Proof

The density can be derived easily from theorem 4.1, noting that $\tilde{z}_i = (z_i/q_i)^{\beta_i}$, for $i = 1,\cdots,n+1$, with Jacobian $J(\tilde{\mathbf{z}} \to \mathbf{z}) = \prod_{i=1}^{n+1}(\beta_i/q_i)(z_i/q_i)^{\beta_i - 1}$ and $\tilde{\mathbf{z}} \sim L_{n+1}(\alpha_1/\beta_1,\ldots,\alpha_{n+1}/\beta_{n+1}; f)$. □

Smith (1994) gives the following theorems regarding generalized Liouville distributions on the positive orthant; the proofs can be found there.

Theorem 4.6 *Let* $\mathbf{z} = (z_1,\cdots,z_{n+1}) \sim GL_{n+1}^{\infty}(\alpha_1,\cdots,\alpha_{n+1},\beta_1,\cdots,\beta_{n+1},q_1,\cdots,q_{n+1}; f)$, *and consider the resulting composition* $\mathbf{x}$. *Split* $\mathbf{x}$ *into two parts,* $(x_1,\cdots,x_d)$ *and* $(x_{d+1},\cdots,x_n,1-x_1\cdots-x_n)$, *and consider the resulting partition of order 1. Then we have the following:*

1) $\mathbf{x}$ *has partition independence with respect to the above partition if and only if* $\beta_1 = \cdots = \beta_{n+1} = 1$, $q_1 = \cdots = q_d$ *and* $q_{d+1} = \cdots = q_{n+1}$.

2) $\mathbf{x}$ *has left neutrality with respect to the above partition if and only if* $\beta_1 = \cdots = \beta_d = 1$ *and* $q_1 = \cdots = q_d$.

3) $\mathbf{x}$ *has right neutrality with respect to the above partition if and only if* $\beta_{d+1} = \cdots = \beta_{n+1} = 1$ *and* $q_{d+1} = \cdots = q_{n+1}$.

4) $\mathbf{x}$ *has subcompositional independence with respect to the above partition if and only if* $\mathbf{x}$ *has either left neutrality or right neutrality.*

Theorem 4.7 *Let* $\mathbf{z} \sim GL_{n+1}^{\infty}(\alpha_1,\cdots,\alpha_{n+1},\beta_1,\cdots,\beta_{n+1},q_1,\cdots,q_{n+1}; f)$, *and consider the resulting composition* $\mathbf{x}$. *Then we have the following:*

1) $\mathbf{z}$ *has compositional invariance if and only if* $\beta_1 = \cdots = \beta_{n+1} = 1$ *and* $q_1 = \cdots = q_{n+1}$.

2) $\mathbf{x}$ *has complete right neutrality if and only if* $\beta_2 = \cdots = \beta_{n+1} = 1$ *and* $q_2 = \cdots = q_{n+1}$.

3) $\mathbf{x}$ *has complete left neutrality if and only if* $\beta_1 = \cdots = \beta_n = 1$ *and* $q_1 = \cdots = q_n$.

4) $\mathbf{x}$ *has complete subcompositional independence if and only if one of the following relationships is satisfied:*

i. $\beta_1 = \cdots = \beta_{n-1} = 1$ and $q_1 = \cdots = q_{n-1}$

ii. $\beta_3 = \cdots = \beta_{n+1} = 1$ and $q_3 = \cdots = q_{n+1}$

iii. $\beta_1 = \cdots = \beta_{d-1} = \beta_{d+1} = \cdots = \beta_{n+1} = 1$ and $q_1 = \cdots = q_{d-1} = q_{d+1} = \cdots = q_{n+1}$ for some $d \in \{3, \cdots, n-1\}$.

5) $\mathbf{x}$ *has complete n-subcompositional independence if and only if* $\beta_1 = \cdots = \beta_{d-1} = \beta_{d+1} = \cdots = \beta_{n+1} = 1$ *and* $q_1 = \cdots = q_{d-1} = q_{d+1} = \cdots = q_{n+1}$ *for some* $d \in \{1, \cdots, n+1\}$.

Smith (1994) notes that generalized Liouville distributions cannot "have subcompositional invariance without also having partition independence." Also, he derived a new class of distributions on the unit simplex by allowing the basis to have a generalized Liouville distribution on the positive orthant and then conditioning on the size of the basis. This new class, which is defined below, has the Dirichlet distribution as a special case.

