REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

# Randomly Weighted Averages on Multivariate Dirichlet Distributions with Generalized Parameters

Authors: Hazhir Homei ✉
– Department of Mathematics, Statistics, and Computer Science, University of Tabriz, Tabriz, Iran (homei@tabrizu.ac.ir)

Saralees Nadarajah
– Department of Mathematics, Howard University, Washington DC 20059, USA (saraleesan.nadarajah@howard.edu)

Ameneh Taherkhani
– Department of Mathematics, Statistics, and Computer Science, University of Tabriz, Tabriz, Iran (a.taherkhani@tabrizu.ac.ir)

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- We study the distributional properties of the product of random stochastic matrices by using the Dirichlet distribution. Our observations widely generalize some known results on the randomly weighted averages about multivariate Dirichlet distributions. We present a new method for approximating the distribution that can be used for the multivariate random variables with a slight change.

Keywords:

- *Cauchy Composition Test; Random Stochastic Matrices; Real Lifetime.*

AMS Subject Classification:

- 65Cxx, 62E15.

✉ Corresponding author

## 1. INTRODUCTION

The product of random stochastic matrices is one of the concepts that has attracted the attention of many mathematicians [15]. The behavior of this concept in science and engineering has been investigated under some specific assumptions. The studies and applications of this concept are closely related to the studies of averaging dynamics. Let us note that some applications are presented by Touri [15].

Here, we consider the independent random matrices each of which has independent rows and are identically distributed with the Dirichlet distribution, and investigate some distributional and statistical properties of the product of those random matrices. For this purpose, we have addressed the mixture random variables and followed their traces in applied fields. Here we list some important applications:

**Cauchy Composition Test** Let $p_i$ be the individual $p$-value for $i = 1, 2, \ldots, d$. We define the Cauchy combination test statistic as

$$T = \sum_{i=1}^{d} \omega_i \tan\{(0.5-p_i)\pi\}, \tag{1.1}$$

where the weights $\omega_i$'s are nonnegative and $\sum_{i=1}^{d} \omega_i = 1$. Given that $p_i$ is uniformly distributed between 0 and 1 under the null, the component $\tan\{(0.5-p_i)\pi\}$ follows a standard Cauchy distribution. To overcome these challenges, Liu and Xie [11] propose a new test that takes advantage of the Cauchy distribution. Their test statistic has a simple form and is defined as a weighted sum of Cauchy transformation of individual $p$-values. We emphasize that some of the results have been examined in a state where $W_i$ are random variables. Here, we define this statistic as a vector and obtain the properties of the statistic under some specific assumptions. Of course, some of the results obtained in our study indicate that these properties also apply to other distributions in addition to the Cauchy distribution. For the notation and discussions we refer the reader to Liu and Xie [11].

**Real Lifetime** Suppose that $X_1, \ldots, X_r$ are random variables, and the coefficient of the impact of environment on lifetime, $Y_1, \ldots, Y_r$, are independent random variables with Gamma distribution that indicate the lifetime in the laboratory conditions. The following random variable is called real lifetime: $T = \sum_{i=1}^{r} Y_i X_i$. In this paper, the random variable $T$ presented in the research by Homei and Nadarajah [9] has been generalized to vector random variables. In this paper, some results and characterizations have been studied in vector random variables for $T$; we also answer some questions asked there (in particular, we have generalized [9, Theorem 2.2] to the multivariate case in Theorem 3.4 below).

**Solving Some Differential Equations** Using the properties of the Beta distribution, Homei [8] and Hadad *et al.* [2] solved some differential equations. Indeed, the given equations may be solved by very long calculations. Let us recall that Theorem 1 of Homei [4] identifies the distribution of (the 1-dimensional) $Z$ from the distributions of $X_i$'s by means of the following differential equation:

$$\frac{(-1)^{n^*-1}d^{n^*-1}}{(n^*-1)!dz^{n^*-1}}S(F_Z,z)=\prod_{i=1}^{n}\frac{(-1)^{m_i-1}}{(m_i-1)!}\frac{d^{m_i-1}}{dz^{m_i-1}}S(F_{X_i},z) \tag{1.2}$$

where $F_Z$ denote the cumulative distribution function of a random variable Y and $S(F_Z,z)$ is Stieltjes transform defined by

$$S(H,z)=\int\frac{1}{z-x}H(dx),\quad z\in\mathbf{C}\cap(\mathsf{supp}H)^c. \tag{1.3}$$

Here, $\mathsf{supp}H$ is the support of $H$; see [4].

