Statistical Papers 48, 265-282 (2007)

Statistical Papers

Perfect aggregation of Bayesian analysis on compositional data

Tzu-Tsung Wong

Institute of Information Management, National Cheng Kung University 1, Ta-Sheuh Road, Tainan City 701, Taiwan, ROC
(e-mail: tzutsung@mail.ncku.edu.tw)

Received: February 18, 2004; revised version March 7, 2005

Sufficiency is a widely used concept for reducing the dimensionality of a data set. Collecting data for a sufficient statistic is generally much easier and less expensive than collecting all of the available data. When the posterior distributions of a quantity of interest given the aggregate and disaggregate data are identical, perfect aggregation is said to hold, and in this case the aggregate data is a sufficient statistic for the quantity of interest. In this paper, the conditions for perfect aggregation are shown to depend on the functional form of the prior distribution. When the quantity of interest is the sum of some parameters in a vector having either a generalized Dirichlet or a Liouville distribution for analyzing compositional data, necessary and sufficient conditions for perfect aggregation are also established.

Keywords. Bayesian analysis, generalized Dirichlet distribution, Liouville distribution, perfect aggregation, sufficiency.

1 Introduction

Sufficiency is a widely used concept for reducing the dimensionality of a data set. Sufficiency implies that a reduced data set contains all useful information in the corresponding original data set. In most cases, collecting reduced data is much easier than collecting the corresponding original data, hence more convenient and less expensive. For instance, in flipping a coin, recording only the number of heads or tails in a sequence of tosses is generally easier than keeping track of the outcome of each toss.

Let the quantity of interest λ be a function of some parameters in $\{\theta_1, \theta_2, ..., \theta_k\}$, and let the data corresponding to λ and the θ_j be aggregate and disaggregate data (denoted by AD and DD), respectively. For instance, let θ_j represent the market share of company j for some particular product, and let Ψ be the set of all companies that produce the product of interest in a particular country. Then $\lambda = \Sigma_{j\in\Psi}\theta_j$ will be the market share for that country. In this case, let Q_j be the number of units of the product produced by company j (say, in a year). Then the aggregate data set is AD = $\{\Sigma_{\text{all } j} Q_j, \Sigma_{j\in\Psi} Q_j\}$, and the disaggregate data set is DD

$= \{Q_j \text{ for all } j\}$. If the posterior distributions for the quantity of interest in Bayesian analyses using the aggregate and disaggregate data are identical, then perfect aggregation holds.

Although collecting aggregate data is generally less costly than collecting disaggregate data, the results of an aggregate analysis can be inaccurate when perfect aggregation does not hold. Mosleh and Bier (1992) showed that using only aggregate data to estimate system reliabilities when disaggregate data are available can lead to significant aggregation error in both series and parallel systems. However, when perfect aggregation holds, the information resulting from an aggregate analysis will be accurate. Thus, the conditions for perfect aggregation are useful in determining whether collecting disaggregate data is necessary.

In this paper, perfect aggregation and sufficiency will first be shown to be equivalent. We will then focus on finding conditions for perfect aggregation for some multivariate distributions that are applicable in analyzing compositional data. In many problems involving nonnegativity and unit-sum constraints, analysts generally use Dirichlet distributions as prior distributions for Bayesian analysis for reasons of convenience. However, the Dirichlet prior is quite restrictive, especially with regard to its covariance structure. In this paper, we consider not only Dirichlet distribution, but also generalized Dirichlet and Liouville distributions, both of which include Dirichlet distribution as a special case.

In section 2, we will show that perfect aggregation and sufficiency are equivalent, and that the conditions for perfect aggregation depend on the functional form of a prior distribution. Some properties of Dirichlet, generalized Dirichlet, and Liouville distributions will be presented in section 3. These properties will then be used in section 4 to establish necessary and sufficient conditions for perfect aggregation when a prior distribution is one of these three distributions. In section 5, an application to the market shares of yogurt products in midwestern United States of America will be given. Conclusions and directions for future research are addressed in section 6.

2 Perfect aggregation

Bayesian analyses of a quantity of interest using aggregate and disaggregate data, respectively, are referred to as aggregate and disaggregate analyses (Bier, 1994). Without loss of generality, suppose that the joint distribution of $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ is $f(\theta)$. Let the quantity of interest λ be a function of some of the parameters in θ, and let Δ be the set of indices of those parameters; i.e., $\lambda = \Phi(\theta_j, j \in \Delta)$ for some function Φ. In an aggregate analysis, an aggregate prior $f(\lambda)$ is first derived from the prior $f(\theta)$; then the aggregate prior is updated using the aggregate data to obtain an aggregate posterior $f(\lambda|AD)$. By contrast, in a disaggregate analysis, we first update the prior $f(\theta)$ using the disaggregate data to obtain a posterior distribution $f(\theta|DD)$; then a disaggregate posterior $f(\lambda|DD)$ is derived from the posterior distribution $f(\theta|DD)$. If the aggregate and disaggregate posteriors are identical, then perfect aggregation holds. Perfect aggregation has also been referred to as exact aggregation (Simon and Ando, 1961) and total consistency (Ijiri, 1971).

Definition 1. Let DD be a set of samples from a distribution with an unknown parameter λ. A function T of DD is said to be a sufficient statistic for λ if the conditional distribution of DD, given $T(DD) = AD$, is independent of λ for all AD; i.e., if $L(DD|AD, \lambda) = L(DD|AD)$.

According to the Neyman factorization criterion (Zacks, 1971), the likelihood function $L(DD|\lambda)$ can be factorized into

$$L(DD|\lambda) = L(AD|\lambda)f(DD)$$

if and only if the aggregate data AD is a sufficient statistic for λ. Thus, the aggregate data AD will be sufficient for λ given DD if and only if $L(AD|\lambda) \propto L(DD|\lambda)$ (Lee, 1989). Sufficiency therefore depends on the functional form of the likelihood function $L(DD|\lambda)$.

Definition 2. Let the quantity of interest λ be a function of some parameters in $\theta = (\theta_1, \theta_2, ..., \theta_k)$, and let AD and DD be aggregate and disaggregate data, respectively. Then perfect aggregation holds if the posterior distributions $f(\lambda|AD)$ and $f(\lambda|DD)$ are identical.

Theorem 1. Perfect aggregation is equivalent to sufficiency.
Proof. If perfect aggregation holds (i.e., $f(\lambda|AD) = f(\lambda|DD)$), then we have

$$f(\lambda \mid DD) = \frac{L(DD \mid \lambda)f(\lambda)}{f(DD)} = \frac{L(DD \mid AD, \lambda)L(AD \mid \lambda)f(\lambda)}{f(DD)}$$

$$f(\lambda \mid AD) = \frac{L(AD \mid \lambda)f(\lambda)}{f(AD)}$$

$$\Rightarrow \frac{L(DD \mid AD, \lambda)}{f(DD)} = \frac{1}{f(AD)}$$

$$\Rightarrow L(DD \mid AD, \lambda) = \frac{f(DD)}{f(AD)}. \tag{1}$$

Since the right-hand side of equation (1) does not depend on λ, function $L(DD|AD, \lambda)$ must be independent of λ; i.e., $L(DD|AD, \lambda) = L(DD|AD)$.

