Chapter 3

Generalizations of the Dirichlet distribution

As discussed earlier, the Dirichlet distribution is not sufficiently general to model many compositional data sets. Because of the unit-sum constraint given in (2.1), compositional data will often tend to be negatively correlated, but compositional data sets can also exhibit positive correlation among some of their variables (see for example Reese and Krzysztofowicz, 1991). Also, the Dirichlet distribution has strong independence properties. Thus, many researchers have searched for improvements on the Dirichlet distribution. In this chapter, we will give known generalizations to the Dirichlet distribution on the simplex. Other known generalizations that were originally defined on the positive orthant will be given in the next chapter.

3.1 Connor-Mosimann distribution ($CM_n(\alpha,\lambda)$)

In this section, we will give a generalization of the Dirichlet distribution that was originally proposed by Connor and Mosimann (1969). Some of its properties are also given here.

Definition 3.1 $\mathbf{x} = (x_1, \cdots x_n) \in S^n$ is said to have a ***Connor-Mosimann distribution*** $\mathbf{x} \sim CM_n(\alpha, \lambda)$ if its density is given by

$$g(\mathbf{x}) = \prod_{i=1}^{n} B^{-1}(\alpha_i, \lambda_i) x_i^{\alpha_i - 1} (1 - \textstyle\sum_{j=1}^{i} x_j)^{\xi_i} , \tag{3.1}$$

where $B(\alpha_i, \lambda_i) = \Gamma(\alpha_i)\Gamma(\lambda_i)/\Gamma(\alpha_i + \lambda_i)$, $\xi_i = \lambda_i - \alpha_{i+1} - \lambda_{i+1}$ for i=1,...,n-1, $\xi_n = \lambda_n - 1$, and $\alpha_i > 0$ and $\lambda_i > 0$ for all i.

Note that if we set $\xi_i = 0$ (i.e., $\lambda_i = \alpha_{i+1} + \lambda_{i+1}$) for all i=1,...,n-1, we obtain a Dirichlet density.

Connor and Mosimann (1969) constructed this generalization by letting $w_1 = x_1 \sim Be(\alpha_1, \lambda_1)$, $w_2 = \frac{x_2}{1-x_1} \sim Be(\alpha_2, \lambda_2)$, $\cdots$, $w_n = \frac{x_n}{1-x_1-\cdots-x_{n-1}} \sim Be(\alpha_n, \lambda_n)$, where $Be(\alpha_i, \lambda_i)$ indicates the beta distribution. They then assumed that the w_i were independent, which yielded the Connor-Mosimann distribution. By theorem 2 of Connor and Mosimann (1969), this distribution always has complete right neutrality, since by definition $x_1, \frac{x_2}{1-x_1}, \cdots, \frac{x_n}{1-x_1-\cdots-x_{n-1}}$ are independent quantities. Remember, however, that complete right neutrality depends on the specified order of the x_i.

Connor and Mosimann (1969) discussed the moments of a random vector that has complete right neutrality, and gave the following relationships between the covariances of the variables in such a vector:

$$\mathrm{cov}(x_1, x_j) = -\frac{\mathrm{E}(x_j)}{\mathrm{E}(1-x_1)}\mathrm{var}(x_1), \ j = 2, \cdots, n+1 \tag{3.2}$$

$$\mathrm{cov}(x_i, x_j) = [L_j \,/\, L_{i+1}]\left[\prod_{m=i+1}^{j-1}(1-L_m)\right]\mathrm{cov}(x_i, x_{i+1}), \quad i < j. \tag{3.3}$$

where $L_r = \mathrm{E}(x_r/[1-x_1-\cdots-x_{r-1}])$ for all r.

