Hypergeometric Function and Generalized Dirichlet Moments

Bret Larget

8/3/2017

This document attempts to use the definition of the hypergeometric function to find moments of the generalized Dirichlet distribution for small k.

Generalized Dirichlet Density

As in other documents, here is the joint density of normalized independent $\text{Gamma}(\alpha_i, \lambda_i)$ random variables.

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

In this document, I introduce the notation

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

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

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

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 $(x)_0 = 1$. By definition, it is obvious that ${}_2F_1(a, b; c; z) = {}_2F_1(b, a; c; z)$. 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. A generalization is the ${}_pF_q$ function which is specified by p rising factorials in the numerator and q rising factorials in the denominator. I hope not to need to use this generalization, but Mathematica may have other ideas.

Another key identity is the following integral representation.

$$B(b, c-b)\,{}_2F_1(a, b; c; z) = \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-zt)^{-a}\,\mathrm{d}t$$

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.

$$\int_0^1 t^{a-1}(1-t)^{b-1}(1-zt)^{-c}\,\mathrm{d}t = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\,{}_2F_1(a,c;a+b;z)$$

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

$$\begin{aligned}
{}_2F_1(a,b;c;z) &= (1-z)^{c-a-b}\,{}_2F_1(c-a,c-b;c;z) \\
{}_2F_1(a,b;c;z) &= (1-z)^{-b}\,{}_2F_1(c-a,b;c;\frac{z}{z-1}) \\
{}_2F_1(a,b;c;z) &= (1-z)^{-a}\,{}_2F_1(a,c-b;c;\frac{z}{z-1})
\end{aligned}$$

The $k=2$ case

Our generalized Dirichlet distribution reduces to the standard Dirchlet 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 α values are integers, there are simpler expressions for the moments when they are integer valued. I will calculate $\mathsf{E}(X_1^m)$ (and drop the subscript) as $\mathsf{E}(X_2)=1-\mathsf{E}(X_1)$. In the $k=2$ case, $G(\alpha)=1/B(a,b)$ where B is the beta function.

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

Note that

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

and thus

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

Using the integral representation of ${}_2F_1$ and rewriting the G and Beta functions as fractions of Gamma functions yields

$$\mathsf{E}(X^m)=r^a\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+m)\Gamma(b)}{\Gamma(a+b+m)}\,{}_2F_1(a+m,a+b,a+b+m,1-r)$$

After some simplification and reordering, we have the following.

$$\mathsf{E}(X^m)=r^a\frac{\Gamma(a+b)}{\Gamma(a+b+m)}\frac{\Gamma(a+m)}{\Gamma(a)}\,{}_2F_1(a+m,a+b,a+b+m,1-r)$$

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

$$\mathsf{E}(X^m)=\frac{\Gamma(a+b)}{\Gamma(a+b+m)}\frac{\Gamma(a+m)}{\Gamma(a)}\,{}_2F_1(b,m,a+b+m,1-r)$$

The mean is

$$\mathsf{E}(X)=\frac{a}{a+b}\;{}_2F_1(b,1;a+b+1;1-r)$$

The series representation is the following.

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

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, $r \to 1/r$ and $a \leftrightarrow b$.

$$\mathsf{E}(X_2) = \frac{b}{a+b} \; {}_2F_1(a, 1; a+b+1; \frac{r-1}{r})$$

Using one of the previous identities, XXX this simplifies to

$$\frac{br}{a+b} \; {}_2F_1(b+1, 1; a+b+1; 1-r)$$

The sum of the two mean expressions is

$$\frac{a}{a+b} {}_2F_1(b, 1; a+b+1; 1-r) + \frac{br}{a+b} {}_2F_1(b+1, 1; a+b+1; 1-r)$$

which simplifies to

$$\frac{r^a \Big(a \; {}_2F_1(a+1, a+b; a+b+1; 1-r) + b \; {}_2F_1(a, a+b; a+b+1; 1-r) \Big)}{a+b}$$

Mathematica does not claim that this is one, so maybe I made an error somewhere.

Here is a numerical check:

```
require(hypergeo)
```

```
## Loading required package: hypergeo
```

```
gdmean1 = function(p)
{
  if ( nrow(p) != 2 )
    stop("kmust 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("kmust 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)))
}

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)
  print(c(m1,m2,m1+m2))
  return(invisible(c(m1,m2)))
}

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

```
## [1] 1
## [1] 0.2311037 0.7688963
## [1] 0.2312585 0.7687415 1.0000000
## [1] 2
## [1] 0.2486196 0.7513804
## [1] 0.2491654 0.7508346 1.0000000
## [1] 3
## [1] 0.1676132 0.8323868
## [1] 0.167391 0.832609 1.000000
## [1] 4
## [1] 0.6608699 0.3391301
## [1] 0.6604415 0.3395585 1.0000000
## [1] 5
## [1] 0.5207054 0.4792946
## [1] 0.5215139 0.4784861 1.0000000
## [1] 6
## [1] 0.8440775 0.1559225
## [1] 0.8440008 0.1559992 1.0000000
## [1] 7
## [1] 0.6976621 0.3023379
## [1] 0.6969153 0.3030847 1.0000000
## [1] 8
## [1] 0.4586049 0.5413951
## [1] 0.4584186 0.5415814 1.0000000
## [1] 9
## [1] 0.8502458 0.1497542
## [1] 0.8502595 0.1497405 1.0000000
## [1] 10
```

```
## [1] 0.1647252 0.8352748
## [1] 0.1648494 0.8351506 1.0000000
```

So, it seems like the theory is right and we need to find another hypergeometric function identity.