Available online at http://sanad.iau.ir/journal/ijim/
Int. J. Industrial Mathematics (ISSN 2008-5621)
Vol.16, No. 3, 2024 Article ID IJIM-1629, 10 pages
Research Article

Science and Research Branch (IAU)

The Distribution of Randomly Weighted Averages on Random Independent Vectors with Dirichlet Distributions

H. Hadad[1], H. Homei[2 *], M.H Behzadi[1], R. Farnoosh[3]
[1] Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran.
[2] Department of Statistics, University of Tabriz, P.O.Box 51666-16471, Tabriz, Iran.
[3] School of Mathematics, Iran University of Science and Technology, Tehran, Iran.

Submission Date: 2021/04/18, Revised Date: 2023/01/20, Date of Acceptance: 2024/08/05

Abstract
Randomly weighted averages of vectors with Dirichlet distributions will have the Dirichlet distribution when some certain conditions are satisfied. This has been recently proved by some statistical methods. In this paper, we present two simpler and shorter proofs for this fact, one by using Basu's theorem and one by mathematical induction.

Keywords: Randomly Weighted Averages; Basu's Theorem; Mathematical Induction.

* Corresponding author: Email: homei@tabrizu.ac.ir

1. Introduction

Randomly weighted averages are useful tools in statistics and probability theory (see e.g. [1]). Here is the formal definition: For arbitrary random variables $X_1,\ldots,X_n$ and an arbitrary random vector $W=(W_1,\ldots,W_n)$ that satisfies $\sum_i w_i=1$ with probability one, the randomly weighted average of X_i's with the weights W_i's is

$$z=\sum_{i=1}^{n} W_i X_i \, .$$

([2]) generalized the randomly weighted average for vectors, where its distribution is obtained by applying complex and lengthy calculations. The method of obtaining the distribution in their work is based on the use of the moment's method. For more detail see ([3, 4, 5, 6, 2]). In this short paper, we intend to present two different methods for calculating this distribution. The presented novel evidence can provide a better understanding of the Dirichlet distribution.

2. Some Alternatives Proofs

A proof of the following theorem is provided by Homei in 2019 ([1]) and a relatively hard and long proof was also suggested by the authors in [2]. Here, we want to present other proofs that may make the reader's view of the issue clearer.

Theorem 2.1. Let $X_1,\ldots,X_n$ be independent vectors with $Dirichlet(\alpha_1^{(1)},\ldots,\alpha_k^{(1)}),\ldots,$ $Dirichlet(\alpha_1^{(n)},\ldots,\alpha_k^{(n)})$ distributions, and the random vector $W=\langle W_1,\ldots,W_n\rangle$ be independent from $X_1,\ldots,X_n$ and have the following distribution:

$$Dirichlet(\sum_{i=1}^{k}\alpha_i^{(1)},\ldots,\sum_{i=1}^{k}\alpha_i^{(n)}) \, .$$

Then the random mixture $Z=\sum_{i=1}^{n} W_i X_i$ has the following distribution:

$$Dirichlet(\sum_{j=1}^{n}\alpha_1^{(j)},\ldots,\sum_{j=1}^{n}\alpha_k^{(j)}) \, .$$

2.1. Applying Basu's Theorem

Here, an alternative method is presented for obtaining a randomly weighted average distribution that uses Basu's theorem. It is worth nothing that this method is a technical one that cannot be considered a common way to solve such problems.

The Fist Method. Since X_j's has Dirichlet distributions, then we let

$$X_j \sim \left(\frac{\Gamma_{1j}(\alpha_1^{(j)})}{\sum_{i=1}^{k}\Gamma_{ij}(\alpha_i^{(j)})}, \dots, \frac{\Gamma_{kj}(\alpha_k^{(j)})}{\sum_{i=1}^{k}\Gamma_{ij}(\alpha_i^{(j)})} \right)$$

$(for 1 \le j \le n)$, and thus, we define W by

$$W \sim \left(\frac{\sum_{i=1}^{k}\Gamma_{ij}(\alpha_i^{(1)})}{\sum_{j}^{n}\sum_{i=1}^{k}\Gamma_{ij}(\alpha_i^{(j)})}, \dots, \frac{\sum_{i=1}^{k}\Gamma_{ij}(\alpha_i^{(n)})}{\sum_{j}^{n}\sum_{i=1}^{k}\Gamma_{ij}(\alpha_i^{(j)})} \right).$$

So,

$$W \sim \left(\frac{\sum_{i=1}^{k}\Gamma_{1j}(\alpha_i^{(j)})}{\sum_{j}^{n}\sum_{i=1}^{k}\Gamma_{ij}(\alpha_i^{(j)})}, \dots, \frac{\sum_{i=1}^{k}\Gamma_{kj}(\alpha_i^{(j)})}{\sum_{j}^{n}\sum_{i=1}^{k}\Gamma_{ij}(\alpha_i^{(j)})} \right).$$

We know that $\Gamma_{ij}(\alpha_i^{(j)})$'s ($for 1 \le i \le k$ and $1 \le j \le n$) are independent random variables with gamma distributions, leading to the independence of the components of W from X_j's (Basu's Theorem).