Note that equations (3.2) and (3.3) are valid for any random vector that has complete right neutrality. Connor and Mosimann noted that equation (3.2) implies $\text{cov}(x_1, x_j) < 0$ for all $j = 2, \cdots, n+1$. Also, since $0 < L_r < 1$ for all *r*, equation (3.3) implies

$$\text{cov}(x_i, x_j) < 0 \ \ (> 0) \Leftrightarrow \text{cov}(x_i, x_{i+1}) < 0 \ \ (> 0), \quad i < j, \quad i, j \in \{2, \cdots, n+1\} \tag{3.4}$$

We now consider the moments of the Connor-Mosimann distribution $CM_n(\alpha, \lambda)$. Let $\alpha_{n+1} = 1$, $\lambda_{n+1} = 0$, and $M_{j-1} = \prod_{m=1}^{j-1}[(\lambda_m + 1)/(\alpha_m + \lambda_m + 1)]$ for $j = 1, \cdots, n+1$. Then we have:

$$\text{E}(x_j) = (\alpha_j / [\alpha_j + \lambda_j]) \prod_{m=1}^{j-1} [\lambda_m / (\alpha_m + \lambda_m)], j = 1, \cdots, n+1; \tag{3.5}$$

$$\text{var}(x_j) = \text{E}(x_j)\{[(\alpha_j + 1)/(\alpha_j + \lambda_j + 1)] M_{j-1} - \text{E}(x_j)\}, j = 1, \cdots, n+1; \tag{3.6}$$

and

$$\text{cov}(x_i, x_j) = \text{E}(x_j)[(\alpha_i / [\alpha_i + \lambda_i + 1]) M_{i-1} - \text{E}(x_i)], i = 1, \cdots, n;$$

$$j = i+1, \cdots n+1. \tag{3.7}$$

All of the above formulae are due to Connor and Mosimann (1969).

Lochner (1975) studied some properties of the Connor-Mosimann distribution, and applied his results to life-testing situations. In particular, he showed that:

$$\text{E}(x_i^m) = \{\prod_{j=1}^{i-1} \frac{(\lambda_j)_m}{(\alpha_j + \lambda_j)_m}\} \frac{(\alpha_i)_m}{(\alpha_i + \lambda_i)_m}, i = 2, \cdots, n \tag{3.8}$$

where $(a)_m = a(a+1) \cdots (a+m-1)$. Lochner noted that one can obtain the moments for $x_{n+1} = 1 - x_1 - \cdots - x_n$ from the relationships $x_{n+1} = \prod_{j=1}^{n}(1 - w_j)$ and $w_1 = x_1 \sim Be(\alpha_1, \lambda_1)$, and also proved the following lemma:

Lemma 3.1 If $\mathbf{x} \in S^n$ has the density function given by equation (3.1) and if $\mathrm{E}(x_i) = 1/(n+1)$ for $i = 1,2,\cdots,n$, then

(1) $\lambda_i = (n+1-i)\alpha_i$.

(2) $$\mathrm{var}(x_i) = \left[\prod_{j=1}^{i-1}\left\{\frac{(n+1-j)\alpha_j + 1}{(n+2-j)\alpha_j + 1}\right\}\right]\frac{\alpha_i + 1}{\{(n+2-i)\alpha_i + 1\}(n+1)} - \frac{1}{(n+1)^2}$$

for $i = 1,2,\cdots,n$.

(3) $\mathrm{var}(x_{n+1}) = \mathrm{var}(x_n)$.

(4) $\mathrm{E}(w_i) = 1/(n+2-i)$. Therefore, $\mathrm{E}(w_i)$ is monotonically increasing in *i*.

(5) $\mathrm{var}(x_i)$ is monotonically decreasing in α_j for $j \le i \le n+1$.

(6) $\mathrm{var}(x_i) \ge (\le)\, \mathrm{var}(x_{i+1})$ if $\alpha_i \le (\ge)\, \alpha_{i+1}$ for $i \le n-1$.

