Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 5, No. 2, 2023

Generalized Dirichlet Distribution Based on Confluent Hypergeometric Series

Ruixin Zhao, Hongmei Liu, Yu Tang

School of Science, Dalian Minzu University, Liaoning 116600, P.R.China

Abstract: Dirichlet distribution is a kind of high-dimensional continuous probability distribution, which has important applications in the fields of statistics, machine learning and bioinformatics. In this paper, based on gamma distribution we study two two-dimensional random variables. Then we derive the properties of these two two-dimensional random variables by using the properties of non-central gamma distribution and confluent hypergeometric series. From these properties, we find the two random variables follow generalized Dirichlet distributions. Applying hypergeometric series to Dirichlet distribution broadens the research of Dirichlet distribution.

Keywords: Dirichlet distribution, Confluent hypergeometric function, Gamma distribution, Non-central gamma distribution.

1. Introduction

Dirichlet distribution is a kind of high-dimensional continuous probability distribution with positive simplex as the support set in the real fields, and is a generalization of the beta distribution in the high-dimensional case [1]. In Bayesian analysis, the Dirichlet distribution as a conjugate prior for multinomial distributions is used for parameter estimation of multinomial, binomial and type distributions [1]. In the field of machine learning, Dirichlet and generalized Dirichlet distributions are mostly applied to build mixture models to deal with unsupervised learning problems such as high-dimensional clustering and feature empowerment [2]. In natural language processing and bioinformatics research, the implicit Dirichlet distribution is used to identify potential subject word information in a document set [3][4].

In this article, we study the joint probability density, marginal density, and the distribution of sum, product and quotient for two two-dimensional random variables by using the properties of generalized Dirichlet distribution, non-central gamma distribution and confluent hypergeometric function. So we firstly introduce the generalized Dirichlet type 1 distribution and generalized Dirichlet type 2 distribution.

Let (Z_1,Z_2) be a two-dimensional random variable, if its p.d.f is

$$D1(z_1,z_2;a_1,a_2,a_3)=\frac{z_1^{a_1-1}z_2^{a_2-1}(1-z_1-z_2)^{a_3-1}}{B(a_1,a_2,a_3)},z_1>0,z_2>0,z_1+z_2<1,$$

where $B(a_1,a_2,a_3)=\dfrac{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)}{\Gamma(a_1+a_2+a_3)}$, then (Z_1,Z_2) follows the Dirichlet type 1 distribution.

Let (T_1,T_2) be a two-dimensional random variable, if its p.d.f is

$$D2(t_1,t_2;a_1,a_2,a_3)=\frac{t_1^{a_1-1}t_2^{a_2-1}(1+t_1+t_2)^{-(a_1+a_2+a_3)}}{B(a_1,a_2,a_3)},t_1>0,t_2>0,$$

then (T_1,T_2) follows the Dirichlet type 2 distribution.

Random variable U has a non-central gamma distribution with shape parameter $k(>0)$ and non-central parameter $\delta(\geq 0)$, denoted by $U\sim NCG(k;\delta)$, the probability density is given by [5]

$$f_{NCG}(u;k;\delta)=\{\Gamma(\kappa)\}^{-1}\exp(-\delta-u)u^{k-1}{}_0F_1(\kappa;\delta u),u>0,$$

where ${}_0F_1(m;n)=\sum_{j=0}^{\infty}\frac{1}{(m)_j}\frac{n^j}{j!}$. When $\delta=0$, the non-central gamma distribution reduces to the standard gamma distribution, denoted by $U\sim G(\kappa)$.

Then we introduce hypergeometric series. The definition of generalized hypergeometric function is given by [6]

$${}_pF_q(a_1\cdots a_p;b_1\cdots b_q;z)=\sum_{k=0}^{\infty}\frac{(a_1)_k\cdots(a_p)_k z^k}{(b_1)_q\cdots(b_q)_k k!} \tag{1.1}$$

where Pochhammer symbol $(a)_n=a(a+1)\ldots(a+n-1),n=1,2,3\cdots$, $(a)_0=1$, the convergence condition of this series is shown in [7].

