ON THE INADMISSIBILITY OF SOME STANDARD ESTIMATES IN THE PRESENCE OF PRIOR INFORMATION[1]

By M. Skibinsky[2] and L. Cote

University of California, Berkeley and Purdue University

0. Introduction. A common formulation of statistical decision problems involves a sample space and a class of probability distributions on this space indexed by a parameter θ. The loss consequent upon action to resolve the problem is regarded as a function of θ. Thus in a sense, to choose a proper action, we must "know about θ". Ordinarily, the parameter is not considered to be random. When this view prevails, the decision rule may be chosen so as to keep the expected loss from assuming its worst possible value. In many physical problems, however, certain extreme values of the parameter, though not disallowed completely, are held by the experimenter to be rather unlikely. Because such prior information is rarely used to formulate the decision procedure (it may be judged too crude), experimenters often find they must modify its recommendations to suit their judgement. Some statisticians may accept this with good grace, but the existence of such disparity with its attendant misunderstandings provides a powerful incentive to reformulate the standard procedures so as to allow for a more efficient and exacting use of prior information. For an interesting general discussion of basic theories involved here, see De Finetti (1951).

In the first example which follows, we take into account prior information about the probability of success in a sequence of independent Bernoulli trials. We are to estimate this probability from an observation on the number X of successes in n such trials. The conventional estimate is of course X/n. We will assume below that the probability of success in our trials is the value of a random variable Θ. This value is what we would like to "know about". If nothing is known about the distribution of Θ, we can do no better than the usual formulation. We will assume, however, that Θ has a distribution which belongs to a subclass of the distributions on [0, 1] roughly conforming to the type of prior information that an experimenter might have; e.g., that values of Θ near 0 or 1 are unlikely. If we now regard the binomial distribution to be conditional on Θ, the members of this subclass generate a family of joint distributions for X and Θ. With this as background we may view our problem as a special case of conventional prediction theory. In the following paragraphs, we present a generalized maximum likelihood principle as applied to this example and investigate a class of predictors which it suggests. Under appropriate conditions to be discussed below each of these has a uniformly smaller mean square error than the conventional estimate.

Received July 17, 1962.

[1] Prepared with the partial support of the National Science Foundation, Grant G-14648 (for M. Skibinsky).

[2] Now at the University of Minnesota.

The second example concerns the mean of a normal distribution with known variance and is patterned after the first. We find it preferable from the point of view of readability to present these results at this time in terms of the two given examples rather than in the more general context that is possible.

1. Inadmissibility theorem. Let n, δ, α, be given; n a positive integer,

$$0 < \delta \leqq \tfrac{1}{2}, \qquad 0 < \alpha < 1.$$

We are concerned with random variables Θ, X, the former distributed on the unit interval, the latter discretely over the numbers 0, 1, $\cdots$, n. We suppose that

$$\text{Prob}\,(X = x \mid \Theta = \theta) = f(x, \theta), \tag{1}$$

where

$$f(x,\theta) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x}, & x = 0, 1, \cdots, n, \qquad 0 < \theta < 1 \\ 1 \text{ or } 0, \text{ according as } \begin{matrix} x = 0 \text{ or } x > 0, \\ x = n \text{ or } x < n, \end{matrix} & \begin{matrix} \theta = 0 \\ \theta = 1 \end{matrix} \end{cases}$$

is the value at x of the binomial frequency function with parameters n and θ. Let ν be a c.d.f. on the unit interval. We shall write P_ν to indicate any probability measure on the domain of Θ and X which satisfies (1) and has ν as marginal c.d.f. for Θ; and E_ν for expectation relative to P_ν. Let

$$\mathfrak{M}(\delta, \alpha) = \{\nu: \nu(1 - \delta) - \nu(\delta - 0) \geqq 1 - \alpha\}.$$

THEOREM. *Let ν belong to $\mathfrak{M}(\delta, \alpha)$, then for $\alpha > 0$ and sufficiently small, X/n is an inadmissible predictor of Θ relative to the squared difference loss function, in the sense that there exists a predictor which is uniformly better over $\mathfrak{M}(\delta, \alpha)$. In fact there exists a mapping ξ from the range of X to the unit interval such that*

$$E_\nu[(\xi(X) - \Theta)^2] < E_\nu[(X/n - \Theta)^2],$$

for all $\nu \in \mathfrak{M}(\delta, \alpha)$. We reserve the proof of the above theorem to Section 3.

