Jordan Journal of Mathematics and Statistics (JJMS), 16(2), 2023, pp 335 - 360
DOI: https://doi.org/10.47013/16.2.9

AN ANALYTICAL AND NUMERICAL STUDY OF THE LINEAR COMBINATION OF THE DIRICHLET COMPONENTS

M. GHORBEL

Abstract. The aim of the present paper is to derive the exact distribution and the corresponding moment function of the linear combination $S := \alpha S_1 + \beta S_2$, where (S_1, S_2) has the Dirichlet distribution with positive parameters θ_1, θ_2 and θ_3. We also provide the percentile points associated with the linear combination. Next, a measure of entropy of the linear combination is investigated. Further, we give approximations for this distribution and discuss evidence for their robustness. Finally, we study the application of the considered linear combination to precipitation data in France and to the financial returns volatility domain.

1. Introduction

For given random variables S_1 and S_2, the linear combination $S := \alpha S_1 + \beta S_2$ is of interest in problems pertaining to automation, control, fuzzy sets, neurocomputing, and other areas of computer science. In neurocomputing, linear combinations are used to combine multiple probabilistic classifiers on different feature sets. In order to achieve the improved classification performance, a generalized finite mixture model is proposed, as a linear combination scheme, and implemented based on radial basis function networks (see Chen and Chi (1998)). Furthermore, the theory of congruence equations (see for example, Cerruti (1993)) has applications in computer science. There is a wide literature about congruence equations and the last twenty years have witnessed interesting derived formulas and functions; among these, expressions yielding the number of solutions of linear congruences. Counting such solutions is

---

2010 *Mathematics Subject Classification.* 60E05-62E15-62E17.
Key words and phrases. Entropy, Linear combination, Dirichlet distribution, Gauss hypergeometric function.

Received: May 9, 2022 Accepted: Dec. 5, 2022 .

associated with statistical problems, such as the distribution of the values taken by particular sums.

Linear combination of random variables is one of the significant research areas, both from a theoretical and an application points of view. At this stage, we present some applications proposed by authors concerning the linear combination of some very specific distributions:

The weighted sums of uniform random variables have applications in stochastic processes, which, in many cases, can be modeled by these weighted sums. In computer vision algorithms, Kamgar-Parsi et al. (1995) discussed the applications of the weighted sum, and applied the results to a problem in reliability analysis of digital image processing algorithms, and then compared the analytical results with a Monte Carlo simulation approach.

Linear combinations of inverted gamma random variables are applied to the computation of the generalized p-values used for exact significance testing and interval estimation of the parameter of interest in the Behrens–Fisher problem and for variance components in balanced mixed linear model (see Witkovsky (2001)).

Concerning the Beta distributions, their linear combinations are used for detecting changes in the location of the distribution of a sequence of observations in quality control problems (see Lai (1974)). Furthermore, the distribution of the sum of two beta random variables can be used in several important applications, especially in the area of bayesian optimal allocation and reliability (see for example, Pham-Gia and Turkkan (1993)).

Weighted sums of the Poisson parameters are used in medical applications for directly standardized mortality rates. Another example of the need for confidence limits for weighted sums of Poisson parameters occurs in meta-analysis requiring the aggregation of outcome rates from several different studies, using weights related to the numbers of subjects in each study (see Dobson et al. (1991)).

In this paper, we aim at studying the linear combination of the Dirichlet distribution, considered by a large number of authors (see for example Kotz et al. (2000)) and appearing in a set of applications namely order statistics, probabilistic constrained programming models, and delivery problems. Such applications may be found in Barrera et al. (2005), Fabius (1973), Johnson (1960) and Phillips (1988). Furthermore, Provost and Cheong (2000) showed the importance of the distribution of linear combinations of components of a Dirichlet random vector to quadratic forms, and their ratios in statistics which can be applied in a variety of contexts. It remains to point out that the same problem was raised by Gupta and Nadarajah (2006) in the study of the linear combination of the inverted Dirichlet distribution. Here, the authors found the exact expressions for the pdf and moments of this linear combination. Note that the calculation reveals the hypergeometric Gaussian function as the one found for our distribution.

With $\boldsymbol{\theta} :=(\theta_1, \theta_2,\theta_3) > 0$, we shall assume that (S_1, S_2) is distributed according to the joint pdf

$$f(s_1, s_2) = \frac{\Gamma(\theta_1 + \theta_2 + \theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} s_1^{\theta_1 - 1} s_2^{\theta_2 - 1} (1 - s_1 - s_2)^{\theta_3 - 1}, \tag{1.1}$$

where $s_1 > 0$, $s_2 > 0$, and $s_1 + s_2 \leq 1$. This is known as the Dirichlet distribution $D(\boldsymbol{\theta})$. Alternatively, (S_1, S_2) is characterized by its joint moment function

$$\phi(q_1; q_2) := \mathbf{E}(S_1^{q_1} S_2^{q_2}) = \frac{\Gamma(\theta_1 + \theta_2 + \theta_3)}{\Gamma(\theta_1 + \theta_2 + \theta_3 + q_1 + q_2)} \frac{\Gamma(\theta_1 + q_1)\Gamma(\theta_2 + q_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}. \tag{1.2}$$

Note that Eq. (1.1) and Eq. (1.2) are special cases of the Dirichlet distribution treated by Kotz, Balakrishnan and Johnson (2000, pp 487-488).

The organization of the paper is as follows: In Section 2, we derive explicit expressions for the density function of the linear combination $S := \alpha S_1 + \beta S_2$ by using directly the joint density function of (S_1, S_2). The calculations involve the Gauss hypergeometric function defined by

$$ {}_2F_1(a, b; c; x) := \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!}, \quad |x| < 1,$$

where $(i)_k := i(i+1)...(i+k-1)$ denotes the ascending factorial. Next, various mathematical, statistical and reliability properties of the linear combinations of Dirichlet components (like moments, percentiles) are discussed in Sections 3 and 4. In section 5, we investigate a measure of entropy of the linear combination in the general case and we discuss Rényi and Shannon entropies for particular case $\theta_1 + \theta_2 + \theta_3 = 1$. In section 6, we propose with a similar manner to that suggested by Gupta and Nadarajah (2006), approximations for the distribution of S by the classical beta distribution, and discuss the evidence for their robustness. It consists in giving a simple approximation in terms of the beta distribution so that one can use the known procedures for inference, prediction, etc. In section 7, we provide an application of the results derived in this article to precipitation data about rain and snow from France. The data are available at the website https://en.tutiempo.net/climate. Next, we study another application of the considered linear combination in the field of financial returns volatility of two French car manufacturers: Peugeot and Renault. Here, the data set can be downloaded from the website https://finance.yahoo.com. Finally, section 8 offers some conclusions.

For a frequent use of the beta distribution, we recall that: a random variable A with the beta distribution with parameters $a, b > 0$ (say $A \overset{d}{\sim} \beta(a,b)$), has density $f_A\left(x\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}\left(1-x\right)^{b-1}$, $x \in [0,1]$ and moment function $\mathbf{E}\left(A^q\right) = \frac{\Gamma(a+q)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+q)}$, $q > -a$.

2. Density function of S

In this section, we treat linear combination $S = \alpha S_1 + \beta S_2$, where α and β can be positive or negative constants. Using the symmetry, we restrict our work to the two cases where $\alpha > 0$, $\beta > 0$, and $\alpha < 0$, $\beta > 0$. In the following Theorems 2.1 and 2.2, we derive the exact density functions of the linear combination S for these two cases.

Theorem 2.1. *If S_1 and S_2 are jointly distributed according to Eq. (1.1), then the density function of $S = \alpha S_1 + \beta S_2$, for $\alpha > 0$, $\beta > 0$, is given by*

$$f_S(s) = \frac{\Gamma\left(\theta_1+\theta_2+\theta_3\right)}{\Gamma\left(\theta_1+\theta_3\right)\Gamma\left(\theta_2\right)} \frac{\left(s-\alpha\right)^{\theta_2-1}\left(\beta-s\right)^{\theta_1+\theta_3-1}}{\beta^{\theta_3}(\beta-\alpha)^{\theta_1+\theta_2-1}} \tag{2.1}$$

$$\times {}_2F_1(1-\theta_2, \theta_3; \theta_1+\theta_3; \frac{\alpha\left(s-\beta\right)}{\beta\left(s-\alpha\right)}), \ \text{for } \alpha < s < \beta.$$

