The Annals of Statistics
1982, Vol. 10, No. 3, 1017–1020

STEIN'S PARADOX IS IMPOSSIBLE IN PROBLEMS WITH FINITE SAMPLE SPACE

BY SAM GUTMANN

Northeastern University

The use of admissible procedures in each of several problems may be inadmissible when the problems are combined, in the sense of summed loss. Thus apparently irrelevant information can sometimes be used to reduce risk over the entire parameter space. This is known as Stein's Paradox. We prove here that this cannot occur when the sample spaces for the problems are finite.

The inadmissibility of the usual estimator for three normal means (Stein, 1956) provides an example of three procedures (the MLE's for the individual means) each of which is admissible (with squared error loss) but which are inadmissible when the problems are combined. Johnson (1971) showed that this does not happen in binomial estimation, that the MLE for any number of binomial parameters is admissible (e.g. for squared error loss). In both cases the "connection" between individual problems is that the loss function in the joint problem is the sum of the individual losses.

The purpose of this note is to show that Stein's Paradox does not occur in any problem with finite sample space. Problems with finite sample space have been singled out by Brown (1981), who proved complete class theorems for many such problems. When the loss functions in the problems are strictly convex and bounded (or suitably normalizable) Brown's result can be used to derive the present result.

A statistical decision problem will be written (as in Brown, 1976, for example)

$$P_1 = (X, \Theta, \mathscr{A}, \mathscr{S}(X), \mathscr{S}(\Theta), \mathscr{S}(\mathscr{A}), F_\theta, L_1), \tag{1}$$

where X, Θ, and $\mathscr{A}$ are sample, parameter, and action spaces with given σ-algebras $\mathscr{S}(X)$, $\mathscr{S}(\Theta)$, and $\mathscr{S}(\mathscr{A})$, F_θ is a measure on $\mathscr{S}(X)$ for each $\theta \in \Theta$ such that $F_\theta(S)$ is $\mathscr{S}(\Theta)$-measurable for each $S \in \mathscr{S}(X)$, and the loss

$$L_1 : \Theta \times \mathscr{A} \to [0, \infty]$$

is $\mathscr{S}(\Theta) \times \mathscr{S}(\mathscr{A})$-measurable. Risk functions in this problem are real valued functions of the form

$$\int_X F_\theta(dx) \int_{\mathscr{A}} L_1(\theta, a)\delta(x, da),$$

where for each x, $\delta(x, \cdot)$ is a measure on $\mathscr{S}(\mathscr{A})$ and for fixed $S \in \mathscr{S}(\mathscr{A})$, $\delta(\cdot, S)$ is $\mathscr{S}(X)$-measurable.

Given a second problem

$$P_2 = (Y, \Phi, \mathscr{B}, \mathscr{S}(Y), \mathscr{S}(\Phi), \mathscr{S}(\mathscr{B}), G_\phi, L_2) \tag{2}$$

we define the joint problem P_1P_2 by

$$(X \times Y, \Theta \times \Phi, \mathscr{A} \times \mathscr{B}, \mathscr{S}(X) \times \mathscr{S}(Y), \mathscr{S}(\Theta) \times \mathscr{S}(\Phi), \mathscr{S}(\mathscr{A}) \times \mathscr{S}(\mathscr{B}), F_\theta \times G_\theta, L)$$

where the loss L is given by

$$L((\theta, \phi), (a, b)) = L_1(\theta, a) + L_2(\phi, b).$$

Received April 1981; revised February 1982.
AMS 1980 *subject classifications*. Primary, 62C15; secondary, 62F10.
Key words and phrases. Admissible procedures, Stein's effect, Statistical decision problems.

The product measure $F_\theta \times G_\phi$ is used, of course, to make the data in the two problems independent. Similarly the joint problem $P_1 P_2 \cdots P_K$ can be defined given problems P_1, $P_2, \cdots, P_K$.

We can now state the main result.

THEOREM. *Let P_1 and P_2 be two statistical decision problems, and assume that the sample space in P_1 is finite. Let δ_i be admissible procedures in P_i, $i = 1, 2$. Then the joint procedure (δ_1, δ_2) is admissible in the joint problem $P_1 P_2$.*

Formally, the joint procedure (δ_1, δ_2) is given by

$$(\delta_1, \delta_2)((x, y), \cdot) = \delta_1(x, \cdot) \times \delta_2(y, \cdot).$$

COROLLARY. *Let $P_1, \cdots, P_K$ be statistical problems and assume that the sample spaces in $P_1, P_2, \cdots, P_{K-2}$, and P_{K-1} are finite. Let δ_i be admissible in P_i, $i = 1, \cdots, K$. Then the joint procedure $(\delta_1, \delta_2, \cdots, \delta_K)$ is admissible in the joint problem $P_1 P_2 \cdots P_K$.*

The corollary follows immediately from the theorem by induction on K. Using this corollary, Johnson's (1971) result that the MLE is admissible for k binomial means follows immediately from the admissibility of the MLE for 1 binomial mean. Also, the MLE is admissible for estimating k binomial and 2 normal means, with summed squared error loss, for example.

