Journal of Statistical Physics, Vol. 7, No. 3, 1973

A Class of Loss Functions of Catenary Form

D. E. Raeside[1,2] and R. J. Owen[1,3]

Received June 18, 1971; revised June 12, 1972

The catenary form of loss function is considered in the framework of Bayesian decision theory. The mathematical tractability of this form seems to be unrecognized; it contains quadratic loss as a limiting case. For various probability distributions expressions are given for posterior analysis, and limiting properties are investigated.

KEY WORDS: Parameter estimation; Bayesian estimation theory; loss functions; non-mean-square error criterion; quadratic loss function.

1. INTRODUCTION

In the Bayesian approach to statistics [(1–4)] linear and quadratic loss functions have been widely discussed and their engineering applications well treated [(5–8)]. In addition, a limited literature characterizing broad classes of loss functions has also appeared[(9–11)]. The purpose of this note is to give results on a specific one-parameter family of loss functions.

In this note a quantity thought of as a random variable will be denoted by a capital letter, the lower case form of the letter being reserved for a realization or fixed value of that random variable.

Let W denote the unknown (scalar) parameter of interest and $F(\cdot)$ the

[1] The University of Michigan, Ann Arbor, Michigan.
[2] Present address: Department of Radiological Sciences, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma.
[3] Present address: Department of Statistics, University College of Wales, Penglais, Aberystwyth, Wales, U. K.

cumulative distribution function for W (this may be a prior or a posterior distribution, depending on the context). It is assumed that the moment generating function of W,

$$M(t) \equiv \int_{-\infty<w<\infty} e^{wt}\, dF(w) \tag{1}$$

exists in some interval containing 0, say $-\alpha < t < \beta$ ($\alpha, \beta > 0$). Leibniz's rule for differentiation under the integral sign (extended to Stieltjes integrals) applies in the same interval (Ref. 12, p. 240); hence, for $r = 0, 1,\ldots$ the r^{th} derivative of $M(t)$, $M^{(r)}(t)$, exists and is given by

$$M^{(r)}(t) = \int_{-\infty<w<\infty} w^r e^{wt}\, dF(w) \qquad (-\alpha < t < \beta) \tag{2}$$

In particular, for $r = 0, 1,\ldots$ the r^{th} moment of W, μ_r', exists and equals $M^{(r)}(0)$. The r^{th} ($r = 0, 1,\ldots$) central moment of W, $E_W\{(W - \mu_1')^r\}$, will be denoted by μ_r. The alternative notation μ and σ^2 will be used for μ_1' and μ_2, respectively.

It will be more convenient to work with the moment generating function of $W - \mu$:

$$N(t) = e^{-\mu t} M(t) \tag{3}$$

Of course the derivatives of $N(t)$, like those of $M(t)$, exist for $-\alpha < t < \beta$.

The loss function to be considered is

$$L(w, d) = [\cosh a(w - d)] - 1 \tag{4}$$

where a is a specified positive constant. Notice that the function of x, $(\cosh ax) - 1$, is nonnegative (vanishing only at $x = 0$), strictly increasing in $|x|$, symmetric about $x = 0$, and strictly convex.

In the usual Bayesian decision-theoretic framework let ρ denote the Bayes risk,

$$\rho \equiv \inf_d E_W\{L(W, d)\} \tag{5}$$

and $\hat{d}$ a value of d (if it exists) which attains this infimum; $\hat{d}$ is called a Bayes estimator of w.

2. POSTERIOR ANALYSIS

Taking the usual differentiation approach for minimization and using (2) for $r = 1, 2$, it follows that for $a < \min(\alpha, \beta)$ the Bayes estimator $\hat{d}$ for the loss function (4) is unique and

$$\hat{d} = \mu + (1/2a) \ln[N(a)/N(-a)], \qquad \rho = [N(a)N(-a)]^{1/2} - 1 \tag{6}$$

In the case where the distribution of W is symmetric about the mean, $N(a) = N(-a)$ and (6) simplifies to

$$\hat{d} = \mu, \qquad \rho = N(a) - 1 \tag{7}$$

The first of these results, simply that $\hat{d} = \mu$, is in agreement with the result in Ref. 9 for a convex, symmetric loss function and a distribution symmetric about its mean.

