BAYESIAN POINT ESTIMATION AND PREDICTION*

ROBERT R. BRITNEY AND ROBERT L. WINKLER

(Received Sept. 28, 1971)

Abstract

In the Bayesian viewpoint, point estimation and prediction are treated from a decision-making standpoint. If a loss function can be determined which associates a loss with every possible error of estimation or prediction, then the optimal estimator or predictor is that value which minimizes expected loss. In most applications, the loss function is assumed to be linear or quadratic in the error of estimation or prediction, although there are many practical situations in which these simple functions are quite inappropriate. In this paper, we investigate the properties of Bayesian point estimates under other loss functions; both the general case and two special cases (power and exponential loss functions) are considered. For the special cases, we also investigate the sensitivity of Bayesian point estimation and prediction to misspecification in the loss function and discuss the practical implications of the results.

1. Introduction

In the Bayesian viewpoint, the choice of a point estimate or prediction is considered to be a decision-making problem, generally an infinite-action problem in which the space of actions, A, and the parameter space, Ω, coincide. The decision maker's objective is to maximize his expected utility, or, equivalently, to minimize his expected loss (with losses expressed in terms of utility rather than in terms of some other *numéraire* such as money). In Bayesian point estimation problems it is more convenient to work with losses than with utilities. Formally, suppose that the decision maker wants to estimate a real-valued parameter θ or to predict a future sample outcome θ under the following conditions:

1. The decision maker's probability judgments about θ are quantified

* An earlier version of this paper was presented at the Annual Meeting of the American Statistical Association in New York, August 1968.

in terms of a continuous density function f defined on Ω. In estimation problems, f represents a prior or posterior distribution in Bayesian terminology. In prediction problems, f represents a predictive distribution, which is simply the marginal distribution of a future event, given the current state of (prior) information*.

2. A loss function can be determined which associates a loss $L(\boldsymbol{a}, \theta)$ with every pair of values $(\boldsymbol{a}, \theta)$, where $\boldsymbol{a} \in A=\Omega$ is an estimate of θ.

An optimal value of $\boldsymbol{a}$ is a value which minimizes the decision maker's expected loss,

$$EL(\boldsymbol{a})=\int_{\Omega} L(\boldsymbol{a}, \theta)f(\theta)d\theta \tag{1.1}$$

(such a value may not always exist). Assume that Ω is the real line, R. It is well known that if the loss function is a quadratic function of the form

$$L(\boldsymbol{a}, \theta)=k(\boldsymbol{a}-\theta)^2 \tag{1.2}$$

with $k>0$ (a squared-error loss function), then the mean of the decision maker's distribution, $E(\theta)=\mu$, is an optimal point estimate ([9], Ch. 6). If the loss function is a linear function of the form

$$L(\boldsymbol{a}, \theta)=\begin{cases} k_0(\boldsymbol{a}-\theta) & \text{if } \boldsymbol{a}\geqq\theta\ , \\ k_u(\theta-\boldsymbol{a}) & \text{if } \boldsymbol{a}\leqq\theta\ , \end{cases} \tag{1.3}$$

where k_0 and k_u are positive constants, then the $k_u/(k_u+k_0)$ fractile of $f(\theta)$ is an optimal point estimate. Here k_0 and k_u can be thought of as the per unit costs of overestimation and underestimation, respectively.

In most applications, $L(\boldsymbol{a}, \theta)$ is assumed to be of one of the above forms. However, while it is true that the linear and quadratic loss functions have wide applicability, there are situations in which they are quite inappropriate. For instance, even if the loss function is linear or quadratic when losses are expressed in terms of money, it may be of a completely different form when the losses are expressed in terms of utility. This phenomenon is likely to occur whenever the decision maker's utility function is not linear as a function of money (for a general discussion, see Baron [2]; for concrete examples, see Gould [6] and Sections 4 and 5 of this paper). Of course, even if the utility function is linear in terms of money, the linear and quadratic loss functions may be incapable of adequately representing the potential losses in any given situation (for examples of such situations, see Granger [7]). Raiffa

* Henceforth, the problem at hand will generally be referred to as an estimation problem, although it should be emphasized that the framework applies equally well to prediction problems.

and Schlaifer ([9], pp. 205–207) note the possibility of using "modified" linear and quadratic loss structures to allow the decision maker more flexibility in the choice of a loss function. Marshall and Olkin [8] apply quadratic and exponential loss functions to screening and classification problems.

In this paper, we investigate the properties of Bayesian point estimates under loss functions other than the simple (or modified) linear and quadratic functions. In Section 2 the symmetric case (symmetric loss function and distribution) is considered, and the more general case is discussed in Section 3. In Sections 4 and 5, numerical results are presented for the normal distribution and two special classes of loss functions, power loss functions and exponential loss functions. A sensitivity analysis is discussed in Section 6, and Section 7 contains a brief summary.

2. The symmetric case

It can be shown that for a wide class of loss functions and distributions, the mean of the distribution is a Bayesian point estimate. A number of conditions regarding $L(\boldsymbol{a}, \theta)$ and $f(\theta)$ will be introduced at this point.

CONDITION 1. $L(\boldsymbol{a}, \theta)$ is a monotone nondecreasing function of $(\boldsymbol{a}-\theta)$ for $\boldsymbol{a} \geqq \theta$ and a monotone nondecreasing function of $(\theta-\boldsymbol{a})$ for $\boldsymbol{a} \leqq \theta$.

This implies that $L(\boldsymbol{a}, \boldsymbol{a}) \leqq L(\boldsymbol{a}, \theta)$ for all $\theta \in \Omega$. Thus, using the opportunity loss concept, $L(\boldsymbol{a}, \boldsymbol{a})=0$. Situations in which Condition 1 does not hold are possible (e.g., Savage's example of William Tell and the apple shot, for which a downward error of 10° might be preferred to one of 1° ([10], pp. 230–231)) but are very unusual. Furthermore, unless there is some interval over which the loss function is constant, the above inequalities for L are strict. To maintain flexibility, this will not be assumed, but it *will* be assumed that L is not *everywhere* constant.

CONDITION 2. $L(\boldsymbol{a}, \theta)$ is symmetric about $\theta=\boldsymbol{a}$. That is, for any $\boldsymbol{a}$ and e,

$$L(\boldsymbol{a}, \boldsymbol{a}+e)=L(\boldsymbol{a}, \boldsymbol{a}-e) . \tag{2.1}$$

This is equivalent to saying that $L(\boldsymbol{a}, \theta)$ can be expressed as a function of the absolute difference between $\boldsymbol{a}$ and θ:

$$L(\boldsymbol{a}, \theta)=L(|\boldsymbol{a}-\theta|)=L(|e|) , \tag{2.2}$$

where $e=\boldsymbol{a}-\theta$.

CONDITION 3. $f(\theta)$ is symmetric about its mean, μ.

CONDITION 4. For each $\boldsymbol{a} \in A$, $L(\boldsymbol{a}, \theta)$ is a convex function of θ.

