

# An Introduction to a Generalized Dirichlet Distribution with Flexible Marginal Variances

*Bret Larget*

*10/16/2017*

## Purpose

This document is aimed at an advanced undergraduate level to describe a generalization to the Dirichlet distribution and to highlight existing open problems.

## The Dirichlet Distribution

The Dirichlet distribution is a very commonly used probability distribution on sets of positive random variables constrained to sum to one, or a set of random variables on a simplex. The random variables  $X_1, \dots, X_k$  are said to have a Dirichlet distribution when they have the joint density

$$\frac{\Gamma(\alpha_1 + \dots + \alpha_k)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i - 1}, \quad \text{where } x_i > 0 \text{ for all } i \text{ and } \sum_{i=1}^k x_i = 1$$

where the parameters  $\alpha_i > 0$  for  $i = 1, \dots, k$ .

## Derivation

The Dirichlet distribution arises by taking  $k$  independent gamma random variables with the same rate parameter and dividing each by their sum. Let the random variables  $Y_1, \dots, Y_k$  be independent with  $Y_i \sim \text{Gamma}(\alpha_i, \lambda)$  for some positive parameter  $\lambda$ , let  $S = \sum_i Y_i$ , and define  $X_i = Y_i/S$  for  $i = 1, \dots, k$ . Then the joint distribution of  $(X_1, \dots, X_k)$  is Dirichlet with parameters  $\alpha_1, \dots, \alpha_k$ . Note that the joint density of the  $\{X_i\}$  does not depend on  $\lambda$ , which is often chosen to be equal to 1 without loss of generality. The reason is that if  $Y \sim \Gamma(\alpha, \lambda)$ , then  $\lambda Y \sim \Gamma(\alpha, 1)$ , and  $\lambda$  cancels out in the normalization that defines the  $\{X_i\}$ .

## Gamma Distribution

As a reminder, the gamma density is

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

where the gamma function  $\Gamma(\cdot)$  is defined as

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

for argument  $z > 0$ . In the case of integer argument,  $\Gamma(n) = (n-1)!$  for  $n = 1, 2, 3, \dots$ . Another fundamental identity is  $\Gamma(z+1) = z\Gamma(z)$ .

The mean and variance of this distribution are  $E(Y) = \alpha/\lambda$  and  $\text{Var}(Y) = \alpha/\lambda^2$ . When  $\alpha = 1$ , the gamma distribution is known as the exponential distribution. The parameter  $\alpha$  is called the *shape parameter* as it alone determines the shape of the density. The parameter  $\lambda$  is the *scale parameter* as the density looks identical when  $\lambda$  changes other than changing the scale of the axis.

Each random variable  $X_i$  of the Dirichlet distribution has a marginal  $\text{Beta}(\alpha_i, \theta - \alpha_i)$  distribution where  $\theta = \sum_{i=1}^k \alpha_i$ .

## Beta Distribution

As an aside, here is a description of the Beta distribution. There are two parameters  $\alpha$  and  $\beta$ , each positive. The density is

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1$$

Hence, the Dirichlet distribution with  $k = 2$  is the Beta distribution and the Dirichlet distribution may be thought of as a multivariate extension of the Beta distribution.

## Moments

If  $X \sim \text{Beta}(\alpha, \beta)$ , then  $\mathbf{E}(X) = \alpha/(\alpha + \beta)$  and  $\text{Var}(X) = \alpha\beta/((\alpha + \beta)^2(\alpha + \beta + 1))$ . Thus, the marginal moments of the Dirichlet distribution are  $\mathbf{E}(X_i) = \alpha_i/\theta$  and variance  $\text{Var}(X_i) = \alpha_i(\theta - \alpha_i)/(\theta^2(\theta + 1))$ .

A consequence is that when attempting to select a Dirichlet distribution to fit the distribution of a given set of positive random variables constrained to equal one, while it is possible to select the parameters  $\{\alpha_i\}$  to match the marginal means by letting  $\alpha_i$  be proportional to the desired marginal mean, there remains only a single scale parameter which determines all of the marginal variances. We seek a generalization with a larger parameterization that is flexible enough to match, at least approximately, the means and variances of each marginal distribution.

## A Generalized Dirichlet Distribution

Define the flexible marginal variance generalized Dirichlet distribution on  $X_1, \dots, X_n$  to be the distribution of  $(X_1, \dots, X_k)$  where  $X_i = Y_i / \sum_{j=1}^k Y_j$  for  $i = 1, \dots, k$  where the random variables  $\{Y_i\}$  are mutually independent and  $Y_i \sim \text{Gamma}(\alpha_i, \lambda_i)$ . As the distribution of the  $\{X_i\}$  would be the same if all  $\{Y_i\}$  were multiplied by a common constant, we add the constraint that  $\sum_{i=1}^k \lambda_i = k$  so that the average values of the  $\{\lambda_i\}$  parameters is one.

