THE PROBLEM OF ESTIMATION

By H. Steinhaus

University of Wrocław

0. Introduction. A treatment has proved successful in m cases out of n; what is its efficacy? In other words: What will be the frequency of successes if we continue to apply the same treatment in all cases of the disease it was invented for? A lot has been examined by sampling and m items out of n proved defective; what is the best estimate for the fraction defective in the whole lot? This is another example of the same question. A hundred people of a certain tribe have been classified as to their belonging to the blood groups A, B, AB and O; how are we to estimate the frequencies of the 4 groups in the whole tribe? It is not an overstatement to say that the abundance of applications gives the problem of estimation an importance sufficient to rank it among the principal problems of science. This is the reason why, for more than a century, it has not ceased to haunt the ingenuity of mathematicians.

To make things quite plain let us imagine a statistician (S) compelled by the devil (D) to play the following game:

The devil has a collection of coins and he knows for each of them its probability p of showing heads as a result of tossing; he is rich enough to have specimens for any p in (0, 1). He chooses a coin suiting his fancy and lets the statistician throw it n times; it shows heads m times and it is up to the statistician to give an estimate p' of the value p which is known to D but unknown to him. This being done, S pays to D \$$(p' - p)^2$. S tries his best to reduce his loss by an appropriate method of guessing, as far as possible. If he succeeds in finding the best method he will not regret the money lost: his will be the fame of having solved our problem of the best estimate, if "best" is understood as minimizing the expected square error of the estimation; the rules of the game have been fixed by the devil in accordance with such an interpretation of our problem.

The following remark may explain the link connecting our game with the problem of point-estimation. The classical method solves this problem assuming that the distribution of coins employed by the devil is known to the statistician. He computes his guess, combining by Bayes' rule this knowledge with the observed result of tossing; the guess p' will be equal to the a posteriori value Ep. It can be defined also as the value of p' that minimizes the expected loss $E(p' - p)^2$. Thus the rules of our game correspond to the problem of point-estimation in the case of a known prior distribution; it is not artificial to employ them in the case of an unknown prior distribution.

I proposed this problem in 1954 [10] in Prague and in Berlin, calling it "Das statistische Spiel," but it is only in 1956 that I have been informed by L. J. Savage that it has already been solved: J. L. Hodges, Jr., and E. L. Lehmann [4]

attribute the priority to Herman Rubin in their important paper published in 1950. Another paper [3] by M. A. Girshick and L. J. Savage deals thoroughly with the same question; the results have already been included in books [2] and [8]. The generality of the results reached by the American school and their priority make it necessary to explain the reasons for publishing this paper: the first is that my exposition will perhaps be easier to read for some readers who do not feel comfortable when spoken to in the language of consummate specialists; the second is to be found in the last sections, which deal with some cases not covered by the authors quoted above.

1. A statistician who does not believe in probabilities when he is faced with a single game does not even look at the gambling table. He says: "The number m gives me no information whatever about p. My method is to declare p equal to $\frac{1}{2}$. In the worst case the true p will be 0 or 1 and I will lose 25 cents. Any gambler adopting a system different from mine exposes himself to a greater loss than a quarter of a dollar: if his p' is greater than $\frac{1}{2}$ he loses more than 25 cents for a very small p and if his p' surpasses $\frac{1}{2}$ his loss is greater than 25 cents for a p near unity. This is the reason why I abstain from counting heads and tails."

The method described above denies the possibility of statistical estimation, since its estimate is not influenced by the results of experiments. It has, however, the advantage of the certainty it gives to S of never losing more than 25 cents, a guarantee that no other method is capable of. From this point of view we can define 25 cents as the "value of the game." The devil's point of view, however, is different. As he is never sure that S will not guess the exact p, in which case his gain is nil, the "value of the game" is 0 when computed by D. In both cases we consider only the gains of D; nevertheless we get two different values: 25 cents and 0. Such games may be called "open."