The confluent hypergeometric function is a special case of the generalized hypergeometric function, which is expressed as follows [6]:

$${}_1F_1(a;b;x)=\sum_{k=0}^{\infty}\frac{(a)_k x^k}{(b)_k k!}.$$

Its integral form is

$${}_1F_1(a;b;x)=\frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)}\int_0^1 t^{a-1}(1-t)^{b-a-1}\exp(xt)dt.$$

Let independent random variable $X_i,i=1,2,3$ follow gamma distributions with shape parameter $v_i,i=1,2,3$, $(Z_1,Z_2)=(X_1/(X_1+X_2+X_3),X_2/(X_1+X_2+X_3))$ and $(Y_1,Y_2)=(X_1/X_3,X_2/X_3)$, then (Z_1,Z_2) and (Y_1,Y_2) follow Dirichlet type 1 distribution and Dirichlet type 2 distribution respectively. Orozco-Castañeda, Nagar and Gupta [6] generalized the above distribution by hypergeometric series. Let the random variables U,V,W be independent, U and V follow gamma distributions with shape parameter a and b, W follows the non-central gamma distribution with shape parameter c and non-central parameter δ, $P=U/(U+W)$, $Q=V/(V+W)$, Gupta, Orozco-Castañeda and Nagar [5] studied the joint probability density of P and Q as well as other statistical properties. Based on the previous work, in this paper we define

$(X,Y)=(U/(U+V+W),V/(U+V+W))$ and $(X_1,Y_1)=(U/W,V/W)$, then study their joint probability density, marginal probability density, and statistical properties of sum, product, quotient. From these properties, authors find that they follow generalized Dirichlet type 1 distribution and generalized Dirichlet type 2 distribution respectively.

2. Generalized Dirichlet Type 1 Distribution

Theorem 2.1: Let U,V and W be independent random variables, $U\sim G(a),V\sim G(b)$ and $W\sim NCG(\mathrm{c},\delta)$, define $X=U/(U+V+W)$, $Y=V/(U+V+W)$, then the joint density of X and Y is give by

$$\frac{x^{a-1}y^{b-1}(1-x-y)^{c-1}e^{-\delta}}{B(a,b,c)}{}_1F_1\big(a+b+c;c;\delta(1-x-y)\big),\ x>0,y>0,x+y<1 \quad (2.1)$$

The p.d.f of $S=U+V+W$ is given by

$$\frac{e^{-\delta-s}s^{a+b+c-1}}{\Gamma(a+b+c)}{}_0F_1(a+b+c;\delta s).$$

Proof: Since U,V and W are independent random variables, the joint density is given by

$$\frac{e^{-\delta-u-v-w}u^{a-1}v^{b-1}w^{c-1}}{\Gamma(a)\Gamma(b)\Gamma(c)}{}_0F_1(c;\delta w),u>0,v>0,w>0. \quad (2.2)$$

Making the transformation $X=U/(U+V+W),Y=V/(U+V+W)$ and $S=U+V+W$ with the Jacobian $J(u,v,w\to x,y,s)=s^2$, then we obtain the joint p.d.f of X,Y and S as:

$$\frac{x^{a-1}y^{b-1}(1-x-y)^{c-1}e^{-\delta}}{\Gamma(a)\Gamma(b)\Gamma(c)}e^{-s}s^{a+b+c-1}{}_0F_1\big(c;\delta s(1-x-y)\big), \quad (2.3)$$

where $x>0,y>0,x+y<1$ and $s>0$, integrating s, then the joint p.d.f of (X,Y) is given by

$$\frac{x^{a-1}y^{b-1}(1-x-y)^{c-1}e^{-\delta}}{\Gamma(a)\Gamma(b)\Gamma(c)}\int_0^{+\infty}e^{-s}s^{a+b+c-1}{}_0F_1\big(c;\delta s(1-x-y)\big)ds$$

$$=\frac{x^{a-1}y^{b-1}(1-x-y)^{c-1}e^{-\delta}}{\Gamma(a)\Gamma(b)\Gamma(c)}\sum_{j=0}^{\infty}\frac{\delta^j(1-x-y)^j}{(c)_j\,j!}\int_0^{+\infty}e^{-s}s^{a+b+c+j-1}ds$$

$$=\frac{x^{a-1}y^{b-1}(1-x-y)^{c-1}e^{-\delta}}{\Gamma(a)\Gamma(b)\Gamma(c)}\sum_{j=0}^{\infty}\frac{\delta^j(1-x-y)^j\Gamma(a+b+c+j)}{(c)_j\,j!}.$$

From $\Gamma(a+k)=(a)_k\Gamma(a)$, (2.1) is proved.

