Computation of Carlson's Multiple Hypergeometric Function R for Bayesian Applications

by

Tom J. Jiang, Univ. of Massachusetts, Lowell
Joseph B. Kadane[1], Carnegie-Mellon Univ.
James M. Dickey[2], Univ. of Minnesota

Technical Report No. 566
October 1991

[1]Research of Prof. Kadane was supported in part by NSF grants DMS-9005858, DMS-8705646, DMS 8701770, SES-8900025 and ONR contract N00014-89-J-1851.

[2]Research of Prof. Dickey was supported in part by US National Science Foundation Grant DMS-8911548.

Abstract

Carlson's multiple hypergeometric functions arise in Bayesian inference, including methods for multinomial data with missing category distinctions and for local smoothing of histograms. To use these methods one needs to calculate Carlson functions and their ratios. We discuss properties of the functions and explore computational methods for them, including closed form methods, expansion methods, Laplace approximations, and monte carlo. Examples are given to illustrate and compare the methods.

1 Introduction

Let $\mathbf{u} = (u_1, u_2, \ldots, u_I)$ be a Dirichlet random vector with parameter $\mathbf{b} = (b_1, b_2, \ldots, b_I)$, denoted by $\mathbf{u} \sim D(\mathbf{b})$, where every $b_i > 0$. That is, $u_1, \ldots, u_{I-1}$ have the density function on the probability simplex $S_I = \{(u_1, \ldots, u_I) : u_i \geq 0, \sum_{i=1}^I u_i = 1\}$,

$$f(\mathbf{u}; \mathbf{b}) = B(\mathbf{b})^{-1} \prod_{i=1}^{I} u_i^{b_i - 1}, \tag{1.1}$$

where $B(\mathbf{b}) = [\prod_{i=1}^I \Gamma(b_i)]/\Gamma(b_+)$ and $b_+ = \sum_{i=1}^I b_i$. The following integral is a special case of Carlson's (1971, 1974) two-way multiple hypergeometric function, R_a. If we define $R(\mathbf{b}, G, -\mathbf{c}) = R_a(\mathbf{b}, G, -\mathbf{c})$, where $a = c_+$, then,

$$\begin{aligned} R(\mathbf{b}, G, -\mathbf{c}) &= E_{u|b} \textstyle\prod_{j=1}^J (\sum_{i=1}^I g_{ij} u_i)^{c_j} \\ &= \textstyle\int_{S_I} \frac{1}{B(\mathbf{b})} \left(\prod_{i=1}^I u_i^{b_i - 1} \right) \prod_{j=1}^J \left(\sum_{i=1}^I g_{ij} u_i \right)^{c_j} du_1 du_2 \ldots du_{I-1}. \end{aligned} \tag{1.2}$$

See also Dickey (1983) for an introduction to Carlson's functions for statisticians.

Relation (1.2) has suggested several statistical uses for Carlson's R (Dickey 1983; Dickey, Jiang and Kadane 1987; Dickey and Jiang 1991; and Dickey, Garthwaite, and Bian 1989). In particular, we discuss its usefulness for Bayesian local smoothing and Bayesian multinomial censored data in section 2. Properties of the R function are introduced in section 3. Section 4 gives several relevant computational methods for R and ratios of R. Examples are provided in Section 5. Section 6 summarizes our findings.

2 Bayesian applications

2.1 Local smoothness

Let $\mathbf{v} = (v_1, v_2, \ldots, v_J)$ be the unknown cell probabilities for multinomial histogram sampling. If the sample count data is reported as $x = (x_1, x_2, \ldots, x_J)$, then the likelihood function is

$$l_1(\mathbf{v}) = \binom{\mathbf{x}_+}{\mathbf{x}} \prod_{j=1}^{J} v_j^{x_j} \tag{2.1}$$

It is well known that the corresponding natural conjugate family is the Dirichlet distributions, $\mathbf{v} \sim D(\mathbf{b})$. The Dirichlet coordinate random variables are nearly independent, with a slight negative association, because of the constraint on their sum,

$$\text{Corr}\,(v_i, v_j) = -\left[\frac{w_i}{1-w_i}\right]\left[\frac{w_j}{1-w_j}\right], \tag{2.2}$$

where $\mathbf{w} = E\mathbf{v} = \mathbf{b}/b_+$. However, in the sense of prior belief, the sampling probabilities of adjacent cells would usually be positively prior correlated. Hence, a Dirichlet prior distribution is not appropriate for its lack of such smoothness.

To deal with this problem, Dickey and Jiang (1991) introduced filtered-variate Dirichlet distributions as priors for the likelihood function (2.1). That is, the prior-distributed random vector $\mathbf{v}$ is expressed as a linear transformation, $\mathbf{v} = \mathbf{u} \cdot G$, of a Dirichlet vector, $\mathbf{u} \sim D(\mathbf{b})$, where $\mathbf{u}$ and $\mathbf{b}$ are I-dimensional row vectors, and G is an $I \times J$ transformation matrix. Note that each of the rows of the matrix G must sum to one, as the sum of v_i's must be one. Denote this distribution by $\mathbf{v} \sim F_G D(\mathbf{b})$. Since the range of $\mathbf{v}$, which is the convex closure of the set of row-vector points of G, is a subset of the probability simplex S_J, it is too complicated to work directly with a density for $\mathbf{v}$. Therefore, we express and work with the distribution of $\mathbf{v}$ in terms of $\mathbf{u}$. We first reexpress the likelihood function (2.1) in terms of $\mathbf{u}$ as,

$$l_2(\mathbf{u}) = \binom{\mathbf{x}_+}{\mathbf{x}} \prod_{j=1}^{J} \left(\sum_{i=1}^{I} g_{ij} u_i\right)^{x_j}. \tag{2.3}$$

The prior predictive distribution for the count data $\mathbf{x}$ is the prior expectation of (2.3)

$$Pr(\mathbf{x}) = \binom{\mathbf{x}_+}{\mathbf{x}} R(\mathbf{b}, G, -\mathbf{x}), \tag{2.4}$$

with R as defined in (1.2). With this filtered-variate Dirichlet prior, the posterior density of $\mathbf{u}$, where $\mathbf{v} = \mathbf{u} \cdot G$ under multinomial sampling, is

$$f(\mathbf{u}; \mathbf{b}, G, \mathbf{x}) = B(\mathbf{b})^{-1} \left(\prod_{i=1}^{I} u_i^{b_i - 1}\right) \cdot \prod_{j=1}^{J} \left(\sum_{i=1}^{I} g_{ij} u_i\right)^{x_j} / R(\mathbf{b}, G, -\mathbf{x}). \tag{2.5}$$

The normalizing constant in (2.5) is an R function from (2.4). This posterior distribution of $\mathbf{u}$ is a generalized Dirichlet distribution denoted by $\mathbf{u} \mid \mathbf{x} \sim D(\mathbf{b}, G, \mathbf{x})$ (Dickey 1983). The

posterior distribution of $\mathbf{v}$ is then the filtered-variate generalized Dirichlet distribution $\mathbf{v}|\mathbf{x} \sim F_G D(\mathbf{b}, G, \mathbf{x})$. A general moment of $\mathbf{v}$ can be defined by taking the expectation of

$$\prod_{j=1}^{J} v_j^{d_j}. \tag{2.6}$$

For various choices of $\mathbf{d}$, this can specify coordinate means and second moments, mixed second moments, etc. The posterior expectation of (2.6) is simply the ratio,

$$E\left(\prod_{j=1}^{J} v_j^{d_j} \mid \mathbf{x}\right) = \frac{R(\mathbf{b}, G, -\mathbf{x}-\mathbf{d})}{R(\mathbf{b}, G, -\mathbf{x})}. \tag{2.7}$$

See Dickey and Jiang (1991) on Bayesian use of these prior and posterior distributions for problems of local smoothing.

2.2 Bayesian analysis of censored discrete data

Consider multinomial sampling when some of the observations suffer missing distinctions between categories. Let $\mathbf{y} = (y_1, y_2, \ldots, y_I, y_{12}, \ldots, y_{12\ldots I})$ be the vector of frequency counts for sampling from a distribution on a finite sample space having the unknown probability vector $\mathbf{u} = (u_1, u_2, \ldots, u_I)$, where y_i is the frequency count of observations in category i and y_{ij} is the frequency count of observations reported not as a unique category, such as i or j, but rather as the set of two categories $\{i, j\}$. And in general, y_σ is the frequency count of observations that are reported as falling in the set of categories σ, where σ is a non-empty subset of $1, 2, \ldots, I$. If $g_{i,\sigma}$ is the probability that an observation that is actually in category i will be reported as being in set σ, then the likelihood function is

$$\binom{\mathbf{y}+}{\mathbf{y}} \prod_{\sigma} \left[\sum_{i=1}^{I} g_{i,\sigma} u_i\right]^{y_\sigma}, \tag{2.8}$$

where σ varies over a specified class of sets. Typically, many of the $g_{i,\sigma}$'s would be zero. For example, if $g_{i,\sigma} = 0$ unless $\sigma = \{i\}$, there are no category confusions. If $g_{i,\sigma} = g_{j,\sigma}$ for all $i, j \in \sigma$, and zero otherwise, then the data are non-informatively censored. Otherwise, the pattern of censoring in the data, itself, is informative.