2. As is generally known, many mathematical theories were invented to master our problem by the calculus of probability. Laplace did not hesitate to apply to it the so-called inverse probabilities or "probability of causes" derived from the observation of effects. To avoid misunderstandings we will speak of frequencies of different coins instead of the probabilities of their having been chosen by D. Such an approach implies the substitution of a sequence of games for a single game, and of the mean loss for the single loss. Following this principle, we have to imagine S being condemned by D to perpetual gambling. The devil is allowed to choose a new coin in each round of the infinite sequence Q of rounds, but n remains unaltered. It is the statistician's business to make his mean loss as small as possible. He could do it easily if he knew the frequency of different coins employed. To be more explicit, let us assume that D adheres to a certain function $f(p)$, the "density," and that S knows this function. In other words, he knows that for any two numbers u, v ($0 \leqq u \leqq v \leqq 1$) the relative frequency of the coins with $u \leqq p \leqq v$ is $\int_u^v f(p) \, dp$. The function $f(p)$ is integrable, nonnegative and $\int_0^1 f(p) \, dp = 1$. S puts the following problem to himself: the whole sequence Q of games can be split up into $n + 1$ subsequences Q_m ($m = 0, 1, 2, \cdots, n$),

Q_m being the subsequence consisting of all rounds showing the same number m of heads. The mean loss in such a subsequence is proportional to the integral

$$(1) \qquad I(x, m) = \int_0^1 f(p) \binom{n}{m} p^m q^{n-m} (p - x)^2 \, dp \qquad (q = 1 - p),$$

where x is the statistician's answer to m heads. Thus he has to find a function $p' = g(m, n)$ which gives to $I(x, m)$ the least value when substituted for x in the integral (1). Since the $(n + 1)$ values p' takes for $m = 0, 1, \cdots, n$ are independent, S has to solve the problem $I(x, m) = $ minimum separately for each m. This leads to a function $g(m, n)$ that minimizes the mean loss in the whole sequence Q. The problem is easily solved, for x is the root of the equation $dI(x, m) \,/\, dx = 0$. Calling it x_m we get immediately

$$(2) \qquad x_m = \int_0^1 f(p) p^{m+1} q^{n-m} \, dp \Big/ \int_0^1 f(p) p^m q^{n-m} \, dp = g(m, n).$$

Formula (2) is the clue to the problem: S has to declare $p' = g(m, n)$ in every game, $g(m, n)$ being defined by the ratio (2). It is to be noted that the integral (1) is a nonnegative quadratic function of x, which implies that the root of $I' = 0$ is a perfect minimum of I: for any x different from x_m, we get $I(x, m) > I(x_m, m)$.

A nontrivial case is given by the Bayes hypothesis $f(p) \equiv 1$. Formula (2) then leads immediately to

$$(3) \qquad g(m, n) = \frac{m + 1}{n + 2}$$

as the best method for S. Laplace employed such a hypothesis and (3), and he has been rebuked for it by his successors. An example of another kind is the favorite method of many naturalists and other scientists, who put $p' = m/n$. We now have to determine $f(p)$ so as to satisfy (2) identically in m (n is fixed once for all) with $g(m, n)$ replaced by m/n. Putting $m = n$ we immediately get the condition

$$(4) \qquad \int_0^1 f(p) p^{n+1} \, dp = \int_0^1 f(p) p^n \, dp.$$

The functions integrated being nonnegative and p^{n+1} less than p^n everywhere in $(0, 1)$, (4) implies $f(p) \equiv 0$ (almost everywhere), which contradicts the condition $\int_0^1 f(p) \, dp = 1$, essential for any density. Thus the estimate m/n does not correspond to any density $f(p)$ whatever. It is a remarkable fact that some surgeons have reached the same conclusion without being mathematicians and without knowing how to improve the estimate [7].

