[2] HERBERT ROBBINS, "Asymptotically subminimax solutions of compound statistical decision problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951.
[3] P. FRANK AND J. KIEFER, "Almost subminimax and biased minimax procedures," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 465–468.
[4] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, 1950.
[5] A. WALD, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947), pp. 279–313.

ALMOST SUBMINIMAX AND BIASED MINIMAX PROCEDURES[1]

BY P. FRANK AND J. KIEFER

Columbia University

Robbins [1] emphasized the notion of an "almost subminimax" procedure[2] and gave an example of such a procedure. The examples in this paper have been constructed with a view to simplicity and to the indication of the underlying mechanism which makes subminimax solutions exist in certain decision problems. At the same time we point out another potentially undesirable property of a minimax procedure—biasedness.

All our examples fall within the following framework. A sample of one is taken from a population whose distribution is one of n given distributions: $F_1(x), F_2(x), \cdots, F_n(x)$. There are n decisions: $d_1, \cdots, d_n$. The weight function is $W(F_i, d_j) = 0$ if $i = j$ and $= 1$ otherwise. Instead of a finite number of F's, we may have a sequence of F's with a corresponding sequence of decisions. In all our examples each of the F's will be a uniform distribution over a finite interval of the x-axis, and our decision procedures will be randomized. These restrictions are made only for arithmetical simplicity.

With this setup, the risk when F_i is the true distribution is equal to the probability of not making decision d_i, which we will denote $r(F_i)$. We will not give an exact definition of an almost subminimax procedure, but just say that a procedure is almost subminimax if its maximum risk is "a little greater" than that of the minimax procedure (which risk is the same for all minimax procedures in our examples) and on the other hand its risk is "a lot less" than that of the minimax for "most of" the F's. Our examples will conform with this "definition" for almost any reasonable interpretation of the phrases in the quotes.

The first example will give an indication of the mechanism which makes a subminimax example possible. Let $F_1(x)$ be the uniform distribution on the interval $1 - a$ to 1, where $a > 0$ and small. Let $F_2(x)$ be the uniform distribution on the interval 0 to 1. An admissible minimax procedure to decide between d_1

[1] Research done under a contract with the Office of Naval Research.

[2] The examples of this paper fall into the framework of the definition in [1] of an "asymptotically subminimax solution" if each example is replaced by a sequence of examples whose a's approach zero. The present nomenclature was suggested as more suitable here.

and d_2 is to accept d_2 for $0 \leq x \leq 1 - a$ and for $1 - a < x \leq 1$ to accept d_1 with probability p_1 and accept d_2 with probability $1 - p_1$, where $p_1 = \frac{1}{1+a}$. With this procedure $r(F_1) = r(F_2) = \frac{a}{1+a}$.

Let us compare this procedure with the procedure which tells us to accept d_2 for $0 \leq x \leq 1 - a$ and to accept d_1 for $1 - a < x \leq 1$. For this procedure we have $r(F_1) = 0$, and $r(F_2) = a$. Thus we see that we can reduce the risk under F_1 to its absolute minimum 0, while increasing the risk under F_2 only slightly. (In fact, the ratio of the two risks under F_2, namely, a and $\frac{a}{1+a}$, approaches 1 as $a \to 0$.) This example, which may seem meaningless because $\frac{a}{1+a} \to 0$ as $a \to 0$, was given mainly to help in understanding the underlying mechanism in the almost subminimax example which follows. In the latter the maximum risk will be $> \frac{1}{2}$ for all a.

Let $F_3(x)$ be a uniform distribution from $-a$ to $1 - a$. In deciding between $F_2(x)$ and $F_3(x)$, an admissible minimax procedure is to accept d_2 for $1 - a < x \leq 1$, to accept d_3 for $-a \leq x < 0$, and to accept d_2 and d_3 with probability $\frac{1}{2}$ each for $0 \leq x \leq 1 - a$. The minimax risk is $\frac{1}{2}(1 - a)$. For a small, this is near $\frac{1}{2}$ and the two distributions are so intermeshed that there is little hope to disentangle them. When we now consider the problem of deciding between F_1, F_2, and F_3, we expect that the addition of F_1 can not do much to aggravate the difficulty already present in trying to decide between F_2 and F_3.

The following is an admissible minimax procedure for deciding between F_1, F_2, and F_3:

for $-a \leq x < 0$, we accept d_3;
for $0 \leq x < 1 - a$, we accept d_2 with probability p_2 and d_3 with probability $1 - p_2$;
for $1 - a \leq x \leq 1$, we accept d_1 with probability p_1 and d_2 with probability $1 - p_1$;

where $p_1 = \frac{1+a}{2+a}$ and $p_2 = \frac{1}{(2+a)(1-a)}$. For this procedure, $r(F_1) = r(F_2) = r(F_3) = \frac{1}{2+a}$.