This is reasonable when small estimation errors are not too serious but larger errors are much more serious. One possible argument against convex loss functions is that $L \to \infty$ as $|\boldsymbol{a}-\theta| \to \infty$. In some situations it is more reasonable to assume that there is an upper bound for L. A slightly stronger condition than Condition 4 is strict convexity of L. DeGroot and Rao discuss the case in which L is symmetric and convex in [3] and consider multivariate extensions in [4].

CONDITION 5. The cumulative distribution function $F(\theta)$ is a convex function over $(-\infty, \mu]$.

This implies that $f(\theta)$ is non-decreasing on $(-\infty, \mu]$. If $f(\theta)$ is strictly increasing on this interval, then $F(\theta)$ is strictly convex.

Using the above conditions, two important results can be derived.

PROPOSITION 2.1. If Conditions 1–4 are satisfied, μ is an optimal point estimate.

This proposition states that for any symmetric, non-decreasing, convex loss function and any symmetric distribution, the mean of the distribution is optimal. A proof is given in [5] (see also [6]).

PROPOSITION 2.2. If Conditions 1–3 and 5 are satisfied, μ is an optimal point estimate.

Here the requirement that $L(\boldsymbol{a}, \theta)$ be convex is replaced with the requirement that $F(\theta)$ be convex for $\theta < \mu$. Note that Conditions 3 and 5, taken together, imply that $f(\theta)$ is either a uniform distribution or a symmetric, unimodal distribution. The proof of Proposition 2.2 follows from a theorem of Anderson [1], p. 170; see also [7], [11], [12].

The above results imply that for a wide class of loss functions and distributions, μ is an optimal Bayesian point estimate. Examples of loss functions satisfying Conditions 1, 2, and 4 are illustrated in Fig. 1. For any of these functions and a symmetric distribution, μ is an optimal estimate. Furthermore, if the distribution is unimodal in addition to being symmetric, loss functions such as those illustrated in Fig. 2 (as well as those in Fig. 1) lead to the choice of μ. The use of the posterior mean as an estimator reflects in part the fact that for any posterior distribution it is optimal under a squared-error loss function. The results of this section indicate that the mean is optimal for an extended class of symmetric loss functions and distributions.

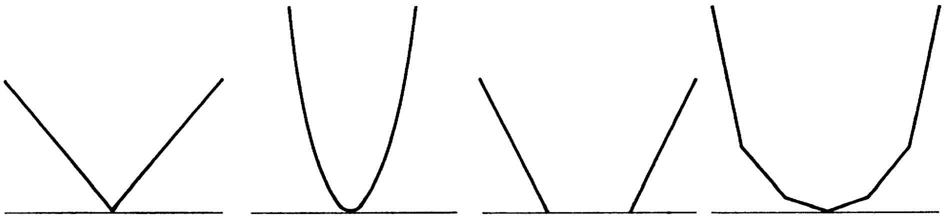

Fig. 2 Examples of Nonconvex Loss Functions

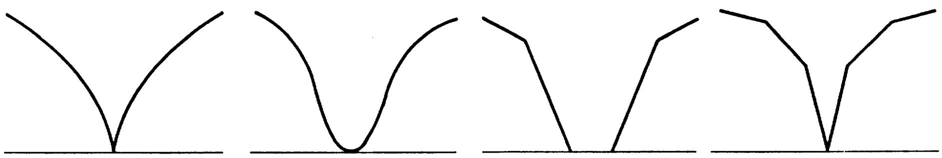

Fig. 1 Examples of Convex Loss Functions

3. The general case

"It is flatly impossible to find *any* one estimator which is 'good'—let alone 'best'—for *all* applications in any absolute sense," ([9], p. 181). The results of the preceding section are of much interest, but they hold only when both the loss function and the distribution are symmetric. Violations of these conditions are quite common; it is easy to imagine situations in which the losses due to underestimation are different from the losses due to overestimation (e.g., deciding how many units of a perishable commodity to stock) or the distribution is skewed (e.g., a distribution of the demand for a commodity).

To allow for asymmetric loss functions, we will break $L(\boldsymbol{a}, \theta)$ up into two functions, one for overestimation and one for underestimation:

$$L(\boldsymbol{a}, \theta)=L_0(\boldsymbol{a}-\theta)+L_u(\theta-\boldsymbol{a}) , \tag{3.1}$$

where $L_0(e)=L_u(e)=0$ for $e \leqq 0$. Assume that L_0 and L_u are monotone nondecreasing functions, so that L satisfies Condition 1 of Section 2.

The expected loss associated with action $\boldsymbol{a}$ is

$$EL(\boldsymbol{a})=\int_{-\infty}^{\boldsymbol{a}} L_0(\boldsymbol{a}-\theta)f(\theta)d\theta+\int_{\boldsymbol{a}}^{\infty} L_u(\theta-\boldsymbol{a})f(\theta)d\theta . \tag{3.2}$$

This expected loss may not exist for some or all values of $\boldsymbol{a}$, due to nonconvergence. Furthermore, even if it exists, the optimal value of $\boldsymbol{a}$ may not be unique.

PROPOSITION 3.1. If L_0 and L_u are convex, and $EL(\boldsymbol{a})<\infty$ for all $\boldsymbol{a} \in \Omega$, the set of optimal values of $\boldsymbol{a}$ is a bounded interval.

It is easily shown that $EL(\boldsymbol{a})$ is convex. Furthermore, since L_0 and L_u are nondecreasing and are not everywhere constant, $EL(\boldsymbol{a})$ approaches ∞ as $\boldsymbol{a}$ approaches $\pm\infty$. Thus, the set of optimal values of $\boldsymbol{a}$ must be a bounded interval.

COROLLARY. *If L_0 and L_u are convex and either is strictly convex and $EL(\boldsymbol{a})<\infty$ for all $\boldsymbol{a} \in \Omega$, the optimal value of $\boldsymbol{a}$ is unique.*

From the above proof, if either L_0 or L_u is strictly convex, then EL is strictly convex and assumes its minimum value at a single point.

Of course, if L_0 and L_u are not convex, the set of optimal values need not be a bounded interval, and the determination of the optimal value(s) could be more difficult. Then we have the following:

PROPOSITION 3.2. If L_0 and L_u are twice differentiable on $(0, \infty)$, then necessary conditions for $\boldsymbol{a}$ to be optimal are

$$E_{-\infty}^{\boldsymbol{a}}[L_0'(\boldsymbol{a}-\theta)]=E_{\boldsymbol{a}}^{\infty}[L_u'(\theta-\boldsymbol{a})] \tag{3.5}$$

and

$$E_{-\infty}^{\boldsymbol{a}}[L_0''(\boldsymbol{a}-\theta)]+E_{\boldsymbol{a}}^{\infty}[L_u''(\theta-\boldsymbol{a})]+L_0'(0)f(\boldsymbol{a})+L_u'(0)f(\boldsymbol{a})\geqq 0 , \tag{3.6}$$

where the expectations are partial expectations of the form

$$E_x^y[L(e)]=\int_x^y L(e)f(\theta)d\theta$$

and the primes denote differentiation with respect to e*. Furthermore, under the assumptions of Proposition 3.1, (3.5) is a sufficient condition for optimality.