Further, integrating x and y in (2.3), we can get the marginal density of S:

$$\int_0^1\int_0^{1-x}\frac{x^{a-1}y^{b-1}(1-x-y)^{c-1}}{\Gamma(a)\Gamma(b)\Gamma(c)}e^{-\delta-s}s^{a+b+c-1}\sum_{j=0}^{\infty}\frac{\delta^j s^j(1-x-y)^j}{(c)_j}dxdy$$

$$=\frac{e^{-\delta-s}s^{a+b+c-1}}{\Gamma(a)\Gamma(b)\Gamma(c)}\sum_{j=0}^{\infty}\frac{(\delta s)^j}{(c)_j}\int_0^1x^{a-1}(1-x)^{c+j-1}\Big(\int_0^{1-x}y^{b-1}\Big(1-\frac{y}{1-x}\Big)^{c+j-1}dy\Big)dx.$$

Setting $u=y/(1-x)$, from the definition of the beta function, the above formula can be calculated as:

$$=\frac{e^{-\delta-s}s^{a+b+c-1}}{\Gamma(a)\Gamma(b)\Gamma(c)}\sum_{j=0}^{\infty}\frac{(\delta s)^j}{(c)_j}\int_0^1x^{a-1}(1-x)^{c+b+j-1}\frac{\Gamma(b)\Gamma(c+j)}{\Gamma(b+c+j)}dx$$

$$=\frac{e^{-\delta-s}s^{a+b+c-1}}{\Gamma(a)\Gamma(b)\Gamma(c)}\sum_{j=0}^{\infty}\frac{(\delta s)^j\Gamma(b)\Gamma(c+j)\Gamma(a)\Gamma(c+b+j)}{(c)_j\Gamma(b+c+j)\Gamma(a+b+c+j)}.$$

From $\Gamma(a+k)=(a)_k\Gamma(a)$, the p.d.f of S is proved.

From Theorem 2.1, we find that (X,Y) follows the generalized Dirichlet type 1 distribution, $S=U+V+W$ follows the non-central gamma distribution, denoted by $S\sim NCG(a+b+c;\delta)$.

Theorem 2.2: If the p.d.f of (X,Y) is given by Theorem 2.1, then the marginal density of X is

$$\frac{e^{-\delta}x^{a-1}(1-x)^{b+c-1}}{B(a,b+c)}{}_1F_1\big(a+b+c;b+c;\delta(1-x)\big).$$

Proof: Integrating y in (2.1), one obtains

$$\frac{e^{-\delta}}{B(a,b,c)}\int_0^{1-x}x^{a-1}y^{b-1}(1-x-y)^{c-1}{}_1F_1\big(a+b+c;c;\delta(1-x-y)\big)dy$$

.

Inserting the series form of ${}_1F_1$ into this formula, the marginal density of X is

$$\frac{e^{-\delta}x^{a-1}}{B(a,b,c)}\sum_{j=0}^{\infty}\frac{(a+b+c)_j\delta^j}{(c)_j\,j!}\int_0^{1-x}y^{b-1}(1-x-y)^{c+j-1}dy$$

$$=\frac{e^{-\delta}x^{a-1}}{B(a,b,c)}\sum_{j=0}^{\infty}\frac{(a+b+c)_j\delta^j}{(c)_j\,j!}(1-x)^{c+j-1}\int_0^{1-x}y^{b-1}\Big(1-\frac{y}{1-x}\Big)^{c+j-1}dy.$$

Setting $t=y/(1-x)$, then the marginal density of X is

$$\frac{e^{-\delta}x^{a-1}}{B(a,b,c)}\sum_{j=0}^{\infty}\frac{(a+b+c)_j\delta^j}{(c)_j\,j!}(1-x)^{c+j-1+b}\frac{\Gamma(b)\Gamma(c+j)}{\Gamma(b+c+j)}$$

$$=\frac{e^{-\delta}x^{a-1}(1-x)^{b+c-1}}{B(a,b+c)}{}_1F_1\big(a+b+c;b+c;\delta(1-x)\big).$$

From Theorem 2.2, we find that X follows the generalized Beta type 1 distribution.

Theorem 2.3: If the p.d.f of (X,Y) is given by Theorem 2.1, then the density of $Z=X+Y$ is

$$\frac{z^{a+b-1}(1-z)^{c-1}e^{-\delta}}{B(c,a+b)}{}_1F_1\big(a+b+c;c;\delta(1-z)\big),\ \ 0<z<1.$$