Suppose now that $\mathbf{u}$ has the prior Dirichlet distribution, $\mathbf{u} \sim D(\mathbf{b})$. Then the Bayesian predictive distribution for the reported vector $\mathbf{y}$ is

$$Pr(\mathbf{y}) = \binom{\mathbf{y}_+}{\mathbf{y}} R(\mathbf{b}, G, -\mathbf{y}), \tag{2.9}$$

where G is the probabilities $g_{i,\sigma}$ arranged in matrix form, so that G is an $I \times J$ matrix, where J is the number of censoring subsets in the model. The posterior distribution for $\mathbf{u}$ is $D(\mathbf{b}, G, \mathbf{y})$, with *p.d.f.*,

$$f(\mathbf{u} \mid \mathbf{y}) = B(\mathbf{b})^{-1} \left(\prod_{i=1}^{I} u_i^{b_i - 1} \right) \cdot \left(\prod_{\sigma} \left[\sum_{i=1}^{I} g_{i,\sigma} u_i \right]^{y_\sigma} \right) / R(\mathbf{b}, G, -\mathbf{y}). \tag{2.10}$$

Thus, the Dirichlet family is not conjugate for the multinomial missing data likelihood. However, using the generalized Dirichlet prior distribution $D(\mathbf{b}, G, \mathbf{c})$, yields the posterior distribution $D(\mathbf{b}, G, \mathbf{c} + \mathbf{y})$. Thus, the generalized Dirichlet is conjugate to the multinomial missing data model. Under the Dirichlet prior, the posterior moment is

$$E\left(\prod_{i=1}^{I} u_i^{d_i} \mid \mathbf{y} \right) = \frac{B(\mathbf{b} + \mathbf{d})}{B(\mathbf{b})} \cdot \frac{R(\mathbf{b} + \mathbf{d}, G, -\mathbf{y})}{R(\mathbf{b}, G, -\mathbf{y})}. \tag{2.11}$$

See Dickey, Jiang and Kadane (1987) for further discussion on Bayesian inference for censored data.

To use these methods effectively, it is necessary to compute R and ratios of R. Before presenting methods for computing the functions R and their ratios (section 4), we first give relevant properties of R.

3 Properties of Carlson's R

In this section, we give properties of Carlson's R function, which simplify its computation.

If the set of columns of matrix G and the vector $\mathbf{c}$ are permuted conformably, the corresponding R retains its value. Formally,

Lemma 3.1 *If $\tilde{G} = GP$ and $\tilde{\mathbf{c}} = \mathbf{c} \cdot P$, where P is a permutation matrix, then*

$$R(\mathbf{b}, G, -\mathbf{c}) = R(\mathbf{b}, \tilde{G}, -\tilde{\mathbf{c}}).$$

The following consequence of the definition (1.2) of R, is useful in Bayesian missing-distinction problems in which the high-dimentional vector $\mathbf{y}$ is sparse.

Lemma 3.2 *If, without loss of generality, the vector* $\mathbf{c}$ *is taken in the form* $\mathbf{c} = (\mathbf{c}^{(1)}, \mathbf{o})$, *where* $\mathbf{c}^{(1)}$ *has* $J^{(1)} < J$ *coordinates, then*

$$R(\mathbf{b}, G, -\mathbf{c}) = R(\mathbf{b}, G^{(1)}, -\mathbf{c}^{(1)}),$$

where $G^{(1)}$ *consists of the first* $J^{(1)}$ *columns of* $G = (G^{(1)}, G^{(2)})$.

The following corollary is an extreme case of Lemma 3.2.

Corollary 3.3 *If* $\mathbf{c} = \mathbf{0}$, *then*

$$R(\mathbf{b}, G, -\mathbf{c}) = 1.$$

Two dimension-reduction lemmas are given next.

Lemma 3.4 *Conformably partition* $G = (G^{(1)}, G^{(2)})$ *and* $\mathbf{c} = (\mathbf{c}^{(1)}, \mathbf{c}^{(2)})$. *If each entry of* $G^{(2)}$ *is a one, then*

$$R(\mathbf{b}, G, -\mathbf{c}) = R(\mathbf{b}, G^{(1)}, -\mathbf{c}^{(1)}).$$

Lemma 3.5 *Conformably partition* $G = (G^{(1)}, G^{(2)}, G^{(3)})$ *and* $\mathbf{c} = (\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \mathbf{c}^{(3)})$. *If* $G^{(1)} = G^{(3)}$, *then*

$$R(\mathbf{b}, G, -\mathbf{c}) = R(\mathbf{b}, \tilde{G}, -\tilde{\mathbf{c}}), \text{ where } \tilde{G} = (G^{(1)}, G^{(2)}) \text{ and } \tilde{\mathbf{c}} = (\mathbf{c}^{(1)} + \mathbf{c}^{(3)}, \mathbf{c}^{(2)}).$$

The next lemma shows the relationship between R's when their corresponding matrices have proportional columns. The asterisk in a subscript denotes the list of possible values of that subscript. Hence, $\mathbf{g}_{*j} = (g_{1j}, g_{2j}, \ldots, g_{Ij})^T$ and $\mathbf{g}_{i*} = (g_{i1}, g_{i2}, \ldots, g_{iJ})$. We use this notation throughout the paper.

Lemma 3.6 *Let* $\mathbf{g}_{*j}$ *and* $\mathbf{h}_{*j}$ *be the* j*-th column vectors of matrices* G *and* H *respectively and assume* $\mathbf{g}_{*j} = e_j \cdot \mathbf{h}_{*j}$, *for scalar* e_j *for all* j, *then*

$$R(\mathbf{b}, G, -\mathbf{c}) = (\prod_{j=1}^{J} e_j^{c_j}) \cdot R(\mathbf{b}, H, -\mathbf{c}).$$

□

If the last I columns of the matrix G form an identity matrix, the following dimension-reduction lemma applies.

Lemma 3.7 *Define* $G = (G^{(1)}, G^{(2)})$ *and conformably,* $\mathbf{c} = (\mathbf{c}^{(1)}, \mathbf{c}^{(2)})$. *If* $G^{(2)}$ *is an identity matrix, then*

$$R(\mathbf{b}, G, -\mathbf{c}) = \frac{B(\mathbf{b} + \mathbf{c}^{(2)})}{B(\mathbf{b})} R(\mathbf{b} + \mathbf{c}^{(2)}, G^{(1)}, -\mathbf{c}^{(1)}).$$

Our final lemma extends the applicability of the preceding and other properties of R by interchanging the roles of the rows and columns of the matrix argument G. This identity, which generalizes Picard's classical identity regarding Lauricella's F_D, was given by Dickey (1968) before the introduction of two-way R by Carlson (1971). It can have an important effect on the dimension of integration, but because it can create a pole in the density, it is most often useful as a tool for accessing further simplifying relations, series expansions, and methods of computing R. For the Carlson function, in general, $R_a(b, G, -c) = R_a(-c, G^T, b)$. Our integral form R requires $a = -c_+$ and then, for the right-hand side, $a = b_+$. This is obtained by extending the matrix G according to Lemma 3.4 yielding the following.

Lemma 3.8 Define $c^* = (c, -(b_+ + c_+))$ and $G^* = (G, 1_J)$, where $1_J = (1, \ldots, 1)^T$. Then

$$R(b, G, -c) = R(-c^*, G^{*T}, b).$$

4 Computation of R functions and their ratios

As defined in subsection 2.2, if, for any $g_{i,\sigma}$,

$$g_{i,\sigma} = \begin{cases} e_\sigma & \text{all } i \in \sigma \\ 0 & \text{otherwise} \end{cases}$$

then the pattern of observed missing category distinctions is not itself informative. The matrix G has the same quantity in each nonzero entry of a column. We define a matrix as an *indicator matrix* if each entry is either 0 or 1. By lemma 3.6, R with this matrix G is proportional to an R having an indicator matrix parameter. Therefore, the discussion of the computation of R, when it refers to Bayesian noninformatively missing data problems, can be restricted to R with an indicator matrix parameter. When G takes the special form of a nested partition indicator, defined in subsection 4.1 below, R is expressible in a closed form. When G does not take such a special form, R can be represented by a summation of closed forms, as in our expansion method

of sub-section 4.2. When neither of these two methods is feasible, one may use an approximation method or a simulation method, discussed in sub-sections 4.3 and 4.4, respectively.

4.1 Closed form methods

Carlson's $R(\mathbf{b}, G, -\mathbf{y})$ function is considerably simplified when its matrix parameter G is an *n-level nested-partition indicator*, to be defined later. But first, we define *the j-th level nested-partition subsets for* $j = 1, \ldots, n$.

Let $S = \{1, 2, \ldots, I\}$, then $\{S_{k_1}\}_{\text{all } k_1} = \{S_1, S_2, \ldots, S_K\}$ is said to be the set of 1^{st} *level nested-partition subsets of* S, if all S_{k_1}'s are mutually exclusive and exhaustive subsets of S. That is,

1. $S_i \cap S_j$, for all $i \neq j$
2. $\cup_{\text{all } k_1} S_{k_1} = S$.

Similarly, for $j = 2, \ldots, n$, we define $\{S_{k_1:k_2:\ldots:k_{j-1}:k_j}\}_{\text{all } k_j}$ to be the set of *j-th level nested-partition subsets of* $S_{k_1:k_2:\ldots:k_{j-1}}$ if all subsets are mutually exclusive and exhaustive subsets of $S_{k_1:k_2:\ldots:k_{j-1}}$. If each of the n^{th} level nested-partition subsets contains a single entry, so that the number of n^{th} level nested-partition subsets is I, then we may say that any of the subsets defined above is an *n-level nested-partition subset of* S. These subsets can be shown in the following tree diagram.

.....insert figure 1 here.....

By these definitions, we have $S \supset S_{k_1} \supset S_{k_1:k_2} \supset \ldots \supset S_{k_1:k_2:\ldots:k_n}$. Before defining an n-level nested-partition indicator matrix, we need to define an indicator vector of an n-level nested-partition subset S_σ of S. An $I \times 1$ vector is said to be *an indicator vector of* S_σ, where S_σ is a subset of S, if i-th entry of the vector is 1, if $i \in S_\sigma$, and is 0, otherwise.