Alternatively, if the aggregate data AD are a sufficient statistic for λ given the disaggregate data DD, then $L(DD|AD, \lambda) = L(DD|AD)$, which implies

$$f(\lambda \mid DD) = \frac{f(\lambda, DD)}{f(DD)} = \frac{L(DD \mid \lambda) f(\lambda)}{\int L(DD \mid \lambda) f(\lambda) d\lambda}$$

$$= \frac{f(DD \mid AD) L(AD \mid \lambda) f(\lambda)}{\int f(DD \mid AD) L(AD \mid \lambda) f(\lambda) d\lambda} \quad \text{(by sufficiency)}$$

$$= \frac{L(AD \mid \lambda) f(\lambda)}{\int L(AD \mid \lambda) f(\lambda) d\lambda} = \frac{f(\lambda, AD)}{f(AD)} = f(\lambda \mid AD).$$

Thus, perfect aggregation holds if the aggregate data AD are a sufficient statistic for θ. ♦

Note that the likelihood function $L(DD|\lambda)$ can be expressed as

$$L(DD \mid \lambda) = \frac{f(DD, \lambda)}{f(\lambda)} = \frac{\int_{\Phi(\theta_j, j \in \Delta) = \lambda} f(DD, \lambda, \theta) d\theta}{f(\lambda)}$$

$$= \frac{\int_{\Phi(\theta_j, j \in \Delta) = \lambda} f(DD, \theta) d\theta}{f(\lambda)} \tag{2}$$

In expression (2), both $f(DD,\theta)$ and $f(\lambda)$ depend on the functional form of the prior distribution $f(\theta)$. Furthermore, since the aggregate data AD are sufficient for λ if and only if condition $L(AD|\lambda) \propto L(DD|\lambda)$ holds, the conditions for perfect aggregation will also depend on the functional form of the prior distribution $f(\theta)$.

Example 1. Consider a two-component Bernoulli system in which the components are independent and connected in parallel. Let θ_j be the failure probability of component j for j = 1, 2; hence, the system failure probability is given by $\lambda = \theta_1\theta_2$, and we have $\Delta = \{1, 2\}$. Suppose that component 2 is tested only when component 1 fails. Let M_0 be the number of tests of component 1, and let M_1 and M_2 be the number of failures of components 1 and 2, respectively. So, we have $M_0 \geq M_1 \geq M_2$, and the aggregate and disaggregate data are $\{M_0, M_2\}$ and $\{M_0, M_1, M_2\}$, respectively. It has been shown that perfect aggregation holds if and only if θ_j has a beta distribution with parameters a_j and b_j for j = 1, 2 such that $a_1 = a_2+b_2$ (Bier, 1994; Gupta and Nadarajah, 2004). Since the aggregate data set is $\{M_0, M_2\}$, we have

$$L(AD \mid \lambda) = \binom{M_0}{M_2} \lambda^{M_2} (1-\lambda)^{M_0 - M_2}.$$

If θ_j has a beta distribution with parameters a_j and b_j for j = 1, 2, then by expression (2), it can be shown that $L(DD \mid \lambda) \propto \lambda^{M_2}(1-\lambda)^{M_0-M_2}$, hence the aggregate data AD is a sufficient statistic for the system failure probability λ, when $a_1 = a_2+b_2$. ♦

Since the definitions of sufficiency in both frequentist and Bayesian analysis are equivalent, the property $L(AD|\lambda) \propto L(DD|\lambda)$ can in principle be used to find conditions for perfect aggregation. However, both $\int_{\Phi(\theta_j, j\in\Delta)=\lambda} f(DD,\theta)d\theta$ and $f(\lambda)$ in expression (2) are generally difficult to obtain from $f(\theta)$. Thus, the property $L(AD|\lambda) \propto L(DD|\lambda)$ is often not helpful in finding the conditions for perfect aggregation for a specific prior distribution $f(\theta)$. Another approach to identify conditions for perfect aggregation when function Φ is the sum of some parameters in θ will be presented in section 4.

3 Some probability distributions defined on the unit simplex

In this section, we will introduce some multivariate distributions that can be used as prior distributions for compositional data that are subject to unit-sum and nonnegativity constraints. Note that these distributions have different covariance structures. In the rest of this paper, let the number of possible outcomes of an experiment be k+1, and let θ_j be the probability that a trial of the experiment results in outcome j. Hence, we have $\theta_1+\theta_2+...+\theta_{k+1} = 1$.

3.1 Dirichlet distribution

Definition 3. A parameter vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$ has a k-variate Dirichlet distribution with parameters $\alpha_j > 0$ for $j = 1, 2, ..., k+1$ if it has density

$$f(\theta) = \frac{\Gamma(\alpha_1 + \alpha_2 + ... + \alpha_{k+1})}{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_{k+1})} \prod_{j=1}^{k} \theta_j^{\alpha_j - 1} (1 - \theta_1 - ... - \theta_k)^{\alpha_{k+1}-1}$$

for $\theta_1+\theta_2+...+\theta_k \leq 1$ and $\theta_j \geq 0$ for $j = 1, 2, ..., k$. This distribution will be denoted $D_k(\alpha_1, \alpha_2, ..., \alpha_k; \alpha_{k+1})$.

The properties of the Dirichlet distribution can be found in Wilks (1962). In addition, the general moment function given in Lemma 1 below can also be used to prove the results in Lemma 2.

Lemma 1. If $\theta = (\theta_1, \theta_2, ..., \theta_k) \sim D_k(\alpha_1, \alpha_2, ..., \alpha_k; \alpha_{k+1})$, then the general moment function of θ is given by

$$E(\theta_1^{r_1}\theta_2^{r_2}...\theta_k^{r_k}) = \frac{\prod_{j=1}^{k}\Gamma(\alpha_j + r_j)\Gamma\left(\sum_{j=1}^{k+1}\alpha_j\right)}{\prod_{j=1}^{k}\Gamma(\alpha_j)\Gamma\left(\sum_{j=1}^{k+1}\alpha_j + \sum_{j=1}^{k} r_j\right)}.$$

Lemma 2. Let $\theta = (\theta_1, \theta_2, ..., \theta_k)$ be a parameter vector having a k-variate Dirichlet distribution $D_k(\alpha_1, \alpha_2, ..., \alpha_k; \alpha_{k+1})$.

(1) The marginal distribution of $(\theta_1, \theta_2, ..., \theta_s)$ for any $s < k$ is an s-variate Dirichlet distribution $D_s(\alpha_1, \alpha_2, ..., \alpha_s; \alpha_{s+1}+\alpha_{s+2}+...+\alpha_{k+1})$.

