BAYESIAN ANALYSIS OF "RANDOM ROUTES"

by

M.J. Bayarri and B. Font
Universitat de Valéncia, Purdue University
Universitat de Valéncia

Technical Report #94-31

Department of Statistics
Purdue University

November 1994

Bayesian Analysis of "Random Routes"

M.J. Bayarri and B. Font

Abstract

In many practical situations involving sampling from finite populations, it is not possible (or it is prohibitively expensive) to access, or to even produce, a listing of all of the units in the population. In these situations, inferences can not be based in random samples from the population. Random routes are widely used procedures to collect data in absence of a well defined sampling frame, and they usually have either been improperly analyzed as random samples or entirely ignored as useless. We present here a Bayesian analysis of random routes that incorporates the information provided but carefully takes into account the non-randomness in the selection of the units.

KEYWORDS: Finite populations; hierarchical models; imperfect sampling frames; non-random samples.

1. Introduction

When sampling finite populations it is explicitly or implicitly assumed that there exists a well defined sampling frame from which units or elements are selected. In fact, probability sampling methods are not possible without it. However, it is very often the case in practice that inferences are desired about a finite population for which a listing of elements (or units) is not available (for example, censuses of populations in Spain can hardly be obtained), or it is impossible or prohibitively expensive to produce (as a listing of trees, wild animals, .. etc. in a forest). Still in some other situations the finite population is, in a way, temporary in nature and can not have a listing attached to it; an example is provided by the population of "tourists" in a certain touristic site.

In these situations, practitioners do still make inferences about quantities of interest, since inferences are required and the lack of a well defined sampling frame is not a excuse good enough not to produce them.

What they usually do is to select the elements from the population in some haphazard way that usually involves elaborated instructions to "guarantee" the "randomness" of the elements. A very popular, and common, practice when sampling in a city, say, consists in somehow selecting a path in the city and selecting people living in that path. The mechanisms by which the path is defined and the people from the path are selected can be quite elaborate and are meant to make the final sample "random". We first found the name "random routes" to refer to this type of sampling in Kish(1965)[1]. Practitioners do use this type of sampling widely, and we have read many variants and many ways of defining and selecting these "random routes" in the reports produced by survey sampling companies. Random routes are not only restricted to cities (can also refer to paths in a forest, for instance), nor to actual paths (can also refer to a geographic entity that can be defined from a map, say). Also, they can be defined in a casual way (walk two blocks, turn left, walk three blocks, ... etc.) or they can be well defined entities (a city street, a city square, a registered path in a forest, ... etc.).

What the users of these random routes usually do is to treat the resulting sample as if it were a random sample from the population, which usually results in an overestimation of the precision of the inferences. On the other hand, academic statisticians are usually extremely critic with the procedure and many conclude that this type of data is entirely useless and that they simply can't be analyzed in a rigorous way. This looks again as still another confrontation of "theory" versus "practice" of statistics. We show however that such an analysis is possible when Bayesian methods are used, an that the added uncertainty can be modelled and incorporated into the analysis.

Under the usual (Bayesian) approach to finite populations and when simple random sampling is used, the values of the quantity of interest $Y_1, Y_2, \ldots, Y_N$ in the finite population are assumed to be the realization of N independent, identically distributed random variables. The common distribution is usually taken to be normal with mean μ. Hence, the so-called *superpopulation model* is N-variate normal with diagonal covariance matrix.

In this paper we assume that the finite population is going to be sampled by random routes. We assume that there is a total of K possible routes with M_i elements in each route, that we sample k of them, and that m_i observations for $i = 1, 2, \ldots, k$ are taken in each sampled route. A natural way to model random routes is by means of hierarchical multistage models, in the spirit of Scott & Smith(1969)[3] and Malec & Sedransk(1985)[2]. However, the usual simplifying assumptions, namely that the size of each sampling unit (in this case, a route) M_i is known, and that the variance of the elements within a unit (route) is roughly the same for all of the units, can *not* be maintained to hold in this scenario, and this is intrinsic to random routes. (Notice also, that it would similarly not make any sense to assume that the values of some covariate are known for the whole finite population.)

