Estimating the Loss of Estimators of a Binomial Parameter

by

Andrew L. Rukhin

Technical Report #86-12

Department of Statistics
Purdue University

April 1986

Research supported by the National Science Foundation Grant DMS-8401996.

Estimating the Loss of Estimators of a Binomial Parameter

by

Andrew L. Rukhin

Purdue University

A loss function is introduced, which combines the estimation error of a statistical procedure with a measure of its accuracy. The properties of this loss function are discussed and the problem of estimating a binomial parameter is considered.

Key words: binomial parameter, minimax estimator, posterior loss, unbiased estimator

1. Introduction

We start with the general statistical decision problem, as described by possible states of nature θ, decisions d, and a loss function $W(\theta,d)$. Classical decision theory advocates making some decision $d = \delta(x)$ where x is the observation, resulting in the frequentist risk $R(\theta,\delta)$. This approach has been often criticized for the lack of a data-dependent measure of the accuracy of the procedure used. Numerous examples of classical procedures are known which assign the same decision and the same numerical measure of accuracy (say, a constant confidence coefficient) for two different sample values, one of which seems intuitively much more informative than the other (see Kiefer (1977), Berger (1985a,b)). Thus there are many important situations where one would like to accompany the decision γ by an estimate $\gamma = \gamma(x)$ of its accuracy or of the loss function $W(\theta,\delta)$. Exactly this necessity of measuring the accuracy is behind the idea of confidence intervals (cf. Savage (1954), ch. 17). Indeed while the center of the confidence interval may represent a point estimate of the unknown parameter, its width gives a measure of the precision of this estimate.

To develop a decision-theoretic approach to the loss estimation problem, one must specify an appropriate loss function. "Any serious attempt to take account of the consequences of unreliability in not capturing the true parametric value and of lack of usefulness in excessive width should, we feel, involve the specification of some reasonable loss function and the subsequent examination of the problem in terms of decision theory" (Aitchison and Dunsmore (1968)). In fact a variety of loss functions in interval estimation has been considered (see Pratt (1971), Winkler (1972), Cohen and Strawderman (1973)). Other loss estimation problems are less studied. Lehmann (1959) mentions the estimated power of a test, and Sandred (1968) investigates unbiased estimators of quadratic loss in some estimation situations.

In this paper we give a loss function which combines the general decision problem error with an accuracy estimate. Namely, if γ is an estimator of the loss $W(\theta,\delta)$ with a non-negative loss function W, then

$$\mathcal{L}(\theta;\delta,\gamma) = W(\theta,\delta)\gamma^{-\frac{1}{2}} + \gamma^{\frac{1}{2}} \tag{1.1}$$

represents such combined loss.

The important feature of this loss function is that for a fixed δ, the unique minimum in γ is attained at $W(\theta,\delta)$. On the other hand for a fixed γ, this loss function is just a linear transform of $W(\theta,\delta)$, so that the Bayes procedure δ_B for $\mathcal{L}$ is the Bayes procedure for W. The Bayes estimator γ_B has the form

$$\gamma_B = E\{W(\theta,\delta_B)|x\}$$

i.e. γ_B coincides with the posterior loss of δ_B. This result allows one to interpret the posterior loss of the Bayes procedure δ_B as the Bayes estimator of the risk function of δ_B under (1.1). In the next section we illustrate further properties of the loss (1.1), in the problem of estimating a binomial parameter.

2. Estimators of a Binomial Parameter and Estimators of their Loss

Let X be a binomial random variable with unknown probability θ of success, $0 \leq \theta \leq 1$,

$$P_\theta(X = x) = \binom{n}{x} \theta^x(1-\theta)^{n-x}, \ x = 0,1,\dots,n.$$

We consider the problem of simultaneous estimation of θ by, say, an estimate δ and an estimate of the loss of δ, under the combined loss function (1.1), with $W(\theta,d) = (\theta-d)^2$.

