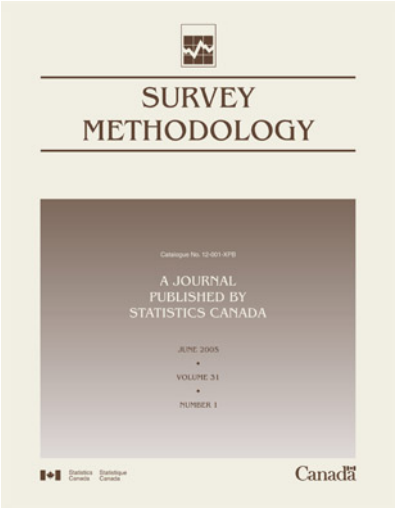
Catalogue no. 12-001-XIE

Survey Methodology

2005



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada
Business Survey Methods Division

Survey Methodology

2005

Published by authority of the Minister responsible for Statistics Canada

November 2005

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 29, No. 1, pp. 91-97
Statistics Canada, Catalogue No. 12-001

Minimum Risk, Fixed Cost Sampling Designs for Independent Poisson Processes

Brad C. Johnson and John Deely [1]

Abstract

Optimal and approximately optimal fixed cost Bayesian sampling designs are considered for simultaneous estimation in independent homogeneous Poisson processes. We develop general allocation formulae for a basic Poisson-Gamma model and compare these with more traditional allocation methods. We then discuss techniques for finding representative gamma priors under more general hierarchical models and show that, in many practical situations, these provide reasonable approximations to the hierarchical prior and Bayes risk. The methods developed are general enough to apply to a wide variety of models and are not limited to Poisson Processes.

Key Words: Optimal sampling allocations; Poisson processes; Poisson-Gamma hierarchy.

1. Introduction

The topic of Bayesian survey sampling techniques is well represented in the literature. A number of articles focus on sampling from finite populations and most make use of normality or a "posterior linearity" property (*cf.* Godambe 1955; Ericson 1988; Ericson 1969; Scott and Smith 1971; Tiwari and Lahiri 1989). An excellent review of recent Bayesian methods for sampling finite populations is contained in (Ghosh and Meeden 1997) as well as some interesting new approaches. Lindely and Deely (1993) discuss optimal allocation in stratified sampling under a normal model when only partial information is available. In terms of Poisson models, Clevenson and Zidek (1975) discuss the simultaneous estimation of means in independent Poisson processes and Leite, Rodrigues and Milan (2000) discuss a Bayesian analysis when estimating the number of species in a population using a non-homogeneous Poisson process. Little work has been done on model specific sampling designs from a Bayesian perspective.

In the present paper we take a model based approach to develop optimal and approximately optimal fixed cost sampling allocations for simultaneous estimation in multiple independent Poisson processes. Section 2 introduces the model and some notation. Section 3 presents the general allocation problem and gives the minimum Bayes risk allocations when independent conjugate gamma priors are assumed for each process. Comparisons are made with classical stratified random sampling allocations. In section 4 we describe techniques for finding "representative" conjugate priors under more general hierarchical models thus allowing (at least approximately) optimal sampling allocations to be determined for this larger class of models. In many situations, these representative conjugate priors can be used to reduce the hierarchical model for the purposes of posterior analysis as well. A full numerical example is presented in section 5.

2. Model and Notation

To avoid the necessity for subscripting, we first present the model and notation in terms of a single homogeneous Poisson process. Let (Ω, F, ν) be a measure space, let $\{N(A): A \in F\}$ be a homogeneous Poisson Process on (Ω, F, ν) with unknown intensity $\theta \in \Theta = (0, \infty)$ and, for any $A \in F$, let $\boldsymbol{X} = (X, m) = (N(A), \nu(A))$ denote a complete sufficient statistic with realization $\boldsymbol{x} = (x, m)$. Less formally, x is the realization of a Poisson count from a sample of "size" m. The p.m.f. of $\boldsymbol{X}$ is given by

$$f(\boldsymbol{x} \mid \theta) = \frac{(m\theta)^x e^{-m\theta}}{\Gamma(x+1)} I_{\{0,1,2,\ldots\}}(x), \quad \theta \in (0, \infty). \quad (1)$$

We express our prior beliefs about the parameter θ by a conjugate gamma distribution with shape parameter α and scale parameter β, denoted Gamma (α, β), with density

$$\pi(\theta \mid \boldsymbol{\lambda}) = \frac{\theta^{\alpha-1} e^{-\theta/\beta}}{\beta^{\alpha}\Gamma(\alpha)} I_{(0,\infty)}(\theta), \quad \boldsymbol{\lambda} = (\alpha, \beta) \in (0, \infty)^2. \quad (2)$$

We presently restrict our attention to the case when $\boldsymbol{\lambda}$ can be specified; the addition of hyper-priors on $\boldsymbol{\lambda}$ is considered in section 4.

For an arbitrary action a in the action space $A = \Theta$, we consider the loss functions

$$L_k(\theta, a) = \frac{(\theta - a)^2}{\theta^k}, \qquad k = 0, 1. \quad (3)$$

L_0 is the ordinary squared error loss and L_1 is the relative squared error loss. For L_1 we require that $\alpha > 1$ which implies the gamma prior is unimodal.

1. Brad C. Johnson, Department of Statistics, Purdue University, West Lafayette, IN 47907. E-mail: bradj@stat.purdue.edu; John Deely, Department of Statistics, Purdue University, West Lafayette, IN 47907.