**Random Convex Combination** A stochastic linear combination

$$\hat{C}_1\cdot Z_1+\hat{C}_2\cdot Z_2+\cdots+\hat{C}_m\cdot Z_m \tag{1.4}$$

of random variables $Z_1,\ldots,Z_m$ where $\hat{C}_i, 1\le i\le m$, are random variables such that

$(i)$ $\hat{C}_i\ge 0,\ 1\le i\le m,$ and

$(ii)$ $\sum_{i=1}^m\hat{C}_i=1, a.s.,$

is called a random convex combination of the random variables $Z_1,\ldots,Z_m$ (for more details see [4]).

Of course, another form of real lifetime is provided by Homei [4], which is not far from the statistic defined by Liu and Xie [11]. Let $Z_i, i=1,\ldots,n$, be the lifetime measured in a laboratory and $0\le C_i\le 1$ be the random effect of the environment on it, so $C_iZ_i\le Z_i$ and thus $\sum_{i=1}^n C_iZ_i$ is the average lifetime in the environment, see Homei [8], Homei and Nadarajah [9]. If $Y_i$ is the real lifetime in the $i$th area, $C_i=\dfrac{Y_i}{\sum_{i=1}^n Y_i}$ is the random effect ratio in the $i$th area. Therefore, it is clear that a good choice for the distribution of $\mathbf{C}=\langle C_1,\ldots,C_n\rangle$ can be Dirichlet distribution. It is important that the product of random stochastic matrices connect us directly to stochastic linear combination.

The structure of the paper is as follows: the next subsection gives the motivation of the research in this paper, and lists some innovations. In Section 2, the mean, variance and moments of the randomly weighted averages on random vectors with Dirichlet distributions are obtained. In Section 3, a new method for calculating the distribution of randomly weighted averages are presented, and the distribution of this random vector is calculated under some specific assumptions. In Section 4 using simulation we suggest an approximation for the distribution of randomly weighted averages.

## 1.1. Motivation and Innovation

For obtaining the distribution of the randomly weighted averages, one suggested method is using Stieltjes transforms, which is a very complex strategy. In this paper, a novel method for obtaining those distributions in the multivariate case is presented. For evaluating this new method we have reproved Theorem 3.4, for which we have also proved some new auxiliary theorems. In spite of the fact that some special cases of those auxiliary results are well known theorems, but these results in the general form that are presented here seem to be new, with much simpler and more elementary proofs. For keeping the copyright of the results, some preliminary versions had been put on the arXiv ([5]), and re-emphasized in [9] that we would generalize the earlier results in the future. At the end we also presented some of them in a national conference [7].

### 1.1.1. Some History and Earlier Results

Two ideas of [14], i.e., (i) defining the random coefficients, and (ii) using Stieltjes transformations for obtaining the distribution of the randomly weighted averages, have been taken from [13]. In fact, the class of randomly weighted averages defined in those two papers are much more restricted in comparison to the ones defined in this paper, since their defined coefficients have Dirichlet distributions with limited parameters (which are positive integer numbers). But in this paper, the random vectors are taken to have Dirichlet distributions without any limitations on their parameters; notice that the assumption $\sum_i \alpha_i = 1$ in [14, Theorem 3.1] shows that the results of [14] (both Theorems 2 and 5) are weaker than ours here.

To make a long story short, the main result of [14], which is their Theorem 2.1, is not useful for a large part of the class of randomly weighted averages defined in the present paper; and their Theorem 3.1 is a very special case of our Theorem 3.4 below, which was published before in [5]. Therefore, our method is much stronger, and even more elementary at the same time; indeed, Theorem 2.1 in [14] should be improved for being usable in our class of randomly weighted averages. A complete and general proof for that theorem is in the possession of the first author, and is planned for a publication in future. For further references, we invite the readers to consult [6] and [12, p. 57].

## 2. Product Moments of Random Convex Combination

The concept of product of random stochastic matrices motivated us to discuss the distributional properties of random convex combinations. These properties include the product moments, and the mean and the variance of components. Throughout the paper, $\langle W_1, \ldots, W_r \rangle$ is called random coefficient vector of environmental effect in $r$-position. As mentioned in the introduction, the rows are independent and have Dirichlet distribution in random stochastic matrices.

