *GBN* density function for selected parameter values with $\mu = 0$ and $\sigma = 1$

Generalized Beta-gamma distribution

By inserting (1.32) in (1.43), we obtain the generalized Beta-gamma distribution with five positive parameters a , b , c , α and β , whose density function is given by

$$g_{F(x;\alpha,\beta)}^{GBG}(x; a, b, c, \alpha, \beta) = \frac{c\beta^\alpha}{B(a, b+1)\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \left\{ \frac{\gamma_{\beta x}(\alpha)}{\Gamma(\alpha)} \right\}^{ac-1} \left\{ 1 - \frac{\gamma_{\beta x}(\alpha)^c}{\Gamma(\alpha)} \right\}^b, \quad (1.45)$$

where $x > 0$.

When Y is a random variable following the \mathcal{GBGa} distribution, it will be denoted by $Y \sim \mathcal{GBGa}(a, b, c, \alpha, \beta)$.

Figure 1.13 illustrates some possible shapes of the \mathcal{GBGa} density function.

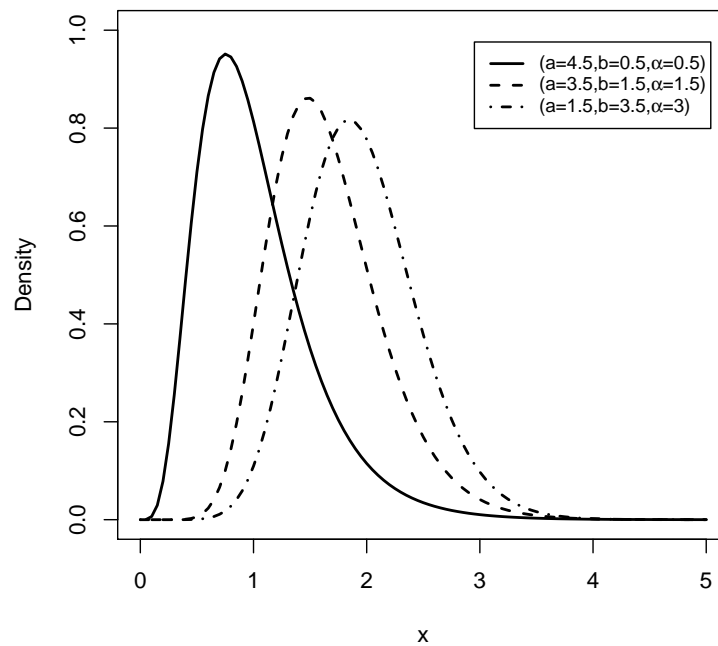


Figure 1.13: The \mathcal{GBGa} density function for selected parameter values with $c = 2.5$ and $\beta = 1.5$

Generalized Beta half-normal distribution

We now introduce the three-parameter generalized Beta half-normal ($GBHN$) distribution by taking $F(x)$ in (1.43) to be the cdf of the standard half-normal distribution. Hence, the density function of $GBHN$ distribution with three

parameters a , b and c has the form

$$g_{2\Phi(x)-1}^{\mathcal{GBG}}(x; a, b, c) = \frac{2c}{B(a, b)} \phi(x) (2\Phi(x) - 1)^{ac-1} [1 - (2\Phi(x) - 1)^c]^{b-1}, \quad x > 0. \quad (1.46)$$

If X is a random variable with density (1.46), we write $X \sim \mathbf{GBHN}(a, b, c)$.

Plots of the density function (1.46) are illustrated in figure 1.14.

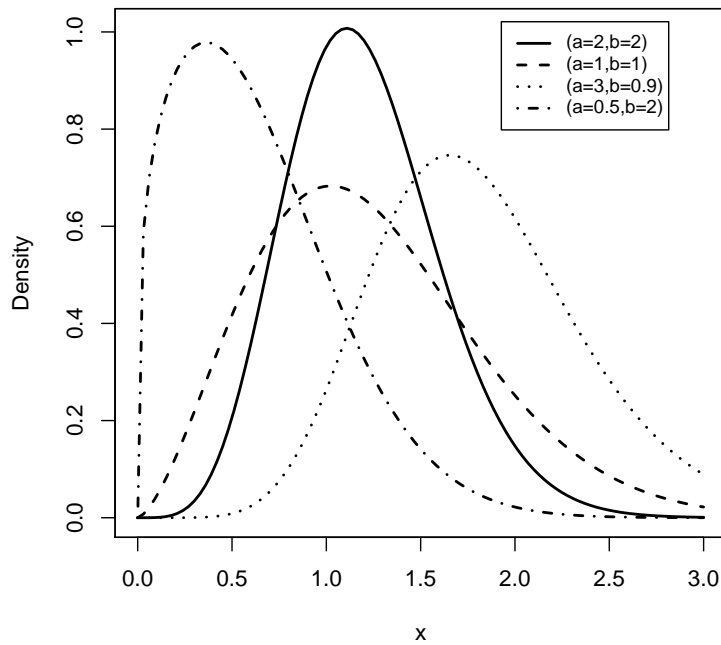


Figure 1.14: *The GBHN density function for selected parameter values with $c = 2.5$*

Chapter 2

Large sample confidence intervals for the skewness parameter

As evidenced in section 1 of chapter 1, the skew-normal model presents some inferential problems linked to the estimation of the skewness parameter. In particular, its maximum likelihood estimate can be infinite especially for moderate sample size and is not clear how to calculate confidence intervals for this parameter. In this chapter we show how these inferential problems of the skew-normal distribution can be solved if we are interested in the distribution of extreme statistics of two random variables with joint normal distribution. Loperfido proved (see theorem 1) that such extreme statistics have a skew-normal distribution with skewness parameter that can be expressed as a function of the correlation coefficient between the two initial variables. It is then possible, using some theoretical results involving the correlation coefficient, to find approximate confidence intervals for the parameter of skewness. These theoretical intervals are then compared with

parametric bootstrap intervals by means of a simulation study. Two applications are given using real data. These results are new and can be found in [48].

2.1 Approximate Confidence Intervals (ACI) for skewness parameter

We denote a random vector (X, Y) having a bivariate normal distribution by $(X, Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, where ρ is the correlation coefficient. Its density is then

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right\}.$$

As pointed out in the first section of chapter 1, Loperfido (2008) [44] has shown that any weighted average of the extremes of an exchangeable and bivariate normal random vector is skew-normal.

In particular, the distribution of $\min(X, Y)$ is

$$SN \left[\xi, \psi, \lambda = -\sqrt{\frac{1-\rho}{1+\rho}} \right], \quad (2.1)$$

and the distribution of $\max(X, Y)$ is

$$SN \left[\xi, \psi, \lambda = \sqrt{\frac{1-\rho}{1+\rho}} \right], \quad (2.2)$$

where $\xi = \mu_X = \mu_Y$ and $\psi^2 = \sigma_X^2 = \sigma_Y^2$.

Suppose now to be interested in constructing a confidence intervals for $\lambda = \sqrt{\frac{1-\rho}{1+\rho}}$ on the basis of a random sample of n pairs (X, Y) .

When all five parameters in $\theta = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ are assumed to be unknown, the MLE of the coefficient of correlation ρ is the sample correlation

coefficient

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

Using the invariance property of MLE, we know that the statistic $\hat{\lambda} = \sqrt{\frac{1-R}{1+R}}$ is the MLE of λ .

It is customary to base tests concerning ρ on the statistic $\frac{1}{2} \ln \left(\frac{1+R}{1-R} \right)$. This is the Fisher transformation of R (see for instance Kendall and Stuart [37], [38]). It can be shown (see for instance [37], [38]) that the distribution of this statistic, for $n > 50$, is approximately normal with mean $\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)$ and variance $\frac{1}{n-3}$. Then the variable

$$Z = \frac{\frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) - \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)}{\frac{1}{\sqrt{n-3}}} \quad (2.3)$$

has approximately standard normal distribution. Using the above approximation we can calculate $1 - \alpha$ confidence intervals for the parameter $\lambda = \sqrt{\frac{1-\rho}{1+\rho}}$.

We have:

$$P \left(-z_{\frac{\alpha}{2}} \leq \frac{\frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) - \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)}{\frac{1}{\sqrt{n-3}}} \leq z_{\frac{\alpha}{2}} \right) \approx 1 - \alpha,$$

which is equivalent to

$$P \left(\exp \left(\frac{-z_{\frac{\alpha}{2}}}{\sqrt{n-3}} - \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \right) \leq \lambda \leq \exp \left(\frac{z_{\frac{\alpha}{2}}}{\sqrt{n-3}} - \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \right) \right) \approx 1 - \alpha.$$

Then the random set

$$C(R) = \left[\exp \left(\frac{-z_{\frac{\alpha}{2}}}{\sqrt{n-3}} - \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \right), \exp \left(\frac{z_{\frac{\alpha}{2}}}{\sqrt{n-3}} - \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \right) \right] \quad (2.4)$$

is an approximate $1 - \alpha$ confidence interval for λ .

This approximation can also be used to test hypotheses concerning $\lambda = \sqrt{\frac{1-\rho}{1+\rho}}$.

If we are interested in testing

$$H_0 : \lambda = \lambda_0 \quad \text{versus} \quad H_1 : \lambda \neq \lambda_0, \quad (2.5)$$

we find that an appropriate critical region of size α for testing the null hypothesis against the alternative, is $|Z| \geq z_{\frac{\alpha}{2}}$, where Z is defined as in (2.3) and $z_{\frac{\alpha}{2}}$ is defined by $P(Z \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$. Then we can write the rejection region as

$$\left\{ r : \left| \frac{\frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) + \ln(\lambda_0)}{\frac{1}{\sqrt{n-3}}} \right| \geq z_{\frac{\alpha}{2}} \right\}. \quad (2.6)$$

The same procedure can be applied to compute the confidence intervals of level $1 - \alpha$ and the critical region for the hypothesis (2.5) for $\lambda = -\sqrt{\frac{1-\rho}{1+\rho}}$. For instance, an approximate confidence interval is

$$C(R) = \left[-\exp \left(\frac{z_{\frac{\alpha}{2}}}{\sqrt{n-3}} - \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \right), -\exp \left(\frac{-z_{\frac{\alpha}{2}}}{\sqrt{n-3}} - \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \right) \right].$$

Using theorem 1 of chapter 1 and these procedures, we can compute confidence intervals and critical regions for the unknown skewness parameter λ when the others unknown parameters (means and variances) of the random variables X and Y are estimated by the corresponding MLEs.

Note that the length of the $1 - \alpha$ confidence interval (2.4)

$$L(R, n) = \exp \left(-\frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \right) \left[\exp \left(\frac{z_{\frac{\alpha}{2}}}{\sqrt{n-3}} \right) - \exp \left(\frac{-z_{\frac{\alpha}{2}}}{\sqrt{n-3}} \right) \right]$$

is a decreasing function of R for fixed n and a decreasing function of n for fixed values of R . We expect to have shorter intervals for R close to 1 and for large samples.

2.2 Parametric Bootstrap Confidence Intervals (BCI)

In this section we use the parametric bootstrap method for constructing confidence intervals (see Efron and Tibshirani (1993) [22]). This method

relies on resampling with replacement from an estimated parametric model and calculating the required statistic from these repeated samples. The values of the statistic from the repeated sampling can then be used to generate standard errors and confidence intervals for the statistic of interest.

In our specific case, we consider a random sample of n pairs $Z = (X, Y)$, where $(X, Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$.

For the parametric bootstrap, instead of estimating the theoretical distribution function F by the empirical distribution function, we estimate the five parameters of the bivariate normal by the corresponding MLEs. We denote the bivariate normal distribution with these values for the parameters by \hat{F}_{norm} .

Suppose that our functional of interest is $\Theta = \Theta(F)$, which we estimate by the statistic: $\hat{\Theta} = \hat{\Theta}(Z_1, \dots, Z_n)$. In order to construct a confidence interval for Θ we introduce the bootstrap random variables $Z_1^*, Z_2^*, \dots, Z_n^*$ i.i.d. with distribution \hat{F}_{norm} . Then we generate B bootstrap samples from $Z_1^*, Z_2^*, \dots, Z_n^*$, denoted by $\mathbf{z}^{*1}, \mathbf{z}^{*2}, \dots, \mathbf{z}^{*B}$, and for each we compute the bootstrap replication $\hat{\Theta}^*(b) = \hat{\Theta}(\mathbf{z}^{*b})$, $b = 1, \dots, B$. Let $\hat{\Theta}_B^{(\alpha)}$ be the $100 \cdot \alpha$ -th empirical percentile of the $\hat{\Theta}^*(b)$ values, that is the $B \cdot \alpha$ -th value in the ordered list of the B replications of $\hat{\Theta}^*$. Likewise, let $\hat{\Theta}_B^{(1-\alpha)}$ be the $100 \cdot (1 - \alpha)$ -th empirical percentile. The approximate $1 - 2\alpha$ percentile interval is

$$[\hat{\Theta}_{\%,lo}, \hat{\Theta}_{\%,up}] \approx [\hat{\Theta}_B^{(\alpha)}, \hat{\Theta}_B^{(1-\alpha)}]. \quad (2.7)$$

In our case $\Theta = \lambda = \pm \sqrt{\frac{1-\rho}{1+\rho}}$ and $\hat{\Theta} = \hat{\Lambda} = \pm \sqrt{\frac{1-R}{1+R}}$.

2.3 Simulation study

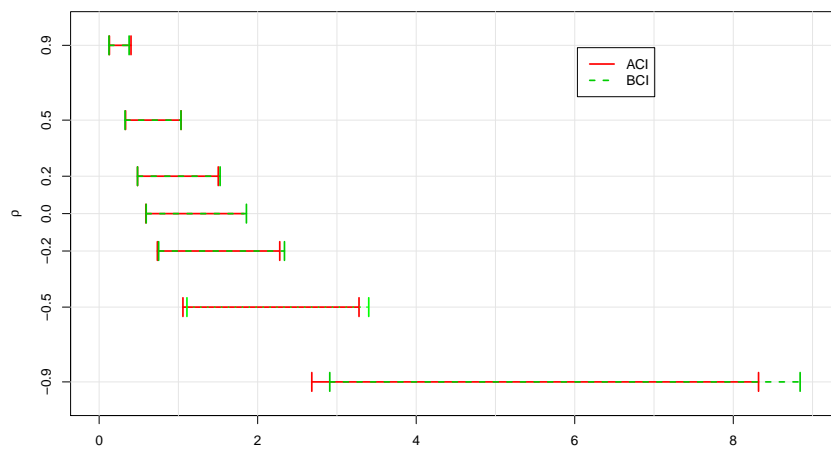
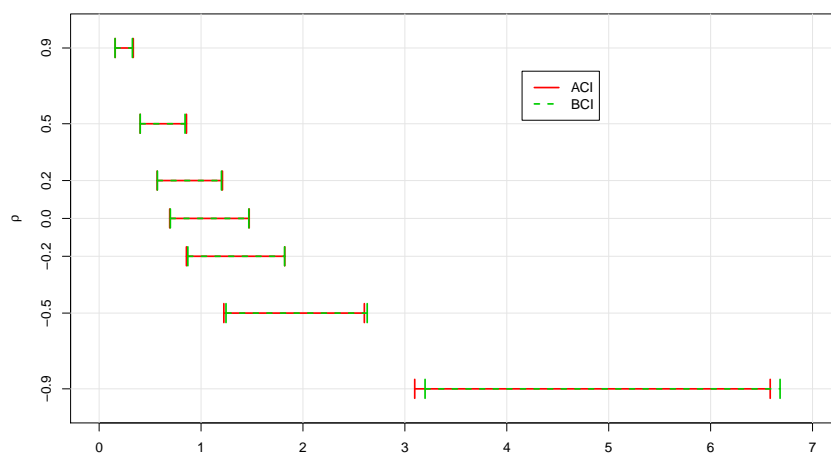
Typically, studies of the comparative performance of confidence intervals rely on simulations.

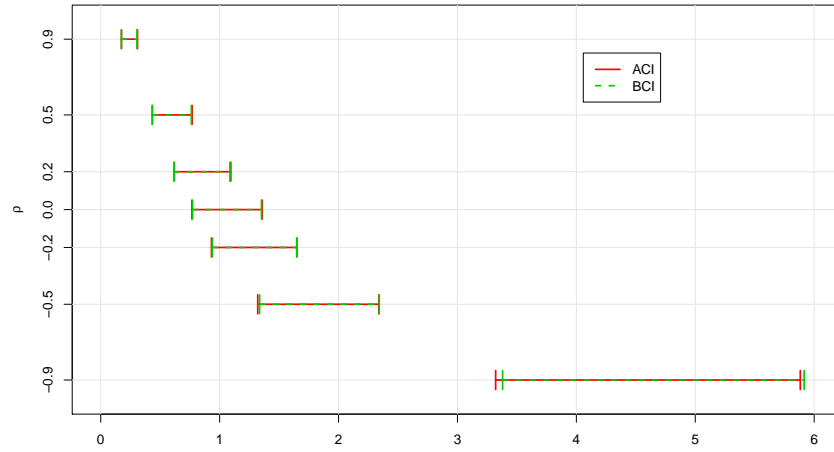
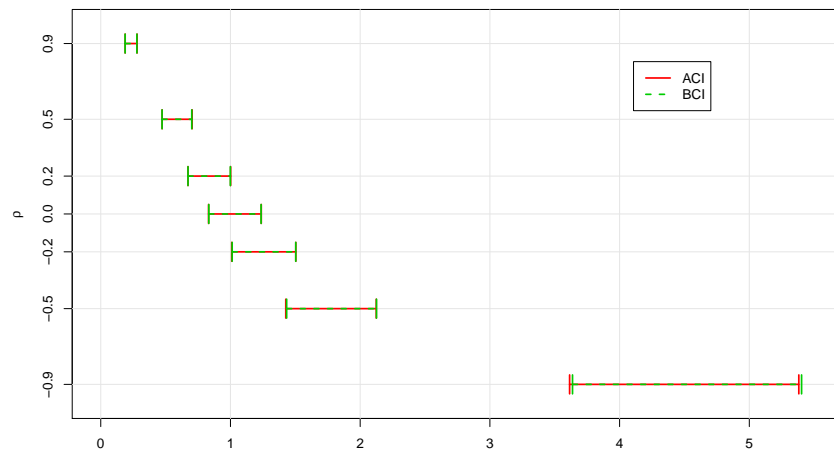
In this section we have performed a simulation study to compare coverage probability and expected length of the ACI and BCI methods, for constructing confidence intervals for $\lambda = \sqrt{\frac{1-\rho}{1+\rho}}$ (of course a similar study can be provided for $\lambda = -\sqrt{\frac{1-\rho}{1+\rho}}$).

Samples of size $n = 15, 30, 40, 50, 80, 100, 500, 1000$ were simulated from the bivariate normal distribution $N_2(0, 0, 1, 1, \rho)$ for the values $\rho = -0.9, -0.8, -0.5, -0.2, 0, 0.2, 0.5, 0.8, 0.9$ of the correlation coefficient.

For each sample size n and each value of ρ , we generate 1000 ACI and BCI intervals for the parameter λ and then we compute AVL and AVU, the average lower and upper confidence bounds, CP, the actual coverage probability of the two-sided confidence intervals (obtained as the ratio of the number of intervals containing the true values over the total number of simulations) and EL, the estimate of the expected length. Then, for these values of n and ρ , the bootstrap distribution of $\hat{\Theta}^* = \sqrt{\frac{1-R}{1+R}}$ was calculated, based on $B = 1000$ bootstrap replications.

Partial results of the simulation study are summarized in figure 2.1 and are reported in tables 2.1 and 2.2.

(a) $n=15$ (b) $n=30$

(c) $n=50$ (d) $n=100$

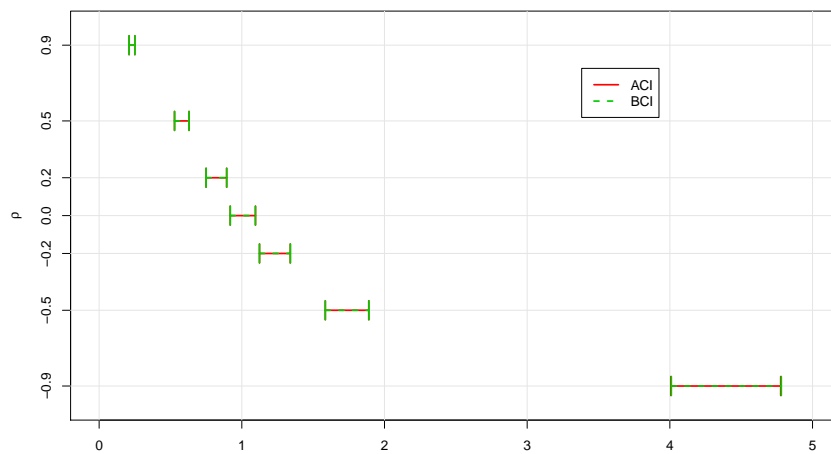
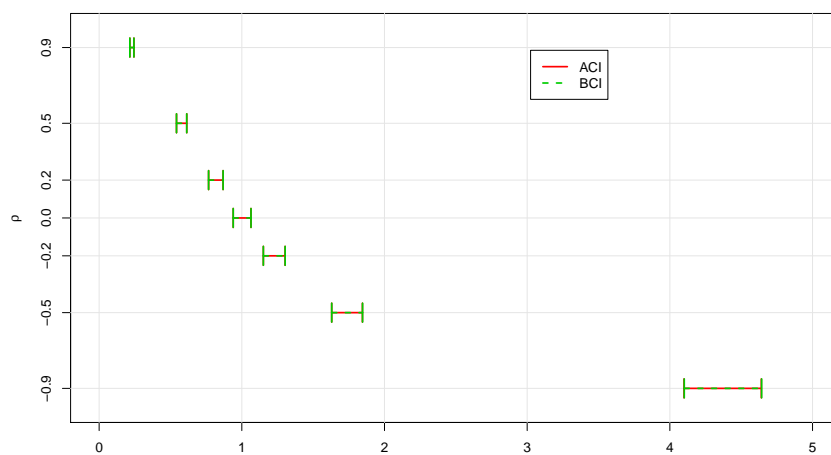
(e) $n=500$ (f) $n=1000$

Figure 2.1: Results of the simulation study

n	Method	$\rho = 0.5$		$\lambda = 0.5774$	
		AVL	AVU	EL	CP
15	ACI	0.3337	1.0347	0.7010	0.9390
	BCI	0.3267	1.0322	0.7055	0.9180
30	ACI	0.3963	0.8427	0.4464	0.944
	BCI	0.3965	0.8293	0.4328	0.923
40	ACI	0.4196	0.7992	0.3797	0.952
	BCI	0.4176	0.7934	0.3758	0.923
50	ACI	0.4339	0.7687	0.3347	0.9470
	BCI	0.4335	0.7619	0.3283	0.9380
80	ACI	0.4646	0.7263	0.2617	0.95
	BCI	0.4639	0.7234	0.2595	0.948
100	ACI	0.4732	0.7045	0.2313	0.9540
	BCI	0.4725	0.7017	0.2292	0.9410
500	ACI	0.5286	0.6302	0.1016	0.9470
	BCI	0.5282	0.6298	0.1017	0.9500
1000	ACI	0.5424	0.6141	0.0717	0.9480
	BCI	0.5422	0.6140	0.0718	0.9470

Table 2.1: Results of the simulations with $\rho = 0.5$

n	Method	$\rho = -0.5$		$\lambda = 1.7321$	
		AVL	AVU	EL	CP
15	ACI	1.0574	3.2786	2.2212	0.9490
	BCI	1.1078	3.4013	2.2935	0.9150
30	ACI	1.2247	2.6041	1.3794	0.947
	BCI	1.2495	2.6192	1.3697	0.922
40	ACI	1.2877	2.4530	1.1653	0.947
	BCI	1.3036	2.4604	1.1567	0.929
50	ACI	1.3202	2.3386	1.0184	0.9570
	BCI	1.3360	2.3421	1.0061	0.9370
80	ACI	1.4027	2.1927	0.7899	0.955
	BCI	1.4115	2.1967	0.7852	0.941
100	ACI	1.4265	2.1238	0.6973	0.9480
	BCI	1.4332	2.1260	0.6928	0.9440
500	ACI	1.5850	1.8897	0.3047	0.9560
	BCI	1.5855	1.8912	0.3058	0.9560
1000	ACI	1.6299	1.8453	0.2154	0.9530
	BCI	1.6301	1.8463	0.2161	0.9540

Table 2.2: Results of the simulations with $\rho = -0.5$

A confidence interval with a narrower expected length implies a more accurate estimate of the parameter and thus is always preferred to a longer one. The actual coverage probability should be near to the nominal coverage 0.95. Table 2.1, table 2.2 and other data not presented indicate that, for $n \geq 50$, the simulation study gives similar results for the two methods and for different values of ρ , both in terms of coverage probability and expected length. For small and moderate sample sizes the ACI has actual coverage probabilities reasonably close to the nominal value of 0.95. In contrast, the intervals based on the BCI method have poor coverage when n is small or moderate. We notice that, in general, the bootstrap method has a coverage rate slightly less than 95%. As expected, with larger sample sizes the confidence intervals become narrower. For both methods, the expected length becomes larger for negative values of ρ . This behaviour is particularly evident when ρ is close to -1 and n is small or moderate. This is not surprising and it is in agreement with other results in literature. In fact, as $\rho \rightarrow -1$, $\lambda \rightarrow \infty$ and estimation problems can arise. As expected, the simulation study confirms that, for all sample sizes, the length of the interval decreases as ρ increases.

2.4 Examples

To highlight the applicability of the method presented in section 1 of this chapter we consider two situations of different nature.

2.4.1 PM_{10} concentrations

In environmental or epidemiological studies it is relevant to estimate the distribution of extreme statistics. If you are monitoring the pollution in different areas of a region or a town it is important to model appropriately

the order statistic maximum and/or minimum or the range. In this example we analyse data from PM₁₀ concentrations recorded daily between the 1st of December 2003 and the 1st of February 2005 in two different stations in Cagliari, Italy. After removing missing values, from each station we have 111 observations. Our interest rests on the natural logarithm of the maximum value of PM₁₀ concentrations in the two stations. We assume that their joint distribution is bivariate normal. We standardized the variables using the MLEs of the unknown means and standard deviations. We are interested in the distribution of the maximum of such standardized random variables. The conditions of theorem 1 are satisfied. Then we know that our extreme statistic has a skew-normal distribution with location parameter equal to 0, scale parameter equal to 1 and skewness parameter λ equal to $-\sqrt{\frac{1-\rho}{1+\rho}}$. To evaluate a confidence interval for λ we apply the procedure described in section 2.1. In order to check the fit of the bivariate normal distribution to the data we use the Shapiro-Wilk Multivariate Normality Test (see Shapiro and Wilk (1965) [55]). This is based on the Shapiro-Wilk statistic defined as the ratio of two estimates of the variance of a normal distribution based on a random sample of ordered n observations $y_1 \leq y_2 \leq \dots \leq y_n$. Analytically, $W = \frac{(\sum a_i y_i)^2}{\sum (y_i - \bar{y})^2}$, where $a = (a_1, \dots, a_n)^T$ is such that $\sqrt{n-1} \sum a_i y_i$ is the best unbiased estimate of the standard deviation of the y_i assuming normality. The observed value is $W = 0.9823$ and the corresponding p-value is 0.1474. The estimated value for R is 0.5145. In table 2.3 (left size) are reported the estimated value of λ together with an approximate 95% confidence interval. This confidence interval is then compared with the bootstrap interval constructed as described in section 2. ACI provides slightly better results than BCI.

2.4.2 Creatinine clearance

In our second example we consider a dataset concerning the follow-up of 145 patients who had an operation for a renal cancer in the University hospital of Strasbourg. The follow-up consists of several medical examinations (1, 3, 6, 12 and 24 months after the operation) with blood tests and further investigations. Glomerular filtration rate is a measure of renal function using the flow rate of filtered fluid through the kidney. Creatinine clearance rate is the volume of blood plasma that is cleared of creatinine per unit time and is a common measure for approximating the glomerular filtration rate. When the patient has his medical consultation six months after operation, the maximum of the creatinine clearance rate between the value at 1 month and the value at 3 months can be considered as the value of his new renal function after recovery. Statistical tests confirm that the two measures (creatinine clearance at 1 month and at 3 months) have the same mean and the same variance. The Shapiro-Wilk test can not reject the hypothesis that the distribution is joint normal ($W = 0.9893$ and p-value is 0.3302). The estimated correlation coefficient is $R = 0.9266$. Table 2.3 (right side) summarizes the point estimate of λ and its confidence interval at the 95% level using Fisher's transformation and the bootstrap technique. As previously, the length of this interval is narrower with the theoretical approximate method than with the bootstrap.

Method	Example 1: $\hat{\lambda} = 0.5661$		Example 2: $\hat{\lambda} = 0.1951$	
	CI	Length	CI	Length
ACI	(0.4688, 0.6836)	0.2148	(0.1655, 0.2300)	0.0645
BCI	(0.4401, 0.6701)	0.23	(0.1607, 0.2352)	0.07454

Table 2.3: ACI and BCI of level 0.95 for λ using data from example 1 and 2

Chapter 3

The Beta skew-normal distribution

The main task of this chapter is to introduce a new class of distributions, which we call Beta skew-normal since it is a special case of the Beta-generated distribution. The moment generating function and some important theorems about the moments of this distribution are derived in section 1. Furthermore, we give some bimodality properties. In section 2 we link the distributions introduced in section 1.2 with the Beta skew-normal. In section 3 we provide bounds for the moments and the variance of the Beta-generated distribution. In the last section the estimation of the parameters is investigated by maximum likelihood method. The results presented in this chapter are new and some of these can be found in [\[45\]](#).

3.1 The Beta skew-normal

3.1.1 Definition and simple properties

Replacing in (1.18) $F(x)$ by $\Phi(x; \lambda)$, we obtain the Beta skew-normal distribution, with distribution function given by

$$G_{\Phi(x; \lambda)}^B(x; \lambda, a, b) = \frac{1}{B(a, b)} \int_0^{\Phi(x; \lambda)} z^{a-1} (1-z)^{b-1} dz, \quad x \in \mathbb{R}, \quad (3.1)$$

and probability density function

$$g_{\Phi(x; \lambda)}^B(x; \lambda, a, b) = \frac{2}{B(a, b)} (\Phi(x; \lambda))^{a-1} (1 - \Phi(x; \lambda))^{b-1} \phi(x) \Phi(\lambda x). \quad (3.2)$$

Throughout this thesis, we denote the Beta skew-normal distribution with vector of parameters $\boldsymbol{\xi} = (\lambda, a, b)$ by $BSN(\lambda, a, b)$.

The class of the Beta skew-normal can be generalized by the inclusion of the location and scale parameters which we identify as μ and $\sigma > 0$. Thus if $X \sim BSN(\lambda, a, b)$ then $Y = \mu + \sigma X$ is a Beta skew-normal with vector of parameters $\boldsymbol{\xi} = (\mu, \sigma, \lambda, a, b)$. We denote Y by $Y \sim BSN(\mu, \sigma, \lambda, a, b)$.

We now present some properties concerning the $BSN(\lambda, a, b)$.

Properties of $BSN(\lambda, a, b)$:

- a. $g_{\Phi(x; \lambda)}^B(x; \lambda, 1, 1) = \phi(x; \lambda)$, for all $x \in \mathbb{R}$, i.e. $BSN(\lambda, 1, 1) = SN(\lambda)$.
- b. $g_{\Phi(x; 0)}^B(x; 0, a, b) = g_{\Phi(x)}^B(x; a, b)$, for all $x \in \mathbb{R}$, i.e. $BSN(0, a, b) = BN(a, b)$.
- c. $g_{\Phi(x; 0)}^B(x; 0, 1, 1) = \phi(x)$, for all $x \in \mathbb{R}$, i.e. $BSN(0, 1, 1) = N(0, 1)$.
- d. $g_{\Phi(x; 1)}^B(x; 1, \frac{1}{2}, 1) = \phi(x)$, for all $x \in \mathbb{R}$, i.e. $BSN(1, \frac{1}{2}, 1) = N(0, 1)$.
- e. $g_{\Phi(x; -1)}^B(x; -1, 1, \frac{1}{2}) = \phi(x)$, for all $x \in \mathbb{R}$, i.e. $BSN(-1, 1, \frac{1}{2}) = N(0, 1)$.
- f. If $X \sim BSN(\lambda, a, b)$, then $-X \sim BSN(-\lambda, b, a)$.

- g. If $X \sim BSN(\lambda, a, b)$, then $Y = \Phi(X; \lambda)$ is a $Beta(a, b)$.
- h. If $X \sim BSN(\lambda, a, b)$, then $Y = 1 - \Phi(X; \lambda)$ is a $Beta(b, a)$.
- i. As $\lambda \rightarrow +\infty$, $g_{\Phi(x; \lambda)}^B(x; \lambda, a, b)$ tends to the Beta half-normal density.

Remark 10. *Properties from a to e establish that the family of $BSN(\lambda, a, b)$ contains the standard normal distribution, the skew-normal distribution and the Beta-normal distribution as special cases.*

Proof. Points from a to h follow directly from (3.2) and from elementary properties of the skew-normal distribution.

