9th National Convention on Statistics (NCS)
EDSA Shangri-La Hotel
October 4-5, 2004

Generalization of the Dirichlet Distribution

by
Epimaco A. Cabanlit, Jr.
Roberto N. Padua
Khursheed Alam

For additional information, please contact:

Author's name:	Epimaco A. Cabanlit, Jr.
Designation:	Director
Agency:	Research and Development Center Mindanao State University
Address:	General Santos City
Telefax:	(083) 380-7167
E-mail:	maco_727 @ yahoo.com

Generalization of the Dirichlet Distribution

by

Epimaco A. Cabanlit, Jr.[1]
Roberto N. Padua[2]
Khursheed Alam[3]

ABSTRACT

The DirIchlet distribution is a generalization of the beta distribution. In Bayesian analysis the Dirichlet distribution is used as a conjugate prior distribution for the parameters of a multinomial distribution. However, the dirichlet family is not sufficiently rich in scope to represent many important distributional assumptions, because the DirIchlet distribution has few number of parameters. We provide a generalization of the DirIchlet distribution with added number of parameters.

Keywords: Dirichlet Distribution, Multinomial Distribution, Bayesian Analysis

I. Introduction

The Dirichlet distribution finds many uses in practical situations apart from its inherently interesting theoretical properties. Its main attraction is the fact that it is often used as a conjugate prior to the multinomial distribution in Bayesian analysis. It is well known, for instance, that using appropriate conjugate priors result in easy computation of Bayesian estimates of parameters (Lehmann, 1984).

The distribution may be thought of as a natural generalization of the beta distribution. The beta distribution is given by:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \text{ , } 0 < x < 1, a > 0, b>0. \qquad (1.1)$$

The parameters a and b determine the shape of the distribution. Note, in particular, that if a = 1 and b = 1, then we obtain the uniform distribution on the interval [0,1]. Higher values of a and b result in high peaked distributions on the interval. The variety of shapes that the beta distribution can take when a and b change makes this particular distribution quite interesting as a prior conjugate distribution to the binomial distribution. The magnitude of a and b can

[1] Director, Research and Development Center, Mindanao State University, General Santos City
[2] Vice President for Research, Extension and International Affairs, Mindanao Polytechnic State College, Cagayan de Oro City
[3] Professor, Clemson University, South Carolina, USA

accommodate a host of the Bayesian's belief of the actual shape of the distribution of the parameters.

It can be noted that defining x over the interval (0,1) does not restrict its use. If $c \le x \le d$, then $x^* = (x-c)/(d-c)$ will define a new variable defined on the interval (0,1).

A natural way to extend the beta distribution is to consider a vector $(z_1, \ldots z_K)$ distributed as:

$$f(z_1, \ldots z_K) = C z_1^{\lambda_1 - 1} \cdots z_K^{\lambda_K} (1 - \sum_{i=1}^{K} z_i)^{\lambda_{K+1} - 1} \tag{1.2}$$

where the z_i's range from 0 to 1 such that $\sum_{i=1}^{K+1} z_i$ =1 and $\lambda_i > 0$ for $i = 1, \ldots, K+1$.

The constant C can be found by actual integration and is given by:

$$C = \frac{\Gamma(\lambda)}{\Gamma(\lambda_1) \cdots \Gamma(\lambda_{K+1})}, \text{ where } \lambda = \sum_{i=1}^{K+1} \lambda_i \tag{1.3}$$

The distribution given above is called the Dirichlet distribution with k+1 parameters and reduces to the beta distribution if k = 1. Just as the beta distribution is the conjugate prior of the binomial distribution, the Dirichlet distribution is the conjugate prior of the multinomial distribution.

Consider $U_1, \ldots, U_{K+1}$ which are independent Gamma random variables with shape parameters $\lambda_1, \ldots, \lambda_{K+1}$ and identical scale parameter β. Let

$$z_i = \frac{U_i}{\sum_{i=1}^{K+1} U_i}, \tag{1.4}$$

then the joint distribution of the z_i's is given by (1.2), (Wilks (1962), Sec 7.7). Thus, the Dirichlet distribution appears naturally as the distribution of normalized (by the sum) gamma random variables. The denominator of (1.4) is seen immediately as having a gamma distribution with parameters $\sum_{i=1}^{K+1} \lambda_i$ and β.