**Theorem 2.1.** *Suppose that the independent random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ have identical distributions with mean $\mu$ and variance $\boldsymbol{S}$ and that the random vector $W = \langle W_1, \ldots, W_r \rangle$ is independent from $\mathbf{X}_1, \ldots, \boldsymbol{X}_r$ such that $\sum_{j=1}^{r} W_j = 1$, a.s. Then the mean and the variance of $\boldsymbol{Z} = \sum_{j=1}^{r} W_j \boldsymbol{X}_j$ are*

$$E(\boldsymbol{Z}) = \mu \quad \textit{and} \quad Var(\boldsymbol{Z}) = \sum_{j=1}^{r} EW_j^2 \mathbf{S},$$

*where $\boldsymbol{S}$ is variance-covariance matrix.*

**Proof:** By using the double conditional expectation and the conditional variance, the result is proved. □

The following theorem results in product moments of random convex combination when we consider the random vectors by Dirichlet distribution.

**Theorem 2.2.** *Suppose that the independent random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ have respectively,*

$$Dirichlet(n_{11,}, \ldots, n_{1k}), \ldots, Dirichlet(n_{r1}, \ldots, n_{rk})$$

*distributions and that the random vector $W = \langle W_1, \ldots, W_r \rangle$ is independent from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ and has $Dirichlet(\alpha_1, \ldots, \alpha_r)$ distribution.*

*Then the product moments in $(s_1, \ldots, s_k)$ of $\boldsymbol{Z} = \sum_{j=1}^{r} W_j \boldsymbol{X}_j$ are*

$$E(L_1^{s_1} L_2^{s_2} \cdots L_k^{s_k}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + h)} \sum_{h_1} \cdots \sum_{h_k} \left( \prod_{j=1}^{k} \binom{s_j}{h_{1j}, \ldots, h_{rj}} \times \prod_{i=1}^{r} \frac{\Gamma(\alpha_i + h_i)}{\Gamma(\alpha_i)} \right.$$

$$\left. \frac{\Gamma(n_i)}{\Gamma(n_i + h_i)} \prod_{i=1}^{r} \prod_{j=1}^{k} \frac{\Gamma(n_{ij} + h_{ij})}{\Gamma(n_{ij})} \right),$$

*where $L_j$'s are components of vector $Z$, $\sum_{i=1}^{r} h_i = h$ and $\sum_{i=1}^{r} \alpha_i = \alpha$.*

**Proof:** We find the general moments $(s_1, s_2, \ldots, s_k)$ of $Z$ as follows

$$E(L_1^{s_1} L_2^{s_2} \cdots L_k^{s_k}) = E\left( \prod_{j=1}^{k} (\sum_{i=1}^{r} W_i \mathbf{X}_{ij})^{s_j} \right)$$

$$= E\Big(\prod_{j=1}^{k}(\sum_{h_j}\binom{s_j}{h_{1j}, h_{2j}, \ldots, h_{nj}} \times \prod_{i=1}^{r}(W_i X_{ij})^{h_{ij}})\Big),$$

where the expression $\sum h_j$ denotes the summation over all the nonnegative integers $h_j = (h_{1j}, h_{2j}, \ldots, h_{rj})$ subject to

$$\sum_{i=1}^{r} h_{ij} = s_j, \quad (j = 1, 2, \ldots, k).$$

This can be rearranged as

$$E\Big(\sum_{h_1}\sum_{h_2}\cdots\sum_{h_k}(\prod_{j=1}^{k}\binom{s_j}{h_{1j}, h_{2j}, \ldots, h_{rj}}\prod_{j=1}^{k}\prod_{i=1}^{r}(W_i\mathbf{X}_{ij})^{h_{ij}})\Big) =$$

$$E\Big(\sum_{h_1}\sum_{h_2}\cdots\sum_{h_k}(\prod_{j=1}^{k}\binom{s_j}{h_{1j}, h_{2j}, \ldots, h_{rj}}\prod_{i=1}^{r}W_i^{h_{i.}}\prod_{j=1}^{k}\prod_{i=1}^{r}X_{ij}^{h_{ij}})\Big),$$