We now show point i. From property a of section 1.1 we have $\phi(x; \lambda) \xrightarrow{\lambda \rightarrow \infty} 2\phi(x)$, with $x > 0$. Then

$$\Phi(x; \lambda) = \int_{-\infty}^x \phi(t; \lambda) dt = \int_{-\infty}^{\infty} H(x-t) \phi(t; \lambda) dt \xrightarrow{\lambda \rightarrow \infty} 2\Phi(x) - 1, \quad \text{with } x > 0. \quad (3.3)$$

This is the distribution function of a variable with half-normal distribution.

Here H is Heaviside's function defined as

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0. \end{cases}$$

Then we have

$$\lim_{\lambda \rightarrow \infty} g_{\Phi(x; \lambda)}^B(x; \lambda, a, b) = \frac{2}{B(a, b)} (2\Phi(x) - 1)^{a-1} (2(1 - \Phi(x)))^{b-1} \phi(x). \quad (3.4)$$

The right side of (3.4) is the density function of a variable with Beta half-normal distribution. \square

The BSN distribution is easily simulated using property g as follows: if Y has a Beta distribution with parameters a and b , then the variable $X = \Phi^{-1}(Y; \lambda)$

has $BSN(\lambda, a, b)$ distribution, where $\Phi^{-1}(\cdot; \lambda)$ is the quantile function of the skew-normal distribution.

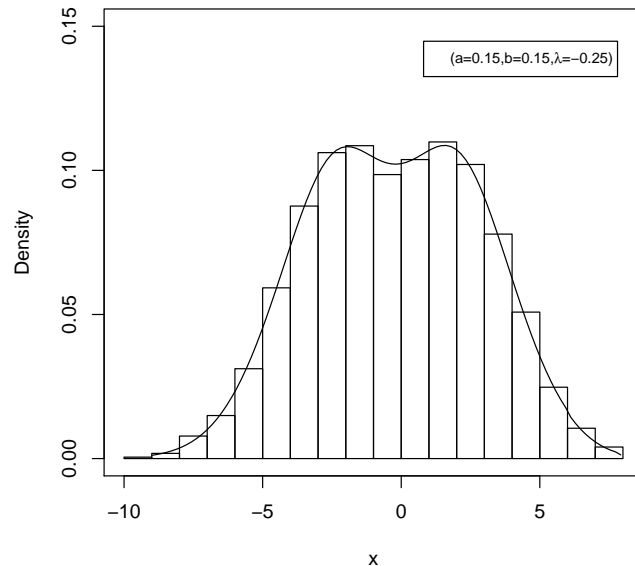
In figure 3.1 are plotted random samples generated by the BSN distribution for some a , b and λ with the respective curve of the density function obtained using the R-package “sn” (see Azzalini (2010) [8]).

From this plot we can observe that, for values of a and b close to zero, the distribution can be bimodal.

From remark 5, we know that, if $a \geq 1$ and $b \geq 1$, the density (3.2) is strongly unimodal, i.e. $\log g_{\Phi(x; \lambda)}^B(x; \lambda, a, b)$ is a concave function of x (see figure 3.2).

We do not have general results for $a < 1$ and/or $b < 1$.

A numerical study has shown that, when at least one of the two parameters a and b is close to zero (0.10, 0.20), the density can be bimodal (see figure 3.3).



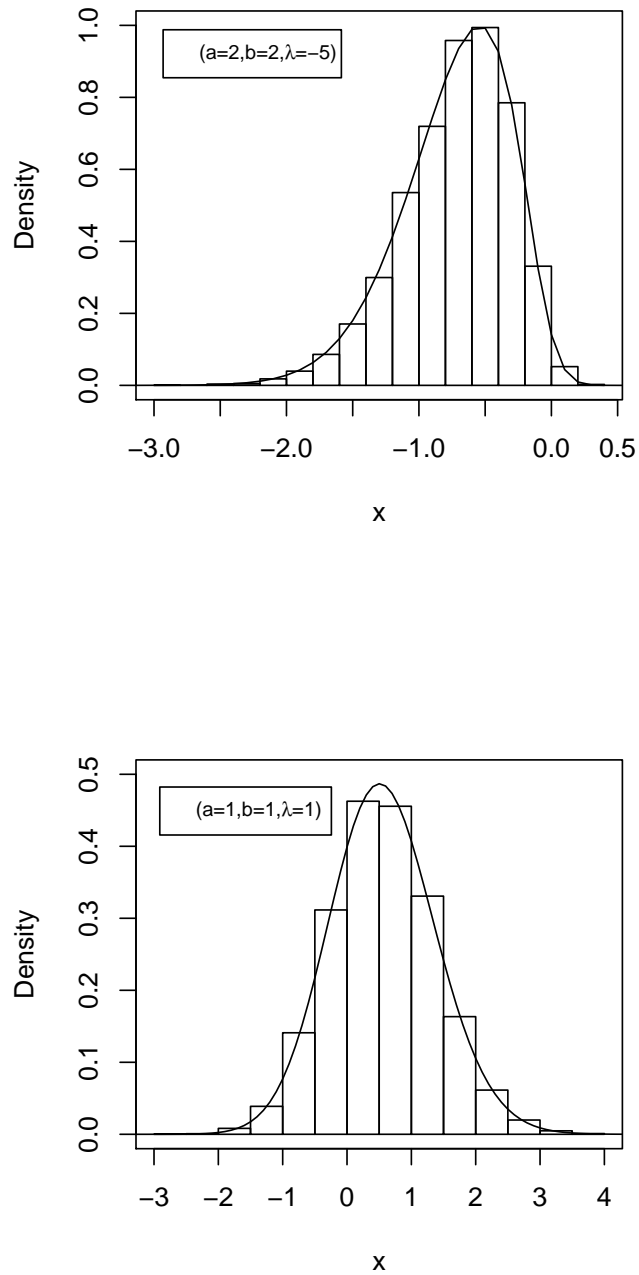


Figure 3.1: Generated samples of the *BSN* distribution for some values of λ , a and b

3.1.2 Moment generating function and moments

Now we find the moment generating function of X which has density (3.2).

Property 8. *The moment generating function of $X \sim \text{BSN}(\lambda, a, b)$ is given by*

$$M_X(t) = \frac{2}{B(a, b)} e^{\frac{t^2}{2}} E_Z \left((\Phi(Z; \lambda))^{a-1} (1 - \Phi(Z; \lambda))^{b-1} \Phi(\lambda Z) \right), \quad (3.5)$$

where $Z \sim N(t, 1)$.

Proof. Using integration by parts, it follows that

$$\begin{aligned} M_X(t) &= \frac{1}{B(a, b)} \int_{-\infty}^{\infty} e^{tx} \phi(x; \lambda) (\Phi(x; \lambda))^{a-1} (1 - \Phi(x; \lambda))^{b-1} dx = \\ &= \frac{2}{B(a, b)} \int_{-\infty}^{\infty} e^{tx} \phi(x) \Phi(\lambda x) (\Phi(x; \lambda))^{a-1} (1 - \Phi(x; \lambda))^{b-1} dx = \\ &= \frac{2e^{\frac{t^2}{2}}}{B(a, b)} \int_{-\infty}^{\infty} \phi(x-t) \Phi(\lambda x) (\Phi(x; \lambda))^{a-1} (1 - \Phi(x; \lambda))^{b-1} dx, \end{aligned}$$

and the proof is complete. \square

We have the following recursion formula:

Property 9. *Let $k \in \mathbb{N}$ and $k \geq 1$. If $X \sim \text{BSN}(\lambda, a, b)$, with $a > 1$ and $b > 1$, then*

$$\begin{aligned} E_X(X^k) &= (k-1)E_X(X^{k-2}) + \lambda E_X \left(X^{k-1} \frac{\phi(\lambda X)}{\Phi(\lambda X)} \right) + \\ &\quad + (a+b-1)E_U \left(U^{k-1} \phi(U; \lambda) \right) - (a+b-1)E_V \left(V^{k-1} \phi(V; \lambda) \right), \end{aligned}$$

where $U \sim \text{BSN}(\lambda, a-1, b)$ and $V \sim \text{BSN}(\lambda, a, b-1)$ are independent random variables.

Proof. The proof follows easily from application of the formula for integration by parts and by using the well note relation $\frac{\partial \phi(x)}{\partial x} = -x\phi(x)$ (see Arnold et al. (1992) [4]). \square

By property f of *BSN* we can deduce the following proposition.

Proposition 3. *Let $X \sim \text{BSN}(\lambda, a, b)$ and $Y \sim \text{BSN}(-\lambda, b, a)$. We have the following statements:*

- $E_X(X) = -E_Y(Y)$;
- $\text{var}_X(X) = \text{var}_Y(Y)$;
- $\gamma_1(X) = -\gamma_1(Y)$;
- $\gamma_2(X) = \gamma_2(Y)$;

with γ_1 and γ_2 we indicate the skewness and the kurtosis, respectively.

The following lemma is an application to the *BSN* distribution of lemma 4 of Zografos and Balakrishnan (2009) [60].

Lemma 2. *Let $X \sim \text{BSN}(\lambda, a, b)$. Then the following sentences hold.*

1. $E_X(\Phi(X; \lambda)) = \frac{a}{a+b}$;
2. $E_X(\ln \Phi(X; \lambda)) = \psi(a) - \psi(a+b)$;
3. $E_X(1 - \Phi(X; \lambda)) = \frac{b}{a+b}$;
4. $E_X(\ln(1 - \Phi(X; \lambda))) = \psi(b) - \psi(b+a)$;

where $\psi(t) = \frac{d \ln(\Gamma(t))}{dt}$ is the di-gamma function (called also the logarithmic derivative of the gamma function).

We refer to [1] for details concerning on the di-gamma function.

The Beta skew-normal density is in general asymmetric (see figures 3.2 and 3.3). We have a partial result concerning symmetry:

Proposition 4. *If $a = b$ and $\text{BSN}(\lambda, a, b)$ is symmetric about 0 then $\lambda = 0$.*

Proof. We consider the density of a random variable $X \sim \text{BSN}(\lambda, a, a)$:

$$g_{\Phi(x;\lambda)}^B(-x; \lambda, a, a) = \frac{2}{B(a, a)} \phi(x) \Phi(-\lambda x) (1 - \Phi(x; -\lambda))^{a-1} (\Phi(x; -\lambda))^{a-1},$$

this is equal to $g_{\Phi(x;\lambda)}^B(x; \lambda, a, a)$ if $\Phi(\lambda x) = \Phi(-\lambda x)$ and $\Phi(x; \lambda) = \Phi(x; -\lambda)$.

However for property 3 we find that $\Phi(x; \lambda) = \Phi(x)$ which implies that $\lambda = 0$. \square

Remark 11. Eugene et al. (2002) [23] have shown that the $\text{BN}(a, b) = \text{BSN}(0, a, b)$ is symmetric about 0 when $a = b$.

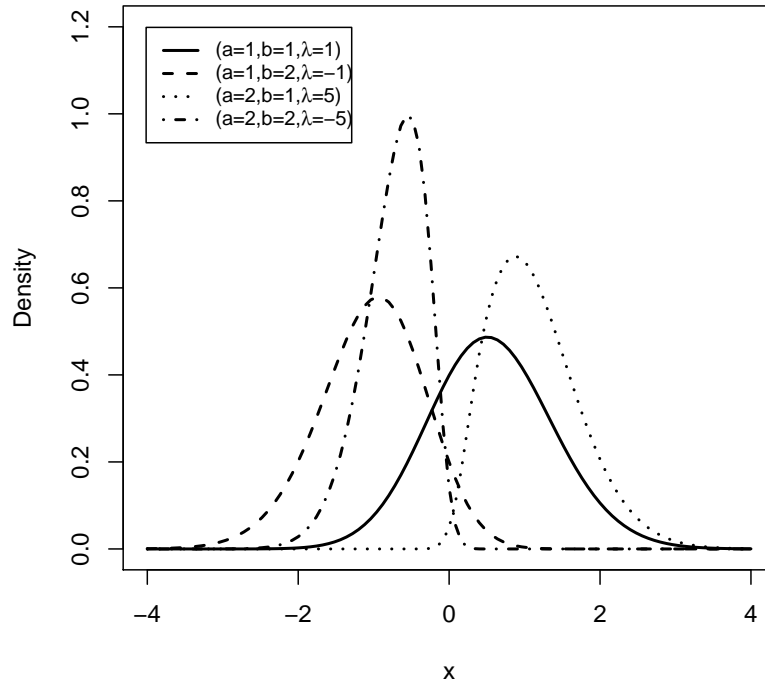


Figure 3.2: The $\text{BSN}(\lambda, a, b)$ for values of $a \geq 1$ and $b \geq 1$

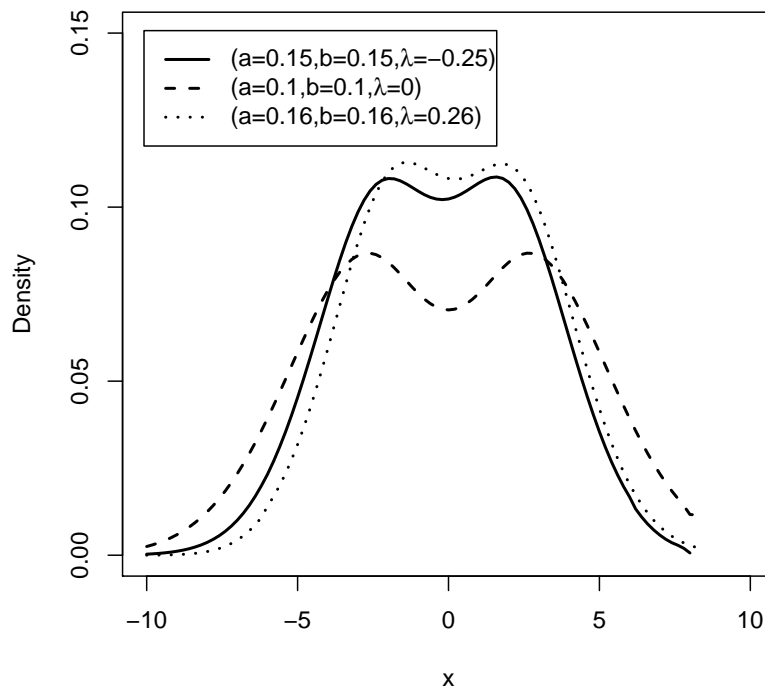


Figure 3.3: The $BSN(\lambda, a, b)$ for values of $a < 1$ and $b < 1$

Moments of the BSN can not be evaluated exactly in closed form. We have computed them numerically using the software R.

In table 3.1 we have reported the values of the mean μ_{BSN} , standard deviation σ_{BSN} , skewness γ_1 and kurtosis γ_2 for different values of the parameters a , b and λ . From this numerical study we have noted that:

- for fixed values of a and b the mean μ_{BSN} and skewness γ_1 are both increasing function of λ ;

- for fixed values of b and λ the mean μ_{BSN} and skewness γ_1 are both increasing function of a ;
- for fixed values of a and λ the mean μ_{BSN} is a decreasing function of b .

(a)							(b)						
a	b	λ	μ_{BSN}	σ_{BSN}	γ_1	γ_2	a	b	λ	μ_{BSN}	σ_{BSN}	γ_1	γ_2
0.25	0.25	-10	-1.1579	1.4029	-1.1329	3.7648	0.5	10	-10	-2.3678	0.7314	-0.7505	3.7967
		-1	-0.6501	1.9679	-0.2378	2.7777			-1	-2.3617	0.7389	-0.7188	3.7849
		0	0	2.3382	-0.0004	2.6217			0	-2.0809	0.8033	-0.6173	3.5736
		1	0.6484	1.9649	0.2306	2.1362			1	-1.0893	0.6117	-0.5642	3.4799
		10	1.1580	1.4027	1.1329	3.7632			10	-0.0182	0.1429	0.3706	3.8635
0.25	0.5	-10	-1.5906	1.3469	-0.7185	2.7580	1	0.5	-10	-0.4873	0.5777	-1.3200	4.8570
		-1	-1.4424	1.6716	-0.3284	3.0202			-1	0	1	0	3
		0	-0.9631	1.9061	-0.0849	2.8029			0	0.7043	1.2479	0.1372	2.9831
		1	-0.1772	1.5265	0.0938	2.8543			1	1.1664	1.0704	0.3086	3.1161
		10	0.5446	0.8728	1.5054	5.1988			10	1.3018	0.9148	0.8262	3.4814
0.5	0.25	-10	-0.5447	0.8727	-1.5061	5.2003	1	1	-10	-0.7939	0.6080	-0.9556	3.8232
		-1	0.1773	1.5265	-0.0938	2.8541			-1	-0.5642	0.8256	-0.1369	3.0617
		0	0.9625	1.9051	0.0819	2.7927			0	0	1	0	3
		1	1.4411	1.6694	0.3203	2.9849			1	0.5642	0.8256	0.1369	3.0617
		10	1.6339	1.3974	0.8434	3.2655			10	0.7939	0.6080	0.9556	3.8232
0.5	0.5	-10	-0.8979	0.8874	-0.9703	3.3176	10	1	-10	-0.0839	0.1364	-0.7082	4.2018
		-1	-0.5882	1.2659	-0.1811	2.9514			-1	0.6744	0.4536	0.3597	3.2722
		0	0	1.5253	0	2.8615			0	1.5388	0.5868	0.4099	3.3314
		1	0.5882	1.2659	0.1811	2.9514			1	1.8675	0.5251	0.5005	3.4685
		10	0.9179	0.9153	1.0703	3.7747			10	1.8807	0.5124	0.5744	3.5243
0.5	1	-10	-1.3018	0.9148	-0.8262	3.4815	1	10	-10	-1.8807	0.5124	-0.5744	3.5243
		-1	-1.1664	1.0704	-0.3085	3.1159			-1	-1.8675	0.5251	-0.5005	3.4685
		0	-0.7043	1.2479	-0.1372	2.9831			0	-1.5388	0.5868	-0.4099	3.3314
		1	0	0.9999	0	2.9999			1	-0.6744	0.4536	-0.3597	3.2722
		10	0.4873	0.5778	1.3199	4.8561			10	0.0839	0.1364	0.7082	4.2018

Table 3.1: The first moment, the standard deviation, the skewness and the kurtosis of $BSN(\lambda, a, b)$ for different values of a , b and λ

3.1.3 Order statistics from the skew-normal distribution

We now give some results concerning the distribution of order statistics from a skew-normal distribution:

Proposition 5. *Let X_1, \dots, X_n be a random sample from a $SN(\lambda)$. Then the j -th order statistic is a $BSN(\lambda, j, n - j + 1)$, where $j = 1, \dots, n$.*

Proof. The proof follows easily using the standard formula of the density of $X_{(i)}$, the i -th order statistic of a random sample of size n from the distribution $SN(\lambda)$. \square

From proposition 5 follows immediately that the family of BSN contains the distributions of the order statistics of the skew-normal distribution.

In particular, we have the following corollaries:

Corollary 5. *Let X_1, \dots, X_n be a random sample from a $SN(1)$. Then*

$$X_{(n)} = \max \{X_1, \dots, X_n\}$$

is a $BSN(1, n, 1)$.

Corollary 6. *Let X_1, \dots, X_n be a random sample from a $SN(-1)$. Then*

$$X_{(1)} = \min \{X_1, \dots, X_n\}$$

is a $BSN(-1, 1, n)$.

Corollary 7. *Let $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ be the order statistics from a sample of size n from a $SN(\lambda)$ distribution. Then $X_{(i)}$, $i = 1, \dots, n$, has log-concave density.*

Proof. From property d of section 1.1 we know that X_i has a log-concave density. We conclude the proof using the following result due to Gupta (2004) [31]: *Suppose $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ be the order statistics from a sample of size n from a distribution having a log-concave density function. Then $X_{(i)}$, $i = 1, \dots, n$, has log-concave density.* \square

3.1.4 Some interesting properties

Here, we present some properties of the *BSN* distribution of general interest.

Theorem 7. *Let $X \sim \text{BSN}(\lambda, a, b)$ be independent of a random sample (Y_1, \dots, Y_n) from $\text{SN}(\lambda)$, then*

- i. $X | (Y_{(n)} \leq X) \sim \text{BSN}(\lambda, a + n, b)$,
- ii. $X | (Y_{(1)} \geq X) \sim \text{BSN}(\lambda, a, b + n)$,

where $Y_{(n)}$ and $Y_{(1)}$ are the largest and the smallest order statistics, respectively.

Proof. We shall prove point i. If $W = X | (Y_{(n)} \leq X)$, then we have

$$P(W \leq w) = \frac{\int_{-\infty}^w (\Phi(x; \lambda))^n \frac{2}{B(a, b)} \phi(x) \Phi(\lambda x) (\Phi(x; \lambda))^{(a-1)} (1 - \Phi(x; \lambda))^{(b-1)} dx}{P(Y_{(n)} \leq X)}. \quad (3.6)$$

Also

$$\begin{aligned} P(Y_{(n)} \leq X) &= P(Y_1 \leq X, \dots, Y_n \leq X) = \\ &= \int_{-\infty}^{\infty} (\Phi(x; \lambda))^n \frac{2}{B(a, b)} \phi(x) \Phi(\lambda x) (\Phi(x; \lambda))^{(a-1)} (1 - \Phi(x; \lambda))^{(b-1)} dx = \\ &= \frac{B(a + n, b)}{B(a, b)}. \end{aligned}$$

Taking derivative from (3.6) with respect to w , we obtain the $BSN(\lambda, a+n, b)$ density function.

The proof of point *ii* is similar. \square

The following theorem is a generalization of the above one.

Theorem 8. *Let $X \sim BSN(\lambda, a, b)$ be independent of $Y \sim BSN(\lambda, c, 1)$ and of $Z \sim BSN(\lambda, 1, d)$. Then*

i. $X|(Y \leq X) \sim BSN(\lambda, a+c, b),$

ii. $X|(Z \geq X) \sim BSN(\lambda, a, b+d),$

where c and d are positive real numbers.

Theorem 9. *If $X \sim BSN(\lambda, a, b)$ is independent of $U_1, \dots, U_n, V_1, \dots, V_m$ i.i.d. random variables having $SN(\lambda)$ distribution, then*

$$X|(U_{(n)} \leq X, V_{(1)} \geq X) \sim BSN(\lambda, a+n, b+m), \quad (3.7)$$

where $U_{(n)} = \max\{U_1, \dots, U_n\}$ and $V_{(1)} = \min\{V_1, \dots, V_m\}$.

Proof. The proof is quite similar to the one of theorem 7. \square

We can generalize the above theorems for the family of the Beta-generated distributions in the following way.

Theorem 10. *Let $X \sim \text{Beta} - F(a, b)$ be independent of a random sample (Y_1, \dots, Y_n) from $F(\cdot)$ with density function $f(\cdot) = F'(\cdot)$, then*

$X|(Y_{(n)} \leq X) \sim \text{Beta} - F(a+n, b)$ and $X|(Y_{(1)} \geq X) \sim \text{Beta} - F(a, b+n)$, where $Y_{(n)}$ and $Y_{(1)}$ are the largest and the smallest order statistics, respectively.

Theorem 11. *Let $X \sim \text{Beta} - F(a, b)$ be independent of $Y \sim \text{Beta} - F(c, 1)$ and of $Z \sim \text{Beta} - F(1, d)$. Then $X|(Y \leq X) \sim \text{Beta} - F(a+c, b)$ and $X|(Z \geq X) \sim \text{Beta} - F(a, b+d)$, where c and d are positive real numbers.*

Theorem 12. *If $X \sim \text{Beta} - F(a, b)$ is independent of $U_1, \dots, U_n, V_1, \dots, V_m$ i.i.d. random variables with pdf $f(\cdot) = F'(\cdot)$, then*

$$X | (U_{(n)} \leq X, V_{(1)} \geq X) \sim \text{Beta} - F(a + n, b + m), \quad (3.8)$$

where $U_{(n)} = \max\{U_1, \dots, U_n\}$ and $V_{(1)} = \min\{V_1, \dots, V_m\}$.

Theorem 7 can be used to generate $X \sim \text{BSN}(\lambda, n, 1)$ by extending the acceptance-rejection technique, due to Azzalini, as follows (see Azzalini (1985) [6] and Sharafi and Behboodian (2008) [57]): first we generate a random sample $T, U_1, U_2, \dots, U_{n-1}$ from $SN(\lambda)$, if $\max(U_1, U_2, \dots, U_{n-1}) \leq T$ we put $X = T$. Otherwise, we generate a new random sample, until the above inequality is satisfied.

The same procedure can be used to generate $X \sim \text{Beta} - F(n, 1)$.

3.1.5 Bimodal properties

Motivated by the work of Famoye et al. (2004) [24], we prove, in this section, bimodality properties of the Beta skew-normal.

Theorem 13. *The mode(s) of $\text{BSN}(\mu, \sigma, \lambda, a, b)$ is any point $x_0 = x_0(\lambda, a, b)$ that satisfies (satisfy)*

$$x_0 = \sigma \left\{ \lambda \frac{\phi\left(\frac{\lambda(x_0 - \mu)}{\sigma}\right)}{\Phi\left(\frac{\lambda(x_0 - \mu)}{\sigma}\right)} + (a - 1) \frac{\phi\left(\frac{x_0 - \mu}{\sigma}; \lambda\right)}{\Phi\left(\frac{x_0 - \mu}{\sigma}; \lambda\right)} - (b - 1) \frac{\phi\left(\frac{x_0 - \mu}{\sigma}; \lambda\right)}{1 - \Phi\left(\frac{x_0 - \mu}{\sigma}; \lambda\right)} \right\} + \mu. \quad (3.9)$$

Proof. Differentiating the density of a random variable with $\text{BSN}(\mu, \sigma, \lambda, a, b)$ distribution with respect to x and setting this derivative equal to zero, and solving it for x , we obtain the stated result. \square

Corollary 8. *If $\text{BSN}(\mu, \sigma, \lambda, a, b)$ has a mode at x_0 , then $\text{BSN}(\mu, \sigma, -\lambda, b, a)$ has a mode at the point $2\mu - x_0$.*

Proof. It is sufficient to show that equation (3.9) remains the same if x_0 is replaced with $2\mu - x_0$, a with b and λ with $-\lambda$. By making these substitutions, it follows that

$$\mu - x_0 = \sigma \left\{ -\lambda \frac{\phi\left(\frac{-\lambda(\mu-x_0)}{\sigma}\right)}{\Phi\left(\frac{-\lambda(\mu-x_0)}{\sigma}\right)} + (b-1) \frac{\phi\left(\frac{(\mu-x_0)}{\sigma}; -\lambda\right)}{\Phi\left(\frac{(\mu-x_0)}{\sigma}; -\lambda\right)} - (a-1) \frac{\phi\left(\frac{(\mu-x_0)}{\sigma}; -\lambda\right)}{1 - \Phi\left(\frac{(\mu-x_0)}{\sigma}; -\lambda\right)} \right\},$$

and, using $\phi(-x; \lambda) = \phi(x; -\lambda)$ and $1 - \Phi(x; \lambda) = \Phi(-x; \lambda)$, we get the result in (3.9). \square

Corollary 9. *The modal point x_0 is an increasing function of a and a decreasing function of b .*

Proof. Differentiating the result in (3.9) with respect to a and b , we get respectively:

$$\frac{\partial x_0}{\partial a} = \sigma \left(\frac{\phi\left(\frac{x_0-\mu}{\sigma}; \lambda\right)}{\Phi\left(\frac{x_0-\mu}{\sigma}; \lambda\right)} \right) > 0; \quad (3.10)$$

$$\frac{\partial x_0}{\partial b} = -\sigma \left(\frac{\phi\left(\frac{x_0-\mu}{\sigma}; \lambda\right)}{1 - \Phi\left(\frac{x_0-\mu}{\sigma}; \lambda\right)} \right) < 0. \quad (3.11)$$

Hence x_0 is an increasing function of a and a decreasing function of b . \square

Corollary 10. *The bimodal property of $BSN(\mu, \sigma, \lambda, a, b)$ is independent of the parameters μ and σ .*

Proof. The mode(s) of $BSN(\mu, \sigma, \lambda, a, b)$ is at the point $x_0 = x_0(\lambda, a, b)$ that satisfies equation (3.9) and can be rewritten in the following way:

$$\frac{x_0 - \mu}{\sigma} = \left\{ \lambda \frac{\phi\left(\frac{\lambda(x_0-\mu)}{\sigma}\right)}{\Phi\left(\frac{\lambda(x_0-\mu)}{\sigma}\right)} + (a-1) \frac{\phi\left(\frac{x_0-\mu}{\sigma}; \lambda\right)}{\Phi\left(\frac{x_0-\mu}{\sigma}; \lambda\right)} - (b-1) \frac{\phi\left(\frac{x_0-\mu}{\sigma}; \lambda\right)}{1 - \Phi\left(\frac{x_0-\mu}{\sigma}; \lambda\right)} \right\}, \quad (3.12)$$

so we replace $\frac{x_0-\mu}{\sigma}$ by z_0 and we obtain

$$z_0 = \left\{ \lambda \frac{\phi(\lambda z_0)}{\Phi(\lambda z_0)} + (a-1) \frac{\phi(z_0; \lambda)}{\Phi(z_0; \lambda)} - (b-1) \frac{\phi(z_0; \lambda)}{1 - \Phi(z_0; \lambda)} \right\}, \quad (3.13)$$

which is independent of parameters μ and σ . \square

3.1.6 Expansion for the density function

Here, we give a simple expansion for the *BSN* density function.

Application of (1.23) to equation (3.1), if b is real non-integer, gives

$$G_{\Phi(x;\lambda)}^B(x; a, b) = \sum_{i=0}^{\infty} w_i(a, b) \Phi(x; \lambda)^{a+i}, \quad (3.14)$$

where $w_i(a, b) = \frac{1}{B(a, b)} (-1)^i \binom{b-1}{i} \frac{1}{a+i}$. Correspondingly, the density function (3.2) can be written as

$$g_{\Phi(x;\lambda)}^B(x; \lambda, a, b) = \sum_{i=0}^{\infty} w_i(a, b) g_{\Phi(x;\lambda)}^B(x; \lambda, a+i, 1), \quad (3.15)$$

where the weights $w_i(a, b)$ are such that $\sum_{i=0}^{\infty} w_i(a, b) = 1$.

However, it is clear from the last equation that $g_{\Phi(x;\lambda)}^B(x; \lambda, a, b)$ can be expressed as an infinite mixture of *BSN*($\lambda, a+i, 1$) densities with constant weights $w_i(a, b)$. For b integer, the previous sums stop at $b-1$. If a is real non-integer the distribution function takes the following expression

$$\begin{aligned} G_{\Phi(x;\lambda)}^B(x; a, b) &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} s_j(a+i) \Phi(x; \lambda)^j = \\ &= \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} w_i(a, b) s_j(a+i) \Phi(x; \lambda)^j = \sum_{j=0}^{\infty} t_j(a, b) \Phi(x; \lambda)^j, \end{aligned} \quad (3.16)$$

where

$$s_j(a+i) = \sum_{k=j}^{\infty} (-1)^{k+j} \binom{a+i-1}{k} \binom{k}{j}$$

and

$$t_j(a, b) = \sum_{i=0}^{\infty} w_i(a, b) s_j(a+i).$$

The density for a real non-integer can be easily obtained from the above equation by differentiation.

Remark 12. *The density function (1.1) of model $SN(\lambda)$ can be represented in the following way:*

$$\begin{aligned}\phi(z; \lambda) &= 2\phi(z)\Phi(\lambda z) = 2\phi(z)\Phi(\lambda z)(1 - \Phi(z; \lambda) + \Phi(z; \lambda)) = \\ &= 2\phi(z)\Phi(\lambda z)(1 - \Phi(z; \lambda)) + 2\phi(z)\Phi(\lambda z)\Phi(z; \lambda) = \\ &= \frac{1}{2} \left(g_{\Phi(z; \lambda)}^B(z; \lambda, 1, 2) + g_{\Phi(z; \lambda)}^B(z; \lambda, 2, 1) \right).\end{aligned}\quad (3.17)$$

In other words the density function of the skew-normal with parameter λ is a mixture between a Beta skew-normal density with parameters λ , $a = 1$ and $b = 2$ and a Beta skew-normal density with parameters λ , $a = 2$ and $b = 1$, which are the density function of the smallest and the largest statistic from a sample of size 2 of a skew-normal distribution with parameter λ , respectively. In general, we can see the density function of the skew-normal with parameter λ as mixture of Beta skew-normal distributions with the same parameter λ in the following way:

$$\phi(x; \lambda) = \frac{1}{b} g_{\Phi(x; \lambda)}^B(x; \lambda, 1, b) - \sum_{i=1}^{\infty} (-1)^i \binom{b-1}{i} \frac{1}{1+i} g_{\Phi(x; \lambda)}^B(x; \lambda, 1+i, 1).\quad (3.18)$$

The above formula is obtained setting $a = 1$ in (3.15) and using the property a of the BSN.

We use the preceding expansion (3.15) to present a formula for the moments of the BSN when a and b are integers values.