Under the loss functions L_k the above model is extremely well understood. To simply notation somewhat, let $\pi^{\lambda} = \pi(\theta \mid \boldsymbol{\lambda})$ and let δ_k^{λ} denote the Bayes procedure for π^{λ} under the loss function L_k. We recall that the posterior distribution of θ given $\boldsymbol{x}$ is

$$\theta \mid \boldsymbol{x} \sim \text{Gamma}\left(\alpha + x, \frac{\beta}{m\beta + 1}\right).$$

The Bayes procedure for loss function L_k is given by

$$\delta_k^{\lambda}(\boldsymbol{x}) = \frac{\beta(\alpha + x - k)}{m\beta + 1}, \qquad \alpha > k. \tag{4}$$

The posterior expected loss in using δ_k^{λ} under the loss function L_k is

$$\rho(\pi^{\lambda},\ \delta_k^{\lambda},\ L_k) = \left(\frac{\beta}{m\beta + 1}\right)^{2-k} (\alpha + x - k)^{1-k}, \quad \alpha > k; \tag{5}$$

with Bayes risk

$$r(\pi^{\lambda},\ \delta_k^{\lambda},\ L_k) = \frac{\alpha^{1-k}\,\beta^{2-k}}{m\beta + 1}, \qquad \alpha > k. \tag{6}$$

It is interesting to note that under L_1, (4) and (5) imply that the Bayes procedure $\delta_1^{\lambda}(\boldsymbol{x})$ is the mode of the posterior and that $\rho(\pi^{\lambda},\ \delta_1^{\lambda},\ L_1)$ does not depend on the observed count *x* and hence is constant.

It is often more convenient, in terms of the elicitation process, to allow the shape parameter α of the gamma prior for θ to depend on the scale parameter β. In particular, the following alternate parameterizations are used throughout:

$$\theta \mid \boldsymbol{\lambda} \sim \text{Gamma}(\mu/\beta,\ \beta),\ \boldsymbol{\lambda} = (\mu,\ \beta),\ E(\theta \mid \boldsymbol{\lambda}) = \mu; \tag{7}$$

$$\theta \mid \boldsymbol{\lambda} \sim \text{Gamma}(\eta/\beta + 1, \beta), \boldsymbol{\lambda} = (\eta, \beta), \text{Mode}(\theta \mid \boldsymbol{\lambda}) = \eta. \tag{8}$$

Unless specified otherwise, results and formulae for these alternate parameterizations can be obtained by simply substituting the proper value for α. For $\boldsymbol{\lambda}$ as in (7) or (8) we substitute $\alpha = \mu/\beta$ or $\eta/\beta + 1$ respectively.

3. Optimal Allocation

We now discuss the allocation of sampling effort when $\{N_s(A) : A \in F_s\}$, for $s = 1, ..., S$ are independent homogeneous Poisson processes on corresponding measure spaces (Ω_s, F_s, v_s) with unknown intensities θ_s. The realization of a sample is now denoted $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_S)$ where the $\boldsymbol{x}_s = (x_s, m_s)$ have the same meanings as $\boldsymbol{x}_s = (x, m)$ in section 2. For each process, $s = 1, ..., S$, we assume that

$$\boldsymbol{X}_s \mid \theta_s \sim \text{Poisson}(m_s\ \theta_s);$$
$$\theta_s \mid \boldsymbol{\lambda}_s \sim \text{Gamma}(\alpha_s, \beta_s), \qquad \boldsymbol{\lambda}_s = (\alpha_s,\ \beta_s).$$

Notice that we have not assumed that the θ_s are exchangeable so that prior information about one process is not influenced by the others.

Let $\boldsymbol{\delta}_k^{\lambda} = \boldsymbol{\delta}_k^{\lambda}(\boldsymbol{x}) = (\delta_k^{\lambda_1}(\boldsymbol{x}), ..., \delta_k^{\lambda_S}(\boldsymbol{x}_S))$ be the component-wise vector of Bayes procedures for estimating $\boldsymbol{\theta} = (\theta_1, ..., \theta_S)$ under the loss function L_k and let $\pi^{\boldsymbol{\lambda}}$ denote the overall prior specification. We assume that the overall loss for estimating some (possibly vector valued) function $g(\boldsymbol{\theta})$ with $g(\boldsymbol{\delta}_k^{\lambda})$ can be expressed as

$$L_k(g(\boldsymbol{\theta}),\ g(\boldsymbol{\delta}_k^{\lambda})) = \sum_{s=1}^{S} w_s\ L_k(\theta_s,\ \delta_k^{\lambda_s}(\boldsymbol{x}_s)), \tag{9}$$

where the w_s are known arbitrary non-negative weights. In particular this covers the case when we are interested in the simultaneous estimation of $\boldsymbol{W\theta}$ where $\boldsymbol{W} = (w_{js})$ is a $J \times S$ matrix and the loss structure is of the form

$$L_k(\boldsymbol{W\theta}, \boldsymbol{W\delta}_k^{\lambda}) = \sum_{j=1}^{J} \sum_{s=1}^{S} L_k(w_{js}\theta_s, w_{js}\delta_k^{\lambda_s})$$
$$= \sum_{s=1}^{S} \left(\sum_{j=1}^{J} w_{js}^{2-k}\right) L_k(\theta_s, \delta_k^{\lambda_s}). \tag{10}$$

