© Journal of Applied Mathematics & Decision Sciences, 2(2), 107–117 (1998)
Reprints Available directly from the Editor. Printed in New Zealand.

The Predictive Distribution in Decision Theory: A Case Study

GEOFF JONES g.jones@massey.ac.nz
Institute of Information Sciences and Technology
College of Sciences, Massey University, New Zealand

Abstract. In the classical decision theory framework, the loss is a function of the decision taken and the state of nature as represented by a parameter θ. Information about θ can be obtained via observation of a random variable X. In some situations however the loss will depend not directly on θ but on the observed value of another random variable Y whose distribution depends on θ. This adds an extra layer to the decision problem, and may lead to a wider choice of actions. In particular there are now two sample sizes to choose, for X and for Y, leading to a range of behaviours in the Bayes risk. We illustrate this with a problem arising from the cleanup of sites contaminated with radioactive waste. We also discuss some computational approaches.

Keywords: Decision Theory, Bayes Rule, Predictive Distribution, Monte Carlo Integration

1. Introduction

Consider the following consulting problem: the client is involved in the cleanup of sites contaminated with radioactive waste, which involves sending bins of radioactive material to a nuclear reprocessing plant. Two such plants are available, one of which is less expensive but which will only accept a bin if the level of radioactivity of the material is below a threshold level. The actual level of radioactivity in the bin is determined by sampling at the reprocessing plant; as far as the client is aware only one such sample is taken. If the measured level exceeds the threshold, the material is returned to the client and must then be sent to the second, more expensive reprocessing plant which will accept material at any level of contamination.

The client wishes to base the decision of which reprocessing plant to use on a sample or samples taken from each bin on site before the material is dispatched. The cost of sampling is small relative to the difference in reprocessing costs, but is not negligible. How many samples should be taken, and how should this information be used?

The Loss Function in this problem clearly has the form

$$\begin{aligned} L(a_1, y) &= \begin{cases} S & \text{if } y < c \\ S + K + P & \text{otherwise} \end{cases} \\ L(a_2, y) &= S + K \end{aligned}$$

Here a_1 is the decision to send to the less expensive reprocessing plant at a cost S, K is the extra cost of sending to the other plant, and P the extra penalty incurred by sending first to the less expensive plant but having the material rejected. This

occurs if the sample taken at the plant has an observed value y greater than the threshold level c. Since the value of S plays no part in the decision, we can without loss of generality take $S = 0$.

The client expects that there will be considerable variation in levels of radioactivity of the material within each bin, much of it being at quite a low level but with some "hotspots" of high radioactivity, suggesting a highly skewed distribution. It seems reasonable then to model the result for a single sample as a random variable X with an Exponential distribution. We write $X \sim \text{Exp}(\theta)$ to denote that X has density

$$f_{X|\theta}(x \mid \theta) = \theta e^{-\theta x}, \qquad x > 0 \tag{1}$$

Here θ parametrises the "state of nature" and relates to the average level of radioactivity of the material in the bin. Note however that the mean level is $1/\theta$. The alternative parametrisation of the Exponential distribution is more intuitive, but we keep the present form for mathematical convenience. Because of the high skewness it is intuitively clear that a single sample will not provide reliable information about the average radioactivity level. It may be advantageous for the client to persuade the reprocessor, through financial inducement or otherwise, to take further samples before accepting or rejecting a bin. Thus there may be two sample sizes to consider, relating to the sampling before and after dispatch. Intuitively one might expect that increasing either sample size would be advantageous to the client, and that for a given cost of sampling the total sample size might be shared equally between the two stages. This turns out not to be the case.

In Section 2 we develop and analyse a Bayesian framework for this problem using a conjugate prior distribution for θ. In Section 3 we consider the determination of optimal sample sizes at each stage of sampling. In Section 4 we reconsider the choice of prior and discuss some numerical strategies for incorporating a non-conjugate prior.

The basic principles of statistical decision theory, as used here, are described by DeGroot [2], although our notation is closer to that of Ferguson [3]. In the classical approach the loss incurred by the decision maker is a function of the action taken and the true value of an unknown parameter, information about which can be obtained by sampling. The situation in which the loss depends not on a parameter but on future observations was considered by Roberts [7] in the context of statistical prediction. Aitchison and Dunsmore [1] and Geisser [4] provide an overview and many applications of the predictive approach, some of which involve decision making but not the sample size determination problem considered here. A related problem in determining a single sample size in the classical framework when there are two "adversaries" with different priors, was considered by Lindley and Singpurwalla [5], and an application in environmental monitoring of radiation levels given by Wolfson et al. [8].