2.2 Mathematical Induction

The Second Method. The theorem follows for $n = 2$ from [7, Lemma 1]. Our proof is by induction; so, let the theorem hold for an arbitrary n (the induction hypothesis) and we prove that it holds for $n+1$ (the induction step). By dividing both sides of

$$\sum_{i=1}^{n+1} Y_i X_i = Y_{n+1} X_{n+1} + \left(\sum_{i=1}^{n} Y_i\right)\left(\sum_{j=1}^{n} \frac{Y_j}{\sum_{i=1}^{n} Y_i} X_j\right)$$

by $\sum_{i=1}^{n+1} Y_i$ and using the induction hypothesis, we will have

$$\sum_{i=1}^{n+1} \frac{Y_j}{\sum_{i=1}^{n} Y_i} X_i = \frac{Y_{n+1}}{\sum_{i=1}^{n+1} Y_i} X_{n+1} + \left(\sum_{i=1}^{n} \frac{Y_j}{\sum_{i=1}^{n} Y_i} \right) \left(\sum_{i=1}^{n} \frac{Y_j}{\sum_{i=1}^{n} Y_i} X_j\right) .$$

The right-hand side of which holds true for $n = 2$, by [7].

3. Simulations

In order to observe the values of the simulation, we consider the mixed model Z in the following particularities:

$$X_1 \sim Dirichlet(1,1,1),$$

$$X\ 2 \sim Dirichlet(1,1,1),$$

and thus, we define W by

$$W \sim Beta(3,3).$$

Firstly, we generate X_1 and X_2 data by means of the R software package. Then we compose them in accordance with the definition of Z. As a result, the values of Z will be observable. The figures show the simulated data.