(2) Variable $\theta_1+\theta_2+...+\theta_k$ has a beta distribution with parameters $\alpha_1+\alpha_2+...+\alpha_k$ and α_{k+1}.

Definition 4. Let $\theta = (\theta_1, \theta_2, ..., \theta_k)$, and let $V_s = 1-\theta_1-...-\theta_s$ for some $s < k$. Then $(\theta_1, \theta_2, ..., \theta_s)$ is said to be neutral if it is independent of $(\theta_{s+1}/V_s, \theta_{s+2}/V_s, ..., \theta_k/V_s)$. If $(\theta_1, \theta_2, ..., \theta_s)$ is neutral for all $s < k$, then θ is said to be completely neutral.

Connor and Mosimann (1969) showed that every permutation of the parameters in a vector having a Dirichlet distribution is completely neutral. This implies that the order of the parameters in a vector having a Dirichlet distribution is arbitrary. For example, suppose that $(\theta_1, \theta_2, \theta_3) \sim D_3(\alpha_1, \alpha_2, \alpha_3; \alpha_4)$, and let $\theta_4 = 1-\theta_1-\theta_2-\theta_3$. Then $(\theta_3, \theta_1, \theta_2) \sim D_3(\alpha_3, \alpha_1, \alpha_2; \alpha_4)$, $(\theta_2, \theta_4, \theta_1) \sim D_3(\alpha_2, \alpha_4, \alpha_1; \alpha_3)$, and so on. Hence, part (1) of Lemma 2 can be generalized as follows:

Lemma 3. Let $\theta = (\theta_1, \theta_2, ..., \theta_k)$ follow a k-variate Dirichlet distribution $D_k(\alpha_1, \alpha_2, ..., \alpha_k; \alpha_{k+1})$. Then the marginal distribution of $(\theta_{n_1}, \theta_{n_2}, ..., \theta_{n_s})$ for $s < k$, $1 \le n_1 < n_2 < ... < n_s \le k$ is an s-variate Dirichlet distribution $D_s(\alpha_{n_1}, \alpha_{n_2}, ..., \alpha_{n_s}; \delta)$, where $\delta = (\alpha_1 + \alpha_2 + ... + \alpha_k) - (\alpha_{n_1} + \alpha_{n_2} + ... + \alpha_{n_s})$.

Note that any two parameters in a vector having a Dirichlet distribution are negatively correlated, and that the θ_j are marginally beta distributed for all j. Although the Dirichlet distribution is relatively tractable, a Dirichlet prior will often not be realistic in practice. For instance, in the case studied by Castillo et al. (1997), the unemployment proportions of the 17 regions of Spain are assumed to have a Dirichlet prior. However, it may be reasonable for the unemployment proportions in adjacent regions to be positively correlated, since people living in one region may work in a neighboring region. Thus, the Dirichlet distribution may not be a wholly appropriate prior for this case.

3.2 Generalized Dirichlet distribution

Definition 5. A parameter vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$ has a k-variate generalized Dirichlet distribution with parameters $\alpha_j > 0$ and $\beta_j > 0$ for $j = 1, 2, ..., k$ if it has density

$$f(\theta) = \prod_{j=1}^{k} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)} \theta_j^{\alpha_j - 1} (1 - \theta_1 - ... - \theta_j)^{\gamma_j}$$

for $\theta_1+\theta_2+...+\theta_k \le 1$ and $\theta_j \ge 0$ for $j = 1, 2, ..., k$, where $\gamma_j = \beta_j - \alpha_{j+1} - \beta_{j+1}$ for $j = 1, 2, ..., k-1$ and $\gamma_k = \beta_k - 1$. This distribution will be denoted $GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$.

For a parameter vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$, let $Z_1 = \theta_1$, and let $\theta_j = \theta_j/V_{j-1}$ for $j = 2, 3, ..., k$, where $V_{j-1} = 1-\theta_1-...-\theta_{j-1}$. If the Z_j are independent, then θ is completely neutral. Connor and Mosimann (1969) assumed that each of the Z_j has a beta distribution with parameters α_j and β_j, and derived the density function for the generalized Dirichlet distribution as shown in Definition 5. Wong (1998) used the concept of complete neutrality to derive the general moment function for the generalized Dirichlet distribution (as given in Lemma 4), and then to establish the property given in Lemma 5.

Lemma 4. Let $\theta = (\theta_1, \theta_2, ..., \theta_k)$ be a parameter vector having a k-variate generalized Dirichlet distribution $GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$. Then the general moment function of $(\theta_1, \theta_2, ..., \theta_k)$ is

$$E(\theta_1^{r_1}\theta_2^{r_2}...\theta_k^{r_k}) = \prod_{j=1}^{k}\frac{\Gamma(\alpha_j+\beta_j)\Gamma(\alpha_j+r_j)\Gamma(\beta_j+\delta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)\Gamma(\alpha_j+\beta_j+r_j+\delta_j)}$$

where $\delta_j = r_{j+1}+r_{j+2}+...+r_k$ for $j = 1, 2, ..., k-1$, and $\delta_k = 0$.

Lemma 5. Let $\theta = (\theta_1, \theta_2, ..., \theta_k)$ be a parameter vector having a k-variate generalized Dirichlet distribution $GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$. The marginal distribution of $(\theta_1, \theta_2, ..., \theta_s)$ for any $s < k$ is an s-variate generalized Dirichlet distribution $GD_s(\alpha_1, \alpha_2, ..., \alpha_s; \beta_1, \beta_2, ..., \beta_s)$.

When $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for $j = 1, 2, ..., k-1$, the generalized Dirichlet distribution reduces to a Dirichlet distribution. If $(\theta_1, \theta_2, ..., \theta_k)$ has a generalized Dirichlet distribution, then $(\theta_1, \theta_2, ..., \theta_k)$ is completely neutral. However, this does not mean that every permutation of $(\theta_1, \theta_2, ..., \theta_k)$ is also completely neutral. For instance, if $(\theta_1, \theta_2, \theta_3) \sim GD_3(\alpha_1, \alpha_2, \alpha_3; \beta_1, \beta_2, \beta_3)$ and $\beta_1 \neq \alpha_2+\beta_2$, then $(\theta_2, \theta_1, \theta_3)$ will not have a generalized Dirichlet distribution. So, when $(\theta_1, \theta_2, ..., \theta_k)$ has a generalized Dirichlet distribution, the order of the θ_j is generally not arbitrary.

In a generalized Dirichlet distribution, θ_1 is always negatively correlated with the other parameters. However, θ_j and θ_m can be positively correlated for $j, m > 1$ (Lochner, 1975). In particular, if there exists some $m > j$ such that θ_j and θ_m are positively (negatively) correlated, then θ_j will be positively (negatively) correlated with θ_i for all $i > j$. Since the generalized Dirichlet distribution has a more general covariance structure than the Dirichlet distribution, this makes the generalized Dirichlet distribution more practical and useful than the Dirichlet distribution, but it is substantially less tractable.