We are now ready to define an n-level nested-partition indicator matrix. A matrix G is said to be an *n-level nested-partition indicator matrix* if there are n-level nested-partition subsets so that each column vector of G is an indicator vector of each of these subsets.

For convenience, in this subsection, we shall refer to the j-th column of G and the j-th entry of $\mathbf{y}$ as column σ of G and y_σ, respectively, if the j-th column vector of G is the indicator

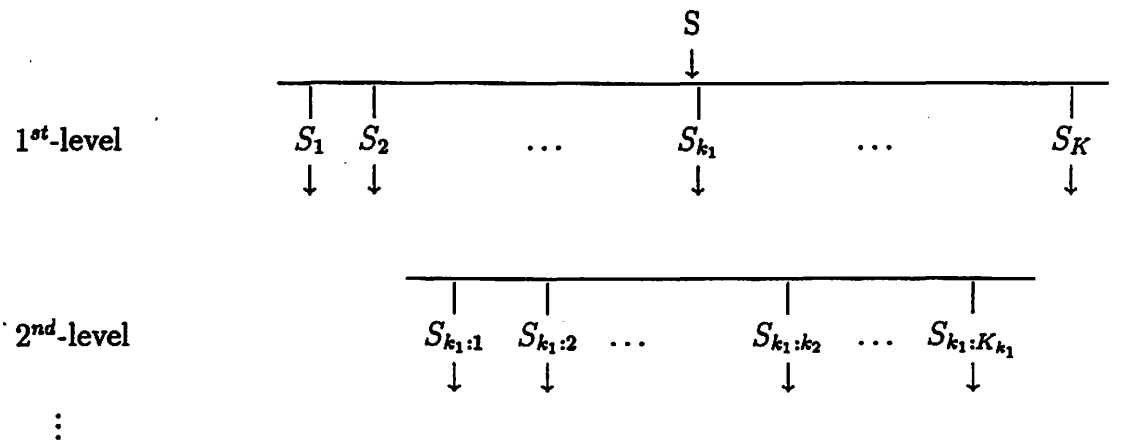

Figure 1

vector of S_σ. For example, we have column 1 or column 2:1, etc and y_1 or $y_{2:1}$, etc. Further, we use (σ), as the subscript of u(or b) to denote the sum of the cell probabilities (or parameters) corresponding to the nested-partition subset S_σ, that is $u_{(\sigma)} = \sum_{i \in S_\sigma} u_i$ (and $b_{(\sigma)} = \sum_{i \in S_\sigma} b_i$). For example, $u_{(2:1)} = \sum_{i \in S_{2:1}} u_i, b_{(1:3:2)} = \sum_{i \in S_{1:3:2}} b_i$. If we define $u_{(\sigma:+)} = \sum_{\text{all } k_j} u_{(\sigma:k_j)}$, then $u_{(\sigma)} = u_{(\sigma:+)}$. For example, $u_{(3:2)} = u_{(3:2:+)} = \sum_{\text{all } k_3} u_{(3:2:k_3)}$ and $b_{(4)} = b_{(4:+)} = \sum_{\text{all } k_2} b_{(4:k_2)}$.

As in section 3, we use an asterisk to denote a vector of possible entries. So, we have $\mathbf{u}_{(\sigma:*)} = (u_{(\sigma:1)}, u_{(\sigma:2)}, \ldots, u_{(\sigma:K_\sigma)})$ and $\mathbf{b}_{(\sigma:*)} = (b_{(\sigma:1)}, b_{(\sigma:2)}, \ldots, b_{(\sigma:K_\sigma)})$.

We also use brackets instead of parentheses in the subscript of u to indicate that $\mathbf{u}_{[\sigma:*]}$ is a probability vector. That is, $\mathbf{u}_{[\sigma:*]} = (u_{[\sigma:1]}, u_{[\sigma:2]}, \ldots, u_{[\sigma:K_\sigma]})$, where $u_{[\sigma:i]} = u_{(\sigma:i)}/u_{(\sigma)}$ for all $i = 1, \ldots, K_\sigma$.

The following theorem can be derived by transformation of variables.

Theorem 4.1 *In terms of n-level nested-partition subsets of* $\{1, 2, \ldots, I\}$, *the random probability vector* $\mathbf{u}$ *has the Dirichlet distributions as defined in (1.1), if and only if, independently,* $\mathbf{u}_{(*)} = (u_{(1)}, \ldots, u_{(K)}) \sim D(\mathbf{b}_{(*)})$, $\mathbf{u}_{[k_1:*]} \sim D(\mathbf{b}_{(k_1:*)})$, *for all* k_1, $\mathbf{u}_{[k_1:k_2:*]} \sim D(\mathbf{b}_{(k_1:k_2:*)})$, *for all* k_1 *and* $k_2, \ldots, \mathbf{u}_{[k_1:k_2:\ldots:k_{n-1}:*]} \sim D(\mathbf{b}_{(k_1:k_2:\ldots:k_{n-1}:*)})$, *for all* $k_1, k_2, \ldots, k_{n-1}$.

Proof: The absolute value of the Jacobian can be shown to be

$$\left(\prod_{k_1} u_{(k_1)}^{K_{k_1}-1}\right) \cdot \left(\prod_{k_1,k_2} u_{(k_1:k_2)}^{K_{k_1:k_2}-1}\right) \cdot \ldots \cdot \left(\prod_{k_1,k_2,\ldots,k_{n-1}} u_{(k_1:k_2:\ldots:k_{n-1})}^{K_{k_1:k_2:\ldots:k_{n-1}}-1}\right),$$

where K_{k_1} is the number of 2^{nd} level nested-partition subsets of S_{k_1}, $K_{k_1:k_2}$ is the number of 3^{rd} level nested-partition subsets of $S_{k_1:k_2}$, and $K_{k_1:k_2,\ldots,k_{n-1}}$ is the number of n^{th} level nested-partition subsets of $S_{k_1:k_2:\ldots:k_{n-1}}$.

□

We now show that $R(\mathbf{b}, G, -\mathbf{y})$ has a closed form expression if G is an n-level nested-partition indicator matrix. As defined earlier, we use $y_{k_1:k_2:\ldots:k_n}$ to denote the entry of $\mathbf{y}$ corresponding to $S_{k_1:k_2:\ldots:k_n}$ in this subsection. $y_{k_1:k_2:\ldots:k_{j-1}:+}$ is defined to be the sum of $y_{k_1:k_2:\ldots:k_{j-1}:k_j}$ over all possible k_j, for $j = 1, 2, \ldots, n$. That is

$$y_{k_1:k_2:\ldots:k_{j-1}:+} = \sum_{\text{all } k_j} y_{k_1:k_2:\ldots:k_{j-1}:k_j}, \text{ for any } j = 1, 2, \ldots, n.$$

Note that $y_{k_1:k_2:\ldots:k_{j-1}}$ may not be the same as $y_{k_1:k_2:\ldots:k_{j-1}:+}$. This is different from the situation $u_{(\sigma)}$'s. Similarly, we also denote $y_{k_1:k_2:\ldots:k_{j-2}:+:+}$ as the sum of $y_{k_1:k_2\ldots:k_{j-2}:k_{j-1}:+}$ over all possible k_{j-1}, for $j = 2, \ldots, n$. Therefore,

$$y_{k_1:k_2:\ldots:k_{j-2}:+:+} = \sum_{\text{all } k_{j-1}, k_j} y_{k_1:k_2:\ldots:k_{j-2}:k_{j-1}:k_j} \quad , \text{ for any } j = 2, \ldots, n.$$

If the colon ":" before the + sign in $y_{k_1:\ldots:k_j:+}$ is replaced by semicolon ";", *i.e.* $y_{k_1:\ldots:k_j;+}$, then it is the sum of $y_{k_1:\ldots:k_j}, y_{k_1:\ldots:k_j:+}, y_{k_1:\ldots:k_j:+:+}, \ldots$ and $y_{k_1:\ldots:k_j:+:\ldots:+}$. That is, $y_{k_1:\ldots:k_{n-1};+} = y_{k_1:\ldots:k_{n-1}} + y_{k_1:\ldots:k_{n-1}:+}$, and $y_{k_1:\ldots:k_j;+} = y_{k_1:\ldots:k_j} + \sum_{\text{all } k_{j+1}} y_{k_1:\ldots:k_{j+1};+}$, for all $j = 1, 2, \ldots, n-2$.

Theorem 4.2 *If G is an n-level nested-partition indicator matrix, then*

$$R(\mathbf{b}, G, -\mathbf{y}) = \left(\frac{B(\mathbf{b}_{(*)} + \mathbf{y}_{*;+})}{B(\mathbf{b}_{(*)})}\right) \cdot \left(\prod_{k_1} \frac{B[\mathbf{b}_{(k_1:*)} + \mathbf{y}_{k_1:*;+}]}{B(\mathbf{b}_{(k_1:*)})}\right) \ldots .$$

$$\cdot \left(\prod_{k_1 k_2 \ldots k_{n-2}} \frac{B[\mathbf{b}_{(k_1:k_2:\ldots:k_{n-2}:*)} + \mathbf{y}_{k_1:k_2\ldots:k_{n-2}:*;+}]}{B[\mathbf{b}_{(k_1:k_2:\ldots:k_{n-2}:*)}]}\right) \cdot \left(\prod_{k_1 k_2 \ldots k_{n-1}} \frac{B[\mathbf{b}_{(k_1:k_2:\ldots:k_{n-1}:*)} + \mathbf{y}_{k_1:k_2:\ldots:k_{n-1}:*}]}{B[\mathbf{b}_{(k_1:k_2:\ldots:k_{n-1}:*)}]}\right),$$

where $\mathbf{y}_{k_1:k_2:\ldots:k_j:*;+} = (y_{k_1:k_2:\ldots:k_j:1;+}, y_{k_1:k_2:\ldots:k_j:2;+}, \ldots, y_{k_1:k_2:\ldots:k_j:K_{k_1,k_2,\ldots,k_j};+})$ *and* $K_{k_1,k_2,\ldots,k_j}$ *is the number of the* $(j+1)^{st}$ *level nested-partition subsets of* $S_{k_1:k_2:\ldots:k_j}$.