The weights in (9) become $w_s = \sum_{j=1}^{J} w_{js}^{2-k}$ and, by the linearity of the expectation operator, the overall Bayes risk is given by

$$r(\pi^{\lambda},\ \boldsymbol{W\delta}_k^{\lambda},\ L_k) = \sum_{s} w_s r(\pi^{\lambda_s},\ \delta_k^{\lambda_s},\ L_k).$$

Let $(\xi = \xi_1, ..., \xi_S)$ denote the full specification where $\xi_s = (\alpha_s, \beta_s, w_s, c_s)$ denotes the specification for process *s* and c_s is the per unit sampling cost within that process. The general allocation problem involves finding an $\boldsymbol{m} = (m_1, ..., m_S)$ that minimizes the total risk $r(\pi^{\lambda},\ g(\boldsymbol{\delta}^{\lambda}),\ L_k)$ of $g(\boldsymbol{\delta}^{\lambda})$ subject to the constraint

$$C = \sum_{s=1}^{S} c_s m_s;$$

where *C* is the fixed total sampling budget. The proof of the following result is routine and deferred to the appendix.

Result 1. Let $\xi = (\xi_1, ..., \xi_S)$ be given. *The allocation* $\boldsymbol{m} = (m_1, ..., m_S)$ *that minimizes* $r(\pi^{\lambda},\ g(\boldsymbol{\delta}_0^{\lambda}),\ L_0)$ *for fixed total cost C is*

$$m_s = \frac{\sqrt{w_s\alpha_s\beta_s/c_s}}{\sum_s \sqrt{w_s\alpha_s\beta_s c_s}}\left(C + \sum_s \frac{c_s}{\beta_s}\right) - \frac{1}{\beta_s}. \tag{11}$$

The allocation that minimizes $r(\pi^{\lambda},\ g(\boldsymbol{\delta}_1^{\lambda}),\ L_1)$ *is*

$$m_s = \frac{\sqrt{w_s/c_s}}{\sum_s \sqrt{w_s c_s}}\left(C + \sum_s \frac{c_s}{\beta_s}\right) - \frac{1}{\beta_s}. \tag{12}$$

Equations (11) and (12) can result in one or more $m_s \le 0$ (*i.e.*, we take no samples from the offending processes) in which case we would remove these processes and re-allocate *C* among the remaining processes. Of course, for the removed processes, our posterior mean and variance are equivalent to the prior mean and variance of θ_s.

We also comment that the allocation which minimizes $r(\pi^{\lambda}, \boldsymbol{\delta}_1^{\lambda}, L_1)$ in (12) also minimizes $\rho(\pi^{\lambda}, \boldsymbol{\delta}_1^{\lambda}, L_1)$ since this latter quantity is free of the observed counts x_s.

3.1 Comparisons with Traditional Frequentist Sampling Allocations

A special case of the above result is when the $\{N_s(A): A \in F_s\}$ can be thought of as a stratification of a single non-homogeneous Poisson process $\{N(A): A \in F\}$ and we are interested in estimating the overall population mean, say θ. To this end, let W_s denote the relative size of each Ω_s (which is assumed to be finite) and consider estimating the overall population mean $\theta = \mathbf{W}\boldsymbol{\theta}$, where $\mathbf{W} = (W_1, ..., W_S)$, with the decision rule $\mathbf{W}\delta_0$. The weights in this case are $w_s = W_s^2$ and, substituting into (11), we obtain

$$m_s = \frac{W_s\sqrt{\alpha_s\beta_s / c_s}}{\sum_s W_s\sqrt{\alpha_s\beta_s c_s}}\left(C + \sum_s \frac{c_s}{\beta_s}\right) - \frac{1}{\beta_s}.$$

Letting $\beta_s \to \infty$ and $\alpha_s \to 0$ such that $\alpha_s\beta_s \to \mu_s$ simultaneously for each process is equivalent to letting $E(\theta_s) \to \mu_s$ and $\mathrm{Var}(\theta_s \to \infty)$ for all *s* and we obtain

$$m_s = \frac{CW_s\sqrt{\mu_s / c_s}}{\sum_s W_s\sqrt{\mu_s c_s}}. \qquad (13)$$

This expression, up to the finite population correction factor, is the traditional frequentist allocation under the parametric model $\boldsymbol{X}_s = (X_s, 1) \sim \mathrm{Poisson}(\mu_s)$ where μ_s represents our "best guess" for the mean (and hence variance) of $\boldsymbol{X}_s$ (*cf.* Cochran 1977). When $c_s = 1$ for all *s*, this becomes the Neyman allocation when the finite population correction factor is ignored.

Assuming that all of the μ_s are the same in (13) yields

$$m_s = \frac{CW_s / \sqrt{c_s}}{\sum_s W_s\sqrt{c_s}}; \qquad (14)$$

and, when $c_s = 1$ for all *s*, we obtain the usual proportional allocation for fixed total sample size $C = N$.

In this sense, we see that the traditional frequentist solutions to the allocation problem can be obtained as the appropriate limit of Bayes solutions just as the traditional frequentist estimates can be obtained as a limit of Bayes procedures.