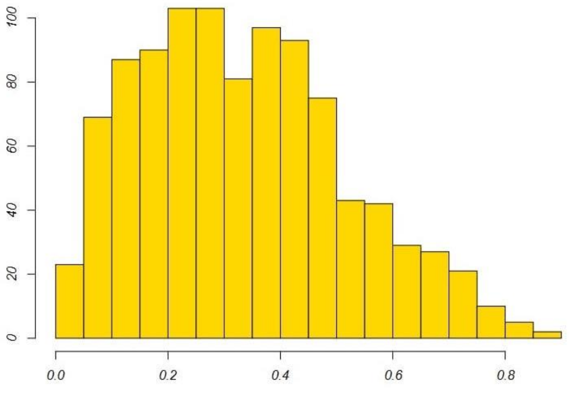

Figure 1. The Histogram of Z_1

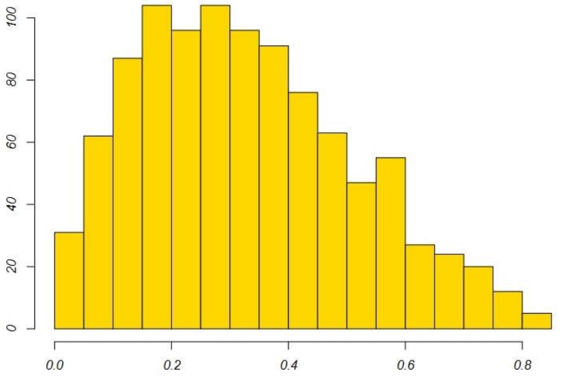

Figure 2. The Histogram of Z_2

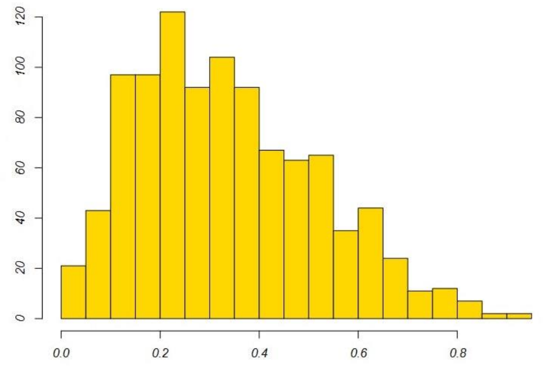

Figure 3. The Histogram of Z_3

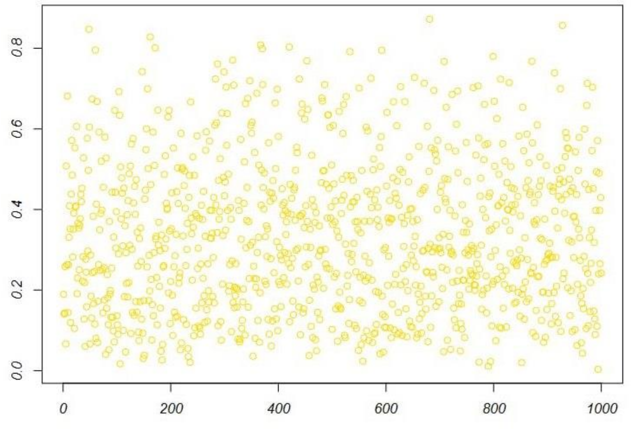

Figure 4. The Scatter Plot of Z_1

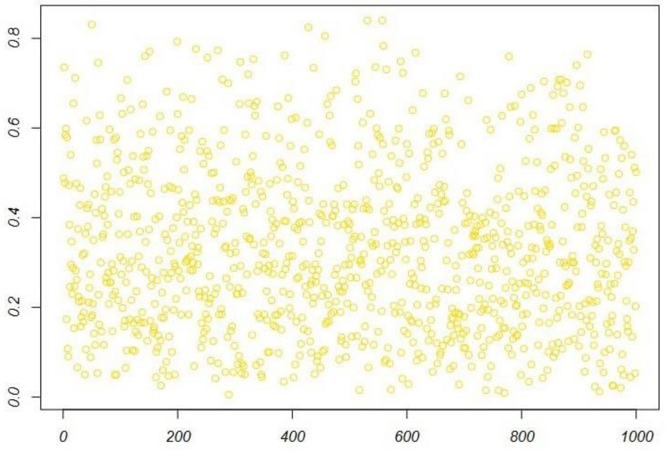

Figure 5. The Scatter Plot of Z_2

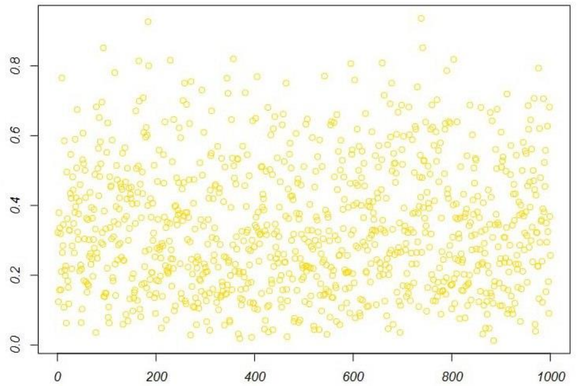

Figure 6. The Scatter Plot of Z_3

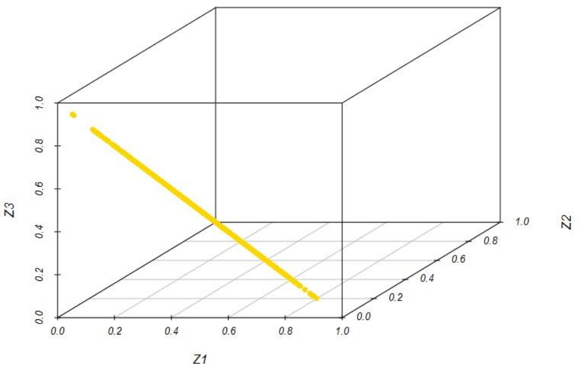

Figure 7. The 3D Plot of Z_1, Z_2, Z_3

4. An Observation

The following was observed by the second named author (who is also the corresponding author).

In the beginning of [8, Section 3], the author (of [8]) promises to “present a new result which extends the result of Van Assche (1987) for the randomly weighted averages of size 2 to the univariate and the multivariate random weighted averages of size m >2”. Then it is emphasized that “[t]his indeed has been an open problem” ([8, p. 3]). Here, we provide a very short solution for this problem. Our definitions and notation follow those of [8].

Theorem ([8, Theorem 3.1] corrected). Let $X_1, X_2, \dots$ be a sequence of [independent] multivariate random vectors with a continuous common distribution.