II. Importance of the Generalization of the Dirichlet distribution

In equation (1.2), we observed that the number of parameters for the K-tuples is K+1. The said parameters are called the shape parameters. These parameters explain the shape of the distribution. To illustrate it clearly, we consider the case when k=1. When $\lambda_1 = \lambda_2 = 1$, we have the uniform distribution on (0,1). When $\lambda_1 = \lambda_2$, $\lambda_1 > 1$ and $\lambda_2 > 1$, we have a function that is similar to the normal distribution on the interval (0,1). As $\lambda_1 = \lambda_2 \to \infty$, the distribution tends to $N(\mu, \sigma^2)$, the normal distribution with mean μ and variance σ^2. When $\lambda_1 < 1$ and $\lambda_2 < 1$, the distribution (1.1) behaves as a function that concaves upward on (0,1). When $\lambda_1 < 1$ and $\lambda_2 > 1$, we have a decreasing function from 0 to 1. Lastly, when $\lambda_1 > 1$ and $\lambda_2 < 1$, we get an increasing function from 0 to 1.

For $K > 1$, we can see that the distribution varies its shapes for different values and restrictions of λ_i. By introducing another set of parameters, more types of probability models are obtained that are also useful.

Another reason for generalizing the distribution is explained below.

The mean, variance and covariance of z_i in (1.2) are given by $Ez_i = \lambda_i / \lambda$

$$\mathrm{Var}(z_i) = \lambda_i(\lambda - \lambda_i)/(\lambda^2(\lambda+1) \tag{2.1}$$

and

$$\mathrm{Cov}(z_i, z_j) = -\lambda_i\lambda_j/(\lambda^2(\lambda+1)). \tag{2.2}$$

(See O'hagan, 1994). It is seen that the relative magnitude of the λ_i's determine the mean of z_i. The overall magnitude λ represents the strength of the prior information when the Dirichlet distribution is used as a prior distribution for the parameters of the multinomial distribution, because the variance of z_i decreases as λ increases, keeping the value of λ_i / λ fixed. There is no other parameter left to provide for important aspects of the prior belief. For example, the correlation between z_i and z_j is completely specified by the values of $E z_i$ and $E z_j$. Thus, we see that the Dirichlet family can represent only a limited range of prior belief. In view of this difficulty, several authors have proposed certain generalizations of the Dirichlet distribution (see Albert and Gupta(1982), Dickey(1983) and Leonard and Novick(1986)).

In the next section, we provide a new generalization of the Dirichlet distribution by introducing a scale parameter. By introducing a new set of scale parameters, we can make the Dirichlet distribution more flexible and it can be used to model different real life situations and phenomena.

III. Generalization of the Dirichlet distribution

In this study we provide a new generalization of the Dirichlet distribution, given by the distribution of $\underset{\sim}{w} = (w_1, \ldots, w_{K+1})$, where

$$w_i = (\theta_i z_i)/(\sum_{i=1}^{K+1} \theta_i z_i) \tag{3.1}$$

$\theta_i > 0$. We can assume without loss of generality that $0 < \theta_i < 1$, $i = 1, \ldots, K+1$. Note that $\sum_{i=1}^{K+1} w_i = 1$ and that $\underset{\sim}{w} = \underset{\sim}{z}$ for $\theta_1 = \theta_2 = \cdots = \theta_{K+1} = \theta$. The distribution of $\underset{\sim}{w}$ contains 2K+3 parameters including the K+2 parameters of (1.2). Therefore the family of distributions represented by $\underset{\sim}{w}$ is more flexible and richer than the family of Dirichlet distributions.

The inverse transformation of (3.1) gives

$$z_i = (\beta_i w_i)/(\sum_{i=1}^{K+1} \beta_i w_i) \tag{3.2}$$

where $\beta_i = 1/\theta_i,\ i = 1,\ldots,K+1$. From (3.1) and (3.2) we derive the density function of $\underset{\sim}{w}$. Its form is given by

$$g_{\underset{\sim}{\beta},\underset{\sim}{\lambda}}(\underset{\sim}{w}) = \frac{\Gamma(\lambda)}{\Gamma(\lambda_1)\cdots\Gamma(\lambda_{K+1})}((\prod_{i=1}^{K+1} \beta_i^{\lambda_i})/(\sum_{i=1}^{K+1} \beta_i w_i)^{\lambda})w_1^{\lambda_1-1}\cdots w_K^{\lambda_K-1}$$

$$(1-\sum_{i=1}^{K} w_i)^{\lambda_{K+1}-1}$$

$$w_{K+1} = 1-\sum_{i=1}^{K} w_i,\ 0 < w_i < 1, \tag{3.3}$$

where $\underset{\sim}{\beta} = (\beta_1,\ldots,\beta_{K+1})'$ and $\underset{\sim}{\lambda} = (\lambda_1,\ldots,\lambda_{K+1})'$.