It is known that the distribution of the sum  $S = \sum_{i=1}^k Y_i$  may be written as an infinite mixture of Gamma densities, but does not have a closed-form density. However, the joint density of  $(X_1, \dots, X_k)$  does have a closed-form solution.

$$f(x_1, \dots, x_k) = \frac{\Gamma(\sum_{i=1}^k \alpha_i) (\prod_{i=1}^k \lambda_i^{\alpha_i})}{\prod_{i=1}^k \Gamma(\alpha_i)} \times \frac{\prod_{i=1}^k x_i^{\alpha_i-1}}{(\sum_{i=1}^k \lambda_i x_i)^{\sum_{i=1}^k \alpha_i}}, \quad \text{where } x_i > 0 \text{ for all } i \text{ and } \sum_{i=1}^k x_i = 1$$

## Marginal Moments

We have not been able to derive closed form solutions for the marginal means and variances of the generalized Dirichlet distribution for general  $k$ , but when all of the  $\alpha_i$  values are large, the marginal means are numerically close to  $(\alpha_i/\lambda_i) / \sum_{j=1}^k (\alpha_j/\lambda_j)$ . However, we have found a system of linear equations that relate the marginal means on lattices in the parameter space. We present the following relationship between marginal means from generalized Dirichlet distributions with common  $\lambda$  vectors and  $\alpha$  vectors that differ by one in a single dimension. To simplify notation, we rewrite the density as

$$f(x|\alpha, \lambda) = \frac{\Gamma(\theta)\lambda^\alpha}{\Gamma(\alpha)} \times \frac{x^{\alpha-1}}{(\lambda \cdot x)^\theta}$$

where  $x = (x_1, \dots, x_k)$ ,  $\alpha = (\alpha_1, \dots, \alpha_k)$ ,  $\lambda = (\lambda_1, \dots, \lambda_k)$ ,  $\theta = \sum_{i=1}^k \alpha_i$ , an expression of vectors of the form  $a^b$  is short for  $\prod_{i=1}^k a_i^{b_i}$ ,  $\Gamma(\alpha)$  represents  $\prod_{i=1}^k \Gamma(\alpha_i)$ , and  $\lambda \cdot x$  is the dot product  $\sum_{i=1}^k \lambda_i x_i$ . Furthermore, let  $\alpha^{(j)}$  represent the vector where the  $i$ th element equals  $\alpha_i$  if  $i \neq j$  and equals  $\alpha_j + 1$  when  $i = j$ . The integral

expression  $\int_{\Delta} \cdot dx$  refers to integration over the simplex  $\sum_{i=1}^k x_i = 1$  and  $0 < x_i < 1$  for  $i = 1, \dots, k$ . We introduce the notation

$$m_j(\alpha, \lambda) = \mathbf{E}(X_j | \alpha, \lambda) = \int_{\Delta} x_j f(x | \alpha, \lambda) dx$$

for the marginal means. Using the fact that  $a\Gamma(a) = \Gamma(a+1)$ , we note that

$$\begin{aligned} x_j f(x | \alpha, \lambda) &= x_j \times \frac{\Gamma(\theta)\lambda^\alpha}{\Gamma(\alpha)} \times \frac{x^{\alpha-1}}{(\lambda \cdot x)^\theta} \\ &= \frac{(\Gamma(\theta+1)/\theta)(\lambda^{\alpha^{(j)}}/\lambda_j)}{\Gamma(\alpha^{(j)})/\alpha_j} \times \frac{x^{\alpha^{(j)}-1}(\lambda \cdot x)}{(\lambda \cdot x)^{\theta+1}} \\ &= \left( \frac{\alpha_j/\lambda_j}{\theta} \right) (\lambda \cdot x) f(x | \alpha^{(j)}, \lambda) \end{aligned}$$

Integrating over the simplex on both sides shows that

$$m_j(\alpha, \lambda) = \left( \frac{\alpha_j/\lambda_j}{\theta} \right) \sum_{i=1}^k \lambda_i m_i(\alpha^{(j)}, \lambda)$$

### Open problem 1

Is there a similar expression for the marginal variance? A similar derivation should provide an expression for  $\mathbf{E}(X_j)^2$ . Then use the fact that  $\text{Var}(X_j) = \mathbf{E}(X_j)^2 - (\mathbf{E}(X_j))^2$ .

### Ratios of marginal means

The previous section shows that while we do not have a closed-form solution for each marginal mean, we can express the marginal mean of  $X_j$  as a linear combination of the marginal means from the generalized Dirichlet distribution with the same  $\lambda$  vector and an  $\alpha$  vector that is incremented by one in the  $j$ th dimension. Another relationship involves ratios of marginal means from neighboring points in the parameter space.

#### Proposition

$$\frac{m_j^i}{m_i^j} = \frac{\alpha_i/\lambda_i}{\alpha_j/\lambda_j} \quad \text{for all } 1 \leq i, j \leq k \text{ and for all } k$$

The above expression uses the following notation simplification. As we are considering distributions that share a common vector  $\lambda$ , we suppress  $\lambda$  and let

$$m_j^0 = m_j(\alpha, \lambda) \quad \text{and} \quad m_j^i = m_j(\alpha^{(i)}, \lambda)$$

so the superscript 0 refers to means with the base vector  $\alpha$  and the superscript  $i$  refers to means where the base  $\alpha$  vector has been incremented by one in the  $j$ th dimension. Using the new notation, the recursion expression is

$$m_j^0 = \frac{\alpha_j}{\lambda_j \theta} \sum_{i=1}^k \lambda_i m_i^j$$

where  $\theta = \sum_{i=1}^k \alpha_i$ .