For example, consider a class of loss functions which are power functions such that

$$L_0(\boldsymbol{a}-\theta)=\begin{cases} k_0(\boldsymbol{a}-\theta)^r & \text{if } \boldsymbol{a}-\theta\geqq 0 , \\ 0 & \text{if } \boldsymbol{a}-\theta<0 , \end{cases}$$

(3.7) and

$$L_u(\theta-\boldsymbol{a})=\begin{cases} k_u(\theta-\boldsymbol{a})^s & \text{if } \boldsymbol{a}-\theta\leqq 0 , \\ 0 & \text{if } \boldsymbol{a}-\theta>0 , \end{cases}$$

where k_u and k_0 are positive constants as in (1.3) and r and s are also positive constants. The first order condition (3.5) becomes

---

* For example, $L_0'(\boldsymbol{a}-\theta)=\left[\frac{d}{de}L_0(e)\,|\,e=\boldsymbol{a}-\theta\right]$. Throughout this paper all derivatives evaluated at $e=0$ are taken to be right-hand derivatives $\left(\text{e.g., } L_0'(0)=\left[\frac{d+}{de}L_0(e)\,|_{e=0}\right]\right)$.

$$\frac{E_{-\infty}^{\boldsymbol{a}}(\boldsymbol{a}-\theta)^{r-1}}{E_{\boldsymbol{a}}^{\infty}(\theta-\boldsymbol{a})^{s-1}}=\frac{sk_u}{rk_0}\,. \tag{3.8}$$

For the power loss functions, $L(\boldsymbol{a}, \theta)$ is symmetric about $\theta=\boldsymbol{a}$ if and only if (1) $k_0=k_u$ and (2) $r=s$; hence there are two possible types of asymmetry. Note also that $L(\boldsymbol{a}, \theta)$ is strictly convex if $r>1$ and $s>1$. A final point of interest is that it is only necessary to know the ratio, not the magnitudes, of k_u and k_0; thus, if both L_0 and L_u are multiplied by the same positive constant, the entire function L is just multiplied by that constant and the optimal point estimate is unchanged.

For a second example, consider a class of exponential loss functions:

$$L_0(\boldsymbol{a}-\theta)=\begin{cases} k_0\,|e^{r(\boldsymbol{a}-\theta)}-1| & \text{if } \boldsymbol{a}-\theta\geqq 0\,,\\ 0 & \text{if } \boldsymbol{a}-\theta<0\,,\end{cases}$$

(3.9) and

$$L_u(\theta-\boldsymbol{a})=\begin{cases} k_u\,|e^{s(\theta-\boldsymbol{a})}-1| & \text{if } \boldsymbol{a}-\theta\leqq 0\,,\\ 0 & \text{if } \boldsymbol{a}-\theta>0\,.\end{cases}$$

The first order condition (3.5) is

$$\frac{E_{-\infty}^{\boldsymbol{a}}(e^{-r\theta})}{E_{\boldsymbol{a}}^{\infty}(e^{s\theta})}=\left[\frac{|s|\,k_u}{|r|\,k_0}\right]e^{-\boldsymbol{a}(r+s)}\,. \tag{3.10}$$

For this class of exponential loss functions, L is strictly convex if $r>0$ and $s>0$.

It should be noted that if $F(\theta)$ is not continuous, Proposition 3.1 still holds, but Proposition 3.2 is inapplicable. More general necessary conditions for optimality are given by Proposition 3.3.

PROPOSITION 3.3. If L_0 and L_u are (not necessarily twice) differentiable on $(0, \infty)$, and $\boldsymbol{a}$ is optimal, then

$$E_{\theta<\boldsymbol{a}}[L_0'(\boldsymbol{a}-\theta)]\leqq E_{\theta\geqq\boldsymbol{a}}[L_u'(\theta-\boldsymbol{a})] \tag{3.11}$$

and

$$E_{\theta\leqq\boldsymbol{a}}[L_0'(\boldsymbol{a}-\theta)]\geqq E_{\theta>\boldsymbol{a}}[L_u'(\theta-\boldsymbol{a})]\,. \tag{3.12}$$

This follows from examination of the left- and right-hand derivatives of EL at $\boldsymbol{a}$; optimality requires that the former be nonpositive and the latter be nonnegative. If the assumptions of Proposition 3.1 are met, then (3.11) and (3.12) are sufficient for optimality; if the assumptions of the corollary are met, they are sufficient *and* the optimal value of $\boldsymbol{a}$ is unique. Furthermore, if either (1) $L_0'(0)=L_u'(0)=0$ or (2) $F(\theta)$ is continuous, then (3.11) and (3.12) reduce to (3.5). For the power loss functions with $r>1$ and $s>1$, $L_0'(0)=L_u'(0)=0$. If $r=s=1$, (3.11) and (3.12) are

$$k_0 P(\theta < \boldsymbol{a}) \leqq k_u P(\theta \geqq \boldsymbol{a}) \qquad \text{and} \qquad k_0 P(\theta \leqq \boldsymbol{a}) \geqq k_u P(\theta > \boldsymbol{a}) .$$

For the exponential loss functions, $L_0'(0) = k_0|r|$ and $L_u'(0) = k_u|s|$, so that (3.11) and (3.12) are

$$k_0|r|e^{ar}E_{\theta<a}(e^{-r\theta}) \leqq k_u|s|[e^{-as}E_{\theta>a}(e^{s\theta}) + P(\theta = \boldsymbol{a})]$$

and

$$k_0|r|[e^{ar}E_{\theta<a}(e^{-r\theta}) + P(\theta = \boldsymbol{a})] \geqq k_u|s|e^{-as}E_{\theta>a}(e^{s\theta}) .$$

4. Optimal estimates for power loss functions and the normal distribution

In this section we discuss the determination of optimal estimates when the loss function is of the form (3.7) with $r = s = n$ and the distribution $f(\theta)$ is a normal distribution. Notice that, in terms of the discussion following (3.8), we are only considering one of the two types of asymmetry possible for the general class of power loss functions, the asymmetry which occurs when $k_0 \neq k_u$. This situation could arise, for example, when losses are linear in terms of money but utility is not linear in money. Suppose that in terms of money, the loss function is

$$L(\boldsymbol{a}, \theta) = \begin{cases} c_0(\boldsymbol{a} - \theta) & \text{if } \boldsymbol{a} \geqq \theta , \\ c_u(\theta - \boldsymbol{a}) & \text{if } \boldsymbol{a} \leqq \theta , \end{cases}$$

and the utility function for negative *changes* in total monetary wealth is

$$U(\Delta x) = -|\Delta x|^n .$$

Then, in terms of utility, the loss function is of the form (3.7) with $r = s = n$, $k_0 = c_0^n$, and $k_u = c_u^n$.