2. A maximum likelihood method for prediction of Θ. We proceed in two steps.

First we choose corresponding to each x, a c.d.f. $\nu_x \in \mathfrak{M}(\delta, \alpha)$ such that

$$P_{\nu_x}(X = x) \geqq P_\nu(X = x), \qquad \text{all } \nu \in \mathfrak{M}(\delta, \alpha). \tag{2}$$

By definition of P_ν and (1), we have for each c.d.f. ν on [0, 1] that $P_\nu(X = x) = E_\nu f(x, \Theta)$, $x = 0, 1, \cdots, n$. But for each x, the likelihood function $f(x, \cdot)$ is strictly monotone on [0, 1] to each side of a unique maximum at $\theta = x/n$. Hence ν_x defined by

$$P_{\nu_x}(\Theta = x/n) = \alpha, \qquad P_{\nu_x}(\Theta = \delta) = 1 - \alpha, \qquad x \leqq n\delta,$$

$$P_{\nu_x}(\Theta = x/n) = 1, \qquad n\delta \leqq x \leqq n - n\delta,$$

$$P_{\nu_x}(\Theta = 1 - \delta) = 1 - \alpha, \qquad P_{\nu_x}(\Theta = x/n) = \alpha, \qquad x \geqq n - n\delta,$$

satisfies (2) uniquely for $x = 0, 1, \cdots, n$.

For the second step, we obtain corresponding to each x, a value of Θ, we shall call it $\hat{\theta}_{\delta,\alpha}(x)$, whose "a posteriori probability" (given that $X = x$ and that ν_x is the "a priori" c.d.f. of Θ) is maximum, i.e., a value θ which maximizes the conditional probability

$$(3) \qquad P_{\nu_x}(\Theta = \theta \mid X = x) = P_{\nu_x}(\Theta = \theta, X = x)/P_{\nu_x}(X = x).$$

The numerator of the right hand side above can be written

$$(4) \qquad P_{\nu_x}(\Theta = \theta)f(x, \theta),$$

and may be interpreted for each fixed x as an "a posteriori likelihood of θ". Clearly, maximizing this is for fixed x equivalent to maximizing (3). In view of the definition of ν_x : When $x < n\delta$, (4) is equal to $\alpha f(x, x/n)$, $(1 - \alpha)f(x, \delta)$, or 0, according as $\theta = x/n$, $\theta = \delta$, or θ otherwise. Symmetrically, when $x > n - n\delta$, (4) is equal to $\alpha f(x, x/n)$, $(1 - \alpha)f(x, 1 - \delta)$, or 0, according as $\theta = x/n$, $\theta = 1 - \delta$, or θ otherwise. When $n\delta \leqq x \leqq n - n\delta$, (4) is equal to $f(x, x/n)$ for $\theta = x/n$, and 0 otherwise.

Now let $\eta(x, \delta) = f(x, x/n)/f(x, \delta)$ and define $\hat{\theta}_{\delta,\alpha}(x)$ as follows: When $x < n\delta$, take it equal to x/n or δ, according as $\eta(x, \delta) >$ or $\leqq (1 - \alpha)/\alpha$. When $x > n - n\delta$, take it equal to x/n or $1 - \delta$, according as $\eta(x, 1 - \delta) >$ or $\leqq (1 - \alpha)/\alpha$. When $n\delta \leqq x \leqq n - n\delta$, take it equal to x/n. It is clear from the above description of (4) that for each x

$$(5) \qquad P_{\nu_x}(\Theta = \hat{\theta}_{\delta,\alpha}(x) \mid X = x) \geqq P_{\nu_x}(\Theta = \theta \mid X = x), \qquad \text{all } \theta.$$

This completes the second step. To find a simple expression for $\hat{\theta}_{\delta,\alpha}$, we proceed as follows.