We only have a complete proof of the proposition in the case where  $k = 2$ , but simulation suggests that the ratio result is true for general  $k$ . We derive the proof for general  $k$  and note that this final expression is sufficient when  $k = 2$ .

### Proof

Multiply both sides of the recursion by  $\theta$  and sum over  $j$ .

$$\theta \sum_{j=1}^k m_j^0 = \sum_{j=1}^k \frac{\alpha_j}{\lambda_j} \sum_{i=1}^k \lambda_i m_i^j$$

Simplify the left-hand side and separate the double sum on the right hand side into the cases where  $i$  is equal to and not equal to  $j$ .

$$\theta = \sum_{i=1}^k \alpha_i m_i^i + \sum_{i=1}^k \sum_{j \neq i} \frac{\alpha_j \lambda_i}{\lambda_j} m_i^j$$

Write  $m_i^i$  as one minus the sum of the other means from the same distribution.

$$\theta = \sum_{i=1}^k \alpha_i \left(1 - \sum_{j \neq i} m_j^i\right) + \sum_{i=1}^k \sum_{j \neq i} \frac{\alpha_j \lambda_i}{\lambda_j} m_i^j$$

Subtract  $\theta$  from both sides and exchange the order of summation of the second sum.

$$0 = - \sum_{i=1}^k \sum_{j \neq i} \alpha_i m_j^i + \sum_{i=1}^k \sum_{j \neq i} \frac{\alpha_i \lambda_j}{\lambda_i} m_j^i$$

Combine sums and simplify.

$$0 = \sum_{i=1}^k \sum_{j \neq i} \frac{\alpha_i}{\lambda_i} m_j^i (\lambda_j - \lambda_i)$$

Rewrite, combining terms for the same  $i$  and  $j$ .

$$0 = \sum_{i=1}^{k-1} \sum_{j=i+1}^k \left[ \left( \frac{\alpha_i}{\lambda_i} m_j^i - \frac{\alpha_j}{\lambda_j} m_i^j \right) (\lambda_j - \lambda_i) \right]$$

If  $k = 2$  and  $\lambda_1 \neq \lambda_2$ , then

$$\frac{\alpha_1}{\lambda_1} m_2^1 = \frac{\alpha_2}{\lambda_2} m_1^2$$

which proves the proposition in this case. However, for  $k > 2$ , the derivation shows that the entire sum equals zero. However, simulation suggests that each term in the sum equals zero, or that

$$\frac{m_i^j}{m_j^i} = \frac{\alpha_i / \lambda_j}{\alpha_j / \lambda_i}$$

More information beyond the sum constraints  $\sum_{i=1}^k m_i^j = 1$  for  $j = 0, 1, \dots, k$  and the recursion equation is needed to demonstrate this last equation.

### Open Problem 2

Prove the result for the ratios of means for general  $k$ .

### Utility of the proposition

The ratio proposition is useful to show that when the  $\alpha$  vector is large, then  $E(X_j) \approx (\alpha_j / \lambda_j) / \sum_i (\alpha_i / \lambda_i)$ . The rough argument is as follows, using the concise notation.

$$\frac{m_i^0}{m_j^0} = \frac{m_i^0}{m_i^j} \times \frac{m_i^j}{m_j^j} \times \frac{m_j^j}{m_j^0}$$

When the  $\alpha$  vector is large, adding one to one of the elements does not change the gamma density much at all and so we expect the first and last ratios on the right hand expression above to each be close to one. If the proposition is true, then  $m_j^0$  is nearly proportional to  $\alpha_j/\lambda_j$  and the constraint that the sum of these over  $j$  equals one leads to the approximation.

### Open Problem 3

State the approximation rigorously and find bounds on the error in the approximation as a function of the vectors  $\alpha$  and  $\lambda$ .

## A closed-form solution in the case that $k = 2$

There is an expression for the marginal mean in the  $k = 2$  case that uses the hypergeometric function. First, to further simplify the notation, let

$$G(\alpha) = \frac{\Gamma(\theta)}{\Gamma(\alpha)} = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)}$$

for vector argument  $\alpha$ . In addition, as the density is identical up to scalar multiples of the  $\lambda$  parameter vector, I also introduce the notation  $r_j = \lambda_1/\lambda_j$  so that the density of the generalized Dirichlet distribution may be rewritten as follows.

$$f(x_1, \dots, x_k) = G(\alpha) \prod_{i=2}^k r_i^{-\alpha_i} \frac{\prod_{i=1}^k x_i^{\alpha_i-1}}{(\sum_{i=1}^k x_i/r_i)^{\sum_{i=1}^k \alpha_i}}, \quad \text{where } x_i > 0 \text{ for all } i \text{ and } \sum_{i=1}^k x_i = 1$$

Here, assume that all of the  $\lambda_i$  are distinct. Before continuing, we need to present the hypergeometric function and some identities.

### The Hypergeometric Function and Some Identities

The hypergeometric function  ${}_2F_1(a, b; c; t)$  is defined as

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}$$

where the rising factorial  $(x)_n = x(x+1) \cdots (x+n-1)$  when  $n > 0$  and  $(x)_0 = 1$ . Note that the 2 and 1 refer to the number of rising factorial factors in the numerator and denominator. There exist more general hypergeometric functions  ${}_pF_q$  for  $p > q$  that have  $p$  rising factorials in the numerator and  $q$  in the denominator, but we only are using the basic  ${}_2F_1$  version here, which we will just write as  $F$ . Many common functions are special cases of hypergeometric functions or are closely related, such as  $\ln(1+z) = zF(1, 1; 2, -z)$ .