If $r = s = n$, (3.8) becomes

$$\frac{k_u}{k_0} = \frac{E_{-\infty}^{a}(\boldsymbol{a} - \theta)^{n-1}}{E_a^{\infty}(\theta - \boldsymbol{a})^{n-1}} . \tag{4.1}$$

The right-hand side of (4.1) can be evaluated for any choice of $\boldsymbol{a}$. But this result is precisely the loss ratio k_u/k_0 for which $\boldsymbol{a}$ is optimal. By following this procedure for a large number of values of $\boldsymbol{a}$, a table can be developed relating the optimal value of $\boldsymbol{a}$ to the loss ratio for the standard normal distribution.

From (4.1),

$$\frac{k_u}{k_0}=\frac{\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i a^{n-1-i}E_{-\infty}^{a}(\theta^i)}{(-1)^{n-1}\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i a^{n-1-i}E_{a}^{\infty}(\theta^i)}\,,$$

where

$$E_a^{\infty}(\theta^i)=[E(\theta^i)-E_{-\infty}^{a}(\theta^i)]\,.$$

But $E(\theta^i)$ is zero if i is odd and $1\cdot 3\cdots(i-1)$ if i is even. Furthermore, the partial moments $E_{-\infty}^{a}(\theta^i)$ are

$$E_{-\infty}^{a}(\theta^i)=\begin{cases} F_{N^*}(a) & \text{if } i=0\,,\\ -f_{N^*}(a) & \text{if } i=1\,,\\ -a^{i-1}f_{N^*}(a)+(i-1)E_{-\infty}^{a}(\theta^{i-2}) & \text{if } i\geqq 2\,,\end{cases}$$

where F_{N^*} and f_{N^*} are the cumulative distribution function and density function of the standard normal distribution (see [15]). To simplify the notation, let $G=F_{N^*}(a)$ and $g=f_{N^*}(a)$. Then

$$\frac{k_u}{k_0}=\begin{cases} \dfrac{G}{1-G} & \text{if } n=1\,,\\[2ex] \dfrac{aG+g}{a(1-G)-g} & \text{if } n=2\,,\\[2ex] \dfrac{a^2G+ag+G}{a^2(1-G)-ag+(1-G)} & \text{if } n=3\,.\end{cases} \tag{4.3}$$

In general, if we let t_n and u_n represent the numerator and denominator on the right-hand side of (4.3) for a particular value of n (ignoring the sign preceding the entire term), then it can be shown that the following recursive relationship holds:

$$\frac{k_u}{k_0}=(-1)^{n-1}\left[\frac{at_{n-1}+(n-2)t_{n-2}}{au_{n-1}+(n-2)u_{n-2}}\right] \qquad \text{if } n\geqq 3\,. \tag{4.4}$$

Using the preceding formulas, tables of the loss ratios associated with different optimal estimates were found for $a=0(0.01)5$ and $n=1(1)6$. Values of F_{N^*} and f_{N^*} to 15 decimal places were used [13]. The results are summarized in Table 1. Because of the symmetry of the normal distribution, it is only necessary to consider positive values of a. Such values will be associated with loss ratios greater than one; whenever $k_u>k_0$, the optimal estimate should be greater than zero to reduce the higher expected loss of underestimation. If the loss ratio is less than one, then the optimal estimate will simply be the negative of the optimal estimate under the reciprocal loss ratio. For example, if $n=1$, the optimal estimate is 0.97 if $k_u/k_0=5$ and -0.97 if $k_u/k_0=1/5$.

Table 1. Optimal Values of a for the Standard Normal Distribution and Power Loss Functions with $r=s=n$

k_u/k_0	n = 1	2	3	4	5	6
1.0	0.00	0.00	0.00	0.00	0.00	0.00
1.1	0.06	0.04	0.03	0.03	0.02	0.02
1.2	0.11	0.07	0.06	0.05	0.04	0.04
1.3	0.16	0.10	0.08	0.07	0.06	0.06
1.4	0.21	0.13	0.11	0.09	0.08	0.07
1.5	0.25	0.16	0.13	0.11	0.10	0.09
1.6	0.29	0.19	0.15	0.12	0.11	0.10
1.7	0.33	0.21	0.17	0.14	0.12	0.11
1.8	0.37	0.23	0.18	0.16	0.14	0.12
1.9	0.40	0.26	0.20	0.17	0.15	0.14
2.0	0.43	0.28	0.22	0.18	0.16	0.15
2.5	0.57	0.36	0.29	0.24	0.22	0.19
3	0.67	0.44	0.34	0.29	0.26	0.23
4	0.84	0.55	0.43	0.37	0.33	0.29
5	0.97	0.64	0.50	0.43	0.38	0.34
6	1.07	0.71	0.56	0.48	0.42	0.38
7	1.15	0.77	0.61	0.52	0.46	0.41
8	1.22	0.82	0.65	0.55	0.49	0.44
9	1.28	0.86	0.68	0.58	0.51	0.47
10	1.34	0.90	0.72	0.61	0.54	0.49
12	1.43	0.97	0.77	0.66	0.58	0.53
14	1.50	1.03	0.82	0.70	0.62	0.56
16	1.56	1.08	0.86	0.73	0.65	0.59
18	1.62	1.12	0.89	0.76	0.68	0.61
20	1.67	1.16	0.93	0.79	0.70	0.64
25	1.77	1.24	0.99	0.85	0.75	0.68
30	1.85	1.31	1.05	0.90	0.79	0.72
35	1.91	1.36	1.09	0.94	0.83	0.75
40	1.97	1.41	1.13	0.97	0.86	0.78
45	2.02	1.45	1.17	1.00	0.89	0.81
50	2.06	1.49	1.20	1.03	0.91	0.83
60	2.13	1.55	1.25	1.07	0.95	0.87
70	2.19	1.60	1.30	1.11	0.99	0.90
80	2.25	1.65	1.34	1.15	1.02	0.93
90	2.29	1.69	1.37	1.18	1.05	0.95
100	2.33	1.72	1.41	1.21	1.08	0.98
150	2.48	1.85	1.52	1.31	1.16	1.06
200	2.58	1.95	1.60	1.38	1.23	1.12
250	2.65	2.02	1.66	1.44	1.28	1.16
300	2.71	2.07	1.71	1.48	1.32	1.20
350	2.76	2.12	1.75	1.52	1.35	1.23
400	2.81	2.16	1.79	1.55	1.38	1.26
450	2.85	2.20	1.82	1.58	1.41	1.28
500	2.88	2.23	1.85	1.61	1.43	1.31
600	2.94	2.29	1.90	1.65	1.47	1.34
700	2.98	2.33	1.94	1.69	1.51	1.37
800	3.02	2.37	1.98	1.72	1.54	1.40
900	3.06	2.41	2.01	1.75	1.56	1.43
1000	3.09	2.44	2.04	1.78	1.59	1.45

Two points of interest are that (1) as k_u/k_0 increases for a given n, the optimal estimate increases, and (2) as n increases for a given ratio k_u/k_0, the optimal estimate moves closer to the mean (in Proposition 4.1, we prove slightly more general assertions). The first result is intuitively obvious; all other things being equal, a higher ratio causes the decision maker to increase $\boldsymbol{a}$ in order to reduce the relatively higher expected loss of underestimation. The second result indicates that as n increases, he will reduce the amount by which he shifts $\boldsymbol{a}$ away from the mean. This is because as n increases, the seriousness of errors of more than one unit also increases. If the decision maker shifts $\boldsymbol{a}$ far to the right of the mean in order to reduce the expected loss of underestimation, the corresponding increase in expected loss of overestimation increases with increasing n. In terms of the utility example presented at the beginning of this section, the concavity of U when $n>1$ (indicating a risk-avoider) shifts the optimal estimate closer to the mean than it would be with a linear utility function.