Proof.

$$\begin{aligned}
R(\mathbf{b}, G, -\mathbf{y}) &= E_{u|b}\left(\prod_{k_1} u_{(k_1)}^{y_{k_1}}\right)\left(\prod_{k_1,k_2} u_{(k_1:k_2)}^{y_{k_1:k_2}}\right) \cdot \ldots \cdot \prod_{k_1 \ldots k_n} u_{(k_1:k_2:\ldots:k_n)}^{y_{k_1:k_2:\ldots:k_n}} \\
&= E\left(\prod_{k_1} u_{(k_1)}^{y_{k_1}+y_{k_1:+}+\ldots+y_{k_1:+:\ldots:+}}\right) \cdot \left(\prod_{k_1 k_2} u_{[k_1:k_2]}^{y_{k_1:k_2}+y_{k_1:k_2:+}+\ldots+y_{k_1:k_2:+:\ldots:+}}\right) \ldots \\
&\quad \cdot \left(\prod_{k_1 \ldots k_n} u_{[k_1:k_2:\ldots:k_n]}^{y_{k_1:k_2:\ldots:k_n}}\right) \\
&= \left(\frac{B[\mathbf{b}(*)+\mathbf{y}_{*;+}]}{B[\mathbf{b}(*)]}\right) \cdot \left(\prod_{k_1} \frac{B[\mathbf{b}_{(k_1:*)}+\mathbf{y}_{k_1:*;+}]}{B[\mathbf{b}_{(k_1:*)}]}\right) \ldots \left(\prod_{k_1,\ldots,k_{n-1}} \frac{B[\mathbf{b}_{(k_1:\ldots:k_{n-1}:*)}+\mathbf{y}_{k_1:\ldots:k_{n-1}:*}]}{B[\mathbf{b}_{(k_1:\ldots:k_{n-1}:*)}]}\right).
\end{aligned}$$

The last equality follows by Theorem 4.1.

□

The following corollary, a special case with $n = 2$, was given by Dickey, Jiang and Kadane (1987).

Corollary 4.3 *If $n = 2$, and $y_{k_1:k_2} = 0$ for all k_1, k_2, then*

$$R(\mathbf{b}, G, -\mathbf{y}) = \frac{B[\mathbf{b}_{(*)} + \mathbf{y}_{*;+}]}{B[\mathbf{b}_{(*)}]}.$$

We give a final special case in the following.

Corollary 4.4 *If $n = 3$, and $y_{k_1:k_2:k_3} = 0$ for all k_1, k_2, k_3, then*

$$R(\mathbf{b}, G, -\mathbf{y}) = \frac{B[\mathbf{b}_{(*)} + \mathbf{y}_{*;+}]}{B[\mathbf{b}_{(*)}]} \left(\prod_{k_1} \frac{B[\mathbf{b}_{(k_1:*)} + \mathbf{y}_{k_1:*;+}]}{B[\mathbf{b}_{(k_1:*)}]} \right)$$

4.2 Expansion Methods

The method in the previous subsection fails if the parameter G in R is not an n-level nested-partition indicator matrix. By definition (1.2), if $\mathbf{c}$ is a vector of non-negative integers, $R(\mathbf{b}, G, -\mathbf{c})$ is the expectation of the product of linear combinations of u_i's, and we can expand the product of some of these linear combinations to become a linear combination of products of $u_i^{a_i}$'s, where the a_i's are non-negative integers. So, we can reexpress R as a linear combination of expectations of the unexpanded product times a product of $u_i^{a_i}$'s, where the expectation is over the Dirichlet distribution. But the expectation of the unexpanded product times the product of $u_i^{a_i}$'s is another R. Therefore, $R(\mathbf{b}, G, \mathbf{c})$ can be expressed as a linear combination of other R's. We shall use a matrix variable W to indicate the possible expansions of the product.

Consider an arbitrary matrix $G(I \times J$ for arbitrary I and $J)$. Conformably partition $G = (G^{(1)}, G^{(2)})$, $\mathbf{c} = (\mathbf{c}^{(1)}, \mathbf{c}^{(2)})$ and $J = J^{(1)} + J^{(2)}$, where each entry of $\mathbf{c}^{(2)}$ is a non-negative integer. Expanding the expression $(\sum_{i=1}^{I} u_i g_{ij}^{(2)})^{c_j^{(2)}}$, we obtain a sum which we represent symbolically as $\sum_{k=1}^{N} a_k u_1^{m_{1jk}} u_2^{m_{2jk}} \ldots u_I^{m_{Ijk}}$ where $m_{1jk} + m_{2jk} + \ldots + m_{Ijk} = c_j^{(2)}$ for all k. Let W be an $I \times J^{(2)}$ matrix variable whose jth column can be any of the vectors $(m_{1jk}, m_{2jk}, \ldots, m_{Ijk})$, $k = 1, \ldots, N$. Therefore, matrix variable W can be any matrix having the following properties:

$$\begin{aligned} W \;=\; & \{w_{ij} : i = 1, 2, \ldots, I, j = 1, 2, \ldots, J^{(2)}, \\ & w_{ij} = 0, \text{ if } g_{ij}^{(2)} = 0, w_{ij} \text{ a non-negative integer and } w_{+j} = c_j^{(2)}\} \end{aligned} \tag{4.1}$$

For example, if $G^{(2)} = \begin{bmatrix} 0.8 & 0 \\ 0.5 & 0.4 \\ 0 & 1.4 \end{bmatrix}$ and $\mathbf{c}^{(2)} = (6,4)$, then the matrix variable W is

$$W = \left\{ \begin{bmatrix} w_{11} & 0 \\ 6 - w_{11} & w_{22} \\ 0 & 4 - w_{22} \end{bmatrix} : w_{11} = 0,1,\ldots,6 \text{ and } w_{22} = 0,1,\ldots,4 \right\}.$$

The following representation for R can be used to calculate R for general G.

Theorem 4.5 *Consider an arbitrary matrix $G(I \times J)$. Conformably partition $G = (G^{(1)}, G^{(2)})$, $\mathbf{c} = (\mathbf{c}^{(1)}, \mathbf{c}^{(2)})$ and $J = J^{(1)} + J^{(2)}$. Refining only $\mathbf{c}^{(2)}$ and matrix W defined as in (4.1), we have*

$$R(\mathbf{b}, G, -\mathbf{c}) = \sum_{W|\mathbf{c}^{(2)}} \left[\prod_{j=J^{(1)}+1}^{J} \binom{c_j^{(2)}}{\mathbf{w}_{*j}} \cdot \prod_{i=1}^{I} g_{ij}^{(2)^{w_{ij}}} \right] \frac{B(\mathbf{b} + \mathbf{w}_{*+})}{B(\mathbf{b})} R(\mathbf{b} + \mathbf{w}_{*+}, G^{(1)}, -\mathbf{c}^{(1)}), \quad (4.2)$$

where the summation is over any W having the vector of its column sums the same as $\mathbf{c}^{(2)}$, $\mathbf{w}_{*j}$ is the j^{th} column vector of W, and $\mathbf{w}_{*+}$ is the vector of row sums of W.

Proof

$$\begin{aligned} & R(\mathbf{b}, G, -\mathbf{c}) \\ & \quad = E_{u|b} \left[\textstyle\prod_{j=1}^{J^{(1)}} (\mathbf{u} \cdot \mathbf{g}_{*j}^{(1)})^{c_j^{(1)}} \right] \left[\textstyle\prod_{j=J^{(1)}+1}^{J} (\mathbf{u} \cdot \mathbf{g}_{*j}^{(2)})^{c_j^{(2)}} \right] \\ & \quad = \textstyle\sum_{W|\mathbf{c}^{(2)}} \prod_{j=J^{(1)}+1}^{J} \binom{c_j^{(2)}}{\mathbf{w}_{*j}} E_{u|b} \left\{ \left[\textstyle\prod_{j=1}^{J^{(1)}} (\mathbf{u} \cdot \mathbf{g}_{*j}^{(1)})^{c_j^{(1)}} \right] \left[\textstyle\prod_{i=1}^{I} (u_i g_{ij}^{(2)})^{w_{ij}} \right] \right\} \\ & \quad = \textstyle\sum_{W|\mathbf{c}^{(2)}} \left[\prod_{j=J^{(1)}+1}^{J} \binom{c_j^{(2)}}{\mathbf{w}_{*j}} \prod_{i=1}^{I} g_{ij}^{(2)^{w_{ij}}} \right] E_{u|b} \left\{ \left[\textstyle\prod_{j=1}^{J^{(1)}} (\mathbf{u} \cdot \mathbf{g}_{*j}^{(1)})^{c_j^{(1)}} \right] \cdot \left[\textstyle\prod_{i=1}^{I} u_i^{w_{i+}} \right] \right\} \end{aligned}$$

□

Note that, if $J^{(1)} = 0$, the R in the right hand side of (4.2) is one. This is exploited in the following corollary.