The proof of the theorem will be given after one lemma, in which we use the following notion.

DEFINITION. Let R be a risk function in a problem with parameter space Φ. Then for $T \subset \Phi$, R is admissible relative to T iff for any other risk function R',

$$R'(\phi) \le R(\phi) \quad \forall \phi \in T \quad \text{implies} \quad R'(\phi) = R(\phi) \quad \forall \phi \in T.$$

LEMMA. *Let R be an admissible risk function in a problem with finitely many nonrandomized risk functions, but infinite parameter space Φ. Then there exists a finite set $H \subset \Phi$ such that for any $T \supset H$, R is admissible relative to T.*

The set of risk functions is compact, and for each ϕ the set of R' for which $R(\phi) < R'(\phi)$ is open. But these sets do not cover the set of risk functions because R is in none of them. The finite dimensionality as well as the compactness must be used.

PROOF. (LEMMA). Let $R_1, \cdots, R_N$ be the nonrandomized risk functions. Let d_ϕ denote the vector

$$(R(\phi) - R_1(\phi), R(\phi) - R_2(\phi), \cdots, R(\phi) - R_N(\phi)).$$

The admissibility of R is then equivalent to the assertion that for every vector $c = (c_1, c_2, \cdots, c_N)$ with $c_i \ge 0$, $\sum c_i = 1$,

$$\text{if} \quad c \cdot d_\phi \ge 0 \quad \forall \phi \in \Phi \quad \text{then} \quad c \cdot d_\phi = 0 \quad \forall \phi \in \Phi. \tag{3}$$

Let C denote the simplex $\{c : c_i \ge 0, \sum c_i = 1\}$. Let V be the smallest subspace containing $\{d_\phi : \phi \in \Phi\}$. If P is projection onto V, then by (3) we have for any $v \in PC$,

$$\text{if} \quad v \cdot d_\phi \ge 0 \quad \forall \phi \in \Phi \quad \text{then} \quad v \cdot d_\phi = 0 \quad \forall \phi \in \Phi. \tag{4}$$

But if $v \in PC$ and $v \cdot d_\phi = 0$ $\forall \phi \in \Phi$, then v must be 0. Thus for every $v \in PC \backslash \{0\}$ there must be some ϕ such that

$$v \cdot d_\phi < 0 \quad \text{or alternatively} \quad \frac{v}{|v|} \cdot d_\phi < 0. \tag{5}$$

Since PC is a convex polyhedron,

$$M = \{v/|v| : v \in PC\backslash\{0\}\}$$

is a compact subset of the unit sphere. By (5),

$$\{\{u \in M : u \cdot d_\phi < 0\}, \qquad \phi \in \Phi\}$$

is an open cover of M. Let H be a finite subset of Φ corresponding to a finite subcover.

Let $T \supset H$. Assume

$$R(\phi) \geq \textstyle\sum c_i R_i(\phi) \qquad \forall \phi \in T.$$

Thus $c \cdot d_\phi \geq 0 \ \forall \phi \in H$, and either $Pc = 0$ or $Pc \neq 0$ and $Pc/|Pc| \cdot d_\phi \geq 0 \ \forall \phi \in H$. The second alternative is excluded by the choice of H. If $Pc = 0$ then

$$R(\phi) = \textstyle\sum c_i R_i(\phi) \qquad \forall \phi \in \Phi,$$

a fortiori for all $\phi \in T$. Hence R is admissible relative to T.

PROOF (THEOREM). Let P_1 and P_2 be as in (1) and (2) respectively. Let $r(\theta)$ and $R(\phi)$ be the risk functions corresponding to δ_1 and δ_2 respectively. Then the risk function corresponding to (δ_1, δ_2) in the problem P_1P_2 is $r(\theta) + R(\phi)$. If $\rho(\theta, \phi)$ is any other risk function in the problem P_1P_2, we must show that

$$r(\theta) + R(\phi) \geq \rho(\theta, \phi) \qquad \forall \theta, \phi$$

cannot hold with strict inequality for some pair θ, ϕ.