4. Representative Conjugate Priors Under Hierarchical Models

Up until now we have assumed that the $\boldsymbol{\lambda}_s$ were known. Returning to the notation of section 2 we now consider amore general hierarchical model

$$\begin{aligned} \boldsymbol{X} \mid \theta &\sim \mathrm{Poisson}(m\theta); \\ \theta \mid \boldsymbol{\lambda} &\sim \mathrm{Gamma}(\alpha, \beta); \\ \boldsymbol{\lambda} &\sim h(\boldsymbol{\lambda}) \quad \boldsymbol{\lambda} \in H. \end{aligned} \qquad (15)$$

We restrict our attention to choices of $h(\boldsymbol{\lambda})$ where the Bayes risk is finite and this precludes, among other things, the use of improper $h(\boldsymbol{\lambda})$. The unconditional prior for θ under this model can, at least in principle, be obtained as

$$\pi(\theta) = E^{h(\lambda)}\pi(\theta \mid \boldsymbol{\lambda}).$$

In practice however, there is little to be gained since the resulting $\pi(\theta)$ will usually not be expressible in closed form. Indeed, it is usually the case that numeric integration and/or simulation is required to obtain the required posterior quantites and the Bayes risk.

We propose two methods for finding a "representative" *single conjugate prior* which, in most cases, can be substituted for $\pi(\theta)$ for the purposes of allocation. Indeed, for many practical situations, we find that these "representtative" conjugate priors can replace the hierarchical model completely, greatly simplifying the posterior analysis.

We assume that it is relatively easy to simulate a sequence of random variables, $\{\boldsymbol{\lambda}_n\}_{n=1}^N$, from $h\{\boldsymbol{\lambda}\}$ and, as such, a sequence of random variables, $\{\theta_n\}_{n=1}^N$, can be obtained easily from $\pi(\theta)$ by taking $\theta_i \sim \pi(\theta \mid \boldsymbol{\lambda}_i)$.

We now discuss the two techniques for finding the representative conjugate prior.

4.1 The Minimum L_∞ Conjugate Prior

Let $F(\theta)$ and $F(\theta \mid \boldsymbol{\lambda})$ denote the distribution functions of $\pi(\theta)$ and $\pi(\theta \mid \boldsymbol{\lambda})$ respectively. The L_∞ conjugate prior, or $L_\infty C$ prior, is defined to be the prior $\pi^\infty = \pi(\theta \mid \boldsymbol{\lambda}^\infty)$ where $\boldsymbol{\lambda}^\infty$ is chosen such that

$$\left\| F(\theta) - F(\theta \mid \boldsymbol{\lambda}^\infty) \right\|_\infty = \inf_{\boldsymbol{\lambda} \in \mathrm{H}} \| F(\theta) - F(\theta \mid \boldsymbol{\lambda}) \|_\infty .$$

That is, the $L_\infty C$ prior is the prior $\pi(\theta \mid \boldsymbol{\lambda})$ which minimizes the L_∞ distance between $F(\theta)$ and $F(\theta \mid \boldsymbol{\lambda})$.

In order to estimate such a $\pi(\theta \mid \boldsymbol{\lambda}^\infty)$ let $\{\theta_i\}_{i=1}^N$ be N simulated values from the unconditional prior $\pi(\theta)$; let $\theta_{i:N}$ denote the i^{th} ordered value of the $\{\theta_i\}$; and define the function

$$d_N(\boldsymbol{\lambda}) = \max_i \left| F(\theta_{i:N} \mid \boldsymbol{\lambda}) - \frac{i - 0.5}{N} \right|. \qquad (16)$$

It is usually a routine matter to numerically find an (at least approximate) minimizing $\boldsymbol{\lambda}$ for (16) and our $L_\infty C$ prior is $\pi(\theta \mid \boldsymbol{\lambda}^\infty)$ where $\boldsymbol{\lambda}^\infty$ satisfies

$$d_N(\boldsymbol{\lambda}^\infty) = \inf_{\boldsymbol{\lambda} \in \mathrm{H}} d_N(\boldsymbol{\lambda}).$$

Note that we are essentially minimizing the Kolmogorov-Smirnov statistic and the obvious appeal of estimating $\pi(\theta)$ in this manner is that $d_N(\boldsymbol{\lambda}^\infty)$ can be directly interpreted as the estimated maximum difference of

cumulative probabilities under $\pi(\theta)$ and $\pi(\theta|\boldsymbol{\lambda}^{\infty})$. In the sequel, we will denote the Bayes procedure under the prior π^{∞} and loss function L_k as $\delta_k^{\infty}(\boldsymbol{x})$.

4.2 The ML Conjugate Prior

Let $\theta_1, ..., \theta_N$ be N simulated values from $\pi(\theta)$. The ML conjugate prior, or ML-C prior, when it exists, is defined to be the prior $\pi(\theta|\boldsymbol{\lambda}^{\text{ml}})$ where $\boldsymbol{\lambda}^{\text{ml}}$ satisfies

$$\pi(\theta|\boldsymbol{\lambda}^{\text{ml}}) = \sup_{\lambda\in H} L(\boldsymbol{\lambda}|\boldsymbol{\theta}) = \sup_{\lambda\in H} \prod_{i=1}^{N} \pi(\theta_i|\boldsymbol{\lambda});$$

or, equivalently,

$$\ln\ \pi(\theta|\boldsymbol{\lambda}^{\text{ml}}) = \sup_{\lambda\in H} l(\boldsymbol{\lambda}|\boldsymbol{\theta}) = \sup_{\lambda\in H} \sum_{i=1}^{N} \ln\ \pi(\theta_i|\boldsymbol{\lambda}).$$

That is, $\boldsymbol{\lambda}^{\text{ml}}$ is the usual maximum likelihood estimator of $\boldsymbol{\lambda}$ if $\theta_1, ..., \theta_N$ were i.i.d. from $\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$. Again, it is usually a simple matter to obtain $\boldsymbol{\lambda}^{\text{ml}}$ by numerical or simulation techniques. As in the L_{∞} method, we will let π^{ml} and δ_k^{ml} denote the estimated prior and the Bayes procedure under π^{ml} and loss function L_k.