Also, for $1 \le r < s \le n+1$, we have

(7) $$\mathrm{cov}(x_r, x_s) = \frac{1}{n+1}\left\{\left(\prod_{j=1}^{r-1}\frac{\lambda_j + 1}{\alpha_j + \lambda_j + 1}\right)\frac{\alpha_r}{\alpha_r + \lambda_r + 1} - \frac{1}{n+1}\right\}.$$

(8) $\mathrm{cov}(x_r, x_s) > (<)\, \mathrm{cov}(x_{r+1}, x_s)$ if $\alpha_r > (<)\, \alpha_{r+1}$.

Finally, Lochner noted that for the Connor-Mosimann distribution, $\mathrm{cov}(x_r, x_s)$ is independent of *s* for $1 \le r < s \le n+1$, while for the usual Dirichlet distribution $\mathrm{cov}(x_r, x_s)$ is independent of both *r* and *s*. Further, for some parameter values of the Connor-Mosimann distribution, we have $\mathrm{cov}(x_r, x_s) > 0$ for all $1 < r < s \le n+1$. For example, using property (7) above, Lochner considers the case where $\lambda_j = (n+1-j)\alpha_j$ for $j = 1,\cdots,r$ and $\alpha_r = (n+3-r)\alpha_{r-1} + 1$. Smith (1994) stated that for the Connor-Mosimann distribution,

$\operatorname{cov}(x_2, x_3) > 0$ if $\alpha_2 + \lambda_2 > \lambda_1(\lambda_1 + \alpha_1 + 1)/\alpha_1$. Note that in this case by equation (3.4) we have $\operatorname{cov}(x_2, x_j) > 0 \;\; \forall j > 2$.

3.2 Dickey's distributions

Dickey (1968) defined the so-called ***scaled Dirichlet distribution*** by letting $\mathbf{y} \sim D_{n+1}(\boldsymbol{\alpha})$ and considering the transformation

$$x_i = \frac{c_i^{-1} y_i}{\sum_{j=1}^{n+1} c_j^{-1} y_j}, \; i = 1, \ldots, n+1, \tag{3.9}$$

where $\alpha_i > 0$ and $c_i > 0$ for all *i*. The density of **x** was shown to have the form

$$g(\mathbf{x}) = \frac{\Gamma(\sum_{i=1}^{n+1} \alpha_i) \prod_{i=1}^{n+1} c_i^{\alpha_i}}{\prod_{i=1}^{n+1} \Gamma(\alpha_i)} \frac{\prod_{i=1}^{n+1} x_i^{\alpha_i - 1}}{\left(\sum_{i=1}^{n+1} c_i x_i \right)^{\alpha_1 + \cdots \alpha_{n+1}}} \tag{3.10}$$

Aitchison (1986) pointed out that this distribution still retains much of the strong independence structure of the Dirichlet distribution. Dickey (1983) further generalized this distribution to the ***Dickey distribution***, which has a density of the form

$$g(\mathbf{x}) = A \prod_{i=1}^{n+1} x_i^{\alpha_i - 1} \left(\sum_{i=1}^{n+1} c_i x_i \right)^{\beta}, \tag{3.11}$$

where $\alpha_i > 0$, $\beta \in \Re$ and $c_i > 0$ for all *i*. The ***multiple Dickey distribution*** (Dickey, 1983, 1987) is a generalization of the Dickey distribution, and has a density of the form

$$g(\mathbf{x}) = A \left(\prod_{i=1}^{n+1} x_i^{\alpha_i - 1} \right) \prod_{j=1}^{K} \left(\sum_{i=1}^{n+1} c_{ij} x_i \right)^{\beta_j}, \tag{3.12}$$

where $\boldsymbol{\alpha} \in \Re_+^{n+1}$, $\boldsymbol{\beta} \in \Re^K$, and $\boldsymbol{C}$ is an $n+1 \times K$ matrix of nonnegative entries. In symbols, $\mathbf{x} \sim MD_{n+1}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{C})$ on $H_{n+1} = \{(x_1, \ldots, x_{n+1}) | \sum_{i=1}^{n+1} x_i = 1\}$ in $\Re_+^{n+1}$.