Let $S_{m+1} \underline{\underline{d}} R_{m:1} X_1 + \dots + R_{m:m+1} X_{m+1}, m \geq 0,$ where $U_{\langle m:1 \rangle}, \dots, U_{\langle m:m \rangle}$ are the cuts of $(0,1)$ by the uniform $(0,1)$ sample $U_{\langle m:1 \rangle}, \dots, U_{\langle m:m \rangle}$. If for some $m \geq 1$, $S_{m+1} \underline{\underline{d}} X_1$, then X_1

possesses the multivariate Cauchy distribution, which in turn implies [that] for every $m \geq 1$, $S_{m+1} \underline{\underline{d}} X_1$.

Proof. We prove the univariate case by induction on m (see [9, p.3]); the multivariate case follows straightforwardly. The case of $m = 1$ follows from (Van Assche 1987, Theorem 2), cited also in [8]. So let us suppose that the theorem holds for an arbitrary m (the induction hypothesis), and we prove that it holds for $m+1$ too (the induction step). We start from (see [9, p. 6], The Second Method) nothing that

$$S_{m+1} \underline{\underline{d}} U_{\langle m:m \rangle} S_m + (1 - U_{\langle m:m \rangle}) X_{m+1},$$

and so $Ee^{ixS_{m+1}} = E\,(Ee^{ix\left[U_{\langle m:m \rangle} S_m + \left(1 - U_{\langle m:m \rangle}\right) X_{m+1}\right]} \big| U_{\langle m:m \rangle})$.

Thus,

$$Ee^{ixS_{m+1}} = E_{U\langle m:m \rangle}\left(\phi_{S_{m+1}}(U_{\langle m:m \rangle} x)\phi_{X_{m+1}}((1 - U_{\langle m:m \rangle}) x)\right)$$

follows from the independence assumption (cf. [8, p. 3, line – 10]). Since S_{m+1} and S_m and X_{m+1} have identical distributions, $\phi(x) = E_{U_{\langle m:m \rangle}}\left(\phi(U_{\langle m:m \rangle} x)\phi((1 - U_{\langle m:m \rangle}) x)\right)$ holds. So, by using the completeness property, we have $\phi(x) = \phi(ux)\phi\left((1-u)x\right) w.p.1$, for every $u \in (0,1)$ and $x \in R$. We can take $m = 1$, since the equation is free of m. So, the case of $m+1$ follows from (Van Assche 1987, Theorem 2). Finally, the distributions are seen to be either Cauchy or Degenerate. Therfore, more generally, all the equations that are formed as $\phi(x) = \prod_{i=1}^{m+1} \phi(u_i x)$, for some $m \geq 1$ and some vector $(u_1, ..., u_{m+1})$ in which $u_i \rangle 0$ and $\sum_{i=1}^{m+1} u_i = 1$, have two common solutions; namely, the characteristic Cauchy and Degenerate functions.

Let us note that the above equation has two solutions, one of which (the Degenerate case) has been ignored in [8]. Also, the independence assumption (which is dropped but implicitly assumed in Theorem 3.1 of [8]) is necessary, since if we take X_i 's to be equal, then the deduced result holds for all distributions, not only for multivariate Cauchy (or Degenerate) distributions.

Although our method casts a shadow on the long and difficult arguments of [8], but it, not the least, generalizes their results. Lastly if this problem were really open, as claimed by the author of [8], then our solution method should be very useful. A more complete generalization of the above theorem, with an elegant proof, is in the possession of this (second-named) author, planned for a future publication.

Let us also notice that the bi conditional of [10, Theorem 3.2] does not hold (for general n's). The result should be corrected into:

3.2. Theorem. Assume S_n is given by (1.1) and $X_1,...,X_n$ are $i.i.d$. Continues random variables with a common distribution function F. Then S_n has distribution F if ~~and only if~~ F is a Cauchy distribution.

Proving the "only if" in [10] was based on the fact that, by an argument similar to the one given in 1987 by Van Assche, the solution of the differential equation (for $z \in \mathbb{C}$)

$$\frac{(-1)^{n-1}}{(n-1)!}\frac{d^{n-1}}{dz^{n-1}}\varphi(F,z)=\left[\varphi(F,z)\right]^n \text{ , (1)}$$

is

$$\varphi(F,z)=\frac{1}{z-a+ib}, \quad \mathrm{Im}(z)\rangle 0, \quad b \neq 0,$$

which is the Stieltjes Transform of the Cauchy distribution.