Corollary 4.6 *Given an $I \times J$ matrix G and W defined as in (4.1) for refining $\mathbf{c}$, we have*

$$R(\mathbf{b}, G, -\mathbf{c}) = \sum_{W|\mathbf{c}} \left[\prod_{j=1}^{J} \binom{c_j}{w_{*j}} \cdot \prod_{i=1}^{I} g_{ij}^{w_{ij}} \right] \cdot \frac{B(\mathbf{b} + \mathbf{W}_{*+})}{B(\mathbf{b})}.$$

□

Consider the case when G is an indicator matrix. For example, let

$$G = (G^{(1)}, G^{(2)}) = \left(\begin{array}{cccccc|c} 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}\right). \tag{4.3}$$

If $c_3 = c_4 = c_5 = c_6 = 0$, then

$$\begin{aligned} & R(\mathbf{b}, G, -(c_1, \ldots, c_7)) && (4.4a) \\ = \; & \tfrac{1}{B(\mathbf{b})} \textstyle\sum_{w_1=0}^{c_7} \sum_{w_2=0}^{c_7-w_1} \left(\begin{array}{c} c_7 \\ w_1, w_2, w_3 \end{array}\right) B(\mathbf{b}+\mathbf{w}) \cdot R(\mathbf{b}+\mathbf{w}, G^{(1)}, -(c_1, \ldots, c_6)) && (4.4b) \\ = \; & \tfrac{1}{B(\mathbf{b})} \textstyle\sum_{w_1=0}^{c_7} \sum_{w_2=0}^{c_7-w_1} \left(\begin{array}{c} c_7 \\ w_1, w_2, w_3 \end{array}\right) B(\mathbf{b}+\mathbf{w}) \frac{B(b_1+b_2+c_1, b_3+b_4+c_2)}{B(b_1+b_2, b_3+b_4)}. && (4.4c) \end{aligned}$$

where $w_3 = c_7 - w_1 - w_2$, and $\mathbf{w} = (w_1, w_2, w_3, 0)$. The first identity (4.4b) is implied by Theorem 4.5. The second identity (4.4c) is from Corrollary 4.3. Alternatively, for the same R in (4.4a),

$$\begin{aligned} & R(\mathbf{b}, G, -(c_1, \ldots, c_7)) \\ = \; & \tfrac{1}{B(\mathbf{b})} \int \left(\textstyle\prod_{i=1}^4 u_i^{b_i-1}\right) (u_1+u_2)^{c_1} \cdot (u_3+u_4)^{c_2} \left[\textstyle\sum_{w_1=0}^{c_7} \left(\begin{array}{c} c_7 \\ w_1, w_2 \end{array}\right) (u_1+u_2)^{w_1} \cdot u_3^{w_2}\right] d\mathbf{u} && (4.5a) \\ = \; & \tfrac{1}{B(\mathbf{b})} \textstyle\sum_{w_1=0}^{c_7} \left(\begin{array}{c} c_7 \\ w_1, w_2 \end{array}\right) \cdot B(\mathbf{b}+\mathbf{w}^{(I)}) \cdot R(\mathbf{b}+\mathbf{w}^{(I)}, G^{(1)}, -(c_1+w_1, c_2, \ldots, c_6)) && (4.5b) \\ = \; & \tfrac{1}{B(\mathbf{b})} \textstyle\sum_{w_1=0}^{c_7} \left(\begin{array}{c} c_7 \\ w_1, w_2 \end{array}\right) \cdot B(\mathbf{b}+\mathbf{w}^{(I)}) \cdot \frac{B(b_1+b_2+c_1+w_1, b_3+b_4+c_2+w_2)}{B(b_1+b_2, b_3+b_4+w_2)} && (4.5c) \end{aligned}$$

where $w_2 = c_7 - w_1, \mathbf{w}^{(I)} = (0, 0, w_2, 0)$ and $G^{(1)}$ is defined in (4.3).

As the number of summation terms, $c_7 + 1$, of (4.5c) is smaller than $\frac{(c_7+1)(c_7+2)}{2}$, the number for (4.4c), it is usually better for the computation of R to use (4.5c). This motivates the following theorem, useful for calculating R when G is an indicator matrix. But first, we give some definitions.

Consider an $I \times J$ indicator matrix G and a vector of non-negative integers $\mathbf{y}$. Conformably partition $G = (G^{(1)}, G^{(2)}), \mathbf{y} = (\mathbf{y}^{(1)}, \mathbf{y}^{(2)})$ and $J = J^{(1)} + J^{(2)}$ so that $G^{(1)}$ is an n-level nested-partition indicator matrix. If the j-th column of $G^{(2)}$ is an indicator vector of S_j, then there are mutually exclusive and exhaustive subsets of S_j, whose indicator vectors are some columns of $G^{(1)}$, for $j = 1, 2, \ldots, J^{(2)}$. We call these subsets *partition subsets* of S_j. Given one way of partitioning S_j, there is one set of partition subsets of S_j. Since there are usually many ways of partitioning S_j, there are many sets of partition subsets of S_j. In practice, it is useful to have a way of partitioning so that the number of partition subsets is a minumum. Hence, the number of summation terms is small. For the previous example, $G^{(1)}$ in (4.3) is a 2-level nested-partition indicator matrix, $J^{(1)} = 6, J^{(2)} = 1$ and $S_1 = \{1, 2, 3\}$. One set of partition subsets of S_1, is $\{\{1\}, \{2\}, \{3\}\}$. Another set of partition subsets of S_1 is $\{\{1, 2\}, \{3\}\}$. The second way of partitioning, which yields two subsets, is better than the first way, which yields three subsets. This also agrees with our previous experience.

Once we have a way of partitioning for each S_j, we can define the matrix variable W as any $J^{(1)} \times J^{(2)}$ matrix having the following properties:

1. Each row corresponds to an n-level nested-partition subset. (4.6)
2. The $j-th$ column $\mathbf{w}_{*j}$, a vector of non-negative integers, corresponds to S_j and $y_j^{(2)}$, so that

   (a) $w_{ij} = 0$, if i-th row does not correspond to a partition subset of S_j.
   (b) $w_{+j} = y_j^{(2)}$.

Let S_1 and S_2 be subsets of $\{1, 2, \ldots, I\}$. We say an observation, which is reported as falling in the set of categories S_1, is *less-censored* than another observation, which is reported as falling in the set of categories S_2, if $S_1 \subset S_2$. One interpretation of the matrix variable W in censored categorical data problem is that each value of matrix W indicates a possible set of frequency counts refining the observed frequency vector $\mathbf{y}$ into less-censored data.
We then have the following useful theorem for censored data problems.

Theorem 4.7 *Consider an indicator matrix $G, I \times J$ for arbitrary I, J. Conformably partition $G = (G^{(1)}, G^{(2)}), \mathbf{y} = (\mathbf{y}^{(1)}, \mathbf{y}^{(2)})$ and $J = J^{(1)} + J^{(2)}$ so that $G^{(1)}$ is an n-level nested-partition indicator matrix. Then W, defined in (4.6), is a $J^{(1)} \times J^{(2)}$ matrix and*

$$R(\mathbf{b}, G, -\mathbf{y}) =$$

$$\sum_{W|\mathbf{y}^{(2)}} \left[\prod_{j=1}^{J^{(2)}} \begin{pmatrix} y_j^{(2)} \\ \mathbf{w}_{*j} \end{pmatrix} \right] \cdot R(\mathbf{b}, G^{(1)}, -(\mathbf{y}^{(1)} + \mathbf{w}_{*+})). \tag{4.7}$$

The summation is over any W having the vector of its column sums the same as $\mathbf{y}^{(2)}$, $\mathbf{w}_{*j}$ is the j^{th} column vector of matrix W, and $\mathbf{w}_{*+}$ is the vector of row sums of matrix W.

Now, since $G^{(1)}$ is an n-level nested-partition indicator matrix, we have a closed form for R in the right hand side of (4.7). Therefore, for the computation of R with indicator matrix G, it is simpler to use Theorem 4.7 than to use Theorem 4.5.

In summary, given an R with an indicator matrix parameter, we can express it as an R with parameter matrix $G = (G^{(1)}, G^{(2)})$ so that $G^{(1)}$ is an n-level nested-partition indicator. Let the j-th column of $G^{(2)}$ be an indicator vector of S_j. We then determine a way of partititioning each of the S_j's. Thus, we have W in the form (4.6), and we can apply Theorem 4.7 to compute R.

4.3 Approximation methods

For extensive non-nested censored data, the expansions that were discussed in subsection 4.2 may contain too many summation terms, so they may not be practical for computing R. In this subsection, we shall use the approximation formula of Laplace to approximate R and that of Tierney and Kadane (1986) to approximate ratios of R. The Laplace approximation formula is stated as follows:

If h is a function from R^{I-1} to R and has a unique maximum at $\hat{\mathbf{u}}_1$, then

$$\int e^{h(\mathbf{u}_1)} d\mathbf{u}_1 \doteq (2\pi)^{\frac{I-1}{2}} \cdot (\det \Sigma)^{\frac{1}{2}} \cdot e^{h(\hat{\mathbf{u}}_1)}, \tag{4.8}$$

where $\sum$ is the negative of the inverse Hessian of h at the mode $\hat{\mathbf{u}}_1$, and $\mathbf{u}_1$ is a vector which is composed of the first $I-1$ entries of $\mathbf{u}$. The order of the error is $0(\frac{1}{m})$, where m is the sample size.

To approximate $R(\mathbf{b}, G^{(2)}, -\mathbf{c})$, we shall need to reexpress it in the form of (4.8). Let $G = (G^{(1)}, G^{(2)})$ be an $I \times J$ matrix, where $G^{(1)}$ is an $I \times I$ identity matrix, $G^{(2)}$ is an $I \times J^{(2)}$ matrix and $J = I + J^{(2)}$. Then

$$R(\mathbf{b}, G^{(2)}, -\mathbf{c}) = \frac{1}{B(\mathbf{b})} \int \prod_{j=1}^{J} \left(\sum_{i=1}^{I} g_{ij} u_i \right)^{n_j} d\mathbf{u}, \tag{4.9}$$

where $\mathbf{n} = ((\mathbf{b} - \mathbf{1}), \mathbf{c})$ and $\mathbf{1} = (1, 1, \ldots, 1)$. Define the integrand of the right-hand side of (4.9) as $g(\mathbf{u})$, *i.e.*,

$$g(\mathbf{u}) = \prod_{j=1}^{J} (\sum_{i=1}^{I} g_{ij} u_i)^{n_j}.$$

Now R may be expressed in the form of (4.8), *i.e.*

$$R(\mathbf{b}, G^{(2)}, -\mathbf{c}) = \frac{1}{B(\mathbf{b})} \int e^{h(\mathbf{u}_1)} d\mathbf{u}_1, \tag{4.10}$$

where $h(\mathbf{u}_1) = \log g(u_1, \ldots, u_{I-1}, 1 - \sum_{i=1}^{I-1} u_i)$. The theorem that follows shows that, under mild conditions, $g(\mathbf{u})$ has a unique mode.