By definition, it is obvious that  $F(a, b; c; z) = F(b, a; c; z)$ . The notation suggests that the two parameters  $a$  and  $b$  before the first semicolon correspond to the rising factorials in the numerator, the one parameter after the first semicolon corresponds to the rising factorial in the denominator, and the last value  $z$  is the argument of the function. Note that if  $a$  or  $b$  are nonpositive integers, then the hypergeometric function has only a finite number of terms. However, in our application,  $a$  and  $b$  will be positive.

Another key identity is the following integral representation.

$$B(b, c-b)F(a, b; c; z) = \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-zt)^{-a} dt$$

where  $B$  is the Beta function (reciprocal of my  $G$  function in the  $k = 2$  case). I find it useful to show this identity relative to the exponents of the integral expression.

$$\int_0^1 t^{a-1} (1-t)^{b-1} (1-zt)^{-c} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} F(a, c; a+b; z)$$

There is a huge literature of identities involving hypergeometric functions that undoubtedly would be useful. Here are a few more identities.

$$\begin{aligned} F(a, b; c; z) &= (1-z)^{c-a-b} F(c-a, c-b; c; z) \\ F(a, b; c; z) &= (1-z)^{-b} F(c-a, b; c; \frac{z}{z-1}) \\ F(a, b; c; z) &= (1-z)^{-a} F(a, c-b; c; \frac{z}{z-1}) \end{aligned}$$

## The $k = 2$ case

Our generalized Dirichlet distribution reduces to the standard Dirichlet when  $\lambda_1 = \lambda_2$ . Assume here that  $\lambda_1 > \lambda_2$  and let  $r = \lambda_1/\lambda_2 > 1$ . Also, let  $\alpha = (a, b)$ . Although I will not assume that these  $\alpha$  values are integers, there are simpler expressions for the moments when they are integer valued. I will calculate  $E(X_1^m)$  (and drop the subscript) as  $E(X_2) = 1 - E(X_1)$ . In the  $k = 2$  case,  $G(\alpha) = 1/B(a, b)$  where  $B$  is the beta function.

$$E(X^m) = \int_0^1 x^m f(x|a, b, r) dx = G(\alpha) r^{-b} \int_0^1 x^{a+m-1} (1-x)^{b-1} (x + (1-x)/r)^{-(a+b)} dx$$

Note that

$$(x + (1-x)/r) = r^{-1}(1 - (1-r)x)$$

and thus

$$E(X^m) = G(\alpha) r^a \int_0^1 x^{a+m-1} (1-x)^{b-1} (1 - (1-r)x)^{-(a+b)} dx$$

Using the integral representation of  $F$  and rewriting the  $G$  and Beta functions as fractions of Gamma functions yields

$$E(X^m) = r^a \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+m)\Gamma(b)}{\Gamma(a+b+m)} F(a+m, a+b, a+b+m, 1-r)$$

After some simplification and reordering, we have the following.

$$E(X^m) = r^a \frac{\Gamma(a+b)}{\Gamma(a+b+m)} \frac{\Gamma(a+m)}{\Gamma(a)} F(a+m, a+b, a+b+m, 1-r)$$

Using an identity from above, this simplifies to the following.

$$E(X^m) = \frac{\Gamma(a+b)}{\Gamma(a+b+m)} \frac{\Gamma(a+m)}{\Gamma(a)} F(b, m, a+b+m, 1-r)$$

The mean is

$$E(X) = \frac{a}{a+b} F(b, 1; a+b+1; 1-r)$$

The series representation is the following.

$$E(X) = \frac{a}{a+b} \sum_{n=0}^{\infty} \frac{(b)_n}{(a+b+1)_n} (1-r)^n$$

In the original parametrization, this is

$$E(X_1) = \frac{\alpha_1}{\alpha_1 + \alpha_2} F\left(\alpha_2, 1, \alpha_1 + \alpha_2 + 1, 1 - \frac{\lambda_1}{\lambda_2}\right) = \frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{\lambda_1}{\lambda_2} F\left(\alpha_1 + 1, 1, \alpha_1 + \alpha_2 + 1, \frac{\lambda_1 - \lambda_2}{\lambda_1}\right)$$

Note that the hypergeometric function  $F$  is only guaranteed to converge when the magnitude of the argument is less than one. So, if  $\lambda_1 < \lambda_2$ , the first expression is okay, but when  $\lambda_1 > \lambda_2$ , the second argument is okay.

The derivation assumes  $\lambda_1 > \lambda_2$ , but some steps in between may need to be changed to ensure this is always the case.

We should find that the formula for the other mean and this expression sum to one. Although I specified  $r > 1$ , I really only used  $r \neq 1$ . When considering  $X_2$ ,  $r \rightarrow 1/r$  and  $a \leftrightarrow b$ .

$$E(X_2) = \frac{b}{a+b} F(a, 1; a+b+1; \frac{r-1}{r})$$

Using one of the previous identities, this simplifies to

$$\frac{br}{a+b} F(b+1, 1; a+b+1; 1-r)$$

The sum of the two mean expressions is

$$\frac{a}{a+b} F(b, 1; a+b+1; 1-r) + \frac{br}{a+b} F(b+1, 1; a+b+1; 1-r)$$

#### Open Problem 4

Use identities for consecutive neighbors of the hypergeometric function to show that the previous expression actually is equal to one.