To be specific, the conditional (on μ) one-stage model that we will be considering, can be described as follows: In route i, $i = 1, 2, \ldots, K$ there are M_i elements whose values of Y are assumed to be i.i.d normal with mean θ_i and variance σ_i^2. The means across the routes, θ_i's, are assumed to be independent (but not identically distributed) normal with common mean μ and variances $c_i\sigma_i^2$. The σ_i^2's in turn are assumed i.i.d. with an inverse gamma distribution. This election of the joint prior for the θ_i's and the σ_i^2's greatly simplifies the needed calculus and the number of parameters while still providing a flexible model to take into account the uncertainties in the θ_i's. (Of course, a completely general formulation would take the c_i's to also be unknown; thus, our analysis here, with fully assessed c_i's, could be re-interpreted as the needed conditional intermediate steps in a more general analysis. We shall not however pursue this analysis any further in this paper.) The conditional one-stage model is, thus,

$$\begin{aligned} Y_{ij}|\boldsymbol{\theta},\boldsymbol{\sigma^2} &\sim N(\theta_i,\sigma_i^2) \quad j=1,2,\ldots,M_i \quad i=1,2,\ldots,K \\ \theta_i|\mu &\sim N(\mu,c_i\sigma_i^2) \quad \sigma_i^2 \sim Ga^{-1}(\alpha,\beta) \quad i=1,2,\ldots,K, \end{aligned} \tag{1.1}$$

where $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_K)^t$ and $\boldsymbol{\sigma^2}=(\sigma_1^2,\ldots,\sigma_K^2)^t$. It is interesting to note the form of the super-population model as deduced from (1.1). In fact, it is no longer a N-variate normal or Student t (depending or whether or not we condition on the common variance) with diagonal covariance matrix, but a product of K M_i-variate Student t's:

$$(Y_{i1},\ldots,Y_{iM_i})^t|\mu \sim t_{M_i}(\mathbf{1}_{M_i}\mu,\frac{\beta}{\alpha}(\boldsymbol{I}_{M_i}+\mathbf{1}_{M_i}\mathbf{1}_{M_i}^t c_i),2\alpha), \tag{1.2}$$

with marginal moments

$$E(Y_{ij}|\mu) = \mu, \quad Var(Y_{ij}|\mu) = \tfrac{\beta}{\alpha-1}(1+c_i),$$

and covariances

$$Cov(Y_{ij},Y_{i^*j^*}|\mu)=\begin{cases} \frac{\beta}{\alpha-1}c_i & \text{if } i=i^* \quad \text{(same routes)} \\ 0 & \text{if } i\neq i^* \quad \text{(different routes)}. \end{cases}$$

Apart from this Introduction, the rest of the paper is organized in three more sections. In Section 2 we state the inferential goals and carry out the usual hierarchical computations conditional on μ and the M_i's. In Section 3 uncertainty about the M_i's is introduced in a simple, particular case. Finally, in Section 4, uncertainty about μ is also incorporated. The greater uncertainty in the situation, as compared with problems in which simple random sample is used, produces predictive distributions with no moments of order higher than one, as it could have been expected. Another noteworthy aspect of the analysis is that most derivations can be carried out in closed form for the special simple case that we consider here.

2. Conditional posterior and predictive distributions

Here and in the rest of the paper, we shall label, without loss of generality, the k sampled routes as routes $1,2,\ldots,k$. Similarly, we label the m_i elements observed in each route as if they were the first ones; that is, y_{ij}, $j = 1, 2, \ldots, m_i$, $i = 1, 2, \ldots, k$. Also, we denote by n the total number of observations $n = \sum_{i=1}^{k} m_i$, by $\boldsymbol{y}_s$ the $n \times 1$ vector of sampled (observed) values y_{ij}, and by $\boldsymbol{Y}_u$ the $(N-n) \times 1$ vector of unsampled ones. Their arithmetic means will be denoted by $\overline{y}_s$, $\overline{Y}_u$ respectively. Similarly, we shall denote by $\boldsymbol{y}_{s_i}$ the $m_i \times 1$ vector of sampled elements in route i, $i = 1, \ldots, k$, and by $\overline{y}_{s_i}$ its mean; $\boldsymbol{Y}_{u_i}$ will stand for the vector of unsampled elements in route i (it has dimension $(M_i - m_i) \times 1$ for $i = 1, 2, \ldots, k$, and $M_i \times 1$ for $i = k+1, \ldots, K$), and $\overline{Y}_{u_i}$ for its mean.