Once we have found the mode, the next quantity that we need to determine is the determinant $\det \sum$, where $\sum$ is minus the inverse of the Hessian matrix of h at the mode $\hat{\mathbf{u}}_1$. It is sufficient to find $\det(\sum^{-1}) = 1/\det \sum$. Let

$$\beta_{ij} = g_{ij} - g_{Ij}, \text{ for } 1 \le i \le I - 1, \ \ 1 \le j \le J, \tag{4.11}$$

then

$$h(\mathbf{u}_1) = \sum_{j=1}^{J} n_j \log \left(\sum_{i=1}^{I-1} \beta_{ij} u_i + g_{Ij} \right). \tag{4.12}$$

We have

$$\frac{\partial^2 h}{\partial u_i \partial u_k} = \sum_{j=1}^{J} \frac{-n_j \beta_{kj} \beta_{ij}}{\left[\sum_{i=1}^{I-1} \beta_{ij} u_i + g_{Ij} \right]^2}. \tag{4.13}$$

Then

$$\sum^{-1} = \sum_{j=1}^{J} \gamma_j \beta_{*j} \beta_{*j}^T, \text{ where} \tag{4.14}$$

$$\gamma_j = n_j / \left[\sum_{i=1}^{I-1} \beta_{ij} \hat{u}_i + g_{Ij} \right]^2 . \tag{4.15}$$

Therefore, by (4.8), (4.10) and (4.14), we have

$$R(\mathbf{b}, G^{(2)}, -\mathbf{c}) \doteq \frac{1}{B(\mathbf{b})} (2\pi)^{\frac{I-1}{2}} \left[\det \left(\sum_{j=1}^{J} \gamma_j \boldsymbol{\beta}_{*j} \boldsymbol{\beta}_{*j}^T \right) \right]^{-\frac{1}{2}} e^{h(\hat{\mathbf{u}}_1)}, \tag{4.16}$$

where h, β_{ij}'s, γ_j's are defined in (4.12), (4.11) and (4.15) respectively.

Theorem 4.8 *Assume*

(i) $n_j > 0 \qquad \forall j = 1, \ldots, J;$ *(4.17a)*

(ii) the vectors $\mathbf{g}_{*j}, j = 1, \ldots, J$, *span the* I*-dimensional real vector space.* *(4.17b)*

Then, $g(\mathbf{u}) = \prod_{j=1}^{J} (\sum_{i=1}^{I} g_{ij} u_i)^{n_j}$ *has a single local mode.*

□

Proof By (4.13), the Hessian matrix of h is

$$H = \sum_{j=1}^{J} m_j \boldsymbol{\beta}_{*j} \boldsymbol{\beta}_{*j}^T, \text{ where}$$

$m_j = -n_j / \left[\sum_{i=1}^{I-1} \beta_{ij} u_i + g_{Ij} \right]^2$. By (4.17a), $m_j < 0$, for every j. For every non-zero $(I-1)$-dimensional column vector $\mathbf{x}$, $(\mathbf{x}^T \boldsymbol{\beta}_{*j})(\boldsymbol{\beta}_{*j}^T \mathbf{x}) = (\mathbf{x}^T \boldsymbol{\beta}_{*j})^2 \geq 0$. But, by (4.17b), there is at least one j such that $\mathbf{x}^T \boldsymbol{\beta}_{*j} \neq 0$. Hence, for every non-zero vector $\mathbf{x}$, $\mathbf{x}^T H \mathbf{x} < 0$. By definition, H is a negative definite matrix, and so $g(\mathbf{u})$ has a single local mode. □

A posterior moment (e.g. (2.7) or (2.11)) is proportional to a ratio of two R functions. Therefore, we can approximate a posterior moment by applying formula (4.16) separately to each of the numerator and denominator.

$$R(\mathbf{b}+\mathbf{d}, G^{(2)}, -(\mathbf{c}+\mathbf{e})) / R(\mathbf{b}, G^{(2)}, -\mathbf{c})$$

$$\doteq \frac{B(\mathbf{b})}{B(\mathbf{b}+\mathbf{d})} \cdot \left[\frac{\det(\sum_{j=1}^{J} \gamma_j \boldsymbol{\beta}_{*j} \boldsymbol{\beta}_{*j}^T)}{\det(\sum_{j=1}^{J} \gamma_j^* \boldsymbol{\beta}_{*j} \boldsymbol{\beta}_{*j}^T)} \right]^{\frac{1}{2}} \cdot \left[\frac{e^{h^*(\hat{\mathbf{u}}_1^*)}}{e^{h(\hat{\mathbf{u}}_1)}} \right], \tag{4.18}$$

where $\hat{\mathbf{u}}_1, h$ and γ are defined the same as those in (4.16), but $\hat{\mathbf{u}}_1^*, h^*$ and γ^* are defined with $\mathbf{b}$ replaced by $\mathbf{b}+\mathbf{d}$ and $\mathbf{c}$ replaced by $\mathbf{c}+\mathbf{e}$.

Our experience is consistent with Tierney and Kadane's (1986) theorem that the approximation (4.18) for a posterior moment is order $0(\frac{1}{m^2})$, and thus is more accurate than the approximation (4.16), which is order $0(\frac{1}{m})$.

For the parameters we discussed in this subsection, if some of the b_j's are smaller than unity, then their corresponding n_j's, which are $b_j - 1$, would be negative. Theorem 4.8 would not be applicable in this case. The integrand of R, then, would usually not have a unique mode. The approximation formula (4.8) could not be used directly. Therefore, before applying formula (4.8), we need to apply the following transformation.

Theorem 4.9 *Suppose* $\mathbf{u} \sim D(\mathbf{b}, G, \mathbf{c})$. *Let* $v_i = \log(u_i/u_I)$, *for* $i = 1, 2, \ldots, I-1$, *then the p.d.f. of* $\mathbf{v}$ *is*

$g(\mathbf{v}) = h(\mathbf{v})/R(\mathbf{b}, G, -\mathbf{c})$, where,

$$h(\mathbf{v}) = \frac{1}{B(\mathbf{b})} \cdot \exp\Big(\sum_{i=1}^{I-1} b_i v_i\Big) \cdot \left[\prod_{j=1}^{L} \left(\sum_{i=1}^{I-1} g_{ij} e^{v_i} + g_{Ij}\right)^{c_j}\right] \Big/ \left(1 + \sum_{i=1}^{I-1} e^{v_i}\right)^{b_+ + c_+}, \tag{4.19}$$

and the range of each v_i *is* $(-\infty, \infty)$. *Furthermore, if* $b_i > 0$ *and* $c_j > 0$, *for each* i *and* j, *then* $h(\mathbf{v})$ *has a unique mode.*

Proof. It can be shown that

$$u_i = e^{v_i}/(1 + \sum_{i=1}^{I-1} e^{v_i}), i = 1, 2, \ldots, I-1,$$

$$u_I = 1/(1 + \sum_{i=1}^{I-1} e^{v_i}),$$

and the absolute value of the Jacobian is

$$\frac{\exp\left(\sum_{i=1}^{I-1} v_i\right)}{\left(1 + \sum_{i=1}^{I-1} e^{v_i}\right)^I} = \prod_{i=1}^{I} u_i.$$

Hence, the *p.d.f.* of $\mathbf{v}$ can be expressed as $g(\mathbf{v})$. If we express h of (4.19) in terms of $\mathbf{u}$ and let $\mathbf{v} = \mathbf{k}(\mathbf{u})$, then

$$h(\mathbf{v}) = f(\mathbf{u}) = \frac{1}{B(\mathbf{b})}\left(\prod_{i=1}^{I} u_i^{b_i}\right)\prod_{j=1}^{L}\left(\sum_{i=1}^{I} g_{ij}u_i\right)^{c_j}, \tag{4.20}$$

where $f(\mathbf{u}) = h(\mathbf{k}(\mathbf{u}))$. By Theorem 4.8, $f(\mathbf{u})$ has a unique mode and so does $h(\mathbf{v})$. □

Note that the Jacobian makes the exponent of u_i for the p.d.f. of $D(\mathbf{b}, G, \mathbf{c})$ change from "$b_i - 1$" to "b_i" in (4.20). Now, to compute R, we integrate $h(\mathbf{v})$ in (4.19). By Theorem 4.9, we can now apply the approximation formula (4.8).

As noted earlier, although the Laplace approximation to the ratio of R's is accurate to order $0(\frac{1}{m^2})$, the approximation of R, itself, is only accurate to order $0(\frac{1}{m})$. An alternative approach is to use monte carlo methods, explained below.

4.4 Monte Carlo Methods

If we reexpress R in (1.2) as

$$R(\mathbf{b}, G, -\mathbf{c}) = E_{\mathbf{u}|\mathbf{b}}h(\mathbf{u}), \tag{4.21}$$

where

$$h(\mathbf{u}) = \prod_{j=1}^{J}\left(\sum_{i=1}^{I} g_{ij}u_i\right)^{c_j}. \tag{4.22}$$

We may then use the following 3 steps to generate a Monte Carlo value for $h(\mathbf{u})$:

Step 1: generate I gamma random deviates according to gamma $(b_i, 1)$, say $x_i, i = 1, \ldots, I$.