Any risk function $\rho(\theta,\phi)$ can be written

$$r_\phi(\theta) + \int F_\theta(dx) R_x(\phi),$$

where r_ϕ and R_x are risk functions in P_1 for each ϕ and in P_2 for all x, respectively. If $\delta^*((x, y), \cdot)$ denotes the procedure corresponding to ρ, and if $\delta_i^*((x, y), \cdot)$ denotes the projection measure of δ^* onto $\mathscr{A}$ for $i = 1$ and onto $\mathscr{B}$ for $i = 2$, then r_ϕ corresponds to the procedure

$$\int_Y G_\phi(dy) \delta_1^*((x, y), \cdot)$$

and R_x corresponds to the procedure $\delta_2^*((x, y), \cdot)$. The (easy) details may be found in Gutmann (1981); what is crucial is that the loss in the joint problem is the sum of the individual losses.

Since X, the sample space in P_1, was assumed finite, we can write

$$\rho(\theta, \phi) = r_\phi(\theta) + \textstyle\sum_x f_\theta(x) R_x(\phi)$$

where $f_\theta(x) = F_\theta(\{x\})$. We now assume

$$r_\phi(\theta) + \textstyle\sum_x f_\theta(x) R_x(\phi) \leq r(\theta) + R(\phi) \qquad \forall \theta, \phi, \tag{6}$$

with strict inequality for some pair θ, ϕ, and derive a contradiction. For convenience, set $X = \{1, 2, \cdots, N\}$.

R is admissible in the original problem, hence also in the problem consisting of nonrandomized risk functions $R, R_1, \cdots, R_N$. Let $\mathscr{Q}$ denote the set of all convex combinations of $R, R_1, \cdots, R_N$. Choose H as in the lemma. In addition, choose ϕ_0 in Φ such that for some $\theta \in \Theta$,

$$r_{\phi_0}(\theta) + \textstyle\sum_x f_\theta(x) R_x(\phi_0) < r(\theta) + R(\phi_0).$$

Let $T = H \cup \{\phi_0\}$. By the lemma, R is admissible relative to T. The problem has now been reduced to one with finitely many nonrandomized risk functions and a finite parameter space T. For such a problem, every admissible R must be Bayes with respect to some prior π which puts positive mass at every point of T (Blackwell and Girschick, 1954, page 130). That is,

$$\sum_T \pi(\phi)R(\phi) \le \sum_T \pi(\phi)R'(\phi) \qquad \forall R' \in \mathcal{Q}. \tag{7}$$

Now integrate (6) with respect to π, obtaining

$$\sum_T \pi(\phi)r_\phi(\theta) + \sum_T \pi(\phi) \sum_x f_\theta(x)R_x(\phi) \le r(\theta) + \sum_T \pi(\phi)R(\phi), \tag{8}$$

which holds for all θ. In fact, since $\pi(\phi) > 0$ $\forall \phi \in T$ and since (6) was a strict inequality for $\phi_0 \in T$ for some θ, the inequality (8) must also be strict for some θ. For each θ, $\sum_x f_\theta(x)R_x(\phi)$ is in $\mathcal{Q}$, so by (7) and (8), we must have

$$\sum_T \pi(\phi)r_\phi(\theta) \le r(\theta) \qquad \forall \theta \in \Theta \tag{9}$$

with strict inequality for some θ. Since every r_ϕ corresponds to a procedure $\delta_\phi(\cdot, \cdot)$, the left-hand side of (9) is a risk function corresponding to the procedure $\sum_T \pi(\phi)\delta_\phi(\cdot, \cdot)$ and thus (9) contradicts the admissibility of $r(\theta)$.

In the language of Gutmann (1981), we have proved that any admissible risk function in a problem with finite sample space is immune, by proving the superimmunity of any admissible risk function in a problem with finitely many nonrandomized risk functions.

Perhaps the usual caveat is in order. Admissibility is no guarantee of a procedure's usefulness. For the problems considered here, e.g. estimating several binomial parameters or several normal means in a "discretized" setting, we would expect Stein-like estimators to improve over our admissible estimators for large parts of the parameter space at little cost over the rest, if the loss is reasonable.

Acknowledgment. L. D. Brown suggested looking at these questions in the finite sample space case. An Associate Editor and referees suggested valuable improvements.

REFERENCES

BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

BROWN, L. D. (1976). *Notes on Statistical Decision Theory.* (Unpublished lecture notes, Ithaca.)

BROWN, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.* **9,** 1289–1300.

GUTMANN, S. (1981). Decisions immune to Stein's effect. Preprint.

JOHNSON, B. MCK. (1971). On the admissible estimators for certain fixed sample binomial problems. *Ann. Math. Statist.* **42,** 1579–1587.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. 3rd Berkeley Symposium on Math Statist. and Probability* **1** 197–208. University of California Press, Berkeley.

DEPARTMENT OF MATHEMATICS
NORTHEASTERN UNIVERSITY
360 HUNTINGTON AVENUE
BOSTON, MASSACHUSETTS 02115