APPLIED MATHEMATICS AND COMPUTATION

ELSEVIER Applied Mathematics and Computation 97 (1998) 165–181

Generalized Dirichlet distribution in Bayesian analysis

Tzu-Tsung Wong [1]

Department of Industrial Engineering, University of Wisconsin, 1513 University Avenue, Madison, WI 53706, USA

Abstract

Generalized Dirichlet distribution has a more general covariance structure than Dirichlet distribution. This makes the generalized Dirichlet distribution to be more practical and useful. The concept of complete neutrality will be used to derive the general moment function for the generalized Dirichlet distribution, and then some properties of the generalized Dirichlet distribution will be established. Similar to the Dirichlet distribution, the generalized Dirichlet distribution will be shown to conjugate to multinominal sampling. Two experiments are designed for studying the differences between the Dirichlet and the generalized Dirichlet distributions in Bayesian analysis. A method for generating samples from a generalized Dirichlet in presented. When a prior distribution is either a Dirichlet or a generalized Dirichlet distribution, the way for constructing such a prior is discussed.

Keywords: Bayesian analysis; Completely neutral; Conjugate; Generalized Dirichlet distribution; Prior construction

1. Introduction

Dirichlet distribution has been widely used in geology, biology, and chemistry for handling compositional data which are subject to nonnegativity and constant-sum constraints. In Bayesian analysis, Dirichlet distribution can be used as a prior for statistical models. Spiegelhalter et al. [1] used this way to study the frequencies of congenital heart disease. Paulino and Pereira [2] de-

[1] E-mail: tzu-tsun@cae.wisc.edu.

PII: S0096-3003(97)10140-0

veloped a Bayesian approach to analyze incomplete categorical data that does not follow any specific pattern. Lange [3] assumed that the allele frequency has a Dirichlet prior, and constructed a model to compute the forensic match probabilities.

When $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ has a Dirichlet distribution, any two random variables in $\mathbf{X}$ will be negatively correlated. In some practical cases, two random variables may be positively correlated, and hence the Dirichlet distribution will not be a reasonable choice to be a prior distribution in Bayesian analysis. Connor and Mosimann [4] used the concept of complete neutrality to generalize the Dirichlet distribution.

Random vector $\mathbf{X}$ is said to be completely neutral if $(X_1, X_2, \ldots, X_j)$ is independent of $(X_{j+1}, X_{j+2}, \ldots, X_k)/V_j$ for all $j < k$, where $V_j = 1 - X_1 - X_2 - \cdots - X_j$. Let $Z_1 = X_1$, and let $Z_j = X_j/V_{j-1}$ for $j = 2, 3, \ldots, k$. When the Z_j are independent, then $\mathbf{X}$ is also completely neutral. Connor and Mosimann supposed that each Z_j has a beta distribution with parameters α_j and β_j, and derived the density function for the generalized Dirichlet distribution as follows:

$$f(x) = \prod_{i=1}^{k} \frac{1}{B(\alpha_i, \beta_i)} x_i^{\alpha_i - 1}(1 - x_1 - \cdots - x_i)^{\gamma_i}$$

for $x_1 + x_2 + \cdots + x_k \leqslant 1$ and $x_j \geqslant 0$ for $j = 1, 2, \ldots, k$, where $B(\alpha_j, \beta_j) = \Gamma(\alpha_j)\Gamma(\beta_j)/\Gamma(\alpha_j + \beta_j)$, $\gamma_j = \beta_j - \alpha_{j+1} - \beta_{j+1}$ for $j = 1, 2, \ldots, k-1$, and $\gamma_k = \beta_k - 1$.

In a generalized Dirichlet distribution, X_1 is always negatively correlated with the other random variables. However, X_j and X_m can be positively correlated for $j, m > 1$ [5]. If there exists some $m > j$ such that X_j and X_m are positively (negatively) correlated, then X_j will be positively (negatively) correlated with X_n for all $n > j$. Since the generalized Dirichlet distribution has a more general covariance structure than the Dirichlet distribution, this makes the generalized Dirichlet distribution to be more practical and useful.

In Section 2, the concept of complete neutrality will be used to derive the general moment function for the generalized Dirichlet distribution, and some properties of the generalized Dirichlet distribution can then be established. The generalized Dirichlet distribution will be shown to conjugate to multinomial sampling in Section 3. Two experiments are designed for studying the differences between the Dirichlet and the generalized Dirichlet distributions in Bayesian analysis. Then a method to generate samples for the generalized Dirichlet distribution will be presented in Section 4. Since the Dirichlet distribution is a special case of the generalized Dirichlet distribution, the same method can be used to generate samples for the Dirichlet distribution. Finally, the way for constructing a priori that is either a Dirichlet or a generalized Dirichlet distribution is discussed.

2. General moment function

Let $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ be a vector random variable having the generalized Dirichlet distribution $\mathrm{GD}_k(\alpha_1, \alpha_2, \ldots, \alpha_k;\ \beta_1, \beta_2, \ldots, \beta_k)$. Define $Z_1 = X_1$ and $Z_j = X_j / V_{j-1}$ for $j = 2, 3, \ldots, k$, where $V_{j-1} = 1 - X_1 - X_2 - \cdots - X_j$. Then we have

$$X_1 = Z_1,$$

$$X_j = Z_j(1 - X_1 - \cdots - X_{j-1}) = Z_j \prod_{i=1}^{j-1} (1 - Z_i) \quad \text{for } j = 2, 3, \ldots, k.$$

By Connor and Mosimann, since $\mathbf{X}$ is completely neutral, the Z_j are independent, and each Z_j has a beta distribution with parameters α_j and β_j. Hence, the general moment function of the generalized Dirichlet distribution will be

$$\begin{aligned} E[X_1^{r_1} X_2^{r_2} \cdots X_k^{r_k}] &= E\left[Z_1^{r_1} Z_2^{r_2} (1 - Z_1)^{r_2} \cdots Z_k^{r_k} \prod_{i=1}^{k-1} (1 - Z_i)^{r_k} \right] \\ &= E[Z_1^{r_1} (1 - Z_1)^{r_2 + \cdots + r_k} Z_2^{r_2} (1 - Z_2)^{r_3 + \cdots + r_k} \cdots Z_k^{r_k}] \\ &= E[Z_1^{r_1} (1 - Z_1)^{r_2 + \cdots + r_k}] E[Z_2^{r_2} (1 - Z_2)^{r_3 + \cdots + r_k}] \cdots E[Z_k^{r_k}] \\ &\quad \text{(by independence).} \end{aligned}$$

