UNIVERSIDADE DE SÃO PAULO

Bayesian solutions to a class of selections problems using weighted loss functions

JOSEMAR RODRIGUES

Nº 13

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

Bayesian solutions to a class of selections problems using weighted loss functions

JOSEMAR RODRIGUES

Nº 13

N O T A S D O I C M S C
Série Estatística

São Carlos
Out. / 1994

Bayesian Solutions to a Class of Selections Problems Using Weighted Loss Functions

Josemar Rodrigues
ICMSC-USP-13560-970-São Carlos- SP

Abstract

The purpose of this paper is to study a class of selection problems introduced by Marsh and Zellner (1994) using weighted loss functions. The histogram approach as described in Albert (1993) is used via Minitab to assess the prior information and to analyse the quality of predictions. The Rubin's Sampling-Importance-Resampling (SIR) can also be used to simulate the posterior distribution to solve the selection problems. In such problems, we need to decide how many offers to make to obtain a targed number of acceptances. These kind of problems are relevant in planning social gatherings and meetings or organizing an entering class of a graduate program. The selections problems were studied in details by Marsh and Zellner using squared and asymmetric loss functions.

1 Introduction

Marsh and Zellner (1994) discussed the following and interesting problem: Suppose that one wishes to make offers to individuals and are uncertain about how many offers will be accepted or to predict the number of offers given a target number. For example, the dean will be interested in predicting the number of students who enroll in his program given the number of offers or to make a number of offers given his target number of students. There are many different possible applications of selections problems and we suggest the reader to see the introduction of the Marsh and Zellner's paper for more details. In this paper we shall show how such problems can be solved using differents loss functions from the Bayesian decision point of view.

The paper is organized as follows: In Section 2 we obtain the optimal solution of the selection problems which consists in minimizing the prior expectation of loss functions. In this section, we consider different loss functions and try to solve the selection problems when individuals' probabilities of acceptance differ, that is, when heterogeneity is present. A numerical illustration of the optimal solutions using the histogram approach introduced by Berger (1985) and how data can be employed using the Rubin's Sampling-Importance- Resampling (SIR) are presented in Section 3.

2 Selection Problems under different loss functions

In this section we consider particular selection problems based on different loss functions and minimization of expected loss.

2.1 Selection Problems: Loss Functions and Optimal Predictions

To formulate the selection problems and to derive the optimal prediction under different loss functions, we introduce a binary variable X_i, where $X_i = 1$ denotes that the i'th individual enrolls given an offer with probability p and X_i denotes non-enrollment given an offer with probability $1-p$. Then if n offers are made, the random number enrolling, denoted by N, is given by

$$N = \sum_{i=1}^{n} X_i. \tag{1}$$

Usually in selection problems, the value of p is not known and in this situation we can assign a prior mean for p, $Ep = \bar{p}$ and a prior variance $Var(p) = \sigma^2$. Given n offers, Marsh and Zellner introduced a symmetric squared loss function:

$$L_q[\hat{N}, N] = (N - \hat{N})^2, \tag{2}$$

where $\hat{N}$ is some point prediction. Then the expected loss, assuming that the $X_i's$ are uncorrelated is

$$EL_q[\hat{N}, N] = \bar{p}(1-\bar{p}) + n(n-1)\sigma^2 + (\hat{N} - n\bar{p})^2. \qquad (3)$$

Note that if $\sigma^2 = 0$ in the last expression, it reduces to the expression (2.5) in Marsh and Zellner's paper based on the assumption that p is known. From (3) for any given n, it is trivially that the optimal solution is $\hat{N} = \hat{N}_q = n\bar{p}$ and the minimal expected loss given by

$$v = Var[N] = n\bar{p}(1-\bar{p}) + n(n-1)\sigma^2. \qquad (4)$$

The minimal expected loss (4) is substantially inflated by the prior variance. Also, note that this optimal solution is the dean's rule discussed by Marsh and Zellner when taking $\sigma^2 = 0$, that is, when p is known.