Table 1 can also be used to find optimal estimates if θ is normally distributed with mean μ and variance σ^2. An estimate $\boldsymbol{a}$ of θ is optimal if and only if $\boldsymbol{a}^*=(\boldsymbol{a}-\mu)/\sigma$ is an optimal estimate of $\theta^*=(\theta-\mu)/\sigma$. But $f(\theta^*)$ is a standard normal distribution, so $\boldsymbol{a}^*$ can be read from Table 1, and $\boldsymbol{a}=\mu+\sigma\boldsymbol{a}^*$. In general, then, Table 1 gives the optimal estimate in standardized units (standard deviations from the mean) corresponding to various values of k_u/k_0 and n. For example, suppose that $n=3$, $k_0=14$, $k_u=2$, $\mu=100$, and $\sigma=10$. Since $k_u/k_0=1/7<1$, we look in Table 1 under the reciprocal loss ratio, 7, and find the value 0.61. Thus, $\boldsymbol{a}^*=-0.61$, and the optimal estimate is $\boldsymbol{a}=\mu+\sigma\boldsymbol{a}^*=93.9$.

A byproduct of the use of (4.1) is a table of partial and complete moments about an arbitrary point $\boldsymbol{a}$ for various values of $\boldsymbol{a}$. Due to lack of space, the partial moments $E^{\boldsymbol{a}}_{-\infty}(\boldsymbol{a}-\theta)^n$ are presented in Table 2 only for $n=1(1)6$ and for $\boldsymbol{a}=0(0.10)3$. These partial moments are useful in computing the expected loss of the optimal $\boldsymbol{a}$ (or the expected loss associated with nonoptimal values of $\boldsymbol{a}$, as we will discuss in Section 6).

It should be noted that the procedure used to develop Tables 1 and 2 is only applicable if n is a positive integer. An alternative method, one which can be applied for any positive n, is to use numerical integration to evaluate (4.1). To check the accuracy of the numerical procedure, it was carried out for $\boldsymbol{a}=0(0.01)5$ and $n=1(1)6$, with the interval $[-5, 5]$ divided into 1200 subintervals. The results agree with Table 1. However, for larger values of $\boldsymbol{a}$ (i.e., for any n, those values leading to a loss ratio greater than 1,000), the numerical procedure was quite inaccurate.

In this section we have considered the case in which $r=s=n$. If $r\neq s$, a search algorithm can be used to find the optimal value of $\boldsymbol{a}$ (e.g., see [14], Ch. 2). To simplify matters, Table 1 can be used to find either a lower bound or an upper bound for the optimal estimate. Assume that $r\geqq 1$ and $s\geqq 1$, and let $\boldsymbol{a}(r, s, t)$ denote an optimal estimate for given r, s, and $t=k_u/k_0$.

Table 2. Partial Moments $E^{a}_{-\infty}(a-\theta)^{n}$

a	n					
	1	2	3	4	5	6
0.0	.3989	.5000	.7979	1.5000	3.1915	7.5000
0.1	.4509	.5849	.9604	1.8508	4.0265	9.6567
0.2	.5069	.6806	1.1499	2.2719	5.0540	12.3703
0.3	.5668	.7879	1.3699	2.7748	6.3121	15.7676
0.4	.6304	.9076	1.6239	3.3724	7.8446	19.9996
0.5	.6978	1.0404	1.9158	4.0790	9.7026	25.2461
0.6	.7687	1.1870	2.2495	4.9106	11.9444	31.7194
0.7	.8429	1.3481	2.6294	5.8847	14.6369	39.6695
0.8	.9202	1.5243	3.0599	7.0208	17.8561	49.3890
0.9	1.0004	1.7163	3.5455	8.3400	21.6882	61.2193
1.0	1.0833	1.9247	4.0913	9.8653	26.2304	75.5568
1.1	1.1686	2.1498	4.7020	11.6217	31.5920	92.8596
1.2	1.2561	2.3923	5.3829	13.6363	37.8951	113.6554
1.3	1.3455	2.6524	6.1392	15.9381	45.2761	138.5493
1.4	1.4367	2.9306	6.9761	18.5583	53.8863	168.2325
1.5	1.5293	3.2272	7.8993	21.5305	63.8931	203.4920
1.6	1.6232	3.5424	8.9143	24.8901	75.4813	245.2203
1.7	1.7183	3.8765	10.0267	28.6749	88.8540	294.4263
1.8	1.8143	4.2298	11.2421	32.9251	104.2338	352.2464
1.9	1.9111	4.6023	12.5665	37.6831	121.8637	419.9567
2.0	2.0085	4.9942	14.0054	42.9936	142.0089	498.9858
2.1	2.1065	5.4057	15.5649	48.9035	164.9572	590.9279
2.2	2.2049	5.8368	17.2508	55.4624	191.0206	697.5573
2.3	2.3037	6.2877	19.0690	62.7219	220.5364	820.8429
2.4	2.4027	6.7583	21.0254	70.7360	253.8682	962.9640
2.5	2.5020	7.2488	23.1260	79.5614	291.4076	1126.3262
2.6	2.6015	7.7591	25.3767	89.2569	333.5747	1313.5785
2.7	2.7011	8.2894	27.7835	99.8836	380.8197	1527.6311
2.8	2.8008	8.8396	30.3523	111.5053	433.6241	1771.6738
2.9	2.9005	9.4097	33.0892	124.1879	492.5018	2049.1945
3.0	3.0004	9.9998	36.0002	137.9999	558.0002	2363.9998

PROPOSITION 4.1. $a(r, s, t)$ is a monotone increasing function of t for fixed r and s, a monotone decreasing function of r for fixed s and t, and a monotone increasing function of s for fixed r and t.