Here is a numerical check:

```
require(hypergeo)
gdmean1 = function(p)
{
  if ( nrow(p) != 2 )
    stop("k must be equal to 2")
  a = p$alpha[1]
  b = p$alpha[2]
  r = p$lambda[1]/p$lambda[2]
  return(Re(a*hypergeo(b,1,a+b+1,1-r)/(a+b)))
}

gdmean2 = function(p)
{
  if ( nrow(p) != 2 )
    stop("k must be equal to 2")
  a = p$alpha[2]
  b = p$alpha[1]
  r = p$lambda[2]/p$lambda[1]
  return(Re(a*hypergeo(b,1,a+b+1,1-r)/(a+b)))
}

gdmean2.2 = function(p)
{
  if ( nrow(p) != 2 )
    stop("k must be equal to 2")
  a = p$alpha[1]
  b = p$alpha[2]
  r = p$lambda[1]/p$lambda[2]
  return(Re(b*r*hypergeo(b+1,1,a+b+1,1-r)/(a+b)))
}

parameters = function(k=4,mean.a=2000,shape=2)
```

```

{
  alpha = rgamma(k,shape,shape/mean.a)
  lambda = rgamma(k,shape,shape)
  lambda = lambda/mean(lambda)
  return(data.frame(alpha,lambda))
}

generate = function(p,n=1000)
{
  k = nrow(p)
  x = with(p, matrix(rgamma(n*k,rep(alpha,each=n),rep(lambda,each=n)),n,k) )
  s = apply(x,1,sum)
  for ( i in 1:n )
    x[i,] = x[i,] / s[i]
  return(x)
}

test.gdmean = function()
{
  p = parameters(2,mean.a=5,shape=2)
  x = generate(p,10^5)
  print(apply(x,2,mean))
  m1 = gdmean1(p)
  m2 = gdmean2(p)
  m3 = gdmean2.2(p)
  print(c(m1,m2,m1+m2,m3,m1+m3))
  return(invisible(c(m1,m2)))
}

for ( i in 1:10 )
{
  print(i)
  test.gdmean()
}

```

```

## [1] 1
## [1] 0.5920832 0.4079168
## [1] 0.5914889 0.4085106 0.9999995 0.4085111 1.0000000
## [1] 2
## [1] 0.6806991 0.3193009
## [1] 0.6801019 0.3198981 1.0000000 0.3198981 1.0000000
## [1] 3
## [1] 0.5592751 0.4407249
## [1] 0.5602166 0.4397834 1.0000000 0.4397834 1.0000000
## [1] 4
## [1] 0.4968299 0.5031701
## [1] 0.4969854 0.5030146 1.0000000 0.5030146 1.0000000
## [1] 5
## [1] 0.1146064 0.8853936
## [1] 0.1147905 0.8852095 1.0000000 0.8852095 1.0000000
## [1] 6
## [1] 0.2818893 0.7181107
## [1] 0.2819855 0.7180145 1.0000001 0.7180145 1.0000000
## [1] 7

```



```

## [1] 0.2748148 0.7251852
## [1] 0.2745872 0.7254128 1.0000000 0.7254128 1.0000000
## [1] 8
## [1] 0.7567082 0.2432918
## [1] 0.7561833 0.2438167 1.0000000 0.2438167 1.0000000
## [1] 9
## [1] 0.3944256 0.6055744
## [1] 0.3946042 0.6053958 1.0000000 0.6053958 1.0000000
## [1] 10
## [1] 0.1478514 0.8521486
## [1] 0.1476953 0.8523047 1.0000000 0.8523047 1.0000000

```

### $k = 3$ Case

I do not know if there is a closed-form solution using hypergeometric functions when  $k = 3$ . Here, there are two free variables (as  $X_3 = 1 - X_1 - X_2$ ) so the calculus to find  $E(X_1)$  requires a double integral over a simplex. I can solve the inner integral similarly to the  $k = 2$  case after some algebraic manipulation and calculus change of variables. What remains is an integral of something that has pieces of the integral form of the hypergeometric function times a hypergeometric function, evaluated at a constant times the variable of the integral. There are known identities in tables of integrals, but I have not found one that matches exactly what I need, but have found one close enough to make me think that there may be a solution.