2. Bayesian Analysis

We assume that the uncertainty about θ can be expressed as a Gamma distribution, $\theta \sim \Gamma(\alpha, \lambda)$ with prior density

$$\pi_\theta(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta}, \qquad \theta > 0 \tag{2}$$

If we wish to use a non-informative prior we might consider $\alpha = 1$ and $\lambda \to 0$, although this may not be sensible here (see Section 4). The main advantage of the Gamma prior is that it is a conjugate for the Exponential distribution, so that the posterior distribution for θ after observing one or more sample values X will also be Gamma. Specifically, for a single observation $X = x$ the joint distribution of (X, θ) has density

$$f_{X,\theta}(x,\theta) = \pi_\theta(\theta) f_{X|\theta}(x \mid \theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^\alpha e^{-(\lambda+x)\theta}, \qquad x, \theta > 0 \tag{3}$$

so by inspection the posterior density $\pi_{\theta|X}(\theta \mid x) \propto \theta^\alpha e^{-(\lambda+x)\theta}$ and $(\theta \mid x) \sim \Gamma(\alpha+1, \lambda+x)$.

The usual procedure, when the loss is a function of θ, would be to choose the action (a_1 or a_2) which minimizes the expected loss under this posterior distribution, giving the Bayes Rule

$$\delta(x) = \arg\min_a \{\mathrm{E_X}\,[\mathrm{L(a}, \theta \mid \mathrm{x})]\} \tag{4}$$

Here however the loss depends not on θ but on the observed value of a second random variable, say Y, representing the result of the sample taken at the reprocessing plant. The Bayes criterion is now $\mathrm{E_X}\,[\mathrm{L(a, y} \mid \mathrm{x})]$, the expected loss under the predictive distribution for Y given $X = x$. If we assume that both the client and the reprocessor use the same sampling and measurement technique, then Y has the same distribution as X (for a given θ), in which case the predictive distribution has density

$$f_{Y|X}(y \mid x) = \int_\theta f_{Y|\theta}(y \mid \theta)\pi_{\theta|X}(\theta \mid x)d\theta \tag{5}$$

$$= \int_0^\infty \frac{(\lambda+x)^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{\alpha+1}e^{-(\lambda+x+y)\theta}d\theta \tag{6}$$

$$= (\alpha+1)\frac{(\lambda+x)^{\alpha+1}}{(\lambda+x+y)^{\alpha+2}}, \qquad y > 0 \tag{7}$$

The expected loss for a_2, the expensive reprocessor, is fixed at $S + K$ and for a_1

$$\mathrm{E_X}\,[\mathrm{L(a_1, y} \mid \mathrm{x})] = S + (K+P)\mathrm{P_X}\,[\mathrm{y} > \mathrm{c}]$$

$$= S + (K+P)\left(\frac{\lambda+x}{\lambda+x+c}\right)^{\alpha+1}$$

so the Bayes Rule is to choose a_1 if $\left(\frac{\lambda+x}{\lambda+x+c}\right)^{\alpha+1} < \frac{K}{K+P}$, i.e. if

$$x < \xi = \frac{c\zeta}{1-\zeta} - \lambda \qquad \text{where } \zeta = \sqrt[\alpha+1]{\frac{K}{K+P}} \tag{8}$$

The expected loss incurred by using this decision rule, say $\delta(x)$, can now be found by integrating the expected loss at fixed X, as given above, with respect to the marginal distribution f_X of X. This gives the Bayes Risk $B(\pi, \delta)$ of the rule δ with respect to the prior distribution π. Formally we may write

$$B(\pi, \delta) = \mathrm{E}_\pi \left[\mathrm{E}_\theta \left[\mathrm{L}(\delta(\mathrm{x}), \mathrm{y}) \mid \theta\right]\right] = \mathrm{E}_\mathrm{f} \left[\mathrm{E}_\mathrm{X} \left[\mathrm{L}(\delta(\mathrm{x}), \mathrm{y}) \mid \mathrm{x}\right]\right] \tag{9}$$

to show two different ways of calculating the Bayes Risk corresponding to two different forms of iterated expectation. It is more convenient here to use the latter.

