Bayesian Analysis (0000) **00**, Number 0, pp. 1

A generalized Dirichlet distribution with scaled variances

Bret Larget and Claudia Solís-Lemus

Abstract. xxx

MSC 2010 subject classifications: Primary 60K35, 60K35; secondary 60K35.

Keywords: xxx, xxx.

1 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

$$f(x_1, \ldots, x_k) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i - 1}, \qquad \text{where } x_i > 0 \text{ for all } i \text{ and } \sum_{i=1}^k x_i = 1$$

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 match the distribution of a given set of 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 factor which determines all of the marginal variances. We seek a generalization with variation of scale factors in the variances, and with a larger parameterization that is flexible enough to match, at least approximately, the means and variances of each marginal distribution.

We know of another generalization of the Dirichlet distribution (described HERE) that is different than what we propose here in that it is asymmetric in the indices of the random variables and has the property that some correlations may be positive.

1.1 Generation of random variables

To generate random variables $X_1, \ldots, X_k \sim$ Dirichlet$(\alpha_1, \ldots, \alpha_k)$, one simply generates independent random variables $Y_i \sim$ Gamma(α_i, λ) 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

?? Department of Botany 132 Birge Hall, 430 Lincoln Drive, Madison, WI 53706 ?? ??

 DOI: 0000

random variables constrained to sum to one with the desired flexibility in the first and second moments.

2 A Generalized Dirichlet Distribution

Define the symmetric 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 $Y_i \sim \text{Gamma}(\alpha_i, \lambda_i)$ (see also Craiu and Craiu). 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. (CHECK IF SETTING MEAN OF $1/\lambda_i$ TO BE ONE IS ANY MORE CONVENIENT).

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. However, the joint density of $X_1, \ldots, X_k)$ has 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.

I have not been able to derive closed form solutions for the marginal means and variances, but the means are close (if not exactly equal to) $(\alpha_i/\lambda_i) / \sum_{j=1}^k (\alpha_j/\lambda_j)$.

2.1 Parameter Estimation

Suppose that a probability density g on the k-dimensional simplex has marginal mean $\{\mu_i\}$ and marginal variances $\{v_i\}$. We do the following.

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

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

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

I need to provide some more theoretical evidence that these parameter estimates work.

3 Simulations

simulate generalized dirichlet, pick alpha and lambda based on mean/variance, show that it is a good fit

To do:

- justify moments
- what are marginals?
- does this work for small alpha (only tested with big alpha)?
- we need to show that the mean/variance exist, or give boundaries. also, marginal?
- there are possible constraints for what the variances can be given a set of means
- even if we do not have expressions for moments, can we show that the density is unimodal and is there an expression for the marginal mode?

Acknowledgments

xxx