A particulary weighted loss function for the present problem is

$$L_w[N, \hat{N}] = \frac{(N - n\hat{p})^2}{n\hat{p}(1-\hat{p})}, \qquad (5)$$

where $\hat{p}$ is some estimate. Then prior expected loss is given by:

$$\begin{aligned} EL_w[N, \hat{N}] &= \frac{Var[N] + [E[N] - n\hat{p}]^2}{n\hat{p}(1-\hat{p})} = \\ &= \lambda + \frac{v + n^2(\bar{p} - \hat{p})^2 - n\lambda\hat{p}(1-\hat{p})}{n\hat{p}(1-\hat{p})} = \\ &= \lambda + \frac{(n+\lambda)}{\hat{p}(1-\hat{p})}[\hat{p} - \frac{n\bar{p} + \lambda/2}{n+\lambda}], \end{aligned}$$

where λ is the smallest positive solution of

$$\frac{1}{4}\lambda^2 + (n\bar{p} - \frac{v}{n} - n\bar{p}^2)\lambda - v = 0. \qquad (6)$$

The optimal point prediction relative to the weighted loss function in (5) is given by

$$\hat{N}_w = n\hat{p}_w = n\frac{n\bar{p} + \lambda\frac{1}{2}}{n+\lambda}, \qquad (7)$$

where the minimal prior expected loss is given by λ, the solution of equation (6). Note that $\hat{p}_w$ in (7) is a linear combination of the prior mean and 1/2.

Following Marsh and Zellner (1994), we consider $\sigma^2 = 0$ or $p = \bar{p}$ and a possible asymmetry in our selection problem to determine how the dean's rule gets modified. We employ Varian's (1975) asymmetric linex loss function given by

$$L[\Delta] = b[e^{a\Delta} - a\Delta - 1], \tag{8}$$

where a and b are given parameters with $b > 0$ and $a \neq 0$. If $a > 0$, and $\Delta > 0$, loss grows exponentially as Δ grows whereas when $\Delta < 0$, loss grows approximately linearly. The expected loss is given by

$$EL[\Delta] = b[E[e^{a\Delta}] - aE[\Delta] - 1] \tag{9}$$

Taking $\Delta = N - \hat{N}$ (see Marsh and Zellner) and minimizing the expected loss with respect to $\hat{N}$, the optimal prediction is

$$\hat{N}_{MZ} = npA_{MZ}, \tag{10}$$

where

$$A_{MZ} = \frac{1}{ap}\ln\{\frac{1}{pe^{-a} + (1-p)}\}. \tag{11}$$

However, if we take $\Delta = \frac{\hat{N}-N}{np(1-p)}$ we get from (10) that the optimal prediction is

$$\hat{N}_R = npA_R \tag{12}$$

where

$$A_R = \frac{n(1-p)}{a}\ln\{\frac{1}{pe^{-\frac{a}{np(1-p)}} + 1 - p}\} \tag{13}$$

Table 1 shows the values of factor A_{MZ} and A_R for different values of a , $p = 1/2$ and $n = 200$. Note that it is important to take account of asymmetry of loss functions in analyzing selection problems when using Marsh and Zellner's procedure. However, we obtain a very stable solution with respect to the weighted Linex function, that is, we do not to worry about the asymmetry of our loss function (8).

Table 1: A_{MZ} and A_R for $p = 1/2$,$n = 200$ and various values of a.

a	A_R	A_{MZ}
0.1	0.99	0.98
0.3	0.99	0.92
0.5	0.99	0.88
1.0	0.99	0.33
4.0	0.96	0.08
10	0.90	0.03

Shown in Table 2 are the optimal estimates for the quadratic loss function, weighted loss function, LINEX loss function with $\Delta = \hat{N} - N$ and $\Delta = \frac{\hat{N}-N}{np(1-p)}$, respectively.