Definition 4.4 $\mathbf{x} = (x_1, \cdots, x_n) \in S^n$ has a ***conditional generalized Liouville distrib-ution***, $CGL_n(\alpha_1, \cdots, \alpha_{n+1}, \beta_1, \cdots, \beta_{n+1}, q_1, \cdots, q_{n+1}; f)$, if $\mathbf{x}$ has density

$$g(\mathbf{x}) = A \prod_{i=1}^{n} \frac{x_i^{\alpha_i - 1}(1 - x_1 - \cdots - x_n)^{\alpha_{n+1}-1}}{\left(\sum_{i=1}^{n+1}(x_i / q_i)^{\beta_i}\right)^{\theta - 1}} \; f\left((\frac{x_1}{q_1})^{\beta_1} + \cdots + (\frac{x_n}{q_n})^{\beta_n} + (\frac{(1 - x_1 - \ldots - x_n)}{q_{n+1}})^{\beta_{n+1}} \right), \quad (4.10)$$

where $\theta = \sum_{i=1}^{n+1} \alpha_i / \beta_i$, $\alpha_i > 0$, $q_i > 0$, and $\beta_i > 0$ for all *i*, and f is a density function.

Smith noted that, unlike the generalized Liouville distribution on the simplex, this distribution is invariant with respect to the fill-up value.

Also, when $\beta_1 = \cdots = \beta_n = 1$, $q_1 = \cdots = q_n = 1$, and $\alpha_{n+1} = 1$, then the conditional generalized Liouville distribution approaches the standard Liouville distribution as $q_{n+1} \to \infty$. Thus, there should be no problem in finding choices of the function f that can model positive covariance. However, as we discussed in section 4.1, the covariance matrix of the Liouville distribution has a very restrictive form, and it is unclear the extent to which the conditional generalized Liouville distribution is an improvement over the Liouville distribution in this regard. Also, Smith notes that there is an infinite class of conditional generalized Liouville distributions to choose from, and it is not clear which choices of the function f can model both positive and negative covariances, as well as correlations close to 1 or -1. In general, it is currently very difficult to construct CGL distributions with desired covariance structures. In addition, the CGL distribution has a large number of uninterpretable parameters in its density function. This violates some of the criteria already established for successful simplex distributions in chapter 1.

Smith shows that:

1) Conditional generalized Liouville distributions can differentiate between partition independence, left neutrality, right neutrality, and subcompositional independence. Like generalized Liouville distributions on the positive orthant, however, they have no ability to represent subcompositional invariance.

2) Conditional generalized Liouville distributions are quite flexible for modeling complete independence concepts, just as they are for modeling independence concepts of a partition of

order 1. The necessary and sufficient parameter values for satisfying certain independence concepts can be found in Smith (1994).

Because the CGL distribution allows for fairly general covariance structures, it is difficult to get an intuitive sense of the dependence structure underlying this distribution. However, we give the following proposition to illustrate the possible covariance structures of the CGL distribution in the simple case of complete right neutrality.

Proposition 4.1 *Let* $\mathbf{x} \sim CGL_n(f\,;\alpha,\beta,\boldsymbol{q})$, *and let* $x_{n+1} = 1 - x_1 - \cdots - x_n$. *Then, if* $\beta_2 = \cdots = \beta_{n+1} = 1$ *and* $q_2 = \cdots = q_{n+1}$, *we have*

$$\begin{pmatrix} Cov(x_1, x_j) < 0 \quad for\,all \quad j = 2,\cdots,n+1 \\ \\ Cov(x_i, x_j) < 0(>0) \Leftrightarrow Cov(x_i, x_{i+1}) < 0(>0), \quad i < j, \quad i,j \in \{2,\cdots,n+1\} \end{pmatrix} \tag{4.11}$$

Proof

Smith (1994) shows that $\mathbf{x}$ has complete right neutrality if and only if $\beta_2 = \cdots = \beta_{n+1} = 1$ and $q_2 = \cdots = q_{n+1}$. Connor and Mosimann (1969) show that (4.11) holds for any vector that has complete right neutrality. □

Note that (4.11) in general will not hold for Liouville distributions or generalized Liouville distributions on the simplex, since they do not have complete right neutrality, except in the special case where **x** is Dirichlet distributed.

4.4 Conclusions

Fang et al. (1990) stated that "In analysing so-called 'compositional data,' scientists have been handicapped by the lack of known distributions to describe various patterns of variability." Aitchison (1986) has also discussed this problem. We believe that the approach outlined in section 4.1 will be useful in constructing distributions with desired covariance structures. This approach was also noted as promising by Fang et al. (1990). In chapter 6 and 7.5.2, we will extend this approach and then propose several new simplex and positive orthant distributions using this extension, some of which will be used in our applications.

Note that it is difficult to construct either a generalized Liouville distribution on the simplex or a CGL distribution with a desired covariance structure. In addition, these distributions have a large number of uninterpretable parameters in their density functions. Hence, we believe that these distributions are less well suited for practical use than the types of distributions we propose in chapter 5 and 6.