These expressions are now evaluated for various distributions of W which are commonly utilized as prior distributions. In what follows $f(\cdot)$ denotes a probability density function and $p(\cdot)$ a probability mass function.

2.1. Normal

$$f(w) = (h/2\pi)^{1/2} \exp[-\tfrac{1}{2}h(w-\mu)^2],$$
$$-\infty < w < \infty \qquad (-\infty < \mu < \infty, \quad h > 0)$$
$$\hat{d} = \mu, \qquad \rho = \exp(\tfrac{1}{2}a^2h^{-1}) - 1$$

2.2. Uniform

$$f(w) = \tfrac{1}{2}\theta^{-1}, \qquad \mu - \theta < w < \mu + \theta \qquad (-\infty < \mu < \infty, \quad \theta > 0)$$
$$\hat{d} = \mu, \qquad \rho = (1/a\theta)(\sinh a\theta) - 1$$

2.3. Gamma

$$f(w) = [\lambda^r/\Gamma(r)]w^{r-1}e^{-\lambda w}, \qquad w > 0 \qquad (\lambda > 0, \quad r > 0)$$

Then for $a < \lambda$

$$\hat{d} = (r/2a)\ln[(\lambda + a)/(\lambda - a)], \qquad \rho = \lambda^r(\lambda^2 - a^2)^{-r/2} - 1$$

2.4. Beta (First Kind)

$$f(w) = [1/B(p, q)]w^{p-1}(1 - w)^{q-1}, \qquad 0 < w < 1 \qquad (p > 0, \quad q > 0)$$
$$\hat{d} = (1/2a)\ln[\Phi(p, p+q; a)/\Phi(p, p+q; -a)]$$
$$\rho = [\Phi(p, p+q; a)\Phi(p, p+q; -a)]^{1/2} - 1$$

Here we use the notation $\Phi(\cdot, \cdot; \cdot)$ to designate the degenerate hypergeometric function (Ref. 13, p. 1058).

2.5. Binomial

$$p(w) = \binom{n}{w} p^w q^{n-w}, \qquad w = 0, 1, \ldots, n$$

$$(0 < p < 1, \quad q \equiv 1 - p, \quad n = 1, 2, \ldots)$$

$$\hat{d} = (n/2a) \ln[(pe^a + q)/(pe^{-a} + q)]$$

$$\rho = [1 + 4pq \sinh^2(a/2)]^{n/2} - 1$$

2.6. Negative Binomial

$$p(w) = \binom{r + w - 1}{w} p^r q^w, \qquad w = 0, 1, \ldots$$

$$(0 < p < 1, \quad q \equiv 1 - p, \quad r = 1, 2, \ldots)$$

Then for $a < \ln(1/q)$

$$\hat{d} = (r/2a) \ln[(1 - qe^{-a})/(1 - qe^a)]$$

$$\rho = [p^2/(1 - qe^{-a})(1 - qe^a)]^{r/2} - 1$$

2.7. Poisson

$$p(w) = \lambda^w e^{-\lambda}/w!, \qquad w = 0, 1, \ldots \qquad (\lambda > 0)$$

$$\hat{d} = (\lambda/a) \sinh a, \qquad \rho = \exp[\lambda(\cosh a - 1)] - 1$$

3. ASYMPTOTIC PROPERTIES

Here it is shown that asymptotically as $a \to 0$, the loss function (4) "behaves like" the quadratic loss function

$$L_1(w, d) = \tfrac{1}{2}a^2(w - d)^2 \tag{8}$$

In what follows the symbols $\sim$ and O will have their usual meaning. Denote the Bayes estimator and Bayes risk associated with L_1 by $\hat{d}_1$ and ρ_1, respectively. It is well known that

$$\hat{d}_1 = \mu \qquad \text{and} \qquad \rho_1 = \tfrac{1}{2}a^2\sigma^2 \tag{9}$$