Table 2- Optimal estimates for different loss functions

	L_q	L_w	$\Delta = \hat{N} - N$	$\Delta = \frac{\hat{N}-N}{np(1-p)}$
Optimal estimate	$\hat{N}_q = n\bar{p}$	$\hat{N}_w = n\frac{n\bar{p}+\lambda 1/2}{n+\lambda}$	$\hat{N}_{MZ} = npA_{MZ}$	$\hat{N}_R = npA_R$
Prior expected loss	v	λ	-	-

The dependence of $\hat{N}$ on p and a is shown in Figure 1.

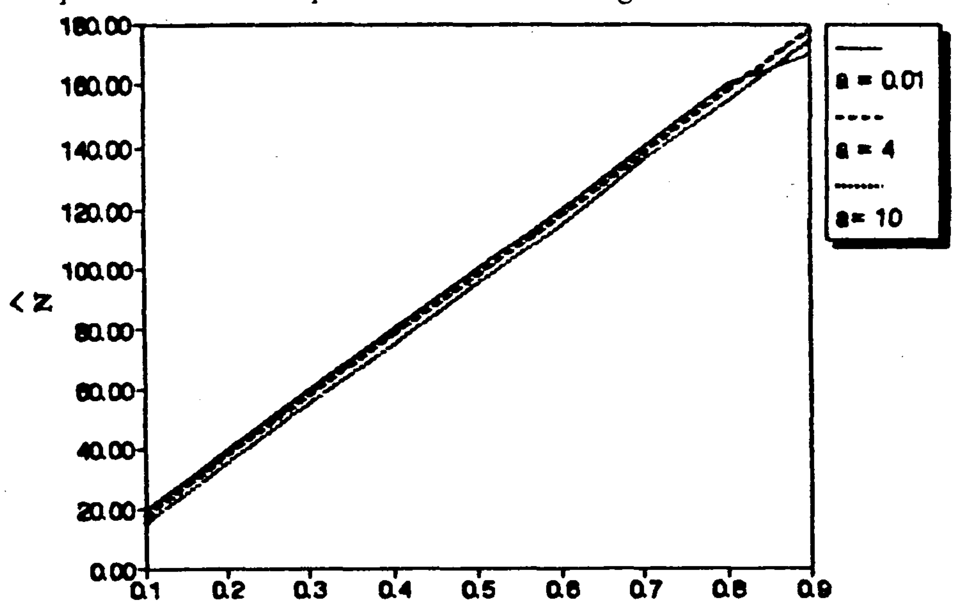

Figure 1: The optimal estimates versus p for various values of a.

2.2 Control Problems

In this section we consider a different problem : Suppose we are interested in predicting the number of offers n, called a control problem by Marsh and Zellner. Given a target number of students, denoted by N_0, they considered the following symmetric loss function:

$$L(N, N_0) = (N - N_0)^2 \tag{14}$$

with $N = \sum_{i=1}^{n} X_i$ as before. They showed that the value of n, number of offers, that minimizes the expected loss in (14) is

$$\begin{aligned} \hat{n}_q &= \frac{N_0}{\bar{p}} B_{MZ} - C_{MZ}, \\ B_{MZ} &= \frac{1}{1 + \sigma^2/\bar{p}^2}, \\ C_{MZ} &= \frac{1}{2}\frac{1-\bar{p}}{\bar{p}}[\frac{1 - \sigma^2/\bar{p}(1-\bar{p})}{1+\sigma^2/\bar{p}^2}], \end{aligned} \tag{15}$$

with $Var(p) = \sigma^2$ and $E[p] = \bar{p}$. We now introduce an alternative weighted loss function given by

$$\begin{aligned} L_W[N, N_0] &= \frac{(N - N_0)^2}{Var[N]} \\ &= \frac{(N - N_0)^2}{an^2 + bn} \end{aligned} \tag{16}$$

with $a = \sigma^2$ and $b = \bar{p}(1 - \bar{p}) - \sigma^2$ Then using the weighted loss function in (16) and the same procedure used in (5) , we have