In the next Section we give certain properties of the distribution (3.3).

IV. Some Properties of the Distribution of $\underset{\sim}{w}$

Consider the generalized Dirichlet random variable $\underset{\sim}{w}$ whose distribution is given by (3.3). Some important summaries of the distribution for application in Bayesian analysis, as a prior distribution for the parameters of a multinomial distribution, are given by

$$E(w_j/w_i) = (\theta_j/\theta_i)(\lambda_j/(\lambda_i-1)),\ \ j \neq i \tag{4.1}$$

$$E(\frac{1}{w_i}) = 1+(\sum_{j\neq i}\theta_j\lambda_j)/(\theta_i(\lambda_i-1)),\ \lambda_i > 1 \tag{4.2}$$

$$= 1+\mu_i,\ \text{say}$$

$$Var(\frac{1}{w_i}) = \sum_{j\neq i}(\theta_j/\theta_i)^2\lambda_j(\lambda_j+1)/((\lambda_i-1)(\lambda_i-2))$$

$$+2\sum_{j\neq l\neq i}(\theta_j\theta_l/\theta_i^2)^2(\lambda_j\lambda_l(\lambda_j+1)(\lambda_l+1)/(\lambda_i-1)(\lambda_i-2))-\mu_i^2,\ \lambda_i > 2 \tag{4.3}$$

$$Cov(\frac{1}{w_i},\frac{1}{w_j}) = 1+\sum_{r\neq i\neq j}(\theta_r^2/\theta_i\theta_j)((\lambda_r)(\lambda_r+1))/(\lambda_i-1)(\lambda_j-1)$$

$$+2\sum_{r\neq s\neq i\neq j}(\theta_r\theta_s/\theta_i\theta_j)(\lambda_r\lambda_s/(\lambda_i-1)(\lambda_j-1))$$

$$+2\sum_{r\neq i\neq j}(\theta_r/\theta_i)(\lambda_r/(\lambda_i)(\lambda_i-1)+(\theta_r/\theta_j)(\lambda_r/(\lambda_j)(\lambda_j-1)$$

$$-\sum_{i\neq j}(\theta_i/\theta_j)(\lambda_i/(\lambda_j-1)-(\sum_{i\neq j}(\theta_i/\theta_j)(\lambda_i/(\lambda_j-1))(\sum_{j\neq i}(\theta_j/\theta_i)(\lambda_j/(\lambda_i-1))$$

$$-\sum_{j\neq i}(\theta_j/\theta_i)(\lambda_j/(\lambda_i-1)-(\sum_{i\neq j}(\theta_i/\theta_j)(\lambda_i/(\lambda_j-1))(\sum_{j\neq i}(\theta_j/\theta_i)(\lambda_j/(\lambda_i-1)). \quad (2.9)$$

We observe that by introducing K+1 additional parameters we have made the family of generalized Dirichlet distribution more flexible compared to the Dirichlet distribution, for the specification of prior belief.

A turning point of the distribution (3.3) is given by the solution of

$$\frac{1}{\beta_1}\left(\frac{\lambda_1-1}{w_1}+c\right)=\cdots=\frac{1}{\beta_{K+1}}\left(\frac{\lambda_{K+1}-1}{w_{K+1}}+c\right),$$
$$w_1+\cdots+w_{K+1}=1$$

where c is an undetermined constant.

Let $\underset{\sim}{\psi}=(\psi_1,\ldots,\psi_{K+1})$ denote the parameters of a multinomial distribution. Since the Dirichlet family does not have enough parameters to represent prior beliefs regarding $\underset{\sim}{\psi}$, whereas the multivariate normal family is much richer, it has been proposed in the literature to assign a multivariate normal prior distribution to a set of log contrasts of the ψ_i (see O'Hagan, 1994). We define

$$\phi_i=\sum_{j=1}^{K+1}c_{ij}\log\psi_j\,,\quad \sum_{j=1}^{K+1}c_{ij}=0\,,\quad i=1,\ldots,K \qquad (4.4)$$