where $h_{i.} = \sum_{j=1}^{k} h_{ij}$ and we have this equal to

$$\sum_{h_1}\cdots\sum_{h_k}\Bigg(\prod_{j=1}^{k}\binom{s_j}{h_{1j}, h_{2j}, \ldots, h_{rj}}E(\prod_{i=1}^{r}W_i^{h_{i.}})E(\prod_{j=1}^{k}\prod_{i=1}^{r}X_{ij}^{h_{ij}})\Bigg), \tag{2.1}$$

now we find two expectations in equation (2.1):

$$E(\prod_{i=1}^{r}W_i^{h_{i.}}) = \frac{\Gamma(\sum_{i=1}^{r}\alpha_i)}{\Gamma(\sum_{i=1}^{r}(\alpha_i + h_{i.}))} \times \prod_{i=1}^{r}\frac{\Gamma(\alpha_i + h_{i.})}{\Gamma(\alpha_i)}.$$

By using the Dirichlet distribution, we have

$$E(\prod_{i=1}^{r}W_i^{h_{i.}}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + h)} \times \prod_{i=1}^{r}\frac{\Gamma(\alpha_i + h_{i.})}{\Gamma(\alpha_i)}. \tag{2.2}$$

Also, we have

$$\begin{aligned} E(\prod_{j=1}^{k}\prod_{i=1}^{r}X_{ij}^{h_{ij}}) &= \prod_{i=1}^{r}E(\prod_{j=1}^{k}X_{ij}^{h_{ij}}) \\ &= \prod_{i=1}^{r}\Big(\frac{\Gamma(\sum_{j=1}^{k}n_{ij})}{\Gamma(\sum_{j=1}^{k}(n_{ij} + h_{ij})} \times \prod_{j=1}^{k}\frac{\Gamma(n_{ij} + h_{ij})}{\Gamma(n_{ij})}\Big), \end{aligned}$$

now we have $\sum_{j=1}^{k} n_{ij} = n_{i.}$ and $\sum_{j=1}^{k} h_{ij} = h_{i.}$, so the above is equal to

$$\prod_{i=1}^{r}\left(\frac{\Gamma(n_{i.})}{\Gamma(n_{i.}+h_{i.})}\times\prod_{j=1}^{k}\frac{\Gamma(n_{ij}+h_{ij})}{\Gamma(n_{ij})}\right) \tag{2.3}$$

and by using the Dirichlet distribution we have

$$E(\prod_{j=1}^{k} X_{ij}^{h_{ij}}) = \frac{\Gamma(\sum_{j=1}^{k}\alpha_j^{(i)})}{\Gamma(\sum_{j=1}^{k}\alpha_j^{(i)}+h_{i.})}\prod_{j=1}^{k}\frac{\Gamma(\alpha_j^{(i)}+h_{ij})}{\Gamma(\alpha_j^{(i)})}.$$

So, by using (2.2) and (2.3) in (2.1) the above is equal to

$$\sum_{h_1}\cdots\sum_{h_k}\left(\prod_{j=1}^{k}\binom{s_j}{h_{1j},h_{2j},\ldots,h_{rj}}\frac{\Gamma(\alpha)}{\Gamma(\alpha+h)}\prod_{i=1}^{r}\frac{\Gamma(\alpha_i+h_{i.})}{\Gamma(\alpha_i)}\right.$$

$$\left.\times\prod_{i=1}^{r}\frac{\Gamma(n_{i.})}{\Gamma(n_{i.}+h_{i.})}\prod_{j=1}^{k}\frac{\Gamma(n_{ij}+h_{ij})}{\Gamma(n_{i)})}\right)$$

$$=\frac{\Gamma(\alpha)}{\Gamma(\alpha+h)}\sum_{h_1}\cdots\sum_{h_k}\prod_{j=1}^{k}\binom{s_j}{h_{1j},\ldots,h_{rj}}$$

$$\times\prod_{i=1}^{r}\frac{\Gamma(\alpha_i+h_{i\cdot})}{\Gamma(\alpha_i)}\frac{\Gamma(n_{i.})}{\Gamma(n_{i.}+h_i)}\prod_{i=1}^{r}\prod_{j=1}^{k}\frac{\Gamma(n_{ij}+h_{ij})}{\Gamma(n_{ij})}.$$

Therefore the proof of the product moments on $(s_1,\ldots,s_k)$ is complete. $\square$

The moments of $Z$ on $(s_1,s_2,s_3)$ are given in Table 1.