Here is the calculus problem to try to solve. As in the  $k = 2$  case, reparametrize the  $\lambda$  vector with  $\lambda_1 > \lambda_2 > \lambda_3$  which implies  $1 < r < s$  where  $r = \lambda_1/\lambda_2$  and  $s = \lambda_1/\lambda_3$ . Also use  $\alpha = (a, b, c)$ . Solve

$$E(X) = \int_0^1 \int_0^{1-x} x \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} r^{-b} s^{-c} \frac{x^{a-1} y^{b-1} (1-x-y)^{c-1}}{(x+y/r + (1-x-y)/s)^{a+b+c}} dy dx$$

I believe this to be a partial solution.

$$E(X) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b+c)} r^c s^{-c} \int_0^1 z^{b+c-1} (1-z)^a (1+(r-1)z)^{-1} F\left(a+b+c, c; b+c; \frac{(s-r)}{s} z\right) dz$$

Here are some more potentially useful identities. The last one is very close, except for the factor  $(s-r)/s$  multiplied by  $z$  in the last argument of  $F$ . A change of variable changes the limits of integration and then also will not match.

$$\int_0^1 x^{\alpha-1} (1-x)^{\beta-1} (1-zx)^{-\gamma} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} F(\gamma, \alpha; \alpha+\beta; z)$$

$$\int_0^u x^{\alpha-1} (u-x)^{\beta-1} (x+z)^{-\gamma} dx = z^{-\gamma} u^{\alpha+\beta-1} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} F(\gamma, \alpha; \alpha+\beta; -u/z)$$

$$\int_0^1 x^{\alpha-1} (1-x)^{\beta-1} (1-zx)^{-\gamma} F(a, b; c; x) dx = (1-z) {}_3F_2(\beta, \gamma, \alpha+\beta-a-b; \alpha+\beta-a, \alpha+\beta-b; \frac{z}{z-1})$$

A partial derivation is in Appendix-2.

### Open Problem 5

Get a simplified expression for  $E(X)$  when  $k = 3$ . If possible, make this closed-form. Check by simulation or other methods that it is correct.

### Open Problem 6

Guess a formula for general  $k$  and show it is true.

## A potentially useful transformation

Define  $\widetilde{X}_j = \frac{\lambda_j X_j}{\sum_{i=1}^k \lambda_i X_i}$ . Then  $\widetilde{X} = (\widetilde{X}_1, \dots, \widetilde{X}_k) \sim \text{Dirichlet}(\alpha)$ .

**Proof** As  $X_j$  has the same distribution as  $Y_j/S$  where  $S = \sum_{i=1}^k Y_i$ , it follows that  $\widetilde{X}_j$  has the same distribution as

$$\frac{\lambda_j Y_j}{S} \bigg/ \sum_{i=1}^k \left( \frac{\lambda_i Y_i}{S} \right) = \frac{\lambda_j Y_j}{\sum_{i=1}^k \lambda_i Y_i}$$

As  $\lambda_j Y_j \sim \text{Gamma}(\alpha_j, 1)$  for all  $j$ , the result follows.

## Parameter Estimation

Suppose that a probability density  $g$  on the  $k$ -dimensional simplex has marginal mean  $\{\mu_i\}$  and marginal variances  $\{v_i\}$ . We wish to find  $\{\alpha_i\}$  and  $\{\lambda_i\}$  so that the generalized Dirichlet distribution in this paper has marginal means and variances that are close to those of the distribution  $g$ .

We do the following.

$$\alpha_i = \frac{\mu_i^2(1 - \mu_i)}{v_i}$$

$$\lambda_i = \frac{\mu_i(1 - \mu_i)}{v_i} \bigg/ \sum_{j=1}^k \frac{\mu_j(1 - \mu_j)}{kv_j}$$

By construction, the mean of the  $\{\lambda_i\}$  is one.

### Open Problem 7

Demonstrate with some theoretical argument, at least when  $\min_i \{\alpha_i\}$  is large, that the estimation scheme above is accurate.

## Appendix

Here is a derivation of the density of the generalized Dirichlet distribution with flexible marginal variances.

The joint density of  $(Y_1, \dots, Y_k)$  where  $Y_i \sim \text{Gamma}(\alpha_i, \lambda_i)$  and are mutually independent is

$$f(y_1, \dots, y_k) = \prod_{i=1}^k \left( \frac{\lambda_i^{\alpha_i}}{\Gamma(\alpha_i)} y_i^{\alpha_i-1} e^{-\lambda_i y_i} \right)$$

Let  $S = \sum_{i=1}^k Y_i$  and  $X_i = Y_i/S$  for  $i = 1, \dots, k$ . Note that  $Y_i = SX_i$  for  $i = 1, \dots, k-1$  and  $Y_k = S(1 - \sum_{i=1}^{k-1} X_i)$ . We find the joint density of  $(S, X_1, \dots, X_{k-1})$ . The Jacobian matrix  $J = \partial(y_1, \dots, y_k)/\partial(x_1, \dots, x_{k-1}, s)$  satisfies

$$J_{ij} = \begin{cases} s & \text{if } i = j, i < k \\ 0 & \text{if } i \neq j, i < k \\ x_j & \text{if } i = k, j < k \\ -s & \text{if } j = k, i < k \\ 1 - \sum_{i=1}^{k-1} x_i & \text{if } i = j = k \end{cases}$$

To determine the determinant, replace the  $k$ th row by itself minus  $x_i/s$  times the  $i$ th row for  $i = 1, \dots, k-1$ , which does not affect the value of the determinant. The resulting  $k$ th row has values 0 in columns  $j = 1, \dots, k-1$

and value 1 in column  $k$  and is a diagonal matrix with diagonal elements  $s$  in the first  $k-1$  rows and 1 in the last row. Thus  $|\det J| = s^{k-1}$ . It follows that the joint density of  $(X_1, \dots, X_{k-1}, S)$  is

$$f(x_1, \dots, x_{k-1}, s) = s^{k-1} \prod_{i=1}^k \left( \frac{\lambda_i^{\alpha_i}}{\Gamma(\alpha_i)} (sx_i)^{\alpha_i-1} e^{-\lambda_i s x_i} \right)$$

where  $x_k = 1 - \sum_{i=1}^{k-1} x_i$ . Rewriting, the joint density is as follows.