The marginal distribution for X has density

$$f_X(x) = \int_0^\infty f_{X,\theta}(x, \theta) d\theta = \frac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}, \qquad x > 0 \tag{10}$$

so the Bayes Risk is

$$\begin{aligned} B(\pi, \delta) &= \int_0^\xi (K+P) \left(\frac{\lambda+x}{\lambda+x+c}\right)^{\alpha+1} \frac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}} dx + \int_\xi^\infty K \frac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}} dx \\ &= \lambda^\alpha \left[\frac{K+P}{(\lambda+c)^\alpha} - \frac{K+P}{(\lambda+\xi+c)^\alpha} + \frac{K}{(\lambda+\xi)^\alpha}\right] \end{aligned}$$

Note that λ is a scale parameter for the marginal distribution of X (and of Y) so that the problem is invariant to transformations $(\lambda, c) \to (k\lambda, kc)$ for $k > 0$. This transformation corresponds to a change in the unit of measurement of radioactivity. Similarly the decision made depends only on the ratio K/P not on the individual values.

Suppose then that with suitable units we take $\alpha = 3$, $\lambda = 10$, $c = 5$, $K = 10$, $P = 15$. The prior distribution for the mean level of radioactivity $1/\theta$ is shown in Figure 1, and the marginal distribution for X (and Y) in Figure 2. We find that the Bayes Rule is

$$\delta(x) = \begin{cases} a_1 & \text{if } x < 9.423 \\ a_2 & \text{otherwise} \end{cases} \tag{11}$$

It is clear from Figure 2 that a_1 will be chosen most (93%) of the time, even though the mean level of radioactivity is often above the critical level c. This occurs because of the extreme skewness of the sampling distribution for $Y \mid \theta$ which makes the "gamble" of using the cheaper reprocessor worthwhile even when the average level of radioactivity in a bin is quite high. In the next section we consider the changes which occur when repeated sampling is used at both ends of the process (i.e. for X and for Y).

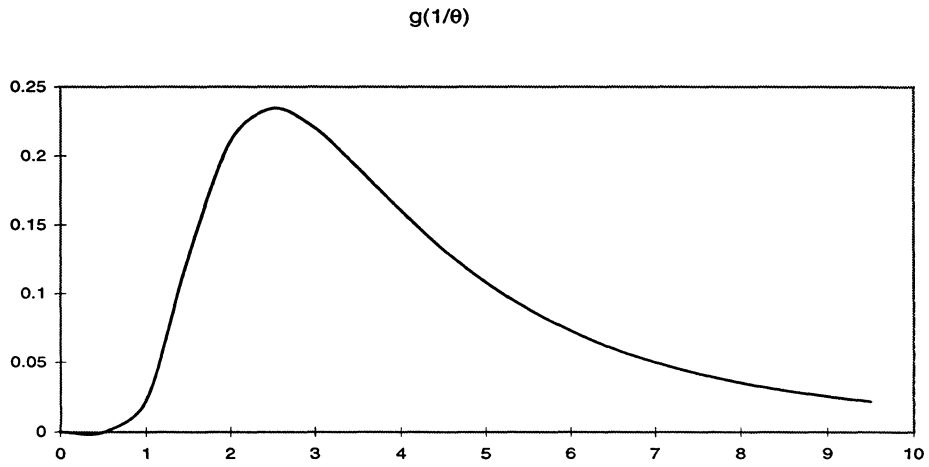

Figure 1. Prior inverted gamma density for mean level of radioactivity $(1/\theta)$ with $\alpha = 3$ and $\lambda = 10$