Let b denote the largest integer less than $n\delta$. It is easy to show that for $x = 0, 1, \cdots, b$, $\eta(x, \delta)$ is strictly decreasing in x and bounded below by 1. It follows that

$$(6) \qquad 1/(1 + \eta(x, \delta)) = c(x, \delta), \quad \text{say.}$$

is strictly increasing for these x, with $c(b, \delta) \leqq \frac{1}{2}$. By the above discussion and simple considerations of symmetry, $\hat{\theta}_{\delta,\alpha}(x) = x/n$, when $\alpha > c(b, \delta)$. For $\alpha \leqq c(b, \delta)$ on the other hand, if we define a to be the smallest non-negative integer x such that $\eta(x, \delta) \leqq (1 - \alpha)/\alpha$ (or equivalently, such that $c(x, \delta) \geqq \alpha$), then

$$\hat{\theta}_{\delta,\alpha}(x) = \begin{cases} \delta, & a \leqq x \leqq b \\ 1 - \delta, & n - b \leqq x \leqq n - a \\ x/n, & \text{otherwise.} \end{cases}$$

$\hat{\theta}_{\delta,\alpha}$ is uniquely optimum in the sense of (5) unless $c(x, \delta) = \alpha$, for some $x \leqq b$, in which case it may be modified at x (and/or at $n - x$), by replacing its value

δ, there, with x/n (and/or its value $1 - \delta$, there, with $1 - x/n$) without affecting the value of the left hand side in (5). It is to be emphasized that the precise sense in which $\hat{\theta}_{\delta,\alpha}(X)$ is optimal as a predictor of Θ relative to the class $\mathfrak{M}(\delta, \alpha)$ is given by (5) and (2). In the following section a comparison is made between this predictor and the more conventional predictor X/n in terms of their mean square deviations from Θ.

3. A class of predictors for Θ. We consider the following class of predictors for Θ which are suggested by the maximum likelihood predictor $\hat{\theta}_{\delta,\alpha}(X)$. Define ξ_j, for $j = 0, 1, \cdots, b$ on the range of X by

$$\xi_j(x) = \begin{cases} \delta, & j \leqq x \leqq b \\ 1 - \delta, & n - b \leqq x \leqq n - j \\ x/n, & \text{otherwise.} \end{cases}$$

Observe that ξ_j depends upon δ although this is suppressed in the notation for convenience in presentation. The relationship of $\hat{\theta}_{\delta,\alpha}$ to the ξ_j follows directly from the definition of a in the preceding section. Indeed for $j = 0, 1, \cdots, b$, we have that

$$\hat{\theta}_{\delta,\alpha}(x) \equiv \xi_j(x), \qquad \text{when } c(j - 1, \delta) < \alpha \leqq c(j, \delta). \tag{7}$$

We take $c(-1, \delta) = 0$.

To compare $\xi_j(X)$ with X/n as a predictor of Θ, we examine, for $\nu \in \mathfrak{M}(\delta, \alpha)$, the difference

$$E_\nu[(\xi_j(X) - \Theta)^2] - E_\nu[(X/n - \Theta)^2] = E_\nu H_j(\Theta), \quad \text{say,} \tag{8}$$

where we take $H_j(\Theta)$ to be the conditional expectation, given Θ, of the difference between the two squares. A simple computation shows that $H_j(\theta) = h_j(\theta) + h_j(1 - \theta)$, where

$$h_j(\theta) = \sum_{x=j}^{b} (\delta - x/n)(\delta + x/n - 2\theta) f(x, \theta).$$

Thus, the H_j are polynomials in θ each of which is symmetric about $\theta = \frac{1}{2}$. In addition, for $j = 0, 1, \cdots, b$, $H_j(\theta) < 0$, when $\delta \leqq \theta \leqq 1 - \delta$, and >0, for sufficiently small $\theta > 0$ (e.g. whenever $0 < \theta < \frac{1}{2}[1 - (1 - \delta)^{\frac{1}{2}}]$). Note that $H_0(0) = \delta^2$, while $H_j(0) = 0$, $j = 1, 2, \cdots, b$. These properties are readily verified and have as an immediate consequence that the largest value attained by (8) for any $\nu \in \mathfrak{M}(\delta, \alpha)$ is

$$\alpha \max\nolimits_{0 \leqq \theta < \delta} H_j(\theta) + (1 - \alpha) \max\nolimits_{\delta \leqq \theta \leqq \frac{1}{2}} H_j(\theta). \tag{9}$$

The first term above is positive and the second, is negative. The theorem of Section 1 follows immediately. Clearly, any one of the predictors $\xi_j(X)$ is uniformly better over $\mathfrak{M}(\delta, \alpha)$ than X/n relative to the squared difference loss function provided only that α is sufficiently small.