3.3 Liouville distribution

Definition 6. A parameter vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$ has a k-variate Liouville distribution with parameters $\alpha_j > 0$ for $i = 1, 2, ..., k$ and density generator $g(u)$ if it has density

$$f(\theta) = C_0 g(u) \prod_{j=1}^{k} \theta_j^{\alpha_j - 1}$$

for $\theta_1+\theta_2+...+\theta_k \leq 1$ and $\theta_j \geq 0$ for $j = 1, 2, ..., k$, where $u = \theta_1+\theta_2+...+\theta_k$ and C_0 is a normalizing constant. This distribution will be denoted $L_k(g(u); \alpha_1, \alpha_2, ..., \alpha_k)$.

Let $\mathbf{Z} \sim D_{k-1}(\alpha_1, \alpha_2, ..., \alpha_{k-1}; \alpha_k)$, and let U defined on [0,1] be an independent parameter with probability density function f(u). Fang et al. (1990) showed that $\theta = U\mathbf{Z}$ has a Liouville distribution $L_k(g(u); \alpha_1, \alpha_2, ..., \alpha_k)$, where $g(u) \propto u^{-(\alpha-1)} f(u)$ and $\alpha = \alpha_1+\alpha_2+...+\alpha_k$. This result can be used to derive the general moment function of the Liouville distribution, as given in Lemma 6 below.

Lemma 6. Let μ_r be the r^{th} moment of U; i.e., $\mu_r = E(U^r)$. If θ has a Liouville distribution $L_k(g(u); \alpha_1, \alpha_2, ..., \alpha_k)$, then the general moment function of θ is

$$E(\theta_1^{r_1} \theta_2^{r_2} ... \theta_k^{r_k}) = \mu_r \frac{\prod_{j=1}^{k} \Gamma(\alpha_j + r_j) \Gamma\left(\sum_{j=1}^{k} \alpha_j\right)}{\prod_{j=1}^{k} \Gamma(\alpha_j) \Gamma\left(\sum_{j=1}^{k} \alpha_j + r\right)},$$

where $r = r_1+r_2+...+r_k$.

When the density generating variate U has a beta distribution with parameters γ and ω such that $\gamma = \alpha_1+\alpha_2+...+\alpha_k$, by Lemma 6, we have $g(u) = (1-u)^{\omega-1}$ and

$$E(\theta_1^{r_1} \theta_2^{r_2} ... \theta_k^{r_k}) = \frac{\Gamma(\gamma+\omega)\Gamma(\gamma+r)}{\Gamma(\gamma+\omega+r)\Gamma(\gamma)} \cdot \frac{\prod_{j=1}^{k} \Gamma(\alpha_j + r_j) \Gamma\left(\sum_{j=1}^{k} \alpha_j\right)}{\prod_{j=1}^{k} \Gamma(\alpha_j) \Gamma\left(\sum_{j=1}^{k} \alpha_j + r\right)}$$

$$= \frac{\prod_{j=1}^{k} \Gamma(\alpha_j + r_j) \Gamma(\alpha_1 + \alpha_2 + ... + \alpha_k + \omega)}{\prod_{j=1}^{k} \Gamma(\alpha_j) \Gamma(\alpha_1 + \alpha_2 + ... + \alpha_k + \omega + r)}$$

which is the general moment function of the Dirichlet distribution $D_k(\alpha_1, \alpha_2, ..., \alpha_k; \omega)$. This means that when the density generating variate U has a beta distribution with parameters γ and ω, the Liouville distribution will reduce to a Dirichlet distribution if $\gamma = \alpha_1+\alpha_2+...+\alpha_k$. The following lemma proposed by Sivazlian (1981) presents some properties of the Liouville distribution.

Lemma 7. Let $\theta = (\theta_1, \theta_2, ..., \theta_k)$ be a parameter vector having a k-variate Liouville distribution $L_k(g(u); \alpha_1, \alpha_2, ..., \alpha_k)$. Then

(1) Subvector $(\theta_1, \theta_2, ..., \theta_s)$ for any $s < k$ has an s-variate Liouville distribution $L_s(h(u); \alpha_1, \alpha_2, ..., \alpha_s)$, where $h(u) = \int_0^{1-u} g(u+\tau)\tau^{\alpha_{s+1}+\alpha_{s+2}+...+\alpha_k-1} d\tau$.

(2) Variable $\theta_1+\theta_2+...+\theta_k$ has a univariate Liouville distribution $L_1(g(u); \alpha_1+\alpha_2+...+\alpha_k)$.

Fang et al. (1990) showed that the covariance of θ_i and θ_j in a parameter vector having a Liouville distribution is

$$\mathrm{Cov}(\theta_i, \theta_j) = \frac{\alpha_i \alpha_j}{\alpha}\left(\frac{\mu_2}{\alpha+1} - \frac{\mu_1^2}{\alpha}\right) \quad \text{for } i \neq j,$$

where $\alpha = \alpha_1+\alpha_2+...+\alpha_k$. Since α, α_i, and α_j are all positive, we have

$$\mathrm{Cov}(\theta_i, \theta_j) = \frac{\alpha_i \alpha_j}{\alpha}\left(\frac{\mu_2}{\alpha+1} - \frac{\mu_1^2}{\alpha}\right) > 0$$

$$\Leftrightarrow \sigma_U / \mu_1 > 1/\sqrt{\alpha},$$

where $\sigma_U^2 = \mathrm{Var}(U)$. Thus, for any $i \neq j$, θ_i and θ_j will be positively correlated if and only if the variate U has a coefficient of variation greater than $1/\sqrt{\alpha}$. Note that if there exist $i \neq j$ such that θ_i and θ_j are positively (negatively) correlated, then θ_m and θ_t must be positively (negatively) correlated for any $m \neq t$.

Wong (1998) notes that the generalized Dirichlet distribution allows different degrees of uncertainty about parameters with the same mean. This cannot happen for the Dirichlet distribution, nor for the Liouville distribution (except for parameter θ_{k+1}). Hence, the generalized Dirichlet distribution is substantially more flexible than the other two distributions considered here, although it is generally somewhat less tractable.

4 Conditions for perfect aggregation

When the prior distribution of a parameter vector is assumed to have a multivariate distribution defined on the unit simplex, the quantity of interest is often the sum of some parameters in the vector. Hence, the case where $\lambda = \Sigma_{j\in\Delta}\theta_j$ is likely to be widely applicable when θ has a multivariate distribution defined on the unit simplex. For instance, in the model of multibrand purchasing behavior developed by Chatfield and Goodhardt (1975), each brand of regular ground coffee has several package sizes. Purchase rates for the package sizes of each brand were computed across a group of customers. If we are interested in the purchase rate of one particular brand, the quantity of interest will be the sum of the purchase rates of the various package sizes of that brand.

