J. R. Statist. Soc. B (1985)
47, *No.* 3, *pp.* 519–527

Bayesian Estimation of Finite Population Parameters in Categorical Data Models Incorporating Order Restrictions

By J. SEDRANSK†, J. MONAHAN and H. Y. CHIU

State University of New York at Albany, USA · *North Carolina State University, USA* · *Rolm Corporation, USA*

[Received June 1983. Final revision July 1985]

SUMMARY

This note describes a Bayesian method for estimation of finite population parameters in general population surveys where acceptable regression-type models are typically unavailable. A categorical data model is adopted as in Ericson (1969, Section 4). However, specifications of smoothness are incorporated into the prior distribution. These smoothness conditions are expressed as unimodal or, possibly, multi-modal order relations among the category probabilities. Emphasis is placed on posterior inference about the finite population mean. Of independent interest is the methodology for evaluating the posterior moments and probabilities using Monte Carlo integration with importance sampling.

Keywords: DIRICHLET DISTRIBUTION; MONTE CARLO INTEGRATION; IMPORTANCE SAMPLING; ISOTONIC REGRESSION

1. INTRODUCTION

For sample surveys of establishments (e.g., banks, colleges, retail stores), there are often acceptable models relating the variable of interest, Y, to one or more covariates. Also, many establishment surveys are repeated over time at regular or irregular intervals. In such situations there is a sound basis for Bayesian inference for finite population parameters. In particular, prior distributions for the parameters can be selected using the data available from earlier surveys. (See Sedransk, 1977, for an illustration.) However, the situation is not so felicitous for some general population surveys where, for example, the sampling unit is a household. In such surveys, analysts often are reluctant or unwilling to accept the regression-type models that are more readily adopted in establishment surveys. Thus, a categorical data approach appears to be indicated.

Ericson (1969, Section 4) postulates that for each unit in the finite population the random variable of interest, Y, assumes one of a finite set of values, $y = \{y_1, \ldots, y_t\}$ where $y_1 < \ldots < y_t$. For unit i, the probability that $Y = y_j$ is denoted by p_j ($\Sigma\, p_j = 1, j = 1, \ldots, t$). A Dirichlet prior distribution is assigned to $p = (p_1, \ldots, p_{t-1})$. Binder (1982) has extended Ericson's results by adopting Dirichlet process priors. However, neither the Dirichlet nor the Dirichlet process prior distribution can capture known smoothness relationships among $(p_1, \ldots, p_t)$. For example, in a survey of households, income is typically categorized. Letting y_j denote the central measure of income assigned to the jth category, and p_j the proportion of units with income falling in the jth category, there may be a known relationship among $(p_1, \ldots, p_t)$ such as

$$p_1 \leqslant \ldots \leqslant p_k \geqslant p_{k+1} \geqslant \ldots \geqslant p_t.$$

In this note, Ericson's (1969) categorized data model is assumed. However, we add to the prior specification known smoothness relationships among $(p_1, \ldots, p_t)$. For specificity, we assume $p_1 \leqslant \ldots \leqslant p_k \geqslant p_{k+1} \geqslant \ldots \geqslant p_t$, although multi-modal specifications can be handled without much additional difficulty (see Chiu, 1982).

† *Present address*: Professor J. Sedransk, Dept of Math. & Stat., SUNY-Albany, 1400 Washington Ave., Albany, N.Y. 12222, USA.

 0035–9246/85/47519 $2.00

First, if the modal value, k, is known, then the prior distribution for p is assumed to be given by

$$f_1'(p) = C_k(\beta_1, \ldots, \beta_t) \cdot \prod_{i=1}^{t} p_i^{\beta_i - 1} \quad \text{for} \quad p \in R_t^{(k)} \qquad (1)$$
$$= 0 \qquad \text{otherwise,}$$

where $\beta_i > 0$ for $i = 1, \ldots, t$,

$$R_t^{(k)} = \{p : p_1 \leqslant \ldots \leqslant p_k \geqslant \ldots \geqslant p_t;\ 0 \leqslant p_i \leqslant 1,\ \sum_{j=1}^{t} p_j = 1\},$$

and

$$\{C_k(\beta_1, \ldots, \beta_t)\}^{-1} = \int_{R_t^{(k)}} \prod_{i=1}^{t} p_i^{\beta_i - 1}\, dp_i.$$

Since $R_t^{(k)}$ has dimension $t-1$, the density (1) and integrals are with respect to Lebesgue measure on $R_t^{(k)}$.