Theorem 14. *Let $X \sim BSN(\mu, \sigma, \lambda, a, b)$ for integers values of a and b , then*

$$\begin{aligned}E(X^n) &= \mu^n + \frac{2\mu^n}{B(a, b)} \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} \sum_{i=1}^n \binom{n}{i} \left(\frac{\sigma}{\mu}\right)^i * \\ &* \left\{ \sum_{k=0}^{a+j-1} (-1)^k \binom{a+j-1}{k} J_{i, k, \lambda} + (-1)^i J_{i, a+j-1, -\lambda} \right\},\end{aligned}\quad (3.19)$$

where

$$J_{i,k,\lambda} = \int_0^\infty z^i \phi(z) \Phi(\lambda z) (1 - \Phi(z; \lambda))^k dz. \quad (3.20)$$

Proof. The proof follows the same lines of that of theorem 1 in [32]. The density of the random variable X can be written as:

$$g_{\Phi(x;\lambda)}^B(x; \mu, \sigma, \lambda, a, b) = \frac{2}{\sigma B(a, b)} \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} h_j(x), \quad (3.21)$$

where

$$h_j(x) = \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{\lambda(x-\mu)}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}; \lambda\right)^{a+j-1}. \quad (3.22)$$

It follows that

$$E(X^n) = \frac{2}{\sigma B(a, b)} \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} \int_{-\infty}^{\infty} x^n h_j(x) dx. \quad (3.23)$$

Substituting $z = \frac{x-\mu}{\sigma}$ and using the binomial expansion for $(\sigma z + \mu)^n$, we find that the above integral can be written in this way:

$$\int_{-\infty}^{\infty} x^n h_j(x) dx = \sigma \mu^n \sum_{i=0}^n \binom{n}{i} \left(\frac{\sigma}{\mu}\right)^i \int_{-\infty}^{\infty} z^i \phi(z) \Phi(\lambda z) \Phi(z; \lambda)^{a+j-1} dz. \quad (3.24)$$

The integral term in the above equation can be expressed as

$$\begin{aligned} \int_{-\infty}^{\infty} z^i \phi(z) \Phi(\lambda z) \Phi(z; \lambda)^{a+j-1} dz &= \int_0^\infty z^i \phi(z) \Phi(\lambda z) \Phi(z; \lambda)^{a+j-1} dz + \\ &+ (-1)^i \int_0^\infty z^i \phi(z) \Phi(-\lambda z) (1 - \Phi(z; -\lambda))^{a+j-1} dz = \\ &= \sum_{k=0}^{a+j-1} (-1)^k \binom{a+j-1}{k} J_{i,k,\lambda} + (-1)^i J_{i,a+j-1,-\lambda}. \end{aligned} \quad (3.25)$$

On substituting (3.24) and (3.25) into (3.23) and rearranging, we obtain

$$\begin{aligned} E(X^n) &= \frac{2\mu^n}{B(a, b)} \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} \sum_{i=0}^n \left(\frac{\sigma}{\mu}\right)^i \binom{n}{i} * \\ &* \left\{ \sum_{k=0}^{a+j-1} (-1)^k \binom{a+j-1}{k} J_{i,k,\lambda} + (-1)^i J_{i,a+j-1,-\lambda} \right\}, \end{aligned} \quad (3.26)$$

where $J_{i,k,\lambda}$ is given by formula (3.20).

At this point, we shall confine attention to the term corresponding to $i = 0$ and we shall show that it equals μ^n . Employing (3.1) and (3.2), one obtains

$$J_{0,k,\lambda} = \frac{1}{2(k+1)} \int_0^\infty (k+1)\phi(z;\lambda)(1-\Phi(z;\lambda))^k dz = \frac{(1-\Phi(0;\lambda))^{k+1}}{2(k+1)}.$$

So the term inside the brackets in (3.26) for $i = 0$ reduces to

$$\begin{aligned} \sum_{k=0}^{a+j-1} (-1)^k \binom{a+j-1}{k} J_{0,k,\lambda} + J_{0,a+j-1,-\lambda} &= \sum_{k=0}^{a+j-1} (-1)^k \binom{a+j-1}{k} \frac{(1-\Phi(0;\lambda))^{k+1}}{2(k+1)} + \\ &+ \frac{(1-\Phi(0;-\lambda))^{a+j}}{2(a+j)} = \frac{1}{2(a+j)}, \end{aligned} \quad (3.27)$$

where the last equality follows from lemma 1 of Gupta and Nadarajah (2005) [32]. On applying lemma 2 in [32], the term corresponding to $i = 0$ of (3.26) reduces to

$$\frac{2\mu^n}{B(a,b)} \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} \left\{ \frac{1}{2(a+j)} \right\} = \frac{2\mu^n}{B(a,b)} \frac{B(a,b)}{2} = \mu^n. \quad (3.28)$$

The theorem is proved. \square

Remark 13. Clearly, this theorem when $\lambda = 0$ reduces to theorem 1 in [32]. Furthermore, we can note that the authors, in the cited theorem, defined the function

$$I_{i,k} = \int_0^\infty z^i \phi(z)(1-\Phi(z))^k dz, \quad (3.29)$$

which is related to the function $J_{i,k,\lambda}$, when $\lambda = 0$, by the following relation:

$$J_{i,k,0} = \frac{1}{2} I_{i,k}. \quad (3.30)$$

3.1.7 The $BSN(1, n, b)$

As previously noted, general expansions for the moment generating function and the k -th moment of a variable with Beta skew-normal distribution are difficult to find. Exact closed form expressions for the moments can be obtained in certain special cases. One of these cases is discussed in this section.

Theorem 15. *The moment generating function of $X \sim B(1, n, b)$ is*

$$M_X(t) = \frac{2}{B(n, b)} \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} e^{\frac{t^2}{2}} E(\Phi^{2(n+j)-1}(V)), \quad (3.31)$$

where $V \sim N(t, 1)$.

Proof. By applying the binomial expansion and property 2 in section 1.1 it follows that, for $t \in \mathbb{R}$, the moment generating function (m.g.f.) of X is

$$\begin{aligned} M_X(t) &= \frac{2}{B(n, b)} \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} \int_{-\infty}^{\infty} e^{tx} \phi(x) \Phi(x) \Phi(x)^{2(n+j-1)} dx = \\ &= \frac{2}{B(n, b)} \sum_{j=0}^{\infty} (-1)^j \frac{\binom{b-1}{j}}{2(n+j)} \int_{-\infty}^{\infty} 2(n+j) e^{tx} \phi(x) \Phi(x)^{2(n+j)-1} dx = \\ &= \frac{2}{B(n, b)} \sum_{j=0}^{\infty} (-1)^j \frac{\binom{b-1}{j}}{2(n+j)} M_Y(t), \end{aligned} \quad (3.32)$$

where Y is a Balakrishnan skew-normal with parameters 1 and $2(n+j) - 1$. The result in (3.31) then follows by the use of the m.g.f. of the Balakrishnan skew-normal. \square

We can obtain the moments of $X \sim B(1, n, b)$ readily from the derivatives of $M_X(t)$ in (3.32). For example, we get the first moment as

$$\begin{aligned} E(X) &= \frac{1}{B(n, b)} \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} \frac{1}{n+j} E(Y) = \\ &= \frac{1}{B(n, b)} \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} \frac{1}{n+j} \frac{(2(n+j)-1)(n+j)}{\sqrt{\pi}} \frac{1}{c_{(2(n+j)-2)}\left(\frac{1}{\sqrt{2}}\right)}. \end{aligned}$$

Remark 14. Note that in the special case $b = 1$ and $n = 2$, we have

$$E(X) = \frac{6}{\sqrt{\pi}} \left[\arctan(\sqrt{2}) \right], \quad (3.33)$$

which is exactly the mean of the maximum from a sample of size 2 from a $SN(1)$ obtained by Chiogna (1998) [14].

The following theorem provides a recursion formula for the moments of the $BSN(1, n, b)$.

Theorem 16. Let $X \sim BSN(1, n, b)$. Then

$$E(X^k) = \frac{1}{B(n, b)} \sum_{j=0}^{\infty} (-1)^j \frac{\binom{b-1}{j}}{(n+j)} \left\{ (k-1)E(Y^{k-2}) + \frac{2n+2j-1}{2^{\frac{k+1}{2}} \sqrt{\pi}} \frac{2(n+j)}{c_{(2(n+j)-2)\left(\frac{1}{\sqrt{2}}\right)}} E(W^{k-1}) \right\},$$

where $W \sim SNB_{(2(n+j)-2)}\left(\frac{1}{\sqrt{2}}\right)$ and $Y \sim SNB_{(2(n+j)-1)}(1)$.

Proof. The proof follows by combining (1.11) with

$$E(X^k) = \frac{2}{B(n, b)} \sum_{j=0}^{\infty} (-1)^j \frac{\binom{b-1}{j}}{2(n+j)} E(Y^k),$$

where $Y \sim SNB_{(2(n+j)-1)}(1)$. □

Remark 15. It should be noted that similar results can be provided for the $BSN(-1, b, n)$ distribution. This is due to the fact that, as previously noted, if $X \sim BSN(1, n, b)$ then $-X \sim BSN(-1, b, n)$.

3.2 Further results

In this section we present some results concerning the SNB distribution and link the distributions introduced in section 1.2 with the Beta skew-normal. First we consider two results about the Balakrishnan skew-normal.

We study the distribution of the largest order statistic from $SNB_m(1)$ and subsequently the distribution of the smallest order statistic from $SNB_m(-1)$. We found that these distributions belong to the family of SNB .

Proposition 6. *Let X_1, \dots, X_n be a random sample from a $SNB_m(1)$. Then*

$$X_{(n)} = \max \{X_1, \dots, X_n\}$$

is a $SNB_k(1)$, where $k = n(m+1) - 1$.

Proof. The proof follows easily using the standard formula for the density of $X_{(n)}$, the largest order statistic of a random sample of size n from the distribution $SNB_m(1)$. \square

In particular, the following corollary holds:

Corollary 11. *Let X_1, \dots, X_n be a random sample from a $SN(1)$. Then*

$$X_{(n)} = \max \{X_1, \dots, X_n\}$$

is a $SNB_{2n-1}(1)$.

Proof. The skew-normal distribution with parameter $\lambda = 1$ is a Balakrishnan skew-normal with parameters $\lambda = 1$ and $m = 1$.

The same result can be established making use of the well-known result for the density of the largest order statistic from the distribution $SN(1)$ and property 2. If $X \sim SN(1)$ then its density function is $\phi(x; 1) = 2\phi(x)\Phi(x)$ and its distribution function is $\Phi(x; 1) = \Phi(x)^2$, for the property 2. The distribution of $X_{(n)}$ is

$$F_{X_{(n)}}(x) = (\Phi(x; 1))^n = (\Phi(x)^{2n}), \quad (3.34)$$

and the relative density function is

$$f_{X_{(n)}}(x) = n\phi(x; 1)\Phi(x; 1)^{n-1} = 2n\phi(x)\Phi(x)^{2n-1}, \quad (3.35)$$

which is the density function of a variable with Balakrishnan skew-normal distribution with parameters $2n - 1$ and $\lambda = 1$. \square

Corollary 12. *Let X_1, \dots, X_n be a random sample from a $SNB_m(-1)$. Then*

$$X_{(1)} = \min\{X_1, \dots, X_n\}$$

is a $SNB_k(-1)$, where $k = n(m + 1) - 1$.

It follows immediately from corollaries 5, 6, 11 and 12 that the *BSN* distribution is related to the skew-normal generalizations introduced in section 1.2. In fact, its density simplifies to the Balakrishnan skew-normal when $b = 1$, $\lambda = 1$ and $a \geq 1$ integer (or $a = 1$, $\lambda = -1$ and $b \geq 1$ integer). Further, if $\lambda = 0$ the *BSN* density reduces to the generalized Balakrishnan skew-normal when a and b are both integers. These consideration have been summarized in the following proposition.

Proposition 7. *The BSN distribution satisfies the following properties:*

- $g_{\Phi(x;1)}^B(x; 1, n, 1) = f_{2n-1, m}(x; 1, 0)$, for all $x \in \mathbb{R}$, *i.e.*
 $BSN(1, n, 1) = TBSN_{2n-1, m}(1, 0)$;
- $g_{\Phi(x;-1)}^B(x; -1, 1, m) = f_{n, 2m-1}(x; 0, -1)$, for all $x \in \mathbb{R}$, *i.e.*
 $BSN(-1, 1, m) = TBSN_{n, 2m-1}(0, -1)$;
- $g_{\Phi(x;0)}^B(x; 0, n, m) = f_{n-1, m-1}(x; 1, -1)$, for all $x \in \mathbb{R}$, *i.e.*
 $BSN(0, n, m) = TBSN_{n-1, m-1}(1, -1)$;

where n and m are positive integer numbers.

Given a random variable $X \sim BSN(\lambda, a, b)$ we are interested in constructing a random variable Y with Kumaraswamy distribution. This goal can be achieved using the below properties which follow easily from properties *g* and *h* of the *BSN*, respectively.

Property 10. If $X \sim \text{BSN}(\lambda, 1, b)$ then $Y = (\Phi(X; \lambda))^{\frac{1}{a}}$ is a $\text{Kum}(a, b)$. In particular, if $X \sim \text{SNB}_{2b-1}(-1)$ then $Y = (1 - \Phi(-X))^{\frac{1}{a}}$ is a $\text{Kum}(a, b)$.

Property 11. If $X \sim \text{BSN}(\lambda, a, 1)$ then $Y = (1 - \Phi(X; \lambda))^{\frac{1}{b}}$ is a $\text{Kum}(b, a)$. In particular, if $X \sim \text{SNB}_{2a-1}(1)$ then $Y = (1 - \Phi(X))^{\frac{1}{b}}$ is a $\text{Kum}(b, a)$.

We now present a theorem about the $\text{BSN}(\lambda, a, b)$ distribution.

Theorem 17. If $X \sim \text{BSN}(\lambda, a, b)$, then $X^2 \xrightarrow{L} \text{B}\chi^2(1, a, b)$, as $\lambda \rightarrow \infty$, where $\text{B}\chi^2(1, a, b)$ is a Beta chi-square random variable with parameters 1, a and b .

Proof. Let $Y = X^2$. We can easily check that the density of the random variable Y is

$$\begin{aligned} f_Y(y) &= \frac{\phi(\sqrt{y})}{B(a, b)\sqrt{y}} \left\{ \Phi(\lambda\sqrt{y}) (\Phi(\sqrt{y}; \lambda))^{a-1} (1 - \Phi(\sqrt{y}; \lambda))^{b-1} + \right. \\ &\quad \left. + \Phi(-\lambda\sqrt{y}) (\Phi(-\sqrt{y}; \lambda))^{a-1} (1 - \Phi(-\sqrt{y}; \lambda))^{b-1} \right\} = \\ &= \frac{1}{B(a, b)} f_{\chi^2(1)}(y) h(y; \lambda, a, b), \quad y > 0, \end{aligned}$$

with

$$\begin{aligned} h(y; \lambda, a, b) &= \left\{ \Phi(\lambda\sqrt{y}) (\Phi(\sqrt{y}; \lambda))^{a-1} (1 - \Phi(\sqrt{y}; \lambda))^{b-1} + \right. \\ &\quad \left. + \Phi(-\lambda\sqrt{y}) (\Phi(-\sqrt{y}; \lambda))^{a-1} (1 - \Phi(-\sqrt{y}; \lambda))^{b-1} \right\}, \end{aligned}$$

and $f_{\chi^2(1)}(\cdot)$ is the chi-square density function. We can note that

$$\begin{aligned} h(y; \lambda, a, b) &\xrightarrow{\lambda \rightarrow \infty} (2\Phi(\sqrt{y}) - 1)^{a-1} (2(1 - \Phi(\sqrt{y})))^{b-1} = \\ &= F_{\chi^2(1)}^{a-1}(y) \left(1 - F_{\chi^2(1)}(y)\right)^{b-1}, \end{aligned} \quad (3.36)$$

where $F_{\chi^2(1)}(\cdot)$ is the chi-square distribution function. Consequently, the density $f_Y(\cdot)$ converges to the density of a Beta chi-square random variable with parameters 1, a and b as $\lambda \rightarrow \infty$. \square

3.2.1 Skewing mechanism

Recently, Ferreira and Steel (2006) [26] have presented a general approach which allows to generate skew distributions. They show that every univariate continuous skew distribution can be obtained from a “perturbation” of a symmetric one as it explained in the following definition:

Definition 6. *A distribution S is said to be a skewed version of the symmetric distribution $F(\cdot)$, generated by the skewing mechanism P , if its pdf is of the form*

$$s(y|f, P) = f(y)p(F(y)), \quad y \in \mathbb{R}, \quad (3.37)$$

where $f(\cdot)$ and $F(\cdot)$ are the pdf and cdf of a symmetric distribution on the real line, respectively, and $p(\cdot)$ ($P(\cdot)$) is the pdf (cdf) of a distribution on $(0, 1)$.

Note that, if $F(\cdot)$ is the standard normal distribution and $p(\cdot)$ on $(0, 1)$ is given by

$$p(u; \lambda, a, b) = \frac{2}{B(a, b)} \Phi(\lambda \Phi^{-1}(u)) (\Phi(\Phi^{-1}(u); \lambda))^{a-1} (1 - \Phi(\Phi^{-1}(u); \lambda))^{b-1}, \quad (3.38)$$

formula (3.37) reduces to a Beta skew-normal with parameters λ , a and b .

Then the pdf of a Beta skew-normal with parameters λ , a and b can be seen as a weighted version of $\phi(y)$, with weight function given by $p(\Phi(y); \lambda, a, b)$.

Abtahi et al. (2011) [2] give the following definition:

Definition 7. *A random variable $X_{f,p}$ is said to have a unified skewed distribution with functional parameters f and p , if its pdf is of the form (3.37).*

We denote a random variable with this unified skewed distribution by $X_{f,p} \sim USD(f, p)$.

Here, we recall a proposition given by Abtahi et al. (2011) [2].

Proposition 8. *Let U and V be two independent random variables with pdfs (cdf) f (F) on the real line and p on $(0,1)$, respectively.*

- *When $W = V - F(U)$, the conditional distribution of U given $(W = 0)$ is $USD(f, p)$.*
- *$F(X_{f,p}) \stackrel{d}{=} V$, i.e. $F(X_{f,p})$ and V have the same distribution p .*

The following corollary arises naturally from the above proposition.

Corollary 13. *Let U and V be two independent random variables with pdfs (cdf) ϕ (Φ) on the real line and p on $(0,1)$ given by equation (3.38), respectively.*

- *When $W = V - \Phi(U)$, the conditional distribution of U given $(W = 0)$ is $BSN(\lambda, a, b)$.*
- *Let $X \sim BSN(\lambda, a, b)$. Then $\Phi(X) \stackrel{d}{=} V$.*

3.3 Bounds of the moments and the variance of the Beta-generated distribution

Several authors have given methods of finding bounds for the moments of order statistics. One of the earliest result is that derived by Gumbel (1954) [30] and Hartley and David (1954) [33]. Different methods are required for the variance of the order statistics. Following the idea of these works, we apply Hölder's inequalities and Hoeddfing's identity to find inequalities for the moments and the variance of the Beta-generated distribution.

3.3.1 Bounds of the moments

In this section we assume that X and Y have distributions $G_{F(\cdot)}^B$ and $F(\cdot)$, respectively.

Theorem 18. *Let $k > 0$, $p > 1$ and $E(Y^{kp}) < \infty$. Then we have*

$$E(X^k) \leq \frac{1}{B(a,b)} \left(E(Y^{kp}) \right)^{\frac{1}{p}} \left(B\left(\frac{pa-1}{p-1}, \frac{pb-1}{p-1}\right) \right)^{1-\frac{1}{p}}. \quad (3.39)$$

Proof. Proof is based on Hölder's inequality. For an arbitrary distribution function $F(\cdot)$ the k -th moment of the Beta-generated distribution is given by the following formula:

$$E(X^k) = \int_{-\infty}^{\infty} \frac{1}{B(a,b)} x^k (F(x))^{a-1} (1-F(x))^{b-1} dx. \quad (3.40)$$

The latter integral, after the change of variable $y = F(x)$, can be rewritten as

$$E(X^k) = \int_0^1 \frac{1}{B(a,b)} (F^{-1}(y))^k y^{a-1} (1-y)^{b-1} dy. \quad (3.41)$$

Now we apply Hölder's inequality to last formula and obtain the following expression:

$$\begin{aligned} E(X^k) &\leq \frac{1}{B(a,b)} \left(\int_0^1 (F^{-1}(y))^{kp} dy \right)^{\frac{1}{p}} \left(\int_0^1 y^{\frac{p}{p-1}(a-1)} (1-y)^{\frac{p}{p-1}(b-1)} dy \right)^{1-\frac{1}{p}} = \\ &= \frac{1}{B(a,b)} \left(E(Y^{kp}) \right)^{\frac{1}{p}} \left(B\left(\frac{p(a-1)}{p-1} + 1, \frac{p(b-1)}{p-1} + 1\right) \right)^{1-\frac{1}{p}} = \\ &= \frac{1}{B(a,b)} \left(E(Y^{kp}) \right)^{\frac{1}{p}} \left(B\left(\frac{pa-1}{p-1}, \frac{pb-1}{p-1}\right) \right)^{1-\frac{1}{p}}. \end{aligned}$$

□

3.3.2 Bounds of the variance of the Beta-generated distribution

Let $X \sim G_{F(\cdot)}^B(\cdot, a, b)$, with $a > 1$ and $b > 1$. We are interested in finding a bound for the variance of X in function of the variance of $Y \sim F(\cdot)$.

Let us introduce the notations:

$$\begin{aligned} G(x) &= I_x(a, b), & g(x) &= G'(x), \\ t_1(x) &= \frac{G(x)}{x}, & t_2(y) &= \frac{1 - G(y)}{1 - y}, \\ t(x, y) &= t_1(x)t_2(y), & t(x) &= t(x, x), \end{aligned}$$

with $0 < x \leq y < 1$.

We will need the following lemma which is a trivial extension of lemma 2.1 of Papadatos (1995) [51].

Lemma 3. *Let $a > 1$ and $b > 1$. Then there exist unique numbers $\rho_1 = \rho_1(a, b)$, $\rho_2 = \rho_2(a, b)$ satisfying*

$$0 < \rho_1 < \frac{a-1}{a+b-2} < \rho_2 < 1, \quad (3.42)$$

such that, for $0 < x < y < 1$:

1. $t_1(x)$ strictly increases in $(0, \rho_2)$ and strictly decreases in $(\rho_2, 1)$ and similarly $t_2(y)$ strictly increases in $(0, \rho_1)$ and strictly decreases in $(\rho_1, 1)$.
2. If $x \geq \rho_1$ or $y \leq \rho_2$, then $t(x, y) < \max\{t(x), t(y)\}$.
3. If $x < \rho_1$ and $y > \rho_2$, then $t(x, y) < t(\rho_1, \rho_2) < \max\{\rho_1, \rho_2\}$.
4. There exists a unique $x_0 = x_0(a, b) \in (\rho_1, \rho_2)$ such that the function $t(x)$ strictly increases in $(0, x_0)$ and strictly decreases in $(x_0, 1)$.

Proof. The proof follows the same lines of that given by Papadatos (1995) [51].

1. The derivative of the function $t_1(x)$ is $t_1'(x) = \frac{xg(x) - G(x)}{x^2}$. The numerator of $t_1'(x)$ has derivative $xg'(x)$ which attains its maximum at the unique

point $x = \frac{a-1}{a+b-2}$ and by taking account that $\lim_{x \rightarrow 0^+} xg(x) - G(x) = 0$, and $\lim_{x \rightarrow 1^-} xg(x) - G(x) = -1$, we readily see that $xg(x) - G(x) = 0$ has a unique root $\rho_2 = \rho_2(a, b)$ which lies in $(\frac{a-1}{a+b-2}, 1)$.

Then $xg(x) - G(x) > 0$ for $x \in (0, \rho_2)$, and $xg(x) - G(x) < 0$ for $x \in (\rho_2, 1)$.

Similar arguments show that t_2 strictly increases in $(0, \rho_1)$ and strictly decreases in $(\rho_1, 1)$. In fact, the derivative of $t_2(y)$ is $t_2'(y) = \frac{1-G(y)-g(y)(1-y)}{(1-y)^2}$.

The function $1 - G(y) - g(y)(1 - y)$ has derivative $-g'(y)(1 - y)$ which is positive if $y < \frac{a-1}{a+b-2}$, and negative if $y > \frac{a-1}{a+b-2}$. Since

$$\lim_{y \rightarrow 0^+} 1 - G(y) - g(y)(1 - y) = 1 \text{ and } \lim_{y \rightarrow 1^-} 1 - G(y) - g(y)(1 - y) = 0,$$

we deduce that $1 - G(y) - g(y)(1 - y) = 0$ has a unique root $\rho_1 = \rho_1(a, b)$ that is on the interval $(0, \frac{a-1}{a+b-2})$. Hence, $1 - G(y) - g(y)(1 - y) > 0$ for $y \in (0, \rho_1)$ and $1 - G(y) - g(y)(1 - y) < 0$ if $y \in (\rho_1, 1)$.

2. Let $x < y$. If $\rho_1 \leq x$, then $t(x, y) = t_1(x)t_2(y) < t_1(x)t_2(x) = t(x)$.

Similarly, if $y \leq \rho_2$, it follows that $t(x, y) = t_1(x)t_2(y) < t_1(y)t_2(y) = t(y)$.

3. If $x < \rho_1$ and $y < \rho_2$, we have $t(x, y) = t_1(x)t_2(y) < t_1(\rho_1)t_2(\rho_2) = t(\rho_1, \rho_2)$.

4. Clearly, $\lim_{x \rightarrow 0^+} t(x) = \lim_{x \rightarrow 1^-} t(x) = 0$. Furthermore, the function $t(x)$ strictly increases in $(0, \rho_1]$ and strictly decreases in $[\rho_2, 1)$. Hence, we have only to study $t(x)$ in (ρ_1, ρ_2) . To do that we verify the log-concavity of t_1 and t_2 in the intervals $(0, \rho_2)$ and $(\rho_1, 1)$, respectively.

We observe that

$$\left(\log \left(\frac{G(x)}{x} \right) \right)'' = -\frac{1}{x^2 G^2(x)} [x^2 g(x)^2 - G^2(x) - x^2 g'(x)G(x)]. \quad (3.43)$$

Furthermore, the function $x^2 g(x)^2 - G^2(x) - x^2 g'(x)G(x)$, for $x \in (0, \rho_2)$, majorizes the function $x^2 g^2(x) - xg(x)G(x) - x^2 g'(x)G(x)$, which can be rewritten as

$$x^2 g^2(x) - xg(x)G(x) - x^2 g'(x)G(x) = \frac{xg(x)}{1-x} \{x(1-x)g(x) - [a - (a+b-1)x]G(x)\}. \quad (3.44)$$

The function $r(x) = x(1-x)g(x) - [a - (a+b-1)x]G(x)$ is positive for all x . In fact, it increases because its derivative is $(a+b-1)G(x) - xg(x)$, which is a positive and an increasing function for all x , and moreover,

$$x(1-x)g(x) - (a - (a+b-1)x)G(x) > \lim_{x \rightarrow 0^+} x(1-x)g(x) - (a - (a+b-1)x)G(x) = 0. \quad (3.45)$$

Hence, $\left(\log\left(\frac{G(x)}{x}\right)\right)'' < 0$, for $0 < x < \rho_2$, i.e. $t_1(x)$ is strictly log-concave in $(0, \rho_2)$.

In a similar way one can show that t_2 is log-concave in $(\rho_1, 1)$.

First we note that

$$\left(\log\left(\frac{1-G(y)}{1-y}\right)\right)'' = \frac{-1}{(1-y)^2(1-G(y))^2} \left[(1-y)^2 g(y)^2 + (1-G(y))^2 - (1-y)^2 g'(y)(1-G(y)) \right],$$

hence, we observe that, for $\rho_1 < y < 1$, the function inside the brackets majorizes the function

$$g^2(y)(1-y)^2 + g'(y)(1-G(y))(1-y)^2 - (1-y)(1-G(y))g(y). \quad (3.46)$$

The following relation holds:

$$g^2(y)(1-y)^2 + g'(y)(1-G(y))(1-y)^2 - (1-y)(1-G(y))g(y) = \frac{g(y)(1-y)}{y} * \{y(1-y)g(y) + (1-G(y))[a-1-(a+b-1)y]\}.$$

It is immediate to verify that the function

$$f(y) = y(1-y)g(y) + (1-G(y))[a-1-(a+b-1)y]$$

decreases because its derivative is $g(y)(1-y) - (a+b-1)G(y)$, which is negative for all y . Obviously,

$$f(y) > \lim_{y \rightarrow 1^-} f(y) = 0,$$

and consequently, $f(y)$ is positive for all y .

Then $\left(\log\left(\frac{1-G(y)}{1-y}\right)\right)'' < 0$, $\rho_1 < x < 1$, that is, $t_2(x)$ is strictly log-concave in $(\rho_1, 1)$. Hence, $t(x) = t_1(x)t_2(x)$ is a strictly log-concave function in (ρ_1, ρ_2) , and the lemma is proved.

□

Definition 8. The maximum variance function $\sigma_b^2(a)$ is defined by the following relation

$$\sigma_b^2(a) = \sup_{0 < x < 1} \left(\frac{G(x)(1-G(x))}{x(1-x)} \right), \quad a > 1 \text{ and } b > 1. \quad (3.47)$$

Remark 16. It is of interest to point out that $\sigma_b^2(a)$, as $\sigma_n^2(k)$ defined in [51], does not have a closed form. However, it is possible to identify the following behaviour of $\sigma_b^2(a)$:

- if $a = 1$, then $\sigma_b^2(a) = b$;
- if $b = 1$, then $\sigma_b^2(a) = a$;
- if $a = b = 1$, then $\sigma_b^2(a) = 1$.

Theorem 19. Let $X \sim G_{F(\cdot)}^B(\cdot; a, b)$, $Y \sim F(\cdot)$, $a > 1$ and $b > 1$. Then

$$\text{Var}(X) \leq \sigma_b^2(a)\text{Var}(Y). \quad (3.48)$$

Proof. The proof proceeds along the same lines as that of theorem 3.1 of Papadatos at page 189 (see [51]), which is based on Hoeffding's identity for the covariance. We remind Hoeffding's identity for the covariance of two random variable X and Y :

$$\text{Cov}(X, Y) \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H(x, y) - H(x, \infty)H(\infty, y)) dy dx, \quad (3.49)$$

where H is the bivariate distribution function of the random vector (X, Y) . □

We have carried out a numerical study in order to compare σ_{BSN}^2 , the variance of the variable $X \sim BSN(\lambda, a, b)$, and $\sigma_b^2(a)Var(Y)$.

The results found are reported in table 3.2. We observe that if the parameter a takes the value 1 and if b is large then $\sigma_b^2(a)Var(Y)$ is not close to σ_{BSN}^2 . The same situation occurs when $b = 1$ and a is large.

Moreover, if $\lambda = 0$ then $var(Y) = 1$ and the maximum variance coincides with $\sigma_b^2(a)$.

All computations have been done using the software R.

a	b	λ	σ_{BSN}^2	$\sigma_b^2(a)Var(Y)$
1	1	-10	0.3696834	0.3696834
1	1	0	1	1
1	1	10	0.3696834	0.3696834
1	10	-10	0.2625293	3.696335
1	10	0	0.3443438	9.99865
1	10	10	0.01859684	3.696335
2	10	-10	0.139195	0.8695345
2	10	0	0.2051976	2.352106
2	10	10	0.02108108	0.8695345
10	1	-10	0.01859684	3.696335
10	1	0	0.3443438	9.99865
10	1	10	0.2625293	3.696335
10	10	-10	0.03145790	0.3696834
10	10	0	0.08079098	1
10	10	10	0.03145790	0.3696834

Table 3.2: The variance of the $BSN(\lambda, a, b)$ and $\sigma_b^2(a)Var(Y)$ for different values of a , b and λ

3.4 Maximum likelihood estimation

We now determine the maximum likelihood estimates (MLEs) of the parameters of the *BSN* distribution. Let x_1, \dots, x_N be a random sample of size N from a *BSN*($\mu, \sigma, \lambda, a, b$) distribution. The log-likelihood function $l(\boldsymbol{\xi})$ for the vector of parameters $\boldsymbol{\xi} = (\mu, \sigma, \lambda, a, b)$ can be written as

$$\begin{aligned} l(\boldsymbol{\xi}) &= N \log 2 - N \log(\sigma) - N \log B(a, b) + \sum_{i=1}^N \log \left(\phi \left(\frac{x_i - \mu}{\sigma} \right) \right) + \\ &+ \sum_{i=1}^N \log \left(\Phi \left(\lambda \left(\frac{x_i - \mu}{\sigma} \right) \right) \right) + (a-1) \sum_{i=1}^N \log \left(\Phi \left(\frac{x_i - \mu}{\sigma}; \lambda \right) \right) + \\ &+ (b-1) \sum_{i=1}^N \log \left(1 - \Phi \left(\frac{x_i - \mu}{\sigma}; \lambda \right) \right). \end{aligned} \quad (3.50)$$

The components of the score vector $U(\boldsymbol{\xi})$ are given by

$$\begin{aligned} U_a(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial a} = -N(\psi(a) - \psi(a+b)) + \sum_{i=1}^N \log v_i; \\ U_b(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial b} = -N(\psi(b) - \psi(a+b)) + \sum_{i=1}^N \log(1 - v_i); \\ U_\mu(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial \mu} = \frac{1}{\sigma} \sum_{i=1}^N z_i - \frac{\lambda}{\sigma} \sum_{i=1}^N y_i - \frac{a-1}{\sigma} \sum_{i=1}^N w_i + \frac{b-1}{\sigma} \sum_{i=1}^N t_i; \\ U_\sigma(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^N z_i^2 - \frac{\lambda}{\sigma} \sum_{i=1}^N z_i y_i - \frac{a-1}{\sigma} \sum_{i=1}^N z_i w_i + \frac{b-1}{\sigma} \sum_{i=1}^N z_i t_i; \\ U_\lambda(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial \lambda} = \sum_{i=1}^N z_i y_i + (a-1) \sum_{i=1}^N \frac{\partial v_i}{\partial \lambda} - (b-1) \sum_{i=1}^N \frac{\partial v_i}{\partial \lambda}; \end{aligned}$$

where $\psi(t) = \frac{d \log(\Gamma(t))}{dt}$ is the di-gamma function and

$$\begin{aligned} z_i &= \frac{x_i - \mu}{\sigma}; & v_i &= \Phi \left(\frac{x_i - \mu}{\sigma}; \lambda \right); & y_i &= \frac{\phi \left(\lambda \left(\frac{x_i - \mu}{\sigma} \right) \right)}{\Phi \left(\lambda \left(\frac{x_i - \mu}{\sigma} \right) \right)}; \\ w_i &= \frac{\phi \left(\frac{x_i - \mu}{\sigma}; \lambda \right)}{\Phi \left(\frac{x_i - \mu}{\sigma}; \lambda \right)}; & t_i &= \frac{\phi \left(\frac{x_i - \mu}{\sigma}; \lambda \right)}{1 - \Phi \left(\frac{x_i - \mu}{\sigma}; \lambda \right)}. \end{aligned}$$

We can find the estimates of the unknown parameters by maximum likelihood method by setting the above expressions equal to zero and solving them simultaneously.