Proof. From Eq. (1.1), the joint density function of $(S_2, S) = (S_2, \alpha S_1 + \beta S_2)$ can be written as

$$\begin{aligned} f(s_2, s) &= \frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \frac{s_2^{\theta_2-1}}{\alpha} \left(\frac{s-\beta s_2}{\alpha}\right)^{\theta_1-1} (1 - s_2 - \frac{s-\beta s_2}{\alpha})^{\theta_3-1} \\ &= \frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \frac{s_2^{\theta_2-1}\beta^{\theta_1-1}(\beta-\alpha)^{\theta_3-1}}{\alpha^{\theta_1+\theta_3-1}} \left(\frac{s}{\beta} - s_2\right)^{\theta_1-1} (s_2 - \frac{s-\alpha}{\beta-\alpha})^{\theta_3-1}, \end{aligned}$$

where $\frac{s-\alpha}{\beta-\alpha} < s_2 < \frac{s}{\beta}$ and $\alpha < s < \beta$. Thus, the density function of S becomes

$$\begin{aligned} f_S(s) = {} & \frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \frac{\beta^{\theta_1-1}(\beta-\alpha)^{\theta_3-1}}{\alpha^{\theta_1+\theta_3-1}} \\ & \times \int_{\frac{s-\alpha}{\beta-\alpha}}^{\frac{s}{\beta}} s_2^{\theta_2-1} \left(\frac{s}{\beta} - s_2\right)^{\theta_1-1} (s_2 - \tfrac{s-\alpha}{\beta-\alpha})^{\theta_3-1} ds_2 \end{aligned} \tag{2.2}$$

for $\alpha < s < \beta$. Using equation (2.2.6.1) in Prudnikov et al. (1986), the integral in Eq. (2.2) is equal to

$$\frac{\Gamma(\theta_1)\Gamma(\theta_3)}{\Gamma(\theta_1+\theta_3)} \left(\frac{s}{\beta} - \frac{s-\alpha}{\beta-\alpha}\right)^{\theta_1+\theta_3-1} \left(\frac{s-\alpha}{\beta-\alpha}\right)^{\theta_2-1}$$

$$\times {}_2F_1(1-\theta_2, \theta_3; \theta_1+\theta_3; \frac{\alpha(s-\beta)}{\beta(s-\alpha)}).$$

Inserting this quantity in Eq. (2.2), the result follows. □

Theorem 2.2. *If S_1 and S_2 are jointly distributed according to Eq. (1.1), then the density function of $S = \alpha S_1 + \beta S_2$, for $\alpha < 0$, $\beta > 0$, is given by*

$$f_S(s) = \frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1+\theta_3)\Gamma(\theta_2)} \frac{s^{\theta_2-1}(\beta-s)^{\theta_1+\theta_3-1}}{\beta^{\theta_2+\theta_3-1}(\beta-\alpha)^{\theta_1}} \tag{2.3}$$

$$\times {}_2F_1(1-\theta_2, \theta_1; \theta_1+\theta_3; \frac{\alpha(\beta-s)}{s(\beta-\alpha)})$$

for $0 < s < \beta$, and

$$f_S(s) = \frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_2+\theta_3)\Gamma(\theta_1)} \frac{(-s)^{\theta_1-1}(s-\alpha)^{\theta_2+\theta_3-1}}{(-\alpha)^{\theta_1+\theta_3-1}(\beta-\alpha)^{\theta_2}} \tag{2.4}$$

$$\times {}_2F_1(1-\theta_1, \theta_2; \theta_2+\theta_3; \frac{\beta(s-\alpha)}{s(\beta-\alpha)})$$

for $\alpha < s < 0$.

Proof. If $s > 0$, the joint density function of $(S_1, S) = (S_1, \alpha S_1 + \beta S_2)$ can be written as

$$\begin{aligned} f(s_1, s) &= \frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \frac{s_1^{\theta_1-1}}{\beta}\left(\frac{s-\alpha s_1}{\beta}\right)^{\theta_2-1}(1 - s_1 - \frac{s-\alpha s_1}{\beta})^{\theta_3-1} \\ &= \frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \frac{s_1^{\theta_1-1}(\beta-\alpha)^{\theta_3-1}}{\beta^{\theta_2+\theta_3-1}}(s-\alpha s_1)^{\theta_2-1}(\frac{\beta-s}{\beta-\alpha} - s_1)^{\theta_3-1} \end{aligned}$$

where $0 < s_1 < \frac{\beta-s}{\beta-\alpha}$ and $0 < s < \beta$. Thus, the density function of S becomes

$$\begin{aligned} f_S(s) &= \frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \frac{(\beta-\alpha)^{\theta_3-1}}{\beta^{\theta_2+\theta_3-1}} \\ &\quad \times \int_0^{\frac{\beta-s}{\beta-\alpha}} s_1^{\theta_1-1}(s-\alpha s_1)^{\theta_2-1}(\frac{\beta-s}{\beta-\alpha} - s_1)^{\theta_3-1} ds_1 \end{aligned}$$

for $0 < s < \beta$. Making the change of variable $u = \frac{\beta-\alpha}{\beta-s}s_1$, we obtain

$$\begin{aligned} f_S(s) = &\frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \frac{s^{\theta_2-1}(\beta-\alpha)^{\theta_3-1}}{\beta^{\theta_2+\theta_3-1}}\left(\frac{\beta-s}{\beta-\alpha}\right)^{\theta_1+\theta_3-1} \\ &\times \int_0^1 u^{\theta_1-1}(1-u)^{\theta_3-1}\left(1 - u\frac{\alpha(\beta-s)}{s(\beta-\alpha)}\right)^{\theta_2-1} du. \end{aligned} \tag{2.5}$$

Using equation (2.2.6.15) in Prudnikov et al. (1986), the integral in Eq. (2.5) is equal to

$$\frac{\Gamma(\theta_1)\Gamma(\theta_3)}{\Gamma(\theta_1+\theta_3)} \times {}_2F_1(1-\theta_2, \theta_1; \theta_1+\theta_3; \frac{\alpha(\beta-s)}{s(\beta-\alpha)}).$$

Inserting this quantity in Eq. (2.5), the Eq. (2.3) follows. The proof of Eq. (2.4) can be treated similarly. □

Plots of the density function for some selected values of $\boldsymbol{\theta}$, α and β are shown in Figures 1 and 2, respectively. The four curves in each plot correspond to the selected values of θ_1 and θ_2. It is clear that the densities become flatter by increasing the value of β.

3. Moments

In this section, we give the moments of $S = \alpha S_1 + \beta S_2$, when S_1 and S_2 are jointly distributed according to Eq. (1.1). Now we establish the following result:

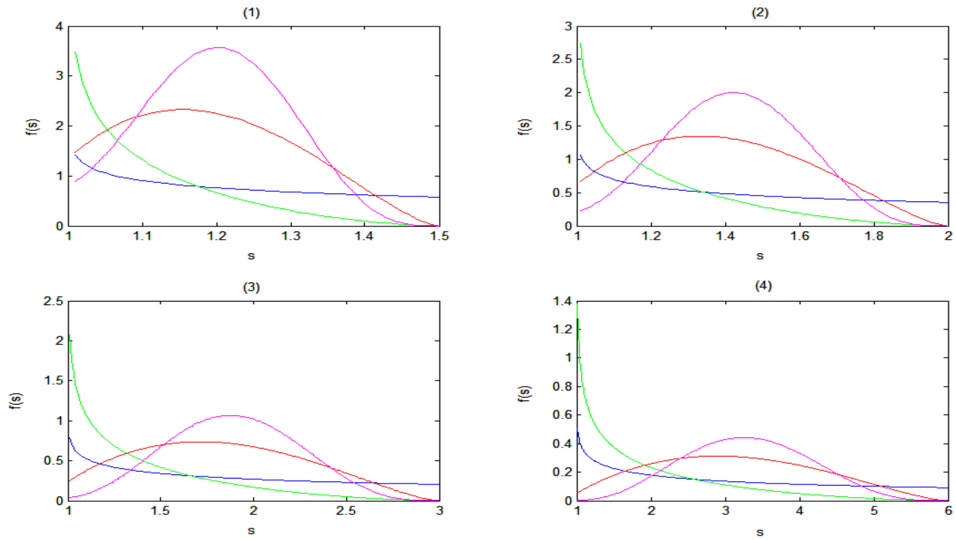

FIGURE 1. The density function (2.1) for $\alpha = 1$, $\theta_3 = 0.5$ and (1): $\beta = 1.5$; (2): $\beta = 2$; (3): $\beta = 3$; (4): $\beta = 6$. The four curves in each plot are: the blue curve ($\theta_1 = 0.5$, $\theta_2 = 0.5$), the green curve ($\theta_1 = 2$, $\theta_2 = 0.5$), the red curve ($\theta_1 = 2$, $\theta_2 = 2$), and the magenta curve ($\theta_1 = 4$, $\theta_2 = 4$).