If the value of k is uncertain one may first assign a prior distribution, $\{g'(k): k = 1, \ldots, t\}$, to the possible values of k. Then, given k, take

$$f'(p \mid k) = C_k(\beta_1^{(k)}, \ldots, \beta_t^{(k)}) \prod_{i=1}^{t} p_i^{\beta_i^{(k)} - 1} \quad \text{for} \quad p \in R_t^{(k)} \qquad (2)$$
$$= 0 \qquad \text{otherwise.}$$

Thus, the prior distribution for p is given by

$$f'(p) = \sum_{k=1}^{t} f'(p \mid k)\, g'(k). \qquad (3)$$

2. POSTERIOR INFERENCE, RANDOM SAMPLING

To simplify the presentation we assume that a random sample of n units has been selected with replacement, and denote by n_i the number of sample units having $Y = y_i$. First assuming that k is known, the posterior distribution of p is given by

$$f''(p \mid n) = \{C_k(n_1 + \beta_1, \ldots, n_t + \beta_t)\} \prod_{i=1}^{t} p_i^{n_i + \beta_i - 1} \quad \text{for} \quad p \in R_t^{(k)} \qquad (4)$$
$$= 0 \qquad \text{otherwise,}$$

where $n = (n_1, \ldots, n_t)$.

One may easily obtain the posterior mode of $(p_1, \ldots, p_t)$ by using, for the unimodal case as in (4), the algorithm in Cryer and Robertson (1975), and, for the multi-modal case, the algorithm in Chiu (1982, Chapter 1). Assuming (4), the modal value $\hat{p}^* = \{\hat{p}_1^*, \ldots, \hat{p}_t^*\}$ is the isotonic regression of $\{\hat{p}_i = (n_i + \beta_i - 1)/\Sigma\,(n_j + \beta_j - 1): i, j = 1, \ldots, t\}$ with weights $W_j = 1$ for $j = 1, \ldots, t$. The procedure is to evaluate

$$A(s_1, s_2) = \sum_{j=s_1}^{s_2} \hat{p}_j W_j \Big/ \sum_{j=s_1}^{s_2} W_j \text{ for } \{(s_1, s_2): 1 \leqslant s_1 \leqslant k, k \leqslant s_2 \leqslant t\}.$$

The value of (s_1, s_2) yielding the maximal value of $A(s_1, s_2)$ is denoted by (b, d), and $\hat{p}_b = \hat{p}_{b+1} = \ldots = \hat{p}_d = A(b, d)$. Formula (5) gives the complete solution:

$$\hat{p}_i^* = \begin{cases} \max\limits_{1 \leqslant s_1 \leqslant k} \; \max\limits_{k \leqslant s_2 \leqslant t} \; A(s_1, s_2) \equiv A(b, d) & \text{for } i = b, \ldots, d \\ \max\limits_{1 \leqslant s_1 \leqslant i} \; \min\limits_{i \leqslant s_2 \leqslant b-1} \; A(s_1, s_2) & \text{for } i = 1, \ldots, b-1 \\ \min\limits_{d+1 \leqslant s_1 \leqslant i} \; \max\limits_{i \leqslant s_2 \leqslant t} \; A(s_1, s_2) & \text{for } i = d+1, \ldots, t \end{cases} \tag{5}$$

The principal objective is posterior inference about the finite population mean, $\mu \equiv \mu(p)$, where

$$\mu \equiv \mu(p) = \sum_{i=1}^{t} y_i p_i.$$

One easily calculated point estimator of μ is $\hat{\mu} = \Sigma\, y_i \hat{p}_i^*$, $i = 1, \ldots, t$. Alternatively, posterior moments or probabilities (values of the distribution function) may be desired. Since

$$E''\{h(p) \mid n\} = \int_{R_t^{(k)}} h(p) f''(p \mid n)\, dp, \tag{6}$$

for the mth posterior moment take $h(p) = \{\mu(p)\}^m$ and for the probabilities take $h(p) = I\{\mu(p) \leqslant z\}$ where $I(\cdot)$ is the indicator function. Taking $h(p) = 1$ permits calculation of normalization constants.