Let

(4.5)
$$q=\frac{E^{a}_{-\infty}(a-\theta)^{r-1}}{E^{\infty}_{a}(\theta-a)^{s-1}} .$$

From (3.8), an optimal a is such that

$$q=st/r .$$

Thus,

$$dq = -\frac{st}{r^2}dr + \frac{t}{r}ds + \frac{s}{r}dt \ .$$

But

$$dq = \frac{dq}{d\boldsymbol{a}}d\boldsymbol{a} = \frac{dq}{d\boldsymbol{a}}\left[\frac{\partial \boldsymbol{a}}{\partial r}dr + \frac{\partial \boldsymbol{a}}{\partial s}ds + \frac{\partial \boldsymbol{a}}{\partial t}dt\right] ,$$

so we have

$$\frac{\partial \boldsymbol{a}}{\partial r} = -\frac{st}{r^2}\frac{dq}{d\boldsymbol{a}} , \qquad \frac{\partial \boldsymbol{a}}{\partial s} = \frac{t}{r}\frac{dq}{d\boldsymbol{a}} \qquad \text{and} \qquad \frac{\partial \boldsymbol{a}}{\partial t} = \frac{s}{r}\frac{dq}{d\boldsymbol{a}} ,$$

where $dq/d\boldsymbol{a}$ is evaluated at $\boldsymbol{a}(r, s, t)$. Since $dq/d\boldsymbol{a}>0$ (from (4.5) and earlier continuity assumptions), $r \geqq 1$, $s \geqq 1$, and $t>0$, the proof of the proposition is completed. It should be noted that in the general case given by (3.1), if L_0 and L_u are multiplied by positive constants k_0 and k_u, then the optimal $\boldsymbol{a}$ is a monotone increasing function of $t=k_u/k_0$.

COROLLARY. *For a given loss ratio $k_u/k_0=t$, $\boldsymbol{a}(r, r, t)$ and $\boldsymbol{a}(s, s, t)$ are lower bounds for $\boldsymbol{a}(r, s, t)$ if $r \leqq s$ and upper bounds for $\boldsymbol{a}(r, s, t)$ if $r \geqq s$.*

The results in Table 1 indicate that $\boldsymbol{a}(r, r, t)$ is a tighter lower or upper bound than $\boldsymbol{a}(s, s, t)$ if $t>1$ and $\boldsymbol{a}(s, s, t)$ is a tighter lower or upper bound than $\boldsymbol{a}(r, r, t)$ if $t<1$. For example, suppose that $k_u/k_0=2$, $\mu=0$, $\sigma=1$, $r=2$, and $s=3$. From Table 1 and Proposition 4.1, the optimal estimate must be greater than 0.28. The expected loss associated with any estimate $\boldsymbol{a}$ is

$$EL(\boldsymbol{a}) = k_0\sigma^r E_{-\infty}^{\boldsymbol{a}^*}(\boldsymbol{a}^*-\theta^*)^r + k_u\sigma^s E_{\boldsymbol{a}^*}^{\infty}(\theta^*-\boldsymbol{a}^*)^s , \tag{4.6}$$

where $\boldsymbol{a}^*=(\boldsymbol{a}-\mu)/\sigma$, $\theta^*=(\theta-\mu)/\sigma$, and the partial expectations are taken with respect to the standard normal distribution. Using the tables of partial and complete moments discussed above and the fact that $E(x) = E_{-\infty}^{\boldsymbol{a}}(x)+E_{\boldsymbol{a}}^{\infty}(x)$, $EL(\boldsymbol{a})$ can be determined for any value of $\boldsymbol{a}$. Searching the region above 0.28 yields an optimal value of 0.46. Similarly, suppose that the loss ratio is still 2 but that $r=3$ and $s=2$. An upper bound is $\boldsymbol{a}(3, 3)=0.22$, and a search yields an optimal estimate of 0.02. Note that for this last example, it would not even be possible to predict whether $\boldsymbol{a}$ would be positive or negative. A loss ratio greater than one suggests that $\boldsymbol{a}$ should be positive, but $r>s$ suggests that $\boldsymbol{a}$ should be negative; here the result is very close to zero, although this will not always be the case.

5. Optimal estimates for exponential loss functions and the normal distribution

In this section we discuss the determination of optimal estimates when the loss function is of the form (3.9) with $r=s=n$ and the distribution $f(\theta)$ is a normal distribution. For $r=s=n$, (3.10) becomes

$$\frac{k_u}{k_0}=\left[\frac{E^{a}_{-\infty}(e^{-n\theta})}{E^{\infty}_{a}(e^{n\theta})}\right]e^{2an} . \tag{5.1}$$

Since

$$E^{a}_{-\infty}(e^{t\theta})=e^{t^2/2}F_{N^*}(a-t) , \tag{5.2}$$

we can rewrite (5.1) as follows:

$$\frac{k_u}{k_0}=\left[\frac{F_{N^*}(a+n)}{1-F_{N^*}(a-n)}\right]e^{2an} . \tag{5.3}$$

For any choice of a, the right-hand side of (5.3) can be evaluated, resulting in the loss ratio k_u/k_0 for which a is optimal. Thus, a table of optimal values can be determined, just as in the case of the power loss functions. This was done for $n=-1.5(.5)1.5$ and $a=0(0.01)5$; the results are summarized in Table 3. Once again, as k_u/k_0 increases for a given n, the optimal estimate increases. Also, as n increases for a given ratio k_u/k_0, the optimal estimate moves closer to the mean.

Some of the comments made in Section 4 concerning power loss functions also apply if L is exponential. First, just as for Table 1, Table 3 can be used to find optimal estimates if θ is normally distributed with mean μ and variance σ^2. Second, since (5.3) holds for any value of n (not just integer values), it is unnecessary to use numerical integration when working with exponential loss functions and the normal distribution, since (5.3) is both easier to use and more accurate. Third, if $r>0$, $s>0$, and $r\neq s$, Proposition 4.1 still holds, and its corollary can be used to find a bound for the optimal estimate. Such a situation would occur with the loss function given in terms of money by

$$L(a,\theta)=\begin{cases} c_0(a-\theta) & \text{if } a\geqq\theta , \\ c_u(\theta-a) & \text{if } a\leqq\theta , \end{cases}$$

and the utility function for money given by

$$U(x)=-e^{-\lambda x} .$$

The loss function in terms of utility is then of the form (3.9) with $k_0=k_u=e^{-\lambda w}$, $r=\lambda c_0$, and $s=\lambda c_u$, where w denotes current wealth. The expected loss associated with any estimate a is

Table 3. Optimal Values of a for the Standard Normal Distribution and Exponential Loss Functions with $r=s=n$