## 3. Some Characterizations

In this section, some characterizations of random stochastic linear combinations (real lifetime, random convex combination) in Dirichlet random vectors are introduced.

| $(n_{11}, n_{12}, n_{13})$ | $(n_{21}, n_{22}, n_{23})$ | $(n_{31}, n_{32}, n_{33})$ | $(\alpha_1, \alpha_2, \alpha_3)$ | $(s_1, s_2, s_3)$ | $E(Z)$ |
|---|---|---|---|---|---|
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | 0.001851852 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (1,1,2) | 0.002469136 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (1,3,1) | 0.003086420 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (1,2,3) | 0.004938272 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (2,4,5) | 0.025925926 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (2,2,2) | 0.006172840 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (2,2,1) | 0.003703704 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (2,2,3) | 0.008641975 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (3,3,3) | 0.017901235 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (3,3,1) | 0.006790123 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (3,3,5) | 0.029012346 |
| (1,1,1) | (1,1,1) | (1,1,1) | (1,1,1) | (4,3,5) | 0.038271605 |

**Theorem 3.1.** *Let the $k$-variate random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ be independent with common distributions, and let $Y_1, \ldots, Y_r$ be independent with Gamma$(k\alpha, \frac{1}{\mu})$ distributions, and independent from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$. Then the components of $\boldsymbol{T} = \sum_{i=1}^{r} Y_i \mathbf{X}_i$ have independent Gamma$(r\alpha, \frac{1}{\mu})$ distributions if and only if $\mathbf{X}_i$ have Dirichlet$(\alpha, \ldots, \alpha)$ distributions $(i = 1, \ldots, n)$.*

**Proof:** First we find the moment generating function of T

$$\begin{aligned} E(e^{t^{'}T}) &= E(e^{t^{'}\sum_{i=1}^{r} Y_i \mathbf{X}_i}) \\ &= \prod_{i=1}^{r} E(e^{t^{'}Y_i \mathbf{X}_i}) \\ &= \prod_{i=1}^{r} E(E(e^{t^{'}Y_i \mathbf{X}_i} | \mathbf{X}_i)) \\ &= \prod_{i=1}^{r} E((\frac{1}{1 - t^{'}\mathbf{X}_i})^{\alpha}) \\ &= E^{r}(\frac{1}{1 - t^{'}\mathbf{X}_i})^{\alpha} \end{aligned}$$

The second side of equation is well-known Stieltjes transformation; for more application and properties of this transformation see [3], [4], [5], [7], and [1].

The last statement is the Stieltjes transformation which is unique, by which both of the if part and the only if part can be easily proved. □

The following theorem is a generalization Theorem 1 of Yeo and Milne [16] that leads us to the next theorems.

**Theorem 3.2.** *Suppose that* $\mathbf{U}$ *and* $V$ *are independent (absolutely continuous) nonnegative random variables, respectively, such that* $\mathbf{U}$ *has bounded support and* $\mathbf{Z} = \mathbf{U}V$. *Then for arbitrary positive* $\alpha_i; i = 1, \ldots, k$ *and any two of the following three conditions imply the third.*
*(i)* $\mathbf{Z} \sim < Gamma_1(\alpha_1, \frac{1}{\mu}), \ldots, Gamma_k(\alpha_k, \frac{1}{\mu}) >$ *where* $Gamma(\alpha_i, \frac{1}{\mu})$ *are independent;*
*(ii)* $\mathbf{U} \sim Dirichlet(\alpha_1, \ldots, \alpha_k)$*;*
*(iii)* $V \sim Gamma(\alpha^+, \frac{1}{\mu}), \quad \alpha^+ = \sum \alpha_i$.

**Proof:** For proving $(i), (ii) \Rightarrow (iii)$, it suffices to note that the sum of the components of $\mathbf{Z}$ and $\mathbf{U}V$ are identically distributed; i.e. $V \stackrel{\mathrm{d}}{=} Gamma(\alpha^+, \frac{1}{\mu})$, since random vectors $\mathbf{Z}$ and $\mathbf{U}V$ have identical distribution.

Now we prove the implication $(i), (iii) \Rightarrow (ii)$. By using the general moments of $(s_{11}, \ldots, s_{kk})$ we have

$$E(Z_{11}^{s_{11}} Z_{12}^{s_{12}} \cdots Z_{kk}^{s_{kk}}) = E(V^{\sum_{i=1}^{k} s_i}) \cdot E((U_{11})^{s_{11}} (U_{12})^{s_{12}} \cdots (U_{kk})^{s_{kk}}).$$

By substituting the gamma moments it can be shown that $\mathbf{U}$ has Dirichlet distribution.