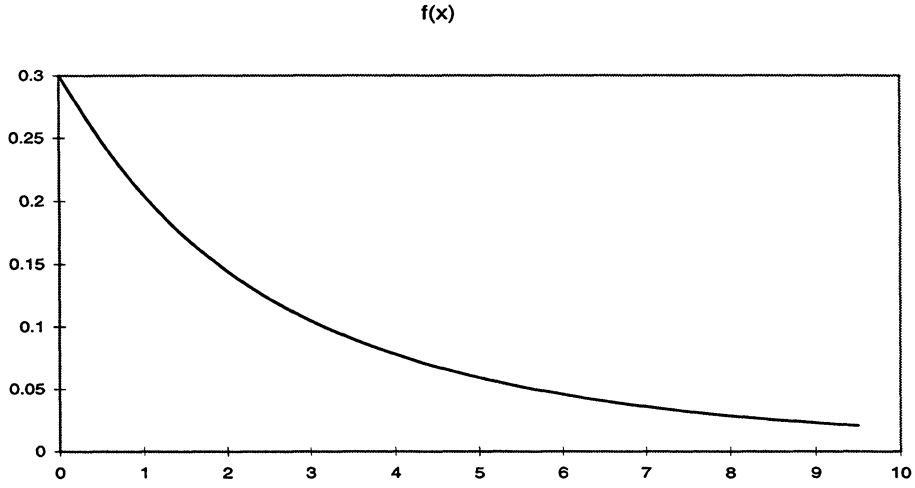

Figure 2. Marginal density for sampled level of radioactivity (X) with $\alpha = 3$ and $\lambda = 10$

3. Optimal Sample Sizes

Using the framework established in the previous section, we now suppose that the client bases his decision on samples $X_1, X_2, \ldots, X_n$ taken from the bin, and

assume that these are iid Exp(θ). It is convenient now to work with the total $X = X_1 + X_2 + \ldots + X_n$ which is a sufficient statistic for θ and is distributed as $\Gamma(n, \theta)$. Similarly the total Y of m samples taken by the cheaper reprocessor, assuming the bin is sent there, will be $\Gamma(m, \theta)$. The decision to accept or reject the bin is then based on the mean of the m samples, so that in the Loss Function c is replaced by mc.

Proceeding as before, we now find that the posterior distribution for $\theta \mid X$ is $\Gamma(n + \alpha, \lambda + x)$. The predictive distribution for $Y \mid X$, following the method of Equations (5)-(7), then has density

$$f_{Y|X}(y \mid x) = \frac{\Gamma(m+n+\alpha)}{\Gamma(m)\Gamma(n+\alpha)} \frac{(\lambda+x)^{n+\alpha} y^{m-1}}{(\lambda+x+y)^{m+n+\alpha}} \tag{12}$$

If we make the substitution $u = \frac{y}{\lambda+x+y}$ we find that the predictive density for $U = \frac{Y}{\lambda+x+Y}$, given $X = x$, has the form of a Beta distribution, and we write:

$$\frac{Y}{\lambda + X + Y} \mid X = x \sim \mathrm{Be(m, n + \alpha)} \tag{13}$$

for the predictive distribution of the transformed variable.

A closed form for the predictive probability P[Y < mc | X = x] is not possible, but the incomplete Beta distribution is easy to calculate numerically (see Press et al. [6]) so we can use Equation (13) to write

$$\mathrm{P[Y < mc \mid X = x] = IB}_{\frac{\mathrm{mc}}{\lambda+\mathrm{x}+\mathrm{mc}}}(\mathrm{m, n} + \alpha). \tag{14}$$

The Bayes Rule is then to choose a_1 if

$$\mathrm{IB}_{\frac{\mathrm{mc}}{\lambda+\mathrm{x}+\mathrm{mc}}}(\mathrm{m, n} + \alpha) > \frac{\mathrm{P}}{\mathrm{K + P}} \tag{15}$$

i.e. if

$$x < \xi = \frac{mc\zeta}{1-\zeta} - \lambda \tag{16}$$

where $1 - \zeta$ is the $\frac{P}{P+K}$ quantile of Be(m, n + α).

To determine the Bayes Risk $B(\pi, \delta)$ for fixed m and n we use the marginal density of X. Proceeding as before we find that

$$\frac{X}{\lambda + X} \sim \mathrm{Be(n, \alpha)} \tag{17}$$

so from Equation (9)

$$\begin{aligned} B(\pi, \delta) &= K\mathrm{P[X > \xi]} + \mathrm{(K + P)P[X < \xi \text{ and } Y > mc]} \\ &= K\left(1 - \mathrm{IB}_{\frac{\xi}{\lambda+\mathrm{xi}}}(\mathrm{n}, \alpha)\right) + (K+P)\int_0^{\xi} \left(1 - \mathrm{IB}_{\frac{\mathrm{mc}}{\lambda+\mathrm{x}+\mathrm{mc}}}(\mathrm{m, n} + \alpha)\right) f_X(x)dx \end{aligned}$$