Since Z_j has a beta distribution for $j = 1, 2, \ldots, k$, we have

$$E[Z_j^{r_j} (1 - Z_j)^{r_{j+1} + \cdots + r_k}] = \frac{\Gamma(\alpha_j + \beta_j)\Gamma(\alpha_j + r_j)\Gamma(\beta_j + r_{j+1} + \cdots + r_k)}{\Gamma(\alpha_j)\Gamma(\beta_j)\Gamma(\alpha_j + \beta_j + r_j + \cdots + r_k)}$$

for $j = 1, 2, \ldots, k - 1$, and

$$E[Z_k^{r_k}] = \frac{\Gamma(\alpha_k + \beta_k)\Gamma(\alpha_k + r_j)}{\Gamma(\alpha_k)\Gamma(\alpha_k + \beta_k + r_k)}.$$

Let $\delta_j = r_{j+1} + r_{j+2} + \cdots + r_k$ for $j = 1, 2, \ldots, k - 1$ and $\delta_k = 0$. Then the general moment function of the generalized Dirichlet distribution $\mathrm{GD}_k(\alpha_1, \alpha_2, \ldots, \alpha_k; \beta_1, \beta_2, \ldots, \beta_k)$ becomes

$$E[X_1^{r_1} X_2^{r_2} \cdots X_k^{r_k}] = \prod_{j=1}^{k} \frac{\Gamma(\alpha_j + \beta_j)\Gamma(\alpha_j + r_j)\Gamma(\beta_j + \delta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)\Gamma(\alpha_j + \beta_j + r_j + \delta_j)}.$$

Property 1. *The distribution of $(X_1, X_2, \ldots, X_s)$ for any $s < k$ is an s-variate generalized Dirichlet distribution.*

Proof. Let $r_{s+1} = r_{s+2} = \cdots = r_k = 0$. Then we have

$$E[X_1^{r_1} X_2^{r_2} \cdots X_s^{r_s}] = \prod_{j=1}^{s} \frac{\Gamma(\alpha_j + \beta_j)\Gamma(\alpha_j + r_j)\Gamma(\beta_j + \delta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)\Gamma(\alpha_j + \beta_j + r_j + \delta_j)},$$

which is the general moment function of a generalized Dirichlet distribution. Since $(X_1, X_2, \ldots, X_s)$ is defined on a compact support, random vector $(X_1, X_2, \ldots, X_s)$ has an s-variate generalized Dirichlet distribution. □

Property 2. *The rth moment of* $1 - X_1 - X_2 - \cdots - X_k$ *is*

$$E[(1 - X_1 - X_2 - \cdots - X_k)^r] = \prod_{j=1}^{k} \frac{\Gamma(\alpha_j + \beta_j)\Gamma(\beta_j + r)}{\Gamma(\beta_j)\Gamma(\alpha_j + \beta_j + r)}.$$

Proof. Note that Z_j has a beta distribution with parameters α_j and β_j, and hence $1 - Z_j$ has a beta distribution with parameters β_j and α_j. Therefore,

$$\begin{aligned} E[(1 - X_1 - X_2 - \cdots - X_k)^r] &= E\left[\prod_{j=1}^{k}(1 - Z_j)^r\right] \\ &= \prod_{j=1}^{k} E[(1 - Z_j)^r] \text{ (by independence)} \\ &= \prod_{j=1}^{k} \frac{\Gamma(\alpha_j + \beta_j)\Gamma(\beta_j + r)}{\Gamma(\beta_j)\Gamma(\alpha_j + \beta_j + r)}. \qquad \square \end{aligned}$$

Property 3. *Let* $V_s = 1 - X_1 - X_2 - \cdots - X_s$. *Then* $(X_{s+1}/V_s, X_{s+2}/V_s, \ldots, X_k/V_s | X_1, X_2, \ldots, X_s)$ *for any* $s < k$ *has a* $(k - s)$*-variate generalized Dirichlet distribution.*

Proof. Since the Z_j are independent, and X_i only depends on the Z_j for $j \leqslant i$, Z_j is independent of the X_i for all $i < j$. Let $\delta_j = r_{j+1} + r_{j+2} + \cdots + r_k$ for $j = s+1, s+2, \ldots, k-1$ and $\delta_k = 0$. Then the general moment function of $(X_{s+1}/V_s, X_{s+2}/V_s, \ldots, X_k/V_s \mid X_1, X_2, \ldots, X_s)$ will be

$$\begin{aligned} &E\left[\left(\frac{X_{s+1}}{V_s}\right)^{r_{s+1}}\left(\frac{X_{s+2}}{V_s}\right)^{r_{s+2}}\cdots\left(\frac{X_k}{V_s}\right)^{r_k}\middle| X_1, X_2, \ldots, X_s\right] \\ &= E\left[Z_{s+1}^{r_{s+1}}\{Z_{s+2}(1-Z_{s+1})\}^{r_{s+2}}\cdots\left\{Z_k\prod_{j=s+1}^{k-1}(1 - Z_j)\right\}^{r_k}\middle| X_1, X_2, \ldots, X_s\right] \\ &= E\left[Z_{s+1}^{r_{s+1}}\{Z_{s+2}(1 - Z_{s+1})\}^{r_{s+2}}\cdots\left\{Z_k\prod_{j=s+1}^{k-1}(1 - Z_j)\right\}^{r_k}\right] \\ &= E[Z_{s+1}^{r_{s+1}}(1 - Z_{s+1})^{r_{s+2}+\cdots+r_k}]E[Z_{s+2}^{r_{s+2}}(1 - Z_{s+2})^{r_{s+3}+\cdots+r_k}]\cdots E[Z_k^{r_k}] \\ &= \prod_{j=s+1}^{k} \frac{\Gamma(\alpha_j + \beta_j)\Gamma(\alpha_j + r_j)\Gamma(\beta_j + \delta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)\Gamma(\alpha_j + \beta_j + r_j + \delta_j)}, \end{aligned}$$

which is the general moment function of a generalized Dirichlet distribution. □

When $\beta_j = \alpha_{j+1} + \beta_{j+1}$ for $j = 1, 2, \ldots, k-1$, a generalized Dirichlet distribution reduces to a Dirichlet distribution. The Dirichlet distribution also has Properties 1 and 3. By setting $\beta_j = \alpha_{j+1} + \beta_{j+1}$ for $j = 1, 2, \ldots, k-1$, it is easy to show that random variable $1 - X_1 - X_2 - \cdots - X_k$ has a beta distribution if $(X_1, X_2, \ldots, X_k)$ follows a Dirichlet distribution.