Finally, assume that $(ii)$ *and* $(iii)$ are satisfied, then we can obtain the distribution of $\mathbf{U}V$ by using the transformation method of random variables (or change of variables),

$$f(z_1, \ldots, z_r) = \frac{1}{\prod_{i=1}^{r} \Gamma(\alpha_i) \mu^{(\sum_{i=1}^{r} \alpha_i)}} e^{-\frac{\sum_{i=1}^{r} z_i}{\mu}} \prod_{i=1}^{r} z_i^{\alpha_i - 1}.$$

So, the proof is complete. □

**Remark 3.1.** Throughout this paper we set $\alpha^+ = \sum \alpha_i$.

**Theorem 3.3.** *Let* $\boldsymbol{X}$ *be any random vector with bounded support and* $Y$ *be independent random variable of* $\boldsymbol{X}$ *with* $Gamma(\sum_{j=1}^{r} \alpha_j, \frac{1}{\mu})$ *distribution. If*

$$\sum_{i=1}^{r} Y_i \boldsymbol{X}_i \stackrel{\mathrm{d}}{=} Y \boldsymbol{X}, \tag{3.1}$$

*where* $Y_i (i = 1, \ldots, r)$ *are independent random variables with* $Gamma(\alpha_i, \frac{1}{\mu})$ *distribution, then* $\boldsymbol{X}$ *and the randomly linear combination* $\boldsymbol{Z} = \sum_{i=1}^{r} W_i \boldsymbol{X}_i$ *have identical distribution, where the random vector* $\mathbf{W} = \langle W_1, \ldots, W_r \rangle$ *is independent from* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ *and has* $Dirichlet(\alpha_1, \ldots, \alpha_k)$ *distribution.*

**Proof:** First we define $Y^+ = \sum_{i=1}^r Y_i$, which has $Gamma(\alpha^+, \frac{1}{\mu})$ distribution, then by using (3.1) we have

$$Y^+ \cdot \frac{\sum_{i=1}^r Y_i \mathbf{X}_i}{Y^+} \stackrel{\mathrm{d}}{=} Y\mathbf{X},$$

the fraction $\frac{\sum_{i=1}^r Y_i \mathbf{X}_i}{Y^+}$ has the same distribution as $\mathbf{Z}$, so we can rewrite the above expression in the form of

$$Y^+\mathbf{Z} \stackrel{\mathrm{d}}{=} Y\mathbf{X}. \tag{3.2}$$

The random vectors $Y^+\mathbf{Z}$ and $Y\mathbf{X}$ in (3.2) have the same moments, so

$$E((Y^+)Z_1{}^{k_1} \cdot (Y^+)^{k_2} Z_2^{k_2} \cdots (Y^+)^{k_r} Z_r^{k_r}) = E(Y^{k_1} X_1^{k_1} \cdot Y^{k_2} X_2^{k_2} \cdots Y^{k_r} X_r^{k_r}),$$

and we have

$$E((Y^+)^{k^+})E(Z_1^{k_1} Z_2^{k_2} \cdots (Z_r^{k_r}) = E(Y^{k^+}(E(X_1^{k_1} \cdot X_2^{k_2} \cdots X_r^{k_r}))),$$

where $k^+ = \sum_{j=1}^r \sum_{i=1}^r k_{ij}$.

Considering the same distribution of $Y^+$ and $Y$, we can omit the first expectations from both sides of the equation

$$E(Z_1^{k_1} Z_2^{k_2} \cdots Z_r^{k_r}) = E(X_1^{k_1} \cdot X_2^{k_2} \cdots X_r^{k_r})$$

as a result of having bounded support variables, the equation of the same moments of two variables conduces to the same distribution, so the proof is completed and $\mathbf{X}$ and $\mathbf{Z}$ have identical distributions. □

The following theorem is a generalization of Theorem 2.2 (Homei and Nadarajah [9]) and we want to provide another perspective to prove Theorem 2.1 of Homei [8].