k_u/k_0	n					
	−1.5	−1.0	−0.5	0.5	1.0	1.5
1.0	0.00	0.00	0.00	0.00	0.00	0.00
1.1	0.11	0.09	0.07	0.05	0.04	0.03
1.2	0.21	0.17	0.14	0.09	0.07	0.06
1.3	0.30	0.25	0.20	0.13	0.10	0.08
1.4	0.38	0.32	0.26	0.17	0.13	0.10
1.5	0.46	0.38	0.31	0.20	0.16	0.12
1.6	0.53	0.44	0.36	0.23	0.18	0.14
1.7	0.60	0.50	0.41	0.26	0.21	0.16
1.8	0.66	0.55	0.45	0.29	0.23	0.18
1.9	0.72	0.60	0.49	0.32	0.25	0.20
2.0	0.78	0.65	0.53	0.34	0.27	0.21
2.5	1.01	0.85	0.70	0.45	0.35	0.28
3.0	1.20	1.01	0.83	0.54	0.42	0.33
4	1.47	1.25	1.03	0.67	0.53	0.42
5	1.68	1.42	1.18	0.78	0.62	0.49
6	1.83	1.56	1.30	0.86	0.68	0.54
7	1.96	1.67	1.40	0.93	0.74	0.59
8	2.06	1.76	1.48	0.99	0.79	0.63
9	2.15	1.84	1.55	1.04	0.83	0.66
10	2.24	1.92	1.62	1.09	0.87	0.70
12	2.37	2.03	1.72	1.16	0.93	0.75
14	2.47	2.13	1.80	1.23	0.99	0.79
16	2.56	2.21	1.87	1.28	1.04	0.83
18	2.63	2.28	1.94	1.33	1.08	0.86
20	2.70	2.34	1.99	1.37	1.11	0.89
25	2.83	2.46	2.10	1.46	1.19	0.96
30	2.93	2.55	2.19	1.53	1.25	1.01
35	3.02	2.63	2.26	1.59	1.30	1.05
40	3.09	2.70	2.33	1.64	1.35	1.09
45	3.15	2.76	2.38	1.68	1.38	1.12
50	3.20	2.81	2.43	1.72	1.42	1.15
60	3.29	2.89	2.50	1.79	1.48	1.20
70	3.37	2.96	2.57	1.84	1.52	1.25
80	3.43	3.02	2.62	1.89	1.57	1.28
90	3.48	3.07	2.67	1.93	1.60	1.32
100	3.53	3.12	2.72	1.97	1.64	1.35
150	3.70	3.28	2.87	2.10	1.76	1.45
200	3.82	3.39	2.98	2.20	1.84	1.53
250	3.90	3.48	3.06	2.27	1.91	1.59
300	3.97	3.54	3.12	2.32	1.96	1.63
350	4.03	3.60	3.17	2.37	2.01	1.67
400	4.08	3.65	3.22	2.41	2.04	1.71
450	4.12	3.69	3.26	2.45	2.08	1.74
500	4.16	3.72	3.29	2.48	2.10	1.76
600	4.22	3.78	3.35	2.53	2.16	1.81
700	4.28	3.84	3.40	2.58	2.20	1.85
800	4.32	3.88	3.45	2.62	2.23	1.88
900	4.36	3.92	3.48	2.65	2.26	1.91
1000	4.40	3.95	3.52	2.69	2.30	1.94

(5.4) $$EL(a)=k_0E^{a}_{-\infty}|e^{r(a-\theta)}-1|+k_uE^{\infty}_{a}|e^{s(\theta-a)}-1| .$$

If θ is normally distributed with mean zero and variance one, we can apply (5.2) to get

$$(5.5) \qquad EL(a)=k_0\left|\exp\left(ar+\frac{r^2}{2}\right)F_{N^*}(a+r)-F_{N^*}(a)\right|$$
$$+k_u\left|\exp\left(-as+\frac{s^2}{2}\right)[1-F_{N^*}(a-s)]-[1-F_{N^*}(a)]\right| .$$

6. Sensitivity analysis

An important point which has not been discussed is the sensitivity of Bayesian point estimates to misspecification in the loss function. For instance, how much does the decision maker stand to lose by using the mean of his distribution as an estimate when in fact some other value is optimal? Using formulas such as (4.6), the appropriate expected losses can be computed and compared.

For example, suppose that θ is normally distributed with mean 100 and variance 100 and that the appropriate loss function is a power loss function with $r=s=n=3$, $k_0=14$, and $k_u=2$. In Section 4 we found the optimal estimate for this problem to be 93.9. Using (4.6),

$$EL(93.9)=7{,}762 .$$

The expected loss associated with $a=\mu=100$ ($a^*=0$) is

$$EL(100)=12{,}768 .$$

Thus, using the mean instead of the optimal estimate causes the decision maker to suffer an increase in expected loss of $12{,}768-7{,}762=5{,}006$, or about 64%.

Suppose that the decision maker realizes that $n=3$ and $k_u/k_0=1/7$, but to simplify his problem he acts as though $n=1$ (linear loss). This means that he simply finds the $k_u/(k_u+k_0)=1/8$ fractile of his distribution, which is 88.5. Thus, his expected loss is

$$EL(88.5)=10{,}934 .$$

This is not as bad as using the mean, but it still represents an increase in expected loss of 3,172, or about 41%, in comparison with the optimal estimate.

Using this approach, the percentage increase in expected loss corresponding to certain entries in Table 1 was computed, (1) using the mean of the distribution instead of the optimal a and (2) using the optimal value for $n=1$ instead of the true n. The results are presented in Tables 4 and 5. For instance, if $k_u/k_0=16$ and $n=4$, EL increases by 141% if the mean is used and by 126% if the correct loss ratio is used but it is assumed that $n=1$. In general, if the mean is used, there is a greater increase in EL (as compared with the EL of the optimal

Table 4. Percentage Increase in Expected Loss for Power Loss Functions and the Normal Distribution if $a=\mu$ is used Instead of the Optimal Value of a

k_u/k_0	n = 1	2	3	4	5	6
1.0	0	0	0	0	0	0
1.2	0.7	0.5	0.5	0.5	0.5	0.5
1.4	2.2	1.8	1.7	1.6	1.6	1.5
1.6	4.4	3.6	3.3	3.1	3.1	3.0
1.8	6.9	5.6	5.2	4.9	4.8	4.7
2.0	9.7	7.8	7.2	6.9	6.7	6.6
2.5	17	14	13	12	12	12
3	26	20	19	18	17	17
4	42	33	30	29	28	28
6	77	59	53	51	49	48
8	111	84	75	71	68	67
10	144	107	95	90	86	84
20	302	214	185	172	164	159
30	452	310	263	241	229	220
40	597	399	334	304	286	275
60	876	565	463	416	388	371
80	1146	719	581	516	479	456
100	1410	867	691	609	563	533

Table 5. Percentage Increase in Expected Loss for Power Loss Functions and the Normal Distribution if it is assumed that $n=1$

k_u/k_0	n = 2	3	4	5	6
1.0	0	0	0	0	0
1.2	0.1	0.4	0.7	1.1	1.5
1.4	0.6	1.6	2.9	4.3	5.8
1.6	1.0	3.0	5.4	8.1	11
1.8	1.8	5.1	9.1	13	18
2.0	2.3	6.7	12	18	24
2.5	4.1	12	21	32	43
3	4.9	15	27	41	56
4	8.0	24	44	67	93
6	12	37	68	107	152
8	15	45	85	136	197
10	16	51	98	158	231
20	22	70	140	235	359
30	23	77	158	270	422
40	25	83	173	302	479
60	26	88	188	333	540
80	27	93	200	361	593
100	27	95	205	374	620