Let DD = $\{y_1, y_2, ..., y_{k+1}\}$ be the disaggregate data collected for parameter vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$. Then the aggregate data will be AD = $\{n, y_0\}$, where $n = y_1+y_2+...+y_{k+1}$ and $y_0 = \Sigma_{j\in\Delta} y_j$. Suppose that the likelihood function of the disaggregate data $L(DD|\theta)$ follows a multinomial distribution. This implies that all tests are independent. Since we have $\lambda = \Sigma_{j\in\Delta}\theta_j$, the likelihood function of the aggregate data $L(AD|\lambda)$ will follow a binomial distribution. Note that the posterior density $f(\theta|DD)$ is proportional to the product $L(DD|\theta)f(\theta)$. It is not difficult to show that the Dirichlet, the generalized Dirichlet, and the Liouville distributions are all conjugate to the multinomial likelihood function. Thus, we give the following results without proof.

Lemma 8. When $\theta = (\theta_1, \theta_2, ..., \theta_k) \sim D_k(\alpha_1, \alpha_2, ..., \alpha_k; \alpha_{k+1})$ and $L(DD|\theta)$ follows a multinomial distribution, the posterior density $f(\theta|DD)$ is $D_k(\alpha_1', \alpha_2', ..., \alpha_k'; \alpha_{k+1}')$, where $\alpha_j' = \alpha_j + y_j$ for $j = 1, 2, ..., k+1$.

Lemma 9. When $\theta = (\theta_1, \theta_2, ..., \theta_k) \sim GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$ and $L(DD|\theta)$ follows a multinomial distribution, the posterior density $f(\theta|DD)$ is $GD_k(\alpha_1', \alpha_2', ..., \alpha_k'; \beta_1', \beta_2', ..., \beta_k')$, where $\alpha_j' = \alpha_j + y_j$ and $\beta_j' = \beta_j + y_{j+1} + \ldots + y_{k+1}$ for $j = 1, 2, ..., k$.

Lemma 10. When $\theta = (\theta_1, \theta_2, ..., \theta_k) \sim L_k(g(u); \alpha_1, \alpha_2, ..., \alpha_k)$ and $L(DD|\theta)$ follows a multinomial distribution, the posterior density $f(\theta|DD)$ is $L_k(h(u); \alpha_1', \alpha_2', ..., \alpha_k')$, where $\alpha_j' = \alpha_j + y_j$ for $j = 1, 2, ..., k$, and $h(u) = g(u)(1-u)^{y_{k+1}}$.

In the rest of this section, we will assume unless indicated otherwise that Δ is a subset of $\{1, 2, ..., k\}$. If $k+1$ is in Δ, then results similar to those presented below can be obtained by analyzing the conditions for perfect aggregation for $1-\lambda$ instead of λ.

When θ has a Dirichlet distribution, Lemmas 2, 3, and 8 can be used to show that both the aggregate and the disaggregate posteriors will have the same (beta) distribution (Azaiez, 1993). Thus, perfect aggregation always holds when θ has a Dirichlet prior. However, perfect aggregation does not always hold for the generalized Dirichlet and the Liouville distributions.

4.1 Generalized Dirichlet distribution

Recall that if θ has the generalized Dirichlet distribution $GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$ and $\beta_j = \alpha_{j+1} + \beta_{j+1}$ for $j = 1, 2, ..., k-1$, then the generalized Dirichlet distribution reduces to a Dirichlet distribution. In this case, by part (2) of Lemma 2, $\theta_1 + \theta_2 + ... + \theta_k$ will have a beta distribution with parameters $\alpha_1 + \alpha_2 + ... + \alpha_k$ and β_k.

Theorem 2. Let the distribution of θ be the generalized Dirichlet distribution $GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$, and let $q = \max\{j, j \in \Delta\}$. Suppose that $L(DD|\theta)$ follows a multinomial distribution. Then perfect aggregation holds if and only if $\beta_j = \alpha_{j+1} + \beta_{j+1}$ for $j = 1, 2, ..., q-1$.

Proof.

Sufficiency:

In an aggregate analysis, by Lemma 5, $(\theta_1, \theta_2, ..., \theta_q)$ has a q-variate generalized Dirichlet distribution. Since $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for $j = 1, 2, ..., q-1$, the joint distribution of $(\theta_1, \theta_2, ..., \theta_q)$ reduces to the Dirichlet distribution $D_q(\alpha_1, \alpha_2, ..., \alpha_q; \beta_q)$. By Lemma 3 and part (2) of Lemma 2, the aggregate prior is a beta distribution with parameters $\Sigma_{j\in\Delta}\alpha_j$ and $\Sigma_{j\notin\Delta, j\le q}\alpha_j + \beta_q$. Updating this prior with the aggregate data, the aggregate posterior $f(\lambda|AD)$ is a beta distribution with parameters $\Sigma_{j\in\Delta}\alpha_j + y_0$ and $\Sigma_{j\notin\Delta, j\le q}\alpha_j + \beta_q + n - y_0$. Alternatively, in a disaggregate analysis, by Lemma 9, the posterior of $\theta|y$ is a generalized Dirichlet distribution with parameters α_j' and β_j' for $j = 1, 2, ..., k$, where $\alpha_j' = \alpha_j+y_j$ and $\beta_j' = \beta_j+ y_{j+1}+...+y_{k+1}$. Since $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for $j = 1, 2, ..., q-1$, we have $\beta_j' = \beta_j + y_{j+1} + ... + y_{k+1} = \alpha_{j+1}' + \beta_{j+1}'$ for $j = 1, 2, ..., q-1$. By Lemma 5 and part (2) of Lemma 2, the disaggregate posterior $f(\lambda|DD)$ will be a beta distribution with parameters $\Sigma_{j\in\Delta}\alpha_j'$ and $\Sigma_{j\notin\Delta, j\le q}\alpha_j' + \beta_q'$. Since $\Sigma_{j\in\Delta}\alpha_j' = \Sigma_{j\in\Delta}\alpha_j + y_0$ and $\Sigma_{j\notin\Delta, j\le q}\alpha_j' + \beta_q' = \Sigma_{j\notin\Delta, j\le q}\alpha_j + \beta_q + n - y_0$, perfect aggregation holds.