$$E[L_W[N, N_0] = \lambda + \frac{Ep^2 - \lambda a}{an^2 + bn}[n - \hat{n}_w]^2 \tag{17}$$

with

$$\hat{n}_w = \frac{N_0}{\bar{p}} \frac{1}{\sqrt{\frac{\sigma^2}{\bar{p}^2}(1 - \lambda)}} \tag{18}$$

and λ is the smallest positive solution of

$$b^2\lambda^2 - 2[(b - 2\bar{p}N_0)b - 2aN_0^2]\lambda + (b - 2\bar{p}N_0)^2 - 4N^2Ep^2 = 0. \tag{19}$$

Note that with $\sigma^2 = 0$ or $\sigma^2 = \overline{p}(1-\overline{p})$ we have the dean's solution discussed in details by Marsh and Zellner.

The control problem can also be solved using the Linex function in (8) with $\Delta = \frac{N}{N_0} - 1$. Considering $\sigma^2 = 0$ and applying the Marsh and Zellner's optimal solution we have that the value of n minimizing the expected loss under the Linex function is

$$\hat{n}_1 = \frac{N_0}{p}\left[\frac{ap}{N_0 ln(A_2)} + \frac{1}{N_0 ln(A_2)} ln\left(\frac{ap}{N_0 ln(A_2)}\right)\right] \tag{20}$$

with

$$A_1 = pe^a + (1-p) \quad \text{and} \quad A_2 = pe^{\frac{a}{N}} + (1-p). \tag{21}$$

2.3 Selection Problems: Heterogeneity case

In this section we allow the probability of acceptance , p, to be different for different individuals or we consider now the heterogeneity situation. In this section, we will consider in details ony the prediction problem. As in Section 2.1, we introduce a independent random variable X_i, $i = 1, \ldots, n$ such that $X_i = 1$ with probability p_i and $X_i = 0$ with probability $1 - p_i$, $i = 1, \ldots, n$. Then the random number of accepting offers is

$$N = \sum_{i=1}^{n} X_i \tag{22}$$

with the following hierarchical model:

$$\begin{array}{lll} (i) & E[N \mid \mathrm{p}] = np & Var[N \mid \mathrm{p}] = np(1-p) - \sigma_p^2 \\ (ii) & E[p] = \overline{p} \ \text{(known)} & Var[p] = \sigma^2 \ \text{(known)} \end{array} \tag{23}$$

with

$$\begin{aligned} \mathrm{p} &= (p_1, p_2, \ldots, p_n), \\ \sigma_p^2 &= \frac{\sum_{i=1}^{n}(p_i - p)^2}{n} \quad \text{(known)} \quad \text{and} \\ p &= \frac{\sum_{i=1}^{n} p_i}{n}. \end{aligned} \tag{24}$$

Then if we use the the weighted loss function

$$L_w[N,\hat{N}] = \frac{(N-\hat{N})^2}{n\hat{p}(1-\hat{p}) - n\sigma_p^2} \tag{25}$$

with $\hat{N} = n\hat{p}$, the optimal predictor is given by

$$\hat{N}_w = n\frac{n\overline{p} + \lambda/2}{n+\lambda}, \tag{26}$$

where the expected loss associated with $\hat{N}_w$ is the smallest positive solution λ of the following equation:

$$(1/4 - \frac{\sigma_p^2}{n})\lambda^2 + (n\overline{p} - v/n - \overline{p}^2)\lambda - (v - n\sigma_p^2) = 0. \tag{27}$$

Note that if $\sigma_p^2 = 0$ in these last results, we have the homogeneity case considered before.