Recall from (7) that for an arbitrary fixed α, $0 < \alpha \leqq c(b, \delta)$, we have

$\hat{\theta}_{\delta,\alpha} = \xi_j$, for some integer j, $0 \leqq j \leqq b$. By the preceding paragraph, it follows for any such α, that for an $\alpha' > 0$ and sufficiently small, $\hat{\theta}_{\delta,\alpha}(X)$ is uniformly better over $\mathfrak{M}(\delta, \alpha')$ than X/n relative to the squared difference loss function. For example, as exhibited by the tables described in the following section, when $n = 30$ and $\delta = .35$, $\hat{\theta}_{.35,.40}(X) = \xi_9(X)$ is uniformly better over $\mathfrak{M}(.35, .22)$ than $X/30$, but it is not uniformly better than $X/30$ over the larger class $\mathfrak{M}(.35, .40)$. Thus, although $\hat{\theta}_{\delta,\alpha}(X)$ is always optimal over $\mathfrak{M}(\delta, \alpha)$ in the sense of (5) and (2), it need not be uniformly better than X/n over all of this class. On the other hand, we have by (7) that when $\alpha > 0$ is sufficiently small,

$$\hat{\theta}_{\delta,\alpha}(X) = \xi_0(X)$$

and is in addition, by the results of the preceding paragraph, uniformly better over $\mathfrak{M}(\delta, \alpha)$ than X/n. Thus, $\hat{\theta}_{\delta,\alpha}$ itself qualifies as an example of ξ in the theorem.

For $j = 0, 1, \cdots, b - 1$, the properties described above for H_j continue to hold if we replace $H_j(\theta)$ by the difference $g_j(\theta) = H_{j+1}(\theta) - H_j(\theta)$. But then, for the same reason,

$$\begin{aligned} \sup_{\nu\in\mathfrak{M}(\delta,\alpha)} \{E_\nu[(\xi_j(X) - \Theta)^2] - E_\nu[(\xi_{j+1}(X) - \Theta)^2]\} \\ = \alpha \max_{0 \leqq \theta < \delta} g_j(\theta) + (1 - \alpha) \max_{\delta \leqq \theta \leqq \frac{1}{2}} g_j(\theta). \end{aligned} \tag{10}$$

Thus we have as a coincident and more explicit development of the theorem in Section 1 the

EXTENSION THEOREM. *For any* $j = 0, 1, \cdots, b$ *and* $\alpha > 0$ *sufficiently small, each predictor (except the last) in the sequence*

$$\xi_j(X), \xi_{j+1}(X), \cdots, \xi_b(X), X/n$$

is uniformly better over $\mathfrak{M}(\delta, \alpha)$ *as a predjtor of* Θ *than the one which follows, according to the squared difference loss function.*

For each fixed δ, the class $\mathfrak{M}(\delta, \alpha)$ increases monotonically with α. Hence it is clear that the largest "sufficiently small" α for which the above theorem will hold cannot decrease as j increases. For example, again considering the case $n = 30$, $\delta = .35$, we find after consulting the tables described below that $b = 10$ and (correct to four decimal places) this largest α is .0000 for $j = 0, 1, 2$; .0001, $j = 3$; .0005, $j = 4$; .0030, $j = 5$; .0141, $j = 6$; .0514, $j = 7$; and .1381 for $j = 8, 9, 10$.