Connor and Mosimann said that if $(X_1, X_2, \ldots, X_k)$ has a Dirichlet distribution, then every permutation of $(X_1, X_2, \ldots, X_k)$ is completely neutral. This implies that if $(X_1, X_2, \ldots, X_k)$ has a Dirichlet distribution, then the order of the X_j can be arbitrary. For example, if $(X_1, X_2, X_3) \sim D_3(\alpha_1, \alpha_2, \alpha_3; \alpha_4)$, then $(X_2, X_3, X_1) \sim D_3(\alpha_2, \alpha_3, \alpha_1; \alpha_4)$, $(X_3, X_2, X_1) \sim D_3(\alpha_3, \alpha_2, \alpha_1; \alpha_4)$, and so on. When $(X_1, X_2, \ldots, X_k)$ has a generalized Dirichlet distribution, then $(X_1, X_2, \ldots, X_k)$ is also completely neutral. However, this does not mean that every permutation of $(X_1, X_2, \ldots, X_k)$ will be completely neutral. For instance, if $(X_1, X_2, X_3) \sim \mathrm{GD}_k(\alpha_1, \alpha_2, \alpha_3; \beta_1, \beta_2, \beta_3)$ and $\beta_1 \neq \alpha_2 + \beta_2$, then (X_2, X_1, X_3) will not have a generalized Dirichlet distribution. So, if $(X_1, X_2, \ldots, X_k)$ has a generalized Dirichlet distribution, the order of the X_j is generally not arbitrary.

3. Bayesian analysis

Suppose that the joint prior of $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ is a Dirichlet distribution $D_k(\alpha_1, \alpha_2, \ldots, \alpha_k; \alpha_{k+1})$, and the sampling model $\mathbf{Y}|\mathbf{X}$ follows a multinomial distribution, where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$. Let $Y_j = y_j$ for $j = 1, 2, \ldots, k$ and $y_{k+1} = n - y_1 - y_2 - \cdots - y_k$. Then the joint posterior of $\mathbf{X}|\mathbf{Y}$ is a Dirichlet distribution $D_k(\alpha'_1, \alpha'_2, \ldots, \alpha'_k; \alpha'_{k+1})$, where $\alpha'_j = \alpha_j + y_j$ for $j = 1, 2, \ldots, k+1$. So, the Dirichlet distribution is conjugate to multinomial sampling. A similar result for the generalized Dirichlet distribution is shown in the following lemma.

Lemma 1. *Suppose that the joint prior of* $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ *is a generalized Dirichlet distribution* $\mathrm{GD}_k(\alpha_1, \alpha_2, \ldots \alpha_k;\ \beta_1, \beta_2, \ldots, \beta_k)$ *and the sampling model* $\mathbf{Y}|\mathbf{X}$ *follows a multinomial distribution, where* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$. *Let* y_j *for* $j = 1, 2, \ldots, k+1$ *be defined as above. Then the joint posterior of* $\mathbf{X}|\mathbf{Y}$ *is a generalized Dirichlet distribution* $\mathrm{GD}_k(\alpha'_1, \alpha'_2, \ldots, \alpha'_k;\ \beta'_1, \beta'_2, \ldots, \beta'_k)$, *where* $\alpha'_j = \alpha_j + y_j$ *and* $\beta'_j = \beta_j + y_{j+1} + y_{j+2} + \cdots + y_{k+1}$ *for* $j = 1, 2, \ldots, k$.

Proof. According to the densities of $f(\mathbf{x})$ and $L(\mathbf{y}/\mathbf{x})$,

$$\begin{aligned} f(\mathbf{x}|\mathbf{y}) &\propto f(\mathbf{x})L(\mathbf{y}|\mathbf{x}) \\ &= x_1^{(\alpha_1+y_1)-1}(1-x_1)^{(\beta_1+y_2+\cdots+y_{k+1})-(\alpha_2+y_2)-(\beta_2+y_3+\cdots+y_{k+1})} \\ &\quad \times x_2^{(\alpha_2+y_2)-1}(1-x_1-x_2)^{(\beta_2+y_3+\cdots+y_{k+1})-(\alpha_3+y_3)-(\beta_3+y_4+\cdots+y_{k+1})} \cdots \\ &\quad \times x_k^{(\alpha_k+y_k)-1}(1-x_1-\cdots-x_k)^{(\beta_k+y_{k+1})-1}. \end{aligned}$$

Let $\alpha'_j = \alpha_j + y_j$ and $\beta'_j = \beta_j + y_{j+1} + y_{j+2} + \cdots + y_{k+1}$ for $j = 1, 2, \ldots k$. Then the posterior density of $\mathbf{X}|\mathbf{Y}$ will be

$$\begin{aligned} f(\mathbf{x}|\mathbf{y}) &\propto x_1^{\alpha'_1 - 1}(1 - x_1)^{\beta'_1 - \alpha'_2 - \beta'_2} x_2^{\alpha'_2 - 1}(1 - x_1 - x_2)^{\beta'_2 - \alpha'_3 - \beta'_3} \cdots x_k^{\alpha'_k - 1} \\ &\quad \times (1 - x_1 - \cdots - x_k)^{\beta'_k - 1} \\ &= \prod_{i=1}^{k} x_i^{\alpha'_1 - 1_i}(1 - x_1 - \cdots - x_i)^{\gamma'_i}, \end{aligned}$$

where $\gamma'_j = \beta'_j - \alpha'_{j+1} - \beta'_{j+1}$ for $j = 1, 2, \ldots, k-1$ and $\gamma'_k = \beta'_k - 1$. This differs from the generalized Dirichlet distribution only by a constant factor. Therefore, $f(\mathbf{x}|\mathbf{y})$ is a generalized Dirichlet distribution. □

When the prior is a Dirichlet distribution $D_k(\alpha_1, \alpha_2, \ldots, \alpha_k;\ \alpha_{k+1})$, by Dirichlet integral, the density function of its prior predictive distribution will be

$$f(\mathbf{y}) = \int L(\mathbf{y}|\mathbf{x}) f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \frac{\Gamma\left(\sum_{j=1}^{k+1} \alpha_j\right) \prod_{j=1}^{k+1} \Gamma(\alpha_j + y_j)}{\prod_{j=1}^{k+1} \Gamma(\alpha_j) \Gamma\left(\sum_{j=1}^{k+1} (\alpha_j + y_j)\right)}.$$

Let $\mathbf{w} = (w_1, w_2, \ldots, w_{k+1})$ be some sampling data from the posterior distribution. Since Dirichlet distribution is conjugate to the multinomial sampling, the density function of the posterior predictive distribution will be

$$\begin{aligned} f(\mathbf{w}|\mathbf{y}) &= \int L(\mathbf{w}|\mathbf{x}, \mathbf{y}) f(\mathbf{x}|\mathbf{y})\, \mathrm{d}\mathbf{x} = \int L(\mathbf{w}|\mathbf{x}) f(\mathbf{x}|\mathbf{y})\, \mathrm{d}\mathbf{x} \\ &= \frac{\Gamma\left(\sum_{j=1}^{k+1} (\alpha_j + y_j)\right) \prod_{j=1}^{k+1} \Gamma(\alpha_j + y_j + w_j)}{\prod_{j=1}^{k+1} \Gamma(\alpha_j + y_j) \Gamma\left(\sum_{j=1}^{k+1} (\alpha_j + y_j + w_j)\right)}. \end{aligned}$$