Consider the alternative procedure which is exactly the same as the minimax procedure except that for $1 - a \leq x \leq 1$ we always accept d_1. For this procedure,

$$r(F_1) = 0; \qquad r(F_2) = \frac{1}{2+a} + \frac{a}{2+a}; \qquad r(F_3) = \frac{1}{2+a}.$$

Thus the alternative procedure reduces the risk from $\frac{1}{2+a}$ to 0 under F_1, increases it under F_2 by $\frac{a}{2+a}\left(\frac{a}{a+2} \to 0 \text{ as } a \to 0\right)$, and leaves it unaltered under F_3. The alternative looks more attractive.

This last example can be altered slightly so as to appear more striking. We can replace the distribution $F_1(x)$ by a sequence of distributions $F^1(x), F^2(x), \cdots, F^n(x), \cdots$, where $F^n(x)$ is the uniform distribution on the interval

$$(1 - a) + \frac{a}{n+1} < x \leq (1 - a) + \frac{a}{n}.$$

Call this interval I_n, $(n = 1, 2, \cdots)$. Corresponding to the distributions $F_2(x)$, $F_3(x), F^1(x), \cdots, F^n(x), \cdots$, there are decisions $d_2, d_3, d^1, \cdots, d^n, \cdots$.

An admissible minimax procedure is described as follows:

for $-a \leq x < 0$, accept d_3;
for $0 \leq x \leq 1 - a$, accept d_2 with probability p_2 and d_3 with probability $1 - p_2$;
for $x \varepsilon I_n$, accept d^n with probability p_1 and d_2 with probability $1 - p_1$;

where $p_1 = \frac{1+a}{2+a}$ and $p_2 = \frac{1}{(2+a)(1-a)}$. For this procedure,

$$r(F_2) = r(F_3) = \frac{1}{2+a} = r(F^j) \qquad \text{for } j = 1, 2, \cdots.$$

Consider the following alternative procedure:

for $-a \leq x < 0$, accept d_3;
for $0 \leq x \leq 1 - a$, accept d_2 with probability p_2 and d_3 with probability $1 - p_2$;
for $x \varepsilon I_n$, accept d^n;

where $p_2 = \frac{1}{(2+a)(1-a)}$. For this procedure,

$$r(F_2) = \frac{1}{2+a} + \frac{a}{2+a}; \qquad r(F_3) = \frac{1}{2+a}; \qquad r(F^j) = 0, \qquad j = 1, 2, \cdots.$$

If a is sufficiently small, the alternative procedure is certainly almost subminimax in the sense of our third paragraph: the maximum of the risk of the alternative procedure is only $\frac{a}{2+a}$ greater than that of the minimax procedure, and the alternative procedure has reduced the risk to zero for all except two of the distributions.

A decision procedure for deciding which of a class of distribution functions is the true distribution of X is said to be unbiased for F_i if $\text{Prob}(d_i \mid F_i) \geq$

$\text{Prob}(d_j \mid F_i)$ for all j. If a procedure is not unbiased for F_i, it will be said to be biased for F_i. In the next example every minimax procedure is biased for F_1.

Let $F_4(x)$ be the uniform distribution on the interval 0 to $1 - a$, with $a \leq \frac{1}{4}$. The problem is to decide between F_1, F_2, F_3, and F_4. An admissible minimax procedure is described as follows:

for $-a \leq x < 0$, accept d_3;
for $0 \leq x < 1 - a$, accept d_2 with probability p_2,
accept d_3 with probability p_3,
accept d_4 with probability $1 - p_2 - p_3$;
for $1 - a \leq x \leq 1$, accept d_1 with probability p_1,
accept d_2 with probability $1 - p_1$;

where $$p_1 = \frac{1 + a}{3}, \qquad p_2 = \frac{1 - a + a^2}{3(1 - a)}, \qquad p_3 = \frac{1 - 2a}{3(1 - a)}.$$

Thus, we have $\text{Prob}(d_2|F_1) = \dfrac{2 - a}{3} > \dfrac{1 + a}{3} = \text{Prob}(d_1|F_1)$.

This shows that the procedure is biased for F_1. By altering the procedure so that $p_2 = p_3 = \frac{1}{3}$ and $p_1 = 1$, we obtain a procedure which is unbiased for all F_i, and whose maximum risk is increased by only $\frac{2}{3}a$ over the minimax risk of $\dfrac{2 - a}{3}$.

The above example may be altered in the same way as the example of an almost subminimax solution so that there are infinitely many distributions for all but three of which the minimax solution is biased. In fact, it is possible to construct an example of a biased minimax solution for deciding among any number of distributions greater than two. It is impossible for a minimax procedure to be biased when there are only two distributions.

Along similar lines an example can be constructed for any $\epsilon > 0$ for a continuum of distributions, where any minimax procedure has constant risk of $\frac{2}{3} - \epsilon$ and is biased for all but three distributions, and where there exists an alternative almost subminimax procedure which is unbiased for all distributions and which reduces the risk to zero for all but three distributions where it is increased by less than 3ϵ.

REFERENCE

[1] H. Robbins, "Asymptotically subminimax solutions of compound statistical decision problems." *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951.