If $\lambda(\theta) = \theta^{\alpha-1}\ (1-\theta)^{\beta-1}$ is a conjugate prior density, then it is well known that the Bayes estimator δ_B has the form

$$\delta_B(X) = (X + \alpha)/(n + \alpha + \beta).$$

An easy calculation shows that the Bayes estimator γ_B of the mean square error of δ_B under (1.1) (i.e. the posterior loss) is

$$\gamma_B(X) = (X + \alpha)(n + \beta - X)/[(n + \alpha + \beta)^2(n + \alpha + \beta + 1)].$$

In particular if $\alpha = \beta = 0$, the best unbiased estimator is

$$\delta_0(X) = X/n \ ,$$

and its risk estimate is

$$\delta_0(X) = X(n - X)([n^2(n+1)].$$

Notice that if $X = 0$ or $X = n$, then $\gamma_0(X) = 0$, i.e. the choice of an improper prior beta-density with $\alpha = \beta = 0$ leads to a rather silly estimate of the risk of δ_0. In fact, for all θ, $0 < \theta < 1$.

$$E_\theta \mathcal{L}(\theta, \delta_0, \gamma_0) = E_\theta[(\delta_0 - \theta)^2 \gamma_0^{-\frac{1}{2}} + \gamma_0^{\frac{1}{2}}] = \infty \tag{2.1}$$

so that the risk is infinite in the open interval (0,1). Thus the loss (1.1) seems to provide a tool for weeding out unreasonable accuracy estimators, even if they are generalized Bayes.

Another interesting case is $\alpha = \beta = 0.5n^{\frac{1}{2}}$, leading to the minimax estimator

$$\delta_m(X) = (X + 0.5n^{\frac{1}{2}})/(n + n^{\frac{1}{2}})$$

which is known to have a constant risk,

$$R(\theta,\delta_m) = E_\theta(\delta_m - \theta)^2 = 1/[4(n^{\frac{1}{2}} + 1)^2]. \tag{2.2}$$

The Bayes estimator γ_m of $R(\theta,\delta_m)$ is not constant,

$$\gamma_m(X) = (X + 0.5n^{\frac{1}{2}})(n + 0.5n^{\frac{1}{2}} - X)/[(n + n^{\frac{1}{2}})^2(n + n^{\frac{1}{2}} + 1)].$$

Because of (2.2), one may consider the constant estimator γ_1 of $R(\theta,\delta_m)$

$$\gamma_1(X) = 1/[4(n^{\frac{1}{2}} + 1)]^2$$

which has the risk

$$R(\theta, \delta_m, \gamma_1) = E_\theta[(\delta_m - \theta)^2\gamma_1^{-\frac{1}{2}} + \gamma_1^{\frac{1}{2}}] = 1/(n^{\frac{1}{2}} + 1).$$

It is instructive to compare this risk with the risk $R(\theta,\delta_m, \gamma_m)$.

One has

$$R(0, \delta_m, \gamma_m) = [2(n + n^{\frac{1}{2}})]^{-1}$$

$$[n^{\frac{1}{2}}(n + n^{\frac{1}{2}} + 1)^{\frac{1}{2}} ((2n + n^{\frac{1}{2}}))^{-\frac{1}{2}} + n^{\frac{1}{2}}(2n + n^{\frac{1}{2}})^{\frac{1}{2}} (n + n^{\frac{1}{2}} + 1)^{-\frac{1}{2}}]$$

$$> [2(n + n^{\frac{1}{2}})]^{-1}\, 2n^{\frac{1}{2}} = 1/(n^{\frac{1}{2}} +1) = R(\theta,\delta_m\gamma_1) ,$$

so that at $\theta = 0$ the total risk of (δ_m, γ_m) exceeds that of the pair (δ_m, γ_1).

A numerical calculation shows that for θ between 0 and 0.5 (because of the symmetry it suffices to consider only these values), one has

$$R(\theta, \delta_m, \gamma_m) < R(\theta, \delta_m, \gamma_1)$$

if and only if $\underline{\theta} < \theta < \bar{\theta}$. The values of $\underline{\theta}$ and $\bar{\theta}$ for small values of n, n = 2,..., 8, are given in Table 1. It is easy to show that as n increases, $\bar{\theta} \to .5$ and $\underline{\theta} \to \theta_0 = .069$. This value θ_0 solves the equation

$$[\theta(1-\theta)]^2 + 1/4 = [\theta(1-\theta)]^{\frac{1}{2}} .$$

From the frequentist point of view (see Kiefer (1977)) it is desirable to have conservatively biased estimators γ, i.e. estimators satisfying