Table 6. Percentage Increase in Expected Loss for Exponential Loss Functions and the Normal Distribution if $\boldsymbol{a}=\mu$ is used Instead of the Optimal Value of $\boldsymbol{a}$

k_u/k_0	n					
	−1.5	−1.0	−0.5	0.5	1.0	1.5
1.0	0	0	0	0	0	0
1.2	1.0	0.7	0.8	0.8	0.8	0.5
1.4	3.4	3.7	2.6	2.2	1.7	1.6
1.6	6.7	6.6	6.2	4.2	3.7	3.1
1.8	11	9.8	8.0	7.0	5.5	4.8
2.0	15	13	10	9.0	7.4	6.8
2.5	27	24	18	15	13	12
3	41	34	28	21	18	17
4	67	58	50	36	32	28
6	124	107	91	65	56	49
8	182	158	133	92	78	69
10	241	208	170	118	99	87
20	535	459	378	240	194	164
30	830	709	577	350	276	229
40	1128	958	773	453	351	288
60	1719	1455	1159	648	488	388
80	2308	1949	1539	830	606	479
100	2898	2442	1917	1005	728	564

estimate) as n decreases and as the loss ratio k_u/k_0 increases. If it is assumed that $n=1$, there is a greater increase in EL as n increases and k_u/k_0 increases, although for large values of k_u/k_0 the percentage increase in EL seems to level off somewhat.

For the exponential loss functions, we can use (5.5) to determine the expected loss of any $\boldsymbol{a}$. For the optimal estimates given in Table 3, the percentage increase in EL was computed under the assumption that the mean was used instead of the optimal estimate. The results are presented in Table 6. For example, if $k_u/k_0=10$ and $n=1.5$, EL increases by 87% if the mean is used rather than the optimal $\boldsymbol{a}=0.70$.

The results of the sensitivity analysis indicate that it may be quite costly to use nonoptimal estimates. This, of course, depends on the particular loss function and distribution and on the particular nonoptimal estimate that is chosen. The question of the sensitivity of EL to the use of nonoptimal estimates is, after all, a relative matter. In some situations, the cost of determining the optimal estimate may be greater than the potential savings unless tables such as Table 1 and Table 3 are available.

7. Summary and discussion

In this paper we have investigated the properties of Bayesian point estimates under loss functions other than the simple linear and quadratic loss functions. The results of Section 2 strengthen the contention that the mean is a good "all-purpose" estimator *in the symmetric case* by demonstrating that it is optimal (in the sense of minimizing expected loss) for a wide class of symmetric loss functions and distributions. Although there are many situations in which both the loss function and the distribution are symmetric, violations of one or both of these conditions are quite common. As a result, the case in which no symmetry assumptions are made is of much interest. In this case, it is convenient if the expected loss, $EL(\boldsymbol{a})$, can be shown to be a convex function of $\boldsymbol{a}$; Proposition 3.1 indicates that EL is convex if it converges and L is convex. The convexity of EL means that any local optimum is a global optimum, thus simplifying the determination of an optimal estimate. Some necessary conditions for optimality are given in Propositions 3.2 and 3.3.

Using the formulas developed for the general case, numerical results are presented for two special cases: power and exponential loss functions and the normal distribution (these results are also applicable for certain "modified" power and exponential loss functions similar to the "modified" linear and quadratic functions discussed in [9]). For the power loss functions, if the overestimation and underestimation loss functions involve the same power (i.e., if $r=s=n$), then the optimal estimate can be determined from Table 1 for $n=1(1)6$. Of course, numerical methods could be used to evaluate (4.1) for any $n>0$. This involves the calculation of certain partial and complete moments involving $\boldsymbol{a}$ and θ. For the exponential loss functions with $r=s=n$, optimal values of $\boldsymbol{a}$ for $n=-1.5(.5)1.5$ are given in Table 3. For a given loss ratio k_u/k_0, the Bayesian point estimate moves closer to the mean as n increases. For a given n, the estimate increases as the loss ratio increases; when $k_u/k_0=1$, the mean is optimal. When $r\neq s$, the determination of $\boldsymbol{a}$ is slightly more difficult; the corollary to Proposition 4.1 gives a bound for $\boldsymbol{a}$, and this can be used as a starting point for a search procedure.

The results of the sensitivity analysis presented in Section 6 indicate that in some situations it may be quite costly to use nonoptimal estimates. This suggests that it is important to carefully specify the loss function for any given Bayesian point estimation problem and to determine the optimal estimate under that loss function, using tables such as Tables 1 and 3 when applicable. The use of the mean as a general "all-purpose" estimator may result in a large increase in EL (as compared with the EL under the optimal estimate) if the loss function is asymmetric. As we have indicated, there are many situations in which symmetric loss functions or even simple asymmetric loss functions such as the linear loss functions of the form (1.3) are inappropriate, particularly when the decision maker's utility function for money is not linear. In this paper we have determined optimal estimates for two rich classes of loss functions (the power and exponential loss functions, which can be viewed as combinations of linear loss and nonlinear utility) and the normal distribution; also, we have discussed how optimal estimates can be determined for other loss functions and distributions.

UNIVERSITY OF WESTERN ONTARIO
INDIANA UNIVERSITY

REFERENCES

[1] Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities, *Proc. Amer. Math. Soc.*, **6**, 170-176.

[2] Baron, D. P. (1968). Demand Uncertainty and the Theory of the Firm, Bloomington, Indiana University, unpublished doctoral dissertation.

[3] De Groot, M. H. and Rao, M. M. (1963). Bayes estimation with convex loss, *Ann. Math. Statist.*, **34**, 839-846.

[4] De Groot, M. H. and Rao, M. M. (1966). Multidimensional Information Inequalities and Prediction, in P. Krishnaiah, ed., *Multivariate Analysis*, Academic Press, New York.

[5] Deutsch, R. (1965). *Estimation Theory*, Prentice-Hall, Englewood Cliffs, N.J.

[6] Gould, J. P. (1969). The expected utility hypothesis and the selection of optimal deductibles for a given insurance policy, *J. Bus.*, **42**, 143-151.

[7] Granger, C. W. J. (1969). Prediction with a generalized cost of error function, *Operat. Res. Quart.*, **20**, 199-207.

[8] Marshall, A. W. and Olkin, I. (1968). A general approach to some screening and classification problems, *J. Roy. Statist. Soc.*, B, **30**, 407-435.

[9] Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*, Graduate School of Business Administration, Harvard University, Boston.

[10] Savage, L. J. (1954). *The Foundations of Statistics*, Wiley, New York.

[11] Sherman, S. (1955). A theorem on Convex sets with applications, *Ann. Math. Statist.*, **26**, 763-767.

[12] Sherman, S. (1958). Non mean-square error criteria, *I.R.E. Transactions on Information Theory*, IT-4, 125-126.

[13] *Tables of Normal Probability Functions* (National Bureau of Standards, Applied Mathematics Series, Vol. 23), (1953), U.S. Government Printing Office, Washington.

[14] Wilde (1964). *Optimum Seeking Methods*, Prentice-Hall, Englewood Cliffs, N.J.

[15] Winkler, R. L., Roodman, G. M. and Britney, R. R. (1972). The determination of partial moments, *Management Science*, **19**, 290-296.