$$f(x_1, \dots, x_{k-1}, s) = \prod_{i=1}^k \left( \frac{\lambda_i^{\alpha_i} x_i^{\alpha_i-1}}{\Gamma(\alpha_i)} \right) s^{\sum_{i=1}^k \alpha_i - 1} e^{-\left(\sum_{i=1}^k \lambda_i x_i\right) s}$$

Holding all of the  $x_i$  constant, we recognize the gamma density in  $s$  up to constants and can thus integrate out  $s$  to find the joint density of the  $x_i$ .

$$f(x_1, \dots, x_{k-1}) = \prod_{i=1}^k \left( \frac{\lambda_i^{\alpha_i} x_i^{\alpha_i-1}}{\Gamma(\alpha_i)} \right) \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{(\sum_{i=1}^k \lambda_i x_i)^{\sum_{i=1}^k \alpha_i}}$$

where  $x_k = 1 - \sum_{i=1}^{k-1} x_i$ . Reorganization yields the equation of the density of the generalized Dirichlet distribution.

## Appendix 2

Reparametrize the  $\lambda$  vector with  $\lambda_1 > \lambda_2 > \lambda_3$  which implies  $1 < r < s$  where  $r = \lambda_1/\lambda_2$  and  $s = \lambda_1/\lambda_3$ . Also use  $\alpha = (a, b, c)$ . Solve and find an expression for a general moment.

$$\mathbb{E}(X^m) = \int_0^1 \int_0^{1-x} x^m \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} r^{-b} s^{-c} \frac{x^{a-1} y^{b-1} (1-x-y)^{c-1}}{(x+y/r + (1-x-y)/s)^{a+b+c}} dy dx$$

Factor out constants and rewrite.

$$\mathbb{E}(X^m) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} r^{-b} s^{-c} \int_0^1 x^{a+m-1} \int_0^{1-x} y^{b-1} (1-x-y)^{c-1} \left( \frac{1+(s-1)x}{s} + \frac{s-r}{rs} y \right)^{-(a+b+c)} dy dx$$

Factor out  $(s-r)/(rs)$  from the last factor and simplify.

$$\mathbb{E}(X^m) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} r^{a+c} s^{a+b} (s-r)^{-(a+b+c)} \int_0^1 x^{a+m-1} \int_0^{1-x} y^{b-1} ((1-x)-y)^{c-1} \left( \frac{r(1+(s-1)x)}{s-r} + y \right)^{-(a+b+c)} dy dx$$

Use an identity for the inner integral.

$$\begin{aligned} \mathbb{E}(X^m) &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} r^{a+c} s^{a+b} (s-r)^{-(a+b+c)} \\ &\quad \times \int_0^1 x^{a+m-1} \left( \frac{r(1+(s-1)x)}{s-r} \right)^{-(a+b+c)} (1-x)^{b+c-1} \frac{\Gamma(b)\Gamma(c)}{\Gamma(b+c)} F\left(a+b+c, b; b+c; -\frac{(1-x)(s-r)}{r(1+(s-1)x)}\right) dx \end{aligned}$$

Simplify.

$$\mathbb{E}(X^m) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b+c)} r^{-b} s^{a+b} \int_0^1 x^{a+m-1} (1-x)^{b+c-1} (1+(s-1)x)^{-(a+b+c)} F\left(a+b+c, b; b+c; -\frac{(s-r)(1-x)}{r(1+(s-1)x)}\right) dx$$

Use the identity

$$F(\alpha, \beta; \gamma; z) = (1 - z)^{-\alpha} F\left(\alpha, \gamma - \beta; \gamma; \frac{z}{z - 1}\right)$$

where

$$z = -\frac{(s - r)(1 - x)}{r(1 + (s - 1)x)}, \quad 1 - z = \frac{s(1 + (r - 1)x)}{r(1 + (s - 1)x)}, \quad \frac{z}{z - 1} = \frac{(s - r)(1 - x)}{s(1 + (r - 1)x)}$$

in order to get rid of the awkward negative sign.

$$\begin{aligned} \mathbb{E}(X^m) &= \frac{\Gamma(a + b + c)}{\Gamma(a)\Gamma(b + c)} r^{-b} s^{a+b} \int_0^1 x^{a+m-1} (1 - x)^{b+c-1} (1 + (s - 1)x)^{-(a+b+c)} \\ &\quad \times \left\{ \left( \frac{s(1 + (r - 1)x)}{r(1 + (s - 1)x)} \right)^{-(a+b+c)} F\left(a + b + c, c; b + c; \frac{(s - r)(1 - x)}{s(1 + (r - 1)x)}\right) \right\} dx \end{aligned}$$

Then simplify.

$$\mathbb{E}(X^m) = \frac{\Gamma(a + b + c)}{\Gamma(a)\Gamma(b + c)} r^{a+c} s^{-c} \int_0^1 x^{a+m-1} (1 - x)^{b+c-1} (1 + (r - 1)x)^{-(a+b+c)} F\left(a + b + c, c; b + c; \frac{(s - r)(1 - x)}{s(1 + (r - 1)x)}\right) dx$$