$$E_\theta \gamma(X) \geq R(\theta, \delta).$$

In our situation such estimators γ_B do not exist. Indeed for positive α, β

$$R(\theta, \delta_B) = (n + \alpha + \beta)^{-2}[\theta(1-\theta)(n - (\alpha+\beta)^2)) + \theta\beta^2 + (1-\theta)\alpha^2] ,$$

and

$$E_\theta \gamma_B(X) = (n + \alpha + \beta)^{-2}(n + \alpha + \beta + 1)^{-1} \times [n(n-1)\theta(1-\theta) + \beta(n+\alpha)\theta + \alpha(n+\beta)(1-\theta)].$$

Comparing these functions shows that if

$$\alpha < (n + \beta)(n + \alpha + \beta + 1)^{-1} , \tag{2.3}$$

then $E_\theta \gamma_B(X) > R(\theta , \delta_B)$ for small θ, and if

$$\beta < (n + \alpha)(n + \alpha + \beta + 1)^{-1} \tag{2.4}$$

then $E_\theta \gamma_B(X) > R(\theta, \delta_B)$ for θ close to 1.

If both inequalities (2.3) and(2.4) are violated, then $E_\theta \gamma_B(X) > R(\theta, \delta_B)$ in the inner part of the unit inverval.

It follows, for instance, that for some values of X,

$$\gamma_m(X) > R(\theta, \delta_m) = 1/[4(n^{\frac{1}{2}} + 1)]^2 ,$$

i.e. the posterior loss for some samples exceeds the constant risk which it estimates. A similar situation for the unknown normal mean was described as "a moment of indecision" by O'Hagan (1981). An easy calculation shows that these "non-informative" values of X are such that

$$|X - 0.5n| < 0.5n^{3/4}.$$

If $\alpha = \beta = 0$,

$$E_\theta \gamma_0(X) = (n - 1)\theta(1 - \theta)/[n(n + 1)]$$

and

$$R(\theta, \delta_0) = \theta(1 - \theta)/n ,$$

so that for all θ, $0 < \theta < 1$

$$E_\theta \gamma_0(X) < R(\theta, \delta_0)$$

i.e. γ_0 systematically underestimates the loss of δ_0. In view of this fact and also (2.1), this loss estimator cannot be recommended in practice, which confirms previous findings (cf. Bernardo (1979), sec. 3.4 and references there).

REFERENCES

Aitchison, J. and Dunsmore, I. R. (1968). Linear-loss interval estimation of location and scale parameters, Biometrika, 55, 141-148.

Berger, J. (1985a). In defence of the likelihood principle: axiomatics and coherency, In Bayesian Statistics II(J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith eds.). North-Holland, Amsterdam, 33-65.

Berger, J. (1985b). The frequentist viewpoint and conditioning, in Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer. (L. LeCam and R. Olshen eds.). Wadsworth, Belmont 15-44.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. J. Roy. Statist. Sec. B 41, 113-147.

Cohen, A. and Strawderman, W. (1973). Admissible confidence interval and point estimation for translation or scale parameters, Ann. Statist. 1, 545-550.

Kiefer, J. (1977). Conditional confidence statements and confidence estimators. Journ. Amer. Stat. Assoc. 72, 789-827.

Lehmann, E. L. (1959), Testing Statistical Hypothesis. New York, J. Wiley & Sons.

O'Hagan, A. (1981). A moment of indecision, Biometrika, 68, 329-330.

Pratt, J. W. (1961). Length of confidence intervals, Journ. Amer. Stat. Assoc. 56, 549-567.

Sandved, E. (1968), Ancillary statistics and estimation of the loss in estimation problems, Ann. Math. Statist. 39, 1755-1758.

Savage, L. J. (1954), The Foundations of Statistics, New York, J. Wiley & Sons.

Winkler, R. L. (1972). Decision-theoretic approach to interval estimation. Journ. Amer. Stat. Assoc. 67, 187-191.

Table 1. The end-points of the interval where the risk of (δ_m, γ_m) is smaller than the risk of (δ_m, γ_1)

n	$\underline{\theta}$	$\overline{\theta}$
2	.000	.288
3	.005	.293
4	.008	.297
5	.013	.304
6	.018	.309
7	.024	.314
8	.030	.321
∞	.069	.500