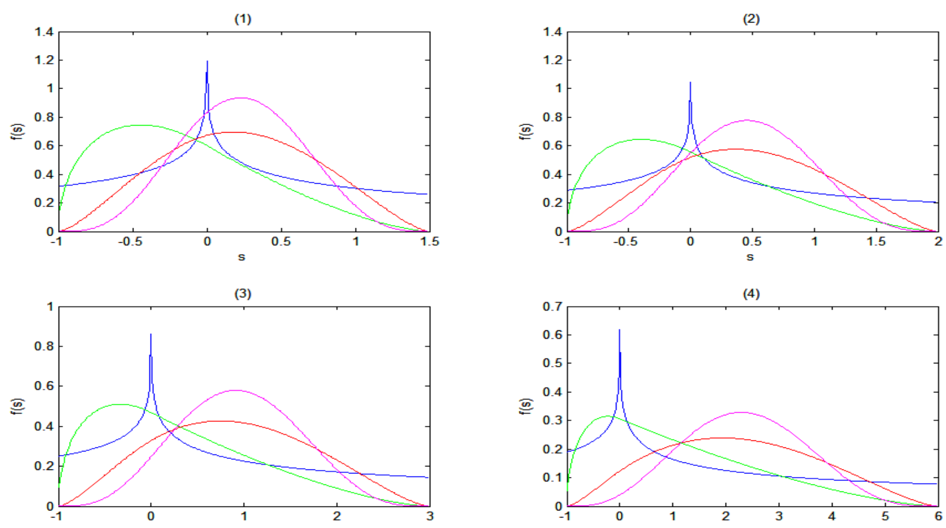

FIGURE 2. The density functions (2.3)-(2.4) for $\alpha = -1$, $\theta_3 = 0.5$ and (1): $\beta = 1.5$; (2): $\beta = 2$; (3): $\beta = 3$; (4): $\beta = 6$. The four curves in each plot are: the blue curve ($\theta_1 = 0.5$, $\theta_2 = 0.5$), the green curve ($\theta_1 = 2$, $\theta_2 = 1$), the red curve ($\theta_1 = 2$, $\theta_2 = 2$), and the magenta curve ($\theta_1 = 4$, $\theta_2 = 4$).

Theorem 3.1. *If S_1 and S_2 are jointly distributed according to Eq. (1.1), then the moment function of $S = \alpha S_1 + \beta S_2$ is given by (for $n \geq 0$)*

(3.1)

$$\mathbf{E}\left(S^n\right) = \frac{\Gamma\left(\theta_1+\theta_2+\theta_3\right)}{\Gamma\left(\theta_1+\theta_2+\theta_3+n\right)} \frac{\alpha^n}{\Gamma\left(\theta_1\right)\Gamma\left(\theta_2\right)} \sum_{k=0}^{n} \binom{n}{k} \Gamma\left(\theta_2+k\right)\Gamma\left(\theta_1+n-k\right)\left(\frac{\beta}{\alpha}\right)^k$$

Proof. The moment function of S can be written as

$$\begin{aligned}\mathbf{E}\left(S^n\right) &= \mathbf{E}\left(\left(\alpha S_1+\beta S_2\right)^n\right)\\ &= \sum_{k=0}^{n}\binom{n}{k}\alpha^{n-k}\beta^k\phi\left(n-k;k\right),\end{aligned}$$

where $\phi\left(n-k;k\right) := \mathbf{E}\left(S_1^{n-k}S_2^k\right)$. By using Eq. (1.2), the result follows. □

From this, we obtain the following corollary, where Eq. (3.1) reduces to particular cases for $n = 1$ and $n = 2$.

Corollary 3.1. *The first two moments of S are*

$$\mathbf{E}\left(S\right) = \frac{\alpha\theta_1+\beta\theta_2}{\theta_1+\theta_2+\theta_3} \tag{3.2}$$

and

$$\mathbf{E}\left(S^2\right) = \frac{\left(\alpha\theta_1+\beta\theta_2\right)^2+\alpha^2\theta_1+\beta^2\theta_2}{\left(\theta_1+\theta_2+\theta_3\right)\left(\theta_1+\theta_2+\theta_3+1\right)}. \tag{3.3}$$

□

Remark: We can give the moment function of $\sum_{m=1}^{k}\alpha_m S_m$, where $(S_1,...,S_k)$ are k random variables distributed according to Dirichlet distribution with parameters $(\theta_1,...,\theta_{k+1})$. Indeed, the joint moment function of $(S_1,...,S_k)$ is given by

$$\mathbf{E}\left(\prod_{m=1}^{k}S_m^{q_m}\right) = \frac{\Gamma\left(\sum_{m=1}^{k+1}\theta_m\right)}{\Gamma\left(\sum_{m=1}^{k+1}\theta_m+\sum_{m=1}^{k}q_m\right)}\prod_{m=1}^{k}\frac{\Gamma\left(\theta_m+q_m\right)}{\Gamma\left(\theta_m\right)}.$$

From this and by using the multinomial formula, we obtain

$$\mathbf{E}\left(\left(\sum_{m=1}^{k}\alpha_m S_m\right)^n\right) = \sum_{\substack{q_1,\dots,q_k\geq 0\\ \sum_{m=1}^{k}q_m=n}}\frac{n!}{\prod_{m=1}^{k}q_m!}\prod_{m=1}^{k}\alpha_m^{q_m}\mathbf{E}\left(\prod_{m=1}^{k}S_m^{q_m}\right)$$

$$= \frac{\Gamma\left(\sum_{m=1}^{k+1}\theta_m\right)}{\Gamma\left(\sum_{m=1}^{k+1}\theta_m + n\right)} \sum_{\substack{q_1,\dots,q_k\geq 0\\ \sum_{m=1}^{k} q_m = n}} \frac{n!}{\prod_{m=1}^{k} q_m!} \prod_{m=1}^{k} \alpha_m^{q_m} \frac{\Gamma\left(\theta_m + q_m\right)}{\Gamma\left(\theta_m\right)}.$$

It is straightforward to found Eq. (3.1) for $k = 2$.

4. Percentiles

In this section, we provide extensive tabulations of the percentiles of the distribution of $S = \alpha S_1 + \beta S_2$. These percentiles are computed numerically by solving the following equations, with respect to s_q,

$$\frac{\Gamma\left(\theta_1+\theta_2+\theta_3\right)}{\Gamma\left(\theta_1+\theta_3\right)\Gamma\left(\theta_2\right)} \frac{1}{\beta^{\theta_3}(\beta-\alpha)^{\theta_1+\theta_2-1}} \int_0^{s_q} \left(s-\alpha\right)^{\theta_2-1}\left(\beta-s\right)^{\theta_1+\theta_3-1}$$

$$\times {}_2F_1(1-\theta_2,\theta_3;\theta_1+\theta_3;\frac{\alpha\left(s-\beta\right)}{\beta\left(s-\alpha\right)})ds = q,$$

$$\frac{\Gamma\left(\theta_1+\theta_2+\theta_3\right)}{\Gamma\left(\theta_1+\theta_3\right)\Gamma\left(\theta_2\right)} \frac{1}{\beta^{\theta_2+\theta_3-1}(\beta-\alpha)^{\theta_1}} \int_0^{s_q} s^{\theta_2-1}(\beta-s)^{\theta_1+\theta_3-1}$$

$$\times {}_2F_1(1-\theta_2,\theta_1;\theta_1+\theta_3;\frac{\alpha\left(\beta-s\right)}{s\left(\beta-\alpha\right)})ds = q$$

and

$$\frac{\Gamma\left(\theta_1+\theta_2+\theta_3\right)}{\Gamma\left(\theta_2+\theta_3\right)\Gamma\left(\theta_1\right)} \frac{1}{\left(-\alpha\right)^{\theta_1+\theta_3-1}(\beta-\alpha)^{\theta_2}} \int_0^{s_q} \left(-s\right)^{\theta_1-1}\left(s-\alpha\right)^{\theta_2+\theta_3-1}$$

$$\times {}_2F_1(1-\theta_1,\theta_2;\theta_2+\theta_3;\frac{\beta\left(s-\alpha\right)}{s\left(\beta-\alpha\right)})ds = q.$$

Evidently, this involves computation of the Gauss hypergeometric function and routines for this are widely available. Table 1 provides the numerical values of s_q for $\alpha = -2, -1.9, ..., 1.9$, $\beta = 2$, $\theta_1 = 4$, $\theta_2 = 4$, $\theta_3 = 1$. We hope that these values will be of use to the practitioners mentioned in Section 1. Similar tabulations could be easily derived for other values of α, β, θ_1, θ_2 and θ_3.