Examples. The following four examples give an indication of how these procedures perform for a few different choices of $h(\boldsymbol{\lambda})$. In all of the examples we consider the general fist stage setup to be

$$\boldsymbol{X}|\theta \sim \text{Poisson}(m\theta)$$

$$\theta|\boldsymbol{\lambda} \sim \text{Gamma}(\eta/\beta + 1,\ \beta) \qquad \boldsymbol{\lambda} = (\eta,\ \beta)$$

Furthermore, we assume η and β are independent so that $h(\boldsymbol{\lambda})$ may be written as $h_1(\eta)h_2(\beta)$. Adopting the notational conventions

$$Y \sim \text{Beta}_{(a,\ b)}(\zeta_1,\ \zeta_2) \Rightarrow f(y) \propto (y-a)^{\zeta_1-1}(b-y)^{\zeta_2-1} I_{(a,\ b)}(y);$$

$$Y \sim \text{InvGamma}(a,\ b) \Rightarrow f(y) \propto y^{-(a+1)} e^{-1/yb} I_{(0,\infty)}(y);$$

the four examples considered are

Example	η	β
(a)	Uniform (4, 6)	$\text{Beta}_{(0.5,\ 2)}(2,\ 5)$
(b)	Gamma (6.25, 0.8)	InvGamma (11, 1/30)
(c)	Uniform (2, 18)	Uniform (0.2, 0.5)
(d)	$\text{Beta}_{(3,\ 15)}(2,\ 1)$	$\text{Beta}_{(0.1,\ 0.5)}(1,\ 2)$

Table 1 gives the estimated $\boldsymbol{\lambda}^{\infty}$ and $\boldsymbol{\lambda}^{\text{ml}}$ with $d_N(\boldsymbol{\lambda}^{\infty})$ and, for comparison, $d_N(\boldsymbol{\lambda}^{\text{ml}})$ for each of these examples where all of the estimates are based on $N = 100{,}000$ simulated values from $\pi(\theta)$. In examples (a) and (b) both methods give very similar results and provide very good fits to $\pi(\theta)$ as indicated by the small values of d_N. Examples (c) and (d) were chosen to illustrate what happens when $\pi(\theta)$ deviates noticeably from a gamma distribution. Examples (c) has a "plateau" distribution and example (d) is skewed in the wrong direction. As expected, the fits are less convincing in these examples. Figure 1 shows the simulated $\pi(\theta)$ along with π^{∞} and π^{ml} for each of these examples.

Table 1
Estimated $\boldsymbol{\lambda}^{\infty}$ and $\boldsymbol{\lambda}^{\text{ml}}$ for examples (a) – (d)

Example	$\boldsymbol{\lambda}^{\infty}$	$d_N(\boldsymbol{\lambda}^{\infty})$	$\boldsymbol{\lambda}^{\text{ml}}$	$d_N(\boldsymbol{\lambda}^{\text{ml}})$
(a)	(4.94, 0.98)	0.003	(4.93, 1.00)	0.006
(b)	(4.42, 3.53)	0.003	(4.35, 3.63)	0.006
(c)	(7.72, 2.92)	0.043	(7.38, 2.93)	0.065
(d)	(10.44, 1.01)	0.040	(10.12, 1.10)	0.068

A more important consideration, for the purposes of the allocations discussed in section 3, is how well the Bayes risks are approximated under π^{∞} and π^{ml}. Table 2 gives the Bayes risk, $r(\pi,\ \delta_k^{\pi},\ L_k)$ under the hierarchical model and the values for $r_k^*(\pi^{\infty})$ and $r_k^*(\pi^{\text{ml}})$ where

$$r_k^*(\pi^{\bullet}) = \frac{r(\pi,\ \delta_k^{\pi},\ L_k) - r(\pi^{\bullet},\ \delta_k^{\bullet},\ L_k)}{r(\pi,\ \delta_k^{\pi},\ L_k)} \qquad (17)$$

and where $\bullet = \infty$ or ml for each of the examples. The values $r(\pi,\ \delta_k^{\pi},\ L_k)$ in this table were obtained by simulation and are subject to a certain amount of variation. Repeated simulations produced similar results. In examples (a) and (b) the correspondence between the Bayes risk under the representative priors is very close, especially for the ML-C priors. In examples (c) and (d) the correspondence is still quite good considering these relatively small sample sizes. Overall, the ML-C prior appears to perform slightly better in the sense that the Bayes risks $r(\pi^{\text{ml}},\ \delta_k^{\text{ml}},\ L_k)$ tend to be closer to $r(\pi,\ \delta_k^{\pi},\ L_k)$ with the exception of example (c) under the loss function L_1 where the L_{∞}-C prior is slightly better.