The usual goal when sampling finite populations is to make inferences about some function of the N values of Y in the population. A very common function of interest is the mean $\overline{Y}$ of the finite population, $\overline{Y} = \frac{\sum_{i=1}^{K} \sum_{j=1}^{M_i} Y_{ij}}{N}$. Since $\overline{Y}$ can be expressed as $\overline{Y} = f\,\overline{y}_s + (1-f)\overline{Y}_u$, where $f = \frac{n}{N}$ is the *sampling fraction*, we shall restrict ourselves to consideration of $\overline{Y}_u$, and state that our inferential aim is to predict $\overline{Y}_u$ based on the observed values $\boldsymbol{y}_s$. We shall next derive all the needed intermediate predictive an posterior distributions.

All of the distributions derived in this Section are conditional on μ and the M_i's. (Uncertainty about M_i's is added in Section 3 and about μ in Section 4.) We think it to be a good strategy, since in some problems with a lot of previous information, it might not be too crude an approximation to assume that μ in (1.1) can be assessed; also, although rare, there might be problems with full information about the routes and the M_i's would then be known. (This would actually be better described as a usual two-stage sampling model.)

The computations, although lengthy, are straightforward and will be omitted.

From the model (1.1) we directly get

$$\begin{aligned} p(\boldsymbol{y}_s|\boldsymbol{\theta}, \boldsymbol{\sigma^2}) &= N_n(\boldsymbol{y}_s| \left(diag_1^k\{\mathbf{1}_{m_i}\}\ \mathbf{0}_{n\times(K-k)}\right)\boldsymbol{\theta}\ ,\ diag_1^k\{\boldsymbol{I}_{m_i}\sigma_i^2\}), \\ p(\boldsymbol{\theta}|\mu, \boldsymbol{\sigma^2}) &= N_K(\boldsymbol{\theta}|\mathbf{1}_K\mu\ ,\ diag_1^K\{c_i\sigma_i^2\}). \end{aligned} \tag{2.1}$$

From (2.1) we derive $p(\boldsymbol{y}_s|\mu, \boldsymbol{\sigma^2})$ as

$$p(\boldsymbol{y}_s|\mu, \boldsymbol{\sigma^2}) = N_n(\boldsymbol{y}_s|\mathbf{1}_n\mu\ ,\ diag_1^k\{(\boldsymbol{I}_{m_i} + \mathbf{1}_{m_i}\mathbf{1}_{m_i}^t c_i)\sigma_i^2\}). \tag{2.2}$$

From (2.2) and the joint prior distribution for $\boldsymbol{\sigma^2}$, $p(\boldsymbol{\sigma^2}) = \prod_{i=1}^{K} Ga^{-1}(\alpha, \beta)$, we get the joint posterior distribution for $\boldsymbol{\sigma^2}$, $p(\boldsymbol{\sigma^2}|\boldsymbol{y}_s, \mu)$, which can be fully described by saying that, given μ (and the M_i's), the components σ_i^2, $i = 1, 2, \ldots, K$ are independent a posteriori, and

$$\begin{aligned} p(\sigma_i^2|\boldsymbol{y}_s, \mu) &= Ga^{-1}(\sigma_i^2|\alpha_i, \beta_i), \quad i = 1, \ldots, k \\ &= Ga^{-1}(\sigma_i^2|\alpha, \beta), \quad i = k+1, \ldots, K, \end{aligned} \tag{2.3}$$

where for $i = 1, 2, \ldots, k$,

$$\begin{aligned} \alpha_i &= \alpha + (m_i/2) \\ \beta_i &= \beta + (Q_i(\mu)/2) \\ Q_i(\mu) &= (\boldsymbol{y}_{s_i} - \mathbf{1}_{m_i}\mu)^t(\boldsymbol{I}_{m_i} + \mathbf{1}_{m_i}\mathbf{1}_{m_i}^t c_i)^{-1}(\boldsymbol{y}_{s_i} - \mathbf{1}_{m_i}\mu). \end{aligned} \tag{2.4}$$