Table 1. Percentile points of S

α	0.7	0.8	0.9	0.95	0.99
-2	0.4196	0.5253	0.5981	0.6179	0.6246
-1.9	0.4315	0.5402	0.6145	0.6342	0.6407
-1.8	0.4397	0.5519	0.6293	0.6502	0.6575
-1.7	0.4521	0.5667	0.6462	0.6680	0.6751
-1.6	0.4618	0.5824	0.6637	0.6866	0.6935
-1.5	0.4738	0.5988	0.6840	0.7055	0.7128
-1.4	0.4865	0.6174	0.7033	0.7258	0.7330
-1.3	0.4918	0.6319	0.7239	0.7470	0.7545
-1.2	0.4968	0.6510	0.7452	0.7692	0.7769
-1.1	0.5009	0.6674	0.7669	0.7925	0.8004
-1	0.4923	0.6789	0.7885	0.8159	0.8246
-0.9	0.4757	0.6864	0.8087	0.8392	0.8490
-0.8	0.4448	0.6789	0.8242	0.8604	0.8717
-0.7	0.3973	0.6539	0.8292	0.8737	0.8887
-0.6	0.3137	0.6039	0.8173	0.8736	0.8923
-0.5	0.2070	0.5170	0.7699	0.8458	0.8689
-0.4	0.1715	0.3794	0.6687	0.7644	0.7984
-0.3	0.0910	0.1944	0.4918	0.6070	0.6515
-0.2	0.0068	0.0230	0.2344	0.3589	0.4113
-0.1	0.0005	0.0069	0.0188	0.0786	0.1218
0	0.8345	1.0148	1.1333	1.1644	1.1746
0.1	0.8803	1.0610	1.1814	1.2134	1.2232
0.2	0.9250	1.1095	1.2327	1.2648	1.2753
0.3	0.9736	1.1625	1.2879	1.3205	1.3314
0.4	1.0266	1.2199	1.3474	1.3807	1.3918
0.5	1.0845	1.2821	1.4119	1.4463	1.4569
0.6	1.1494	1.3491	1.4815	1.5161	1.5272
0.7	1.2195	1.4230	1.5574	1.5922	1.6037
0.8	1.2979	1.5039	1.6398	1.6752	1.6869
0.9	1.3840	1.5934	1.7305	1.7675	1.7777
1	1.4808	1.6918	1.8298	1.8653	1.8769
1.1	1.5889	1.8006	1.9392	1.9739	1.9868
1.2	1.7108	1.9225	2.0607	2.0966	2.1078
1.3	1.8496	2.0584	2.1959	2.2318	2.2430
1.4	2.0099	2.2119	2.3473	2.3828	2.3940

1.5	2.1917	2.3882	2.5197	2.5554	2.5652
1.6	2.4048	2.5940	2.7169	2.7488	2.7595
1.7	2.6655	2.8289	2.9441	2.9707	2.9851
1.8	2.9590	3.1159	3.2155	3.3248	3.2503
1.9	3.0053	3.4177	3.5476	3.5488	3.5488

5. Entropy

An entropy provides an excellent tool to quantify the amount of information (or uncertainty) contained in a random observation regarding its parent distribution (population). A large value of entropy implies greater uncertainty in the data. The concept of entropy is important in different areas such as physics, probability and statistics, communication theory, and economics, etc. Several measures of entropy have been studied and compared in the literature. The simplest known entropy is the Shannon entropy (Shannon (1948)) of the random variable $S = \alpha S_1 + \beta S_2$ defined by

$$\mathbf{E}[-\log f(S)] = -\int f(s)\log f(s)ds. \tag{5.1}$$

One of the main extensions of the Shannon entropy was defined by Rényi (1961). This generalized entropy measure is given by

$$H_R(\mu, f) = \frac{\log G(\mu)}{1-\mu} \text{ (for } \mu > 0 \text{ and } \mu \neq 1), \tag{5.2}$$

where

$$G(\mu) := \int (f(s))^{\mu} \, ds.$$

Note that the Shannon entropy is the particular case of the Rényi entropy for $\mu \uparrow 1$.

Consider calculating the Shannon entropy when S has the density function given by Eq. (2.1). If $\alpha < s < \beta$, then one can write

$$\begin{aligned}\mathbf{E}[-\log f(S)] &= -\log\left[\frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1+\theta_3)\Gamma(\theta_2)}\frac{1}{\beta^{\theta_3}(\beta-\alpha)^{\theta_1+\theta_2-1}}\right] \\ &\quad + (1-\theta_2)\cdot\mathbf{E}[\log(S-\alpha)] + (1-\theta_1-\theta_3)\cdot\mathbf{E}[\log(\beta-S)] \\ &\quad - \frac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1+\theta_3)\Gamma(\theta_2)}\frac{1}{\beta^{\theta_3}(\beta-\alpha)^{\theta_1+\theta_2-1}} I,\end{aligned}$$

where I denotes the integral

$$I = \int_{\alpha}^{\beta} (s-\alpha)^{\theta_2-1}(\beta-s)^{\theta_1+\theta_3-1}\ln\left({}_2F_1(1-\theta_2,\theta_3;\theta_1+\theta_3;\frac{\alpha(s-\beta)}{\beta(s-\alpha)})\right) ds.$$

Unfortunately, this integral I cannot be reduced to a closed form. Thus, one cannot obtain a closed form expression for the entropy measure when S is distributed as in Theorem 2.1 or as in Theorem 2.2. Hence, we performed a numerical study to examine the behavior of Eq. (5.1) with respect to the parameters in Theorem 2.1 and Theorem 2.2 respectively.

Figure 3 below shows the variation of Eq. (5.1) for a range of values of α and θ_1 with $\theta_2 = 2$, $\theta_3 = 1$ and $\beta = 6$. The effect of the parameters is evident: Eq. (5.1) is an increasing function of both $\alpha < 0$ and θ_1 whereas Eq. (5.1) increases with respect to $\alpha > 0$ but with respect to θ_1, it initially increases before decreasing. Note that similar results can be obtained if we change θ_1 by θ_2 or θ_3 and if we fix α and change β. Finally, for the reasons mentioned above, one cannot obtain closed form expressions for the Rényi entropy and the investigation will have to be performed numerically.

Next, we derive exact forms of Rényi and Shannon entropies under the constraint $\theta_1 + \theta_2 + \theta_3 = 1$, for the distribution of S defined by Eq. (2.1) with $0 < \alpha < s < \beta$. Note that the case where $\alpha < 0$ and $\beta > 0$ can be traited similarly.

Firstly, we recall that the density function of S in the particular case $\theta_1+\theta_2+\theta_3 = 1$ is given by

$$f_S(s) = \frac{s^{\theta_1+\theta_2-1} (s-\alpha)^{-\theta_1} (\beta - s)^{-\theta_2}}{\Gamma(1-\theta_2)\Gamma(\theta_2)} \text{ for } 0 < \alpha < s < \beta \tag{5.3}$$

Before continuing, we need the following lemma:

Lemma 5.1. *Let* $g(\alpha, \beta, \theta_1, \theta_2) = \lim_{\mu \to 1} h(\mu)$, *where*

$$h(\mu) = \frac{d}{d\mu}\, {}_2F_1\left(\mu(1-\theta_1-\theta_2), 1-\mu\theta_2; 2-\mu(\theta_1+\theta_2); \frac{\beta-\alpha}{\beta}\right) \tag{5.4}$$

Then,

$$g(\alpha, \beta, \theta_1, \theta_2) = \sum_{j \geq 1} \frac{\Gamma(1-\theta_2+j)\Gamma(1-\theta_1-\theta_2+j)\Gamma(2-\theta_1-\theta_2)}{\Gamma(1-\theta_2)\Gamma(1-\theta_1-\theta_2)\Gamma(2-\theta_1-\theta_2+j)\, j!} \left(\frac{\beta-\alpha}{\beta}\right)^j$$

$$\times\ [\theta_2\Psi(1-\theta_2) - (1-\theta_1-\theta_2)\Psi(1-\theta_1-\theta_2) + (\theta_1+\theta_2)\Psi(2-\theta_1-\theta_2+j)$$

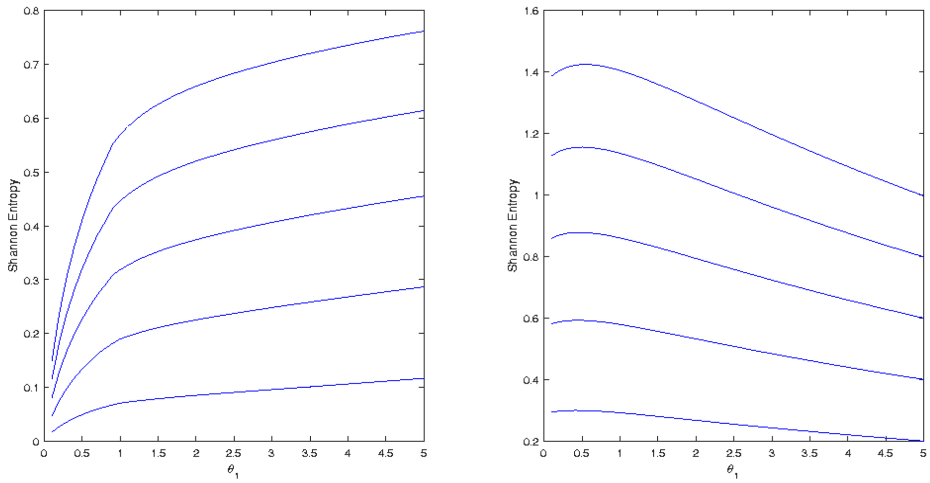

FIGURE 3. Plots of the Shannon entropy for $\theta_2 = 2$ and $\theta_3 = 1$, $\beta = 6$, $\alpha = -5, -4, -3, -2, -1$ and $\theta_1 = 0.1, 0.2, ..., 5$ (left); for $\theta_2 = 2$ and $\theta_3 = 1$, $\beta = 6$, $\alpha = 1, 2, 3, 4, 5$ and $\theta_1 = 0.1, 0.2, ..., 5$ (right).