If a prior is a generalized Dirichlet distribution $\mathrm{GD}_k(\alpha_1, \alpha_2, \ldots, \alpha_k; \beta_1, \beta_2, \ldots, \beta_k)$, then the density function of the prior predictive distribution will be

$$f(\mathbf{y}) = \int L(\mathbf{y}|\mathbf{x}) f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \prod_{j=1}^{k} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)} \prod_{j=1}^{k} \frac{\Gamma(\alpha'_j)\Gamma(\beta'_j)}{\Gamma(\alpha'_j + \beta'_j)},$$

where α'_j and β'_j are the same as defined in Lemma 1. Similarly, the density function of the posterior predictive distribution will be

$$f(\mathbf{w}|\mathbf{y}) = \prod_{j=1}^{k} \frac{\Gamma(\alpha'_j + \beta'_j)}{\Gamma(\alpha'_j)\Gamma(\beta'_j)} \prod_{j=1}^{k} \frac{\Gamma(\alpha''_j)\Gamma(\beta''_j)}{\Gamma(\alpha''_j + \beta''_j)}.$$

where $\alpha''_j = \alpha'_j + w_j$ and $\beta''_j = \beta'_j + w_{j+1} + w_{j+2} + \cdots + w_{k+1}$ for $j = 1, 2, \ldots k$.

The following two experiments are designed to simulate the behavior of an analyst whose prior distribution is either a Dirichlet or a generalized Dirichlet distribution. These experiments will be helpful for understanding the differences between the Dirichlet and the generalized Dirichlet distributions in Bayesian analysis. Suppose that there are some color balls in an urn. The number of colors is known to be $k+1$, but the number of each color ball in the urn is unknown, and hence the proportion of each color ball in the urn is also unknown. Let X_j be the proportion of the balls with color j in the urn.

Experiment 1. Analyst 1 believes that $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ follows a Dirichlet distribution $D_k(\alpha_1, \alpha_2, \ldots, \alpha_k; \alpha_{k+1})$. Without loss of generality, suppose that all α_is are positive integers. According to the Dirichlet prior, analyst 1 makes $k+1$ glass boxes, and puts α_1 balls with color 1 in box 1, α_2 balls with color 2 in box 2, and so on. Then analyst 1 draws one ball from the urn, and puts that ball into the box that has the balls with the same color as the ball drawn from the urn. Since the boxes are made by glass, analyst 1 knows that the ball drawn from the urn should be put into which box. This process continues until n balls are drawn from the urn.

Experiment 2. Analyst 2 believes that $\mathbf{X}$ follows a generalized Dirichlet distribution $\mathrm{GD}_k(\alpha_1, \alpha_2, \ldots, \alpha_k;\ \beta_1, \beta_2, \ldots, \beta_k)$. All parameters are also assumed to be positive integers. First, analyst 2 makes $k+1$ wood boxes, and divides each box into two areas. Then for $j = 1, 2, \ldots, k$, analyst 2 puts α_j balls with color j and β_j marbles in the first and second areas of box j, respectively, and puts a ball with color $k+1$ in the first area of box $k+1$. Unfortunately, analyst 2 never remembers the colors of the balls in the boxes. So, when a ball is drawn from the urn, analyst 2 will first draw a ball from box 1 to determine whether the two balls have the same color or not. If the colors are different, analyst 2 will put the ball drawn from box 1 back and put a marble into the second area of box 1. Then analyst 2 draws a ball from box 2 and compares the colors again. Assume that a ball drawn from box j has the same color as the ball drawn from the urn. Then analyst 2 will put the two balls into box j and go to the urn to draw another ball. This process continues until n balls are drawn from the urn.

In both experiments, the boxes represent current beliefs of the analysts, and drawing balls from the urn represents the collection of data. Let y_j be the number of balls with color j drawn from the urn. In the first experiment, suppose that the first ball drawn from the urn has the same color as the balls in box j. Let $\mu_j = E[X_j]$ for $j = 1, 2, \ldots, k+1$, and let $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_{k+1}$. If $j = 1$, then μ_1 increases from α_1/α to $(\alpha_1 + 1)/(\alpha + 1)$ because $\alpha_1/\alpha < 1$. If $j \neq 1$, then μ_1 decreases from α_1/α to $\alpha_1/(\alpha + 1)$. This argument is true for all colors, and coincides with our instinct. On the other hand, in the second experiment, let the first ball drawn from the urn be finally put into box j. If $j = 1$, then μ_1 increases from $\alpha_1/(\alpha_1 + \beta_1)$ to $(\alpha_1 + 1)/(\alpha_1 + \beta_1 + 1)$ because $\alpha_1/(\alpha_1 + \beta_1) < 1$. If $j \neq 1$, then μ_1 will decrease. However, is this argument also true for all $j > 1$?

If $j = 1$, it is easy to show that μ_2 will decrease. If $j = 2$, then μ_2 increases from $(\alpha_2/(\alpha_2 + \beta_2))(\beta_1/(\alpha_1 + \beta_1))$ to $((\alpha_2 + 1)/(\alpha_2 + \beta_2 + 1))((\beta_1 + 1)/(\alpha_1 + \beta_1 + 1))$. If $j > 2$, the value of μ_2 becomes $(\alpha_2/(\alpha_2 + \beta_2 + 1))((\beta_1 + 1)/(\alpha_1 + \beta_1 + 1))$. Since $(\alpha_2/(\alpha_2 + \beta_2)) > (\alpha_2/(\alpha_2 + \beta_2 + 1))$ but $(\beta_1/(\alpha_1 + \beta_1)) < ((\beta_1 + 1)/(\alpha_1 + \beta_1 + 1))$, the new value of μ_2 can be smaller or larger than the previous one when $j > 2$. As stated in section $1, X_1$ is always negatively correlated with the other random variables. Hence, if a ball drawn from the urn is put into box $1, \mu_1$ will increase and μ_m will decrease for all $m > 1$. If a ball drawn from the urn is put into box j for some $j > 1$, then μ_j will definitely increase. Since some X_m may be positively correlated with X_j when the prior is a generalized Dirichlet distribution, μ_m can increase even when a ball drawn from the urn is not put into box m.

Note that the boxes in the first experiment are made by glass. Hence, when the order of the boxes is changed, any ball drawn from the urn can still be put into the correct box. This means that the order of the boxes has no effect on the first experiment. So, changing the order of the boxes while performing the experiment will have no impact on the result of the first experiment. However, the boxes in the second experiment are made by wood. Analyst 2 will not know that a ball drawn from the urn should be put into which box without comparing. So, if the order of the boxes in the second experiment is changed while performing the experiment, the numbers of comparisons for the same color balls drawn from the urn may be different. This will change the result of the second experiment. So, the order of the variables in a generalized Dirichlet random vector is important.