4. Numerical computations. Computations carried out with Fortran 704 for $n \leqq 50$ and $\delta \leqq \frac{1}{2}$ indicate that for $j = 0, 1, \cdots, b$, H_j is monotone on $[0, \delta]$ to either side of a unique maximum which occurs at 0 if $j = 0$ and is otherwise in $(0, \delta)$. On $[\delta, \frac{1}{2}]$, it either decreases monotonically to a unique minimum at $\theta = \frac{1}{2}$ (this occurs for $n \leqq 6$) or else it is monotone to either side of a minimum in $(\delta, \frac{1}{2})$ (and it then has a relative maximum of negative value at $\theta = \frac{1}{2}$). It follows that

$$\sup_{0 \leqq \theta < \delta} H_j(\theta)$$

TABLE FOR BINOMIAL EXAMPLE[3]

n	δ	j	$\hat{\beta}(j, \delta)$	$\hat{a}(j, \delta)$	$c(j, \delta)$
10	.25	0	.0058	.0698	.0533
		1	.1379	.2424	.3264
		2		.2270	.4825
	.35	0	.0036	.0896	.0133
		1	.0864	.2199	.1576
		2	.2897	.3097	.3677
		3		.3439	.4859
	.45	0	.0023	.0962	.0025
		1	.0556	.2196	.0508
		2	.2312	.3404	.2017
		3	.4225	.4544	.3842
		4		.5599	.4873
20	.25	0	.0000	.0068	.0032
		1	.0003	.0173	.0531
		2	.0044	.0341	.1901
		3	.0331	.0710	.3554
		4		.1515	.4650
	.35	0	.0000	.0268	.0002
		1	.0002	.0549	.0051
		2	.0025	.0867	.0338
		3	.0182	.1376	.1173
		4	.0837	.2191	.2528
		5	.2477	.2649	.3860
		6		.2472	.4719
	.45	0	.0000	.0447	.0000
		1	.0001	.0802	.0003
		2	.0015	.1130	.0029
		3	.0107	.1570	.0162
		4	.0478	.2154	.0600
		5	.1448	.2865	.1527
		6	.2767	.3624	.2802
		7	.3984	.4333	.3983
		8		.4982	.4746
30	.25	0	.0000	.0013	.0002
		1	.0000	.0026	.0048
		2	.0000	.0042	.0299
		3	.0001	.0068	.1021
		4	.0010	.0119	.2236
		5	.0060	.0226	.3528
		6	.0287	.0470	.4477
		7		.1062	.4944
	.35	0	.0000	.0113	.0000

[3] See Section 4 for definitions of table headings.

TABLE FOR BINOMIAL EXAMPLE—*Continued*

n	δ	j	$\hat{\beta}(j, \delta)$	$\hat{\alpha}(j, \delta)$	$c(j, \delta)$
30	.35	1	.0000	.0196	.0001
		2	.0000	.0271	.0011
		3	.0001	.0374	.0065
		4	.0005	.0525	.0261
		5	.0030	.0757	.0757
		6	.0141	.1125	.1644
		7	.0514	.1704	.2772
		8	.1464	.2522	.3826
		9	.2480	.2292	.4579
		10		.1381	.4954
	.45	0	.0000	.0282	.0000
		1	.0000	.0442	.0000
		2	.0000	.0566	.0000
		3	.0000	.0716	.0002
		4	.0003	.0908	.0010
		5	.0017	.1160	.0044
		6	.0076	.1489	.0159
		7	.0268	.1910	.0456
		8	.0759	.2419	.1050
		9	.1703	.2987	.1954
		10	.2724	.3567	.3002
		11	.3649	.4119	.3947
		12	.4450	.4655	.4619
		13		.5311	.4958

which must be positive, is attained at $\theta = 0$ when $j = 0$ and otherwise is attained in $(0, \delta)$, while

$$\max_{\delta \leqq \theta \leqq \frac{1}{2}} H_j(\theta)$$

which must be negative is attained either at $\theta = \delta$ or at $\theta = \frac{1}{2}$. Strictly identical remarks also hold if $H_j(\theta)$ is replaced by $g_j(\theta)$, $j = 0, 1, \cdots, b - 1$.

The table exhibits for a few selected values of n, δ and for $j = 0, 1, \cdots, b - 1$, the largest value of α for which the predictor $\xi_j(X)$ is uniformly better (over $\mathfrak{M}(\delta, \alpha)$) than $\xi_{j+1}(X)$, according to the squared difference loss function, i.e. This is the value of α, call it $\hat{\beta}(j, \delta)$ for which the right hand side of (10) is zero. In the column adjacent is listed for $j = 0, 1, \cdots, b$, the largest value of α for which the predictor $\xi_j(X)$ is uniformly better over $\mathfrak{M}(\delta, \alpha)$ than X/n, according to the squared difference loss function, i.e. this is the value of α, call it $\hat{\alpha}(j, \delta)$ for which the expression in (9) is zero. Also tabulated (last column) are the values of $c(j, \delta)$ given by (6). By (7), these exhibit the values of α for which $\xi_j(X)$ is precisely the two step maximum likelihood **predictor** $\hat{\theta}_{\delta,\alpha}$ **of** Section 2.

