Stat 5101 Lecture Slides: Deck 8
Dirichlet Distribution

Charles J. Geyer
School of Statistics
University of Minnesota

The Dirichlet Distribution

The Dirichlet Distribution is to the beta distribution as the multinomial distribution is to the binomial distribution.

We get it by the same process that we got to the beta distribution (slides 128–137, deck 3), only multivariate.

Recall the basic theorem about gamma and beta (same slides referenced above).

The Dirichlet Distribution (cont.)

Theorem 1. Suppose X and Y are independent gamma random variables

$$
\begin{aligned}
X &\sim \text{Gam}(\alpha_1, \lambda) \\
Y &\sim \text{Gam}(\alpha_2, \lambda)
\end{aligned}
$$

then

$$
\begin{aligned}
U &= X + Y \\
V &= X/(X + Y)
\end{aligned}
$$

are independent random variables and

$$
\begin{aligned}
U &\sim \text{Gam}(\alpha_1 + \alpha_2, \lambda) \\
V &\sim \text{Beta}(\alpha_1, \alpha_2)
\end{aligned}
$$

The Dirichlet Distribution (cont.)

Corollary 1. Suppose X_1, X_2, $\ldots$, are are independent gamma random variables with the same shape parameters

$$X_i \sim \text{Gam}(\alpha_i, \lambda)$$

then the following random variables

$$\begin{aligned}
\frac{X_1}{X_1 + X_2} &\sim \text{Beta}(\alpha_1, \alpha_2) \\
\frac{X_1 + X_2}{X_1 + X_2 + X_3} &\sim \text{Beta}(\alpha_1 + \alpha_2, \alpha_3) \\
&\vdots \\
\frac{X_1 + \cdots + X_{d-1}}{X_1 + \cdots + X_d} &\sim \text{Beta}(\alpha_1 + \cdots + \alpha_{d-1}, \alpha_d)
\end{aligned}$$

are independent and have the asserted distributions.

The Dirichlet Distribution (cont.)

From the first assertion of the theorem we know

$$X_1 + \cdots + X_{k-1} \sim \text{Gam}(\alpha_1 + \cdots + \alpha_{k-1}, \lambda)$$

and is independent of X_k. Thus the second assertion of the theorem says

$$\frac{X_1 + \cdots + X_{k-1}}{X_1 + \cdots + X_k} \sim \text{Beta}(\alpha_1 + \cdots + \alpha_{k-1}, \alpha_k) \qquad (*)$$

and $(*)$ is independent of $X_1 + \cdots + X_k$.

That proves the corollary.

The Dirichlet Distribution (cont.)

Theorem 2. Suppose X_1, X_2, ..., are as in the Corollary. Then the random variables

$$Y_i = \frac{X_i}{X_1 + \ldots + X_d}$$

satisfy

$$\sum_{i=1}^{d} Y_i = 1, \qquad \text{almost surely.}$$

and the joint density of Y_2, ..., Y_d is

$$f(y_2, \ldots, y_d) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_d)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_d)} (1 - y_2 - \cdots - y_d)^{\alpha_1 - 1} \prod_{i=2}^{d} y_i^{\alpha_i - 1}$$

The Dirichlet distribution with parameter vector $(\alpha_1, \ldots, \alpha_d)$ is the distribution of the random vector $(Y_1, \ldots, Y_d)$.

The Dirichlet Distribution (cont.)

Let the random variables in Corollary 1 be denoted $W_2, \ldots, W_d$ so these are independent and

$$W_i = \frac{X_1 + \cdots + X_{i-1}}{X_1 + \cdots + X_i} \sim \text{Beta}(\alpha_1 + \cdots + \alpha_{i-1}, \alpha_i)$$

Then

$$Y_i = (1 - W_i) \prod_{j=i+1}^{d} W_j, \qquad i = 2, \ldots, d,$$

where in the case $i = d$ we use the convention that the product is empty and equal to one.