The elements of the observed information matrix for the vector of parameters $\boldsymbol{\xi} = (\mu, \sigma, \lambda, a, b)$ are

$$U_{aa}(\boldsymbol{\xi}) = -N(\psi'(a) - \psi'(a+b));$$

$$U_{bb}(\boldsymbol{\xi}) = -N(\psi'(b) - \psi'(a+b));$$

$$U_{ab}(\boldsymbol{\xi}) = N\psi'(a+b);$$

$$U_{a\mu}(\boldsymbol{\xi}) = -\frac{1}{\sigma} \sum_{i=1}^N w_i;$$

$$U_{b\mu}(\boldsymbol{\xi}) = \frac{1}{\sigma} \sum_{i=1}^N t_i;$$

$$U_{a\sigma}(\boldsymbol{\xi}) = -\frac{1}{\sigma} \sum_{i=1}^N w_i z_i;$$

$$U_{b\sigma}(\boldsymbol{\xi}) = \frac{1}{\sigma} \sum_{i=1}^N t_i z_i;$$

$$U_{a\lambda}(\boldsymbol{\xi}) = \sum_{i=1}^N \frac{\frac{\partial v_i}{\partial \lambda}}{v_i};$$

$$U_{b\lambda}(\boldsymbol{\xi}) = \sum_{i=1}^N \frac{-\frac{\partial v_i}{\partial \lambda}}{1-v_i};$$

$$U_{\mu\mu}(\boldsymbol{\xi}) = -\frac{N}{\sigma^2} - \frac{\lambda}{\sigma} \sum_{i=1}^N \frac{\partial y_i}{\partial \mu} - \frac{(a-1)}{\sigma} \sum_{i=1}^N \frac{\partial w_i}{\partial \mu} + \frac{b-1}{\sigma} \sum_{i=1}^N \frac{\partial t_i}{\partial \mu};$$

$$U_{\mu\sigma}(\boldsymbol{\xi}) = -\frac{2}{\sigma^2} \sum_{i=1}^N z_i + \frac{\lambda}{\sigma^2} \sum_{i=1}^N y_i - \frac{\lambda}{\sigma} \sum_{i=1}^N \frac{\partial y_i}{\partial \sigma} + \frac{a-1}{\sigma^2} \sum_{i=1}^N w_i - \frac{a-1}{\sigma} \sum_{i=1}^N \frac{\partial w_i}{\partial \sigma} +$$

$$-\frac{b-1}{\sigma^2} \sum_{i=1}^N t_i + \frac{b-1}{\sigma} \sum_{i=1}^N \frac{\partial t_i}{\partial \sigma};$$

$$U_{\mu\lambda}(\boldsymbol{\xi}) = -\frac{1}{\sigma} \sum_{i=1}^N y_i - \frac{\lambda}{\sigma} \sum_{i=1}^N \frac{\partial y_i}{\partial \lambda} - \frac{a-1}{\sigma} \sum_{i=1}^N \frac{\partial w_i}{\partial \lambda} + \frac{b-1}{\sigma} \sum_{i=1}^N \frac{\partial t_i}{\partial \lambda};$$

$$\begin{aligned}
U_{\sigma\sigma}(\boldsymbol{\xi}) &= \frac{N}{\sigma^2} - 3 \sum_{i=1}^N (z_i)^2 + 2 \frac{\lambda}{\sigma^2} \sum_{i=1}^N z_i y_i - \frac{\lambda}{\sigma} \sum_{i=1}^N z_i \frac{\partial y_i}{\partial \sigma} + \frac{2(a-1)}{\sigma^2} \sum_{i=1}^N z_i w_i + \\
&\quad - \frac{(a-1)}{\sigma} \sum_{i=1}^N z_i \frac{\partial w_i}{\partial \sigma} - \frac{2(b-1)}{\sigma^2} \sum_{i=1}^N z_i t_i + \frac{(b-1)}{\sigma} \sum_{i=1}^N z_i \frac{\partial t_i}{\partial \sigma}; \\
U_{\sigma\lambda}(\boldsymbol{\xi}) &= -\frac{1}{\sigma} \sum_{i=1}^N z_i y_i - \frac{\lambda}{\sigma} \sum_{i=1}^N z_i \frac{\partial y_i}{\partial \lambda} - \frac{(a-1)}{\sigma} \sum_{i=1}^N z_i \frac{\partial w_i}{\partial \lambda} + \frac{(b-1)}{\sigma} \sum_{i=1}^N z_i \frac{\partial t_i}{\partial \lambda}; \\
U_{\lambda\lambda}(\boldsymbol{\xi}) &= \sum_{i=1}^N \frac{\partial y_i}{\partial \lambda} + (a-1) \sum_{i=1}^N \left(\frac{\frac{\partial^2 v_i}{\partial \lambda^2}}{v_i} - \frac{\left(\frac{\partial v_i}{\partial \lambda}\right)^2}{v_i^2} \right) - (b-1) \sum_{i=1}^N \left(\frac{\frac{\partial^2 v_i}{\partial \lambda^2}}{1-v_i} + \frac{\left(\frac{\partial v_i}{\partial \lambda}\right)^2}{(1-v_i)^2} \right);
\end{aligned}$$

where $\psi'(\cdot)$ is the derivative of the di-gamma function, which is called tri-gamma function.

Chapter 4

The Kumaraswamy skew-normal distribution

In this chapter we propose another generalization of the skew-normal distribution, referred to as the Kumaraswamy skew-normal, which is a special case of the Kumaraswamy generalized distribution. There is some parallelism between this chapter and chapter 3. In fact, the Kumaraswamy skew-normal and the Beta skew-normal fulfil similar properties. A range of mathematical properties of the Kumaraswamy skew-normal distribution is considered in sections 1 to 2. In section 3 the parameters of the new model are estimated by maximum likelihood and the observed information matrix is derived. The bivariate Kumaraswamy skew-normal distribution is introduced and studied in section 4. In the last section we present the generalized Beta skew-normal distribution.

4.1 The Kumaraswamy skew-normal distribution

We start by defining the Kumaraswamy skew-normal distribution and presenting some of its properties.

4.1.1 Definition and simple properties

Following the procedure of Cordeiro and de Castro (2011) [17] described in section 1.4, we define a generalization of the skew-normal distribution which satisfies some of the properties of the Beta skew-normal one.

Replacing in (1.36) $F(x)$ by $\Phi(x; \lambda)$, we obtain the Kumaraswamy skew-normal distribution, with distribution function given by

$$G_{\Phi(x; \lambda)}^K(x; \lambda, a, b) = 1 - (1 - \Phi(x; \lambda)^a)^b, \quad (4.1)$$

and probability density function

$$g_{\Phi(x; \lambda)}^K(x; \lambda, a, b) = ab\phi(x; \lambda)(\Phi(x; \lambda))^{a-1}(1 - \Phi(x; \lambda)^a)^{b-1}. \quad (4.2)$$

Throughout the chapter, we shall denote the Kumaraswamy skew-normal distribution with vector of parameters $\boldsymbol{\xi} = (\lambda, a, b)$ by $KwSN(\lambda, a, b)$.

This family can be easily generalized by means of linear transformations to introduce a location parameter μ and a scale parameter $\sigma > 0$. Thus if $X \sim KwSN(\lambda, a, b)$, then $Y = \mu + \sigma X$ is a Kumaraswamy skew-normal with vector of parameters $\boldsymbol{\xi} = (\mu, \sigma, \lambda, a, b)$. We indicate Y by $Y \sim KwSN(\mu, \sigma, \lambda, a, b)$.

However, in the following sections we will concentrate on the standard form of the distribution.

The properties derived for the $KwSN$ distribution can be easily extended to the transformed distribution.

We now mention some simple properties of $KwSN(\lambda, a, b)$ density in (4.2):

Properties of $KwSN(\lambda, a, b)$:

- a. $g_{\Phi(x; \lambda)}^K(x; \lambda, 1, 1) = \phi(x; \lambda)$, for all $x \in \mathbb{R}$, i.e. $KwSN(\lambda, 1, 1) = SN(\lambda)$.
- b. $g_{\Phi(x; 0)}^K(x; 0, a, b) = g_{\Phi(x)}^K(x; a, b)$, for all $x \in \mathbb{R}$, i.e. $KwSN(0, a, b) = KwN(a, b)$.
- c. $g_{\Phi(x; 0)}^K(x; 0, 1, 1) = \phi(x)$, for all $x \in \mathbb{R}$, i.e. $KwSN(0, 1, 1) = N(0, 1)$.
- d. $g_{\Phi(x; 1)}^K(x; 1, \frac{1}{2}, 1) = \phi(x)$, for all $x \in \mathbb{R}$, i.e. $KwSN(1, \frac{1}{2}, 1) = N(0, 1)$.
- e. $g_{\Phi(x; -1)}^K(x; -1, 1, \frac{1}{2}) = \phi(x)$, for all $x \in \mathbb{R}$, i.e. $KwSN(-1, 1, \frac{1}{2}) = N(0, 1)$.
- f. If $X \sim KwSN(\lambda, a, b)$, then $Y = \Phi(X; \lambda)$ is a $Kw(a, b)$.
- f. If $X \sim KwSN(\lambda, a, b)$, then $Y = \Phi(X; \lambda)^a$ is a $Kw(1, b)$.
- h. If $X \sim KwSN(\lambda, a, b)$, then $Y = 1 - \Phi(X; \lambda)^a$ is a $Kw(b, 1)$.
- i. As $\lambda \rightarrow +\infty$, $g_{\Phi(x; \lambda)}^K(x; \lambda, a, b)$ tends to the Kumaraswamy half-normal density.

Remark 17. *We note here that the standard normal, the skew-normal and the Kumaraswamy-normal laws are included in this class as special cases. We also observe that item i indicates that as $\lambda \rightarrow \infty$ the $KwSN$ density tends to the Kumaraswamy half-normal one.*

Proof. The results follow immediately taking into account expression (4.2) and the basic properties of the skew-normal distribution. \square

The $KwSN$ distribution can be easily simulated in two ways:

- because its distribution function has closed form and does not involve any special functions we can use the transformation integral: if Y has an uniform distribution then the variable $X = \Phi^{-1} \left(\left(1 - (1 - Y)^{\frac{1}{b}} \right)^{\frac{1}{a}}; \lambda \right)$ has $KwSN(\lambda, a, b)$ distribution;

- if Y has a Kumaraswamy distribution with parameters a and b , then the variable $X = \Phi^{-1}(Y; \lambda)$ has $KwSN(\lambda, a, b)$ distribution;

where $\Phi^{-1}(\cdot; \lambda)$ is the quantile function of the skew-normal distribution. Plots of the density function (4.2) are illustrated in figure 4.1.

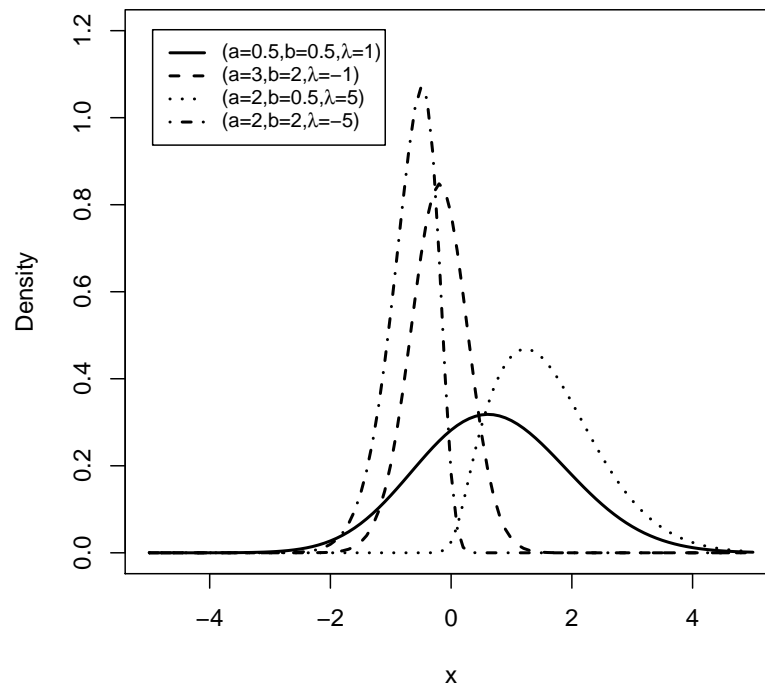


Figure 4.1: The $KwSN(\lambda, a, b)$ for different values of λ , a and b

Numerically, we have noted that the BSN and the $KwSN$ have different shapes. In fact, those values of the parameters a , b and λ , for which the BSN is bimodal, make the $KwSN$ unimodal.

4.1.2 Moment generating function and moments

Let us find the moment generating function of $KwSN(\lambda, a, b)$.

Property 12. *The moment generating function of $X \sim \text{KwSN}(\lambda, a, b)$ is given by*

$$M_X(t) = 2abe^{\frac{t^2}{2}} E_Z \left((\Phi(Z; \lambda))^{a-1} (1 - \Phi(Z; \lambda))^b \Phi(\lambda Z) \right), \quad (4.3)$$

where $Z \sim N(t, 1)$.

We also get a recursive formula for the k -th moment.

Property 13. *Let $k \in \mathbb{N}$ and $k \geq 1$. If $X \sim \text{KwSN}(\lambda, a, b)$, with $a > 1$ and $b > 1$ then*

$$\begin{aligned} E_X(X^k) &= (k-1)E_X(X^{k-2}) + \lambda E_X \left(X^{k-1} \frac{\phi(\lambda X)}{\Phi(\lambda X)} \right) + \\ &+ aE_U \left(U^{k-1} \phi(U; \lambda) \right) - \frac{ba^2}{2a-1} E_V \left(V^{k-1} \phi(V; \lambda) \right), \end{aligned}$$

where $U \sim \text{KwSN}(\lambda, a-1, b)$ and $V \sim \text{KwSN}(\lambda, 2a-1, b-1)$ are independent random variables.

Proof. The statement follows by applying integration by parts and noting that $\frac{\partial \phi(x)}{\partial x} = -x\phi(x)$ (see Arnold et al. (1992) [4]). \square

The following proposition follows by simple changes of variables and by the properties of the Kumaraswamy distribution.

Proposition 9. *Let $X \sim \text{KwSN}(\lambda, a, b)$. Then the following statements hold.*

1. $E_X(1 - \Phi(X; \lambda))^a = \frac{b}{1+b}$;
2. $E_X(\ln(1 - \Phi(X; \lambda))^a) = -\frac{1}{b}$;
3. $E_X(\Phi(X; \lambda)) = bB\left(1 + \frac{1}{a}, b\right)$;
4. $E_X(\ln(\Phi(X; \lambda))) = -\frac{1}{a} \left(\gamma + \psi(b) + \frac{1}{b} \right)$;

where γ is the Euler-Mascheroni constant and $\psi(\cdot)$ is the di-gamma function.

We refer to [1] for details on the Euler-Mascheroni constant.

(a)							(b)						
a	b	λ	μ_{KwSN}	σ_{KwSN}	γ_1	γ_2	a	b	λ	μ_{KwSN}	σ_{KwSN}	γ_1	γ_2
0.25	0.25	-10	-0.75703	1.1661	-1.6750	5.9148	0.5	10	-10	-2.9068	0.7271	-0.5398	3.48441
		-1	-0.0934	1.7790	-0.433	3.2765			-1	-2.9058	0.7297	-0.5421	3.504
		0	0.6542	2.1563	-0.1813	2.9391			0	-2.6716	0.7786	-0.4525	3.3854
		1	1.1915	1.8483	0.0836	2.9637			1	-1.5388	0.5868	-0.4098	3.3310
		10	1.4990	1.4278	0.8594	3.2475			10	-0.1181	0.1143	0.1497	3.6289
0.25	0.5	-10	-1.3451	1.2995	-1.1236	4.04358	1	0.5	-10	-0.4873	0.5778	-1.3200	4.857
		-1	-1.0909	1.6032	-0.4829	3.2765			-1	0	1	0	3
		0	-0.5631	1.8511	-0.2330	2.9385			0	0.7043	1.2479	0.1372	2.9831
		1	0.1421	1.4962	-0.0435	2.9128			1	1.1665	1.0703	0.3086	3.1161
		10	0.709	0.9207	1.2445	4.3209			10	1.3018	0.9148	0.8262	3.4814
0.5	0.25	-10	-0.4352	0.7881	-1.7257	6.3172	1	1	-10	-0.7939	0.6080	-0.9556	3.8232
		-1	0.3528	1.4585	-0.1240	3.0016			-1	-0.5642	0.8256	-0.1369	3.0617
		0	1.1775	1.8327	0.0502	2.9051			0	0	1	0	3
		1	1.624	1.6193	0.2962	3.075			1	0.5642	0.8256	0.1369	3.0617
		10	1.7707	1.3893	0.7667	3.1913			10	0.7939	0.6080	0.9556	3.8232
0.5	0.5	-10	-0.8324	0.8769	-1.2265	4.3886	10	1	-10	-0.084	0.1364	-0.7082	4.2018
		-1	-0.4696	1.2367	-0.2293	3.0442			-1	0.6744	0.4536	0.3597	3.2722
		0	0.1417	1.4969	-0.0472	2.9225			0	1.5388	0.5868	0.4099	3.3314
		1	0.7044	1.248	0.1378	2.9856			1	1.8675	0.5251	0.5005	3.4685
		10	0.9961	0.9305	1.0123	3.752			10	1.8807	0.5124	0.5744	3.5243
0.5	1	-10	-1.3018	0.9148	-0.8263	3.4815	1	10	-10	-1.8807	0.5124	-0.5744	3.5243
		-1	-1.1665	1.0704	-0.3086	3.1067			-1	-1.8675	0.5251	-0.5005	3.4685
		0	-0.7043	1.2479	-0.1372	2.9831			0	-1.5388	0.5868	-0.4099	3.3314
		1	0	1	0	3			1	-0.6744	0.4536	-0.3597	3.2722
		10	0.4873	0.5777	1.3200	4.8570			10	0.084	0.1364	0.7082	4.2018

Table 4.1: The first moment, the standard deviation, the skewness and the kurtosis of $KwSN(\lambda, a, b)$ for different values of a , b and λ

As noted for moments of the BSN , moments of the $KwSN$ involve integrals that can not be solved explicitly. For this reason we have performed a numerical study to compute them numerically using the software R. In table 4.1 we have reported the values of the mean μ_{KwSN} , standard deviation σ_{KwSN} , skewness γ_1 and kurtosis γ_2 for different values of the parameters a , b and λ . It should be noted that:

- for fixed values of a and b the mean μ_{KwSN} is an increasing function of λ ;
- for fixed values of a and λ the mean μ_{KwSN} and the skewness γ_1 are a decreasing and an increasing function of b , respectively.

4.1.3 Some interesting properties

In this subsection, we now derive the main properties of the $KwSN$ distribution.

First we prove the following theorem:

Theorem 20. *Let $X \sim KwSN(\lambda, a, b)$, $Y \sim KwSN(\lambda, a, d)$ be independent then $X|(Y \geq X) \sim KwSN(\lambda, a, b+d)$.*

Proof. Let $W = X|(Y \geq X)$. The cdf of W is then

$$P(W \leq w) = \frac{\int_{-\infty}^w ab(1 - \Phi(x; \lambda)^a)^d \phi(x; \lambda) (\Phi(x; \lambda))^{(a-1)} (1 - \Phi(x; \lambda)^a)^{(b-1)} dx}{P(Y \geq X)}. \quad (4.4)$$

Also

$$\begin{aligned} P(Y \geq X) &= \int_{-\infty}^{\infty} ab(1 - \Phi(x; \lambda)^a)^d \phi(x; \lambda) (\Phi(x; \lambda))^{(a-1)} (1 - \Phi(x; \lambda)^a)^{(b-1)} dx = \\ &= \frac{b}{b+d}. \end{aligned}$$

By taking derivative from the above expression with respect to w , we have

$$f_W(w) = a(b+d)\phi(w; \lambda) (\Phi(w; \lambda))^{(a-1)} (1 - \Phi(w; \lambda)^a)^{(b+d-1)}, \quad (4.5)$$

and the proof is complete. \square

Hence, we can easily derive the following corollary.

Corollary 14. *Let $X, Y \sim KwSN(\lambda, a, b)$ be independent then*

$$X|(Y \leq X) \sim KwSN(\lambda, a, 2b).$$

The proofs of the following theorems are quite similar to that of theorem (20) and are therefore omitted.

Theorem 21. *Let $X \sim KwSN(\lambda, a, 1)$ be independent of $Y \sim KwSN(\lambda, c, 1)$. Then $X|(Y \leq X) \sim KwSN(\lambda, a+c, 1)$, where a and c are positive real numbers.*

As a special case of this theorem, we have the following one.

Theorem 22. *Let $X \sim KwSN(\lambda, a, 1)$ be independent of a random sample (Y_1, \dots, Y_n) from $SN(\lambda)$, then $X|(Y_{(n)} \leq X) \sim KwSN(\lambda, a+n, 1)$, where $Y_{(n)}$ is the largest order statistic.*

We immediately get the following theorem.

Theorem 23. *Let $X \sim KwSN(\lambda, 1, b)$ be independent of a random sample (Y_1, \dots, Y_n) from $SN(\lambda)$, then $X|(X \leq Y_{(1)}) \sim KwSN(\lambda, 1, b+n)$, where $Y_{(1)}$ is the smallest order statistic.*

We give a generalization of the above theorem as follows:

Theorem 24. *Let $X \sim KwSN(\lambda, 1, b)$ be independent of $Y \sim KwSN(\lambda, 1, d)$. Then $X|(X \leq Y) \sim KwSN(\lambda, 1, b+d)$, where b and d are positive real numbers.*

Next we extend the previous theorems for the family of the Kumaraswamy generalized distributions.

Theorem 25. *Let $X \sim Kw-F(a, b)$, $Y \sim Kw-F(a, d)$ be independent then*

$$X|(Y \leq X) \sim Kw-F(a, b+d).$$

We have in particular the following corollary corresponding to the case $d = b$.

Corollary 15. *Let $X, Y \sim Kw - F(a, b)$ be independent then*

$$X|(Y \leq X) \sim Kw - F(a, 2b).$$

Theorem 26. *Let $X \sim Kw - F(a, 1)$ be independent of a random sample (Y_1, \dots, Y_n) from $f(\cdot)$, then $X|(Y_{(n)} \leq X) \sim Kw - F(a + n, 1)$, where $Y_{(n)}$ is the largest order statistic.*

The next result is an extension of the theorem 26.

Theorem 27. *Let $X \sim Kw - F(a, 1)$ be independent of $Y \sim Kw - F(c, 1)$. Then $X|(Y \leq X) \sim Kw - F(a + c, 1)$, where a and c are positive real numbers.*

Theorem 28. *Let $X \sim Kw - F(1, b)$ be independent of a random sample (Y_1, \dots, Y_n) from $f(\cdot)$, then $X|(X \leq Y_{(1)}) \sim Kw - F(1, b + n)$, where $Y_{(1)}$ is the smallest order statistic.*

Theorem 28 can be generalized as follows:

Theorem 29. *Let $X \sim Kw - F(1, b)$ be independent of $Y \sim Kw - F(1, d)$. Then $X|(X \leq Y) \sim Kw - F(1, b + d)$, where b and d are positive real numbers.*

4.1.4 Expansion for the density function

Here, we give a simple expansion for the $KwSN$ density function. Application of (1.23) to equation (4.1), if b is real non-integer, gives

$$G_{\Phi(x;\lambda)}^K(x; \lambda, a, b) = \sum_{i=0}^{\infty} (-1)^i ab \binom{b-1}{i} \Phi(x; \lambda)^{a(1+i)-1}. \quad (4.6)$$

Correspondingly, the density function (4.2) can be written as

$$g_{\Phi(x;\lambda)}^B(x; \lambda, a, b) = \sum_{i=0}^{\infty} (-1)^i \frac{ab}{a(1+i)-1} \binom{b-1}{i} g_{\Phi(x;\lambda)}^K(x; \lambda, a(1+i)-1, 1). \quad (4.7)$$

The density $g_{\Phi(x;\lambda)}^K(x; \lambda, a, b)$ can be seen as an infinite mixture of $KwSN(\lambda, a(1+i)-1, 1)$ densities with constant weights $(-1)^i \frac{ab}{a(1+i)-1}$.

For b integer, the previous sums stop at $b-1$. If a is real integer we can expand $\Phi(x; \lambda)^{a(1+i)-1}$ as follows

$$\begin{aligned} \Phi(x; \lambda)^{a(1+i)-1} &= \sum_{j=0}^{\infty} (-1)^j \binom{a(1+i)-1}{j} (1 - \Phi(x; \lambda))^j = \\ &= \sum_{j=0}^{\infty} \sum_{r=0}^j (-1)^{j+r} \binom{a(1+i)-1}{j} \binom{j}{r} (\Phi(x; \lambda))^r, \end{aligned}$$

and the distribution function takes the following expression

$$G_{\Phi(x;\lambda)}^K(x; a, b) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{r=0}^j (-1)^{i+j+r} ab \binom{b-1}{i} \binom{a(1+i)-1}{j} \binom{j}{r} (\Phi(x; \lambda))^r. \quad (4.8)$$

The density for a real non-integer can be easily obtained from the above equation by differentiation.

Remark 18. *The density function (1.1) of model $SN(\lambda)$, can be represented in the following way:*

$$\begin{aligned} \phi(z; \lambda) &= 2\phi(z)\Phi(\lambda z) = 2\phi(z)\Phi(\lambda z)(1 - \Phi(z; \lambda) + \Phi(z; \lambda)) = \\ &= 2\phi(z)\Phi(\lambda z)(1 - \Phi(z; \lambda)) + 2\phi(z)\Phi(\lambda z)\Phi(z; \lambda) = \\ &= \frac{1}{2} \left(g_{\Phi(z;\lambda)}^K(z; \lambda, 1, 2) + g_{\Phi(z;\lambda)}^K(z; \lambda, 2, 1) \right). \end{aligned} \quad (4.9)$$

In other words the density function of the skew-normal with parameter λ is a mixture between a Kumaraswamy skew-normal density with parameters λ , $a = 1$ and $b = 2$ and a Kumaraswamy skew-normal density with parameters

λ , $a = 2$ and $b = 1$. The density function of the skew-normal with parameter λ can be seen as mixture of Kumaraswamy skew-normal distributions with the same parameter λ in the following way:

$$\phi(x; \lambda) = \frac{1}{b} g_{\Phi(x; \lambda)}^K(x; \lambda, 1, b) - \sum_{i=1}^{\infty} (-1)^i \binom{b-1}{i} g_{\Phi(x; \lambda)}^K(x; \lambda, 1+i, 1). \quad (4.10)$$

The above formula is obtained setting $a = 1$ in (4.7) and using the property a of the KwSN.

The following theorem is nearly identical to the result obtained for the moments of the BSN.

Theorem 30. Let $X \sim \text{KwSN}(\mu, \sigma, \lambda, a, b)$ for integers values of a and b , then

$$E(X^n) = \mu^n + 2ab\mu^n \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} \sum_{i=1}^n \binom{n}{i} \left(\frac{\sigma}{\mu}\right)^i * \left\{ \sum_{k=0}^{a(j+1)-1} (-1)^k \binom{a(j+1)-1}{k} I_{i,k,\lambda} + (-1)^i I_{i,a(j+1)-1,-\lambda} \right\}, \quad (4.11)$$

where

$$I_{i,k,\lambda} = \int_0^{\infty} z^i \phi(z) \Phi(\lambda z) (1 - \Phi(z; \lambda))^k dz. \quad (4.12)$$

Proof. The proof is again analogous to the one given by Gupta and Nadarajah (2005) [32] for theorem 1. If X has the pdf (4.2), then its n -th moment can be written as

$$E(X^n) = 2ab \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} \int_{-\infty}^{\infty} \frac{1}{\sigma} x^n \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\lambda \frac{x-\mu}{\sigma}\right) \left[\Phi\left(\frac{x-\mu}{\sigma}; \lambda\right)\right]^{a(j+1)-1} dx. \quad (4.13)$$

The change of variable $x = \sigma z + \mu$ immediately yields

$$E(X^n) = 2ab\mu^n \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} \sum_{i=0}^n \binom{n}{i} \left(\frac{\sigma}{\mu}\right)^i \int_{-\infty}^{\infty} z^i \phi(z) \Phi(\lambda z) (\Phi(z; \lambda))^{a(j+1)-1} dz. \quad (4.14)$$

In order to obtain (4.11) we split the above integral into two integrals, i.e.

$$\begin{aligned} \int_{-\infty}^{\infty} z^i (\Phi(z; \lambda))^{a(j+1)-1} dz &= \int_0^{\infty} z^i \phi(z) \Phi(\lambda z) (\Phi(z; \lambda))^{a(j+1)-1} dz + \\ &+ (-1)^i \int_0^{\infty} z^i \phi(z) \Phi(-\lambda z) (1 - \Phi(z; -\lambda))^{a(j+1)-1} dz. \end{aligned} \quad (4.15)$$

The first integral in expression (4.15), on using the series representation

$$(\Phi(z; \lambda))^{a(j+1)-1} = \sum_{k=0}^{a(j+1)-1} (-1)^k \binom{a(j+1)-1}{k} (1 - \Phi(z; \lambda))^k, \quad (4.16)$$

becomes

$$\sum_{k=0}^{a(j+1)-1} (-1)^k \binom{a(j+1)-1}{k} \int_0^{\infty} z^i \phi(z) \Phi(\lambda z) (1 - \Phi(z; \lambda))^k dz. \quad (4.17)$$

By putting together expressions (4.14), (4.15) and (4.17) we conclude that

$$\begin{aligned} E(X^n) &= 2ab\mu^n \sum_{j=0}^{b-1} \binom{b-1}{j} (-1)^j \sum_{i=0}^n \left(\frac{\sigma}{\mu}\right)^i \binom{n}{i} * \\ &* \left\{ \sum_{k=0}^{a(j+1)-1} (-1)^k \binom{a(j+1)-1}{k} I_{i,k,\lambda} + (-1)^i I_{i,a(j+1)-1,-\lambda} \right\}. \end{aligned} \quad (4.18)$$

Hence, it follows from lemma 1 of Gupta and Nadarajah (2005) [32] that the term in the brackets in (4.18) for $i = 0$ reduces to

$$\begin{aligned} &\sum_{k=0}^{a+j-1} (-1)^k \binom{a(j+1)-1}{k} I_{0,k,\lambda} + I_{0,a+j-1,-\lambda} = \\ &= \sum_{k=0}^{a(j+1)-1} (-1)^k \binom{a(j+1)-1}{k} \frac{(1 - \Phi(0; \lambda))^{k+1}}{2(k+1)} + \frac{(1 - \Phi(0; -\lambda))^{a+j}}{2(a+j)} = \\ &= \frac{1}{2(a(j+1))}. \end{aligned} \quad (4.19)$$

Finally, from lemma 2 in [32] we conclude that the term for $i = 0$ in (4.18) is equal to μ^n . The proof is complete. \square

Remark 19. The function $I_{i,k,\lambda}$ is the same defined in equation (3.20) in chapter 3.

4.1.5 The $KwSN(1, n, b)$

In this section, we will discuss the moment generating function and the moments of the $KwSN(1, n, b)$ distribution.

Proceeding as in section 3.1.7, we can find the moment generating function of a skew-normal with parameters $\lambda = 1$, a integer and b real using the moment generating function of a Balakrishnan skew-normal.