(5.5)
$$-\theta_2 \Psi(1-\theta_2+j) + (1-\theta_1-\theta_2)\Psi(1-\theta_1-\theta_2+j) - (\theta_1+\theta_2)\Psi(2-\theta_1-\theta_2)],$$

where $\Psi(.) = \Gamma'(.)/\Gamma(.)$ *is the digamma function.*

Proof. Expanding ${}_2F_1$ in series form, we write

$$h(\mu) = \frac{\mathrm{d}}{\mathrm{d}\mu}\sum_{j\geq 0}\frac{\Delta_j(\mu)}{j!}\left(\frac{\beta-\alpha}{\beta}\right)^j = \sum_{j\geq 0}\left[\frac{\mathrm{d}}{\mathrm{d}\mu}\Delta_j(\mu)\right]\frac{(\beta-\alpha)^j}{j!\beta^j} \tag{5.6}$$

where

$$\Delta_j(\mu) = \frac{\Gamma(1-\mu\theta_2+j)\Gamma(\mu(1-\theta_1-\theta_2)+j)\Gamma(2-\mu(\theta_1+\theta_2))}{\Gamma(1-\mu\theta_2)\Gamma(\mu(1-\theta_1-\theta_2))\Gamma(2-\mu(\theta_1+\theta_2)+j)}$$

Now, differentiating the logarithm of $\Delta_j(\mu)$ with respect to μ, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\Delta_j(\mu) = \Delta_j(\mu)\left[\theta_2\Psi(1-\mu\theta_2) - (1-\theta_1-\theta_2)\Psi(\mu(1-\theta_1-\theta_2))\right.$$

$$+ (\theta_1+\theta_2)\Psi(2-\mu(\theta_1-\theta_2)+j) - \theta_2\Psi(1-\mu\theta_2+j)$$

$$\left. + (1-\theta_1-\theta_2)\Psi(\mu(1-\theta_1-\theta_2)+j) - (\theta_1+\theta_2)\Psi(2-\mu(\theta_1-\theta_2))\right] \tag{5.7}$$

Finally, substituting Eq. (5.7) in Eq. (5.6) and taking $\mu \to 1$, we obtain Eq. (5.5). $\square$

Theorem 5.1. *Consider the random variable S defined by its density function given in Eq. (5.3). The Rényi and the Shannon entropies are given respectively by*

$$H_R(\mu, f) = \frac{1}{1-\mu}[\mu \log C(\theta_2) + (1-\mu(\theta_1+\theta_2)) \log(\beta-\alpha)$$

$$+\mu(\theta_1+\theta_2-1)\log(\beta) + \log(\Gamma(1-\mu\theta_1)) + \log(\Gamma(1-\mu\theta_2))$$

$$-\log(\Gamma(2-\mu(\theta_1+\theta_2))) + \log\ {}_2F_1\left(\mu(1-\theta_1-\theta_2), 1-\mu\theta_2; 2-\mu(\theta_1+\theta_2); \frac{\beta-\alpha}{\beta}\right)]$$

and

$$H_{SH}(f) = -\log C(\theta_2) + (\theta_1+\theta_2)\log(\beta-\alpha) - (\theta_1+\theta_2-1)\log(\beta)$$

$$+\theta_1\Psi(1-\theta_1) + \theta_2\Psi(1-\theta_2) - (\theta_1+\theta_2)\Psi(2-(\theta_1+\theta_2))$$

$$-\frac{g(\alpha,\beta,\theta_1,\theta_2)}{{}_2F_1\left(1-\theta_1-\theta_2, 1-\theta_2; 2-\theta_1-\theta_2; \frac{\beta-\alpha}{\beta}\right)},$$

where $C(\theta_2) = (\Gamma(1-\theta_2)\Gamma(\theta_2))^{-1}$ and $g(\alpha,\beta,\theta_1,\theta_2)$ is given in the previous lemma.

Proof. Let us compute $G(\mu)$

$$\begin{aligned} G(\mu) &= \int_\alpha^\beta (f(s))^\mu\, ds \\ &= \int_\alpha^\beta \frac{s^{\mu(\theta_1+\theta_2-1)}(s-\alpha)^{-\mu\theta_1}(\beta-s)^{-\mu\theta_2}}{(\Gamma(1-\theta_2)\Gamma(\theta_2))^\mu} ds \end{aligned}$$

Using again equation (2.2.6.1) in Prudnikov et al. (1986), we obtain

$$\begin{aligned} G(\mu) &= \frac{\Gamma(1-\mu\theta_1)\Gamma(1-\mu\theta_2)}{\Gamma(2-\mu(\theta_1+\theta_2))}(C(\theta_2))^\mu(\beta-\alpha)^{1-\mu(\theta_1+\theta_2)}\alpha^{\mu(\theta_1+\theta_2-1)} \\ &\quad \times {}_2F_1\left(\mu(1-\theta_1-\theta_2), 1-\mu\theta_1, ; 2-\mu(\theta_1+\theta_2); \frac{\alpha-\beta}{\alpha}\right) \end{aligned}$$

Now, we use the following Euler's relation

$$(5.8) \qquad {}_2F_1(a,b;c;z) = (1-z)^{-a}\ {}_2F_1(a, c-b; c; \frac{z}{z-1}),$$

to write

$$\begin{aligned} G(\mu) &= \frac{\Gamma(1-\mu\theta_1)\Gamma(1-\mu\theta_2)}{\Gamma(2-\mu(\theta_1+\theta_2))}(C(\theta_2))^\mu(\beta-\alpha)^{1-\mu(\theta_1+\theta_2)}\beta^{\mu(\theta_1+\theta_2-1)} \\ &\quad \times {}_2F_1\left(\mu(1-\theta_1-\theta_2), 1-\mu\theta_2, ; 2-\mu(\theta_1+\theta_2); \frac{\beta-\alpha}{\beta}\right) \end{aligned}$$

Taking logarithm of $G(\mu)$ and using Eq. (5.2), we obtain the Rényi entropy. The Shannon entropy is obtained from the Rényi entropy by taking $\mu \uparrow 1$ and using L'Hopital's rule. $\square$

6. Approximations

It is clear that if $\alpha > 0$, $\beta > 0$ (respectively if $\alpha < 0$, $\beta > 0$); the random variable $Y = \dfrac{S-\alpha}{\beta-\alpha}$ (respectively $Y = \dfrac{S}{\beta}$ if $S \in (0,\beta)$ or $Y = \dfrac{S}{\alpha}$ if $S \in (\alpha, 0)$) has support in the interval $[0,1]$. Accordingly, it is evident that we will motivate to approximate its distribution by the beta distribution with parameters $a, b > 0$ (say $Y \overset{d}{\sim} \beta(a,b)$) and density function

$$f_Y(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}. \tag{6.1}$$

The idea of approximating distributions including complicated formulae with the beta distribution is frequently tackled in the statistics literature, based on the interesting work of Das Gupta (1968) (see also Sculli and Wong (1985), Fan (1991), Johannesson and Giri (1995), Nadarajah and Kotz (2004) and Gupta and Nadarajah (2006)). Note that recently, Ghorbel (2019) developed this single beta distribution of the form (6.1) for the bivariate beta distribution. Subsequently, this transformation was opted for owing to its popularity, simplicity and its association with the derived distribution. The basic reason for doing this doesn't lie in the fact that the calculation of Eq. (2.1), Eq. (2.3) and Eq. (2.4) cannot be handled; it rather resides in providing a simple approximation in terms of the beta distribution so that one can use the known procedures for inference, prediction, etc. Furthermore, the presented approximation seems to be useful especially to practitioners who not only avoid the use of the Gauss hypergeometric function but also regard the beta distribution as widely accessible in standard statistical packages.

The choice of the beta parameters a and b is made using the method of moments. The first two moments of Y can be written as

$$\mathbf{E}(Y) = \frac{a}{a+b} \text{ and } \mathbf{E}\left(Y^2\right) = \frac{a(a+1)}{(a+b)(a+b+1)}.$$

After some algebraic manipulation, it is straightforward to find that

$$a=\mathbf{E}\left(Y\right)\frac{\mathbf{E}\left(Y\right)-\mathbf{E}\left(Y^{2}\right)}{\mathbf{E}\left(Y^{2}\right)-\mathbf{E}^{2}\left(Y\right)}\tag{6.2}$$

and

$$b=\left(1-\mathbf{E}\left(Y\right)\right)\frac{\mathbf{E}\left(Y\right)-\mathbf{E}\left(Y^{2}\right)}{\mathbf{E}\left(Y^{2}\right)-\mathbf{E}^{2}\left(Y\right)}.\tag{6.3}$$

If $\alpha>0$, $\beta>0$ (approximation 1), we have $\mathbf{E}\left(Y\right)=\left(\mathbf{E}\left(S\right)-\alpha\right)/\left(\beta-\alpha\right)$ and

$$\mathbf{E}\left(Y^{2}\right)=\left(\mathbf{E}\left(S^{2}\right)-2\alpha\mathbf{E}\left(S\right)+\alpha^{2}\right)/\left(\beta-\alpha\right)^{2},$$

and if $\alpha<0$, $\beta>0$ (approximation 2), we have $\mathbf{E}\left(Y\right)=\mathbf{E}\left(S\right)/\beta$ and $\mathbf{E}\left(Y^{2}\right)=\mathbf{E}\left(S^{2}\right)/\beta^{2}$ (respectively $\mathbf{E}\left(Y\right)=\mathbf{E}\left(S\right)/\alpha$ and $\mathbf{E}\left(Y^{2}\right)=\mathbf{E}\left(S^{2}\right)/\alpha^{2}$)) if $S\in(0,\beta)$ (respectively if $S\in(\alpha,0)$). The two moments $\mathbf{E}\left(S\right)$ and $\mathbf{E}\left(S^{2}\right)$ can be computed using Eq. (3.2) and Eq. (3.3), respectively for given values of the parameters $\boldsymbol{\theta}$, α and β.