4. Statistical sampling

Suppose that W_j for $j = 1, 2, \ldots, k + 1$ are independent gamma random variables with parameters α_j and 1. Let $U = W_1 + W_2 + \cdots + W_{k+1}$, and let $X_j = W_j/U$ for $j = 1, 2, \ldots, k$. Then by Wilks [6], $(X_1, X_2, \ldots, X_k)$ has a Dirichlet distribution $D_k(\alpha_1, \alpha_2, \ldots, \alpha_k; \alpha_{k+1})$. So, the way to generate samples for this Dirichlet distribution is as follows:

```
sum = 0
for ( j in 1: k + 1)
{
   w_j = rgamma(α_j, 1)
   sum = sum + w_j
}
for ( j in 1: k) x_j = w_j/sum,
```

where function 'rgamma' is used to generate samples for gamma distribution. The outcome $(x_1, x_2, \ldots, x_k)$ is a sample for distribution $D_k(\alpha_1, \alpha_2, \ldots, \alpha_k;\ \alpha_{k+1})$.

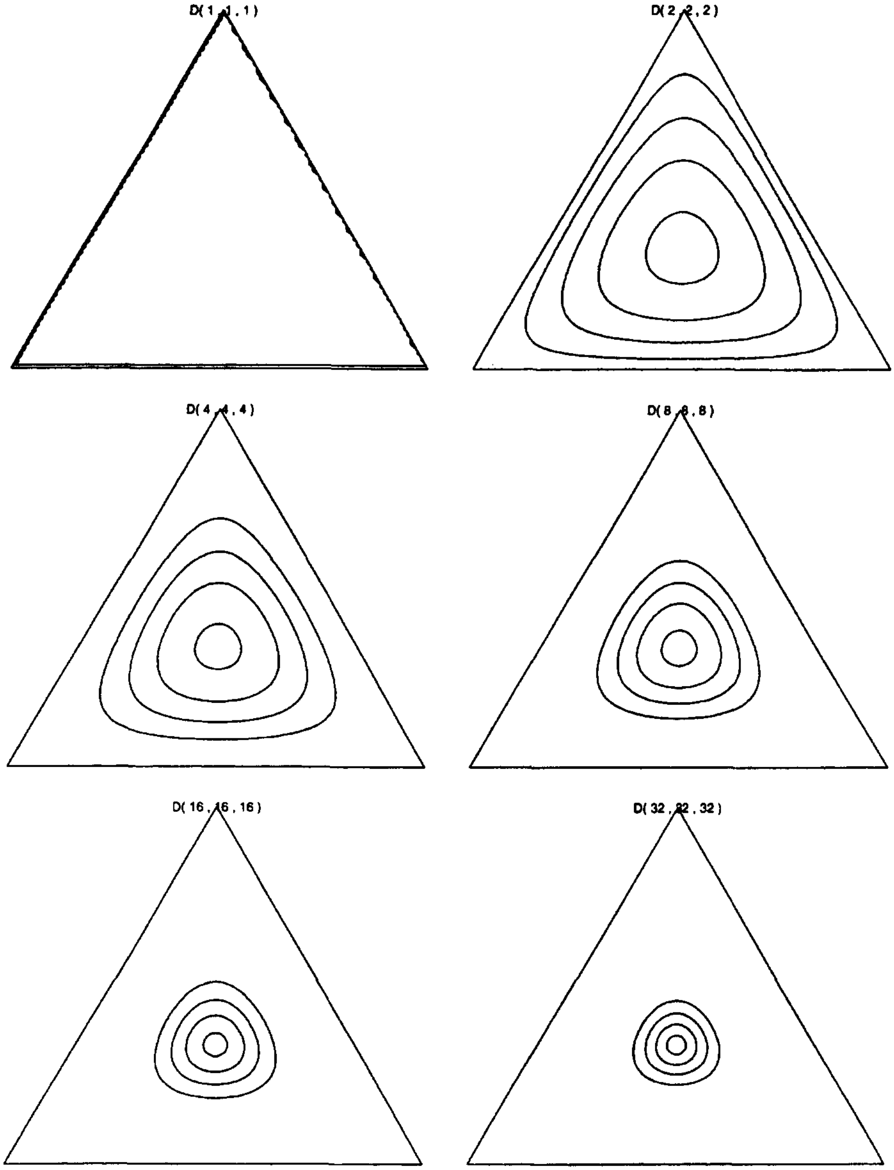

Fig. 1. Contour graphs of Dirichlet distribution.

Some contour graphs for $D_2(\alpha,\ \alpha;\ \alpha), \alpha = 1,\ 2,\ 4,\ 8,\ 16,\ 32$, are shown in Fig. 1. The four contours in each graph are 0.05, 0.2, 0.5, and 0.9 contours comparing to the highest density value. No matter $\alpha = 1,\ 2,\ 4,\ 8,\ 16$ or 32,

$$E[X_1] = E[X_2] = E[1 - X_1 - X_2] = 1/3.$$

For the sake of simplicity, let $X_3 = 1 - X_1 - X_2$. The coordinates of the left, right, and upper extreme points in a triangle are (1,0,0), (0,1,0), and (0,0,1), respectively. This 2-dimensional triangle is used to represent the surface $X_1 + X_2 + X_3 = 1$ in a 3-dimensional space. An interesting result is that when the value of α increases, the contours will be more concentrated and closer to the central point $(1/3, 1/3, 1/3)$. Suppose that the 6 distributions shown in Fig. 1 represent the beliefs of 6 persons. Although they have the same expected value on each variable, the confidence in their beliefs are different. The person with $D_2(1, 1;\ 1)$ is the least confident person, and the person with $D_2(32, 32;\ 32)$ is the most confident one in the group.

Suppose that Z_j for $j = 1, 2, \ldots, k$ are independent beta random variables with parameters α_j and β_j. Let $X_1 = Z_1$ and $X_j = Z_j(1 - X_1 - \cdots - X_{j-1})$ for $j = 2, 3, \ldots, k$. Then by Connor and Mosimann, $(X_1, X_2, \ldots, X_k)$ has a generalized Dirichlet distribution $\mathrm{GD}_k(\alpha_1, \alpha_2, \ldots, \alpha_k;\ \beta_1, \beta_2, \cdots, \beta_k)$. So, the way to generate samples for this generalized Dirichlet distribution is as follows:

$x_1 = \mathrm{rbeta}(\alpha_1, \beta_1)$
sum $= x_1$
for(j in $2 : k$)
{
  $x_j = \mathrm{rbeta}(\alpha_j, \beta_j) * (1 - \mathrm{sum})$
  sum $=$ sum $+ x_j$
}

where function ‘rbeta’ is used to generate samples for beta distribution. The outcome $(x_1, x_2, \ldots, x_k)$ is a sample for distribution $\mathrm{GD}_k(\alpha_1, \alpha_2, \ldots, \alpha_x;\ \beta_1, \beta_2, \ldots, \beta_k)$.