In examples (a) and (b) on may ask why a hierarchical model would be considered in the first place. The answer lies in the relative ease of eliciting absolute or probabilistic bounds on the hyper-paramteres involved and taking $h(\boldsymbol{\lambda})$ to represent this uncertainty. The methods above can then, in many pratical situations, be used to determine a representative single conjugate gamma prior for θ thus greatly simpligying the posterior analysis. The next section illustrates this with an example.

We also point out that it is relatively easy to construct examples where the methods described in this section will fail miserably at not only approximating π but also the Bayes risk. The method is best suited to cases where $h(\boldsymbol{\lambda})$ is chosen to represent uncertainty about $\boldsymbol{\lambda}$. In situations when $h(\boldsymbol{\lambda})$ is being used to change the fundamental behavior of the first stage gamma prior (to create a bimodal prior for example) the presentative priors π^{∞} and π^{ml} would normally not be used as a placement for π in the posterior analysis but may still give suitable approximations of the Bayes risk for the purposes of allocation.

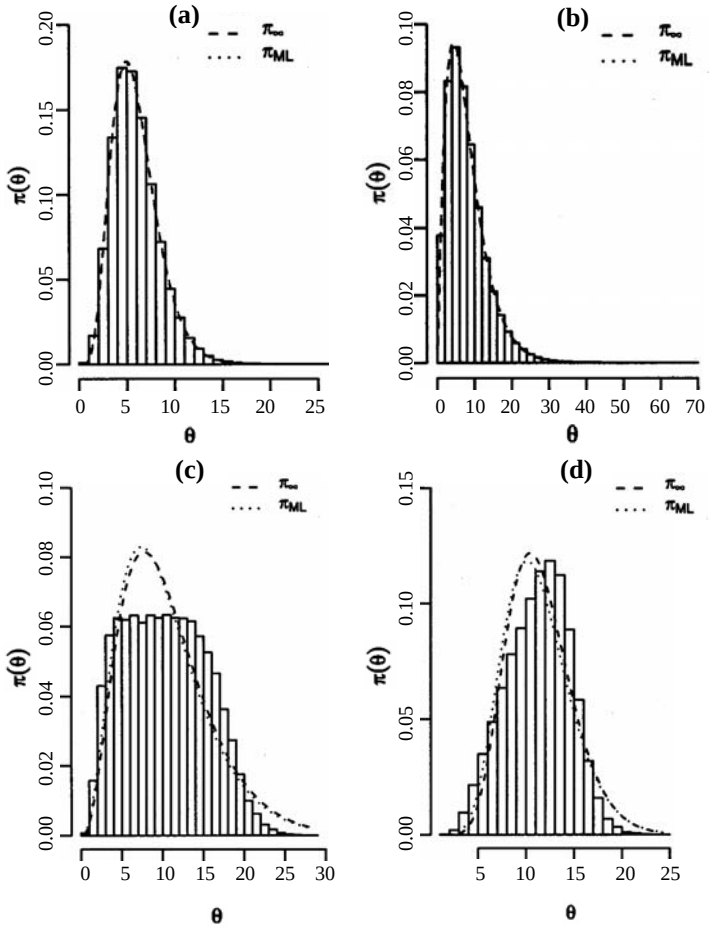


Figure 1. Simulated prior $\pi(\boldsymbol{\theta})$ (histogram) and the representative priors π^{∞} and π^{ml} for examples (a) – (d).

Table 2
Bayes Risks for Examples (a) – (d)

Example (a)						
	L_0			L_1		
m	$r(\pi, \delta^{\pi})$	$r_0^*(\pi^{\infty})$	$r_0^*(\pi^{\mathrm{ml}})$	$r(\pi, \delta^{\pi})$	$r_1^*(\pi^{\infty})$	$r_1^*(\pi^{\mathrm{ml}})$
1	2.997	−0.018	−0.006	0.500	−0.006	−0.002
5	0.986	−0.004	0.001	0.167	−0.003	0.000
10	0.540	−0.006	−0.002	0.091	−0.002	0.000
Example (b)						
	L_0			L_1		
m	$r(\pi, \delta^{\pi})$	$r_0^*(\pi^{\infty})$	$r_0^*(\pi^{\mathrm{ml}})$	$r(\pi, \delta^{\pi})$	$r_1^*(\pi^{\infty})$	$r_1^*(\pi^{\mathrm{ml}})$
1	6.320	−0.021	−0.009	0.791	−0.015	−0.008
5	1.524	−0.014	−0.007	0.190	−0.003	−0.002
10	0.779	−0.008	−0.002	0.097	−0.002	−0.001
Example (c)						
	L_0			L_1		
m	$r(\pi, \delta^{\pi})$	$r_0^*(\pi^{\infty})$	$r_0^*(\pi^{\mathrm{ml}})$	$r(\pi, \delta^{\pi})$	$r_1^*(\pi^{\infty})$	$r_1^*(\pi^{\mathrm{ml}})$
1	6.861	0.154	0.121	0.725	0.027	0.028
5	1.836	0.084	0.052	0.183	0.023	0.024
10	0.968	0.062	0.031	0.095	0.015	0.015
Example (d)						
	L_0			L_1		
m	$r(\pi, \delta^{\pi})$	$r_0^*(\pi^{\infty})$	$r_0^*(\pi^{\mathrm{ml}})$	$r(\pi, \delta^{\pi})$	$r_1^*(\pi^{\infty})$	$r_1^*(\pi^{\mathrm{ml}})$
1	5.251	0.096	0.121	0.523	−0.040	0.002
5	1.778	0.075	0.068	0.165	0.013	0.027
10	0.986	0.057	0.043	0.090	0.013	0.021