Because of the high dimensionality, Monte Carlo integration is the method of choice for evaluating $E''\{h(p) \mid n\}$. The straightforward method is to repeatedly (and independently) sample p from the Dirichlet distribution with parameter $n + \beta$ (by transforming gamma $(n_i + \beta_i, 1)$ variables), test whether $p \in R_t^{(k)}$, and then use as the estimate of $E''\{h(p) \mid n\}$ the sample mean of the accepted p; i.e., for $\{p^{(j)} \in R_t^{(k)}: j = 1, \ldots, M\}$

$$\hat{E}''\{h(p) \mid n\} = M^{-1} \sum_{j=1}^{M} h(p^{(j)}).$$

However, the (posterior) probability that $p \in R_t^{(k)}$ can be so low that this method is inefficient (see Section 3).

A more efficient approach is to use importance sampling. That is, generate $p^{(j)}$ from the density $s(p)$ which has the same support $R_t^{(k)}$ as $f''(p \mid n)$, and then weight each "observation" $h(p^{(j)})$ by the ratio $w(p^{(j)}) = f''(p^{(j)} \mid n)/s(p^{(j)})$. The estimate of $E''\{h(p) \mid n\}$ is now

$$\hat{\hat{E}}''\{h(p) \mid n\} = M^{-1} \sum_{j=1}^{M} h(p^{(j)})\, w(p^{(j)}).$$

Note that the means of $\hat{E}$ and $\hat{\hat{E}}$ are the same, but their variances are different. Details about the computing methods (including the choice of $s(p)$) are given in the Appendix.

When the modal value, k, is uncertain the same technique may be used to evaluate the posterior moment, $E''\{h(p) \mid n\}$, where $n = (n_1, \ldots, n_t)$:

$$E''\{h(p) \mid n\} = \sum_{k=1}^{t} E''\{h(p) \mid n, k\}\, g''(k \mid n),$$

where

$$g''(k \mid n) = \Pr(n \mid k)\, g'(k) \Big/ \sum_{k=1}^{t} \Pr(n \mid k)\, g'(k).$$

Clearly, $E''\{h(p) \mid n, k\}$ can be obtained as described for the case when k is known. Further,

$$\Pr(n \mid k) = n!\, C_k(\beta_1^{(k)}, \ldots, \beta_t^{(k)}) \Big/ \left(\prod_{i=1}^{t} n_i!\right) C_k(\beta_1^{(k)} + n_1, \ldots, \beta_t^{(k)} + n_t)$$

is a simple function of normalization constants. Such constants can be evaluated by taking $h(p) = 1$ in expressions analogous to (6).

3. EXAMPLES

In this section we first illustrate the effect on the posterior expected value and variance of μ of taking (1) as the prior distribution for p rather than the unrestricted

$$f_2'(p) = \Gamma\left(\sum_{j=1}^{t} \beta_j\right) \left\{\prod_{i=1}^{t} p_i^{\beta_i - 1} / \Gamma(\beta_i)\right\}, \; \{p\colon 0 \leqslant p_i \leqslant 1,\; \sum_{j=1}^{t} p_j = 1\}. \tag{7}$$

For the first set of illustrative examples, assume that $t = 5$, the modal value k is known to be 5, $\beta = (1, 1, 1, 1, 2)$ and $n = (2, 2, 2, 2, 4)$ while for the second set, $t = 5$, $k = 4$, $\beta = (2, 3, 4, 5, 3)$ and $n = (1, 1, 1, 1, 5)$. In Table 1 we present in parallel columns for each choice of y: the value of

TABLE 1

Comparisons of posterior expected value (a) and variance (b) corresponding to use and omission of order restrictions on p

y	$E_1''(\mu)$	R_E	$V_1''(\mu)$	R_V	SE	N
(a) $k = 5, \beta = (1, 1, 1, 1, 2)$ and $n = (2, 2, 2, 2, 4)$						
2, 4, 8, 16, 32	18.21	0.86	3.77	2.13	0.0032	38
(b) $k = 4, \beta = (2, 3, 4, 5, 3)$ and $n = (1, 1, 1, 1, 5)$						
12, 16, 25, 30, 50	30.30	1.02	3.20	2.25	0.0042	14
12, 16, 25, 30, 150	53.73	1.15	47.64	2.72	0.0618	14
12, 16, 25, 30, 450	124.02	1.24	499.97	2.89	0.6654	14

(a) $E_1''(\mu)$ and $E_2''(\mu)$ are the posterior expected values of μ corresponding, respectively, to assumption and omission of the order restriction on p; $R_E = E_2''(\mu)/E_1''(\mu)$.
(b) $V_1''(\mu)$ and $V_2''(\mu)$ are the posterior variances corresponding to $E_1''(\mu)$ and $E_2''(\mu)$; $R_V = V_2''(\mu)/V_1''(\mu)$.