Now, for the heterogeneity case, we introduce a symmetric loss function

$$L(N,\hat{N}) = (N-\hat{N})^2 \tag{28}$$

Then the expected loss is given by:

$$\begin{aligned} EL(N,\hat{N}) &= Var(N) + (\hat{N} - E(N))^2 \\ &= n\overline{p}(1-\overline{p}) - n[\sigma_p^2 - (n-1)\sigma^2] + (\hat{n} - n\overline{p})^2. \end{aligned} \tag{29}$$

We note that the optimal point predictor is $\hat{N} = \hat{N}_{MZ} = n\overline{p}$ and the minimal expected loss is $v^* = \overline{p}(1-\overline{p}) - n[\sigma_p^2 - (n-1)\sigma^2]$. Thus in this case, the presence of heterogeneity and uncertainty about p have deflated the minimal expected loss. When $\sigma^2 = 0$, Marsh and Zellner (1994) compared (29) numerically to the deflation in which $\sigma_p^2 = 0$.

2.4 Weighted Balanced Loss Function: Control Problems

Marsh and Zellner (1994) expanded the control loss function to take account of budget considerations in the following way: Assume that each admitted student has an associated net cost c and that a budgeted amount is available, denoted by B. Let N_B the target number of students compatible wih the condition $N_{BC} = B$. Considering a balanced loss function as

$$L_B = w(N - N_0)^2 + (1 - w)(N - N_B)^2, \tag{30}$$

where $0 \leq w \leq 1$ and minimizing EL_B with respect to n with p known, they showed that the optimal solution is

$$\begin{aligned} n_B^\star &= \frac{N_W}{p}[1 - \frac{1-p}{2N_W}] \quad \text{and,} \\ N_W &= wN_0 + (1-w)N_B. \end{aligned} \tag{31}$$

In this section, we introduce the following weighted balanced function:

$$L_W = w\frac{(N - N_0)^2}{np(1-p)} + (1 - w)\frac{(N - N_B)^2}{np(1-p)} \tag{32}$$

Using the same procedure as in (17), the solution that minimizes EL_W with respect to n is given by

$$n_W^\star = \frac{N_W}{p}\sqrt{1 + \frac{\sigma_W^2}{N_W^2}}, \tag{33}$$

with $\sigma_W^2 = N_W^{(2)} - N_W^2$ and $N_W^{(2)} = wN_0^2 + (1-w)N_B^2$. The minimal EL_W is

$$\lambda = 1 - \frac{2N_W(1 - \sqrt{1 + \sigma_W^2/N_W^2})}{1 - p}. \tag{34}$$

Note that the best strategy is $\sigma_W^2 = 0$ which implies that $N_0 = N_B$ and the optimal solution is the dean's solution given by

$$n_W^\star = \frac{N_0}{p}.$$

Further, it is direct to solve this problem when p is unknown with known moments , $Ep = \bar{p}$ and $Var(p) = \sigma^2$. Now we outline the above results in the following table:

Table 3: Optimal predictors under the balanced and weighted balanced loss functions: Control problems.

Loss functions	Optimal predictors	Expected loss
L_B	$n_B^\star = \frac{N_W}{p}[1 - \frac{1-p}{2N_W}]$	EL_B
L_W	$n_W^\star = \frac{N_W}{p}\sqrt{1 + \frac{\sigma_W^2}{N_W^2}}$	$\lambda = 1 - \frac{2N_W(1-\sqrt{1+\sigma_W^2/N_W^2})}{1-p}$

3 Making Predictions and offers via histogram approach (Berger,(1985))

We consider in this section two situations: The first one we suppose that we have only prior information to assess values of p and in the second case we try to use past data in making predictions and offers

3.1 Use of prior information in making prediction and offers via histogram approach

In this section we will use the histogram approach described by Berger (1985) to construct a prior distribution for p, the probability that an individual enrolls given an offer. After constructing this prior we can obtain the prior mean $\bar{p}$ and the prior variance σ^2. The Berger's procedure is the following: One breaks the space of possible values of p into subintervals and then subjectively specifies likelihoods of the different intervals as in Figure 2.