Theorem 31. *The moment generating function of $X \sim KwSN(1, n, b)$ is*

$$M_X(t) = 2nb \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} e^{\frac{t^2}{2}} E(\Phi^{2n(1+j)-1}(V)), \quad (4.20)$$

where $V \sim N(t, 1)$.

Proof. Proceeding as in theorem 15, it follows that

$$\begin{aligned} M_X(t) &= 2nb \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} \int_{-\infty}^{\infty} e^{tx} \phi(x) \Phi(x) \Phi(x)^{2n(1+j)-2} dx = \\ &= 2nb \sum_{j=0}^{\infty} (-1)^j \frac{\binom{b-1}{j}}{2n(j+1)} \int_{-\infty}^{\infty} 2n(1+j) e^{tx} \phi(x) \Phi(x)^{2n(1+j)-1} dx = \\ &= 2nb \sum_{j=0}^{\infty} (-1)^j \frac{\binom{b-1}{j}}{2n(1+j)} M_Y(t), \end{aligned} \quad (4.21)$$

where Y is a Balakrishnan skew-normal with parameters 1 and $2n(1+j) - 1$. Now, the result in (4.20) follows from the moment generating function of the Balakrishnan skew-normal. \square

By taking the first derivative of the moment generating function, it can be easily proven that the mean is given by the following expression

$$\begin{aligned} E(X) &= b \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} \frac{1}{1+j} E(Y) = \\ &= b \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} \frac{1}{1+j} \frac{(2n(1+j) - 1)(n(1+j))}{\sqrt{\pi}} \frac{1}{c_{(2n(1+j)-2)}\left(\frac{1}{\sqrt{2}}\right)}. \end{aligned}$$

The general moments are given by the following recursion formula.

Theorem 32. *Let $X \sim KwSN(1, n, b)$, then*

$$E(X^k) = b \sum_{j=0}^{\infty} (-1)^j \frac{\binom{b-1}{j}}{(1+j)} \left[(k-1)E(Y^{k-2}) + \frac{2n(1+j)-1}{2^{\frac{k+1}{2}} \sqrt{\pi}} * \right. \\ \left. * \frac{2n(1+j)}{c_{(2n(1+j)-2)}(\frac{1}{\sqrt{2}})} E(W^{k-1}) \right],$$

where $W \sim SNB_{(2n(1+j)-2)}(\frac{1}{\sqrt{2}})$.

Proof. The proof follows the same lines as that of theorem 16 given in section 3.1.7. □

4.2 Further results

The Kumaraswamy skew-normal density reduces to the Balakrishnan skew-normal when $b = 1$, $\lambda = 1$ and $a \geq 1$ integer (or $a = 1$, $\lambda = -1$ and $b \geq 1$ integer). These considerations lead to the following proposition:

Proposition 10. *The KwSN distribution satisfies the following properties:*

- $g_{\Phi(x;1)}^K(x; 1, n, 1) = f_{2n-1, m}(x; 1, 0)$, for all $x \in \mathbb{R}$, i.e.
 $KwSN(1, n, 1) = TBSN_{2n-1, m}(1, 0)$;
- $g_{\Phi(x;-1)}^K(x; -1, 1, m) = f_{n, 2m-1}(x; 0, -1)$, for all $x \in \mathbb{R}$, i.e.
 $KwSN(-1, 1, m) = TBSN_{n, 2m-1}(0, -1)$;

where n and m are positive integer numbers.

Following the notations of Ferreira and Steel (2006) [26] given in section 3.2.1, we can notice that the density function of a Kumaraswamy skew-normal with parameters λ , a and b may be expressed as a weighted version of the normal density function, with $p(\cdot)$ on $(0, 1)$ given by

$$p(u; \lambda, a, b) = ab \Phi(\lambda \Phi^{-1}(u)) (\Phi(\Phi^{-1}(u); \lambda))^{a-1} (1 - \Phi(\Phi^{-1}(u); \lambda))^b)^{b-1}. \quad (4.22)$$

Moreover, the following corollary of proposition 8 holds.

Corollary 16. *Let U and V be two independent random variables with pdfs (cdf) ϕ (Φ) on the real line and p on $(0, 1)$ given by equation (4.22), respectively.*

- *When $W = V - \Phi(U)$, the conditional distribution of U given $(W = 0)$ is $KwSN(\lambda, a, b)$.*
- *Let $X \sim KwSN(\lambda, a, b)$. Then $\Phi(X) \stackrel{d}{=} V$.*

Now we present a theorem about the $KwSN(\lambda, a, b)$ distribution.

Theorem 33. *If $X \sim KwSN(\lambda, a, b)$, then $X^2 \xrightarrow{L} Kw\chi^2(1, a, b)$, $\lambda \rightarrow \infty$, where $Kw\chi^2(1, a, b)$ is a Kumaraswamy chi-square random variable with parameters 1, a and b .*

Proof. Let $Y = X^2$. The density of Y is

$$\begin{aligned} f_Y(y) &= ab\phi(\sqrt{y}) \frac{1}{\sqrt{y}} \left\{ \Phi(\lambda\sqrt{y}) (\Phi(\sqrt{y}; \lambda))^{a-1} (1 - \Phi(\sqrt{y}; \lambda))^a)^{b-1} + \right. \\ &\quad \left. + \Phi(-\lambda\sqrt{y}) (\Phi(-\sqrt{y}; \lambda))^{a-1} (1 - \Phi(-\sqrt{y}; \lambda))^a)^{b-1} \right\} = \\ &= abf_{\chi^2(1)}(y)h(y; \lambda, a, b), \quad y > 0, \end{aligned}$$

with

$$\begin{aligned} h(y; \lambda, a, b) &= \left\{ \Phi(\lambda\sqrt{y}) (\Phi(\sqrt{y}; \lambda))^{a-1} (1 - \Phi(\sqrt{y}; \lambda))^a)^{b-1} + \right. \\ &\quad \left. + \Phi(-\lambda\sqrt{y}) (\Phi(-\sqrt{y}; \lambda))^{a-1} (1 - \Phi(-\sqrt{y}; \lambda))^a)^{b-1} \right\}, \end{aligned}$$

and $f_{\chi^2(1)}(\cdot)$ is the density function of a chi-square density function. We can note that

$$h(y; \lambda, a, b) \xrightarrow{\lambda \rightarrow \infty} (2\Phi(\sqrt{y}) - 1)^{a-1} (1 - (2\Phi(\sqrt{y}) - 1)^a)^{b-1} = \quad (4.23)$$

$$= F_{\chi^2(1)}^{a-1}(y) \left(1 - F_{\chi^2(1)}^a(y) \right)^{b-1}, \quad (4.24)$$

where $F_{\chi^2(1)}(\cdot)$ is the chi-square distribution function. Therefore, the density $f_Y(\cdot)$ converges to the density of a Kumaraswamy chi-square random variable with parameters 1, a and b as $\lambda \rightarrow \infty$. \square

4.2.1 An interesting theorem

Theorem 34. *Let $X \sim KwsN(\lambda, a, b)$ and $\xi = (\lambda, a, b)$.*

The distribution of X reduces to a normal distribution if and only if one of the following conditions holds:

1. $\xi = (0, 1, 1)$;
2. $\xi = (1, \frac{1}{2}, 1)$;
3. $\xi = (-1, 1, \frac{1}{2})$.

Proof. It easy to see that if one of the conditions from 1 to 3 is verified then X is a normal random variable, as mentioned in the properties from c to e of the Kumaraswamy skew-normal in the first section of this chapter.

Conversely, since $g_{\Phi(x;\lambda)}^K(x; \lambda, a, b)$ is a normal density for fixed λ , a and b it follows that $g_{\Phi(x;\lambda)}^K(x; \lambda, a, b) = \phi(x)$, for all x . This implies that

$$2ab\Phi(\lambda x) (\Phi(x; \lambda))^{a-1} (1 - \Phi(x; \lambda)^a)^{b-1} = 1, \quad \text{for all } x. \quad (4.25)$$

We can without loss of generality take $x = 0$

$$ab(\Phi(0; \lambda))^{a-1} (1 - \Phi(0; \lambda)^a)^{b-1} = 1. \quad (4.26)$$

Now we impose that the distribution function of the Kumaraswamy skew-normal is equal to the distribution function of the normal distribution:

$$1 - [1 - (\Phi(x; \lambda)^a)]^b = \Phi(x), \quad \text{for all } x. \quad (4.27)$$

For the special case $x = 0$, it becomes

$$[1 - (\Phi(0; \lambda)^a)]^b = \frac{1}{2}. \quad (4.28)$$

The first derivative of $g_{\Phi(x; \lambda)}^K(x; \lambda, a, b)$ with respect to x is

$$\frac{\partial g_{\Phi(x; \lambda)}^K(x; \lambda, a, b)}{\partial x} = g_{\Phi(x; \lambda)}^K(x; \lambda, a, b) \left\{ -x + \lambda \frac{\phi(\lambda x)}{\Phi(\lambda x)} + (a-1) \frac{\phi(x; \lambda)}{\Phi(x; \lambda)} - (b-1)a \frac{\phi(x; \lambda) \Phi(x; \lambda)^{a-1}}{1 - \Phi(x; \lambda)^a} \right\},$$

now we impose that

$$\frac{\frac{\partial g_{\Phi(x; \lambda)}^K(x; \lambda, a, b)}{\partial x}}{g_{\Phi(x; \lambda)}^K(x; \lambda, a, b)} = -x, \quad (4.29)$$

and we obtain the following condition:

$$\lambda \frac{\phi(\lambda x)}{\Phi(\lambda x)} + (a-1) \frac{\phi(x; \lambda)}{\Phi(x; \lambda)} - a(b-1) \frac{\phi(x; \lambda) \Phi(x; \lambda)^{a-1}}{1 - \Phi(x; \lambda)^a} = 0, \quad (4.30)$$

which holds for all x . In particular, for $x = 0$ we obtain

$$2\lambda + \frac{a-1}{\Phi(0; \lambda)} - \frac{a(b-1)\Phi(0; \lambda)^{a-1}}{1 - \Phi(0; \lambda)^a} = 0. \quad (4.31)$$

Let us denote by y the distribution function of the skew-normal distribution evaluated at $x = 0$, i.e. $\Phi(0; \lambda)$ (see property 4 in section 1.1). Therefore equations (4.26), (4.28) and (4.31) can be rewritten as follows:

$$\begin{cases} aby^{a-1} (1-y^a)^{b-1} = 1; \\ (1-y^a)^b = \frac{1}{2}; \\ 2 \tan \left(\pi \left(\frac{1}{2} - y \right) \right) + \frac{a-1}{y} - \frac{a(b-1)y^{a-1}}{1-y^a} = 0. \end{cases} \quad (4.32)$$

From the second equation of (4.32) we get

$$y = \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right)^{\frac{1}{a}}. \quad (4.33)$$

Replacing expression (4.33) in the first equation of (4.32) it follows that

$$ab \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right)^{\frac{a-1}{a}} \left(\frac{1}{2} \right)^{\frac{b-1}{b}} = 1,$$

which, after some straightforward algebraical manipulations, one can write as

$$\frac{1}{a} \log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) e^{\frac{1}{a} \log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right)} = b \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \left(\frac{1}{2} \right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right).$$

The function

$$b \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \left(\frac{1}{2} \right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \quad (4.34)$$

is defined only for positive values of b . The figure 4.2 shows that the above function is always negative. Numerically, we have noticed that this function assumes values on the interval $[-\frac{1}{e}, 0)$ when b belongs to the interval $(0, 1.103724877]$. Consequently, two explicit expressions for a in terms of b , using the Lambert W function (see appendix A), are obtained:

$$a = \frac{\log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right)}{W_0 \left(b \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \left(\frac{1}{2} \right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \right)}, \quad (4.35)$$

$$a = \frac{\log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right)}{W_{-1} \left(b \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \left(\frac{1}{2} \right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \right)}. \quad (4.36)$$

Numerically, we have noted that the functions

$W_i \left(b \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \left(\frac{1}{2} \right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \right)$, for $i = -1, 0$, are negative real numbers when b lies in the interval $(0, 1.103724877]$ (see figures 4.3 and 4.4).

Now we divide the proof into two steps.

The first step consists in replacing (4.35) in the third equation of (4.32). The resulting equation, which depends only on the variable b , has two real zeros: $b = 1$ and $b = \frac{1}{2}$ (see figure 4.5).

Substituting $b = 1$ in the second equation of (4.32), we get

$$y = \left(\frac{1}{2} \right)^{1/a}, \quad (4.37)$$

which replaced in the first equation of (4.32) gives

$$ay^{a-1} = 1. \quad (4.38)$$

The solutions of the above equation, obtained using the Lambert W function, are $a = -\frac{\ln(2)}{W_0\left(\frac{\ln(2)}{2}\right)} = 1$ and $a = -\frac{\ln(2)}{W_{-1}\left(-\frac{\ln(2)}{2}\right)} = \frac{1}{2}$.

Taking successively $a = 1$ and $a = \frac{1}{2}$ on (4.37) we obtain that $\lambda = 0$ and $\lambda = 1$, respectively. Replacing $b = \frac{1}{2}$ in the second equation of (4.32) it follows that

$$y = \left(\frac{3}{4}\right)^{\frac{1}{a}}, \quad (4.39)$$

and consequently, the first equation of (4.32) becomes

$$a \left(\frac{3}{4}\right)^{1-\frac{1}{a}} = 1, \quad (4.40)$$

which has two solutions: $a = \frac{\ln\left(\frac{3}{4}\right)}{W_0\left(\frac{3}{4}\ln\left(\frac{3}{4}\right)\right)} = 1$ and $a = \frac{\ln\left(\frac{3}{4}\right)}{W_{-1}\left(\frac{3}{4}\ln\left(\frac{3}{4}\right)\right)} \simeq 0.12$. Setting $a = 1$ in (4.39) we obtain

$$y = \frac{3}{4}, \quad (4.41)$$

which implies $\lambda = -1$.

Replacing $a \simeq 0.12$ in (4.39) we get the value $\lambda \simeq 3.403728$.

We note that the last values ($\lambda \simeq 3.403728$, $a = 0.12$ and $b = \frac{1}{2}$) satisfy the system but the correspondent density function $g_{\Phi(x;\lambda)}^K(x; \lambda, a, b)$ is not a normal density.

The second step consists in replacing (4.36) in the third equation of (4.32).

The equation obtained, which depends only on the variable b , has a real zero: $b = 1$ (see figure 4.6). Consequently, as before we get the same solutions $a = 1$ and $\lambda = 0$, and $a = \frac{1}{2}$ and $\lambda = 1$.

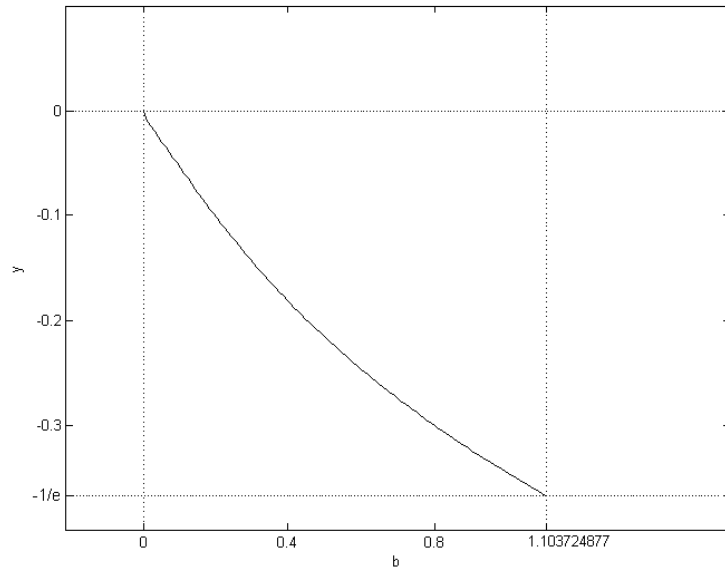


Figure 4.2: The function $b \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right) \left(\frac{1}{2}\right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right)$

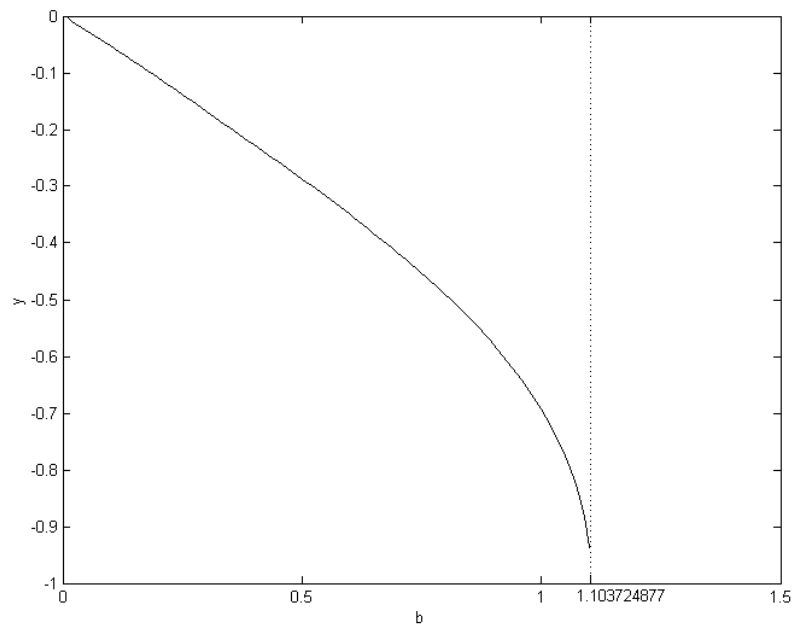


Figure 4.3: The function $W_0 \left(b \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right) \left(\frac{1}{2}\right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{b}}\right) \right)$

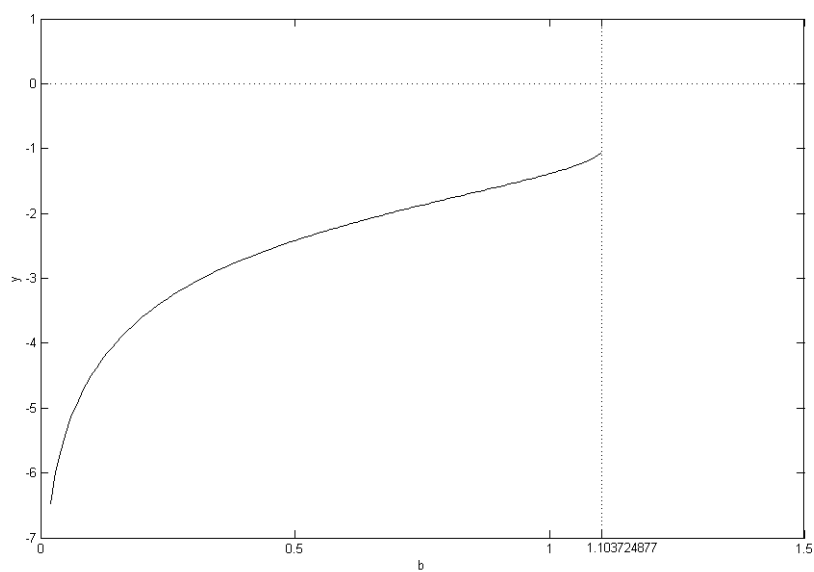


Figure 4.4: The function $W_{-1} \left(b \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \left(\frac{1}{2} \right)^{\frac{b-1}{b}} \log \left(1 - \left(\frac{1}{2} \right)^{\frac{1}{b}} \right) \right)$

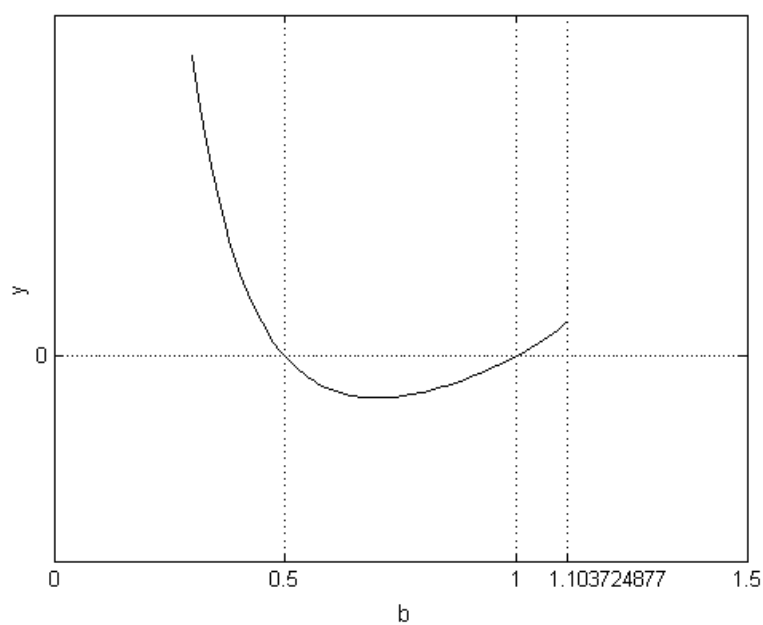


Figure 4.5: The univariate function of b obtained replacing (4.35)

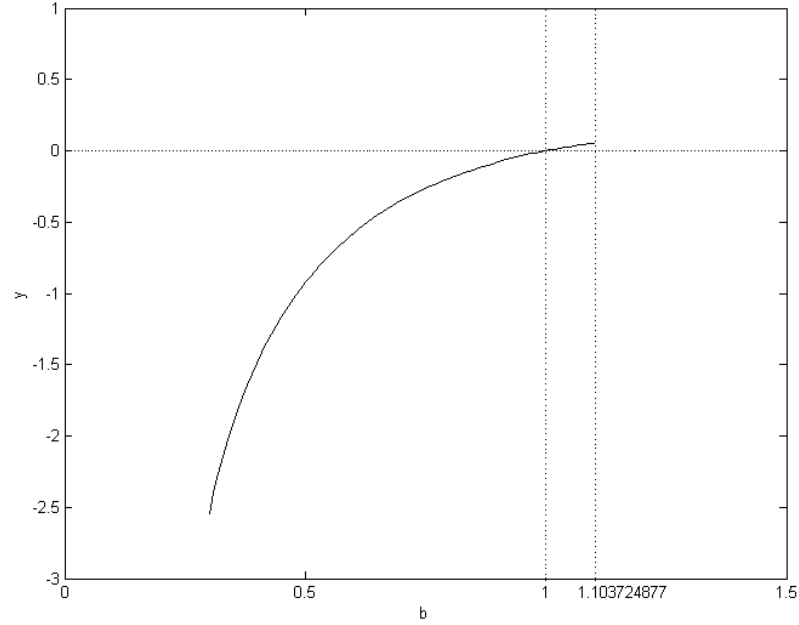


Figure 4.6: The univariate function of b obtained replacing (4.36)

□

4.3 Maximum likelihood estimation

Differentiating (4.2) with respect to the five parameters a , b , λ , μ and σ , a set of five equations is obtained which has to be solved using a numerical root finding algorithm in order to obtain the maximum likelihood estimates of the model parameters. The log-likelihood function $l(\boldsymbol{\xi})$ for the vector of parameters $\boldsymbol{\xi} = (\mu, \sigma, \lambda, a, b)$ can be written as

$$\begin{aligned}
 l(\boldsymbol{\xi}) = & N \log(2) - N \log(\sigma) + N \log(a) + N \log(b) + \sum_{i=1}^N \log \left(\phi \left(\frac{x_i - \mu}{\sigma} \right) \right) + \\
 & + \sum_{i=1}^N \log \left(\Phi \left(\lambda \left(\frac{x_i - \mu}{\sigma} \right) \right) \right) + (a-1) \sum_{i=1}^N \log \left(\Phi \left(\frac{x_i - \mu}{\sigma}; \lambda \right) \right) + \\
 & + (b-1) \sum_{i=1}^N \log \left(1 - \Phi \left(\frac{x_i - \mu}{\sigma}; \lambda \right)^a \right).
 \end{aligned}$$

The components of the score vector $U(\boldsymbol{\xi})$ are given by

$$\begin{aligned}
U_a(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial a} = \frac{N}{a} + \sum_{i=1}^N \log(v_i) - (b-1) \sum_{i=1}^N \frac{v_i^a \log(v_i)}{1-v_i^a}; \\
U_b(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial b} = \frac{N}{b} + \sum_{i=1}^N \log(1-v_i^a); \\
U_\lambda(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial \lambda} = \sum_{i=1}^N z_i y_i + (a-1) \sum_{i=1}^N \frac{\frac{\partial v_i}{\partial \lambda}}{v_i} - (b-1)a \sum_{i=1}^N v_i^{a-1} \frac{\frac{\partial v_i}{\partial \lambda}}{1-v_i^a}; \\
U_\mu(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial \mu} = \frac{1}{\sigma} \sum_{i=1}^N z_i - \frac{\lambda}{\sigma} \sum_{i=1}^N y_i - \frac{(a-1)}{\sigma} \sum_{i=1}^N w_i + \frac{(b-1)a}{\sigma} \sum_{i=1}^N v_i^{a-1} t_i; \\
U_\sigma(\boldsymbol{\xi}) &= \frac{\partial l(\boldsymbol{\xi})}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^N z_i^2 - \frac{\lambda \sum_{i=1}^N z_i y_i}{\sigma} - \frac{(a-1)}{\sigma} \sum_{i=1}^N z_i w_i + \frac{(b-1)a}{\sigma} \sum_{i=1}^N z_i v_i^{a-1} t_i;
\end{aligned}$$

where

$$\begin{aligned}
z_i &= \frac{x_i - \mu}{\sigma}; & v_i &= \Phi\left(\frac{x_i - \mu}{\sigma}; \lambda\right); & y_i &= \frac{\phi\left(\lambda\left(\frac{x_i - \mu}{\sigma}\right)\right)}{\Phi\left(\lambda\left(\frac{x_i - \mu}{\sigma}\right)\right)}; \\
w_i &= \frac{\phi\left(\frac{x_i - \mu}{\sigma}; \lambda\right)}{\Phi\left(\frac{x_i - \mu}{\sigma}; \lambda\right)}; & t_i &= \frac{\phi\left(\frac{x_i - \mu}{\sigma}; \lambda\right)}{1 - \Phi\left(\frac{x_i - \mu}{\sigma}; \lambda\right)^a}.
\end{aligned}$$

The elements of the observed information matrix for the parameters a , b , λ , μ and σ are

$$\begin{aligned}
U_{aa}(\boldsymbol{\xi}) &= -\frac{N}{a^2} - (b-1) \sum_{i=1}^N \log(v_i) \left\{ \frac{v_i^a \log(v_i)}{(1-v_i^a)} + \frac{v_i^{2a} \log(v_i)}{(1-v_i^a)^2} \right\}; \\
U_{bb}(\boldsymbol{\xi}) &= -\frac{N}{b^2}; \\
U_{ab}(\boldsymbol{\xi}) &= -\frac{\sum_{i=1}^N v_i^a \log(v_i)}{1-v_i^a}; \\
U_{a\mu}(\boldsymbol{\xi}) &= -\frac{1}{\sigma} \sum_{i=1}^N w_i + \frac{(b-1)}{\sigma} \sum_{i=1}^N v_i^{a-1} (a \log(v_i) + 1) t_i + \frac{a v_i^{2a-1} \log(v_i) t_i}{1-v_i^a}; \\
U_{b\mu}(\boldsymbol{\xi}) &= \frac{a}{\sigma} \sum_{i=1}^N v_i^{a-1} t_i;
\end{aligned}$$

$$\begin{aligned}
U_{a\sigma}(\boldsymbol{\xi}) &= -\frac{1}{\sigma} \sum_{i=1}^N z_i w_i + \frac{(b-1)}{\sigma} \sum_{i=1}^N \left\{ \frac{(a z_i v_i^{a-1} \log(v_i) t_i + v_i^{a-1} z_i t_i) (1 - v_i^a)}{1 - v_i^a} + \right. \\
&\quad \left. + a \frac{v_i^{2a-1} \log(v_i) z_i t_i}{1 - v_i^a} \right\}; \\
U_{b\sigma}(\boldsymbol{\xi}) &= \frac{a}{\sigma} \sum_{i=1}^N v_i^{a-1} z_i t_i; \\
U_{a\lambda}(\boldsymbol{\xi}) &= \sum_{i=1}^N \frac{\partial v_i}{\partial \lambda} - (b-1) \sum_{i=1}^N \frac{v_i^{a-1} \partial v_i}{1 - v_i^a} - a(b-1) \sum_{i=1}^N \frac{\partial v_i}{\partial \lambda} \left\{ \frac{\log(v_i) v_i^{a-1}}{1 - v_i^a} + \right. \\
&\quad \left. + \frac{\log(v_i) v_i^{2a-1}}{(1 - v_i^a)^2} \right\}; \\
U_{b\lambda}(\boldsymbol{\xi}) &= -a \sum_{i=1}^N \frac{v_i^{a-1} \partial v_i}{1 - v_i^a}; \\
U_{\mu\mu}(\boldsymbol{\xi}) &= -\frac{N}{\sigma^2} - \frac{\lambda}{\sigma} \sum_{i=1}^N \frac{\partial y_i}{\partial \mu} - \frac{(a-1)}{\sigma} \sum_{i=1}^N \frac{\partial w_i}{\partial \mu} + \\
&\quad + \frac{a(b-1)}{\sigma} \sum_{i=1}^N \left\{ -\frac{(a-1)}{\sigma} v_i^{a-1} w_i t_i + v_i^{a-1} \frac{\partial t_i}{\partial \mu} \right\}; \\
U_{\mu\lambda}(\boldsymbol{\xi}) &= -\sum_{i=1}^N \frac{y_i}{\sigma} - \frac{\lambda}{\sigma} \sum_{i=1}^N \frac{\partial y_i}{\partial \lambda} - \frac{(a-1)}{\sigma} \sum_{i=1}^N \frac{\partial w_i}{\partial \lambda} + \\
&\quad + \frac{a(b-1)}{\sigma} \sum_{i=1}^N \left\{ (a-1) v_i^{a-2} \left(\frac{\partial v_i}{\partial \lambda} \right) t_i + v_i^{a-1} \frac{\partial t_i}{\partial \lambda} \right\}; \\
U_{\sigma\sigma}(\boldsymbol{\xi}) &= \frac{N}{\sigma^2} - \frac{3}{\sigma^2} \sum_{i=1}^N z_i^2 + \frac{\lambda}{\sigma^2} \sum_{i=1}^N z_i y_i - \frac{\lambda}{\sigma} \sum_{i=1}^N \left\{ z_i \frac{\partial y_i}{\partial \sigma} - \frac{1}{\sigma} y_i z_i \right\} + \\
&\quad + \frac{(a-1)}{\sigma^2} \sum_{i=1}^N z_i w_i - \frac{(a-1)}{\sigma} \sum_{i=1}^N \left\{ z_i \frac{\partial w_i}{\partial \sigma} - \frac{1}{\sigma} w_i z_i \right\} - \frac{a(b-1)}{\sigma^2} \sum_{i=1}^N v_i^{a-1} z_i t_i + \\
&\quad + \frac{a(b-1)}{\sigma} \sum_{i=1}^N \left\{ \frac{\partial t_i}{\partial \sigma} z_i v_i^{a-1} - \frac{1}{\sigma} z_i t_i v_i^{a-1} - \frac{(a-1)}{\sigma} v_i^{a-1} w_i z_i^2 t_i \right\}; \\
U_{\sigma\mu}(\boldsymbol{\xi}) &= -\frac{2}{\sigma^2} \sum_{i=1}^N z_i - \frac{\lambda}{\sigma} \sum_{i=1}^N \left\{ z_i \frac{\partial y_i}{\partial \mu} - \frac{y_i}{\sigma} \right\} - \frac{(a-1)}{\sigma} \sum_{i=1}^N \left\{ z_i \frac{\partial w_i}{\partial \mu} - \frac{w_i}{\sigma} \right\} + \\
&\quad + \frac{a(b-1)}{\sigma} \left\{ v_i^{a-1} z_i \frac{\partial t_i}{\partial \mu} - \frac{(a-1)}{\sigma} v_i^{a-1} w_i z_i t_i - \frac{1}{\sigma} v_i^{a-1} t_i \right\};
\end{aligned}$$

$$\begin{aligned}
U_{\sigma\lambda}(\boldsymbol{\xi}) &= -\frac{1}{\sigma} \sum_{i=1}^N z_i y_i - \frac{\lambda}{\sigma} \sum_{i=1}^N z_i \frac{\partial y_i}{\partial \lambda} - \frac{(a-1)}{\sigma} \sum_{i=1}^N z_i \frac{\partial w_i}{\partial \lambda} + \\
&\quad + \frac{a(b-1)}{\sigma} \sum_{i=1}^N \left\{ v_i^{a-1} z_i \frac{\partial t_i}{\partial \lambda} + (a-1) v_i^{a-2} z_i t_i \frac{\partial v_i}{\partial \lambda} \right\}; \\
U_{\lambda\lambda}(\boldsymbol{\xi}) &= \sum_{i=1}^N z_i \frac{\partial y_i}{\partial \lambda} + (a-1) \sum_{i=1}^N \frac{\left(\frac{\partial^2 v_i}{\partial \lambda^2} \right) v_i - \left(\frac{\partial v_i}{\partial \lambda} \right)^2}{v_i^2} + \\
&\quad - a(b-1) \sum_{i=1}^N \frac{(a-1) v_i^{a-2} (1-v_i^a) \left(\frac{\partial v_i}{\partial \lambda} \right)^2 + v_i^{a-1} (1-v_i^a) \left(\frac{\partial^2 v_i}{\partial \lambda^2} \right) + a v_i^{2a-2} \left(\frac{\partial v_i}{\partial \lambda} \right)^2}{(1-v_i^a)^2}.
\end{aligned}$$

4.4 Copulas

The Kumaraswamy skew-normal distribution can be generalized to the bivariate case using copulas.