On the other hand, the model for the unsampled elements can be directly seen from (1.1) to be

$$p(\boldsymbol{y}_u|\boldsymbol{\theta}, \boldsymbol{\sigma}^2) = N_{N-n}\left(\boldsymbol{y}_u \left| \begin{pmatrix} \mathbf{1}_{M_1-m_1}\theta_1 \\ \vdots \\ \mathbf{1}_{M_k-m_k}\theta_k \\ \mathbf{1}_{M_{k+1}}\theta_{k+1} \\ \vdots \\ \mathbf{1}_{M_K}\theta_K \end{pmatrix}, \begin{pmatrix} diag_1^k\{\boldsymbol{I}_{M_i-m_i}\sigma_i^2\} & \mathbf{0} \\ \mathbf{0} & diag_{k+1}^K\{\boldsymbol{I}_{M_i}\sigma_i^2\} \end{pmatrix} \right.\right), \tag{2.5}$$

and the conditional on $\boldsymbol{\sigma}^2$, posterior distribution for $\boldsymbol{\theta}$ can be computed from (2.1) to be

$$p(\boldsymbol{\theta}|\boldsymbol{y}_s, \mu, \boldsymbol{\sigma}^2) = N_K\left(\boldsymbol{\theta} \left| \begin{pmatrix} \tilde{\theta}_1 \\ \vdots \\ \tilde{\theta}_k \\ \mathbf{1}_{K-k}\mu \end{pmatrix}, \begin{pmatrix} diag_1^k\left\{\frac{c_i\sigma_i^2}{c_im_i+1}\right\} & \mathbf{0}_{k\times(K-k)} \\ \mathbf{0}_{(K-k)\times k} & diag_{k+1}^K\{c_i\sigma_i^2\} \end{pmatrix} \right.\right), \tag{2.6}$$

where $\tilde{\theta}_i = \frac{c_im_i\bar{y}_{s_i}+\mu}{c_im_i+1}$, $i = 1, \ldots, k$.

Hence, the posterior predictive for $\boldsymbol{Y}_u$, conditional on $\boldsymbol{\sigma}^2$ (and μ and the M_i's) is computed from (2.5) and (2.6) to be

$$p(\boldsymbol{y}_u|\boldsymbol{y}_s, \mu, \boldsymbol{\sigma}^2) = N_{N-n}\left(\boldsymbol{y}_u \left| \begin{pmatrix} \mathbf{1}_{M_1-m_1}\tilde{\theta}_1 \\ \vdots \\ \mathbf{1}_{M_k-m_k}\tilde{\theta}_k \\ \mathbf{1}_{M_u}\ \mu \end{pmatrix}, \begin{pmatrix} diag_1^k\{\boldsymbol{A}_i\sigma_i^2\} & \mathbf{0} \\ \mathbf{0} & diag_{k+1}^K\{\boldsymbol{B}_i\sigma_i^2\} \end{pmatrix} \right.\right), \tag{2.7}$$

where $M_u = \sum_{i=k+1}^K M_i$ and

$$\begin{aligned} \boldsymbol{A}_i &= \boldsymbol{I}_{M_i-m_i} + \frac{c_i\mathbf{1}_{M_i-m_i}\mathbf{1}_{M_i-m_i}^t}{c_im_i+1}, \quad i = 1, \ldots, k \\ \boldsymbol{B}_i &= \boldsymbol{I}_{M_i} + c_i\mathbf{1}_{M_i}\mathbf{1}_{M_i}^t, \quad i = k+1, \ldots, K. \end{aligned} \tag{2.8}$$

Finally, we can integrate $\boldsymbol{\sigma}^2$ out from (2.7) by using (2.3), and we get the joint distribution of $\boldsymbol{Y}_u$, which can be described by saying that the vectors $\boldsymbol{Y}_{u_i}$ are independent, for $i = 1, 2, \ldots, K$, with posterior predictive distributions given by

$$\begin{aligned} p(\boldsymbol{Y}_{u_i}|\boldsymbol{y}_s, \mu) &= t_{M_i - m_i}(\boldsymbol{y}_{u_i}|\boldsymbol{1}_{M_i-m_i}\tilde{\theta}_i, \boldsymbol{A}_i \frac{\beta_i}{\alpha_i}, 2\alpha_i), \ \ i = 1, \ldots, k \\ &= t_{M_i}(\boldsymbol{y}_{u_i}|\boldsymbol{1}_{M_i}\mu, \boldsymbol{B}_i \frac{\beta}{\alpha}, 2\alpha), \ \ i = k+1, \ldots, K, \end{aligned} \tag{2.9}$$

where $\boldsymbol{A}_i$, $\boldsymbol{B}_i$ are given in (2.8), and α_i, β_i in (2.4).