5. Example. Suppose it is known a priori that a certain coin manufacturing process is such that each time a coin is produced, there is *at least an even chance* that its probability of landing heads when tossed in some prescribed manner, will lie between .45 and .55. Suppose we toss such a coin ten times in an effort to "predict" the value of Θ that it has. We may ask whether the standard estimate $X/10$ (X = number of heads in ten tosses) is admissible. The answer is no. For the unconditional distribution of X is given by our prior information, to be induced by a member of $\mathfrak{M}(.45, .50)$, and ξ_4 is a predictor that is uniformly better than $X/10$ over this class. Indeed, for $n = 10$, $\hat{\alpha}(4, .45) = .56$, so that $\xi_4(X)$ is uniformly better than $X/10$ over the larger class $\mathfrak{M}(.45, .56)$. Prior information which is more precise may yield an estimate uniformly better than either $\xi_4(X)$ or $X/10$. If it is known a priori that Θ will at least *four out of five times on the average* lie in the indicated interval, i.e. that the distribution of X is induced by a member of the smaller family $\mathfrak{M}(.45, .20)$, then $\xi_2(X)$ is a uniformly better predictor of Θ than $\xi_3(X)$, $\xi_4(X)$, or $X/10$.

6. Inadmissibility theorem, normal example. Let n, α be given as before, but now take δ to be any fixed positive number. Again we are concerned with random variables Θ, X. This time we suppose both to be distributed on the line and such that

$$P(X \leqq t \mid \Theta = \theta) = \int_{-\infty}^{t} f(x, \theta)\, dx, \tag{11}$$

where

$$f(x, \theta) = (\sqrt{2\pi n})^{-1} e^{-(1/2n)(x-n\theta)^2}$$

is the value at x of the density function of a sum of n independent normal random variables each with mean θ and variance 1. Let ν be a c.d.f. on the line and write P_ν, E_ν with conventions analogous to those of Section 1. Let $\mathfrak{M}(\delta, \alpha) = \{\nu\colon \nu(\delta) - \nu(-\delta - 0) \geqq 1 - \alpha\}$. Then the theorem of Section 1 continues to hold precisely as worded (excepting only that now the mapping ξ is to the line) and with changes in definition as noted above. Proof is deferred to Section 8.

7. Maximum likelihood predictor. By (11) and using Fubini's theorem we have

$$P_\nu(X \leqq t) = E_\nu P(X \leqq t \mid \Theta) = \int_{-\infty}^{t} E_\nu f(x, \Theta)\, dx.$$

Again, we proceed in two steps.

First we choose corresponding to each real x, a c.d.f. $\nu \in \mathfrak{M}(\delta, \alpha)$ such that

$$E_{\nu_x} f(x, \Theta) \geqq E_\nu f(x, \Theta), \qquad \text{all } \nu \in \mathfrak{M}(\delta, \alpha). \tag{12}$$

For reasons strictly analogous to those of the binomial example (see Section 2),

ν_x defined by

$$P_{\nu_x}(\Theta = x/n) = \alpha, \qquad P_{\nu_x}(\Theta = \delta \operatorname{sgn} x) = 1 - \alpha, \qquad |x| > n\delta,$$
$$P_{\nu_x}(\Theta = x/n) = 1, \qquad |x| \leqq n\delta,$$

satisfies (12) uniquely for all x.

For the second step, we obtain a value θ, call it $\hat{\theta}_{\delta,\alpha}(x)$, as before, which maximizes

$$P_{\nu_x}(\Theta = \theta \mid X = x) = P_{\nu_x}(\Theta = \theta) f(x, \theta)/E_{\nu_x} f(x, \Theta).$$

Again, we need only maximize the numerator of the right hand side. In view of the definition of ν_x : When $|x| > n\delta$, the numerator is $\alpha f(x, x/n)$, $(1 - \alpha) f(x, \delta \operatorname{sgn} x)$, or 0, according as $\theta = x/n$, $\theta = \delta \operatorname{sgn} x$, or otherwise. When $|x| \leqq n\delta$, the numerator is $f(x, x/n)$, when $\theta = x/n$ and 0, otherwise.