Copula functions are a useful tool to construct bivariate distributions as well as multivariate ones. In fact, their importance in Statistics is described in Sklar's theorem [58], which states that any multivariate distribution function can be represented as a copula function of its marginals. Inspired by the work of Gupta and Kundu (2012) [41], who used the Clayton copula [15] to introduce a bivariate power normal distribution, we derive a bivariate Kumaraswamy skew-normal distribution (*BKwSN*) using Frank's copula [27].

4.4.1 Definitions and basic properties

In this section we refer to [12], [13] and [50] for notations and background on copulas.

First we remind the definition of the copula function.

Definition 9. *A two-dimensional copula C is a real function defined on $[0, 1] \times [0, 1]$ with range $[0, 1]$. Furthermore, for every element (u, v) in the*

domain,

$$C(u, 0) = C(0, v) = 0, \quad C(u, 1) = u, \quad C(1, v) = v.$$

For every rectangle $[u_1, u_2] \times [v_1, v_2]$ in the domain such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

In other words, a bivariate copula is a bivariate distribution function with univariate margins. Therefore, properties of copulas are analogous to properties of joint distributions.

The following theorem is due to Sklar and describes the relationship between copula and joint distribution function.

Theorem 35. *Let $F(x, y)$ be a joint cumulative distribution function with marginal cumulative distributions $F_1(x)$ and $F_2(y)$. There exists a copula C such that, for all real (x, y) ,*

$$F(x, y) = C(F_1(x), F_2(y)). \quad (4.42)$$

If both F_1 and F_2 are continuous, then the copula is unique; otherwise is uniquely determined on $\text{range}(F_1) \times \text{range}(F_2)$. Conversely, if C is a copula and F_1 and F_2 are cumulative distribution functions, then $F(x, y)$, as defined above, is a joint cumulative distribution function with margins F_1 and F_2 .

This theorem allows to construct a bivariate distribution function having the desired marginal distributions and a given copula.

Let us consider the Frank copula

$$C(u, v; \alpha) = -\frac{1}{\alpha} \ln \left\{ 1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1} \right\}, \quad \text{where } \alpha \in \mathbb{R} \setminus \{0\}, \quad 0 < u, v < 1. \quad (4.43)$$

The following result is consequence of (4.43).

Theorem 36. *If (U, V) has the joint cdf (4.43), then*

1. U, V are uniform random variable in the unit interval.
2. The joint pdf of (U, V) , for $0 \leq u, v \leq 1$, is

$$f_{U,V}(u, v; \alpha) = -\frac{\alpha e^{-\alpha(u+v)}}{(e^{-\alpha} - 1) \left(1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1}\right)^2}. \quad (4.44)$$

3. The joint survival function of (U, V) , for $0 \leq u, v \leq 1$, is

$$S(u, v; \alpha) = 1 - u - v - \frac{1}{\alpha} \ln \left\{ 1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1} \right\}. \quad (4.45)$$

4. The conditional cdf of U given $V = v$ is

$$P(U \leq u | V = v) = \frac{e^{-\alpha v}(e^{-\alpha u} - 1)}{(e^{-\alpha} - 1) \left\{ 1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1} \right\}}. \quad (4.46)$$

Now we remind some dependence measures related to copulas: the Kendall's τ and the Spearman's ρ indexes and the coefficients of upper and lower tail dependence.

Definition 10. *If (X_1, X_2) forms a continuous, 2-dimensional random variable with copula C , then*

$$\text{Kendall's } \tau = 4 \int_0^1 \int_0^1 C(u, v) d(C(u, v)) - 1, \quad (4.47)$$

$$\text{Spearman's } \rho = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3, \quad (4.48)$$

where $u = F_1^{-1}(x_1)$ and $v = F_2^{-1}(x_2)$.

Definition 11. *Let (X_1, X_2) be a vector of continuous random variables with marginals distribution function $F_1(\cdot)$ and $F_2(\cdot)$, respectively. Let $u = F_1(x_1)$*

and $v = F_2(x_2)$. The coefficient of upper tail dependence of (X_1, X_2) is defined as

$$\lim_{u \rightarrow 1^-} P(X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)) = \lambda_U, \quad (4.49)$$

provided that the limit $\lambda_U \in [0, 1]$ exists. If $\lambda_U \in (0, 1]$, X_1 and X_2 are said to be asymptotically dependent in the upper tail; if $\lambda_U = 0$, X_1 and X_2 are said to be asymptotically independent in the upper tail.

In the same way, the coefficient of lower tail dependence of (X_1, X_2) is defined as

$$\lim_{u \rightarrow 0^+} P(X_2 < F_2^{-1}(u) | X_1 < F_1^{-1}(u)) = \lambda_L, \quad (4.50)$$

provided that the limit $\lambda_L \in [0, 1]$ exists. If $\lambda_L \in (0, 1]$, X_1 and X_2 are said to be asymptotically dependent in the lower tail; if $\lambda_L = 0$, X_1 and X_2 are said to be asymptotically independent in the lower tail.

The coefficients of upper and lower tail dependence can be expressed in terms of the copula C between X_1 and X_2 as follows:

Definition 12. Let (X_1, X_2) be a continuous random vector with copula C . Then the coefficients of upper and lower tail dependence are given by the following expressions:

$$\lambda_U = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}, \quad (4.51)$$

$$\lambda_L = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}, \quad (4.52)$$

where $u = F_1^{-1}(x_1)$.

All these measures are completely determined by the copula C .

4.4.2 The bivariate Kumaraswamy skew-normal

Using the Frank copula we define a bivariate Kumaraswamy skew-normal so that the marginals are univariate Kumaraswamy skew-normal distributions.

Let (U, V) be random vector with copula C given in formula (4.43).

Consider the following random variables

$$X_1 = \Phi^{-1} \left(\left[(1 - (1 - U)^{\frac{1}{b}})^{\frac{1}{a}} \right]; \lambda \right), \quad X_2 = \Phi^{-1} \left(\left[(1 - (1 - V)^{\frac{1}{d}})^{\frac{1}{c}} \right]; \lambda \right). \quad (4.53)$$

Then the joint cdf of X_1 and X_2 becomes:

$$\begin{aligned} F_{X_1, X_2}(x_1, x_2) &= P(X_1 \leq x_1, X_2 \leq x_2) = \\ &= P \left(U \leq G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b), V \leq G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d) \right) = \\ &= C(G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b), G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d)) = \\ &= -\frac{1}{\alpha} \ln \left\{ 1 + \frac{\left[e^{-\alpha G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b)} - 1 \right] \left[e^{-\alpha G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d)} - 1 \right]}{e^{-\alpha} - 1} \right\}. \end{aligned} \quad (4.54)$$

We give the following definition:

Definition 13. A random vector (X_1, X_2) with joint distribution function (4.54) is said to have a bivariate Kumaraswamy skew-normal distribution derived from the Frank copula.

Remark 20. It should point out that for any copula C it is possible to define a bivariate Kumaraswamy skew-normal. For this reason, we will denote our bivariate Kumaraswamy skew-normal by $BKwSN(\alpha, \lambda, a, b, c, d, Fr)$, where Fr indicates the copula which has been used.

Using theorem 36 we get the following result.

Corollary 17. *If (X_1, X_2) follows the BKwSN($\alpha, \lambda, a, b, Fr$) distribution, then*

1. $X_1 \sim KwSN(\lambda, a, b)$ and $X_2 \sim KwSN(\lambda, c, d)$.

2. The joint pdf of (X_1, X_2) is

$$f_{X_1, X_2}(x_1, x_2) = \frac{\alpha g_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b) g_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d) e^{-\alpha [G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b) + G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d)]}}{(e^{-\alpha} - 1) \left[1 + \frac{\left[e^{-\alpha G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b)} - 1 \right] \left[e^{-\alpha G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d)} - 1 \right]}{e^{-\alpha} - 1} \right]^2}.$$

3. The joint survival function of (X_1, X_2) is

$$S_{X_1, X_2}(x_1, x_2) = 1 - G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b) - G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d) + \frac{1}{\alpha} \ln \left\{ 1 + \frac{\left[e^{-\alpha G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b)} - 1 \right] \left[e^{-\alpha G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d)} - 1 \right]}{e^{-\alpha} - 1} \right\}.$$

4. The conditional cdf of X_1 given $X_2 = x_2$ is

$$P(X_1 \leq x_1 | X_2 = x_2) = \frac{e^{-\alpha G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d)} \left(e^{-\alpha G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b)} - 1 \right)}{1 + \frac{\left[e^{-\alpha G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b)} - 1 \right] \left[e^{-\alpha G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d)} - 1 \right]}{e^{-\alpha} - 1}}.$$

Proof. The proof of point 1 follows directly from elementary probability theory. In fact, from equation (4.53) it follows that $U = G_{\Phi(x_1; \lambda)}(X_1; \lambda, a, b) = F_1(X_1)$ and $V = G_{\Phi(x_2; \lambda)}(X_2; \lambda, c, d) = F_2(X_2)$, so the variables $X_1 = F_1^{-1}(U)$ and $X_2 = F_2^{-1}(V)$ are distributed according to F_i , for $i = 1, 2$.

To show point 2 we use the following result:

$$f_{X_1, X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1, X_2}(x_1, x_2)}{\partial x_1 \partial x_2} = \frac{\partial u}{\partial x_1} \frac{\partial v}{\partial x_2} \frac{\partial^2 C(u, v)}{\partial u \partial v} = \frac{\partial u}{\partial x_1} \frac{\partial v}{\partial x_2} f_{U, V}(u, v), \quad (4.55)$$

and $u = F_1(x_1)$, $v = F_2(x_2)$ and equation (4.44).

Point 3 follows from

$$\begin{aligned} S_{X_1, X_2}(x_1, x_2) &= P(X_1 \geq x_1, X_2 \geq x_2) = \\ &= P\left(U \geq G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b), V \geq G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d)\right) = \\ &= S_{U, V}(G_{\Phi(x_1; \lambda)}^K(x_1; \lambda, a, b), G_{\Phi(x_2; \lambda)}^K(x_2; \lambda, c, d)), \end{aligned}$$

and relation (4.45).

The statement of point 4 is established using the following relation

$$P(X_1 \leq x_1 | X_2 = x_2) = P(U \leq G_{\Phi(x_1; \lambda)}^K | V = G_{\Phi(x_2; \lambda)}^K),$$

and equation (4.46). □

The cdf and pdf of the maximum and minimum of the bivariate Kumaraswamy skew-normal distribution are given in the following theorem.

Theorem 37. *If $(X_1, X_2) \sim BKwSN(\lambda, a, b, c, d, Fr)$, then*

1. *the cdf and pdf of $\max(X_1, X_2)$ are*

$$\begin{aligned} F_{\max(X_1, X_2)}(x) &= -\frac{1}{\alpha} \ln \left\{ 1 + \frac{\left[e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, a, b)} - 1 \right] \left[e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, c, d)} - 1 \right]}{e^{-\alpha} - 1} \right\}, \\ f_{\max(X_1, X_2)}(x) &= \frac{g_{\Phi(x; \lambda)}^K(x; \lambda, a, b) e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, a, b)} \left(e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, c, d)} - 1 \right)}{(e^{-\alpha} - 1) \left\{ 1 + \frac{\left(e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, a, b)} - 1 \right) \left(e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, c, d)} - 1 \right)}{e^{-\alpha} - 1} \right\}^2} + \\ &+ \frac{g_{\Phi(x; \lambda)}^K(x; \lambda, c, d) e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, c, d)} \left(e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, a, b)} - 1 \right)}{(e^{-\alpha} - 1) \left\{ 1 + \frac{\left(e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, a, b)} - 1 \right) \left(e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, c, d)} - 1 \right)}{e^{-\alpha} - 1} \right\}^2}, \end{aligned}$$

respectively;

2. the cdf and pdf of $\min(X_1, X_2)$ are

$$F_{\min(X_1, X_2)}(x) = G_{\Phi(x; \lambda)}^K(x; \lambda, a, b) + G_{\Phi(x; \lambda)}^K(x; \lambda, c, d) + \frac{1}{\alpha} \ln \left\{ 1 + \frac{\left[e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, a, b)} - 1 \right] \left[e^{-\alpha G_{\Phi(x; \lambda)}^K(x; \lambda, c, d)} - 1 \right]}{e^{-\alpha} - 1} \right\},$$

and

$$f_{\min(X_1, X_2)}(x) = g_{\Phi(x; \lambda)}^K(x; \lambda, a, b) + g_{\Phi(x; \lambda)}^K(x; \lambda, c, d) - f_{\max(X_1, X_2)}(x),$$

respectively.

Proof. The cdf of the maximum between X_1 and X_2 can be easily obtained using

$$\begin{aligned} F_{\max(X_1, X_2)}(x) &= P(X_1 \leq x, X_2 \leq x) = F_{X_1, X_2}(x, x) = C(F_1(x), F_2(x)) = \\ &= C\left(G_{\Phi(x; \lambda)}^K(x; \lambda, a, b), G_{\Phi(x; \lambda)}^K(x; \lambda, c, d)\right). \end{aligned} \quad (4.56)$$

While the distribution function of the minimum between X_1 and X_2 can be found by noting that

$$\begin{aligned} F_{\min(X_1, X_2)}(x) &= 1 - P(X_1 \geq x, X_2 \geq x) = 1 - S_{X_1, X_2}(F_1(x), F_2(x)) = \\ &= 1 - S_{U, V}\left(G_{\Phi(x; \lambda)}^K(x; \lambda, a, b), G_{\Phi(x; \lambda)}^K(x; \lambda, c, d)\right). \end{aligned} \quad (4.57)$$

□

For the bivariate Kumaraswamy skew-normal distribution we get the following results:

Property 14. Let $(X_1, X_2) \sim BKwSN(\alpha, \lambda, a, b, c, d, Fr)$, then Kendall's τ index is given by $\frac{4}{\alpha}(1 - D_1(-\alpha)) - 1$, where the function D_1 is defined as

$$D_1(\alpha) = \frac{1}{\alpha} \int_0^\alpha \frac{t}{e^t - 1} dt \quad (4.58)$$

and is called the first Debye function (see for example [1]).

Proof. If (U, V) is a random vector with Frank's copula then Kendall's τ index is exactly $\frac{4}{\alpha}(1 - D_1(-\alpha)) - 1$ (see for instance [13]). Since Kendall's τ index is independent of the margins the result follows. \square

Property 15. *Let $(X_1, X_2) \sim BKwSN(\alpha, \lambda, a, b, c, d, Fr)$, then Spearman's ρ index is given by $\frac{12}{\alpha}[D_2(-\alpha) - D_1(-\alpha)] - 1$, where D_1 is the first Debye function and the function D_2 is defined as*

$$D_2(\alpha) = \frac{2}{\alpha^2} \int_0^\alpha \frac{t^2}{e^t - 1} dt \quad (4.59)$$

and is known as the second Debye function (see for example [1]).

Proof. Spearman's ρ index of a random vector (U, V) with Frank's copula is $\frac{12}{\alpha}[D_2(-\alpha) - D_1(-\alpha)] - 1$ (see for instance [13]). As the Kendall index, Spearman's ρ one is completely determined by the copula. \square

Property 16. *Let $(X_1, X_2) \sim BKwSN(\alpha, \lambda, a, b, c, d, Fr)$, then the coefficients of upper and lower tail dependence of (X_1, X_2) are null, i.e. X_1 and X_2 are asymptotically independent in both upper and lower tails.*

Proof. The Frank copula has neither lower nor upper tail dependency. These two measures, as the previous ones, do not depend on the marginal probability distributions. \square

In figure 4.7 and 4.8 are given the surface plots of the joint pdf of (X_1, X_2) for different values of the parameters.

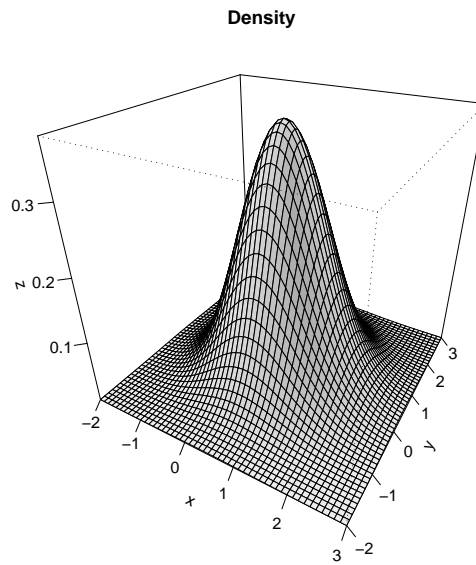


Figure 4.7: The $BKwSN(1, 1, 1, 1, 1, 1)$ density function

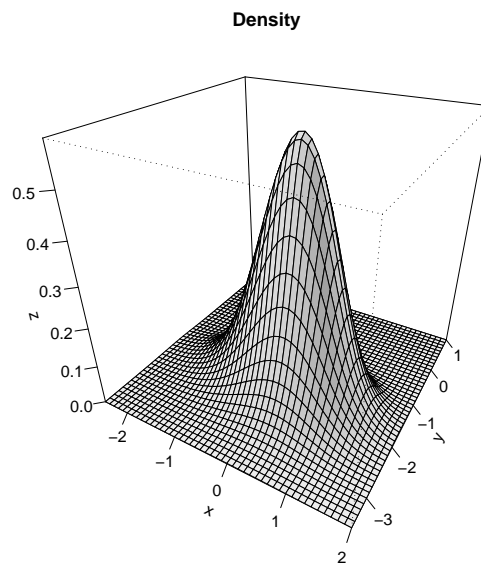


Figure 4.8: The $BKwSN(-1, 1, 2, 4, 1, 2)$ density function

4.5 The generalized Beta skew-normal distribution

We present a new family of distributions that contains the *KwSN* and the *BSN* as special cases. This new distribution is obtained following the procedure described in the last section of chapter 1.

For the sake of completeness, we briefly discuss results concerning this new family which are generalizations of the ones given for the *KwSN* and the *BSN*. We now introduce the four-parameter generalized Beta skew-normal density with parameters $\lambda \in \mathbb{R}$, $a > 0$, $b > 0$ and $c > 0$, say $GBSN(\lambda, a, b, c)$, by taking $F(x)$ in (1.43) to be the cdf of the skew-normal. The *GBSN* density function can be expressed as

$$g_{\Phi(x;\lambda)}^{\mathcal{G}\mathcal{B}\mathcal{G}}(x; \lambda, a, b, c) = \frac{c}{B(a, b)} \phi(x; \lambda) (\Phi(x; \lambda))^{ac-1} (1 - \Phi(x; \lambda))^c)^{b-1}. \quad (4.60)$$

The corresponding cumulative distribution function is given by

$$G_{\Phi(x;\lambda)}^{\mathcal{G}\mathcal{B}\mathcal{G}}(x; \lambda, a, b, c) = I_{\Phi(x;\lambda)^c}(a, b). \quad (4.61)$$

A random variable X having density (4.60) will be indicated by $X \sim GBSN(\lambda, a, b, c)$.

Location and scale parameters may naturally be introduced by setting $Y = \sigma X + \mu$, where $\mu \in \mathbb{R}$ and $\sigma > 0$.

Thus, we denote the random variable Y by $Y \sim GBSN(\mu, \sigma, \lambda, a, b, c)$.

We begin by collecting together some easy results.

Properties of $GBSN(\lambda, a, b, c)$:

- a. $g_{\Phi(x;\lambda)}^{\mathcal{G}\mathcal{B}\mathcal{G}}(x; \lambda, 1, 1, 1) = \phi(x; \lambda)$, for all $x \in \mathbb{R}$, i.e. $GBSN(\lambda, 1, 1, 1) = SN(\lambda)$.

- b. $g_{\Phi(x;\lambda)}^{\mathcal{GBG}}(x; \lambda, a, b, 1) = g_{\Phi(x;\lambda)}^B(x; \lambda, a, b)$, for all $x \in \mathbb{R}$, i.e.
 $GBSN(\lambda, a, b, 1) = BSN(\lambda, a, b)$.
- c. $g_{\Phi(x;\lambda)}^{\mathcal{GBG}}(x; \lambda, 1, b, c) = g_{\Phi(x;\lambda)}^K(x; \lambda, c, b)$, for all $x \in \mathbb{R}$, i.e.
 $GBSN(\lambda, 1, b, c) = KwSN(\lambda, c, b)$.
- d. $g_{\Phi(x;0)}^K(x; 0, a, b, c) = g_{\Phi(x)}^K(x; a, b, c)$, for all $x \in \mathbb{R}$, i.e.
 $GBSN(0, a, b, c) = GBN(a, b, c)$.
- e. $g_{\Phi(x;0)}^K(x; 0, 1, 1, 1) = \phi(x)$, for all $x \in \mathbb{R}$, i.e.
 $GBSN(0, 1, 1, 1) = N(0, 1)$.
- f. $g_{\Phi(x;1)}^K(x; 1, \frac{1}{2}, 1, 1) = \phi(x)$, for all $x \in \mathbb{R}$, i.e.
 $GBSN(1, \frac{1}{2}, 1, 1) = N(0, 1)$.
- g. $g_{\Phi(x;-1)}^K(x; -1, 1, \frac{1}{2}, 1) = \phi(x)$, for all $x \in \mathbb{R}$, i.e.
 $GBSN(-1, 1, \frac{1}{2}, 1) = N(0, 1)$.
- h. If $X \sim GBSN(\lambda, a, b, c)$, then $Y = \Phi(X; \lambda)$ is a $GB(a, b, c)$.
- i. If $X \sim GBSN(\lambda, a, b, c)$, then $Y = \Phi(X; \lambda)^c$ is a $Beta(a, b)$.
- j. If $X \sim GBSN(\lambda, a, b, c)$, then $Y = 1 - \Phi(X; \lambda)^c$ is a $Beta(b, a)$.
- k. As $\lambda \rightarrow +\infty$, $g_{\Phi(x;\lambda)}^K(x; \lambda, a, b, c)$ tends to the generalized Beta half-normal density.

Graphical illustrations of (4.60) are shown in figure 4.9 and 4.10.

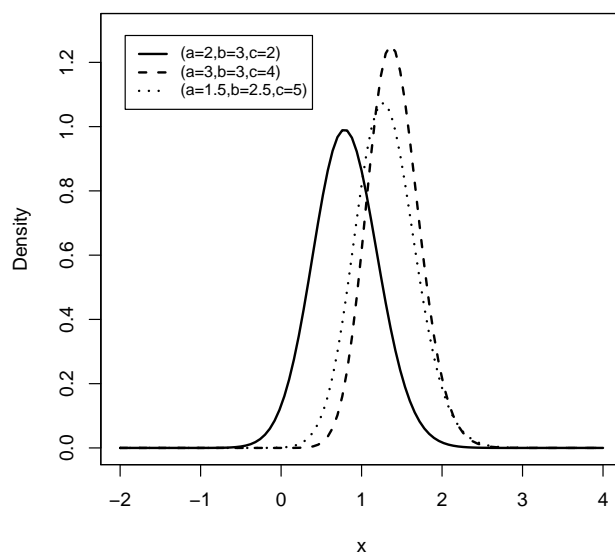


Figure 4.9: The *GBSN* density for different values of a , b , c and $\lambda = 1$

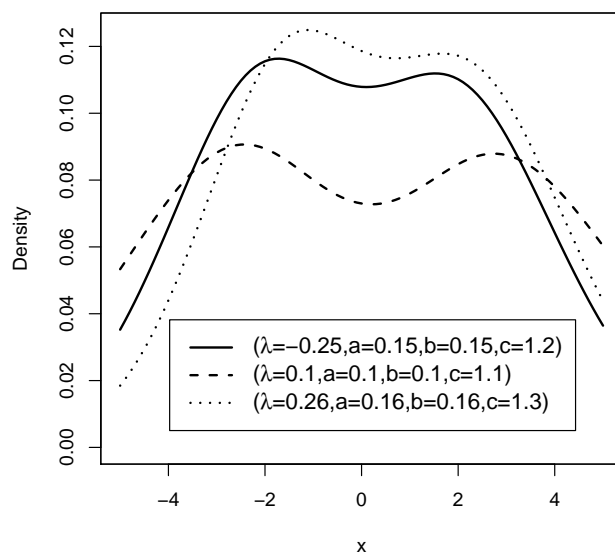


Figure 4.10: The *GBSN* density for different values of c , λ , $a < 1$ and $b < 1$

The following result is a generalization of theorems 17 and 33.

Theorem 38. *If $X \sim GBSN(\lambda, a, b, c)$, then $X^2 \xrightarrow{L} GB\chi^2(1, a, b, c)$ as $\lambda \rightarrow \infty$, where $GB\chi^2(1, a, b, c)$ is a generalized Beta chi-square random variable with parameters 1, a , b and c .*

Proof. Let $Y = X^2$. We can easily check that the density of Y is

$$\begin{aligned} f_Y(y) &= \frac{c}{B(a, b)} \phi(\sqrt{y}) \frac{1}{\sqrt{y}} \left\{ \Phi(\lambda\sqrt{y}) (\Phi(\sqrt{y}; \lambda))^{ac-1} (1 - \Phi(\sqrt{y}; \lambda))^c)^{b-1} + \right. \\ &\quad \left. + \Phi(-\lambda\sqrt{y}) (\Phi(-\sqrt{y}; \lambda))^{ac-1} (1 - \Phi(-\sqrt{y}; \lambda))^c)^{b-1} \right\} = \\ &= \frac{c}{B(a, b)} f_{\chi^2(1)}(y) h(y; \lambda, a, b, c), \quad y > 0, \end{aligned}$$

where $f_{\chi^2(1)}(\cdot)$ is the density function of a chi-square density function. We note that

$$h(y; \lambda, a, b, c) \xrightarrow{\lambda \rightarrow \infty} (2\Phi(\sqrt{y}) - 1)^{ac-1} (1 - (2\Phi(\sqrt{y}) - 1)^c)^{b-1} = \quad (4.62)$$

$$= F_{\chi^2(1)}^{ac-1}(y) \left(1 - \left(F_{\chi^2(1)}(y) \right)^c \right)^{b-1}, \quad (4.63)$$

where $F_{\chi^2(1)}(\cdot)$ is the chi-square distribution function. The density $f_Y(\cdot)$ converges to the density of a generalized Beta chi-square with parameters 1, a , b and c as $\lambda \rightarrow \infty$. \square

4.5.1 Moment generating function and moments

Let us find the moment generating function of $GBSN(\lambda, a, b, c)$.

Property 17. *The moment generating function of $X \sim GBSN(\lambda, a, b, c)$ is given by*

$$M_X(t) = \frac{2c}{B(a, b)} e^{\frac{t^2}{2}} E_Z \left((\Phi(Z; \lambda))^{ac-1} (1 - \Phi(Z; \lambda))^c)^{b-1} \Phi(\lambda Z) \right), \quad (4.64)$$

where $Z \sim N(t, 1)$.

A recursive formula for the k -th moment is obtained using integration by parts.

Property 18. Let $k \in \mathbb{N}$ and $k \geq 1$. If $X \sim \text{GBSN}(\lambda, a, b, c)$, with $a > \frac{1}{c}$ and $b > 1$, then

$$E_X(X^k) = (k-1)E_X(X^{k-2}) + \lambda E_X \left(X^{k-1} \frac{\phi(\lambda X)}{\Phi(\lambda X)} \right) + \\ + \frac{(ac-1)B(a-\frac{1}{c})}{B(a,b)} E_U \left(U^{k-1} \phi(U; \lambda) \right) - \frac{c(b-1)B(a+1-\frac{1}{c})}{B(a,b)} E_V \left(V^{k-1} \phi(V; \lambda) \right),$$

where $U \sim \text{GBSN}(\lambda, a - \frac{1}{c}, b, c)$ and $V \sim \text{GBSN}(\lambda, a + 1 - \frac{1}{c}, b - 1, c)$ are independent random variables.

4.5.2 New properties of the \mathcal{GBG} distribution

We now turn our attention to the $\mathcal{GBG} - F$ model and we extend some of the results, given in section 3.1.4 for the *Beta* - F and in section 4.1.3 for the *Kw* - F distributions.

Definition (1.43) immediately leads to the following theorems, whose proofs are similar to that of theorem (20) and are therefore omitted.

Theorem 39. Let $X \sim \mathcal{GBG} - F(a, b, n)$ independent of a random sample Y_1, Y_2, \dots, Y_n from F . Then $X | (Y_{(n)} \geq X) \sim \mathcal{GBG} - F(a, b + 1, n)$.

The following theorem is a generalization of the above one.

Theorem 40. Let $X \sim \mathcal{GBG} - F(a, b, c)$ independent of a random variable $Y \sim \text{Beta} - F(c, 1)$. Then $X | (Y \geq X) \sim \mathcal{GBG} - F(a, b + 1, c)$.

Theorem 41. Let $X \sim \text{Kw} - F(a, b)$ independent of a random sample Y_1, Y_2, \dots, Y_n from F . Then $X | (Y_{(n)} \leq X) \sim \mathcal{GBG} - F(\frac{a+n}{a}, b, a)$.

The above result has been improved as follows.

Theorem 42. *Let $X \sim Kw - F(a, b)$ independent of $Y \sim Beta - F(c, 1)$. Then $X|(Y \leq X) \sim \mathcal{GBG} - F\left(\frac{a+c}{a}, b, a\right)$.*

Theorem 43. *Let $X \sim \mathcal{GBG} - F(a, b, c)$ independent of a random sample Y_1, Y_2, \dots, Y_n from $Beta - F(c, 1)$. Then $X|(Y_{(1)} \geq X) \sim \mathcal{GBG} - F(a, b+n, c)$.*

Theorem 44. *Let $X \sim \mathcal{GBG} - F(a, b, c)$ independent of a random variable $Y \sim Kw - F(a, c)$. Then $X|(Y \geq X) \sim \mathcal{GBG} - F(a, b+a, c)$.*

Theorem (44) includes theorem (43) as a special case.

Bibliography

- [1] Abramowitz, M., Stegun, I. A., (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications.
- [2] Abtahi, A., Towhidi, M., Behboodian, J. (2011). An appropriate empirical version of skew-normal density. *Stat. Papers*, **52**, 469–489.
- [3] Alexander, C., Cordeiro, G. M., Ortega, E. M. M., Sarabia, J. M., (2011). Generalized Beta-Generated Distributions. *Computational Statistics & Data Analysis*. In press.
- [4] Arnold, B. C., Balakrishnan, N., Nagaraya, H. N. (1992). *A first course in Order statistics*. New York: John Wiley & Sons.
- [5] Arnold, B. C., Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, **11(1)**, 7–54.
- [6] Azzalini, A. (1985). A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*, **12(2)**, 171–178.
- [7] Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal Royal Statist. Soc., series B*, **61**, 579–602.

-
- [8] Azzalini, A. (2010). R package 'sn': The skew-normal and skew-t distributions. URL <http://azzalini.stat.unipd.it/SN>.
- [9] Bahrami, W., Agahi, H., Rangin, H. (2009). A Two-parameter Balakrishnan Skew-normal Distribution. *J. Statist. Res. Iran*, **6**, 231–242.
- [10] Balakrishnan, N. (2002). Discussion of “Skewed multivariate models related to hidden truncation and/or selective reporting”. *Test*, **11**, 37–39.
- [11] Branco, M. D., Dey, D. K. (2001). A general class of skew-elliptical distributions. *Journal Multivariate Analysis*, **79**, 99–113.
- [12] Brigo, D., Mercurio, F. (2006). *Interest rate models: theory and practice: with smile, inflation, and credit*. Heidelberg: Springer Verlag.
- [13] Cherubini, U., Luciano, E., Vecchiato, W. (2004). *Copula Methods in Finance*. Chichester, England: John Wiley & Sons, Ltd.
- [14] Chiogna, M. (1998). Some results on the scalar skew-normal distribution. *Journal of the Italian Statistical Society*, **7**, 1–13.
- [15] Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial dependency in chronic disease incidence. *Biometrika*, **65**, 141–151.
- [16] Cooray, K., Ananda, M. M. A. (2008). A generalization of the half-normal distribution with applications to lifetime data. *Communication in Statistics & Theory and Methods*, **37**, 1323–1337.
- [17] Cordeiro, G. M., de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation & Simulation*, **81(7)**, 883–898.