Note: SE is the estimated standard error of $E_1''(\mu)$, and N is an estimate of the number of replications that would be required using rejective sampling to produce one acceptable sample.

$y, E_1''(\mu \mid n)$, $R_E = E_2''(\mu \mid n)/E_1''(\mu \mid n)$, $V_1''(\mu \mid n)$, $R_V = V_2''(\mu \mid n)/V_1''(\mu \mid n)$ and $SE\{E_1''(\mu \mid n)\}$. Here, E_1'' and V_1'' are the estimated posterior expected value and variance corresponding to (4) while E_2'' and V_2'' are the posterior expected value and variance when (7) is used as the prior distribution (i.e., without the order restriction). $SE\{E_1''(\mu \mid n)\}$, the estimated standard error of our estimate of $E_1''(\mu \mid n)$, reflects the precision of our estimates. Finally, the last column, labelled N, is our estimate of the average number of replications that would be required using rejective sampling (Section 2) to produce *one* acceptable sample; i.e., satisfying $p \in R_t^{(k)}$.

In Table 1, panel (a) presents an example where the data are consistent with the prior specification. While the values of the posterior means, $E_1''(\mu)$ and $E_2''(\mu)$, are similar, the posterior variance of μ associated with the presence of an order restriction on p is less than half of the posterior variance of μ when the order restriction is not imposed. Note also that if one used the rejective sampling method to evaluate the posterior moments, on the average it would have taken 38 samples to obtain *one* sample that satisfied the order restriction.

The three examples in panel (b) have data that are inconsistent with the prior specification; i.e. the prior specification is that $k = 4$ while the data suggest that $k = 5$. As expected, as y_5 is increased from 50 to 150 to 450, R_E increases: Specification that $k = 4$ implies that $E_1''(p_4) > E_2''(p_4)$ and $E_1''(p_5) < E_2''(p_5)$. The effect of these two inequalities becomes more pronounced as y_5 is increased relative to y_4. Analogously, the ratio of posterior variances increases as y_5 increases. In the three examples, the reduction in posterior variance due to imposing the order restriction is much larger than 50 per cent.

For the second example, the number of concerns in business in a U.S. state (or the District of Columbia) has been categorized into six classes. For each of 1970, 1975 and 1980 the percent of the states falling into each class is given in Table 2. We derived the prior specification from the 1970 distribution, and the sample data from the 1975 and 1980 distributions. Five data sets are considered: (a) $n = (3, 2, 2, 1, 1, 1)$, a sample of size 10 distributed in accord with the 1975 distribution and then rounded; (b) $n = (5, 4, 3, 1, 1, 1)$, a sample of size 15 distributed as described in (a); (c) $n = (3, 1, 3, 1, 1, 1)$, a sample of size 10 distributed in accord with the 1980 distribution; (d) $n = (5, 2, 4, 2, 1, 1)$, a sample of size 15 distributed as described in (c); and (e) $n = (1, 3, 3, 1, 1, 1)$ a data set inconsistent with the prior specifications.

TABLE 2

Percent of states with specified numbers of business concerns

No. of concerns (in 000's)	1970	1975	1980
0–20	33.3 (17)†	31.4	33.3
20–40	27.5 (14)	23.5	13.7
40–60	19.6 (10)	19.6	27.5
60–100	7.8 (4)	9.8	11.8
100–150	5.9 (3)	9.8	7.8
150–300	5.9 (3)	5.9	5.9

Source: *State and Metropolitan Area Data Book*, 1982 (p. 519), D.C.: U.S. Government Printing Office.

† Number of states (plus District of Columbia) having specified numbers of concerns.