Necessity:

Let $n_1, n_2, ..., n_s$ be the indices in Δ, and let $n_1 < n_2 < ... < n_s = q$. Then the disaggregate posterior mean of λ given DD will be $E(\lambda \mid DD) = \Sigma_{j=1}^{s} E(\theta_{n_j} \mid DD)$. None of the indices between n_{s-1} and n_s are included in Δ. Let $Z_1 = \theta_1$ and $Z_j = \theta_j/(1-\theta_1-...-\theta_{j-1})$ for $j = 2, 3, ..., k$. Then the disaggregate posterior mean of λ will be a sum of terms including

$$\begin{aligned} E(\theta_{n_{s-1}} \mid DD) &= E(Z_{n_{s-1}} \mid DD) \prod_{j=1}^{n_{s-1}-1} E(1 - Z_j \mid DD) \\ &= \frac{\alpha_{n_{s-1}}'}{\alpha_{n_{s-1}}' + \beta_{n_{s-1}}'} \prod_{j=1}^{n_{s-1}-1} \frac{\beta_j'}{\alpha_j' + \beta_j'} \\ &= \frac{\alpha_{n_{s-1}} + y_{n_{s-1}}}{\alpha_{n_{s-1}} + \beta_{n_{s-1}} + \delta_{n_{s-1}}} \prod_{j=1}^{n_{s-1}-1} \frac{\beta_j + \delta_{j+1}}{\alpha_j + \beta_j + \delta_j} \end{aligned}$$

and

$$E(\theta_{n_s} \mid DD) = \frac{\alpha_{n_s} + y_{n_s}}{\alpha_{n_s} + \beta_{n_s} + \delta_{n_s}} \prod_{j=1}^{n_s-1} \frac{\beta_j + \delta_{j+1}}{\alpha_j + \beta_j + \delta_j},$$

where $\delta_j = y_j+y_{j+1}+...+y_{k+1}$ for $j = 1, 2, ..., k+1$. Since perfect aggregation holds, the moments of the disaggregate posterior mean $E(\lambda|DD)$ must depend on the disaggregate data DD only through the aggregate data AD. Hence, the disaggregate posterior mean $E(\lambda|DD)$ should depend only on AD and the parameters in the prior distribution for θ, and this must be true for any possible DD corresponding to AD.

Note in particular that the term δ_{n_s} does not appear in the aggregate data, which means the disaggregate posterior mean E(λ|DD) cannot depend on δ_{n_s} if perfect aggregation holds. The factors that depend on δ_{n_s} in $E(\theta_{n_s})$ are $\beta_{n_s-1}+\delta_{n_s}$ and $\alpha_{n_s}+\beta_{n_s}+\delta_{n_s}$ in the numerator and denominator, respectively. These two terms must be exactly the same (hence cancel each other out) to have perfect aggregation for all possible disaggregate data sets corresponding to AD = {n, y_0}; i.e., $\alpha_{n_s}+\beta_{n_s}+\delta_{n_s} = \beta_{n_s-1}+\delta_{n_s}$, which implies that $\beta_{n_s-1}=\alpha_{n_s}+\beta_{n_s}$. By the same argument, we will have $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = n_{s-1}, $n_{s-1}+1$, ..., n_s-1. Similarly, by comparing $E(\theta_{n_{s-2}})$ with $E(\theta_{n_{s-1}})+E(\theta_{n_s})$, we will have $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = n_{s-2}, $n_{s-2}+1$, ..., $n_{s-1}-1$ if perfect aggregation holds and $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = n_{s-1}, $n_{s-1}+1$, ..., n_s-1. This process continues until $E(\theta_{n_1})$ is reached. Since E(λ|DD) should not depend on the δ_j for j = 2, 3, ..., n_1 if perfect aggregation holds, we will have $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = 1, 2, ..., n_1-1. Thus, if perfect aggregation holds, we must have $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = 1, 2, ..., q-1. ♦

By Theorem 2, it is possible to have perfect aggregation if $\beta_j \neq \alpha_{j+1}+\beta_{j+1}$ for some j ≥ q, so the joint prior of θ need not be a Dirichlet distribution. Note also that the aggregate prior being a beta distribution is not sufficient to ensure that $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = 1, 2, ..., q-1. For example, suppose that $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = 4, 5, ..., q-1, $\beta_1 = \alpha_3+\beta_3$, $\beta_3 = \alpha_2+\beta_2$, and $\beta_2 = \alpha_4+\beta_4$. Then by Lemma 4 and Lemma 5, it can be shown that $\Sigma_{j\in\Delta}\theta_j$ has a beta distribution with parameters $\Sigma_{j\in\Delta}\alpha_j$ and $\Sigma_{j\notin\Delta, j\le q}\alpha_j+\beta_q$. Although this aggregate prior is the same as the aggregate prior when $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = 1, 2, ..., q-1, perfect aggregation does not hold in this case.

Let Ω = {1, 2, ..., k+1}, and let r = max{j, j ∈ Ω\Δ}. If k+1 ∈ Δ, then we have $\lambda = 1-\Sigma_{j\notin\Delta}\theta_j$. In this case, the conditions for perfect aggregation will be $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = 1, 2, ..., r-1, because the aggregate and disaggregate posteriors for 1-λ will be the same (beta) distribution if $\beta_j = \alpha_{j+1}+\beta_{j+1}$ for j = 1, 2, ..., r-1.

4.2 Liouville distribution

Theorem 3. Let the distribution of θ be the Liouville distribution $L_k(g(u); \alpha_1, \alpha_2, ..., \alpha_k)$. Suppose that L(DD|θ) follows a multinomial distribution. Then perfect aggregation holds if and only if there exists some ω > 0 such that $g(u)=(1-u)^{\omega-1}$.

Proof. Assume without loss of generality that Δ = {1, 2, ..., s} for some s < k.
Sufficiency:
As discussed in section 3.3, when $g(u)=(1-u)^{\omega-1}$ for some ω > 0, the Liouville distribution reduces to a Dirichlet distribution, hence perfect aggregation holds.

Necessity:

In a disaggregate analysis, by Lemma 10, the joint posterior of $(\theta_1, \theta_2, ..., \theta_k)$ is a Liouville distribution $L_k(h_1(u);\ \alpha_1+y_1,\ \alpha_2+y_2,\ ...,\ \alpha_k+y_k)$, where $h_1(u) = g(u)(1-u)^{y_{k+1}}$. Then by part (1) of Lemma 7, the joint density of $(\theta_1, \theta_2, ..., \theta_s)|DD$ is $L_s(h_2(u); \alpha_1+y_1, ..., \alpha_s+y_s)$, where

$$h_2(u) = \int_0^{1-u} h_1(u+\tau)\tau^{\sum_{i=s+1}^{k}(\alpha_i+y_i)-1} d\tau$$

$$= \int_0^{1-u} g(u+\tau)(1-u-\tau)^{y_{k+1}}\tau^{\sum_{i=s+1}^{k}(\alpha_i+y_i)-1} d\tau.$$