However, this solution for (1) works only for n=2, which leads us to the well-known Bernoulli equations, and there are no reason to believe that the same method gives all the solutions for $n\rangle 2$; we only know that one of the solutions is the one provided above and we have no idea about the other possible solutions. Note that the unique solution of the above differential equation of order $n-1$ should satisfy $n-1$ initial conditions. The rest of the paper [10] does not need any changes. A full proof of the above theorem is in the possession of the second named (the corresponding) author, kept for a future publication.

It is worth noting that our methods here are much simpler than those of [11]; see also [12, Subsection 1.1.1], where it is explained that the main result of [11] had been obtained before by the second author [13]. A future work in the line of this research is generalizing the methods of the present paper by using the ideas of [14].

5. Conclusion

The proof given in this paper, along with the poof that appeared in the literature, shows that the theorems of Basu and the induction method can be effective in finding the distribution of such randomly weighted averages. Also, it can be understood why the distribution should be Dirichlet.

References

[1] Homei, H. (2019), "The stochastic linear combination of Dirichlet distributions", Journal of Communications in Statistics: Theory and Methods, Vol. 50 No. 10, pp 2354-2359. https://doi.org/10.1080/03610926.2019.1664588

[2] Hadad, H., Homei, H., Behzadi, M.H. and Farnoosh, R. (2021), "Solving some Stochastic Differential Equation using Dirichlet Distributions", Journal of Computational Methods for Differential Equations, Vol. 9 No. 2, pp. 393-398. https://doi.org//10.22034/cmde.2019.32914.1533

[3] Homei, H. (2012), "Randomly Weighted Averages with Beta Random Proportions", The Journal of Statistics and probability Letters, Vol. 82 NO. 8, pp 1515-1520. https://doi.org/10.1016/j.spl.2012.04.017

[4] Homei, H. (2015), "A Novel Extension of Randomly Weighted Averages",The Journal of Statistical Papers, Vol. 56 No. 4, pp 933-946. https://doi.org/10.1007/s00362-014-0615-5

[5] Homei, H. (2017), "Characterizations of Arcsin and Related Distributions Based on a New Generalized Unimodality", Journal of Communications in Statistics: Theory and Methods, Vol. 46 No. 2, pp 1024-1030. https://doi.org/10.1080/03610926.2015.1006788

[6] Homei, H. and Nadarajah, S. (2018), "On products and mixed Sums of Gamma and Beta Random Variables Motivated by Availability", Journal of Methodology and Computing in Applied Probability, Vol. 20 No. 2, pp 799-810. https://doi.org/10.1007/s11009-017-9591-2

[7] Sethhuraman, J. (1994),"A Constructive Definition on Dirichlet Priors", The Journal of Statistica Sinica, Vol. 4 No. 2, pp 639-650. https://doi.org/10.2307/24305538

[8] Soltani, A. R. (2022),"Recursive Integral Equations for Random Weights Averages: Exponential Functions and Cauchy Distribution", The Journal of Statistics and Probability Letters, Vol. 190. No. 6, pp. (7pp) https://doi.org/10.1016/j.spl.2022.109606

[9] Homei, H. (2017), "Randomly Weighted Averages: a multi-variate case", The Book of the Abstracts of the 11th Seminar on probability and Stochastic Processes, Qazvin, Iran, (30-31 August 2017), Imam Khomeini International University, P. 49. https://doi.org/N.A.

[10] Soltani, A. R. and Homei, H. (2009), "Weighted Averages with Random Proportions that are jointly Uniformly Distributed over the Unit Simplex", The Journal of Statistics and probability Letters, Vol. 79 No. 9, pp. 1215-1218. https://doi.org/10.1016/j.spl.2009.01.009

[11] Soltani, A. R. and Roozegar, R. (2020)," Averages for multivariate random Vectors with random weights", Revstat: Statistical Journal, Vol. 18 No. 4, pp. 453-460. https://doi.org/10.57805/revstat.v18i4.311

[12] Homei, H., Nadarajah, S. and Taherkhani, A. (2023),"Randomly Weighted Averages on Multivariate Dirichlet Distributions with Generalized Parameters", Revstat: Statistical Journal, forthcoming (online: April 2023). https://doi.org/N.A.

[13] Homei, H. (2016), "Randomly weighted averages: a multi-variate case", arXiv:1603.00596 [math.ST]. https://doi.org/10.48550/arXiv.1603.00596

[14] Alemam, S., Homei, H. and Nadarajah, S. (2023), “Generalizations of Rao-Blackwell and Lehmann-Scheffe Theorems with Applications”, Journal of Mathematics, Vol. 11 No. 19, pp. (11pp). https://doi.org/10.3390/math11194146