Referring to (6), it follows by l'Hôspital's rule [the conditions for which are satisfied because of (2)] that

$$\hat{d} \to \hat{d}_1 \qquad \text{as} \qquad a \to 0 \tag{10}$$

Referring to (A.9), (A.10), and (A.12), it may be asserted by Taylor's theorem that for any γ_1 between 0 and γ and any a such that $0 \leqslant |a| \leqslant \gamma_1$ there is some θ between 0 and a such that

$$\rho(a) = \tfrac{1}{2}a^2\mu_2 + (1/4!)\, a^4\rho^{(4)}(\theta) \qquad (0 \leqslant |a| \leqslant \gamma_1 < \gamma) \tag{11}$$

Here θ is a function of a and of the datum x. Regarding $\rho^{(4)}(\theta)$ as a function of a for given x, it follows from (A.12) and (A.10) that $\rho^{(4)}(\theta)$ is bounded on $0 \leqslant |a| \leqslant \gamma_1$ and tends to μ_4 as $a \to 0$. Therefore

$$\rho = \tfrac{1}{2}a^2\mu_2 + O(a^4) \qquad \text{and} \qquad (\rho - \rho_1)/[\tfrac{1}{2}(\rho + \rho_1)] \sim \tfrac{1}{12}a^2\mu_4/\mu_2 \tag{12}$$

It is of interest to note that stemming from the inequality $1 + \frac{1}{2}x^2 \leqslant \cosh x \leqslant 1 + \frac{1}{2}x^2 + (1/24)\, x^3 \sinh x$ (which is true for all x),

$$0 \leqslant \frac{\rho - \rho_1}{\frac{1}{2}(\rho + \rho_1)} \leqslant \frac{a}{12\mu_2} \int_{-\infty < w < \infty} (w - \mu)^3 \sinh a(w - \mu)\, dF(w) \equiv \epsilon(a) \tag{13}$$

$\epsilon(a)$ may be expressed in terms of $N(\cdot)$:

$$\epsilon(a) = (a/24\mu_2)[N^{(3)}(a) - N^{(3)}(-a)] \tag{14}$$

This expression may be used to calculate expressions for $\epsilon(a)$, for example, when W has a normal distribution with mean μ and variance $\sigma^2 \equiv 1/h$,

$$\epsilon(a) = \tfrac{1}{12}a^2(3 + a^2\sigma^2)\sigma^2 \exp(\tfrac{1}{2}a^2\sigma^2)$$

Differentiation of (14) establishes that $\epsilon(a)$ is an increasing function of a and an expansion of $\epsilon(a)$ by Taylor's theorem establishes that

$$\epsilon(a) \sim \tfrac{1}{2}a^2\mu_4/\mu_2 \qquad \text{as} \quad a \to 0 \tag{15}$$

Notice that this is the same as the asymptotic form of $(\rho - \rho_1)/[\frac{1}{2}(\rho + \rho_1)]$ and so, denoting this relative difference by η, it is natural to be curious about the asymptotic behavior of $(\epsilon - \eta)/[\frac{1}{2}(\epsilon + \eta)]$. A Taylor expansion approach establishes that as $a \to 0$

$$(\epsilon - \eta)/[\tfrac{1}{2}(\epsilon + \eta)] \sim (1/15)\, a^2\mu_4^{-1}(2\mu_6 + 5\mu_3{}^2) \tag{16}$$

APPENDIX. THE DERIVATIVES OF ρ

This topic is of interest for a Taylor expansion of ρ and it is presented here to avoid a digression in the body of the article. The definitions and results up to and including Eq. (7) are assumed. Let

$$y \equiv y(a) \equiv N(a)N(-a) \qquad \text{so} \qquad \rho \equiv \rho(a) = y^{1/2} - 1 \tag{A.1}$$

Because of (2) all derivatives of y exist and are continuous for $|a| < \gamma$ where $\gamma = \min(\alpha, \beta)$. For a function f let $(\partial/\partial a)^r f$ be denoted by $f^{(r)}$; then it follows from Leibniz's rule for differentiating a product that for $n = 0, 1, \ldots$

$$y^{(n)}(a) = \sum_{r=0}^{n} \binom{n}{r} (-1)^r N^{(r)}(-a) N^{(n-r)}(a) \qquad (|a| < \gamma) \qquad \text{(A.2)}$$