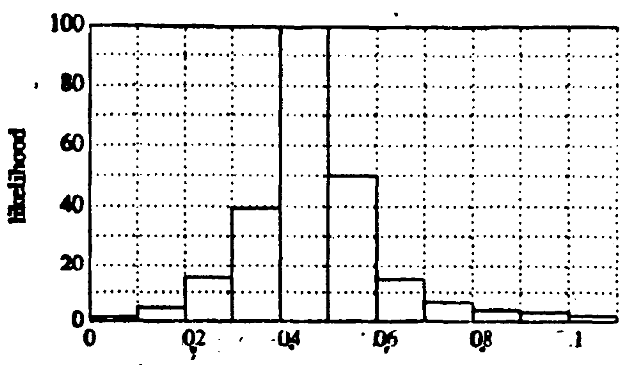

Figure 2: Prior histogram of p

A sample from the prior distribution is easily generated for the above prior histogram using Minitab commands. The discrete subcommands of 'random' chooses an interval at random in the histogram and the 'uniform' subcommand of 'random' randomly chooses a point inside the interval. The prior sample size is 500 which seems to be sufficiently large to provide an acccurate discription of the prior distibution and the prior moments. (see Albert (1993) for details of the Minitab macro used in this section).

The Minitab 'dotplot' command is used to give a graph of the simulated prior which is presented in Figure 3.

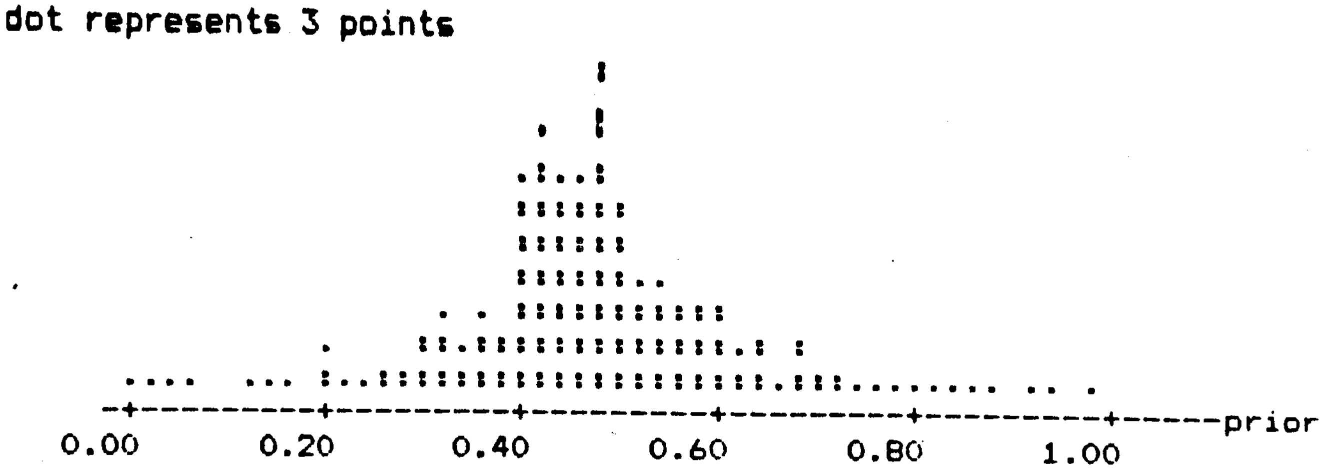

Figure 3: Prior sample for p

The mean prior is 0.467 and prior variance is 0.021, respectively, or, $\overline{p} = 0.467$ and $\sigma^2 = 0.021$.

Table 4: Selection Problems: Homogeneity case and $n = 200$.

	optimal predictions	prior expected loss
quadratic loss	93	900.315
weighted loss	94	18.078

Note that the optimal predictions do not change too much with the loss functions but the prior expected loss is very sensible to the choice of the loss functions.