3. The modern theory of games as codified in [5] and extended, for example, in [12], guarantees the existence of a solution of our problem. It proves, for such two-person games as are considered here, the existence of a density $f(p)$ of a sequence $\{g_j(m, n)\}$ of guessing-methods, and of a number V (called the value of the game) such that

1^0 the statistician applying the guess $p' = g^{(j)}$ (m, n) in the jth round is sure to lose no more than V in the mean whatever the devil does, the mean being taken over the results of the whole sequence Q.

2^0 the devil who mixes his coins in conformity to the frequency function $f(p)$ is sure to win in the mean at least V regardless of what S does. The use of the frequency function $f(p)$ by D is to be understood as the use of a sequence of coins such that the corresponding sequence $\{p^{(j)}\}$ has a relative frequency $\int_u^v f(p)\ dp$ of terms belonging to the interval (u, v) whatever this interval may be $(0 \leqq u \leqq v \leqq 1)$. It follows immediately from 1^0 and 2^0 that

3^0 the "mixed strategy" defined by $\{g^{(j)}\}$ is the best for S in the minimax sense of "best."

It is also clear that V, the "value of the game," is the fixed indemnity which, when paid by D to S before every choice of a coin, makes the play fair; such settlement would not influence our theory.

The importance of the theorem of J. von Neumann leading (through its extensions) to the statements mentioned lies, as far as our endeavor is concerned, in the assurance that it is worthwhile to seek an effective formula for the estimate p'. Nevertheless, our reader is neither supposed to know the theory of games nor the somewhat intricate terminology of the books quoted ([5] and [12]). Our argument will be independent of the general theorem, and it will yield in our special problem more than is promised by 1^0, 2^0, and 3^0. The difference between the result of Sec. 1 and the objective of the following sections is also worth noticing: we could not guarantee to S a loss less than 25 cents in a single round but we can—as will be shown—reduce his mean loss in a sequence of rounds to a quantity V less than 25 cents by finding a minimax solution which admits no further improvement.

It is necessary to explain why we are not entirely satisfied by the promise contained in 1^0, 2^0, and 3^0. What we need is an estimate $p' = g(m, n)$ valid for the whole sequence Q of successive rounds; estimates varying from round to round would introduce an indeterminism of a dangerous kind which would make the scientist utter different judgments in identical situations: if $g_j(m, n) \neq g_h(m, n)$ for the same couple m, n, he is compelled to estimate differently the efficacy of two antibiotics, although both of them gave m positive results in n experiments; he is compelled to such behavior because the first experiment was scheduled "n_0, j" and the other "n_0, h." In Sec. 1 we have seen that no satisfactory estimate can be reached by making it an absolute constant; now we reject estimates which have the drawback of variability, though it is not universally agreed that such estimates are irrational. The case against them is forcibly put by R. A. Fisher, pp. 97–98 of [1].

4. Let us suppose now that D has chosen once for all a particular coin characterized by p, whereas S employs the particular function $g(m, n)$ of (3) as his method of guessing, and let us call E the mean gain of the devil. We get

$$(5) \quad E = \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m} \left(p - \frac{m+1}{n+2} \right)^2$$

$$= \sum_{m=0}^{n} \binom{n}{m} \left[p^2 - 2p \frac{m+1}{n+2} + \left(\frac{m+1}{n+2} \right)^2 \right].$$

To compute E we make use of the trivial equality $\sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m} = 1$ and of the following identities:

$$(6) \quad \sum_{m=0}^{n} \binom{n}{m} m p^m q^{n-m} \equiv np, \qquad \sum_{m=0}^{n} \binom{n}{m} m^2 p^m q^{n-m} \equiv n^2 p^2 + npq;$$

they are, respectively, familiar formulae for the first and second moments of the binomial distribution. Eqs. (5) and (6) lead immediately to the simple result