The (conditional) posterior predictive distribution of the quantity of interest can directly be deduced from (2.9), or more easily from the normal distribution $p(\overline{Y}_u|\boldsymbol{y}_s, \mu, \boldsymbol{\sigma}^2)$ (which can be directly computed from (2.7)) by integrating out $\boldsymbol{\sigma}^2$ with (2.3)). The resulting $p(\overline{Y}_u|\boldsymbol{y}_s, \mu)$ can not be expressed in closed form, but its moments can, and they are given by

$$\begin{aligned} E(\overline{Y}_u|\boldsymbol{y}_s, \mu) &= \frac{1}{N-n} \left[\sum_{i=1}^{k} (M_i - m_i)\tilde{\theta}_i + \mu \sum_{i=k+1}^{K} M_i \right] \\ Var(\overline{Y}_u|\boldsymbol{y}_s, \mu) &= \frac{1}{(N-n)^2} \left\{ \frac{\beta}{\alpha - 1} \sum_{i=k+1}^{K} M_i(1 + M_i c_i) \right. \\ & \left. + \sum_{i=1}^{k} (M_i - m_i) \left(1 + \frac{c_i(M_i - m_i)}{c_i m_i + 1} \right) \frac{2\beta + Q_i(\mu)}{m_i + 2\alpha - 2} \right\}. \end{aligned} \tag{2.10}$$

If the M_i's and μ where known, we could finish here. As argued before, however, it is almost intrinsic to sampling by random routes that the sizes of the routes can not be known in advance. In the following Section we introduce uncertainty about the M_i's into the analysis in a very simple situation which makes computations easy.

3. Uncertainty about M_i

In this Section we explicitly recognize that the sizes of all the K possible routes can not be known in advance. We assume, however, that we are able to observe the sizes M_i for the routes that we do sample. In practice, what we usually can get are accurate estimates of these sizes, and uncertainty of these estimates should also be taken into account; for simplicity, we shall not pursue this venue in this paper but assume that M_i are known for those routes that we get to sample. We change notation in this section and denote by T_i the (unknown in advance) size of route i, $i = 1, 2, \ldots, K$, and assume that we observe $T_i = M_i$, $i = 1, 2, \ldots, k$.

A very simple model for the T_i is to assume that $T_1, T_2, \ldots, T_K$ are i.i.d. Poisson with parameter λ, and give λ the usual non-informative prior $\pi(\lambda) \propto 1/\lambda$. It is now an easy exercise to compute the posterior predictive distribution of $T_{k+1}, T_{k+2}, \ldots, T_K$ given $M_1, M_2, \ldots, M_k$, as

$$p(T_{k+1}, \ldots, T_K | M_1, \ldots, M_k) = \frac{k^{M_s}}{\Gamma(M_s)} \frac{\Gamma(T_u + M_s)}{K^{T_u + M_s}} \tag{3.1}$$

where $T_u = \sum_{i=k+1}^K T_i$ and $M_s = \sum_{i=1}^k M_i$

Notice that the mean and variance of $\overline{Y}_u$ given in (2.10) are really $E(\overline{Y}_u | \boldsymbol{y}_s, \mu, M_1, \ldots, M_k, T_{k+1}, \ldots, T_K)$ and $Var(\overline{Y}_u | \boldsymbol{y}_s, \mu, M_1, \ldots, M_k, T_{k+1}, \ldots, T_K)$ and that $T_{k+1}, \ldots, T_K$ can be very easily integrated out. In fact, since, from (3.1),

$$\begin{aligned} E(T_i | M_1, \ldots, M_k) &= \frac{M_s}{k}, \quad i = k+1, \ldots, K, \\ Var\left(\sum_{i=k+1}^K \gamma_i T_i | M_1, \ldots, M_k\right) &= \frac{M_s}{k} \left[\sum_{i=k+1}^K \gamma_i^2 + \frac{(\sum_{i=k+1}^K \gamma_i)^2}{k}\right], \end{aligned} \tag{3.2}$$