In order to show the robustness of the approximation, we selected fifteen values for each case of the parameters $\boldsymbol{\theta}$, α and β and computed the corresponding estimates for a and b using Eq. (6.2) and Eq. (6.3). From this, we use the selected parameters $\boldsymbol{\theta}$, α and β and the estimates are presented in Tables 2, 3 and 4, respectively. Next, we checked robustness by comparing the exact and the approximated density functions of Y, as given by Eq. (2.1) (respectively Eq. (2.3) and Eq. (2.4)) and Eq. (6.1). These comparisons are illustrated in Figures 4, 5 and 6.

We notice that our approximation is robust for most of the parameter choices. Some examples are illustrated in Figure 4. Besides, we deduce that the robustness is weak when both θ_1 and θ_2 are small (see for example Figure 4). Next, it is clear that approximation is the most robust in terms of increasing the value of β (see Figure 5). Finally, Figure 6 reveals that approximation is the most robust in terms of decreasing the value of α. It is worth noting that the numerical results are obtained by the use of hypergeom function in MATLAB.

TABLE 2. Estimates of (a, b) for selected $(\theta_1, \theta_2, \beta)$ and for $\alpha = 1$ and $\theta_3 = 0.5$

θ_1	θ_2	β	a	b
1	0.5	3	0.0132	0.0921
2	0.5	3	0.0431	0.4741
2	2	3	1.1947	1.8774
2	4	3	2.9228	2.1434
4	4	3	3.1138	3.9442
1	0.5	4	0.1111	0.5556
2	0.5	4	0.1429	1.1429
2	2	4	1.4528	2.1132
2	4	4	3.2857	2.2857
4	4	4	3.4242	4.1685
1	0.5	6	0.2364	0.9455
2	0.5	6	0.2629	1.7086
2	2	6	1.6712	2.2869
2	4	6	3.5786	2.3857
4	4	6	3.6655	4.3234

TABLE 3. Estimates of (a, b) for selected $(\theta_1, \theta_2, \beta)$ and for $\alpha = -1$ and $\theta_3 = 0.5$

θ_1	θ_2	β	a	b
0.5	0.5	3	0.0769	0.2692
0.5	2	3	1.2525	0.7970
0.5	4	3	2.9179	0.8881
2	4	3	1.9728	1.8741
4	4	3	1.1981	2.6208
0.5	0.5	4	0.1518	0.4554
0.5	2	4	1.4205	0.8523
0.5	4	4	3.1683	0.9198
2	4	4	2.3734	2.0343
4	4	4	1.6866	3.0922
0.5	0.5	6	0.2471	0.6424
0.5	2	6	1.6019	0.9054
0.5	4	6	3.4336	0.9497
2	4	6	2.8387	2.1935
4	4	6	2.2999	3.5649

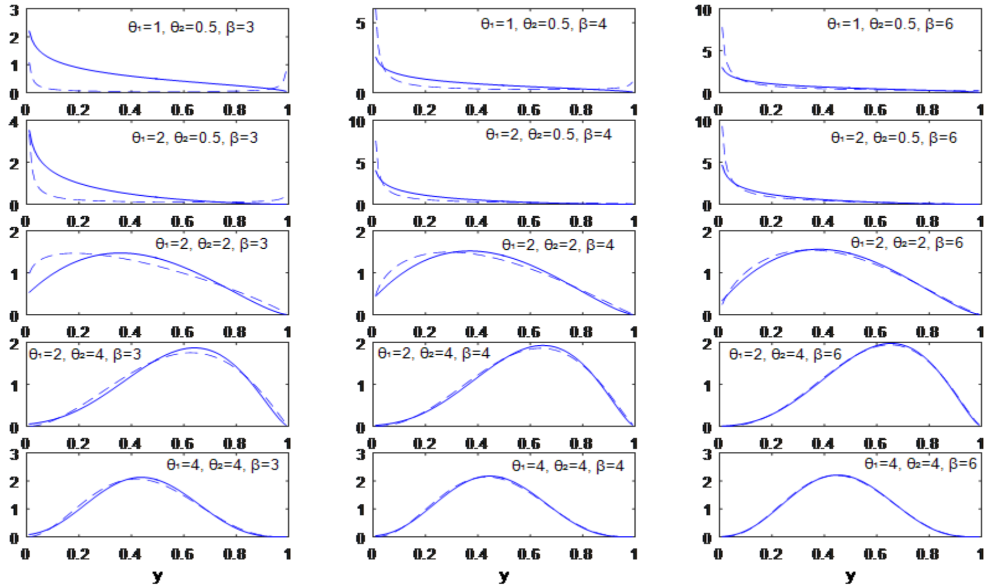

FIGURE 4. The exact density function of S (solid curve) and its approximated function (broken curve) for fixed values $\alpha = 1$, $\theta_3 = 0.5$.

TABLE 4. Estimates of (a, b) for selected $(\theta_1, \theta_2, \alpha)$ and for $\beta = 1$ and $\theta_3 = 0.5$

θ_2	θ_1	α	a	b
0.5	0.5	−5	0.2065	0.5677
0.5	2	−5	1.5278	0.8845
0.5	4	−5	3.3258	0.9380
2	4	−5	2.6442	2.1301
4	4	−5	2.0382	3.3758
0.5	0.5	−6	0.2471	0.6424
0.5	2	−6	1.6019	0.9054
0.5	4	−6	3.4336	0.9497
2	4	−6	2.8387	2.1935
4	4	−6	2.2999	3.5649
0.5	0.5	−8	0.3027	0.7350
0.5	2	−8	1.6974	0.9308
0.5	4	−8	3.5712	0.9637
2	4	−8	3.0985	2.2722
4	4	−8	2.6606	3.8008

7. REAL DATA APPLICATIONS

We initially consider the data of rain and snow collect from five French cities and present an application of the model given by Eq. (1.1).

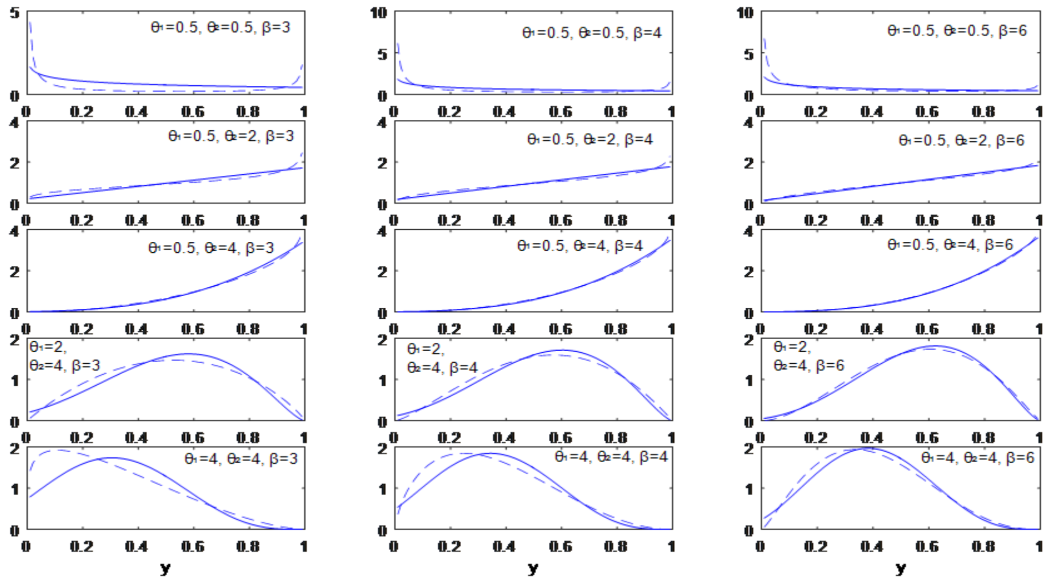

FIGURE 5. The exact density function of S (solid curve) and its approximated function (broken curve) for fixed values $\alpha = -1$, $\theta_3 = 0.5$.