**Theorem 3.4.** *If $\mathbf{X}_1, \ldots, \mathbf{X}_r$ are independent $k$-variate random vectors with respectively $Dirichlet(n_1^{(1)}), \ldots, Dirichlet(n_k^{(r)})$ distributions, for some $k$-dimensional vectors $n_i^{(j)} = \langle n_1^{(j)}, \ldots, n_k^{(j)} \rangle (j = 1, \ldots, r)$, and the random vector $\boldsymbol{W} = \langle W_1, \ldots, W_r \rangle$ is independent from $\mathbf{X}_1, \ldots, \mathbf{X}_r$ and has the distribution*

$$Dirichlet(\sum_{i=1}^k n_i^{(1)}, \ldots, \sum_{i=1}^k n_i^{(r)}),$$

*then the randomly linear combination $\boldsymbol{Z} = \sum_{i=1}^r W_i \mathbf{X}_i$ has the distribution*

$$Dirichlet(\sum_{j=1}^r n_1^{(j)}, \ldots, \sum_{j=1}^n n_k^{(j)}).$$

**Proof:** Let $Y_j$ $(j = 1, \ldots, r)$ be independent random variables and independent from $\langle \mathbf{X}_1, \ldots, \mathbf{X}_r \rangle$ that have the distribution $Gamma(\sum_{i=1}^{k} n_i^{(j)}, \frac{1}{\mu})$, respectively. It can be seen, by some classic ways (e.g. $E(e^{t'T}) = [\Psi(t)]^{(\sum_j n_j)}$ see Kerov and Tsilevich table 2 [10]), that the distribution of $\mathbf{T} = \sum_j \mathbf{T}_j = \sum_j Y_j \mathbf{X}_j$ is the same as the distribution of $\mathbf{T}_j$ with the parameter $(\sum_j n_j, \ldots, \sum_j n_j)$. We can also write $\mathbf{T} \stackrel{\mathrm{d}}{=} Y\mathbf{X}$ in which $Y$ has the Gamma distribution with the parameter $(\sum_j \sum_{i=1}^{k} n_i^{(j)}, \frac{1}{\mu})$, and $Y$ and $\mathbf{X}$ are independent from each other. By using Theorem 3.3 the proof is complete. $\square$

## 4. Simulation and its Application

We present a new method for approximating the distribution that can be used for the multivariate random variables with a slight change in the approximation of Homei and Nadarajah [9]. The previous method of approximation was for a univariate random variable, but we want to introduce a method that is able to approximate the multivariate random variables. Calculating the distribution of $\mathbf{Z}$ may not be easy; in this section we suggest a distribution which may be close to the real distribution of $\mathbf{Z}$. See [9] for more details.

Let $\mathbf{X}_1$ and $\mathbf{X_2}$ be independent random vectors with Dirichlet$(1, 1, 1)$ distribution, and let $w$ be a random variable independent from them with Beta$(1, 1)$ distribution. Calculating the distribution of

$$\mathbf{Z} = w\mathbf{X}_1 + (1 - w)\mathbf{X_2}$$

could be cumbersome, and it could be that there is no closed form. For estimating the distribution of $\mathbf{Z}$ we suggest the following. Firstly, we generate $\mathbf{X}_1$ and $\mathbf{X_2}$ data by the means of Python software package. Then we compose them in accordance with the definition of $\mathbf{Z}$. We simulate random numbers with size 8000 of $\mathbf{Z}$, and assume that it has the Dirichlet distribution; we estimate their parameters by the method of maximum likelihood estimation. As a result, the values of $\mathbf{Z}$ will be observable. Our suggested approximation has the Dirichlet$(2.7, 2.8, 3)$ distribution and we expect it to be useful as mentioned in [9]. Figure 1 shows the approximated distribution of $\mathbf{Z}$, i.e., Dirichlet$(2.7, 2.8, 3)$.