Next, do a change of variable for the integral. Let  $z = (1 - x)/(1 + (r - 1)x)$ . Then

$$dz = -\frac{r}{(1 + (r - 1)x)^2} dx, \quad x = \frac{1 - z}{1 + (r - 1)z}, \quad 1 - x = \frac{rz}{1 + (r - 1)z}, \quad 1 + (r - 1)x = \frac{r}{1 + (r - 1)z}$$

Complete the change of variable where  $z = 1$  when  $x = 0$  and  $z = 0$  when  $x = 1$ , so we can eliminate a negative sign and switch back the limits of the integral.

$$\begin{aligned} \mathbb{E}(X^m) &= \frac{\Gamma(a + b + c)}{\Gamma(a)\Gamma(b + c)} r^{a+c} s^{-c} \int_0^1 \left( \frac{1 - z}{1 + (r - 1)z} \right)^{a+m-1} \left( \frac{rz}{1 + (r - 1)z} \right)^{b+c-1} \left( \frac{r}{1 + (r - 1)z} \right)^{-(a+b+c)+2} \\ &\quad F\left(a + b + c, c; b + c; \frac{(s - r)}{s} z\right) \left( \frac{dz}{r} \right) \end{aligned}$$

Simplify.

$$\mathbb{E}(X^m) = \frac{\Gamma(a + b + c)}{\Gamma(a)\Gamma(b + c)} r^c s^{-c} \int_0^1 z^{b+c-1} (1 - z)^{a+m-1} (1 + (r - 1)z)^{-m} F\left(a + b + c, c; b + c; \frac{(s - r)}{s} z\right) dz$$

If  $m = 0$ , this expression is just integrating the density and should be equal to one.

$$1 = \frac{\Gamma(a + b + c)}{\Gamma(a)\Gamma(b + c)} r^c s^{-c} \int_0^1 z^{b+c-1} (1 - z)^{a-1} F\left(a + b + c, c; b + c; \frac{(s - r)}{s} z\right) dz$$

If  $m = 1$ , this expression is the mean.

$$\mathbb{E}(X) = \frac{\Gamma(a + b + c)}{\Gamma(a)\Gamma(b + c)} r^c s^{-c} \int_0^1 z^{b+c-1} (1 - z)^a (1 + (r - 1)z)^{-1} F\left(a + b + c, c; b + c; \frac{(s - r)}{s} z\right) dz$$

The numerical calculations and simulations below provide evidence that the derivations above are correct (which is a minor miracle given the repeated errors on paper over time).

```

set.seed(25)
alpha = runif(3,1,5)
lambda = rev(sort(runif(3,1,5)))
lambda = 3 * lambda / sum(lambda) ## mean(lambda) = 1
r = lambda[1]/lambda[2]
s = lambda[1]/lambda[3]
B = 1000000
y1 = rgamma(B,alpha[1],lambda[1])
y2 = rgamma(B,alpha[2],lambda[2])
y3 = rgamma(B,alpha[3],lambda[3])
y = cbind(y1,y2,y3)
y.sum = apply(y,1,sum)
x = y
for( i in 1:B )
  x[i,] = x[i,] / y.sum[i]
mx = apply(x,2,mean)
sx = apply(x,2,sd)
cat("alpha = ",alpha,"\n")

## alpha = 2.664474 3.779055 1.595203
cat("lambda = ",lambda,"\n")

## lambda = 1.34403 1.248563 0.4074066
cat("E(X) = ",mx,"\n")

## E(X) = 0.2383242 0.3611444 0.4005314
cat("SD(X) = ",sx,"\n")

## SD(X) = 0.1358071 0.1634416 0.1946092

## numerical integration
z = seq(0,0.999,0.001)
a = alpha[1]
b = alpha[2]
cc = alpha[3]
dx = z[2] - z[1]
## check density integrates to one
num.integral.0 = exp(lgamma(sum(alpha)) - lgamma(a) - lgamma(b+cc)) *
  r^cc * s^(-cc) *
  dx * sum(z^(b+cc-1) * (1-z)^(a-1) * Re(hypergeo(sum(alpha),cc,b+cc,(s-r)*z/s)))
print(num.integral.0)

## [1] 0.9999999

## check E(X_1)
num.integral.1 = exp(lgamma(sum(alpha)) - lgamma(a) - lgamma(b+cc)) *
  r^cc * s^(-cc) * dx *
  sum( z^(b+cc-1) *
    (1-z)^a *
    (1 + (r-1)*z)^(-1) *
    Re(hypergeo(sum(alpha),cc,b+cc,(s-r)*z/s)))
print(mx[1])

##          y1
## 0.2383242

```

```

print(num.integral.1)

## [1] 0.2381038
## check SD(X_1)
num.integral.2 = exp(lgamma(sum(alpha)) - lgamma(a) - lgamma(b+cc)) *
  r^cc * s^(-cc) * dx *
  sum( z^(b+cc-1) *
    (1-z)^(a+1) *
    (1 + (r-1)*z)^(-2) *
    Re(hypergeo(sum(alpha),cc,b+cc,(s-r)*z/s)))
print(sx[1])

##          y1
## 0.1358071
print(sqrt(num.integral.2 - num.integral.1^2))

## [1] 0.1357839

```