and write $\underset{\sim}{\phi}=c\log\underset{\sim}{\psi}$, where c is a $K\times(K+1)$ matrix of rank K. Together with the condition $\sum_{j=1}^{K+1}\psi_j=1$, this defines a 1-1 transformation from $\underset{\sim}{\phi}$ to $\underset{\sim}{\psi}$. We assume that $\underset{\sim}{\phi}$ is distributed a priori according to a multivariate normal distribution $N(\underset{\sim}{h},H)$. Given the multinomial data $\underset{\sim}{n}=(n_1,\ldots,n_{K+1})$, the posterior distribution of ψ $\underset{\sim}{n}$ is given by

$$|\quad f(\psi|\underset{\sim}{n})\propto\left(\prod_{j=1}^{K+1}\psi_j^{n_j-1}\right)\exp(-(c\log\underset{\sim}{\psi}-\underset{\sim}{h})'H^{-1}(c\log\underset{\sim}{\psi}-\underset{\sim}{h})). \qquad (4.5)$$

Aitchison (1985) has shown that the given transformed normal distributions constitute a rich family compared to the family of Dirichlet distribution. One drawback of (4.5) is that the normalizing constant can not be determined analytically through integration.

Suppose that $\underset{\sim}{\psi}$ is distributed a priori according to the generalized Dirichlet distribution (3.3). The posterior distribution of $\underset{\sim}{\psi}$, given the multinomial sample $\underset{\sim}{n}$ is given by the density function

$$f(\underset{\sim}{\psi}|\underset{\sim}{n})=a\left(\sum_{i=1}^{K+1}\beta_i w_i\right)^n g_{\underset{\sim}{\beta},\underset{\sim}{\lambda}+\underset{\sim}{n}}(\underset{\sim}{\psi}) \qquad (4.6)$$

where $a=a(\underset{\sim}{\beta},\underset{\sim}{\lambda},\underset{\sim}{n})$ is a normalizing constant.

References

Aitchison, J. (1985). A general class of distributions on the simplex. J.R. Statist. Soc. B.(47) 136-146.

Alam, K. and Padua, R.N. (1998). An Adaptive Procedure for Goodness-of-fit Test Based on a Sampling Spacings. Journal of Mathematical Sciences (1) 81-95.

Alam, K. Abernathy, R. and Williams, C.L. (1993). Multivariate Goodness-of-fit test based on statistically equivalent blocks. Communication in Statistics, Theory and Methods, 22, In Press.

Albert, J.H. and Gupta, A.K. (1982). Mixture of Dirichlet distributions and estimation in contingency tables. Ann. Statist. (10) 1251-1268.

Anderson, T.W. (1966). "Some parametric procedures based on statistically equivalent blocks," Proceedings of the International Symposium on Multivariate Analysis (P.R. Krishnaiah, ed.) 5-27 Academic Press, New York.

Bartlett, H. S. (1937). Properties of sufficiency and statistical tests. Proceed. Royal Statist. Soc. A, 268-282.

Bartoszynski, R. and Bugaj, M.N. (1996). Probability and Statistical Inference. John Wiley & Sons, Inc., New York.

Brunk, H.D. (1965). An Introduction to Mathematical Statistics, 2nd ed., Blaisdell Publishing Company.

Dickey, J.M. (1983). Multiple hypergeometric functions: probabilistic interpretations and statistical uses. J. Amer. Statist. Assoc. (78) 628-637.

Fraser, D.A.S. (1957). Nonparametric Methods in Statistics. Wiley Series in Statistics, 158-185.

Hartley, H.O. (1950). The maximum F-ratio as a short cut test for homogeneity of variances. Biometrika (37) 308-312.

Lehmann, E.L. (1984). Theory of Point Estimation. John Wiley & Sons, New York.

Leonard, T. and Novick, M.R. (1986). Bayesian full rank marginalization for two-way contingency tables. J. Educ. Statist. (11) 33-56.

O'Hagan, A. (1994). Kendall's Advanced Theory of Statistics, Vol. 2B, Bayesian Inference. University Press, Cambridge.

Sobel, M., Uppuluri, V.R.R. and Frankowski, K. (1977). Dirichlet Distribution –Type 1- Selected Tables In Mathematical Statistics, Vol. IV, AMS, Providence.

Staudte, R.G. and Sheather, S.J. (1990). Robust Estimation and Testing. John Wiley & Sons, Inc. Canada.

Tukey, J. W. (1947). Nonparametric estimation. Statistically equivalent blocks and tolerance regions-the continuous case. Ann. Math. Statist. (18) 529-539.

Wilks, S.S. (1962). Mathematical Statistics. John Wiley & Sons, New York,