As stated in Section 2, when $\beta_j = \alpha_{j+1} + \beta_{j+1}$ for $j = 1, 2, \ldots, k - 1$, a generalized Dirichlet distribution reduces to a Dirichlet distribution. Hence, the method to generate samples for the generalized Dirichlet distribution can be used to generate samples for the Dirichlet distribution. For instance, suppose that $(X_1, X_2, \ldots, X_k)$ has a Dirichlet distribution $D_k(\alpha_1, \alpha_2, \ldots, \alpha_k;\ \alpha_{k+1})$. Then by setting $\beta_j = \alpha_{j+1} + \alpha_{j+2} + \cdots + \alpha_{k+1}$ for $j = 1, 2, \ldots, k$, $(X_1, X_2, \ldots, X_k)$ will have a generalized Dirichlet distribution $\mathrm{GD}_k(\alpha_1, \alpha_2, \ldots, \alpha_k;\ \beta_1, \beta_2, \ldots, \beta_k)$. Now, the method to generate samples for the generalized Dirichlet distribution works for this Dirichlet distribution. In general, the samples from two independent gamma random variables will be used to generate samples for a beta random variable. So, the second method is not faster than the first one in generating samples for the Dirichlet distribution.

Some contour graphs for $\mathrm{GD}_2(\alpha, 2\alpha; 2\alpha, 2\alpha)$, $\alpha = 1, 2, 4, 8, 16, 32$, are shown in Fig. 2. No matter $\alpha = 1, 2, 4, 8, 16$, or 32,

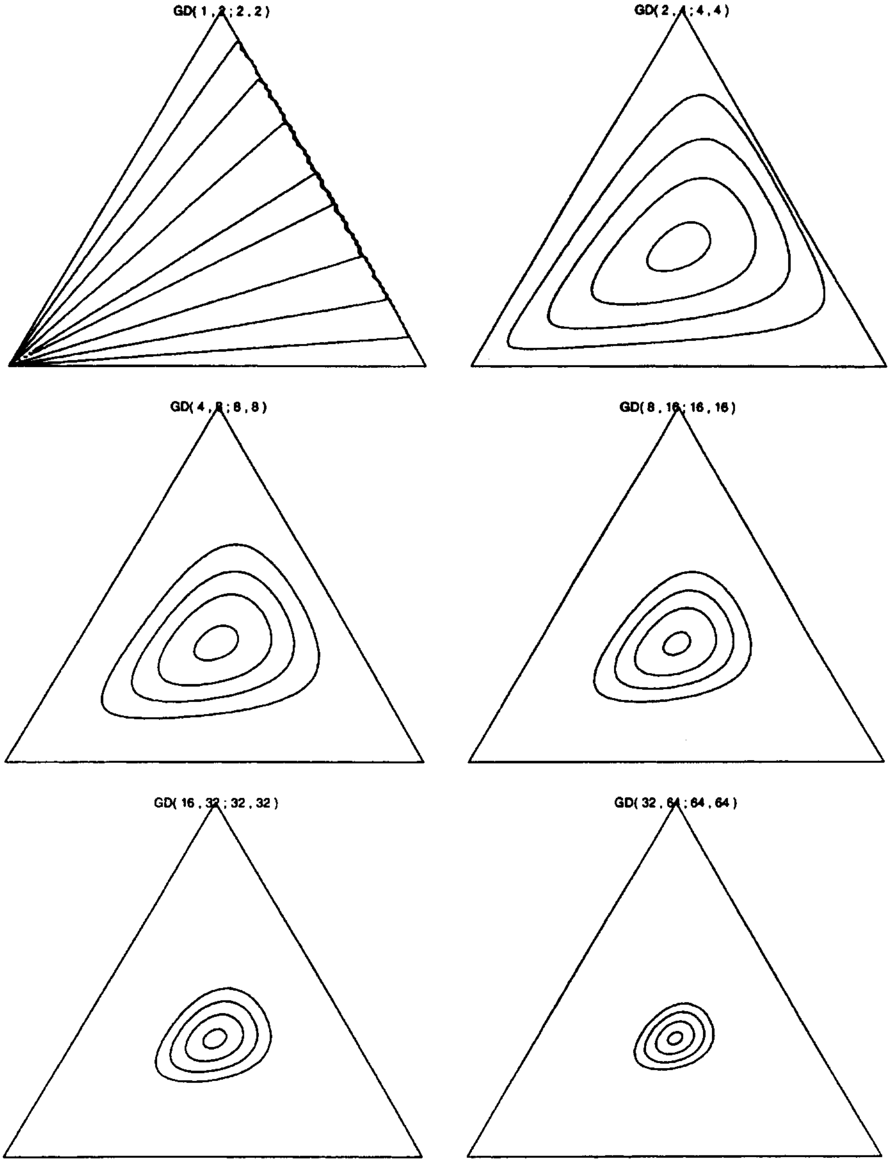

Fig. 2. Contour graphs of generalized Dirichlet distribution.

$$E[X_1] = \alpha/(\alpha + 2\alpha) = 1/3,$$

$$E[X_2] = \frac{2\alpha}{2\alpha + 2\alpha}\frac{2\alpha}{\alpha + 2\alpha} = 1/3,$$

$$E[1 - X_1 - X_2] = 1/3.$$

Let $X_3 = 1 - X_1 - X_2$. Like the Dirichlet distribution, when the value of α increases, the contours are more concentrated and closer to the central point $(1/3, 1/3, 1/3)$. So, the sum of the parameters in the generalized Dirichlet prior represents the confidence in the beliefs.

By comparing the graphs in Figs. 1 and 2, we can see that the contours of the Dirichlet distributions are symmetric, but the contours of the generalized Dirichlet distributions are not. This implies that an analyst whose prior is a generalized Dirichlet distribution can have different degrees of beliefs on the random variables that have the same expected value. This will not happen when the prior is a Dirichlet distribution. Actually, in a generalized Dirichlet distribution, the degree of beliefs on random variable X_j is determined by the sum $\alpha_j + \beta_j$. Consider the graph for GD(4, 8; 8, 8) shown in Fig. 2. An analyst can change the values of α_2 and β_2 without changing the expected values of all random variables. Let $\alpha_2 = \beta_2 = 4$. So, the new distribution is GD(4, 4; 8, 4) that is equivalent to the distribution $D(4, 4; 4)$. Referring to the graph for $D(4, 4; 4)$ shown in Fig. 1, we can see that the curves closer to X_2 in $D(4, 4; 4)$ is not as close and concentrated to the central point as the curves in GD(4, 8; 8, 8). Hence, the degree of beliefs on X_2 in GD(4, 8; 8, 8) is higher. This can also be shown by calculating their probability densities.

5. Prior construction

Krzysztofowicz and Reese [7] proposed a stochastic bifurcation process for generating families of multivariate distributions. They also showed that the outcome of the process governed by independent bifurcations and beta laws will be the generalized Dirichlet family. The bifurcation process recursively divides a unit into two parts according to some probability law. However, recursively dividing a unit into two parts is not always a natural way for an analyst to describe his beliefs. Hence, this process will not be useful for prior construction.