Proof: The p.d.f (2.1) of (X,Y) is known. By using the convolution formula, the probability density of $Z=X+Y$ is

$$\frac{1}{B(a,b,c)}\int_0^z x^{a-1}(z-x)^{b-1}(1-z)^{c-1}e^{-\delta}{}_1F_1(a+b+c;c;\delta(1-z))dx$$

$$=\frac{(1-z)^{c-1}e^{-\delta}}{B(a,b,c)}\sum_{j=0}^{\infty}\frac{(a+b+c)_j\delta^j(1-z)^j}{(c)_j\,j!}\int_0^z x^{a-1}(z-x)^{b-1}dx$$

$$=\frac{(1-z)^{c-1}e^{-\delta}}{B(a,b,c)}\sum_{j=0}^{\infty}\frac{(a+b+c)_j\delta^j(1-z)^j}{(c)_j\,j!}z^{b-1}\int_0^z x^{a-1}\Big(1-\frac{x}{z}\Big)^{b-1}dx.$$

Similar to the proof of the previous theorem, (2.3) is proved.

Theorem 2.4: If the p.d.f of (X,Y) is given by Theorem 2.1, then the density of $Z=XY$ is

$$\frac{z^{b-1}e^{-\delta}}{B(a,b,c)}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(a+b+c)_j\delta^j(-1)^{c+j-1}z^{c+j-1-k}(1-c-j)_k(a-b-c+1)_{k-j}}{(c)_j\,j!(a-b-c+1)_{2k-j+1}}.$$

Proof: According to the probability density formula of two-dimensional random variable product:

$f_{XY}(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}f\left(x,\frac{z}{x}\right)dx$, similar to the proof of the previous theorem, (2.4) is proved.

Theorem 2.5: If the p.d.f of (X,Y) is given by Theorem 2.1, then the density of $Z=Y/X$ is

$$\frac{z^{b-1}e^{-\delta}}{B(a,b,c)}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(a+b+c)_j\delta^j(1-c-j)_k(1+z)^k}{(c)_j j!k!(a+b+k)}.$$

Proof: According to probability density formula of two-dimensional random variable quotient : $f_{\frac{Y}{X}}(z)=\int_{-\infty}^{+\infty}|x|f(x,xz)dx$, similar to the proof of the previous theorem, (2.5) is proved.

3. Generalized Dirichlet Type 2 Distribution

Theorem 3.1: Let U,V and W be independent random variables, $U\sim G(a),V\sim G(b)$ and $W\sim NCG(\mathrm{c},\delta)$, define $X_1=U/W,Y_1=V/W$, then the p.d.f of (X_1,Y_1) is

$$\frac{e^{-\delta}{x_1}^{a-1}{y_1}^{b-1}}{B(a,b,c)(x_1+y_1+1)^{a+b+c}}{}_1F_1\left(a+b+c;c;\frac{\delta}{x_1+y_1+1}\right),x_1>0,y_1>0 \quad (3.1)$$

Proof: Making the transformation $X_1=U/W,Y_1=V/W,W=W$ with the Jacobian $J(u,v,w\to x_1,y_1,w)=w^2$ in (2.2), then the joint p.d.f of X_1,Y_1 and W is

$$\frac{e^{-\delta-(x_1+y_1+1)w}{x_1}^{a-1}{y_1}^{b-1}w^{a+b+c-1}}{\Gamma(a)\Gamma(b)\Gamma(c)}{}_0F_1(c;\delta w),\ \ x_1>0,y_1>0.$$

Integrating w in the above formula, we get the p.d.f of (X_1,Y_1) :

$$\frac{e^{-\delta}{x_1}^{a-1}{y_1}^{b-1}}{\Gamma(a)\Gamma(b)\Gamma(c)}\int_0^{+\infty}e^{-(x_1+y_1+1)w}w^{a+b+c-1}\sum_{k=0}^{\infty}\frac{(\delta w)^k}{(c)_k k!}dw$$

$$=\frac{e^{-\delta}{x_1}^{a-1}{y_1}^{b-1}}{\Gamma(a)\Gamma(b)\Gamma(c)}\sum_{k=0}^{\infty}\frac{\delta^k}{(c)_k k!}\int_0^{+\infty}e^{-(x_1+y_1+1)w}w^{a+b+c+k-1}dw$$

$$=\frac{e^{-\delta}{x_1}^{a-1}{y_1}^{b-1}}{\Gamma(a)\Gamma(b)\Gamma(c)(x_1+y_1+1)^{a+b+c}}\sum_{k=0}^{\infty}\frac{\delta^k(x_1+y_1+1)^{-k}\Gamma(a+b+c+k)}{(c)_k k!}.$$

After simplification, (3.1) is proved.

From Theorem 3.1, we find that (X_1,Y_1) follows the generalized Dirichlet type 2 distribution.