The scaled Dirichlet distribution and the Dickey distribution are special cases of the conditional generalized Liouville distribution (Smith, 1994), while the multiple Dickey distribution is not. Moreover, from the definitions of the conditional generalized Liouville distribution and the multiple Dickey distribution, it is easy to note that the former is not a special case of the latter. Smith mentioned that it is unclear whether Dickey's distributions have the ability to model positive covariance. Multiple Dickey distributions have been used in a variety of applications; see for example Dickey et al. (1987), Paulino and Pereira (1992, 1995), and Queen et al. (1994).

Dickey et al. (1987) proposed the multiple Dickey family as a prior distribution for finite-category sampling when some of the observations suffer missing category distinctions. For more details, see their paper. The output of this sampling can be either a collection of overlapping sets, or a collection of successively nested partitions, in which any two sets are either disjoint or nested. In the latter case, Dickey et al. showed that the multiple Dickey distribution could be expressed as a transformation of a product of independent Dirichlet vectors. A process that gives rise to such a collection of nested partitions can be seen as a *multi-furcation* process that first divides a unit into m_0 parts according to a multinomial process with parameters $\mathbf{y_0} = (y_{01}, y_{02}, \ldots, y_{0m_0})$, where the $\mathbf{y_0}$ are taken to be uncertain and Dirichlet distributed; the i^{th} resultant part is then divided into m_i parts according to another multinomial process independent of the first one with parameters $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im_i})$, where the $\mathbf{y}_i$ are taken to be uncertain and Dirichlet distributed; and so on until this multi-furcation produces the fractions of interest $x_1, x_2, \ldots, x_{n+1}$. This multi-furcation process can be represented graphically as shown in Figure 3.1 below:

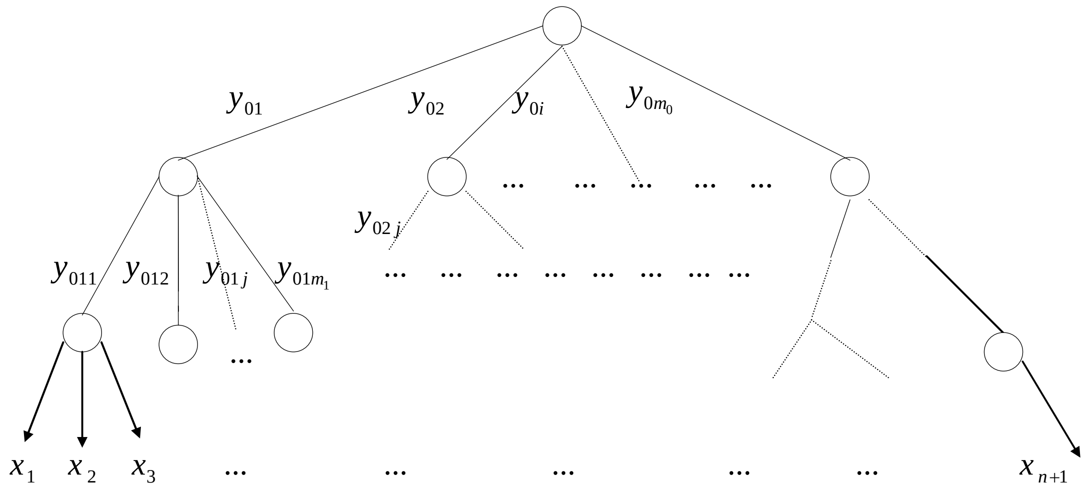

Figure 3.1. Multi-furcation topology underlying a collection of successively nested partitions.

Jiang et al. (1992) give closed-form expressions for the normalizing constant of the multiple Dickey distribution in the case of nested sets. Inference is also considerably simplified in this case. In the next section, we give a special case where all of the multi-furcations are reduced to bifurcations. The resultant family of distributions is called the adaptive Dirichlet family, and was constructed independently in 1993 by Krzysztofowicz and Reese. Therefore, the adaptive Dirichlet family of distributions is a subclass of the multiple Dickey family.