Now $f(x, x/n)/f(x, \delta \operatorname{sgn} x) = \exp[(1/2n)(x - n\delta \operatorname{sgn} x)^2]$ which is bounded below by 1. Hence if we define

$$\hat{\theta}_{\delta,\alpha}(x) = \begin{cases} \begin{cases} \delta \operatorname{sgn} x, & n\delta \leqq |x| \leqq n\delta + (2n \log (1 - \alpha)/\alpha)^{\frac{1}{2}}, \\ x/n, & \text{otherwise,} \end{cases} & 0 < \alpha < \tfrac{1}{2} \\ x/n, \quad \text{all } x, & \tfrac{1}{2} \leqq \alpha \leqq 1, \end{cases}$$

then (5) with appropriate changes in definition as given above, continues to hold.

8. A class of predictors for Θ. Again we consider the obvious class of predictors for Θ which is suggested by $\hat{\theta}_{\delta,\alpha}$. We define ξ_r for each real $r > 0$, by

$$\xi_r(x) = \begin{cases} \delta \operatorname{sgn} x, & n\delta \leqq |x| \leqq n(\delta + r) \\ x/n, & \text{otherwise,} \end{cases}$$

(again δ is suppressed in the notation for convenience in presentation), and observe that $\hat{\theta}_{\delta,\alpha}(x) \equiv \xi_r(x)$, when $\alpha = 1/(1 + e^{nr^2/2})$, $r > 0$. We examine the difference

(13) $$E_\nu[(\xi_r(X) - \Theta)^2] - E_\nu[(X/n - \Theta)^2] = E_\nu H_r(\Theta), \quad \text{say}$$

for $\nu \in \mathfrak{M}(\delta, \alpha)$, and by elementary computations find that

$$H_r(\theta) = h_r(\theta) + h_r(-\theta)$$

where

$$h_r(\theta) = (n/2\pi)^{\frac{1}{2}} \int_\delta^{\delta+r} (\delta - x)(\delta + x - 2\theta) e^{-n(x-\theta)^2/2}\, dx$$

so that

$$H_r(\theta) = (2n/\pi)^{\frac{1}{2}} \int_\delta^{\delta+r} (\delta - t)[(\delta + t) \cosh(n\theta t) - 2\theta \sinh(n\theta t)] e^{-n(t^2+\theta^2)/2}\, dt.$$

Thus, the H_r are continous functions of θ, symmetric about $\theta = 0$. In addition for each $r > 0$, $H_r(\theta) < 0$ when $|\theta| \leqq \delta$, $H_r(\theta) > 0$, for all θ sufficiently large in absolute value, and $H_r(\theta) \to 0$, as $\theta \to \infty$. These properties, which are readily verified, show that for any fixed $r > 0$, the largest value attained by (13) for any $\nu \varepsilon \mathfrak{M}(\delta, \alpha)$ is

$$\alpha \max_{|\theta|>\delta} H_r(\theta) + (1 - \alpha) \max_{|\theta| \leqq \delta} H_r(\theta).$$

In view of the above remarks, this is negative for $\alpha > 0$ sufficiently small and the theorem referred to in Section 6 follows. Analogous to the binomial example, we have the following immediate

EXTENSION THEOREM. *For any $r_0 > 0$, we have that for all sufficiently small $\alpha > 0$, that if r_1, r_2 are any numbers such that*

$$0 < r_1 < r_2 \leqq r_0,$$

then $\xi_{r_2}(X)$ is uniformly better over $\mathfrak{M}(\delta, \alpha)$ as a predictor of Θ (according to the squared difference loss function) than either $\xi_{r_1}(X)$ or X/n.

REFERENCE

DE FINETTI, BRUNO (1951). Recent Suggestions for the Reconciliation of Theories of Probability. *Second Berkeley Symp. Math. Statist. and Prob.* 217–225.