As discussed in Section 3, when $\mathbf{X}$ has a Dirichlet distribution $D_k(\alpha_1, \alpha_2, \ldots, \alpha_k; \alpha_{k+1})$, $E[X_j] = \alpha_j/\alpha$ for $j = 1, 2, \ldots, k+1$, where $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_{k+1}$. Let $\mathbf{y} = (y_1, y_2, \ldots, y_{k+1})$ be the sampling data, and let $n = y_1 + y_2 + \cdots + y_{k+1}$. Then in the posterior distribution,

$$\begin{aligned} E[X_j|\mathbf{y}] &= \frac{\alpha_j + y_j}{\alpha + n} = \frac{\alpha}{\alpha + n}\frac{\alpha_j}{\alpha} + \frac{n}{\alpha + n}\frac{y_j}{n} \\ &= u\frac{\alpha_j}{\alpha} + (1 - u)\frac{y_j}{n}, \quad j = 1, 2, \ldots, k+1, \end{aligned}$$

where $u = \alpha/(\alpha + n)$ and $1 - u$ are the weights of prior and sample means, respectively. Note that the weights for all random variables are identical.

Therefore, the posterior mean is just the sum of the prior and sample means multiplied by their weights.

If the joint distribution of **X** is a generalized Dirichlet distribution $\mathrm{GD}_k(\alpha_1,\alpha_2,\ldots,\alpha_k;\ \beta_1,\beta_2,\ldots,\beta_k)$, then

$$E[X_j] = \frac{\alpha_j}{\alpha_j+\beta_j}\prod_{i=1}^{j-1}\frac{\beta_j}{\alpha_j+\beta_j} = \mu_j\prod_{i=1}^{j-1}(1-\mu_i), \quad j=1,2,\ldots,k,$$

$$E[X_{k+1}] = E[1-X_1-X_2-\cdots-X_k] = \prod_{i=1}^{k}\frac{\beta_j}{\alpha_j+\beta_j} = \prod_{i=1}^{k}(1-\mu_i),$$

where $\mu_j = \alpha_j/(\alpha_j+\beta_j)$ for $j=1,2,\ldots,k$. Assume that the sampling data is the same as before, and let $n_j = y_j + y_{j+1} + \cdots + y_{k+1}$ for $j=1,2,\ldots,k$. Then in the posterior distribution,

$$\begin{aligned} E[X_1|\mathbf{y}] &= \frac{\alpha_1+y_1}{\alpha_1+\beta_1+n_1} = \frac{\alpha_1+\beta_1}{\alpha_1+\beta_1+n_1}\frac{\alpha_1}{\alpha_1+\beta_1} + \frac{n_1}{\alpha_1+\beta_1+n_1}\frac{y_1}{n_1} \\ &= u_1\mu_1 + (1-u_1)\frac{y_1}{n_1}, \end{aligned}$$

$$\begin{aligned} E[X_j|\mathbf{y}] &= \frac{\alpha_j+y_j}{\alpha_j+\beta_j+n_j}\prod_{i=1}^{j-1}\frac{\beta_i+n_{i+1}}{\alpha_i+\beta_i+n_i} \\ &= \left[u_i\mu_j + (1-u_j)\frac{y_j}{n_j}\right]\prod_{i=1}^{j-1}\left[u_i(1-\mu_i) + (1-\mu_i)\frac{n_{i+1}}{n_i}\right], \\ & j=2,3,\ldots,k, \end{aligned}$$

$$E[X_{k+1}|\mathbf{y}] = \prod_{i=1}^{k}\frac{\beta_j+n_{j+1}}{\alpha_j+\beta_j+n_j} = \prod_{i=1}^{k}\left[u_i(1-\mu_i) + (1-u_i)\frac{n_{i+1}}{n_i}\right],$$

where $u_j = (\alpha_j+\beta_j)/(\alpha_j+\beta_j+n_j)$. Let $Z_j = X_j/(1-X_1-\cdots-X_{j-1})$ for $j=1,2,\ldots,k$. Then μ_j is the mean value of Z_j. Note that y_j/n_j is not the sample mean of X_j when $j>1$. So, when **X** has a generalized Dirichlet distribution, the posterior mean of X_j does not have the same interpretation as in the Dirichlet distribution.

In the second experiment, if a ball drawn from the urn is put into box m for some $m<j$, then the content of box j will remain unchanged. So, $n_j = y_{j+1} + y_{j+2} + \cdots + y_{k+1}$ is the number of times that the content of box j is changed. Each change on box j is to put either a ball with color j or a marble into it. Hence, the ratio y_j/n_j represents the percentage that a ball with color j is put into box j given that the content of box j is changed. Therefore, box j in the first experiment represents variable X_j in a Dirichlet random vector, while box j in the second experiment represents variable Z_j corresponding to a generalized Dirichlet random vector. Let $\mu'_j = u_j\mu_j + (1-u_j)(y_j/n_j)$ be the posterior mean of Z_j. Then

$$E[X_j|\mathbf{y}] = \mu_j\prod_{i=1}^{j-1}(1-\mu_i'), \quad j = 1,2,\ldots,k,$$

$$E[X_{k+1}|\mathbf{y}] = \prod_{i=1}^{k}(1-\mu_i').$$

So, u_j should be thought of as the weight for Z_j. When the posterior means of the Z_j are known, calculating the posterior means of the X_j will not be difficult.

If $\mathbf{X}$ has a Dirichlet distribution, then X_j has a beta distribution for all j, and for any $j \neq m, X_j$ and X_m are negatively correlated. In addition, every permutation of $\mathbf{X}$ is completely neutral. This means that if the outcome of a test is not the outcome corresponding to X_j, then the impact on X_j is the same no matter what the outcome is. When an analyst is going to choose the Dirichlet distribution to be his prior, he should consider whether the random variables satisfy these conditions or not.

When an analyst chooses a Dirichlet distribution to be his prior, the remaining work for constructing such a prior is to determine the values of the parameters. First, the analyst must estimate the mean value for each random variable. Let $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_{k+1}$, and let μ_j be the mean value of X_j. Then we have $E[X_j] = \alpha_j/\alpha = \mu_j$ for $j = 1,2,\ldots,k$. There are $k+1$ unknown variables in k equations, so the solutions of the equations are not unique. There is one degree of freedom that can be used to adjust the spread of the distribution.

The variance of X_i is

$$\mathrm{Var}(X_i) = \frac{\alpha_i(\alpha_i+1)}{\alpha(\alpha+1)} - \left(\frac{\alpha_i}{\alpha}\right)^2 = \mu_i\left(\frac{\alpha\mu_i+1}{\alpha+1} - \mu_i\right).$$

Hence, the value of α represents the confidence of the analyst in his beliefs. The larger α is, the more confidence he has. For any $i \neq j$, the covariance between X_i and X_j is given by

$$\mathrm{Cov}(X_i,X_j) = \frac{\alpha_i\alpha_j}{\alpha(\alpha+1)} - \mu_i\mu_j = \mu_i\mu_j\left(-\frac{1}{\alpha+1}\right).$$

Thus, when α gets larger, the correlation between X_i and X_j gets weaker. This is another criterion for determining the value of α.