5. Numeric Example

We now present a numerical example based on data in Lindley and Deely (1993). The data consists of traffic counts between the hours of 7 a.m. and 6 p.m. over a 341 day period (3,751 hours) for a particular street in Auckland, New Zealand. The hours are stratified into $M_1 = 2{,}673$ weekday hours and $M_2 = 1{,}078$ weekend hours and we assume that the number of vehicles per hour can be modeled by two independent Poisson processes. For the purposes of this example we assume a total budget of \$1,500 is to be allocated and that per hour sampling costs are $c_1 = \$10$ and $c_2 = \$5$ for weekdays and weekends respectively. The relative strata sizes in this case are $W_1 = 0.71261$ and $W_2 = 0.28739$ for weekdays and weekends respectively.

The prior belief is that the weekend traffic rate is 40 vehicles per hour and that weekend traffic accounts for 5% of the total weekly traffic which yields a weekday traffic rate of 304 vehicles per hour. Suppose also that, for weekday traffic, we have elicited that the number of vehicles per hour will rarely exceed 400 and that, for weekend days, the number of vehicles per hour will rarely exceed 60, that is, say

$$\Pr(\boldsymbol{X}_1 \le 400) \approx 0.95 \quad \text{and} \quad \Pr(\boldsymbol{X}_2 \le 600) \approx 0.95.$$

Making use of the fact that the marginal distribution of $\boldsymbol{x}_s$ given $\boldsymbol{\lambda}_s$ is a "number of failure" negative binomial distribution of "size" $\alpha = \eta/\beta + 1$ and success probability $1/(m\beta + 1)$ we find that, when $\eta_1 = 304$ and $\eta_2 = 40$, the β_s's that come closest to satisfying these elicited probabilities are $\beta_1 = 7.51$ and $\beta_2 = 1.74$ respectively.

We now assume that the modes of the traffic rates for weekdays and weekends are equally likely to be within approximately 10% of the elicited traffic rates of 304 and 40 respectively and take

$$\eta_1 \sim \text{Uniform}(274, 334) \quad \text{and} \quad \eta_2 \sim \text{Uniform}(36, 44).$$

To represent our uncertainty about the β_s we take

$$\beta_1 \sim \text{InvGamma}(11,\ 0.0136)$$

and

$$\beta_2 \sim \text{InvGamma}(14.25,\ 0.043);$$

which yields $E(\beta_1) = 7.5$ with $\Pr(4 < \beta_1 \le 13.4) \approx 0.95$ and $E(\beta_2) = 1.75$ with $\Pr(1.03 \le \beta_1 \le 2.97) \approx 0.95$.

Using the ML-C technique in section 4 with $N = 100{,}000$, the specifications for weekday $(s = 1)$ and weekend $(s = 2)$ hourly traffic rates along with the values $d_N(\pi^{\text{ml}})$ are

s	c_s	W_s	η_s^{ml}	β_s^{ml}	$d_N(\pi^{\text{ml}})$
1	10	0.71261	302.98	8.303	0.0055
2	5	0.28739	39.876	1.889	0.0060

For the remainder of this section we will dispense with the superscript "ml" and simply refer the prior specification as π^{λ} and let $\boldsymbol{\delta}^{\lambda}(\boldsymbol{x}) = (\delta_1^{\lambda}, ..., \delta_S^{\lambda})$ denote the component-wise vector of Bayes procedures for estimating $\boldsymbol{\theta} = (\theta_1, ..., \theta_S)$ under the prior specification π^{λ} and loss function L_0.

We consider three different allocations based on estimating $\mathbf{W}_1\boldsymbol{\theta}$, $\mathbf{W}_2\boldsymbol{\theta}$ and $\mathbf{W}\boldsymbol{\theta}$ where

$$\mathbf{W}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ \mathbf{W}_2 = [W_1\ W_2] \quad \text{and} \quad \boldsymbol{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix}.$$

With $\mathbf{W}_1$ we are primarily interested in estimating the weekday and weekend traffic rates θ_1 and θ_2 individually; with $\mathbf{W}_2$ we are only interested in estimating the overall traffic rate $\theta = W_1\theta_1 + w_2\theta_2$; and, with $\mathbf{W}$, we are interested in estimating all of these. In the sequel, we will refer to the allocations as $\boldsymbol{m}(\mathbf{W}_1)$, $\boldsymbol{m}(\mathbf{W}_2)$ and $\boldsymbol{m}(\mathbf{W})$ respectively.

Table 3 gives the allocations and corresponding weights w_s for these examples based on (11) and table 4 shows the Bayes risks in estimating θ_1, θ_2, $\mathbf{W}_1\boldsymbol{\theta}$, $\mathbf{W}_2\boldsymbol{\theta}$ and $\mathbf{W}\boldsymbol{\theta}$ under these 3 allocations. While allocation $\boldsymbol{m}(\mathbf{W}_2)$ is optimal for estimating the overall traffic rate θ, it results in large increases in the Bayes risks when estimates for the weekday and weekend traffic rates are also desired – the Bayes risk for estimating θ_2 under $\boldsymbol{m}(\mathbf{W}_2)$ is almost double compared to the Bayes risk under $\boldsymbol{m}(\mathbf{W})$.