Table 5: Selection Problems : Heterogeneity case

σ_p^2	λ	v^*	$\hat{N}_w$	$\hat{N}_{MZ}$
0.00	17.18	850.80	90.62	89.81
0.01	17.14	848.80	90.61	89.91
0.05	16.99	840.80	90.61	89.91
0.10	16.81	830.80	90.60	89.91
0.15	16.62	820.80	90.59	89.91
0.20	16.43	810.80	90.58	89.91
0.25	16.25	800.80	90.57	89.91

Table 5 shows that the deflation of the expected loss is appreciable for the quadratic and weighted loss functions. The optimal predictions are similar for both loss functions but we note an appreciable difference between the expected loss functions.

Table 6: Control Problems : Homogeneity case and $N_o = 100$

loss functions	Optimal predictor	Prior expected loss
quadratic loss	198	1021.92
weighted loss	217	0.95

We note from Table 6 , that the results are quite differents with respect to the squared loss and weighted loss functions.

3.2 Use of past data in making predictions and offers via the SIR algorithm

Marsh and Zellner suggested in their paper how past data can be used in conjunction with judgmental information in solving selection problems. In this section, we look to problem in a different way, that is , we try to use the Albert's idea which consist in simulating a posterior sample from the prior sample using the SIR algorithm. This algorithm can be implemented using MINITAB macros as shown in details by Albert (1993). In this situation all results obtained for selection or controls problems can be applied by replacing the prior moments by their corresponding posterior moments. If we assume that X_i, $i = 1, \ldots, n$ are generated by a binomial process with parameter p, the likelihood function is

$$L(p) \propto p^X(1-p)^{n-X} \tag{35}$$

with $X = \sum_{i=1}^{n} X_i$ representing past data. Suppose that we wish to obtain a sample of values from the posterior density $\pi(p) = K\pi(p)L(p)$, where $\pi(p)$ is the prior density and K a proportionality constant. Suppose we generated a sample from the prior density by using the histogram approach. Then we can apply SIR (see Albert, 1993) to the prior sample to obtain an approximate posterior sample. Let $\{p_1, p_2, \ldots, p_m\}$ be a sample from the prior $\pi(p)$. Then one obtains an approximate sample $\{p_1^\star, p_2^\star, \ldots, p_m^\star\}$ with replacement from $\{p_1, p_2, \ldots, p_m\}$ with unequal probability weights $\{L(p_1), L(p_2), \ldots, L(p_m)\}$. For our purpose we take $m = 500$ and suggest to the reader to see Albert's paper (1993) to see how to implement this algorithm using MINITAB commands.

The Minitab 'dotplot' command is used to give a graph of the simulated posterior which is presented in Figure 4.