Step 2: Let $u_i = x_i / \sum_{k=1}^{I} x_k$, where $i = 1, 2, \ldots, I$. Now, $\mathbf{u}$ follows a Dirichlet distribution with paramater $\mathbf{b}$.

Step 3: Compute $h(\mathbf{u})$ according to (4.22).

We denote the above $h(\mathbf{u})$ as h_1. Repeating the above three steps $n-1$ times independently, we have n Monte Carlo values $h_1, h_2, h_3, \ldots, h_n$, for $h(\mathbf{u})$. Let $\bar{h} = \sum_{i=1}^{n} h_i/n$. Then, $\bar{h}$ is an unbiased estimate of $R(\mathbf{b}, G, -\mathbf{c})$. An estimated standard error of $\bar{h}$ is $\frac{1}{\sqrt{n}}s$, where $s^2 = \frac{\sum_{i=1}^{n}(h_i - \bar{h})^2}{n-1}$ and this is an unbiased estimate of the population variance S^2 of h's. Hence, we may increase the accuracy by increasing the Monte Carlo sample size. One way to determine the

Monte Carlo sample size n is by taking a preliminary sample of size n_1 first. After computing the estimated variance s_1^2, the sample size n can be determined as $\left(\frac{S_1}{d}\right)^2$, where d is the desired standard error of $\bar{h}$. We can then take $(n - n_1)$ further Monte Carlo values.

Consider computation of the ratio of R's,

$$R(\mathbf{b}, G, -(\mathbf{c}+\mathbf{d}))/R(\mathbf{b}, G, -\mathbf{c}) = E_{u|b} f(\mathbf{u})/E_{u|b} h(\mathbf{u}), \tag{4.23}$$

where

$$f(\mathbf{u}) = \prod_{j=1}^{J} (\sum_{i=1}^{I} g_{ij} u_i)^{(c_j+d_j)} \tag{4.24}$$

and $h(\mathbf{u})$ is defined as in (4.22). In the third step of generating a Monte Carlo sample, we may also compute $f(\mathbf{u})$ according to (4.24). The $f(\mathbf{u})$ computed from the i^{th} cycle is denoted by f_i. Then, the ratio of R in (4.23) can be estimated by $r = \bar{f}/\bar{h}$, which is a slightly biased estimate of R_1/R_2, where $R_1 = R(\mathbf{b}, G, -(\mathbf{c}+\mathbf{d}))$ and $R_2 = R(\mathbf{b}, G, -\mathbf{c})$. The bias is of order $0(\frac{1}{n})$. The following results can be used to determine the mean squared error of the ratio.

Theorem 4.10 *The mean squared error of r (to order $0(\frac{1}{n})$) is*

$$MSE(r) = \frac{1}{nR_2^2}(Var(f) + \frac{R_1^2}{R_2^2} Var(h) - 2\frac{R_1}{R_2} Cov(f,h)),$$

where $Var(f) = E_{u|b}(f(\mathbf{u}) - R_1)^2$ and $Cov(f,h) = E_{u|b}(f(\mathbf{u}) - R_1)(h(\mathbf{u}) - R_2)$.

Proof:

Let $d_1 = \frac{\bar{f}-R_1}{R_1}$ and $d_2 = \frac{\bar{h}-R_2}{R_2}$, then

$$\begin{aligned}
\bar{f} &= R_1(1+d_1),\ \bar{h} = R_2(1+d_2) \text{ and} \\
r &= \bar{f}/\bar{h} = R_1(1+d_1)/R_2(1+d_2) \\
&= (R_1/R_2)\cdot(1+d_1)(1-d_2+d_2^2-d_2^3+d_2^4-\ldots), |d_2|<1 \\
&= (R_1/R_2)(1+d_1-d_2+\ldots).
\end{aligned}$$

Therefore, the mean squared error of r is to order $0(\frac{1}{n})$,

$$
\begin{aligned}
MSE(r) &= E(r - R_1/R_2)^2 \\
&= (R_1/R_2)^2 E(d_1 - d_2)^2 \text{ (to } 0(\tfrac{1}{n})) \\
&= (R_1/R_2)^2 (Var(d_1) + Var(d_2) - 2\, Cov(d_1, d_2)) \text{ (Since } E(d_1) = E(d_2) = 0) \\
&= (R_1/R_2)^2 \cdot \tfrac{1}{n} \left(\tfrac{Var(f)}{R_1^2} + \tfrac{Var(h)}{R_2^2} - 2\tfrac{Cov(f,h)}{R_1 R_2} \right) \\
&= \tfrac{1}{nR_2^2} (Var(f) + \tfrac{R_1^2}{R_2^2}\, Var(h) - 2\tfrac{R_1}{R_2}\, Cov(f,h)).
\end{aligned}
$$

□

Corollary 4.11 *The estimated mean squared error for r is*

$$
\hat{MSE}(r) = \frac{1}{n\bar{h}^2} \left[s_f^2 + \frac{R_1^2}{R_2^2} s_h^2 - 2\frac{R_1}{R_2} s_{fh} \right], \text{ where} \tag{4.25}
$$

$$
s_f^2 = \frac{\sum_{i=1}^n (f_i - \bar{f})^2}{n-1} \text{ and } s_{fh} = \frac{\sum_{i=1}^n (f_i h_i - \bar{f}\,\bar{h})^2}{n-1}.
$$

5 Examples

We use three examples to illustrate and compare computation methods. In the first example, the closed-form method is illustrated by computing R with a three-level nested-partition indicator matrix. In the second example, we illustrate our expansion method by computing R and a ratio of R's with an indicator parameter matrix. To compare estimated values and CPU times, the Laplace method and monte carlo method are also used. In the third example, we compare estimated values and CPU times for R and a ratio of R's when neither the closed form nor expansion method is feasible.

Example 1. Three surveys are taken regarding degree of satisfaction of a service. The first, second and third survey questionaires were designed, respectively, to have two (acceptable and not acceptable), four (very good, good, bad and very bad), and eight (1 (excellent), 2, ..., 8 (terrible)) possible outcomes. Hypothetical sample data is shown in Table 1. If the i-th degree of satisfaction corresponds to the i-th category, then there are eight categories. If we also assume a uniform prior distribution, then the posterior *p.d.f.*, after these three samples are combined,

has expression (2.10) with $R(\mathbf{b}, G, -\mathbf{y})$, where $\mathbf{b} = (b_1, b_2, \ldots, b_8), b_1 = b_2 = \ldots = b_8 = 1, \mathbf{y} = (\mathbf{y}^{(1)}, \mathbf{y}^{(2)}) = (3,2,1,5,2,1 \vdots 2,1,5,8,4,3,1,0)$, and

$$G = (G^{(1)}, G^{(2)}) = \left[\begin{array}{cccccc|cccccccc} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Define $\mathbf{b} + \mathbf{y}^{(2)}$ and $(\mathbf{y}^{(1)}, \mathbf{0})$ as $\mathbf{c}$ and $\mathbf{z}$ respectively, i.e. $\mathbf{b} + \mathbf{y}^{(2)} = \mathbf{c}$ and $(\mathbf{y}^{(1)}, \mathbf{0}) = \mathbf{z}$, then

$$\begin{aligned} R(\mathbf{b}, G, -\mathbf{y}) &= \frac{B[\mathbf{b} + \mathbf{y}^{(2)}]}{B[\mathbf{b}]} \cdot R(\mathbf{b} + \mathbf{y}^{(2)}, G^{(1)}, -\mathbf{y}^{(1)}) \\ &= \frac{B[\mathbf{c}]}{B[\mathbf{b}]} R(\mathbf{c}, G, -\mathbf{z}) \\ &= \frac{B[\mathbf{c}]}{B[\mathbf{b}]} \cdot \frac{B[\mathbf{c}_{(*)} + \mathbf{z}_{*;+}]}{B[\mathbf{c}_{(*)}]} \left(\prod_{k_1=1}^{2} \frac{B[\mathbf{c}_{(k_1:*)} + \mathbf{z}_{k_1:*;+}]}{B[\mathbf{c}_{(k_1:*)}]} \right). \end{aligned} \tag{5.1}$$

The first, second and third identity above are from Lemma 3.7, Lemma 3.2 and Corollary 4.4 respectively. Using notation defined in subsection 4.1, we have $z_3 = z_{1:1}$, $z_4 = z_{1:2}, z_5 = z_{2:1}$, $z_6 = z_{2:2}, z_7 = z_{1:1:1}$, $z_8 = z_{1:1:2}, z_9 = z_{1:2:1}$, $z_{10} = z_{1:2:2}, z_{11} = z_{2:1:1}$, $z_{12} = z_{2:1:2}, z_{13} = z_{2:2:1}$ and $z_{14} = z_{2:2:2}$. Hence, $\mathbf{z}_{*;+} = (z_{1;+}, z_{2;+}) = (z_1 + (z_3 + z_4) + (z_7 + z_8 + z_9 + z_{10}), z_2 + (z_5 + z_6) + (z_{11} + z_{12} + z_{13} + z_{14})) = (9,5), \mathbf{z}_{1:*;+} = (z_{1:1;+}, z_{1:2;+}) = (z_3 + (z_7 + z_8), z_4 + (z_9 + z_{10})) = (1,5)$, and $\mathbf{z}_{2:*;+} = (z_{2:1;+}, z_{2:2;+}) = (z_5 + (z_{11} + z_{12}), z_6 + (z_{13} + z_{14})) = (2,1)$. We also have $\mathbf{c}_{(*)} = (c_{(1)}, c_{(2)}) = (\sum_{i=1}^{4} c_i, \sum_{i=5}^{8} c_i) = (20,12), \mathbf{c}_{(1:*)} = (c_{(1:1)}, c_{(1:2)}) = (c_1 + c_2, c_3 + c_4) = (5,15)$, and $\mathbf{c}_{(2:*)} = (c_{(2:1)}, c_{(2:2)}) = (c_5 + c_6, c_7 + c_8) = (9,3)$. Substitute the data vectors into equation (5.1), we have

$$\begin{aligned} R(\mathbf{b}, G, -\mathbf{y}) &= \tfrac{B(3,2,6,9,5,4,2,1)}{B(1,1,1,1,1,1,1,1)} \cdot \tfrac{B(29,17)}{B(20,12)} \left(\tfrac{B(6,20)}{B(5,15)} \cdot \tfrac{B(11,4)}{B(9,3)} \right) \\ &= 5.2172 \times 10^{-28}. \end{aligned}$$

□

Example 2. Neurological complications are one serious sequel associated with meningitis. To evaluate a standard therapy the results of a neurological test were obtained for 33 children with meningitis. Both pre and post-test results were available for 25 children, only pre-test results were available for 6 children, and only post-test results were available for 2 children. This data (Smith and Gunel (1984)) is summarized in Table 2.