By part (2) of Lemma 7, the disaggregate posterior is $L_1(h_2(u); \alpha_1+\alpha_2+...+\alpha_s+y_0)$. Let $\nu = \alpha_{s+1}+\alpha_{s+2}+ ...+\alpha_k$. Then $h_2(u) = \int_u^1 g(\tau)(1-\tau)^{y_{k+1}}(\tau-u)^{\nu+n-y_0-y_{k+1}-1} d\tau$. Since the value of $h_2(u)$ depends on the value of y_{k+1}, we will here write $h_2(u,y_{k+1})$ instead of $h_2(u)$. However, the distribution of the density generating variate in the disaggregate posterior must not depend on the value of y_{k+1}, because perfect aggregation holds. In particular, $h_2(u,0) = \int_u^1 g(\tau)(\tau-u)^{\nu+n-y_0-1} d\tau$ and $h_2(u,1) = \int_u^1 g(\tau)(1-\tau)(\tau-u)^{\nu+n-y_0-2} d\tau$. Using integration by parts, we have

$$h_2(u,1) = \frac{1}{\nu+n-y_0-1}\int_u^1 (\tau-u)^{\nu+n-y_0-1}[g(\tau) - g(\tau)'(1-\tau)]d\tau.$$

In order to have the same density generator, $h_2(u, 0)$ must be proportional to $h_2(u, 1)$; i.e., $h_2(u, 0) = C\times h_2(u, 1)$ for some constant C. Since this equation must hold for all possible values of n and y_0, we will have

$$g(\tau) = \frac{C}{\nu+n-y_0-1}[g(\tau) - g(\tau)'(1-\tau)]$$

$$\Rightarrow \log(g(\tau)) = \frac{C-(\nu+n-y_0-1)}{C}\log(1-\tau)$$

$$\Rightarrow g(\tau) = (1-\tau)^{\omega-1},$$

where $\omega = \dfrac{\nu+n-y_0-1}{C}$. As given in section 3.3, when $g(\tau) = (1-\tau)^{\omega-1}$, we have

$$f(u) = C_0 g(u)u^{\gamma-1} = C_0 u^{\gamma-1}(1-u)^{\omega-1},$$

where C_0 is a normalizing constant and $\gamma = \alpha_1+\alpha_2+...+\alpha_k$. Since f(u) is a probability density function and $\gamma > 0$, ω must be greater than zero. ♦

Theorem 3 implies that perfect aggregation will hold for the Liouville distribution only when the Liouville distribution reduces to a Dirichlet distribution. So, unless the density generating variate has a beta distribution with parameters γ and ω such that $\gamma = \alpha_1+\alpha_2+...+\alpha_k$, perfect aggregation will not hold.

When the density generating variate U has a beta distribution with parameters γ and ω, by Lemma 10, the posterior density $f(\theta|DD)$ is $L_k(h(u); \alpha_1', \alpha_2', ..., \alpha_k')$, where $\alpha_j' = \alpha_j+y_j$ for $j = 1, 2, ..., k$, and

$$h(u) = g(u)(1-u)^{y_{k+1}} = u^{\gamma-\alpha}(1-u)^{\omega+y_{k+1}-1}.$$

The density function of U in the posterior distribution $f(\theta|DD)$ is

$$f(u \mid DD) \propto u^{\alpha+n-y_{k+1}-1}h(u) = u^{\gamma+n-y_{k+1}-1}(1-u)^{\omega+y_{k+1}-1},$$

which implies that the density generating variate U given DD has a beta distribution with parameters $\gamma+n-y_{k+1}$ and $\omega+y_{k+1}$. The disaggregate posterior mean of λ will then be

$$E(\lambda \mid DD) = \sum_{j\in\Delta} E(\theta_j \mid DD) = E(U \mid DD)\frac{\sum_{j\in\Delta}\alpha_j'}{\alpha+n-y_{k+1}}$$

$$= \frac{(\gamma+n-y_{k+1})(\sum_{j\in\Delta}\alpha_j + y_0)}{(\gamma+\omega+n)(\alpha+n-y_{k+1})}, \tag{3}$$

Hence, it is easy to calculate the disaggregate posterior mean $E(\lambda|DD)$ when the density generating variate U has a beta distribution. When U does not follow a beta distribution, the disaggregate posterior mean $E(\lambda|DD)$ generally does not have a simple closed-form expression like expression (3).

5 An illustration

In the marketing study proposed by Balachander and Ghose (2003), Table 1 shows the market shares of yogurt products in midwestern United States of America. Let θ_1 through θ_9 be the parameters corresponding to the nine market shares, as shown in the first column of Table 1. Note that the first four yogurt products are produced by the same company Dannon. Suppose that this company is interested in estimating its whole yogurt market share instead of the market share of each yogurt product. This quantity of interest can be represented as $\lambda = \theta_1+\theta_2+\theta_3+\theta_4$.

In conducting a survey, let y_j be the number of customers favoring the yogurt product corresponding to parameter θ_j. Suppose that $\theta = (\theta_1, \theta_2, \ldots, \theta_8)$ follows the Dirichlet distribution $D_8(117.7, 19.7, 35.8, 1.2, 185.1, 81.8, 114.1, 98; 346.5)$, and that the aggregate data set is $\{n, y_0\} = \{1000, 250\}$ (i.e., after investigating the preference of 1000 customers, 250 of them favor the yogurt products produced by the Dannon company), where $n = y_1+y_2+\ldots+y_9$ and $y_0 = y_1+y_2+y_3+y_4$. It can be shown that the aggregate posterior of λ is a beta distribution with parameters 424.4 and 1575.5. Since perfect aggregation holds when the market shares have a Dirichlet distribution, the disaggregate posterior of λ will also be a beta distribution with parameters 424.4 and 1575.5 no matter what the disaggregate data corresponding the aggregate data AD = {1000, 250} are.

Table 1. The market shares of yogurt products in midwestern United States of America.

Parameter	Product	Market share
θ_1	Dannon low-fat	11.77
θ_2	Dannon non-fat	1.97
θ_3	Dannon fresh flavors	3.58
θ_4	Dannon mini-pack	0.12
θ_5	Nordica low-fat	18.51
θ_6	Wells-Bunny low-fat	8.18
θ_7	Weight Watcher's	11.41
θ_8	Yoplait non-fat	9.80
θ_9	Yoplait	34.65

Next, suppose that the market shares θ_1 through θ_8 have the generalized Dirichlet distribution $GD_8(16.6, 4.6, 1.3, 1.1, 20.3, 5.4, 26.3, 8.5; 124.4, 201.4, 30.1, 764, 70.3, 36.9, 102.6, 30.1)$. This distribution has the same mean values for all market shares as the Dirichlet distribution $D_8(117.7, 19.7, 35.8, 1.2, 185.1, 81.8, 114.1, 98; 346.5)$. Since the aggregate prior is not closed form, it is difficult to evaluate the aggregate posterior mean of λ in this case. Consider two possible disaggregate data sets $DD_1 = \{220, 10, 10, 10, 150, 100, 100, 100, 300\}$ and $DD_2 = \{10, 10, 10, 220, 150, 100, 100, 100, 300\}$ corresponding to the aggregate data set $\{n, y_0\} = \{1000, 250\}$. By Lemma 9, the joint posterior $f(\theta|DD_1)$ is the generalized Dirichlet distribution $GD_8(236.6, 14.6, 11.3, 11.1, 170.3, 105.4, 126.3, 108.5; 904.4, 971.4, 790.1, 1514, 670.3, 536.9, 502.6, 330.1)$, hence the disaggregate posterior mean of λ given DD_1 is

$$E(\lambda|DD_1) = 0.2074 + 0.0117 + 0.0110 + 0.0056 = 0.2357.$$

The same argument can be used to show that $E(\lambda|DD_2) = 0.1676$, which is approximately 71% of $E(\lambda|DD_1)$.