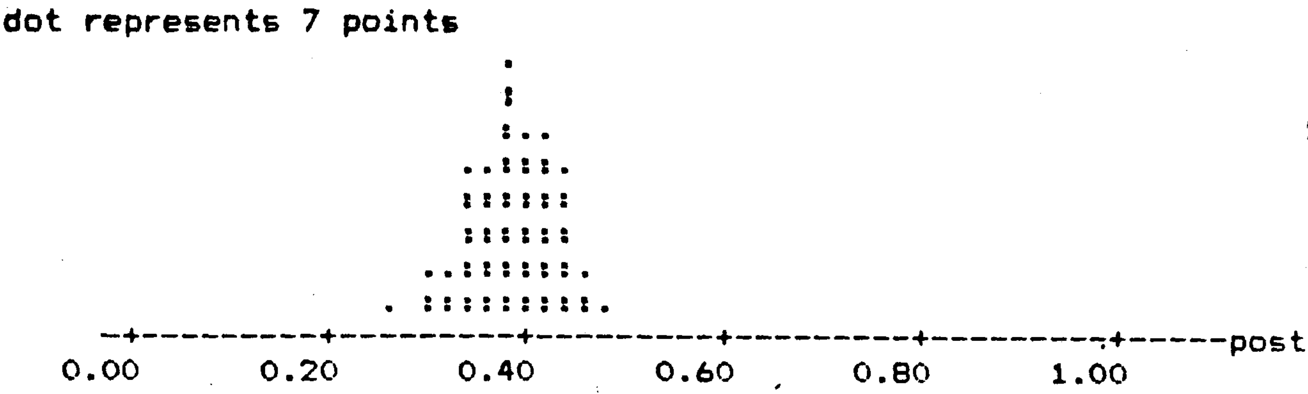

Figure 4: Posterior sample for p

Table 7: Selection problem with past data: Homegeneity case with $n = 200$. Past data: Number of offers=100 and $X = 60$. Posterior mean:0.39 and posterior variance=0.0024

loss functions	Optimal predictors	Posterior expected loss
quadratic loss	78.35	144.61
weighted loss	78.68	3.03

Comparing Table 4 and Table 7, we note a drastic reduction on the posterior expected loss and the optimal point estimates when including past data. However, the optimal point estimates in Table 7 are closed and the posterior expected loss very sensible when considering the quadratic loss function or the weighted loss function.

References

Albert, J.H. (1993). Teaching Bayesian Statistics Using Sampling Methods and MINITAB, The American Statistician, Vol. 47, No.3, 182-191.

Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis, New York: Spring-Verlag.

Marsh, L. and Zellner, A. (1994). Bayesian solutions to a class of selection problems, Technical Report, University of Chicago.

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA

012/94 ACHCAR, J.A.; DAMASCENO, V.L. - *Extreme value models: an useful reparametrization for the survival function*

011/94 ACHCAR, J.A.- *Approximate bayesian analysis for non-normal hierarchical classification model.*

010/94 RODRIGUES, J.; RODRIGUES, E.F. - *Bayesian estimationin the study of tampered Random variables.*

009/94 ACHCAR, J.A.; FOGO, J.C. - *An useful reparametrization for the reliability in the Weibull case.*

008/94 RODRIGUES, J.; LEITE, J.G. - *A note on Bayesian analysis in M/M/1 queues from confidence intervals.*

007/94 ACHCAR, J.A.; DAMACENO, V.L. - *An useful reparametrization for the survival function considering an exponential regression model.*

006/94 RODRIGUES, J. - *Bayesian estimation of a normal mean parameter using linex loss function and robustness considerations.*

005/93 ACHCAR, J.A.; FOGO, J.C.- *Some useful reparametrizations for the reliability function considering an exponential regession model.*

004/93 LEME, E.C. - *A m.v.u. estimator v.s. a m.l. estimator: a way of comparing their precisions by using the Pareto distribution*

003/93 ACHCAR, J.A.- *Some practical aspects of approximate bayesian inference*

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA

012/94 ACHCAR, J.A.; DAMASCENO, V.L. - *Extreme value models: an useful reparametrization for the survival function*

011/94 ACHCAR, J.A.- *Approximate bayesian analysis for non-normal hierarchical classification model.*

010/94 RODRIGUES, J.; RODRIGUES, E.F. - *Bayesian estimationin the study of tampered Random variables.*

009/94 ACHCAR, J.A.; FOGO, J.C. - *An useful reparametrization for the reliability in the Weibull case.*

008/94 RODRIGUES, J.; LEITE, J.G. - *A note on Bayesian analysis in M/M/1 queues from confidence intervals.*

007/94 ACHCAR, J.A.; DAMACENO, V.L. - *An useful reparametrization for the survival function considering an exponential regression model.*

006/94 RODRIGUES, J. - *Bayesian estimation of a normal mean parameter using linex loss function and robustness considerations.*

005/93 ACHCAR, J.A.; FOGO, J.C.- *Some useful reparametrizations for the reliability function considering an exponential regession model.*

004/93 LEME, E.C. - *A m.v.u. estimator v.s. a m.l. estimator: a way of comparing their precisions by using the Pareto distribution*

003/93 ACHCAR, J.A.- *Some practical aspects of approximate bayesian inference*