## 5. Conclusions

In this paper, a novel method for obtaining the distribution of the randomly weighted averages on random vectors is presented, which is simpler and more elementary than the others. Beside that one can obtain the distribution of $T = \sum \mathbf{X}_i Y_i$ by that method, which is left to be done in the future. In case this distribution appears to be complicated, we will approximate it by simulation, and

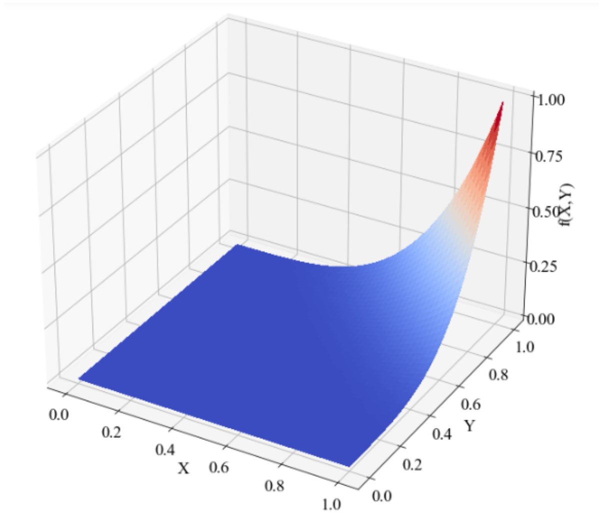


**Figure 1**: The approximate distribution is Dirichlet$(2.7, 2.8, 3)$

we will study some distributional properties of $T$ in general. The four examples illustrated in the Introduction (Cauchy Composition Test, Real Lifetime, Solving Some Differential Equations, and Random Convex Combination) are some visible applications of the research in this paper.

## Acknowledgements

We would like to thank the Referee#2 of the first round for the most useful comments and suggestions, and we also thank the editors and Referee#A of the third round for carefully reading the manuscript and our responses to the referee reports in the previous rounds.

## REFERENCES

[1] Cifarelli, D.M. & Regazzini, E. (1979). A General approach to Bayesian analysis of nonparametric problems. The associative mean values within the framework of the Dirichlet process: II (in Italian). *Rivista di Matematica per le Scienze Economiche e Sociali*, **2**, 95–111.

[2] Hadad H. & Homei H. & Behzadi M. & Farnoosh, R. (2021). Solving some stochastic differential equation using Dirichlet distributions, *Computational Methods for Differential Equations*, **9**(2), 393–398.

[3] Homei, H. (2012). Randomly weighted averages with beta random proportions, *Statistics & Probability Letters*, **82**(8), 1515–1520.

[4] Homei, H. (2015). A novel extension of randomly weighted averages, *Statistical Papers*, **56**(4), 933–946.

[5] Homei, H. (2016). Randomly weighted averages: a multi-variate case, arXiv:1603.00596 [math.ST]. `https://arxiv.org/abs/1603.00596`

[6] Homei, H. (2017a). Characterizations of arcsin and related distributions based on a new generalized unimodality, *Communications in Statistics - Theory and Methods*, **46**(2), 1024–1030.

[7] Homei, H. (2017b). Randomly weighted averages: a multi-variate case, in: *Abstracts of the 11th Seminar on Probability and Stochastic Processes, Qazvin, Iran* (30–31 August 2017), Imam Khomeini International University, page 49.

[8] Homei, H. (2021). The stochastic linear combination of Dirichlet distributions, *Communications in Statistics - Theory and Methods*, **50**(10), 2354–2359.

[9] Homei, H. & Nadarajah, S. (2018). On products and mixed sums of gamma and beta random variables motivated by availability, *Methodology and Computing in Applied Probability*, **20**(2), 799–810.

[10] Kerov, S.V.E. & Tsilevich, N.V. (2004). The Markov-Krein correspondence in several dimensions, *Journal of Mathematical Sciences*, **121**(3), 2345–2359.

[11] Liu, Y. & Xie, J. (2020). Cauchy combination test: a powerful test with analytic $p$-value calculation under arbitrary dependency structure, *Journal of the American Statistical Association*, **115**(529), 393–402.

[12] Pitman, J. (2018). Random Weighted Averages, Partition Structures and Generalized Arcsine Laws, arXiv 1804.07896, to appear in *Probability Surveys*.

[13] Soltani, A.R. & Homei H. (2009). Weighted averages with random proportions that are jointly uniformly distributed over the unit simplex, *Statistics & Probability Letters*, **79**(9), 1215–1218.

[14] Soltani, A.R. & Roozegar R. (2020). Averages for multivariate random vectors with random weights, *REVSTAT–Statistical Journal*, **18**(4), 453–460.

[15] Touri, B. (2012). *Product of Random Stochastic Matrices and Distributed Averaging*, Springer-Verlag, Berlin.

[16] Yeo, G.F. & Milne, R.K. (1989). On characterizations of exponential distributions, *Statistics & Probability Letters*, **7**(4), 303–305.