Finally, suppose that the market shares θ_1 through θ_8 have a Liouville distribution $L_8(g(u); 117.7, 19.7, 35.8, 1.2, 185.1, 81.8, 114.1, 98)$, and that the density generating variate U has a beta distribution with parameters $\gamma = 153.5$ and $\omega = 81.4$. Since $\gamma \neq \alpha = 653.4$, perfect aggregation does not hold even though the mean values of the market shares in this example are the same as those in $D_8(117.7, 19.7, 35.8, 1.2, 185.1, 81.8, 114.1, 98; 346.5)$. Since the aggregate prior is again not closed form, it is difficult to evaluate the aggregate posterior mean of λ. By expression (3), the disaggregate posterior mean for λ depends on the disaggregate data only through y_9. Hence, we consider two possible cases: $y_9 = 50$ and $y_9 = 500$. When $y_9 = 50$, the disaggregate posterior mean of λ is

$$E(\lambda \mid y_9 = 50) = \frac{(\gamma + n - y_9)(\alpha_0 + y_0)}{(\gamma + \omega + n)(\alpha + n - y_9)} = \frac{1103.5 \times 424.4}{1234.9 \times 1603.4} = 0.2365.$$

Similarly, when $y_9 = 500$, we have $E(\lambda|y_9{=}500) = 0.1947$, which is about 82% of $E(\theta|y_9{=}50)$.

By Theorem 2 and Theorem 3, perfect aggregation does not hold for either the generalized Dirichlet distribution $GD_8(16.6, 4.6, 1.3, 1.1, 20.3, 5.4, 26.3, 8.5; 124.4, 201.4, 30.1, 764, 70.3, 36.9, 102.6, 30.1)$ or the Liouville distribution $L_8(g(u); 117.7, 19.7, 35.8, 1.2, 185.1, 81.8, 114.1, 98)$ with $U \sim beta(153.5,81.4)$. Moreover, our examples have shown that different disaggregate data sets corresponding to the same aggregate data set can give rise to significantly different values for the disaggregate posterior means of λ.

6 Conclusions and directions for future research

When perfect aggregation holds, the aggregate data AD will be a sufficient statistic for the quantity of interest λ. Since both $f(DD, \theta)$ and $f(\lambda)$ for the disaggregate likelihood function $L(DD|\lambda)$ depend on $f(\theta)$, the conditions for perfect aggregation will depend on the functional form of the prior distribution $f(\theta)$. In this paper, we consider the case where λ is a sum of some parameters in θ, and assume that $L(DD|\theta)$ follows a multinomial distribution to find conditions for perfect aggregation when the prior distribution $f(\theta)$ is either a Dirichlet, a generalized Dirichlet, or a Liouville distribution.

As pointed out in section 3, the covariance structures of the Dirichlet, the generalized Dirichlet, and the Liouville distributions are quite different. When the parameters are all negatively correlated and each parameter has a beta distribution, the Dirichlet distribution can be an appropriate prior. If the parameters are all positively correlated, then the Liouville distribution may be a more appropriate choice. Finally, if some but not all of the parameters are positively correlated, then the generalized Dirichlet distribution can be a reasonable prior.

When perfect aggregation holds, collecting disaggregate data will not be necessary. If the joint distribution of θ is a Dirichlet distribution, then perfect aggregation always holds. However, when the joint distribution of θ is either a generalized Dirichlet or a Liouville distribution, the conditions for perfect aggregation are fairly restrictive. The aggregation error for these two distributions is almost inevitable, and two different disaggregate data sets corresponding to the same aggregate data set can yield significantly different disaggregate posterior means for λ, as illustrated in section 5. However, when the aggregation error is small, it may not be worthwhile to collect disaggregate data. Thus, an estimate of the aggregation error can be helpful in choosing between aggregate and disaggregate analyses, and methods for estimating aggregation error should be developed.

In general, identifying necessary conditions for perfect aggregation is much harder than identifying sufficient conditions. The general moment functions of the three multivariate distributions discussed in this paper are available in closed form, which greatly facilitates the identification of necessary conditions for perfect aggregation. However, when the general moment function of a multivariate distribution is not closed form, the method used in this paper to identify conditions for perfect aggregation will no longer be applicable. Thus, other methods for finding conditions for perfect aggregation should also be studied.

Acknowledgements

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

References

Azaiez MN (1993) Perfect Aggregation in Reliability Models with Bayesian Updating. Dissertation, University of Wisconsin-Madison.

Balachander S, Ghose S (2003) Reciprocal Spillover Effects: A Strategic Benefit of Brand Extensions. Journal of Marketing **67**, 4-13.

Bier VM (1994) On the Concept of Perfect Aggregation in Bayesian Estimation. Reliability Engineering and System Safety **46**, 271-281.

Castillo E, Hadi AS, Solares C (1997) Learning and Updating of Uncertainty in Dirichlet Models. Machine Learning **26**, 43-63.

Chatfield C, Goodhardt G (1975) Results Concerning Brand Choice. Journal of Marketing Research **12**, 110-113.

Connor RJ, Mosimann JE (1969) Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution. Journal of the American Statistical Association **64**, 194-206.

Fang KT, Kotz S, Ng KW (1990) Symmetric Multivariate and Related Distributions. Chapman and Hall, New York.

Gupta AK, Nadarajah S (2004) Handbook of Beta Distribution and its Applications, Chapter 3. Marcel Dekker, New York.

Ijiri Y (1971) Fundamental Queries in Aggregation Theory. Journal of the American Statistical Association **66**, 766-782.

Lee PM (1989) Bayesian Statistics: An Introduction. Oxford University Press, London.

Lochner RH (1975) A Generalized Dirichlet Distribution in Bayesian Life Testing. Journal of the Royal Statistical Society, Series B **37**, 103-113.

Mosleh A, Bier VM (1992) On Decomposition and Aggregation Error in Estimation: Some Basic Principles and Examples. Risk Analysis **22**, 203-214.

Simon H, Ando A (1961) Aggregation of Variables in Dynamic Systems. Econometrica **29**, 111-113.

Sivazlian BD (1981) On a Multivariate Extension of the Gamma and Beta Distributions. SIAM Journal on Applied Mathematics **41**, 205-209.

Wilks SS (1962) Mathematical Statistics. John Wiley, New York.

Wong TT (1998) Generalized Dirichlet Distribution in Bayesian Analysis. Applied Mathematics and Computation **97**, 165-181.

Zacks S (1971) The Theory of Statistical Inference. John Wiley & Sons, New York.