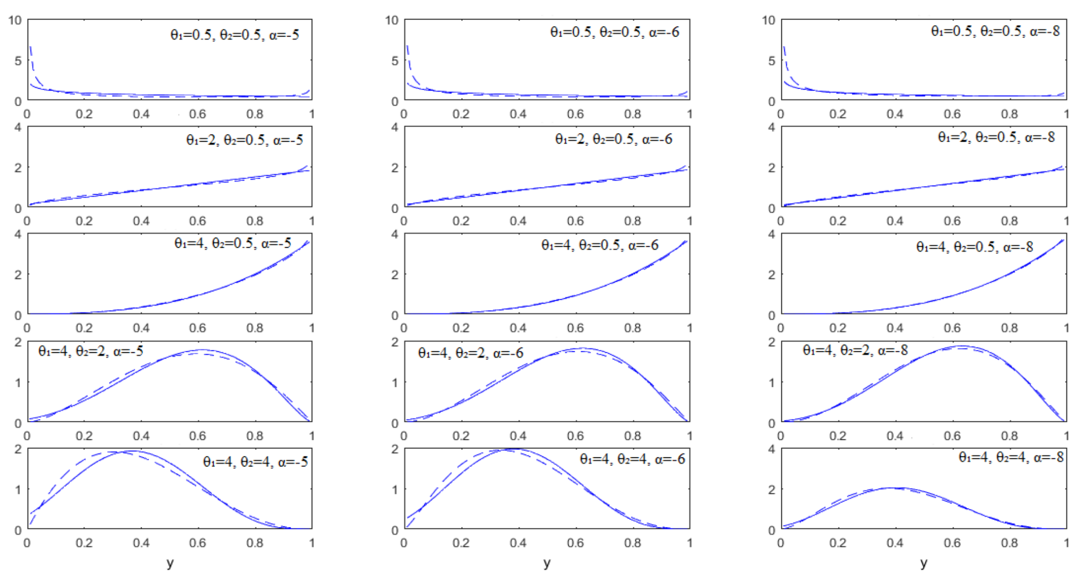

FIGURE 6. The exact density function of S (solid curve) and its approximated function (broken curve) for fixed values $\beta = 1$, $\theta_3 = 0.5$.

Precipitation is any type of water that forms in the Earth's atmosphere and then drops onto the surface of the Earth. Clouds eventually get too full of water vapor, and the precipitation turns into a liquid (rain) or a solid (snow). The rain stands for precipitation that occurs in different sizes, from big, heavy drops to light ones,

TABLE 5. Estimated values of $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ and of $\mathbf{E}(S)$

City	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\mathbf{E}(S)$
Dijon	27.0001	2.9755	25.1356	0.5439
Lille	36.0639	2.7709	30.1578	0.5629
Rennes	26.7630	1.3470	24.2960	0.5364
Paris	46.9868	3.1354	40.8413	0.5510
Toulouse	51.1998	2.0747	66.8744	0.4434

but snow is one of the solid types of precipitation. It is made of water that has been frozen.

We consider data available at the website https://en.tutiempo.net/climate collected from the following five French cities (in different parts of the country) regarding the rain and snow: Dijon, Lille, Rennes, Paris and Toulouse. The dataset comprises the number of days in each year where rain and snow appeared during the period from 1990 to 2018. We consider the following variables:

- S_1 : proportion of days with rain.
- S_2 : proportion of days with snow.
- $S = S_1 + S_2$: proportion of days with precipitation (rain or snow).

The vector (S_1, S_2) would follow the Dirichlet distribution with parameters θ_1, θ_2 and θ_3. These unknown parameters of Eq. (1.1) are estimated by using the maximum likelihood method, and by implementing Fisher scoring method (for more details, see Ronning (1989) and the algorithm used by Narayanan (1991, 1992)). The estimates of parameters and the estimated values of the moments $\mathbf{E}(S)$ (given by Eq. (3.2)) for five cities are outlined in Table 5.

It was reported that the proportion of days with precipitation phenomenon (rain or snow) was similar for the five cities. Toulouse city presented the lowest proportion.

The second application is from finance and we will use the distribution of the linear combination in financial data fitting and show its computability of some financial tools.

If p_t is the closing price of a stock at time t , then the simple rate of return between times t and $t-1$ is defined as $R_t := (p_t - p_{t-1})/p_{t-1}$, which is the percentage price difference. It is well-known that (see for example Danielsson(2011)) a financial portfolio's random rate of return is a linear combination of the random rate of returns on the individual assets in the portfolio. For example, a portfolio composed of two assets can be given as $R_p = \alpha R_1 + \beta R_2$, where $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are the portfolio weights associated with the assets R_1 and R_2, respectively, and $\alpha + \beta = 1$. Note that the choice of weighted index α is 0.001 and β is 0.999 to respect Eq. (2.1) where $\alpha < s < \beta$. From this, the variance of the rate of return on the two-asset portfolio is

$$S^2 := \sigma_p^2 = \alpha^2 S_1^2 + \beta^2 S_2^2 + 2\alpha\beta\rho_{12} S_1 S_2, \tag{7.1}$$

where $S_i; i = 1, 2$ represent standard deviations of R_1 and R_2 respectively and ρ_{12} is the correlation coefficient between the returns on asset R_1 and asset R_2. Portfolio variance is a statistical value that assesses the degree of dispersion of the returns of a portfolio. It is an important concept in modern investment theory. The calculation of portfolio variance considers not only the riskiness of individual assets but also the correlation between each pair of assets in the portfolio. Thus, the statistical variance analyzes how assets within a portfolio tend to move together. Furthermore, we can calculate the standard deviation of the portfolio using portfolio variance. This represents a measure of volatility: the more a stock's returns vary from the stock's average return, the more volatile the stock. This volatility is often used by investors to measure the risk of a stock or a stock portfolio and it has a large negative impact on investment performance and is one of the major reasons investors' long term returns fall far short of expectations. The exploration of return volatility is a significant subject for investors and decision makers, because it is a matter of great account in evaluating risks, modeling market dynamics and enabling portfolios to be optimized. Many authors have analyzed return and volatility of portfolios and have evaluated relationship between return and risk (see for example Robert T. and Rossi (2006), Iyiola et al. (2012) and recently Aliu et al. (2017)).

In this financial application, we will use two monthly closing adjusted prices of stock returns of two French car manufacturers: Peugeot and Renault from 01/01/2010 to 01/01/2017. We may assume that the closing prices of the two stocks are dependent since they are in same French car industries. The data set can be downloaded from the website https://finance.yahoo.com, among othors. By using MSExcel, we can calculate the empirical correlation coefficients between the returns on the two assets and we remark that the majority of correlation coefficients are between 0.9 and 1.0 which indicate that returns can be considered very highly correlated. Then, we can assume that $\rho_{12} = 1$, which would mean that returns are perfectly positively correlated, the righthand side of Eq. (7.1) is a perfect square and simplifies to

$$\begin{aligned} S^2 &= \alpha^2 S_1^2 + \beta^2 S_2^2 + 2\alpha\beta S_1 S_2 \\ &= (\alpha S_1 + \beta S_2)^2 , \end{aligned}$$

or $S = \alpha S_1 + \beta S_2$. A correlation coefficient of 1 means that there is a perfect positive linear association between the two variables. This means that if one metric increases, the other metric increases too. The vice versa is true: the decrease of one metric signifies the decrease for another one too. Therefore, the portfolio standard deviation is a weighted average of the component security standard deviations. In this circumstance, there are no gains to diversification; it is a technique of allocating portfolio resources or capital to a mix of different investments. The ultimate goal of diversification is to reduce the volatility of the portfolio by offsetting losses in one asset class with gains in another asset class (For more details see for example chapter 4 of Szylar (2013)).

Moreover, the histogram of volatility of returns of Peugeot, S_1 and Renault, S_2 is sketched in Figure 7. Next, by using the maximum likelihood method, and the Fisher scoring method mentioned in the first application, we are able to give the estimates of the unknown parameters of Eq. (1.1). We obtain $\hat{\theta}_1 = 2.3431$, $\hat{\theta}_2 = 2.0841$ and $\hat{\theta}_3 = 17.1626$. The data for S are given for convenience as:

0.14842	0.11493	0.14469	0.11354	0.09352	0.14906	0.13447	0.06413
0.04031	0.03462	0.08517	0.08827	0.09176	0.10045	0.06482	0.07953
0.10036	0.09667	0.17629	0.22420	0.15378	0.12761	0.12472	0.11291
0.12931	0.08663	0.06058	0.08519	0.09509	0.05364	0.03618	0.08679
0.08927	0.03288	0.01694	0.04461	0.05523	0.04844	0.12908	0.17636
0.17034	0.14496	0.08796	0.04159	0.07175	0.12149	0.10649	0.06915
0.06965	0.01018	0.00536	0.01821	0.01892	0.00652	0.03526	0.04510
0.08070	0.12121	0.11494	0.14993	0.14834	0.08084	0.05282	0.05606
0.05634	0.01206	0.23799	0.26312	0.13377	0.09658	0.13540	0.12099
0.04196	0.04046	0.08433	0.17450	0.18524	0.10762	0.05179	0.08238
0.12136	0.09849						