we get, from (2.10) and (3.2),

$$\begin{aligned} E(\overline{Y}_u | \boldsymbol{y}_s, \mu, \{M_i\}_1^k) &= \frac{1}{N-n} \left[\sum_{i=1}^k (M_i - m_i)\tilde{\theta}_i + \mu \frac{K-k}{k} M_s\right] \\ Var(\overline{Y}_u | \boldsymbol{y}_s, \mu, \{M_i\}_1^k) &= \frac{1}{(N-n)^2} \frac{\beta}{\alpha - 1} \sum_{i=k+1}^K \left[\frac{M_s}{k} + \frac{c_i M_s}{k} \left(\frac{k+1+M_s}{k}\right)\right] \\ &+ \frac{1}{(N-n)^2} \sum_{i=1}^k (M_i - m_i) \left(1 + \frac{c_i(M_i - m_i)}{c_i m_i + 1}\right) \frac{2\beta + Q_i(\mu)}{m_i + 2\alpha - 2} \\ &+ \frac{1}{(N-n)^2} \mu^2 \frac{M_s}{k} (K-k) \left(\frac{K}{k}\right). \end{aligned} \tag{3.3}$$

4. Uncertainty about μ

Finally, we incorporate uncertainty about μ into the analysis in the form of a non-informative prior $p(\mu)$. For ease of calculus, we take the special case $\alpha = 1$ in the distribution of the σ_i^2's so that we in fact take them to have an inverse exponential distribution. Hence, we have the two-stage hierarchical model:

$$\begin{aligned} Y_{ij} | \boldsymbol{\theta}, \boldsymbol{\sigma^2} &\sim N(\theta_i, \sigma_i^2) \quad j = 1, 2, \ldots, M_i \quad i = 1, 2, \ldots, K \\ \theta_i | \mu &\sim N(\mu, c_i \sigma_i^2) \quad \sigma_i^2 \sim Ex^{-1}(\beta) \quad i = 1, 2, \ldots, K \\ p(\mu) &\propto \text{constant}. \end{aligned} \tag{4.1}$$

The conditional (on $\boldsymbol{\sigma}^2$) posterior distribution of μ is easily obtained from (2.2) and the constant prior giving

$$p(\mu|\boldsymbol{y}_s, \boldsymbol{\sigma}^2) = N\left(\mu\Big|\frac{Z_2}{Z_1}, \frac{1}{Z_1}\right), \tag{4.2}$$

where $Z_1 = \sum_{i=1}^k a_i\sigma_i^{-2}$, $Z_2 = \sum_{i=1}^k a_i\overline{y}_{s_i}\sigma_i^{-2}$, $a_i = \frac{m_i}{c_im_i+1}$, $i = 1, 2, \ldots, k$.

If we could derive the joint posterior distribution of (Z_1, Z_2), then the marginal posterior distribution of μ, $p(\mu|\boldsymbol{y}_s)$ would merely by given in terms of a two-dimensional integral,

$$p(\mu|\boldsymbol{y}_s) = \int\int p(\mu|Z_1, Z_2)p(Z_1, Z_2)dZ_1dZ_2,$$

where $p(Z_1, Z_2)$ is derived from $p(\sigma_1^2, \ldots, \sigma_k^2|\boldsymbol{y}_s)$ and μ could easily be integrated out from all of the expressions in the previous sections. Also, an estimate of $\overline{Y}_u$ could be given, from (3.3) as

$$\begin{aligned} E(\overline{Y}_u|\boldsymbol{y}_s, M_1, \ldots, M_k) &= \frac{1}{N-n}\sum_{i=1}^k (M_i - m_i)a_ic_i\overline{y}_{s_i} \\ &+ \frac{1}{N-n}\left[\sum_{i=1}^k \frac{M_i - m_i}{m_i}a_i + \frac{K-k}{k}M_s\right]E(\mu|y_s) \end{aligned} \tag{4.3}$$

where $E(\mu|\boldsymbol{y}_s) = E^{Z_1,Z_2|y_s}\left(\frac{Z_2}{Z_1}\right)$. Notice that $Var(\overline{Y}_u|\boldsymbol{y}_s, \mu, M_1, \ldots, M_k)$ does no longer exist.