Pairing off terms from the ends of the sum (A.2), it follows that

$$y^{(n)}(0) = 0 \qquad (n \quad \text{odd}) \qquad \text{(A.3)}$$

$$y^{(n)}(0) = 2 \sum_{r=0}^{(n/2)-1} (-1)^r \binom{n}{r} \mu_r \mu_{n-r} + \binom{n}{\frac{1}{2}n} (-1)^{n/2} \mu_{n/2}^2 \qquad (n \quad \text{even}) \qquad \text{(A.4)}$$

The derivatives of y will now be used to calculate the derivatives of ρ. The approach is to write (A.1) in the form

$$y = (1 + \rho)^2 \qquad \text{(A.5)}$$

from which is obtained, by Leibniz's rule, denoting $\partial/\partial a$ by D,

$$y^{(n)} = \sum_{r=0}^{n} \binom{n}{r} D^r(1 + \rho) D^{n-r}(1 + \rho) \qquad (|a| < \gamma) \qquad \text{(A.6)}$$

Pairing off terms from the ends, this reduces to (for $|a| < \gamma$)

$$y^{(n)} = 2 \sum_{r=0}^{(n-1)/2} \binom{n}{r} (1 + \rho)^{(r)}(1 + \rho)^{(n-r)} \qquad (n \quad \text{odd}) \qquad \text{(A.7)}$$

$$y^{(n)} = 2 \sum_{r=0}^{(n/2)-1} \binom{n}{r} (1 + \rho)^{(r)}(1 + \rho)^{(n-r)} + \binom{n}{\frac{1}{2}n} [(1 + \rho)^{n/2}]^2 \qquad (n \quad \text{even}) \qquad \text{(A.8)}$$

Using (A.3) and (A.7), it follows by induction that

$$\rho^{(n)}(0) = 0 \qquad (n \quad \text{odd}) \qquad \text{(A.9)}$$

Using (A.4), (A.8), and (A.9), the even derivatives are calculated sequentially and it is found that

$$\rho^{(2)}(0) = \mu_2, \qquad \rho^{(4)}(0) = \mu_4, \qquad \rho^{(6)}(0) = \mu_6 - 10\mu_3{}^2,$$
$$\rho^{(8)}(0) = 280\mu_2\mu_3{}^2 - 56\mu_3\mu_5 + \mu_8 \qquad \text{(A.10)}$$

When the distribution of W is symmetric about the mean, all odd central moments are zero and referring to (7) expanded as a Taylor series, there follows the simple solution for the even central moments

$$\rho^{(2r)}(0) = \mu_{2r} \qquad (r = 1, 2, \ldots) \tag{A.11}$$

This result may also be arrived at by induction from (A.4) and (A.8).

Equating the expressions for $y^{(n)}(a)$ in (A.2) and (A.6), it follows by an induction argument on n that

$$\rho^{(n)}(a) \text{ exists and is continuous,} \quad [n = 0, 1, \ldots; \quad (0 \leqslant |a| < \gamma)] \tag{A.12}$$

REFERENCES

1. M. H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, New York (1970).
2. I. H. LaValle, *An Introduction to Probability, Decision and Inference*, Holt, Rinehart and Winston, New York (1970).
3. D. V. Lindley, *Introduction to Probability and Statistics from a Bayesian Viewpoint; Part 1, Probability; Part 2, Inference*, Cambridge University Press, London (1965).
4. H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business Administration, Harvard University, Boston (1961).
5. C. W. Helstrom, *Statistical Theory of Signal Detection*, Pergamon Press, Oxford (1968).
6. H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Wiley, New York (1968).
7. A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, New York (1966).
8. R. Deutsch, *Estimation Theory*, Prentice-Hall, Englewood Cliffs, N. J. (1965).
9. M. H. De Groot and M. M. Rao, "Bayes Estimation with Convex Loss," *Ann. Math. Stat.* **34**:839–846 (1963).
10. S. Sherman, "Non-Mean-Square Error Criteria," *IRE Trans. Information Theory* IT-4:125–126 (1958).
11. M. Zahai, General Error Criteria, *IEEE Trans. Information Theory* **IT-10**:94–95 (1964).
12. D. V. Widder, *The LaPlace Transform*, Princeton University Press, Princeton N.J. (1946).
13. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York (1965).