Although numerical integration is now required this can be accomplished quite easily using standard routines (see Press et al. [6]), and evaluation over a range of values of m and n gives a criterion for choosing the optimal (from the client's viewpoint) sampling plan. Using the parameter values from Section 2, Table 1 gives the Bayes Risk for values of m and n ranging from 1 to 6; note that these values do not include the cost of obtaining the samples. The optimal choices for m and n will depend on the sampling cost: if for example each sample determination has a cost of 0.1, we add $0.1(n+m)$ to each value in the table and find that the optimum is $n = m = 5$. The advantage of not including the sampling cost explicitly in the table is that we can observe the behaviour of the Bayes Risk as n and m are varied. Notice that for $n = 1$ the Bayes Risk initially increases as m increases from 1 to 2, the increased accuracy of determination by the reprocessor being disadvantageous to the client, but thereafter an increase in m results in a lower expected cost. For n however a higher value will always decrease the Bayes Risk, as one would expect: more information for the client should always result in a better decision.

Table 1. Bayes Risk for various sample sizes, $\theta \sim \Gamma(3, 10)$

	m=1	2	3	4	5	6
n=1	7.056	7.165	7.130	7.085	7.046	7.014
2	6.885	6.878	6.782	6.699	6.633	6.580
3	6.781	6.709	6.577	6.470	6.387	6.322
4	6.711	6.596	6.439	6.316	6.221	6.147
5	6.661	6.515	6.340	6.205	6.101	6.019
6	6.623	6.453	6.266	6.120	6.009	5.922

Table 2. Cutoff point for $\bar{x}$, $\theta \sim \Gamma(3, 10)$

	m=1	2	3	4	5	6
n=1	9.423	7.398	6.862	6.623	6.490	6.407
2	7.430	6.158	5.828	5.684	5.605	5.557
3	6.768	5.747	5.487	5.375	5.315	5.279
4	6.438	5.543	5.318	5.223	5.173	5.142
5	6.240	5.422	5.217	5.132	5.088	5.062
6	6.109	5.341	5.151	5.073	5.033	5.009

Table 3. Probability of choosing a_2, $\theta \sim \Gamma(3, 10)$

	m=1	2	3	4	5	6
n=1	0.136	0.190	0.209	0.218	0.223	0.226
2	0.182	0.239	0.257	0.266	0.271	0.274
3	0.205	0.262	0.280	0.288	0.293	0.295
4	0.219	0.276	0.293	0.301	0.305	0.307
5	0.229	0.285	0.302	0.309	0.313	0.315
6	0.235	0.291	0.308	0.315	0.318	0.321

Table 2 shows the cutoff point for the Bayes Rule, expressed in relation to the sample mean, i.e. if $\bar{x} = x/n$ is greater than the tabulated value, the bin should be

sent to the expensive reprocessor. As the number of samples increases, the cutoff converges quite quickly to the critical value c. Table 3 shows the proportion of bins which would be sent to the expensive reprocessor for each sampling plan.

Several different behaviours are possible, depending on the parameter values. In some cases the Bayes risk may decrease very slowly, or even increase, as m increases from 1; in other cases it decreases quite markedly. It is important therefore to get accurate information about costs and the prior before deciding whether it is worthwhile obtaining extra samples, and whether the extra effort should be devoted to X or Y or both equally.

4. Non-conjugate Prior

The prior Gamma distribution for θ employed in Sections 2 and 3 was chosen mainly for mathematical convenience. We now re-examine its appropriateness and how a wider class of priors might be incorporated into the analysis.

To obtain a reasonable prior distribution from the client, he must be invited to speculate on the likelihood of a range of values of θ. This may be difficult since θ is itself not a particularly meaningful parameter. A far more natural parametrization of the problem would be to use $1/\theta$ which represents the mean level of radioactivity in the bin; this is something about which the client might reasonably be expected to speculate. We could still proceed by showing the client graphs of the density of $1/\theta$ for various choices of α and λ, as in Figure 1, but even so we are restricting ourselves to a class of distributions, the Inverted Gamma, which might be thought inappropriate. These distributions are very long-tailed, having less than $\alpha - 1$ finite moments.

Suppose that instead we decide to use a general prior distribution specified for $1/\theta$. We can still denote the implied prior for θ by $\pi_\theta(\theta)$, but the integrals needed to evaluate the marginal distribution for X and the predictive distribution for Y will not now involve simple special functions like the Gamma and Beta. It has become commonplace is such situations to employ some form of Monte Carlo integration.

There are essentially three stages to the calculation:

- Evaluate the risk for fixed cutoff ξ and fixed sample sizes n,m.
- Choose ξ to minimize the risk for fixed n,m.
- Choose n,m to minimize the Bayes Risk.