In the 2×2 table there are four categories taken in the respective order, $(1,1),(1,2),(2,1)$ and $(2,2)$. Again, assuming a uniform prior, we have that the normalized constant for the posterior *p.d.f.* is $B(\mathbf{b}) \cdot R(\mathbf{b}, G, -\mathbf{y})$, where $\mathbf{b} = (7,9,4,9), \mathbf{y} = (\mathbf{y}^{(1)}, \mathbf{y}^{(2)}) = (2,4,0,0,0,0,2)$ and

$$G = (G^{(1)}, G^{(2)}) = \left[\begin{array}{cccccc|c} 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}\right].$$

If we partition $S_1^{(2)} = \{1,2\}$ into $\{1\}$ and $\{2\}$. Then, by (4.6), the possible W's are vectors $\mathbf{w}_1, \mathbf{w}_2$ and $\mathbf{w}_3$, where

$$\mathbf{w}_1 = \begin{bmatrix} 0 \\ 0 \\ - \\ 0 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 0 \\ 0 \\ - \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \mathbf{w}_3 = \begin{bmatrix} 0 \\ 0 \\ - \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

By Theorem 4.7 and $y_1^{(2)} = 2$, we have

$$R(\mathbf{b}, G, -\mathbf{y}) = \sum_{i=1}^{3} \left(\begin{pmatrix} 2 \\ \mathbf{w}_i \end{pmatrix} \right) R(\mathbf{b}, G^{(1)}, -(\mathbf{y}^{(1)} + \mathbf{w}_i))$$

$$= R\left(\begin{pmatrix} 7 \\ 9 \\ 4 \\ 9 \end{pmatrix}, G^{(1)}, - \begin{pmatrix} 2 \\ 4 \\ 0 \\ 2 \\ 0 \\ 0 \end{pmatrix} \right) + 2 \cdot R\left(\begin{pmatrix} 7 \\ 9 \\ 4 \\ 9 \end{pmatrix}, G^{(1)}, - \begin{pmatrix} 2 \\ 4 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right).$$

$$+R\left(\begin{pmatrix} 7 \\ 9 \\ 4 \\ 9 \end{pmatrix}, G^{(1)}, - \begin{pmatrix} 2 \\ 4 \\ 2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right)$$

Using the closed form method (as we did for example 1) for each term, we have $R(\mathbf{b}, G, -\mathbf{y}) = 0.0060975$. Denote this R by R_0 and let $R_1 = R(\mathbf{b}', G, -\mathbf{y})$ where $\mathbf{b}' = (8, 9, 4, 9)$. Using the expansion method again, we can compute R_1 and obtain R_1/R_0. Note that the posterior moments are proportional to ratios of R's. To compare relative errors, *i.e.* the absolute value of error over the true value given by the expansion method, and CPU times, the Laplace and monte carlo methods were also used to estimate R_0 and R_1/R_0. Results are given in Table 3. In this example, the monte carlo method is more accurate than the Laplace method for estimating a single R. However, to estimate a ratio of R's, the Laplace method is more accurate. In terms of CPU time, the Laplace method is uniformly better. Note that we used sample size $n = 4,000$ for each of the monte carlo calculations. □

Example 3. In this example, we compare the estimated values of R and ratio of R's by using the Laplace and monte carlo methods. Although we use hypothetical data, this is the type of R we are likely to encounter in Bayesian smoothing problems. Here, we consider $R_0 = R(\mathbf{b}, G, -\mathbf{c})$ and $R_1 = R(\mathbf{b}, G, -\mathbf{c}')$, where $\mathbf{b} = (b_1, \ldots, b_6), b_1 = \ldots = b_6 = 1.46, \mathbf{c} = (5, 15, 3, 5, 1, 1), \mathbf{c}' = (6, 15, 3, 5, 1, 1)$ and

$$G = \begin{bmatrix} .252636 & .346373 & .186692 & .165444 & .048276 & .000580 \\ .249421 & .303337 & .157317 & .144730 & .084169 & .061026 \\ .236832 & .291551 & .143567 & .117765 & .091937 & .118348 \\ .230493 & .295390 & .142317 & .128301 & .093455 & .110044 \\ .236654 & .327355 & .162942 & .141962 & .076223 & .054865 \\ .261564 & .331194 & .183567 & .164998 & .052740 & .005937 \end{bmatrix}.$$

The results are shown in Table 4. The estimated values of R_1/R_0 based on the Laplace and monte carlo methods are very close. But, the estimated values of each of R_0 and R_1 based on the two methods are quite different. Since the expansion method is not feasible, we do not have an exact value available. However, we do have the estimated standard errors, which are small, for R_0 and R_1 based on the monte carlo method. Again, in terms of the CPU time, the Laplace method is more efficient. We used sample size $n = 10,000$ for each of the monte carlo calculations.

Table 1. Three surveys: Degree of Satisfaction

Degree of satisfaction	1	2	3	4	5	6	7	8
Survey 1			3				2	
Survey 2		1			5			2			1	
Survey 3	2	1	5	8	4	3	1	0

Table 2. Data on neurological complications (Smith and Gunel, 1981)

	Pre-test		Supplemental data on post-test
Post-test	S	F	
S	6	8	2
F	3	8	0
Supplemental data on pre-test	2	4	Total sample size 33

Table 3. Neurological complications

Exact value (expansion): $R_0 = .0060975, R_1/R_0 = 1.0278967$

		Estimated value	Relative Error	CPU time (seconds)
Laplace	R_0	.0058201	.0455	.05
	R_1/R_0	1.0291576	.0012	
Monte carlo	R_0	.0061280	.0050	2.58
	R_1/R_0	1.0243962	.0034	

Table 4. General R: Non-indicator matrix G

		Estimated value	CPU Time (seconds)	Standard error
Laplace	R_0	$2.2601394 \times 10^{-20}$	0.100	
	R_1	$.5544338 \times 10^{-20}$	0.090	
	R_1/R_0	.2453096		
Monte Carlo	R_0	3.15694×10^{-20}	10.56	$.00090 \times 10^{-20}$
	R_1	$.774512 \times 10^{-20}$	10.26	$.00023 \times 10^{-20}$
	R_1/R_0	.2453363		

6 Conclusion

The preferred method to compute $R(\mathbf{b}, G, -\mathbf{c})$ or its ratios depends on the parameters of R. We summarize as follows:

1. If G is an n-level nested-partition indicator matrix, use the closed form (section 4.1).

2. If G is an indicator matrix, c_+ is not very large and J' (the number of nonzero c_i's) is small, or if G is not an indicator matrix but c_+ and J' are small, we use the expansion method (section 4.2).

3. If $b_+ + c_+$ is very large, when computing R, or $b_+ + c_+$ is not small, when computing a ratio of R's, we can use the approximation method (section 4.3).

4. If $b_+ + c_+$ is not very large, when computing R, or $b_+ + c_+$ is small, when computing a ratio of R's, it is best to use the Monte Carlo method (section 4.4).

From our experience, it is very likely in practice that, simultaneously, more than one of the above methods can be used very effectively, thus providing comparisons, as in the examples of section 5.

References

Carlson, B.C. (1971), "Appell Functions and Multiple Averages", *SIAM Journal on Mathematical Analysis*, **2**, 420-430.

Carlson, B.C. (1974), "Inequalities for Jacobi Polynomial and Dirichlet Averages", *SIAM Journal on Mathematical Analysis*, **5**, 586-596.

Dickey, J.M. (1968), "Three Multidimensional-Integral Identities with Bayesian Applications," *The Annals of Mathematical Statistics*, **39**, 1615-1627.

Dickey, J.M. (1983), "Multiple Hypergeometric Functions: Probabilistic Interpretations and Statistical Uses", *Journal of the American Statistical Association*, **78**, 628-637.

Dickey, J.M., Garthwaite, P.H., and Bian, G. (1989), "An Elementary Continuous-type Nonparametric Distribution Estimate", School of Statistics, University of Minnesota, Technical Report No. 525.

Dickey, J.M., and Jiang, T.J. (1991), "Filtered-Variate Prior Distributions For Local Smoothing of Histograms", *Journal of the American Statistical Association.* (To appear).

Dickey, J.M., Jiang, J.-M. and Kadane, J.B. (1987), "Bayesian Methods for Censored Categorical Data", *Journal of the American Statistical Association*, **82**, 773-781.

Smith, P.J., and Gunel, E. (1984), "Practical Bayesian approaches to the analysis fo a two by two contingency table with incompletely categorized data", *Communications in Statistics A, Theory and Methods*, **13**, 1941-1963.

Tierney, L., and Kadane, J.B. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities", *Journal of the American Statistical Association*, **81**, 82-86.