Alternatively, if $\mathbf{X}$ has a generalized Dirichlet distribution, $\mathbf{X}$ is also completely neutral, i.e., let $Z_i = X_i/(1 - X_1 - X_2 - \cdots - X_{i-1})$ for $i = 1,2,\ldots,k$, then the Z_i are independent. For any $j < m$,

$$X_j = Z_j\prod_{i=1}^{j-1}(1-Z_i),$$

$$\frac{X_m}{1-X_j} = \frac{Z_m \prod_{i=1}^{m-1}(1-Z_i)}{\prod_{i=1}^{j}(1-Z_i)} = Z_m \prod_{i=j+1}^{m-1}(1-Z_i).$$

Since the Z_i are independent, it follows that X_j and $X_m/(1-X_j)$ are independent for any $j < m$. Hence, a necessary condition for the analyst to assume that a random vector has a generalized Dirichlet distribution is: for any $j \neq m$, either X_j is independent of $X_m/(1-X_j)$, or X_m is independent of $X_j/(1-X_m)$, or both.

Since the order of the variables in a generalized Dirichlet random vector is not arbitrary, constructing a generalized Dirichlet distribution will be more difficult. We propose the following procedure to construct a generalized Dirichlet prior.

1. Let the number of random variables be $k+1$, and let $\Phi = \{1, 2, \ldots, k+1\}$. Set $I_j \leftarrow 0$ for $j = 1, 2, \ldots, k, d \leftarrow 1, \delta \leftarrow 1$.
2. If $d = k+1$, then stop. Otherwise, go to step 3.
3. Suppose that the values of all random variables not in set Φ are known. Let Δ be the set of the random variables in Φ that are negatively correlated with all other random variables in Φ; i.e., the set of random variables such that increasing the value of one of them will decrease the values of all other random variables in Φ. If Δ is empty, then it is inappropriate to assume that the prior distribution is a generalized Dirichlet distribution. If Δ has more than one element, then go to step 4; else let the variable in Δ be X_j and go to step 5.
4. Suppose that the values of all random variables not in set Φ are known. Then for all variables in Δ, pick one (say, X_j) such that if the outcome of a test is not the outcome corresponding to X_j, but is one of the outcomes corresponding to a different random variable in Φ, then the impact on X_j is the same no matter what the outcome is.
5. Set $I_d \leftarrow j, \Phi \leftarrow \Phi/\{j\}$. Estimate the mean value $\mu_j (< \delta)$ for X_j.
6. Let α_d and β_d be the parameters corresponding to variable X_j. Since both μ_j and δ are known, we can set $\alpha_d/(\alpha_d + \beta_d) = \mu_j/\delta$. Note that larger values of $\alpha_d + \beta_d$ imply greater confidence about the estimate of μ_j. Thus, for the specified value of the ratio $\alpha_d/(\alpha_d + \beta_d)$, choosing a value for $\alpha_d + \beta_d$ consistent with the confidence in μ_j will determine the values for α_d and β_d. Set $\delta \leftarrow \beta_d \delta/(\alpha_d + \beta_d)$, and $d \leftarrow d + 1$. If $d = k+1$, then stop. Otherwise, go to step 2.

By Property 3, $(X_{s+1}/V_s, X_{s+2}/V_s, \ldots, X_k/V_s \mid X_1, X_2, \ldots, X_s)$ has a generalized Dirichlet distribution, where $V_s = 1 - X_1 - X_2 - \cdots - X_s$. Hence, when the values of $X_1, X_2, \ldots, X_s$ are known, X_{s+1} will be negatively correlated with $X_{s+2}, X_{s+3}, \ldots$, and X_k. This result is used in step 3 to search for the variables that can be in position $s+1$. However, there may be more than one element in set Δ. Since for any $j \neq m, X_j$ and $X_m/(1-X_j)$ are independent if $j < m$. This concept is used in step 4 to find X_{s+1}. Since the analyst must determine the mean value and the sum of two parameters for each random variable except X_{k+1} in steps 5 and 6, respectively, the total number of values determined by the analyst in this approach is $2k$, where $k+1$ is the number of variables.

6. Summary

Although the generalized Dirichlet distribution has a more general covariance structure than the Dirichlet distribution, analysts did not pay too much attention on it. The two experiments introduced in Section 3 are hopefully useful in distinguishing the differences between the two distributions. This may makes the analysts to know when they should choose the generalized Dirichlet distribution instead of the Dirichlet distribution to be their priors.

The completely neutral property implies conditional independencies among random variables. When the distribution of a random vector is a Dirichlet distribution, every permutation of the variables in the random vector is completely neutral. So, the conditions for a random vector to have a Dirichlet distribution is restrictive, although the Dirichlet distribution is easy to construct and has some good properties in computation. The conditional independencies in a generalized Dirichlet random vector are much weaker. This suggests that the generalized Dirichlet distribution is a more reasonable prior for realistic cases, but the constructing and the computing for such a priori are more complex.

Since the generalized Dirichlet distribution is conjugate to the multinomial sampling, the posterior distribution is still a generalized Dirichlet distribution if the prior distribution is a generalized Dirichlet distribution. In addition, the way to generate samples for the generalized Dirichlet distribution is easy and applicable. Therefore, the work for simulating a random vector that has a generalized Dirichlet distribution in Bayesian analysis is generally tractable.

Acknowledgements

The author is very thankful to Dr. Vicki M. Bier for her valuable comments during the preparation of this paper.

References

[1] D.J. Spiegelhalter, N.L. Harris, K. Bull, R.C.G. Franklin, Empirical-evaluation of prior beliefs about frequencies – Methodology and a case-study in congenital heart disease, J. Amer. Statist. Assoc. 89 (1994) 435–443.

[2] C.D.M. Paulino, C.A.D. Pereira, Bayesian methods for categorical-data under informative general censoring, Biometrika 82 (1995) 439–446.

[3] K. Lange, Applications of the Dirichlet distribution to forensic match probabilities, Genetica 96 (1995) 107–117.
[4] R.J. Connor, J.E. Mosimann, Concepts of independence for proportions with a generalization of the Dirichlet distribution, J. Amer. Statist. Assoc. 64 (1969) 194–206.
[5] R.H. Lochner, A generalized Dirichlet distribution in Bayesian life testing, J. Royal Statist. Soc. B 37 (1975) 103–113.
[6] S.S. Wilks, Mathematical Statistics, Wiley, New York, 1962.
[7] R. Krzysztofowicz, S. Reese, Stochastic bifurcation processes and distributions of fractions, J. Amer. Statist. Assoc. 88 (1993) 345–354.