-
- [18] Cordeiro, G. M., Lemonte, A. J. (2011). The β -Birnbaum-Saunders distribution: an improved distribution for fatigue lifemodelling. *Computational Statistics and Data Analysis*, **55(3)**, 1445–1461.
- [19] Cordeiro, G. M., Pescim, R. R., Ortega, E. M. M. (2012). The Generalized Half-Normal Distribution for Skewed Positive Data. *Journal of Data Science*, **10**, 195–224.
- [20] Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational Mathematics*, **5**, 329–359.
- [21] David, H. A., Nagaraya, H. N. (2003). *Order statistics*. Hoboken, NJ: Wiley.
- [22] Efron, B., Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York; London: Chapman and Hall.
- [23] Eugene, N., Lee, C., Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics - Theory and methods*, **31(4)**, 497–512.
- [24] Famoye, F., Lee, C., Eugene, N. (2004). Beta-Normal Distribution: Bimodality Properties and Application. *Journal of Modern Applied Statistical Methods*, **3(1)**, 85–103.
- [25] Fang, K. T., Kotz, S., Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. New York: Chapman and Hall.
- [26] Ferreira, J. T. A. S., Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions, *Journal of the American Statistical Association*, **101(474)** , 823–829.

-
- [27] Frank, M. J. (1979). On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$. *Aequationes Math.*, **19(2-3)**, 194-226.
- [28] Genton, M. G. (2004). *Skew-elliptical distributions and their applications: a journey beyond normality*. Boca Raton, FL: Chapman & Hall.
- [29] Gradshteyn, I. S., Ryzhik, I. M. (2000). *Table of integrals, Series and Products*. San Diego: Academic Press.
- [30] Gumbel, E. (1954). The maxima of the mean largest value and of the range. *Ann. Math. Statist.*, **25**, 76–84.
- [31] Gupta, R. C., Gupta, R. D. (2004). Generalized skew-normal model. *Test*, **13(2)**, 501–524.
- [32] Gupta, A. K., Nadarajah, S. (2005). On the Moments of the Beta-normal distribution. *Communications in Statistics - Theory and Methods*, **33(1)**, 1–13.
- [33] Hartley, H., David, H. (1954). Universal bounds for the mean range and extreme observation. *Ann. Math. Statist.*, **25**, 85–99.
- [34] Jeffrey, D. J., Hare, D. E. G., Corless, R. M. (1998). Exact rational solutions of a transcendental equation. *C.R. Math. Rep. Acad. Sci. Canada*, **20(3)**, 71–76.
- [35] Jones, M. C. (2004). Families of distributions of order statistics. *Test*, **13(1)**, 1–43.
- [36] Jones, M. C. (2009). Kumaraswamy's distribution: a Beta-type distribution with some tractability advantages. *Statistical Methodology*, **6(1)**, 70–81.

-
- [37] Kendall, M., Stuart, A. (1973). *The advanced theory of statistics, Volume 2: Inference and relationship, Third edition*. London: Charles Griffin & Co Ltd .
- [38] Kendall, M., Stuart, A. (1977). *The advanced theory of statistics, Volume 1: Distribution Theory, Fourth edition*. London: Charles Griffin & Co Ltd.
- [39] Kong, L., Lee, C., Sepanski, J. H. (2007). On the Properties of Beta-Gamma Distribution. *Journal of Modern Applied Statistical Methods*, **6(1)**, 187–211.
- [40] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, **46**, 79–88.
- [41] Kundu, D., Gupta, D. R. (2012). Power normal distribution. *Statistics*, to appear.
- [42] Liseo, B., Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference*, **136**, 373–389.
- [43] Loperfido, N. (2002). Statistical implications of selectively reported inferential results. *Statistics & Probability Letters*, **56**, 13–22.
- [44] Loperfido, N. (2008). A note on skew-elliptical distributions and linear functions of order statistics. *Statistics & Probability Letters*, **78**, 3184–3186.
- [45] Mameli, V., Musio, M.. A new generalization of the skew-normal distri-

- bution: the Beta skew-normal. *Communications in Statistics - Theory and Methods*, to appear.
- [46] Mameli, V.. The Kumaraswamy skew-normal distribution. Manuscript.
- [47] Mameli, V.. Further result on the Beta skew-normal distribution. Manuscript.
- [48] Mameli, V., Musio, M., Sauleau, E., Biggeri, A.. Large sample confidence intervals for the skewness parameter of the skew-normal distribution based on Fisher's transformation. *Journal of Applied Statistics*, to appear.
- [49] McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, **52**, 647–664.
- [50] Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer.
- [51] Papadatos, N. (1995). Maximum variance of order statistics. *Ann. Inst. Statist. Math.*, **47(1)**, 185–193.
- [52] Pescim, R. R., Demetrio, C. G. B., Cordeiro, G. M. E., Ortega, M. M., Urbanoa, M. R. (2010). The Beta generalized half-normal distribution. *Computational Statistics and Data Analysis*, **54(4)**, 945–957.
- [53] Roberts, C. (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association*, **61**, 1184–1190.
- [54] Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew-normal and skew-t distributions. *J. Stat. Plan Inference*, **136**, 4259–4275.

-
- [55] Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52** (3-4), 591–611.
- [56] Sharafi, M., Behboodian, J. (2006). A new skew-normal density. *J. Statist. Res. Iran*, **3**, 47–61.
- [57] Sharafi, M., Behboodian, J. (2008). The Balakrishnan skew-normal density. *Statistical Papers*, **49**, 769–778.
- [58] Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite de Paris.*, **8**, 229–231.
- [59] Yadegari, I., Gerami, A., Khaledi, M. J. (2008). A generalization of the Balakrishnan skew-normal distribution. *Statistics and Probability Letters*, **78**, 1165–1167.
- [60] Zografos, K., Balakrishnan, N. (2009). On families of Beta- and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, **6**, 344–362.

Part II

Chapter 5

Scoring rules

The organization of this chapter is as follows. Section 1 describes the basics of convex analysis, as well as notions of sub-gradient vectors and Gateaux differentiability. Furthermore, we review briefly basic notions that we will borrow from the theory of normed spaces. Section 2 is devoted to the most important features of the theory of scoring rules, special attention is given to McCarthy's characterization theorem. In section 3 we explore two theorems, one relative to bounded loss functions, the other one to unbounded loss functions. In section 4 new generalizations of McCarthy's theorem, for unbounded scores and countable infinite sample spaces, are given.

5.1 Basic concepts

The concepts presented in this section are of fundamental importance and all the subsequent material is based on them.

5.1.1 Concave functions

We refer to Rockafellar (1970) [17] for notations and background on convex analysis. We shall permit functions to take values on the extended real line $\overline{\mathbb{R}} = [-\infty, +\infty]$. Since arithmetic calculations involving $+\infty$ and $-\infty$ must be performed, we adopt the following conventions, used in [17]:

$$\begin{aligned} \alpha + \infty &= \infty + \alpha = \infty, & \text{where } -\infty < \alpha \leq \infty; \\ \alpha - \infty &= -\infty + \alpha = -\infty, & \text{where } -\infty \leq \alpha < \infty; \\ \alpha(\infty) &= (\infty)\alpha = \infty, & \alpha(-\infty) = (-\infty)\alpha = -\infty, & \text{where } 0 < \alpha \leq \infty; \\ \alpha(\infty) &= (\infty)\alpha = -\infty, & \alpha(-\infty) = (-\infty)\alpha = \infty, & \text{where } -\infty \leq \alpha < 0; \\ 0(\infty) &= (\infty)0 = \infty, & 0(-\infty) = (-\infty)0 = -\infty, & -(-\infty) = \infty; \\ \inf \emptyset &= \infty & \text{and } \sup \emptyset = -\infty; \\ +\infty - \infty &= \infty. \end{aligned}$$

Let \mathcal{X} be a vector space and $f: \mathcal{X} \rightarrow \overline{\mathbb{R}}$ be a function. The hypograph of a function f , denoted by $\text{hyp}(f)$, consists of all points in $\mathcal{X} \times \mathbb{R}$ that lie below the function, i.e.

$$\text{hyp}(f) := \{(x, c) \in \mathcal{X} \times \mathbb{R} : c \leq f(x)\}. \quad (5.1)$$

Definition 14. A subset C of \mathcal{X} is said to be convex if $(1 - \lambda)x + \lambda y \in C$ whenever $x, y \in C$ and $0 < \lambda < 1$.

Let us recall the definition of concave function.

Definition 15. The function f is called concave if $\text{hyp}(f)$ is a convex set.

A function f is convex if $-f$ is concave.

The effective domain of a convex function $f: \mathcal{X} \rightarrow \overline{\mathbb{R}}$ is the set

$$\text{dom}(f) = \{x \in \mathcal{X} : f(x) < \infty\}. \quad (5.2)$$

Definition 16. A subset A of \mathcal{X} is said to be affine if $(1 - \lambda)x + \lambda y \in A$ for $x, y \in A$ and $\lambda \in \mathbb{R}$.

Definition 17. A function $f : X \rightarrow \mathbb{R}$ on a vector space is affine if it is of the form $f(v) = l(v) + \alpha$ for some linear function $l \in \mathcal{X}^*$ and some real α .

All affine sets are convex. Convex sets have a lot of important theoretical properties.

Definition 18. A convex function f is called closed if its epigraph, i.e.,

$$\text{epi}(f) := \{(x, c) \in \mathcal{X} \times \mathbb{R} : c \geq f(x)\}, \quad (5.3)$$

is a closed convex set.

Let us now focus on a few operations that preserve convexity of functions. Using these rules and several elementary convex functions, it is possible to build more complex convex functions or prove convexity of a given function.

- Let $\lambda > 0$ and f be a closed convex function. Then the function $f_1(x) = \lambda f(x)$ is closed and convex.
- Let f_1 and f_2 be closed and convex functions. Then the function $f(x) = f_1(x) + f_2(x)$ is closed and convex.
- A weighted combination with positive weights of convex functions is convex. If $w_i > 0$ and f_1, \dots, f_n are convex then $\sum_i w_i f_i$ is convex.
- Let the functions $\{f_i(x)\}_{i \in I}$ be closed and convex. Then the function f defined by $f(x) = \sup \{f_i(x), i \in I\}$ is closed and convex, i.e. the point-wise supremum of an arbitrary collection of convex functions is convex.

Definition 19. A convex function $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ is said to be proper if $\text{dom}(f) \neq \emptyset$, and $f(x) > -\infty$, for all $x \in \mathcal{X}$.

Now, an important theorem may be presented (see for instance [17]).

Theorem 45. *A proper closed convex function f is the pointwise supremum of the collection of all affine functions h such that $h \leq f$.*

Definition 20. *Let $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ be a concave proper function. The closure of f is defined as the function $\bar{f} : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ such that $\text{hyp}(\bar{f}) = \overline{\text{hyp}(f)}$.*

Remark 21. *The closure of a concave proper function exists and is unique. The closure \bar{f} of f is closed and concave and is a majorant of f , i.e., $f(x) \leq \bar{f}(x)$, for all $x \in \mathcal{X}$. The function \bar{f} is the smallest closed concave majorant of f : if ϕ is any closed concave majorant of f then ϕ also majorizes \bar{f} .*

We denote by \mathcal{X}^* the space of linear functionals on \mathcal{X} .

Now we introduce the concept of sub-gradient.

Definition 21. *A vector $x^* \in \mathcal{X}^*$ is called a sub-gradient of f at $x \in \mathcal{X}$ if*

$$f(y) \geq f(x) + x^*(y - x), \quad (5.4)$$

for all $y \in \mathcal{X}$.

The set of all sub-gradients of f at x is denoted by $\partial f(x)$. If $\partial f(x)$ is not empty, f is said to be sub-differentiable at x .

Theorem 46. *Let \mathcal{X} be a linear normed space and $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ be a convex function finite and continuous at $x_0 \in \mathcal{X}$. Then $\partial f(x_0) \neq \emptyset$, i.e., f is sub-differentiable at x_0 .*

The following three well known theorems play a key role to prove one of the main results of this chapter.

Theorem 47. *Let \mathcal{X} be a real linear normed space and $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ be a proper functional having at each point of \mathcal{X} a sub-gradient. Then f is convex and weak-lower semi-continuous (over the whole \mathcal{X}).*

In particular, Hendrickson and Buehler [13] have proved that

Theorem 48. *If f has sub-gradient x^* at each point x in a convex set D , then f is convex in D .*

Theorem 49. *If a convex function $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$, with \mathcal{X} a normed space (more general locally convex space), has a neighbourhood of a point $x \in \mathcal{X}$, where f is bounded above by a finite constant, then f is continuous at x .*

Let us remind the definition of Gateaux differentiability.

Definition 22. *The Gateaux directional derivative of the function f at x in the direction of y is defined as*

$$f'_G(x;y) = \lim_{\lambda \rightarrow 0^+} \frac{f(x + \lambda y) - f(x)}{\lambda}, \quad (5.5)$$

if it exists. The function f is Gateaux differentiable at x if it has a Gateaux derivative at x for all y , and $f'_G(x;y)$ is a linear continuous function of y .

We finish this subsection with some useful definitions, referring to [14] for an overview of this topic.

Definition 23. *A subset A of a metric space \mathcal{X} is said to be dense if its closure \overline{A} coincides with \mathcal{X} .*

Definition 24. *A subset A of a metric space \mathcal{X} is said to be nowhere dense if its closure \overline{A} has empty interior.*

Definition 25. *A subset A of a space is said to be of first category in \mathcal{X} if there exist nowhere dense subsets $F_1, F_2, \dots, F_n, \dots$ such that $A = \cup_{i=1}^{\infty} F_i$. Otherwise, A is said to be of second category in \mathcal{X} .*

Definition 26. *A space is said to be separable if it contains a countable dense subset.*

5.1.2 Banach spaces

In this subsection we introduce basic notions and concepts in Banach space theory. Let us first recall the formal definition of a Banach space.

Definition 27. *A Banach space is a normed linear space that is a complete metric space with respect to the metric derived from its norm.*

Definition 28. *A basis (or Schauder basis) for a Banach space \mathcal{X} is a sequence $(e_n : n > 1)$ of members of \mathcal{X} which has the property that, for each x in \mathcal{X} , there is exactly one sequence of scalars (x_i) for which $x = \sum_{i=1}^{\infty} x_i e_i$ in the sense that $\lim_{n \rightarrow \infty} \|x - \sum_{i=1}^n x_i e_i\| = 0$.*

An important class of Banach spaces is given by the spaces l_p for $1 \leq p < \infty$, which is the space of all sequences $x = \{x_i\}_{i=1}^{\infty}$ for which $\sum_{i=1}^{\infty} |x_i|^p$ is convergent and the norm $\|x\|_p$ is defined as $(\sum_{i=1}^{\infty} |x_i|^p)^{\frac{1}{p}}$. Let e_i be an element of l_p that consists of zeros except for 1 in position i . It is easy to show that $(e_i)_{i=1}^{\infty}$ is a basis for l_p . In fact,

$$\|x - \sum_{i=1}^n x_i e_i\|_p = \left(\sum_{i=n+1}^{\infty} |x_i|^p \right)^{\frac{1}{p}}, \quad (5.6)$$

so $\lim_{n \rightarrow \infty} \|x - \sum_{i=1}^n x_i e_i\|_p = 0$.

To end this section we remind the theorem proved by Mazur (1933) [15], which is useful to obtain one of the most important results of this chapter.

Theorem 50. *Let \mathcal{X} be a separable real Banach space. Let f be a real-valued convex continuous function defined on an open convex subset $\Omega \subset \mathcal{X}$. Then there is a subset $A \subset \Omega$ of the first category such that f is Gateaux differentiable on $\Omega \setminus A$.*

All of the theory developed up until this point can be analogously applied to concave functions, with obvious modifications.

5.2 Scoring rules

In this section, we describe briefly relevant concepts from the literature on scoring rules and state some of the notations that will be used throughout the chapter. For notations and background on this subject we refer to [2] and [12].

5.2.1 Decision problems

Consider a statistical decision problem $(\mathcal{X}, \mathcal{A}, L)$, defined in terms of an outcome space \mathcal{X} , an action space \mathcal{A} and a loss function L .

Let the loss function be given by $L: \mathcal{X} \times \mathcal{A} \rightarrow (-\infty, \infty]$. Let \mathcal{P} be a convex class of distributions over \mathcal{X} such that $L(P, a) := E_{X \sim P} L(X, a)$ exists for all $a \in \mathcal{A}$, and $P \in \mathcal{P}$. The combination $\mathcal{G} = (\mathcal{X}, \mathcal{A}, L)$ is called a basic game. Consider a Decision Maker (DM) who has to make a decision whose consequences will depend on the outcome of a random variable X defined on \mathcal{X} . More formally, a DM has to take some actions a selected from a given action space \mathcal{A} , after which Nature will reveal the value $x \in \mathcal{X}$ of a quantity X and DM will then suffer a loss $L(x, a)$ in $(-\infty, \infty]$.

An act $a_P \in \mathcal{A}$ will be optimal if it minimizes $L(P, a)$ over all $a \in \mathcal{A}$. If a such act a_P exists, it will be called a Bayes act against P .

The Bayes loss $H(P) \in [-\infty, \infty]$ of a distribution $P \in \mathcal{P}$ is defined by

$$H(P) := \inf_{a \in \mathcal{A}} L(P, a). \quad (5.7)$$

A scoring rule is a loss function measuring the quality of a quoted probability distribution Q for a random variable X , in the light of the realised outcome

x of X ; specifically, if the forecaster quotes the predictive distribution Q and the event x materializes, the loss suffered is $S(x, Q)$.

The function $S(x, Q)$ takes values in $(-\infty, \infty]$ and the expected value of $S(x, Q)$ under P is denoted by $S(P, Q)$.

Definition 29. *The scoring rule S is proper relative to the class \mathcal{P} if*

$$S(P, Q) \geq S(P, P), \quad \text{for all } P, Q \in \mathcal{P}. \quad (5.8)$$

It is strictly proper relative to \mathcal{P} if $S(P, Q) \geq S(P, P)$ with equality if and only if $Q = P$.

An arbitrary statistical decision problem can be reduced to one based on a proper scoring rule.

Let $L : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$ be a loss function, defined for an outcome space \mathcal{X} , and an action space \mathcal{A} . Let \mathcal{P} be a class of distributions over \mathcal{X} such that $L(P, a) := E_{X \sim P} L(X, a)$ exists for all $a \in \mathcal{A}$, and $P \in \mathcal{P}$, define for $P, Q \in \mathcal{P}$, and $x \in \mathcal{X}$

$$S(x, Q) := L(x, a_Q), \quad (5.9)$$

where $a_P := \arg \inf_{a \in \mathcal{A}} L(P, a)$ is a Bayes act with respect to P . It is not difficult to see that S is a proper scoring rule, and the associated entropy function is the Bayes loss $H(P) = \inf_{a \in \mathcal{A}} L(P, a)$.

If S is proper, the function d on $\mathcal{P} \times \mathcal{P}$ defined as $d(P, Q) := S(P, Q) - S(P, P)$, for $P, Q \in \mathcal{P}$, is called divergence function. This function is non negative, and if S is strictly proper then $d(P, Q)$ is strictly positive unless $P = Q$.

5.2.2 Finite outcomes

In this subsection we restrict our treatment of the score to the case of a finite sample space $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$. Let \mathcal{B} be the set of real vectors

$\boldsymbol{\alpha} = (\alpha_x : x \in \mathcal{X})$, with each $\alpha_x > 0$, and $\mathcal{P} = \{\mathbf{p} \in \mathcal{B} : \sum_x p_x = 1\}$ the set of such vectors corresponding to strictly positive probability distributions on \mathcal{X} . A distribution P over \mathcal{X} can be represented by its probability vector $\mathbf{p} = (p_1, p_2, \dots, p_N)$, so we will use P for the distribution determined by \mathbf{p} (similarly Q for \mathbf{q}), and generally not distinguish between them.

Consider a game between Forecaster and Nature, where Forecaster quotes a distribution $Q \in \mathcal{P}$ as representing his uncertainty about a quantity X taking values in \mathcal{X} , and Nature then reveals $X = x$. For $P \in \mathcal{P}$, let $S(P, Q) := \sum_x p_x S(x, \mathbf{q})$ be the expected score when Forecaster quotes Q , and Nature generates X from P . The generalized entropy function, or uncertainty function, $H : \mathcal{P} \rightarrow \mathbb{R}$, associated with a proper scoring rule S is given by

$$H(\mathbf{p}) := S(\mathbf{p}, \mathbf{p}) = \sum_x p_x S(x, \mathbf{p}), \quad (5.10)$$

and the corresponding divergence function is defined as

$$d(\mathbf{p}, \mathbf{q}) := S(\mathbf{p}, \mathbf{q}) - H(\mathbf{p}) = \sum_x p_x S(x, \mathbf{q}) - \sum_x p_x S(x, \mathbf{p}). \quad (5.11)$$

5.2.3 Examples of proper scoring rules

A wide variety of proper scoring rules has been proposed. Here we present two special examples of particular interest, on which the literature has focused mainly.

- The quadratic score or Brier score is defined by

$$S(i, \mathbf{p}) = \sum_{j=1}^m (\delta_{ij} - p_j)^2 = -2p_i + \sum_{j=1}^m p_j^2 + 1, \quad (5.12)$$

where $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$, otherwise.

Then $S(\mathbf{p}, \mathbf{q}) = \sum_j q_j^2 - 2\sum_j p_j q_j + 1$, which is uniquely minimized for $\mathbf{q} = \mathbf{p}$, so that is a strictly proper scoring rule. The corresponding

entropy function and divergence function are $H(\mathbf{p}) = 1 - \sum_{j=1}^n p_j^2$ and $d(\mathbf{p}, \mathbf{q}) = \sum_j (p_j - q_j)^2$, respectively. This well-known scoring rule was proposed by Brier (1950) [1].

- The logarithmic scoring rule is $S(i, \mathbf{p}) = -\log(p_i)$. Correspondingly, the entropy function is $H(\mathbf{p}) = -\sum_j p_j \log p_j$, called Shannon entropy, and the divergence function, called the Kullback-Leibler divergence, is $d(\mathbf{p}, \mathbf{q}) = \sum_j p_j \log \left(\frac{q_j}{p_j} \right)$. This is one of the most interesting example of unbounded scoring rules and was proposed by Good (1952) [8].

From these strictly proper scoring rules, it is possible to create infinitely many more strictly proper scoring rules by taking positive affine transformation of the said rules.

For the simple case $\mathcal{X} = \{0, 1\}$, the Brier and the Shannon entropies are depicted in figure 5.1 and 5.2, respectively.

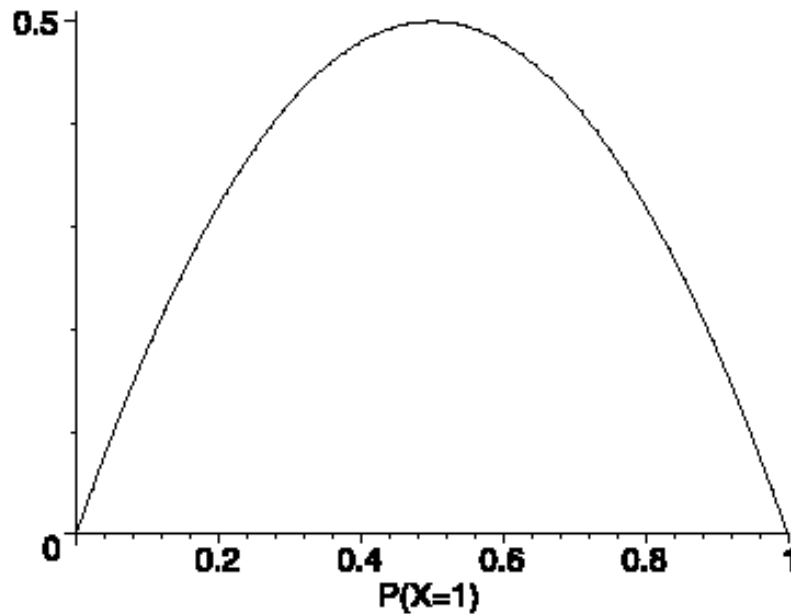


Figure 5.1: The Brier entropy

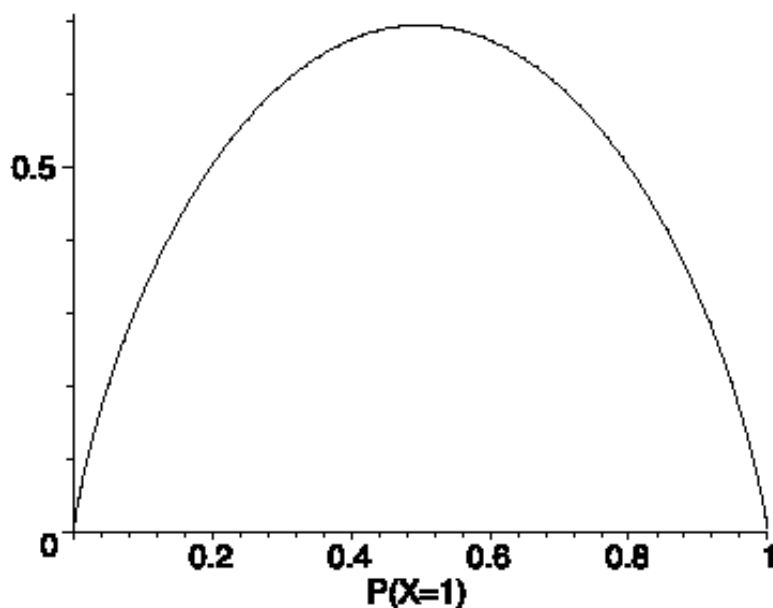


Figure 5.2: The Shannon entropy

5.2.4 Homogeneous scoring rules

In this section we recall the most important features of homogeneous functions which will be used in the characterization theorem for scoring rules, provided by Dawid et al. (2011) [4].

Definition 30. A function $f : \mathcal{A} \rightarrow \mathbb{R}$ is called (positive) homogeneous of order h , or h -homogeneous, if

$$f(\lambda \alpha) = \lambda^h f(\alpha), \quad \text{for all } \lambda > 0. \quad (5.13)$$

If f is differentiable, the above equation will hold if and only if f satisfies Euler's equation:

$$\sum_x \alpha_x \frac{\partial f}{\partial \alpha_x} = hf. \quad (5.14)$$

When f is differentiable at α the super-gradient $\nabla f(\alpha)$ at α coincides with

the gradient vector $\left(\frac{\partial f}{\partial \alpha_x} : x \in \mathcal{X}\right)$. The following lemma extends Euler's equation (5.14) to homogeneous functions with a super-gradient.

Lemma 4. *Suppose f is h -homogeneous, and has a super-gradient $\nabla f(\boldsymbol{\alpha})$ at $\boldsymbol{\alpha}$. Then*

$$\boldsymbol{\alpha}^T \nabla f(\boldsymbol{\alpha}) = hf(\boldsymbol{\alpha}). \quad (5.15)$$

Corollary 18. *Suppose f is 1-homogeneous. Then S is a super-gradient of f at $\boldsymbol{\alpha}$ if and only if*

$$\boldsymbol{\beta}^T S \geq f(\boldsymbol{\beta}), \quad (5.16)$$

for all $\boldsymbol{\beta} \in \mathcal{A}$, with equality when $\boldsymbol{\beta} = \boldsymbol{\alpha}$.

McCarthy (1956) [16] states that a scoring rule S is proper if and only if it can be expressed as the super-gradient of a concave function which is homogeneous of degree 1. This theorem was proved by Hendrickson and Buehler (1971) [13].

In order to conclude this section, we state McCarthy's theorem in terms of homogeneous functions as given in Dawid et al. (2011) [4].

Theorem 51. *Suppose $H : \mathcal{A} \rightarrow \mathbb{R}$ is concave and 1-homogeneous. Let ∇H be a super-gradient of H , and for $x \in \mathcal{X}$, $\boldsymbol{p} \in \mathcal{P}$, define $S(x, \boldsymbol{p})$ to be the x -component of the vector $S(\boldsymbol{p}) := \nabla H(\boldsymbol{p})$. Then S is a proper scoring rule, and the associated entropy at \boldsymbol{p} is $H(\boldsymbol{p})$.*

Theorem 52. *Suppose that $S(x, \boldsymbol{\alpha})$ is a 0-homogeneous proper scoring rule. Define $H(\boldsymbol{\alpha}) := \boldsymbol{\alpha}^T S(\boldsymbol{\alpha})$. Then H is 1-homogeneous and concave, and $S(\boldsymbol{\alpha})$ is a super-gradient of H at $\boldsymbol{\alpha}$.*

However, this characterization is limited because it deals only with a certain subset of scoring rules. In fact, in the simple form stated here the precedent theorems can not be applied when \mathcal{X} is infinite or S assumes

values not finite. Our purpose is to provide a generalization of these theorems. Broadly, we show that, under appropriate regularity conditions on the scoring rules, the previous theorems continue to hold for infinite countable sample spaces and unbounded scoring rules.

5.3 Conjugacy

In the following we intend to recall the Fenchel conjugate function, our treatment is based on Rockafellar (1970) [17].

In this section we suppose that the space \mathcal{X} is finite.

We indicate by \mathcal{P} the set of all probability distributions over \mathcal{X} , then \mathcal{P} is in one-to-one correspondence with the unit simplex in \mathbb{R}^n , and inherits its algebraic and topological structure.

We denote by \mathcal{L} the set of all bounded functions from \mathcal{X} to \mathbb{R} , by \mathcal{L}^+ the set of all functions from \mathcal{X} to \mathbb{R}^+ . The expected value of $l(\cdot)$ under P is denoted by $l(P)$, i.e., $l(P) = E_{X \sim P}(l(X)) = \sum_{x \in \mathcal{X}} p_x l(x)$. The support of $l \in \mathcal{L}^+$ is the set $S(l) := \{x : l(x) < \infty\}$. By the support of a we shall mean the support of $L(\cdot, a)$, i.e. $S(a) = \{x | L(x, a) < \infty\}$.

Definition 31. *Let $H : \mathcal{P} \rightarrow \overline{\mathbb{R}}$ be concave and closed. The conjugate (or Legendre transform, or Legendre-Fenchel transform or Fenchel conjugate) of H is the function $H^* : \mathcal{L} \rightarrow \overline{\mathbb{R}}$ given by*

$$H^*(l) := \inf_{P \in \mathcal{P}} \{l(P) - H(P)\}. \quad (5.17)$$

Geometrically, $H^*(l)$ is the minimum height, over \mathcal{P} , of the function l above the function H .

The conjugate function H^* is always concave and closed. H^* is proper if and only if H is proper.

In a parallel relationship, H may be defined as

$$H(P) := \inf_{l \in \mathcal{L}} \{l(P) - H^*(l)\}. \quad (5.18)$$

By definition, for any $P \in \mathcal{P}$ and $l \in \mathcal{L}$ we have

$$H(P) + H^*(l) \leq l(P). \quad (5.19)$$

Grünwald and Dawid (2003) [11] introduce the function D^H on $\mathcal{P} \times \mathcal{L}$ by

$$D^H(P, l) := l(P) - H(P) - H^*(l), \quad (5.20)$$

the inequality (5.19) can be rewritten in this way

$$D^H(P, l) \geq 0, \quad (5.21)$$

for all $P \in \mathcal{P}$, and $l \in \mathcal{L}$. $P \in \mathcal{P}$ and $l \in \mathcal{L}$ are called conjugate relative to H if $D^H(P, l) = 0$, and it will be denoted by $P \leftrightarrow_H l$. In this case both $H(P)$ and $H^*(l)$ are finite.

Since the simplex \mathcal{P} is bounded, it follows from Theorem 27.3 in [17] that for all $l \in \mathcal{L}$ there exists a conjugate $P \in \mathcal{P}$. It further follows from Theorem 23.4 in [17] that whenever P lies in the relative interior of the effective domain of H there will exist a conjugate $l \in \mathcal{L}$.

For any $l \in \mathcal{L}$, $c \in \mathbb{R}$, define $l^c \in \mathcal{L}$ by $l^c(x) = l(x) - c$, all $x \in \mathcal{X}$ and $l^H := l^{H^*(l)}$. l^c is called a translate of l . Since $H^*(l^c) = H^*(l) - c$, it follows immediately that $H^*(l^H) = 0$, hence, for any $c \in \mathbb{R}$,

$$D^H(P, l^c) = l^c(P) - H(P) - H^*(l^c) = l(P) - H^*(l) - H(P) = l^H(P) - H(P). \quad (5.22)$$

In particular, $P \leftrightarrow_H l \Leftrightarrow P \leftrightarrow_H l^c \Leftrightarrow P \leftrightarrow_H l^H$.

From (5.21) and (5.22), we see that

$$l^H(P) \geq H(P), \quad (5.23)$$

while $l^H(P) = H(P)$ if and only if $P \leftrightarrow_H l$. Geometrically, the last two equations say that l^H is a super-gradient or (upper) supporting hyperplane, for the concave function H on \mathcal{P} , and any such super-gradient has the form l^H . In general, $D^{H(P,l)}$ can be interpreted geometrically as the vertical distance, at $P \in \mathcal{P}$, of the function H below its unique super-gradient that is a translate of l .

As in Grünwald and Dawid (2004) [12], the generalized entropy function $\mathcal{H}^{\mathcal{G}} : \mathcal{P} \rightarrow \overline{\mathbb{R}}$ associated with the game \mathcal{G} is defined by

$$\mathcal{H}^{\mathcal{G}}(P) := \inf_{a \in \mathcal{A}} L(P, a). \quad (5.24)$$

Since this definition displays $\mathcal{H}^{\mathcal{G}}(P)$ as the infimum of a collection of concave functions on \mathcal{P} , it follows easily that $\mathcal{H}^{\mathcal{G}}(P)$ is itself a concave function on \mathcal{P} . The discrepancy function $D^{\mathcal{G}}$ associated with a decision problem $\mathcal{G} = (\mathcal{X}, \mathcal{A}, L)$ is defined by

$$D^{\mathcal{G}}(P, a) := L(P, a) - \mathcal{H}^{\mathcal{G}}(P). \quad (5.25)$$

Then $D^{\mathcal{G}}(P, a) \geq 0$, with equality if and only if a is a Bayes act against P in \mathcal{G} .