Table 3
Weights and Allocations for $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}$

	$\boldsymbol{m}(\mathbf{W}_1)$		$\boldsymbol{m}(\mathbf{W}_2)$		$\boldsymbol{m}(\mathbf{W})$	
s	w_s	m_s	w_s	m_s	w_s	m_s
1	1	119.33	0.5078	136.04	1.5078	123.20
2	1	61.35	0.0826	27.92	1.0826	53.60

Table 4
Bayes Risks Under Allocations $\boldsymbol{m}(\mathbf{W}_1)$, $\boldsymbol{m}(\mathbf{W}_1)$ and $\boldsymbol{m}(\mathbf{W}_1)$

	Estimand				
$\boldsymbol{m}$	θ_1	θ_2	$\mathbf{W}_1\boldsymbol{\theta}$	$\mathbf{W}_2\boldsymbol{\theta}$	$\mathbf{W}\boldsymbol{\theta}$
$\boldsymbol{m}(\mathbf{W}_1)$	2.61	0.68	3.28	1.38	4.66
$\boldsymbol{m}(\mathbf{W}_2)$	2.29	1.47	3.75	1.28	5.04
$\boldsymbol{m}(\mathbf{W})$	2.52	0.77	3.29	1.35	4.64

6. Concluding Comments

The techniques employed in the present paper are general enough to apply to a wide variety of Bayesian models. Optimal allocation equations for other Bayesian models. Optimal allocation equations for other Bayesian models in which the prior beliefs can, at least approximately, be modeled by conjugate priors are usually easy to obtain. The idea of finding "representative" conjugate priors, as discussed in section 4, is also applicable to a wide variety of hierarchical models with first stage conjugate priors. Areas of additional research in this area include allocations under

loss functions other that L_0 and L_1 as well as more complicated cost functions.

7. Acknowledgements

The authors would like to thank Dennis Lindley for his comments on an earlier version of this paper. We would also like to thank the referees for their valuable input and comments. Partially supported by National Science Foundation grant NFS/DMS 9983601.

Appendix A. Proof of Result 1

Proof of Result 1. Introducing the Lagrange multiplier λ, we wish to minimize, for loss function L_k,

$$\sum_{s=1}^{S} \frac{w_s \alpha_s^{1-k} \beta_s^{2-k}}{m_s \beta_s + 1} + \lambda \left(\sum_s m_s c_s - C \right).$$

Differentiating with respect to m_s, setting equal to 0 and solving for m_s yields

$$m_s = \frac{\sqrt{w_s \alpha_s^{1-k} \beta_s^{1-k} / c_s}}{\sqrt{\lambda}} - \frac{1}{\beta_s}.$$

Therefore, to minimize the risk for fixed cost, we take

$$C = \sum_s m_s c_s = \sum_{s=1}^{S} \frac{\sqrt{w_s \alpha_s^{1-k} \beta_s^{1-k} c_s}}{\sqrt{\lambda}} - \sum_{s=1}^{S} \frac{c_s}{\beta_s}$$

or

$$\sqrt{\lambda} = \frac{\sum_s \sqrt{w_s \alpha_s^{1-k} \beta_s^{1-k} c_s}}{C + \sum_s c_s / \beta_s}.$$

Substituting this back into the equation for m_s yields

$$m_s = \frac{\sqrt{w_s \alpha_s^{1-k} \beta_s^{1-k} / c_s}}{\sum_s \sqrt{w_s \alpha_s^{1-k} \beta_s^{1-k} c_s}} \left(C + \sum_s \frac{c_s}{\beta_s} \right) - \frac{1}{\beta_s}.$$

Taking $k = 0$ or 1 gives the desired result.

References

Clevenson, M.L., and Zidek, J.V. (1975). Simultaneous estimation of the means in independent Poisson laws. *Journal of the American Statistical Association*, 70(351), 698-705.

Cochran, W.G. (1977). *Sampling Techniques* (Third ed.). New York: John Wiley & Sons, Inc.

Ericson, W.A. (1969). Subjective Bayesian models in sampling finite populations. In *New Developments in Survey Sampling*, (Eds. N.L. Johnson and H. Smith Jr.). New York: John Wiley & Sons, Inc. 326-357.

Ericson, W.A. (1988). Bayesian inference in finite populations. In *Handbook of Statistics*, (Eds. P. Krishnaiah and C. Rao). Amsterdam: Elsevier Science. 6, 213-246.

Ghosh, M., and Meeden, G. (1997). *Bayesian Methods for Finite Population Sampling*. Monographs on Statistics and Applied Probability. New York: Chapman & Hall. 79.

Godambe, V.P. (1955). A unified theory for sampling from finite populations. *Journal of the Royal Statistical Society Series B*, 17, 267-278.

Leite, J.G., Rodrigues, J. and Milan, L.A. (2000). A Bayesian analysis for estimating the number of species in a population using a nonhomogenoeous Poisson process. *Statistics & Probability Letters*, 48, 153-161.

Lindley, D.V., and Deely, J.J. (1993). Optimal allocation in stratified sampling with partial information. *Test*, 2(1), 147-160.

Scott, A., and Smith, T.M.F. (1971). Bayes estimates for subclasses in stratified sampling. *Journal of the American Statistical Association*, 66(336), 834-836.

Tiwari, R.C., and Lahiri, P. (1989). On robust empirical Bayes analysis of means and variances from stratified samples. *Communication in Statistics, Part A – Theory and Methods*, 18(3), 913-926.