For illustration, the prior distribution for k is taken to be $g'(1) = 0.6$, $g'(2) = 3$ and $g'(3) = 0.1$. Defining $\beta^{(k)} = (\beta_1^{(k)}, \ldots, \beta_t^{(k)})$, we took $\beta^{(1)} = (2.7, 2.4, 2.0, 1.4, 1.3, 1.3)$, $\beta^{(2)} = (2.5, 2.6, 2.0, 1.4, 1.3, 1.3)$ and $\beta^{(3)} = (2.3, 2.4, 2.4, 1.4, 1.3, 1.3)$. First, $\beta^{(1)}$ was obtained by assuming a random sample of size 5 selected from the 1970 distribution (e.g. $\beta_1^{(1)} = 0.33(5) + 1 \doteq 2.7$). We obtained $\beta^{(2)}$ by first smoothing the observed number of states having 0–20 and 20–40 concerns; i.e. we assigned 15 states to the former and 16 states to the latter class. Then, we proceeded as described for $\beta^{(1)}$ (e.g. $\beta_1^{(2)} = (15/51)(5) + 1 \doteq 2.5$). For $\beta^{(3)}$ we assigned 13, 14 and 14 states, respectively, to the classes 0–20, 20–40 and 40–60, and then proceeded as described for $\beta^{(1)}$

TABLE 3

Conditional on k and unconditional posterior expected values and variances of μ for prior distributions and data (a) from Table 2

	$E''(\mu \mid k)$			$V''(\mu \mid k)$			*Posterior Probs*				
n	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	$g''(1)$	$g''(2)$	$g''(3)$	$E''(\mu)$	$V''(\mu)$
3,2,2,1,1,1	49.18	53.70	60.40	63.13	69.29	81.83	.75	.21	.04	50.59	72.73
5,4,3,1,1,1	46.56	49.58	54.48	51.94	56.18	69.01	.80	.18	.02	47.26	55.36
3,1,3,1,1,1	49.83	54.74	61.25	64.44	69.43	82.07	.74	.19	.08	51.62	77.99
5,2,4,2,1,1	48.57	52.72	57.85	55.50	59.60	66.98	.83	.13	.04	49.49	61.39
1,3,3,1,1,1	50.88	56.52	62.24	63.39	71.12	81.89	.34	.51	.15	55.44	84.74

(a) $y = (10, 30, 50, 80, 125, 225)$; $\beta^{(1)} = (2.7, 2.4, 2.0, 1.4, 1.3, 1.3)$, $\beta^{(2)} = (2.5, 2.6, 2.0, 1.4, 1.3, 1.3)$, $\beta^{(3)} = (2.3, 2.4, 2.4, 1.4, 1.3, 1.3)$; $g'(1) = 0.6$, $g'(2) = 0.3$ and $g'(3) = 0.1$.

The conditional and unconditional posterior moments of μ are summarized in Table 3 where $y = (10, 30, 50, 80, 125, 225)$, the y_i's being the midpoints of the classes in Table 2. As expected, the posterior variance decreases as the sample size increases, and $E''(\mu \mid 1) < E''(\mu \mid 2) < E''(\mu \mid 3)$. Next, compare the data set most inconsistent with the prior specification, $n = (1, 3, 3, 1, 1, 1)$, with $n = (3, 2, 2, 1, 1, 1)$. The posterior expected values are similar, but the posterior variances are quite different (84.74 vs. 72.73) reflecting the greater uncertainty associated with $n = (1, 3, 3, 1, 1, 1)$. Finally, for $n = (1, 3, 3, 1, 1, 1)$ the posterior probabilities $\{g''(k)\}$ change markedly from the prior probabilities, $\{g'(k)\}$.

REFERENCES

Binder, D. (1982) Non-parametric Bayesian models for samples from finite populations. *J. R. Statist. Soc.* B, **44**, 388–393.
Cheng, R. and Feast, G. (1980) Gamma variate generators with increased shape parameter range. *Commun. Ass. Comp. Mach.*, **23**, 389–394.
Chiu, H. Y. (1982) A new Bayesian approach to estimating finite population parameters. Ph.D. Thesis, State University of New York at Albany.
Cryer, J. and Robertson, T. (1975) Isotonic estimation of the probability of extinction of a branching process. *J. Amer. Statist. Ass.*, **70**, 905–911.
Ericson, W. A. (1969) Subjective Bayesian models in sampling finite populations (with Discussion). *J. R. Statist. Soc.* B, **31**, 195–224.
Lewis, P., Goodman, A. and Miller, J. (1969) A pseudorandom number generator for the system/360. *IBM Systems J.*, **8**, 136–146.
Schrage, L. (1979) A more portable Fortran random number generator. *ACM Transactions on Mathematical Software*, **5**, 132–138.
Sedransk, J. (1977) Sampling problems in the estimation of the money supply. *J. Amer. Statist. Ass.*, **72**, 516–522.