The Dirichlet Distribution (cont.)

$$Y_i = \frac{X_i}{X_1 + \ldots + X_d}$$
$$W_i = \frac{X_1 + \cdots + X_{i-1}}{X_1 + \cdots + X_i}$$

The inverse transformation is

$$W_i = \frac{Y_1 + \cdots + Y_{i-1}}{Y_1 + \cdots + Y_i} = \frac{1 - Y_i - \cdots - Y_d}{1 - Y_{i+1} - \cdots - Y_d}, \qquad i = 2, \ldots, d,$$

where in the case $i = d$ we use the convention that the the sum in the denominator of the fraction on the right is empty and equal to zero, so the denominator itself is equal to one.

The Dirichlet Distribution (cont.)

$$w_i = \frac{1 - y_i - \cdots - y_d}{1 - y_{i+1} - \cdots - y_d}$$

This transformation has components of the Jacobian matrix

$$\frac{\partial w_i}{\partial y_i} = -\frac{1}{1 - y_{i+1} - \cdots - y_d}$$

$$\frac{\partial w_i}{\partial y_j} = 0, \qquad j < i$$

$$\frac{\partial w_i}{\partial y_j} = -\frac{1}{1 - y_{i+1} - \cdots - y_d} + \frac{1 - y_i - \cdots - y_d}{(1 - y_{i+1} - \cdots - y_d)^2}, \qquad j > i$$

The Dirichlet Distribution (cont.)

Since this Jacobian matrix is triangular, the determinant is the product of the diagonal elements

$$|\det \nabla h(y_2, \ldots, y_d)| = \prod_{i=2}^{d-1} \frac{1}{1 - y_{i+1} - \cdots - y_d}.$$

The Dirichlet Distribution (cont.)

The joint density of W_2, ..., W_d is

$$\prod_{i=2}^{d} \frac{\Gamma(\alpha_1 + \cdots + \alpha_i)}{\Gamma(\alpha_1 + \cdots + \alpha_{i-1})\Gamma(\alpha_i)} w_i^{\alpha_1 + \cdots + \alpha_{i-1} - 1}(1 - w_i)^{\alpha_i - 1}$$

$$= \frac{\Gamma(\alpha_1 + \cdots + \alpha_d)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_d)} \prod_{i=2}^{d} w_i^{\alpha_1 + \cdots + \alpha_{i-1} - 1}(1 - w_i)^{\alpha_i - 1}$$

The Dirichlet Distribution (cont.)

PMF of W's	$\frac{\Gamma(\alpha_1+\cdots+\alpha_d)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_d)} \prod_{i=2}^d w_i^{\alpha_1+\cdots+\alpha_{i-1}-1}(1-w_i)^{\alpha_i-1}$
Jacobian	$\prod_{i=2}^{d-1} \frac{1}{1-y_{i+1}-\cdots-y_d}$
transformation	$w_i = \frac{1-y_i-\cdots-y_d}{1-y_{i+1}-\cdots-y_d}$

The PMF of Y_2, ..., Y_d is

$$
\begin{aligned}
&\frac{\Gamma(\alpha_1+\cdots+\alpha_d)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_d)} \prod_{i=2}^{d} \frac{(1-y_i-\cdots-y_d)^{\alpha_1+\cdots+\alpha_{i-1}-1} y_i^{\alpha_i-1}}{(1-y_{i+1}-\cdots-y_d)^{\alpha_1+\cdots+\alpha_i-1}} \\
&\qquad = \frac{\Gamma(\alpha_1+\cdots+\alpha_d)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_d)} (1-y_2-\cdots-y_d)^{\alpha_1-1} \prod_{i=2}^{d} y_i^{\alpha_i-1}
\end{aligned}
$$

Univariate Marginals

Write $I = \{1, \ldots, d\}$. By definition

$$Y_i = \frac{X_i}{X_1 + \ldots + X_d}$$

has distribution the beta distribution with parameters α_i and

$$\sum_{\substack{j \in I \\ j \neq i}} \alpha_j$$

by Theorem 1 because

$$X_i \sim \text{Gam}(\alpha_i, \lambda)$$

$$\sum_{\substack{j \in I \\ j \neq i}} X_j \sim \text{Gam}\left(\sum_{\substack{j \in I \\ j \neq i}} \alpha_j, \lambda \right)$$

Multivariate Marginals

Multivariate Marginals are "almost" Dirichlet.