Theorem 3.2: If the p.d.f of (X_1,Y_1) is given by Theorem 3.1, then the marginal density of X_1 is

$$\frac{e^{-\delta}{x_1}^{a-1}}{B(a,c)(1+x_1)^{a+c}}{}_1F_1\left(a+c;c;\frac{\delta}{1+x_1}\right).$$

Proof: Integrating y_1 in (3.1), (3.2) is proved.

From Theorem 3.2, we find that X_1 follows the generalized Beta type 2 distribution.p;

Theorem3.3: If the p.d.f of (X_1,Y_1) is given by Theorem 3.1, then the density of $\mathrm{S}=X_1+Y_1$ is

$$\frac{\mathrm{e}^{-\delta}s^{a+b-1}}{B(a+b,c)(s+1)^{a+b+c}}{}_1F_1\left(a+b+c;c;\frac{\delta}{s+1}\right),\ s>0.$$

Proof: By using the convolution formula, (3.3) is proved.

4. Conclusion

This paper makes a preliminary study on two two-dimensional random variables based on confluent hypergeometic series. From the above theorems, we find that (X,Y) and (X_1,Y_1) follow the generalized Dirichlet type 1 distribution and generalized Dirichlet type 2 distribution , which provides new methods for the future study of Dirichlet distribution.

References

[1] Kai Wang Ng, Guo-Liang Tian, Man-Lai Tang. Dirichlet and Related Distributions: Theory, Methods and Applications. New York: John Wiley & Sons, 2011.

[2] Mohamed Maher Ben Ismaila, Hichem Friguib. Unsupervised clustering and feature weighting based on Generalized Dirichlet mixture modeling. Information Sciences, 2014, 274: 35-54.

[3] David M. Blei, Andrew Y. Ng, Michael I. Jordan. Latent Dirichlet Allocation. Journal of Machine Learning Research, 2003, 3: 993-1022.

[4] S. Shivashankar, S. Srivathsan, B. Ravindran, Ashish V. Tendulkar. Multi-view methods for protein structure comparison using latent dirichlet allocation. Bioinformatics, 2011, 27: i61-i68.

[5] Arjun K. Gupta, Johanna Marcela Orozco-Castañeda, Daya K. Nagar. Non-central bivariate beta distribution. Springer, 2011, 52: 139-152.

[6] Johanna Marcela Orozco-Castañeda, Daya K. Nagar, Arjun K. Gupta. Generalized bivariate beta distributions involving Appell's hypergeometric function of the second kind. Computers and Mathematics with Applications, 2012, 64: 2507-2519.

[7] Yudell L. Luke. The Special Functions and Their Approximations. Mathematics in Science and Engineering, 1969,53: 1-349.

[8] D. K. Nagar, Johanna Marcela Orozco-Castañeda, Arjun K. Gupta. Product and quotient of correlated beta variables. Appl. Math. Lett., 2009, 22: 105-109.

[9] Ingram Olkin, Ruixue Liu. A bivariate beta distribution. Statistics & Probability Letters, 2003, 62: 407-412.

[10] Saralees Nadarajah, Samuel Kotz. Some bivariate beta distributions. Statistics, 2005, 39: 457-466.

[11] James J. Chen, Melvin R. Novick, Bayesian Analysis for Binomial Models with Generalized Beta Prior Distributions. Journal of Educational and Behavioral Statistics, 1984, 9: 163-175.

[12] James M. Dickey. Multiple Hypergeometric Functions: Probabilistic Interpretations and Statistical Uses.Journal of the American Statistical Association, 1983, 78: 628-637.

[13] Mohamed Abdalla, Muajebah Hidan. Analytical properties of the two-variables Jacobi matrix polynomials with applications. Demonstratio Mathematica, 2021, 54: 178-188.

[14] Daya K. Nagar, Fabio Fabio Sepúlveda-Murillo. Properties of confluent hypergeometric function kind 1 distribution. Journal of Interdisciplinary Mathematics, 2008, 11: 495-504.

[15] Jiayu Lin. On The Dirichlet Distribution by Jiayu Lin. Canada: Mathematics, 2016.

[16] A. M. Mathai, R. K. Saxena. On a generalized hypergeometric distribution. Springer, 1967, 11: 127-132.

[17] W. N. Bailey. Generalized hypergeometric series .London: Mathematics, 1953.

[18] Wentao Fan, Hassen Sallay, Nizar Bouguila. Online Learning of Hierarchical Pitman-Yor Process Mixture of Generalized Dirichlet Distributions With Feature Selection. IEEE Trans Neural Netw Learn Syst, 2017, 28: 2048-2061.