APPENDIX

Computing Methods

As described in Section 2, Monte Carlo integration using importance sampling was employed to compute posterior expectations of $h(p)$. The major hurdle is to find a distribution that has the desired support, yields a well behaved weighting function and permits easy generation of random variables. A poor choice may lead to estimates that have infinite variance.

To obtain $p \in R_t^{(k)}$ first eliminate the restriction

$$\sum_{i=1}^{t} p_i = 1$$

by moving to a non-singular distribution of

$$X \in S_t^{(k)} = \{X = (X_1, \ldots, X_t): X_i \geqslant 0, X_1 \leqslant \ldots \leqslant X_k \geqslant \ldots \geqslant X_t\}$$

and then using the transformation

$$p_i(X) = X_i / \sum_{j=1}^{t} X_j.$$

Using this transformation, $p: S_t^{(k)} \to R_t^{(k)}$, integrals having the same form as (b), i.e.,

$$E\{h(p)\} = \int_{R_t^{(k)}} h(p)\, C_k(\alpha) \prod_{i=1}^{t} p_i^{\alpha_i - 1}\, dp_i$$

can be rewritten as

$$E\{h(p)\} = \int_{S_t^{(k)}} h\{p(x)\}\, b(x)\, dx, \tag{A1}$$

where

$$b(x) = C_k^*(\alpha) \left(\prod_{i=1}^{t} x_i^{\alpha_i - 1} \right) \exp\left(- \sum_{i=1}^{t} x_i \right),$$

$$C_k^*(\alpha) = C_k(\alpha)/\Gamma(\Sigma\alpha_i).$$

Take $\alpha_i = \beta_i$ for integration with respect to the prior distribution and $\alpha_i = \beta_i + n_i$ for the posterior distribution.

We use importance sampling as described in Section 2: Using the algorithm given below, generate random vectors $\{X^{(j)}: j = 1, \ldots, M\}$ from the density, $s(x)$, on $S_t^{(k)}$ rather than from $b(x)$ in (A1). Here,

$$s(x) = t(k-1)!\,(t-k)! \prod_{i=1}^{t} s^*(x_i)$$

where

$$s^*(x) = \gamma^{\alpha^*} x^{\alpha^* - 1} e^{-\gamma x}/\Gamma(\alpha^*).$$

Then, $E\{h(p)\}$ is estimated by

$$M^{-1} \sum_{j=1}^{M} h\{p(X^{(j)})\}\, w(X^{(j)}) \tag{A2}$$

where

$$\begin{aligned} w(x) &= b(x)/s(x) \\ &= \frac{C_k^*(\alpha)\{\Gamma(\alpha^*)\}^t}{t(k-1)!\,(t-k)!\,\gamma^{t\alpha^*}} \prod_{i=1}^{t} x_i^{(\alpha_i - \alpha^*)} \exp\{-(1-\gamma)\,\Sigma\, x_i\}. \end{aligned} \tag{A3}$$

Efficient choices of $\alpha^* = \min\{\alpha_i\}$ and $\gamma = t\alpha^*/\Sigma\alpha_i$ yield a weight function $w(x)$ that is bounded above, so that the variance of the estimate in (A2) will become reasonably small for practical values of the replication size M.

Since the normalization constant $C_k^*(\alpha)$ is unknown, the weight function $w(x)$ in (A3) is computed without this term, and $E\{h(p)\}$ is estimated by the ratio of (A2) to a denominator of the same form with $h \equiv 1$.

The following algorithm generates random vectors X having the joint density s, with support on the set $S_k^{(t)}$, using i.i.d. random variables with density s^*:

(1) Generate X_i, $i = 1, \ldots, t$, i.i.d with density s^*.
(2) Find i^* so that $X_{i^*} \geqslant X_i$, $i = 1, \ldots, t$.
(3) Interchange X_{i^*} and X_k.
(4) Sort $X_1, \ldots, X_k$ in increasing order.
(5) Sort $X_k, \ldots, X_t$ in decreasing order.

The examples in Section 3 were computed using $M = 900$ replications for integration with respect to the prior to obtain its normalization constant, and $M = 2{,}500$ replications for the posterior. Gamma variables were generated using algorithm GBH of Cheng and Feast (1980). Uniform pseudorandom variables were obtained from Schrage's (1979) implementation of the Lewis, Goodman and Miller (1969) method. Fortran code for the computation will be available (for the cost of transmission) from the second author for one year following publication.