As was the case with the multinomial, if we collapse categories, we get a Dirichlet. Let $\mathcal{A}$ be a partition of I, and define

$$\begin{aligned} Z_A &= \sum_{i \in A} Y_i, \qquad && A \in \mathcal{A}. \\ \beta_A &= \sum_{i \in A} \alpha_i, \qquad && A \in \mathcal{A}. \end{aligned}$$

Then the random vector having components Z_A has the Dirichlet distribution with parameters β_A.

Conditionals

$$Y_i = \frac{X_i}{X_1 + \ldots + X_d}$$

$$\begin{aligned} Y_i &= \frac{X_i}{X_1 + \cdots + X_k} \cdot \frac{X_1 + \cdots + X_k}{X_1 + \cdots + X_d} \\ &= \frac{X_i}{X_1 + \cdots + X_k} \cdot (Y_1 + \cdots + Y_k) \\ &= \frac{X_i}{X_1 + \cdots + X_k} \cdot (1 - Y_{k+1} - \cdots - Y_d) \end{aligned}$$

Conditionals (cont.)

$$Y_i = \frac{X_i}{X_1 + \cdots + X_k} \cdot (1 - Y_{k+1} - \cdots - Y_d)$$

When we condition on Y_{k+1}, ..., Y_d, the second term above is a constant and the first term a component of another Dirichlet random vector having components

$$Z_i = \frac{X_i}{X_1 + \cdots + X_k}, \qquad i = 1, \ldots, k$$

So conditionals of Dirichlet are constant times Dirichlet.

Moments

From the marginals being beta, we have

$$
\begin{aligned}
E(Y_i) &= \frac{\alpha_i}{\alpha_1 + \cdots + \alpha_d} \\
\operatorname{var}(Y_i) &= \frac{\alpha_i}{(\alpha_1 + \cdots + \alpha_d)^2 (\alpha_1 + \cdots + \alpha_d + 1)} \sum_{\substack{j \in I \\ j \neq i}} \alpha_j
\end{aligned}
$$

Moments (cont.)

From the PMF we get the "theorem associated with the Dirichlet distribution."

$$\int \cdots \int (1 - y_2 - \cdots - y_d)^{\alpha_1 - 1} \left(\prod_{i=2}^{d} y_i^{\alpha_i - 1} \right) dy_2 \cdots dy_d$$
$$= \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_d)}{\Gamma(\alpha_1 + \cdots + \alpha_d)}$$

so

$$\begin{aligned} E(Y_1 Y_2) &= \frac{\Gamma(\alpha_1 + 1)\Gamma(\alpha_2 + 1)\Gamma(\alpha_3) \cdots \Gamma(\alpha_d)}{\Gamma(\alpha_1 + \cdots + \alpha_d + 2)} \cdot \frac{\Gamma(\alpha_1 + \cdots + \alpha_d)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_d)} \\ &= \frac{\alpha_1 \alpha_2}{(\alpha_1 + \cdots + \alpha_d + 1)(\alpha_1 + \cdots + \alpha_d)} \end{aligned}$$

Moments (cont.)

The result on the preceding slide holds when 1 and 2 are replaced by i and j for $i \neq j$, and

$$
\begin{aligned}
\operatorname{cov}(Y_i, Y_j) &= E(Y_i Y_j) - E(Y_i)E(Y_j) \\
&= \frac{\alpha_i \alpha_j}{(\alpha_1 + \cdots + \alpha_d + 1)(\alpha_1 + \cdots + \alpha_d)} - \frac{\alpha_i \alpha_j}{(\alpha_1 + \cdots + \alpha_d)^2} \\
&= \frac{\alpha_i \alpha_j}{\alpha_1 + \cdots + \alpha_d} \left[\frac{1}{\alpha_1 + \cdots + \alpha_d + 1} - \frac{1}{\alpha_1 + \cdots + \alpha_d} \right] \\
&= - \frac{\alpha_i \alpha_j}{(\alpha_1 + \cdots + \alpha_d)^2 (\alpha_1 + \cdots + \alpha_d + 1)}
\end{aligned}
$$