The surprise here is that $p(Z_1, Z_2)$ and $E\left(\frac{Z_2}{Z_1}\right)$ can be obtained in closed form. The derivations are, however, very lengthly and cumbersome. We shall illustrate the kind of results obtained in a simple particular case.

Assume that $k = 3$ and that data is such that the following inequalities hold (other cases are handled in a completely similar way):

$$\begin{gathered} \overline{y}_{s_1} > \overline{y}_{s_2} > \overline{y}_{s_3} \\ (\overline{y}_{s_1} - \overline{y}_{s_3})a_1a_3 - (\overline{y}_{s_2} - \overline{y}_{s_3})a_2a_3 - (\overline{y}_{s_1} - \overline{y}_{s_2})a_1a_2 > 0 \\ a_1 > a_3 \text{ and } a_2 > a_3 \end{gathered}$$

Then, the distribution $p(Z_1, Z_2)$, after much algebra, can be computed to be

$$p(Z_1, Z_2) \propto \begin{cases} (\prod_{i=1}^2 \exp\{\beta(u_i - b_i)Z_i\})[\exp\{\beta(u_3 + b_3 - 1)d_2\} - 1] \\ \qquad \text{if } \overline{y}_{s_1} > \frac{Z_2}{Z_1} > \overline{y}_{s_2} \\ (\prod_{i=1}^2 \exp\{\beta(u_i - b_i)Z_i\})[\exp\{\beta(u_3 + b_3 - 1)d_2\} - \exp\{\beta(u_3 + b_3 - 1)d_1\}] \\ \qquad \text{if } \overline{y}_{s_2} > \frac{Z_2}{Z_1} > \overline{y}_{s_1} \\ 0 \qquad \text{otherwise} \end{cases} \tag{4.4}$$

where

$$
\begin{array}{lll}
d_1 = \frac{u_2 Z_2 - u_1 Z_1}{u_3} & d_2 = \frac{b_1 Z_1 - b_2 Z_2}{b_3} & \\
u_1 = \overline{y}_{s_2} u_2 & u_2 = \frac{1}{(\overline{y}_{s_1} - \overline{y}_{s_2}) a_1} & u_3 = (\overline{y}_{s_3} - \overline{y}_{s_2}) a_3 u_2 \\
b_1 = \overline{y}_{s_1} b_2 & b_2 = \frac{1}{(\overline{y}_{s_1} - \overline{y}_{s_2}) a_2} & b_3 = (\overline{y}_{s_1} - \overline{y}_{s_3}) a_3 b_2
\end{array}
$$

The proportionality constant in (4.4) can also be given in closed form and it turns out to be

$$
[\text{constant}]^{-1} = \frac{1}{\beta^2}[(\overline{y}_{s_1} - \overline{y}_{s_3}) a_1 a_3 - (\overline{y}_{s_2} - \overline{y}_{s_3}) a_2 a_3 - (\overline{y}_{s_1} - \overline{y}_{s_2}) a_1 a_2].
$$

Finally, the needed posterior expected value of μ is

$$
\begin{aligned}
E\left(\frac{Z_2}{Z_1}\right) &= \beta^{-2}[\text{constant}] \times \left\{ \frac{(\overline{y}_{s_1} - \overline{y}_{s_3}) a_1 a_3}{a_1 - a_3} (\overline{y}_{s_1} a_1 - \overline{y}_{s_3} a_3) - \right. \\
&- \left. \sum_{i=1}^{2} \frac{(\overline{y}_{s_i} - \overline{y}_{s_{i+1}}) a_i a_{i+1}}{a_i - a_{i+1}} \left[\overline{y}_{s_i} a_i - \overline{y}_{s_{i+1}} a_{i+1} - \frac{(\overline{y}_{s_i} - \overline{y}_{s_{i+1}}) a_i a_{i+1}}{a_i - a_{i+1}} \log \frac{a_i}{a_{i+1}} \right] \right\}.
\end{aligned}
$$

References

[1] Kish, L. (1965). *Survey Sampling.* Wiley. New York.

[2] Malec, D. & Sedransk, J. (1985). Bayesian Inference for Finite Population Parameters in Multistage Cluster Sampling. *J. Amer. Statist. Association.* **80**, 897-902.

[3] Scott, A. & Smith, T.M.F. (1969). Estimation in Multi-stage Surveys. *J. Amer. Statist. Association.* **64**, 830-840.