If we denote the rule with cutoff ξ by δ_ξ, i.e.

$$\delta_\xi(x) = \begin{cases} a_1 & \text{if } x < \xi \\ a_2 & \text{otherwise} \end{cases} \tag{18}$$

then we need to evaluate

$$R(\pi, \delta_\xi) = K\mathrm{P}[\mathrm{X} > \xi] + (\mathrm{K} + \mathrm{P})\mathrm{P}[\mathrm{X} < \xi \text{ and } \mathrm{Y} > \mathrm{mc}] \tag{19}$$

One approach would be to sample from the joint distribution of (θ, X, Y). Provided that the prior for $1/\theta$ is reasonably easy to simulate, we invert a randomly drawn value to get θ, then draw X and Y from their conditional distributions $\Gamma(n, \theta)$ and $\Gamma(m, \theta)$ respectively. The conditional independence of X and Y given θ means that we do not need iterative methods such as the Gibbs sampler. Given a sample (θ_i, X_i, Y_i), $i = 1, \ldots, N$ we can approximate the risk for given ξ by

$$R(\pi, \xi) \simeq \sum_{i=1}^{N} K\mathcal{I}_{\{x_i \leq \xi\}} + (K + P)\mathcal{I}_{\{x_i > \xi \text{ and } y_i > \text{mc}\}} \tag{20}$$

where $\mathcal{I}$ is the indicator function.

Because of the dimensionality problem this method requires a huge sample size to achieve even reasonable accuracy, and repeated computation for varying ξ becomes very inefficient. A better approach is to simulate for θ only and to calculate directly $\mathrm{P}_\theta[\mathrm{X} > \xi]$ and $\mathrm{P}_\theta[\mathrm{Y} > \mathrm{mc}]$. These probabilities are incomplete gamma functions and can be calculated quite efficiently (see Press et al. [6]), giving

$$R(\pi, \xi) \simeq \sum_{i=1}^{N} K\left(1 - \mathrm{IG}_{\theta_i \xi}(\mathrm{n})\right) + (K + P)\mathrm{IG}_{\theta_i \xi}(\mathrm{n})\left(1 - \mathrm{IG}_{\mathrm{mc}\theta_i}(\mathrm{m})\right) \tag{21}$$

where $\mathrm{IG}_z(\mathrm{k})$ denoted the incomplete gamma function

$$\mathrm{IG}_z(\mathrm{k}) = \int_0^z \mathrm{u}^{\mathrm{k}-1}\mathrm{e}^{-\mathrm{u}}\mathrm{du} \tag{22}$$

Note that only the X probabilities depend on ξ, so the Y probabilities for each θ_i may be stored and re-used. This method requires a much smaller sample of θ values to achieve reasonable accuracy, and is therefore more efficient than use of the full multivariate joint distribution.

Using the prior and parameter values from Section 3 it was found that $N = 10,000$ gave sufficient accuracy (2 dp) and a reasonable computation time (about 90s). Now however we are no longer restricted to a small class of priors. The calculation was repeated using a $\Gamma(4, 1)$ prior for $1/\theta$. This is shown in Figure 3 for comparison with Figure 1; it is similar but much less long-tailed. The estimated Bayes Risk and cutoff value using this prior are given in Tables 4 and 5. We now find that with a cost of sampling of 0.1 the best option is $n = m = 1$.

Table 4. Bayes Risk for various sample sizes, $1/\theta \sim \Gamma(4, 1)$

	m=1	2	3	4	5	6
n=1	6.509	6.606	6.554	6.495	6.443	6.400
2	6.469	6.450	6.408	6.321	6.249	6.190
3	6.435	6.418	6.298	6.193	6.107	6.037
4	6.406	6.354	6.214	6.094	5.998	5.920
5	6.382	6.302	6.146	6.016	5.911	5.827
6	6.361	6.260	6.091	5.952	5.840	5.751