We define $\mathcal{L}^H = \{l \in \mathcal{L} : H^*(l) \geq 0\}$, $\mathcal{L}_i^H = \{l \in \mathcal{L} : H^*(l) > 0\}$, and $\mathcal{L}_b^H = \{l \in \mathcal{L} : H^*(l) = 0\}$.

For any of the games having $\mathcal{A} \subseteq \mathcal{L}^H$, and $\mathcal{L}_b^H \subseteq \overline{\mathcal{A}}$, we have $H_{\mathcal{G}} \equiv H$, and setting $a = l \in \mathcal{L}^H$, $L(P, a) = l(P)$, we get

$$D^{\mathcal{G}}(P, l) = D^H(P, l) + H^*(l), \quad (5.26)$$

$$D^H(P, l) = D^{\mathcal{G}}(P, l^H). \quad (5.27)$$

Then $D^{\mathcal{G}}(P, l) \geq D^H(P, l)$, for all $l \in \mathcal{L}^H$. Further, $D^{\mathcal{G}}(P, l) = D^H(P, l)$, for all $P \in \mathcal{P}$, if and only if l is a Bayes act and $D^{\mathcal{G}}(P, l) = 0$ if and only if l is Bayes against P (in which case also $D^H(P, l) = 0$). In particular, $D^{\mathcal{G}}$ and D^H coincide for the game $\mathcal{G} = \mathcal{G}_b^H$, for which all acts are Bayes.

5.3.1 Finite Loss

The results were obtained in [11], and due to its importance for the thesis we shall present them with its proofs. Here we treat the simple case of finite loss, and we give the generalization subsequently in the next subsection.

Theorem 53. *A function $H : \mathcal{P} \rightarrow \overline{\mathbb{R}}$ is the generalized entropy function $\mathcal{H}^{\mathcal{G}}$ arising from a decision problem \mathcal{G} with everywhere finite loss if and only if H is a closed concave function on \mathcal{P} .*

Proof. First suppose that H is the generalized entropy function $\mathcal{H}^{\mathcal{G}}$ associated with some game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, L)$ having L everywhere finite. If $\mathcal{A} = \emptyset$, then by definition of infimum $\mathcal{H}^{\mathcal{G}} = \infty$, which is the unique improper concave closed function. Otherwise, as a pointwise infimum of a non empty family of upper-bounded closed concave functions on \mathcal{P} , H is itself an upper-bounded closed concave function. Conversely, suppose that H is a closed concave function on \mathcal{P} . From theorem 45, $H(P)$ is the pointwise infimum of the collection \mathcal{L}^H of finite affine functions majorizing H . Let $\mathcal{A} = \mathcal{L}^H$ be the action space, with loss function $L(x, a) := a(x)$. Then $\inf_{a \in \mathcal{A}} L(P, a) \equiv H(P)$, and H is the generalized entropy function $\mathcal{H}^{\mathcal{G}}$ of the statistical decision problem $\mathcal{G} = (\mathcal{X}, \mathcal{A}, L)$, which has everywhere finite loss. \square

5.3.2 Infinite Loss

In this subsection we consider the case when the loss function L is unbounded. For any subset $\mathcal{Y} \subseteq \mathcal{X}$, letting $\mathcal{P}_{\mathcal{Y}}$ the set of $P \in \mathcal{P}$ having support $\mathcal{S}(P) = \mathcal{Y}$. This is a relatively open convex subset of the simplex \mathcal{P} , and its closure $\overline{\mathcal{P}}_{\mathcal{Y}}$ is the set of $P \in \mathcal{P}$, which put probability 1 on \mathcal{Y} , i.e., with support $\mathcal{S}(P) \subseteq \mathcal{Y}$. It is clear that \mathcal{P} is the disjoint union of the collection $\{\mathcal{P}_{\mathcal{Y}} : \mathcal{Y} \subseteq \mathcal{X}\}$. Indeed, $\overline{\mathcal{P}}_{\mathcal{Y}}$ being a face of the simplex \mathcal{P} ,

which is the standard representation of a convex set as the disjoint union of its open faces.

Let $H : \mathcal{P} \rightarrow \overline{\mathbb{R}}$ be concave. For any $\mathcal{Y} \subseteq \mathcal{X}$, let $H_{\mathcal{Y}} : \overline{\mathcal{P}_{\mathcal{Y}}} \rightarrow \overline{\mathbb{R}}$ be the closure of the function H when its full domain is restricted to the face $\overline{\mathcal{P}_{\mathcal{Y}}}$.

Definition 32. A function H is said to be internally closed if, for each $\mathcal{Y} \subseteq \mathcal{X}$:

$$H(P) = H_{\mathcal{Y}}(P), \quad \text{for all } P \in \overline{\mathcal{P}_{\mathcal{Y}}}. \quad (5.28)$$

A closed concave function is internally closed.

Theorem 54. A function $H : \mathcal{P} \rightarrow \overline{\mathbb{R}^+}$ is the generalized entropy function from a decision problem \mathcal{G} with loss in \mathbb{R} if and only if H is concave and internally closed.

Proof. First suppose that H is the generalized entropy function $H^{\mathcal{G}}$ of some game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, L)$. Then H is concave. For any $\mathcal{Y} \subseteq \mathcal{X}$, define $\mathcal{A}(\mathcal{Y}) := \{a \in \mathcal{A} : \mathcal{Y} \subseteq \mathcal{S}(a)\}$, i.e., the set of actions a such that the loss $L(x, a)$ is finite, for $x \in \mathcal{Y}$. Define a function $H^{\mathcal{Y}} : \mathcal{P} \rightarrow \overline{\mathbb{R}}$ by

$$H^{\mathcal{Y}}(P) := \inf_{a \in \mathcal{A}(\mathcal{Y})} L(P, a). \quad (5.29)$$

Then by definition of infimum

$$H(P) \leq H^{\mathcal{Y}}(P). \quad (5.30)$$

If its full domain is restricted to the family $\overline{\mathcal{P}_{\mathcal{Y}}}$ of distribution on \mathcal{Y} , $H^{\mathcal{Y}}$ is the generalized entropy function for the restricted game $\mathcal{G}_{\mathcal{Y}} = (\mathcal{X}, \mathcal{A}(\mathcal{Y}), L)$, which has everywhere finite loss. From theorem 53, it follows that $H^{\mathcal{Y}}$, so restricted, is a closed concave function; and so for (5.30), $H^{\mathcal{Y}}$ majorize H .

Hence, if $H_{\mathcal{Y}}$ denotes the closure of the function H relative to $\overline{\mathcal{P}_{\mathcal{Y}}}$, we have:

$$H(P) \leq H_{\mathcal{Y}}(P) \leq H^{\mathcal{Y}}(P), \quad \text{with } P \in \overline{\mathcal{P}_{\mathcal{Y}}}. \quad (5.31)$$

From inequalities (5.30) and (5.31), we deduce that

$H(P) = H_{\mathcal{Y}}$, for $P \in \mathcal{P}_{\mathcal{Y}}$, and it follows that H is internally closed. Conversely, let $H : \mathcal{P} \rightarrow [-\infty, \infty]$ be an internally closed concave function. For any $\mathcal{Y} \subseteq \mathcal{X}$, define $H_{\mathcal{Y}}$ to be the closure of H when its full domain is restricted to $\overline{\mathcal{P}_{\mathcal{Y}}}$. In particular,

$$H(P) \leq H_{\mathcal{Y}}(P), \quad \text{if } \mathcal{S}(P) \subseteq \mathcal{Y}. \quad (5.32)$$

By definition of function internally closed

$$H(P) \leq H_{\mathcal{Y}}(P), \quad \text{if } \mathcal{S}(P) = \mathcal{Y}. \quad (5.33)$$

From theorem 53, it is possible to find an action space $\mathcal{A}_{\mathcal{Y}}$, and a loss function $L_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{A}_{\mathcal{Y}} \rightarrow (-\infty, \infty)$ such that, for all $P \in \overline{\mathcal{P}_{\mathcal{Y}}}$,

$$H_{\mathcal{Y}}(P) = \inf \{L(P, a) : a \in \mathcal{A}_{\mathcal{Y}}\}. \quad (5.34)$$

Let \mathcal{A} be the disjoint union $\dot{\bigcup} \{\mathcal{A}_{\mathcal{Y}} : \mathcal{Y} \subseteq \mathcal{X}\}$ and consider the decision problem $(\mathcal{X}, \mathcal{A}, L)$ with loss function given by:

$$L(x, a) = \begin{cases} L_{\mathcal{Y}}(x, a) & \text{if } x \in \mathcal{Y}_a \\ \infty & \text{otherwise,} \end{cases} \quad (5.35)$$

where \mathcal{Y}_a is the unique subset of \mathcal{X} such that $a \in \mathcal{A}_{\mathcal{Y}}$. We shall then have $\mathcal{S}(a) = \mathcal{Y}_a$, and for any $P \in \mathcal{P}$, the following equalities hold

$$\begin{aligned} \inf_{a \in \mathcal{A}} L(P, a) &= \inf_{a \in \mathcal{A} : \mathcal{S}(P) \subseteq \mathcal{S}(a)} L(P, a) = \\ &= \min_{\mathcal{Y} \subseteq \mathcal{X} : \mathcal{S}(P) \subseteq \mathcal{Y}} \left(\inf_{a \in \mathcal{A} : \mathcal{S}(a) = \mathcal{Y}} L(P, a) \right) = \\ &= \min_{\mathcal{Y} \subseteq \mathcal{X} : \mathcal{S}(P) \subseteq \mathcal{Y}} H_{\mathcal{Y}}(P). \end{aligned} \quad (5.36)$$

□

5.4 New results

We are now in the right position to proceed to our generalisation of the previous theorems.

5.4.1 Infinite countable sample space

We will now consider the case of an infinite countable sample space, identified with the positive integers. We use e_i to denote the standard basis of l_1 . The space l_1 is a separable Banach space.

We indicate by \mathcal{P} the set of all probability on \mathcal{X} . This set is an affine subset of the space l_1 , the space of all infinite sequences $a = (a_1, a_2, \dots, a_i, \dots)$ such that $\sum_{i=1}^{\infty} |a_i| < \infty$.

Theorem 55. *Let \mathcal{Q} an open convex subset of l_1 contained in \mathcal{P} .*

Let $S(i, \mathbf{p}) : \mathcal{X} \times \mathcal{Q} \rightarrow \mathbb{R}$ be a proper scoring rule, 0-homogeneous with respect to $\mathbf{p} \in \mathcal{Q}$. Suppose that the function $S(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{\infty} p_i S(i, \mathbf{q})$ is a bounded function for all \mathbf{p} and \mathbf{q} . Then $H(\mathbf{p}) := S(\mathbf{p}, \mathbf{p})$ is a 1-homogeneous, concave function and is Gateaux differentiable on $\mathcal{Q} \setminus A$, where A is a subset of first category.

Proof. First, we easily check that $H(\mathbf{p})$ is a 1-homogeneous function, using the 0-homogeneity and the linearity of S . By straightforward computation, we obtain

$$H(\lambda \mathbf{p}) = \lambda \sum_{i=1}^{\infty} p_i S(i, \lambda \mathbf{p}) = \lambda \sum_{i=1}^{\infty} p_i S(i, \mathbf{p}) = \lambda H(\mathbf{p}).$$

By the definition of proper scoring rule, for any $\mathbf{q} \in \mathcal{Q}$, we have that

$$S(\mathbf{p}, \mathbf{q}) \geq S(\mathbf{p}, \mathbf{p}). \quad (5.37)$$

Adding the quantity $-H(\mathbf{q})$ side by side in the above relation we get

$$\sum_{i=1}^{\infty} p_i S(i, \mathbf{q}) - q_i S(i, \mathbf{q}) \geq H(\mathbf{p}) - H(\mathbf{q}), \quad (5.38)$$

or, equivalently,

$$\sum_{i=1}^{\infty} (p_i - q_i) S(i, \mathbf{q}) \geq H(\mathbf{p}) - H(\mathbf{q}). \quad (5.39)$$

The last equation shows that $S(i, \mathbf{q})$ is a super-gradient of H at \mathbf{q} . Then by theorems 48 and 49, H is concave and continuous at \mathbf{p} . Moreover, by Mazur's theorem, there exists a subset of first category $A \subset \mathcal{Q}$ such that H is Gateaux differentiable in $\mathcal{Q} \setminus A$. \square

Another important result is the following.

Theorem 56. *Let \mathcal{Q} an open convex subset of l_1 contained in \mathcal{P} .*

Let $H : \mathcal{Q} \rightarrow \mathbb{R}$ be 1-homogeneous, concave, Gateaux differentiable function, bounded for all $\mathbf{p} \in \mathcal{Q}$. Then $S(i, \mathbf{p}) = H'_G(\mathbf{p}; e_i)$ is a 0-homogeneous proper scoring rule.

Proof. Firstly, we shall show that S is 0-homogeneous. Further computations, involving the 1-homogeneity of H , lead to

$$S(i, \lambda \mathbf{p}) = \lim_{\mu \rightarrow 0^+} \frac{H(\lambda \mathbf{p} + \mu e_i) - H(\lambda \mathbf{p})}{\mu} = \lim_{\mu \rightarrow 0^+} \frac{H(\mathbf{p} + \frac{\mu}{\lambda} e_i) - H(\mathbf{p})}{\frac{\mu}{\lambda}} = S(i, \mathbf{p}).$$

Let us prove that S is a proper scoring rule.

It follows from the concavity of H that

$$H(\lambda(\mathbf{q} - \mathbf{p}) + \mathbf{p}) - H(\mathbf{p}) \geq \lambda(H(\mathbf{q}) - H(\mathbf{p})). \quad (5.40)$$

Diving by $\lambda > 0$, and letting $\lambda \rightarrow 0^+$, we obtain that

$$H'_G(\mathbf{p}; \mathbf{q} - \mathbf{p}) \geq H(\mathbf{q}) - H(\mathbf{p}). \quad (5.41)$$

The 1-homogeneity of H immediately yields

$$H'_G(\mathbf{p}; \mathbf{p}) = \lim_{\lambda \rightarrow 0^+} \frac{H(\mathbf{p} + \lambda \mathbf{p}) - H(\mathbf{p})}{\lambda} = H(\mathbf{p}), \quad (5.42)$$

which together with (5.41) and the linearity of $H'_G(\mathbf{p}; \cdot)$ gives

$$H'_G(\mathbf{p}; \mathbf{q}) \geq H(\mathbf{q}). \quad (5.43)$$

Moreover, by the continuity of $H'_G(\mathbf{p}; \cdot)$

$$\begin{aligned} H'_G(\mathbf{p}; \mathbf{q}) &= H'_G\left(\mathbf{p}; \sum_{i=0}^{\infty} q_i e_i\right) = H'_G\left(\mathbf{p}; \lim_{n \rightarrow \infty} \sum_{i=0}^n q_i e_i\right) = \\ &= \lim_{n \rightarrow \infty} H'_G\left(\mathbf{p}; \sum_{i=0}^n q_i e_i\right) = \sum_{i=0}^{\infty} q_i H'_G(\mathbf{p}; e_i) = \\ &= \sum_{i=0}^{\infty} q_i S(i, \mathbf{p}) = S(\mathbf{p}, \mathbf{q}). \end{aligned} \quad (5.44)$$

Finally, combining (5.42), (5.43) and (5.44) we conclude that S is a proper scoring rule. The proof is complete. \square

We remind that Hendrickson and Buehler (1971) [13] and Gneiting and Raftery (2007) [10] give extensions of McCarthy's theorem in the continuous case.

5.4.2 Finite sample space

Using convex analysis tools, we provide an extension of McCarthy's theorem for unbounded scoring rules.

We denote by \mathcal{P} the set of all probability distributions over a set \mathcal{X} . Throughout this subsection we shall consider statistical decision problems such that: the state space \mathcal{X} has finite cardinality, the action space is the space \mathcal{P} , which is in one-to-one correspondence with the unit simplex in \mathbb{R}^n ,

and the score function S on $\mathcal{X} \times \mathcal{P}$ can take values in $(-\infty, \infty)$ or $(-\infty, \infty]$. We first treat the important special case that the score function $S(x, Q)$ is bounded for each $Q \in \mathcal{P}$.

For any $Q \in \mathcal{P}$, the affine function $P \rightarrow S(P, Q)$ is a bounded closed concave function on \mathcal{P} . Following Grünwald and Dawid (2004) [12], we call

$$\mathcal{H}^{\mathcal{G}}(P) := \inf_{Q \in \mathcal{P}} S(P, Q), \quad (5.45)$$

the generalized entropy associated with the score S ; if S is a proper scoring rule then $\mathcal{H}^{\mathcal{G}}(P) := S(P, P)$.

Theorem 57. *A function $H : \mathcal{P} \rightarrow \overline{\mathbb{R}}$ is the generalized entropy function $\mathcal{H}^{\mathcal{G}}$ arising from a decision problem $\mathcal{G} = (\mathcal{X}, \mathcal{P}, S)$ with everywhere finite proper score if and only if H is a closed concave function on \mathcal{P} .*

Proof. The proof follows along the same lines as that of theorem 53.

It follows directly from theorem 53 that if H is the generalized entropy function of some game $\mathcal{G} = (\mathcal{X}, \mathcal{P}, S)$ then H is an upper-bounded, closed and concave function. Conversely, suppose that H is a closed concave function on \mathcal{P} . Then from theorem 45, we conclude that $H(P)$ is the pointwise infimum of the collection \mathcal{L}^H of all affine functions majorizing H .

Let $\mathcal{P} = \mathcal{L}^H$ be the action space and $S(x, Q) := l(x) - H^*(l)$ be the score function, where l and Q are conjugate (since \mathcal{P} is bounded for all $l \in \mathcal{L}$ there exists a conjugate $P \in \mathcal{P}$ (see section 5.3)). From equation (5.19) we have that $\inf_{Q \in \mathcal{P}} S(P, Q) \geq H(P)$, for $P, Q \in \mathcal{P}$. Since l and Q are conjugate we can rewrite $S(P, Q)$ as $S(P, Q) = l(P) + H(Q) - l(Q)$ and when $Q = P$ we have $S(P, P) = H(P)$. This implies that H is the generalized entropy function $\mathcal{H}^{\mathcal{G}}$ of the statistical decision problem $\mathcal{G} = (\mathcal{X}, \mathcal{P}, S)$, whose score function S is proper and everywhere finite. \square

We use the preceding theorem to obtain the following result.

We shall denote by $\text{Supp}(P)$ the support of P , i.e., $\text{Supp}(P) = \{x \in \mathcal{X} | p(x) > 0\}$, and by $\text{Supp}(S_Q)$ the support of S , i.e., $\text{Supp}(S_Q) = \{x \in \mathcal{X} | S(x, Q) < \infty\}$.

Theorem 58. *A function $H : \mathcal{P} \rightarrow \overline{\mathbb{R}}$ is the generalized entropy function from a decision problem $\mathcal{G} = (\mathcal{X}, \mathcal{P}, S)$ with score S in \mathbb{R} if and only if H is concave and internally closed.*

Proof. The proof proceeds along the same lines as the proof of theorem 54. It follows directly from theorem 54 that if $H : \mathcal{P} \rightarrow \overline{\mathbb{R}}$ is the generalized entropy function from a decision problem $\mathcal{G} = (\mathcal{X}, \mathcal{P}, S)$ then H is concave and internally closed. Conversely, let $H : \mathcal{P} \rightarrow [-\infty, \infty]$ be an internally closed concave function. For any $\mathcal{Y} \subseteq \mathcal{X}$, let $H_{\mathcal{Y}}$ be the closure of H when its full domain is restricted to $\overline{\mathcal{P}_{\mathcal{Y}}}$, i.e.,

$$H(P) \leq H_{\mathcal{Y}}(P), \quad \text{if } \text{Supp}(P) \subseteq \mathcal{Y} \text{ (or equivalently } P \in \overline{\mathcal{P}_{\mathcal{Y}}}). \quad (5.46)$$

Further, if $\text{Supp}(P) = \mathcal{Y}$ then

$$H(P) = H_{\mathcal{Y}}(P). \quad (5.47)$$

According to theorem 57, there exist an action space $\mathcal{Q}_{\mathcal{Y}}$, and a proper score function $S_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{Q}_{\mathcal{Y}} \rightarrow (-\infty, \infty)$ such that, for all $P \in \overline{\mathcal{P}_{\mathcal{Y}}}$,

$$H_{\mathcal{Y}}(P) = \inf_{Q \in \mathcal{Q}_{\mathcal{Y}}} \{S(P, Q)\}. \quad (5.48)$$

Now we shall consider the decision problem $(\mathcal{X}, \mathcal{P}, S)$ where \mathcal{P} is the disjoint union $\dot{\cup} \{\mathcal{Q}_{\mathcal{Y}} : \mathcal{Y} \subseteq \mathcal{X}\}$, and S is given by:

$$S(x, Q) = \begin{cases} S_{\mathcal{Y}}(x, Q) & \text{if } x \in \mathcal{Y}_Q \\ \infty & \text{otherwise,} \end{cases} \quad (5.49)$$

where \mathcal{Y}_Q is the unique subset of \mathcal{X} such that $Q \in \mathcal{Q}_{\mathcal{Y}}$, and $\text{Supp}(S_Q) = \mathcal{Y}_Q$.

Then we have

$$\begin{aligned}
 \inf_{Q \in \mathcal{P}} S(P, Q) &= \inf_{Q \in \mathcal{P}: \text{Supp}(P) \subseteq \text{Supp}(S_Q)} S(P, Q) = \\
 &= \min_{\mathcal{Y} \subseteq \mathcal{X}: \text{Supp}(P) \subseteq \mathcal{Y}} \left\{ \inf_{Q \in \mathcal{P}: \text{Supp}(S_Q) = \mathcal{Y}} S(P, Q) \right\} = \\
 &= \min_{\mathcal{Y} \subseteq \mathcal{X}: \text{Supp}(P) \subseteq \mathcal{Y}} H_{\mathcal{Y}}(P). \tag{5.50}
 \end{aligned}$$

The last term is $H(P)$, and the scoring rule above defined is proper. The proof is complete. \square

Bibliography

- [1] Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, **78**, 1–3.
- [2] Dawid, A. P. (1998). Coherent measures of discrepancy, uncertainty and dependence, with applications to Bayesian predictive experimental design. Technical Report 139, Department of Statistical Science, University College London.
- [3] Dawid, A. P. (2007). The Geometry of Proper Scoring Rules. *Annals of the Institute of Statistical Mathematics*, **59**, 77–93.
- [4] Dawid, A. P., Lauritzen, S., Parry, M. (2011). Proper local scoring rules on discrete sample spaces. Submitted to *The Annals of Statistics*.
- [5] Dawid, A. P., Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. *Annals of Statistics*, **27**, 65–81.
- [6] Eidelman, Y., Milman, V., Tzolomitis, A. (2004). *Functional Analysis: An Introduction. Graduate Studies in Mathematics (66)*. Providence, Rhode Island: American Mathematical Society.
- [7] Fang, F., Stinchcombe, M. B., Andrew B. Whinston, A. B. (2010). Proper scoring rules with arbitrary value functions. *Journal of Mathematical Economics*, **46**, 1200–1210.

-
- [8] Good, I. J. (1952). Rational decisions. *J. Roy. Statist. Soc. Ser. B.*, *14*, 107–114.
- [9] Good, I. J. (1971). Comment on “Measuring information and uncertainty” by Robert J. Buehler. In V.P. Godambe, D. A. Sprott (Eds.) *Foundations of Statistical Inference*, (337–339) Toronto: Holt, Rinehart and Winston.
- [10] Gneiting, T., Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.*, **102(477)**, 359–378.
- [11] Grünwald, P. D., Dawid, A. P. (2003). Generalized entropy functions and Bregman divergence. Manuscript.
- [12] Grünwald, P. D., Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy and robust bayesian decision theory. *Ann. Statist.*, **32(4)**, 1367–1433.
- [13] Hendrickson, A. D., Buehler, R. J. (1971). Proper Scores for Probability Forecasts. *Annals of Mathematical Statistics*, **42**, 1916–1921.
- [14] Joshi, K. D. (1983). *Introduction to general topology*. New Delhi: New Age International.
- [15] Mazur, S. (1933). Über konvexe Menge in lineare normierte Räumen. *Studia Math* , **4**, 70–84.
- [16] McCarthy, J. (1956). Measures of the Value of Information. *Proc. Nat. Acad. Sci.*, **42**, 654–655.
- [17] Rockafellar, R. T. (1970). *Convex Analysis*. Princeton: Princeton University Press.

Chapter 6

Conclusions

In chapter 1 we have presented some results concerning the skew-normal distribution and its generalizations. Furthermore, we have introduced Jones' family of distributions, the generalized Kumaraswamy and the generalized Beta-generated distributions.

In the second chapter we have shown that, in the specific case described, the problem of finding the maximum likelihood estimate of the skewness parameter, which in general is not an easy task, can be solved easily using the Fisher transformation if the parameters of the bivariate normal are supposed to be all unknown and we fix both means and standard deviations equal to their MLEs. It is well known that the Fisher transformation is adequate for $n > 50$ and that this approximation is more accurate, for small n , when ρ is close to zero [37]. This begs the question as to how well the ACI method perform when $|\rho|$ is close to 1 and the sample size n is small or moderate. To investigate this dependence we have conducted a simulation study to compare the ACI with another procedure to construct confidence intervals, the percentile parametric bootstrap method (BCI). Comparison of the performance of the confidence intervals is conducted in terms of their: (1) coverage probability,

(2) length. The simulation study has revealed that ACI performs better in terms of coverage probability. We see that, for the most part, actual coverage levels vary but the ACI coverage is little larger and closer to the nominal coverage. The differences between the two methods are particularly important for small and moderate sample size. For example, for $n = 15$ and $\rho = -0.2$ the 95% confidence intervals cover the true value only the 91% of times if we use the BCI and the 94.2% using the ACI. The two methods are comparable in terms of expected length. Results provided allow us to adopt the ACI procedure to compute confidence intervals for λ . The approximation used is good enough to lead us always to prefer the ACI method. Results are not satisfactory in terms of expected width when n is small and ρ is closed to -1 .

The main advantages of the ACI method are that it is based on a theoretical approximation, and it give rapid solutions, even for very large sample sizes n , whereas the percentile bootstrap can take hours.

Results of both examples are in agreement with the findings of the simulation study.

The approximate method proposed here could also be applied to other types of data, for instance to data coming from double measurements with the same instrument, such as spirometry. Another potential application is epidemiological studies on twins (see Roberts (1966) [53]).

The results presented in this chapter has led to the writing of the article [48]. In chapter 3 we have worked with the class of Beta-generated distributions, introduced recently in the literature.

We have introduced a new class of distributions, referred to as the Beta skew-normal (*BSN*), which extends the skew-normal and the Beta-normal distributions. For special values of the parameters this distribution also in-

cludes the Balakrishnan skew-normal (SNB), the generalized Balakrishnan skew-normal ($GBSN$) and a two-parameter generalization of the Balakrishnan skew-normal ($TBSN$). We have provide a mathematical treatment of the new distribution. We have derived various properties of the BSN , including the moment generating function, recurrence relations for moments and two methods for simulating. Some results presented, for example theorems from 7 to 9, bounds for the moments and for the variance, can be adapted for other distributions belonging to the family of Beta-generated distribution, such as the Beta-normal.

The results obtained in this chapter are presented in the article [45] and in the manuscript [47].

In chapter 4 we have introduced a new distribution which is defined by means of a Kumaraswamy distribution. This new distribution is called Kumaraswamy skew-normal and is an important alternative model to the Beta skew-normal. The $KwSN$ represents a generalization of several distributions previously considered in literature such as the Kumaraswamy-normal, the skew-normal and the normal distributions. Some properties of the proposed distribution are discussed. These properties include explicit expansions for the density and the distribution functions, moment generating functions and relationship with other distributions. The estimation of parameters is approached by the method of maximum likelihood and the elements of the observed information matrix are derived.

The study of the $KwSN$ distribution has led to the writing of the manuscript [46].

However, much more work is in order, related to the investigation of the usefulness of the proposed models (BSN and $KwSN$) to analyse real data.

In the last chapter we have reviewed the theory of convex analysis and the

theory of scoring rules.

A scoring rule is a special kind of loss function that measures the quality of probabilistic forecasts based on the predictive distribution Q and on the event that materializes x : $S(x, Q)$.

Any proper scoring rule S has an associated generalized entropy function H . In 1956 McCarthy ([16] of part 2) characterized scoring rules and their entropy functions when the sample space is finite or the scoring rule takes finite values. McCarthy's theorem states that a scoring rule is proper if and only if it can be expressed as the super-gradient of a concave function.

Subsequently, in 1971 Hendrickson and Buehler ([13] of part 2) proved this theorem and gave a generalization in the continuous case.

In this chapter we use convex analysis tools to generalize McCarthy's characterization. We have given generalizations of McCarthy's theorem for countable infinite sample spaces but with bounded score and for finite sample spaces but with unbounded scoring rules.

Appendix A

The Lambert W function

In this appendix a brief description of the Lambert W function is provided. A detailed definition of W as a complex variable function, as well as some historical background and various applications of it in Mathematics and Physics, can be found in [20] of part 1, to which we refer.

The Lambert W function is defined to be the multivalued inverse of the function $f(x) = xe^x$, i.e. the function satisfying

$$W(x)e^{W(x)} = x. \tag{A.1}$$

This function has two real branches, which are represented in figure A.1. The branch satisfying $W(x) \geq -1$ is denoted by $W_0(x)$ and it is referred to as the principal real branch of the W function. The other one, satisfying $W(x) \leq -1$ is known as the secondary real branch and is denoted by $W_{-1}(x)$.

As we can see from figure A.1, if x is real in the interval $-\frac{1}{e} < x < 0$, there are two real values for $W(x)$: $W_0(x)$ and $W_{-1}(x)$. If $x \geq 0$ there is a single real value for $W(x)$ which belongs to the principal branch. If $x = -\frac{1}{e}$ then there is only one negative real value, $W_0(-\frac{1}{e}) = W_{-1}(-\frac{1}{e}) = -1$. Finally, if $x < -\frac{1}{e}$ then there are no real values.

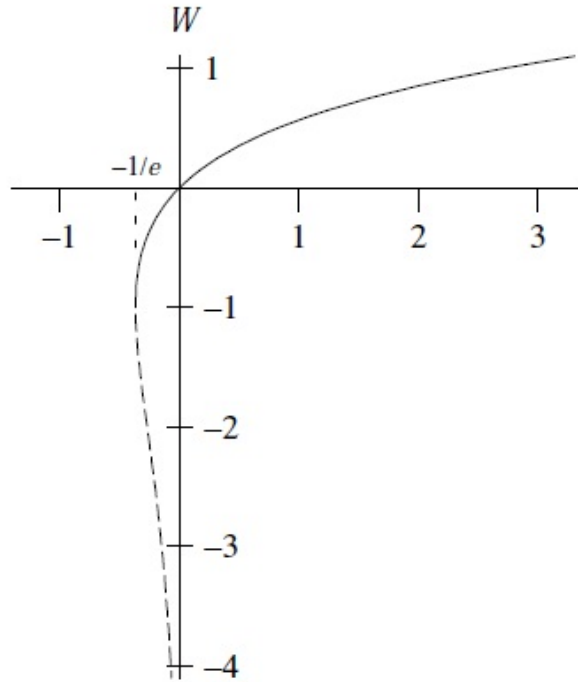


Figure A.1: The solid line shows W_0 and the dashed line W_{-1}

In the following theorem we intend to remind the general solutions of the equations $x^n b^x = c$, which are expressed in terms of the Lambert W function (see [34] of part 1).

Theorem 59. *Let $b, c \in \mathbb{R}$, $b > 1$ and $n \in \mathbb{Z}$. The solutions $x \in \mathbb{R}$ of the equations $x^n b^x = c$ are as follows.*

- If n is odd and $c > -\left(\frac{n}{e \ln(b)}\right)^n$ or if n is even and $c \geq 0$,

$$x = \frac{n}{\ln(b)} W_0 \left(\frac{\ln(b)}{n} c^{\frac{1}{n}} \right).$$

- If n is odd and $0 > c > -\left(\frac{n}{e \ln(b)}\right)^n$,

$$x = \frac{n}{\ln(b)} W_{-1} \left(\frac{\ln(b)}{n} c^{\frac{1}{n}} \right).$$

-
- If n is even and $0 < c < \left(\frac{n}{e \ln(b)}\right)^n$,

$$x = \frac{n}{\ln(b)} W_0\left(-\frac{\ln(b)}{n} c^{\frac{1}{n}}\right) \quad \text{or} \quad x = \frac{n}{\ln(b)} W_{-1}\left(-\frac{\ln(b)}{n} c^{\frac{1}{n}}\right).$$