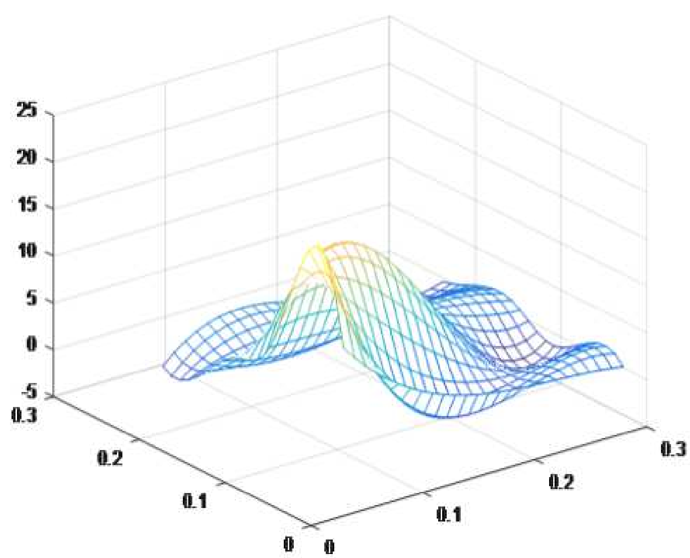

FIGURE 7. Histogram of volatility of returns of two stocks

Further, we also fitted the distribution of S to this data set. For computational convenience, as the data fits above also suggest, we assume $\theta_1 = \theta_2$ fitting. We then use again hypergeom function in Matlab to find the maximum likelihood estimates of the parameters of the exact distribution of S (involving the Gauss hypergeometric function), given by Eq. (2.1). The maximum likelihood estimates are $\hat{\theta}_1 = \hat{\theta}_2 = 2.2978$ and $\hat{\theta}_3 = 19.7721$. The histogram of the volatility of returns and the fit of the distribution of S is sketched in Figure 8 showing a good fit. In order to verify the goodness of fitting, we have performed a Kolmogorov–Smirnov (KS) test. The probability value (p-value) is 0.2883 at the default 5% significance level.

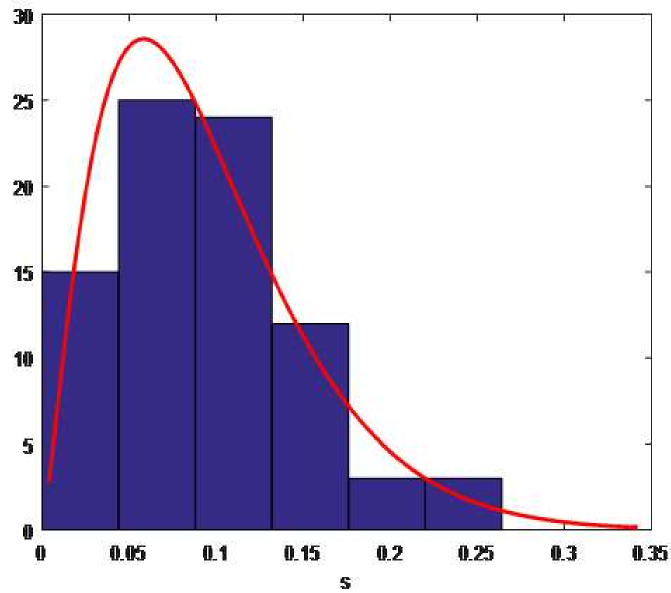

FIGURE 8. Histogram of volatility of returns and superimposed fitted density

8. CONCLUSIONS

In this note, we derived the distribution of the linear combination of two random variables distributed according the Dirichlet distribution. We have derived various properties, including the expressions of the probability density, the ordinary moments, percentiles and entropies. Since the formulas involve the Gauss hypergeometric function, we have proposed an approximation for the distribution of the linear combination based on the standard beta distribution. Finally, two applications of the considered linear combination are used to assess the model and to show its importance. In each application, we analyze two real data sets and the distribution of the linear combination provides an adequate fit to data sets.

Acknowledgement

We express our gratitude to the anonymous referees and to the Editor whose comments significantly improved this manuscript.

References

[1] F. Aliu, D. Pavelkova and B. Dehning, Portfolio risk-return analysis: The case of the automotive industry in the Czech Republic, *J. Inter. Studies*, 10(4)(2017), 72–83.

[2] J. Barrera, T. Huillet and C. Paroissin, Size-biased permutation of Dirichlet partitions and search-cost distribution, *Probab. Eng. Inform. Sc.* 19(1)(2005), 83–97.

[3] U. Cerruti, Counting the number of solutions of congruences, *Appli. Fibon. Numb.* 5(1993), 85–101.

[4] K. Chen and H. Chi, A method of combining multiple probabilistic classifiers through soft competition on different feature sets, *Neurocomput.* 20(1998), 227–252.

[5] J. Danielsson(2011), *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab.* John Wiley, Hoboken.

[6] P. Das Gupta, Two approximations for the distribution of double non-central beta, *Sankhyā,* 30(1968), 83–88.

[7] A. J. Dobson, K. Kuulasmaa, E. Eberle and J. Scherer, Confidence intervals for weighted sums of Poisson parameters, *Stat. Med.* 10(1991), 457–462.

[8] J. Fabius, Two characterizations of the Dirichlet distributions. *Ann. Statist.* 3(1973), 583–587.

[9] D.-Y. Fan, The distribution of the product of independent beta variables, *Comm. Stat.: Theo. Meth.* 20(1991), 4043–4052.

[10] M. Ghorbel, On the product of the bivariate beta components, *Jpn. J. Stat. Data Sci.* 2(2019), 71–80.

[11] A. K. Gupta and S. Nadarajah, Exact and approximate distributions for the linear combination of inverted Dirichlet components, *J. Japan Statist. Soc.* 36(2006), 225–236.

[12] O. Iyiola, Y. Munirat and C. I. Nwufo, The modern portfolio theory as an investment decision tool, *J. Account. Taxation,* 4(2)(2012), 19–28.

[13] B. Johannesson and N. Giri, On approximations involving the beta distribution, *Comm. Stat.: Simul. Comp.* 24(1995), 489–503.

[14] N. L. Johnson, An approximation to the multinomial distribution: some properties and applications, *Biometrika*, 47(1960), 93–102.

[15] B. Kamgar-Parsi, B. Kamgar-Parsi and M. Brosh, Distribution and moments of weighted sum of uniform random variables with applications in reducing Monte Carlo simulations, *J. Statist. Comp. Simul.* 52(1995), 399–414.

[16] S. Kotz, N. Balakrishnan and N. L. Johnson, *Continuous multivariate distributions (volume 1): Models and applications* (second edition). John Wiley and Sons, New York (2000).

[17] T. L. Lai, Control charts based on weighted sums, *Ann. Statist.* 2(1974), 134–147.

[18] S. Nadarajah and S. Kotz, Exact and approximate distributions for the product of Dirichlet components, *Kybernetika,* 40(6)(2004), 735–744.

[19] A. Narayanan, Algorithm AS 266: Maximum likelihood estimation of the parameters of the Dirichlet distribution, *J. R. Stat. Soc.: Ser. C. (Appl. Stat.)* 40(2)(1991), 365–374.

[20] A. Narayanan, A note on parameter estimation in the multivariate beta distribution, *Computers Math. Applic.* 24(10)(1992), 11–17.

[21] T. Pham-Gia and N. Turkkan, Bayesian analysis of the difference of two proportions, *Comm. Stat.: Theo. Meth.* 22(1993), 1755–1771.

[22] P. C. B. Phillips, The characteristic function of the Dirichlet and multivariate F distributions, *Cow. Found. Resea. Econ.* Discussion paper, 865(1988), 1–17.

[23] S. B. Provost and Y.-H. Cheong, On the distribution of linear combinations of the components of a Dirichlet random vector, *Can. J. Stat.* 28(2)(2000), 417–425.

[24] A. P. Prudnikov, Y. A. Brychkov and O. I. Marichev(1986), *Integrals and series (volume 1).* Gordon and Breach Science Publishers, Amsterdam.

[25] A. Rényi, On measures of entropy and information, *Proced. 4th Berkeley Sympos. Math. Statist. Prob.* University of California Press, Berkeley, California (1961), 547–561.

[26] T. D. Robert and L. Rossi, A portfolio of stocks and volatility, *J. Investing*, 15(2)(2006), 99–106.

[27] G. Ronning, Maximum likelihood estimation of Dirichlet distribution, *J. Statist. Comput. Simul.* 32(1989), 215–221.

[28] C. E. Shannon, A mathematical theory of communication, *Bell Sys. Techn. J.* 27(1948), 379–423, 623–656.

[29] D. Sculli and K. L. Wong, The maximum and sum of two beta variables in the analysis of PERT networks, *Omega-Int. J. Manage. S.* 13(1985), 233–240.

[30] C. Szylar(2013), *Handbook of Market Risk,* John Wiley and Sons.

[31] V. Witkovsky, Computing the distribution of a linear combination of inverted gamma variables, *Kybernetika,* 37(1)(2001), 79–90.

laboratory leBiomat, university of Sfax, Faculty of Sciences of Sfax, B P 802, 3018, Sfax, TUNISIA.

Email address: mohamedali.ghorbel@fss.usf.tn, ghorbel.mmed.ali@gmail.com