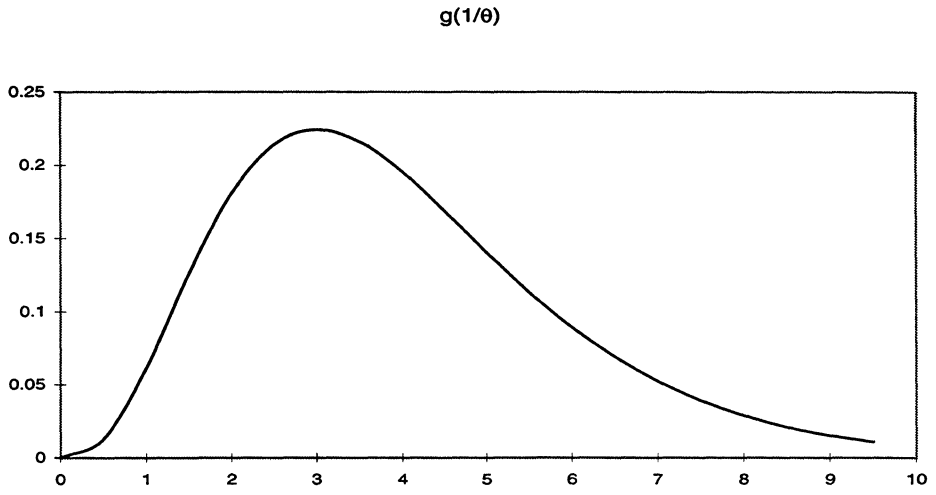

Figure 3. Prior inverted gamma density for mean level of radioactivity $(1/\theta)$ with $\alpha = 3$ and $\lambda = 10$

Table 5. Cutoff point for $\bar{x}$, $1/\theta \sim \Gamma(4, 1)$

	m=1	2	3	4	5	6
n=1	14.538	10.36	9.374	8.907	8.635	8.523
2	9.813	7.494	6.909	6.685	6.561	6.469
3	8.257	6.529	6.137	5.974	5.894	5.816
4	7.514	6.085	5.758	5.619	5.541	5.487
5	7.075	5.829	5.532	5.413	5.354	5.315
6	6.777	5.645	5.390	5.283	5.240	5.200

5. Discussion

In the above analysis we have considered the problem only from the client's point of view, assuming that he can pay the reprocessor to take extra samples, as well as deciding to take more samples himself, if this is to his advantage. We have also assumed that the critical value c used by the reprocessor is kept fixed for different sample sizes. If we now consider the reprocessor's point of view, he clearly does not want to accept material which has too high a level of radioactivity. We assume that he requires the mean level for each bin to be less than c, but that he does not allow for sampling variability in making his test. Were he to do so, he would want to adjust the critical value depending on the sample size m. It would also be to his advantage to take extra samples. Rather than use decision theory merely to improve the decision-making of one side in the process, as we have done here, it would be more appropriate to use an agreed decision theory framework as

a negotiating tool in establishing an optimal sampling scheme which would be of benefit to both parties.

It has already been noted that the optimal solution for the problem we have considered seems to be quite sensitive to the parameter values and prior information. This shows the need for estimated costs to be as accurate as possible, and for prior data to be incorporated in choosing the prior distribution, possibly through an empirical Bayes approach. This sensitivity is probably due in part to the use of long-tailed distributions. There is a considerable range of behavior for different parameter values and distributions. In particular the fact that increasing the sample size for Y may either increase or decrease the risk is interesting, and is the subject of further investigation.

Acknowledgments

The author would like to thank the Editor and the referees for their support and helpful suggestions.

References

1. J. Aitchison and I. R. Dunsmore. *Statistical Prediction Analysis.* University Press, Cambridge, 1975.
2. M. H. DeGroot. *Optimal Statistical Decisions.* McGraw-Hill,New York, 1970.
3. T. S. Ferguson. *Mathematical Statistics: a Decision Theoretic Approach.* Academic Press, New York, 1967.
4. S. Geisser. *Predictive Inference: an Introduction.* Chapman and Hall, London, 1993.
5. D. V. Lindley and N. D. Singpurwalla. On the evidence needed to reach agreed action between adversaries, with application to acceptance sampling. *J. Amer. Statist. Assoc.*, 86, 993–937, 1991.
6. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing*, 2nd ed. University Press, Cambridge, 1992.
7. H. V. Roberts. Probabilistic prediction. *J. Amer. Statist. Assoc.*, 60, 50–62, 1965.
8. L. J. Wolfson, J. B. Kadane and M. J. Small. A subjective Bayesian approach to environmental sampling. In: *Case Studies in Bayesian Statistics* Vol.3 (C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi and N. D. Singpurwalla, eds.) Springer-Verlag, New York, 457–468, 1997.