A generalized Dirichlet distribution with flexible marginal variances

Claudia Solís-Lemus, Steve Goldstein, and Bret Larget

7/13/2017

The Dirichlet Distribution

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

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

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

Each random variable X_i has a marginal Beta$(\alpha_i, \theta - \alpha_i)$ distribution where $\theta = \sum_{i=1}^k \alpha_i$. It follows that X_i has mean $\mathsf{E}(X_i) = \alpha_i/\theta$ and variance $\mathsf{Var}(X_i) = \alpha_i(\theta - \alpha_i)/(\theta^2(\theta + 1))$. 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 α_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.

There are multiple alternative generalizations of the Dirichlet distribution. [Cite the Kotz Johnson book and other examples.] The generalization we describe here is a special case of a generalization in a Romanian language publication [citation], which derives the joint density.

Generation of random variables

To generate random variables $X_1, \ldots, X_k \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$, one simply generates independent random variables $Y_i \sim \text{Gamma}(\alpha_i, \lambda)$ for $i = 1, \ldots, k$ and any arbitrary $\lambda > 0$ (typically $\lambda = 1$) and letting $X_i = Y_i / \sum_{j=1}^k Y_j$. This suggests that by allowing the value of λ to vary with i that we may be able to create a distribution on positive random variables constrained to sum to one with the desired flexibility in the first and second moments.

A Generalized Dirichlet Distribution

Define the flexible marginal variance generalized Dirichlet distribution on $X_1, \ldots, X_n$ to be the distribution of $(X_1, \ldots, X_k)$ where $X_i = Y_i / \sum_{j=1}^k Y_j$ for $i = 1, \ldots, k$ 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 (REFERENCES) 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 $X_1, \ldots, X_k)$ does have a closed-form solution.

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

The derivation is shown in the appendix.

Marginal Moments

We have not been able to derive closed form solutions for the marginal means and variances, but when all of the α_i values are large, the marginal means are numerically close to $(\alpha_i/\lambda_i)\big/\sum_{j=1}^k(\alpha_j/\lambda_j)$, which we use for parameter estimation.

We do present the following relationship between marginal means from generalized Dirichlet distributions with common λ vectors and α vectors that differ by one in a single dimension. To simplify notation, we write

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

where $x = (x_1,\ldots,x_k)$, $\alpha = (\alpha_1,\ldots,\alpha_k)$, $\lambda = (\lambda_1,\ldots,\lambda_k)$, $A = \sum_{i=1}^k \alpha_i$, an expression of vectors of the form a^b is short for $\prod_{i=1}^k a_k^{b_i}$, $\Gamma(\alpha)$ 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 ith element equals α_i if $i \neq j$ and equals α_j+1 when $i=j$. The integral expression $\int_\triangle \cdot\, \mathrm{d}x$ refers to integration over the simplex $\sum_{i=1}^k x_i = 1$ and $0 < x_i < 1$ for $i = 1,\ldots,k$. We introduce the notation

$$m_j(\alpha,\lambda) = \mathsf{E}(X_j|\alpha,\lambda) = \int_\triangle x_j f(x|\alpha,\lambda)\,\mathrm{d}x$$

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

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

Integrating over the simplex on both sides shows that

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

For Steve

On July 12, Steve and I thought that $m_j(\alpha,\lambda)$ was proportional to α_j/λ_j. But this is false. Here is the proof.

Suppose that a solution to the previous system of equations takes the form

$$m_j(\alpha,\lambda) = c(\alpha,\lambda)\frac{\alpha_j/\lambda_j}{A}$$

where $c(\alpha,\lambda)$ is a constant which may depend on α and λ. Then

$$\begin{aligned} c(\alpha,\lambda)\frac{\alpha_j/\lambda_j}{A} &= \left(\frac{\alpha_j/\lambda_j}{A}\right)\left(\sum_{i\neq j}\lambda_i c(\alpha^{(j)},\lambda)\frac{\alpha_i/\lambda_i}{A+1} + \lambda_j c(\alpha^{(j)},\lambda)\frac{(\alpha_j+1)/\lambda_j}{A+1}\right) \\ &= c(\alpha^{(j)},\lambda)\frac{\alpha_j/\lambda_j}{A}\frac{\sum_{i=1}^k \alpha_i + 1}{A+1} \\ &= c(\alpha^{(j)},\lambda)\frac{\alpha_j/\lambda_j}{A} \end{aligned}$$

which holds for all j. Thus, if $m_j(\alpha, \lambda)$ has this form, the constant is the same for all distributions where the λ vector is identical and the α vectors differ by integer differences by dimension.

Using the fact that $\sum_{j=1}^k m_j(\alpha) = 1$ allows us to find the value of the constant c.

$$\begin{aligned} 1 &= \sum_{j=1}^k m_j(\alpha) \\ &= \sum_{j=1}^k \frac{c\alpha_j/\lambda_j}{A} \end{aligned}$$

This implies

$$c = \frac{A}{\sum_{j=1}^k (\alpha_j/\lambda_j)}$$

However, this expression is not constant when α changes to $\alpha^{(j)}$ for some j, unless all of the λ_j are equal to each other. Hence, the proposed solution $m_j(\alpha) \propto \alpha_j/\lambda_j$ is not a solution in general. In my simulation studies, the proposed solution is very accurate when the α_j values are all large (in the hundreds or larger). But when the alpha values are small (average a single digit), then the equation has larger error.

Appendix

Here is a derivation of the density of the symmetric generalized Dirichlet distribution.

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

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

Let $S = \sum_{i=1}^k Y_i$ and $X_i = Y_i/S$ for $i = 1, \ldots, k$. Note that $Y_i = SX_i$ for $i = 1, \ldots, k-1$ and $Y_k = S(1 - \sum_{i=1}^{k-1} X_i)$. We find the joint density of $(S, X_1, \ldots, X_{k-1})$. The Jacobian matrix $J = \partial(y_1, \ldots, y_k)/\partial(x_1, \ldots, x_{k-1}, s)$ 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 kth row by itself minus x_i/s times the ith row for $i = 1, \ldots, k-1$, which does not affect the value of the determinant. The resulting kth row has values 0 in columns $j = 1, \ldots, k-1$ 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 $(X_1, \ldots, X_{k-1}, S)$ is

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

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

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

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.

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

where $x_k = 1 - \sum_{i=1}^{k-1} x_i$. Reorganization yields the equation at the bottom of page 1.