$$E = \frac{(n-4)pq + 1}{(n+2)^2}.$$

Let us consider first the case $n = 4$. Then E ceases to depend on p and becomes equal to $1/36$. This circumstance makes the statistician who adopted the method $p' = (m+1)/6$ independent of all diabolic cunning: whatever D does, whether he sticks to one coin or changes them as he likes, S will lose in the long run $1/36$ as his mean expenditure, never more nor less. It follows therefrom that any method is the best for D as an answer to the method $p' = (m+1)/6$. In particular, the method defined by the density $f(p) = 1$ is such a relatively best method against the statistician's rule $p' = (m+1)/6$. What is, on the other hand, the statistician's adequate answer to the devil's rule $f(p) \equiv 1$? This answer has already been found in Sec. 2: $g(m, n) = (m+1)/(n+2)$, which is equal in our special case $n = 4$ to $(m+1)/6$. Thus we have two antagonistic methods for D and S, each of them being relatively best against the other. We can easily prove that the rule of guessing $p' = (m+1)/6$ is not only relatively best but also absolutely best, i.e., that it fulfils the minimax criteria for best strategies. The proof runs as follows: The rule $p' = (m+1)/6$ reduces the loss of S to $1/36$ whatever D does. On the other hand, the same rule is the best answer against D's employing the system $f(p) \equiv 1$. This last statement implies the impossibility of S's reducing his loss below $1/36$ against the devil's playing as mentioned. Therefore no method gives a certainty to S of a loss below $1/36$; as the rule above gives a guarantee of a loss equal to $1/36$, no rule is better in the minimax sense, Q.E.D. The devil's rule $f(p) \equiv 1$ is also absolutely best for D, a theorem which can also be proved almost immediately. Both proofs are superfluous for a reader knowing the general theorem about games that two methods relatively best against each other are also absolutely best for the opponents employing them. The existence of such a pair of methods shows, in our terminology, that the game is closed. The result of their being constantly applied by both sides is a mean loss equal to the value of the game, which is $1/36$ in our case.

Thus we have solved our problem of the best estimate for $n = 4$. The solution is $p' = (m + 1)/6$.

Let us try to free ourselves from the restriction $n = 4$ by setting more generally,

$$(7) \qquad\qquad p' = g(m, n) = \frac{m + a}{n + b}$$

and assuming the independence of a and b from m but not from n. To compute E in this case we have to write

$$(8) \qquad\qquad E = \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m} \left(p - \frac{m + a}{n + b} \right)^2$$

instead of (5), and to repeat almost verbally the trivial computations of the special case $a = 1, b = 2$; they give now

$$(9) \qquad\qquad E = \frac{npq + (bp - a)^2}{(n + b)^2} \qquad\qquad (q = 1 - p).$$

The numerator $np(1 - p) + (bp - a)^2$ ceases to depend on p just if we put $a = \sqrt{n}/2,\ b = \sqrt{n}$. These values when substituted in (7) for a and b yield the formula

$$(10) \qquad\qquad p' = g(m, n) = \frac{m + \sqrt{n}/2}{n + \sqrt{n}} = G(m, n)$$

as a guessing method for S. When substituted in (9), they give

$$(11) \qquad\qquad \Lambda = \frac{1}{4(\sqrt{n} + 1)^2}$$

as the mean loss of the statistician, whatever the sequence of coins employed by the devil. To follow the path traced in the special case $n = 4$, we have to find a density $f(p)$ which satisfies (2) when we write for $g(m, n)$ the function G defined by (10). Let us try to this effect

$$f(p) \equiv c(pq)^s (s > -1) \quad \text{with} \quad 1/c = \int_0^1 (pq)^s\, dp.$$

The integrals in (2) now take the form

$$c \int_0^1 p^{s+m+1}(1 - p)^{s+n-m}\, dp, \qquad c \int_0^1 p^{s+m}(1 - p)^{s+n-m}\, dp,$$

and their respective values are

$$\frac{\Gamma(s + m + 2)\, \Gamma(s + n - m + 1)}{\Gamma(2s + n + 3)}, \qquad \frac{\Gamma(s + m + 1)\, \Gamma(s + n - m + 1)}{\Gamma(2s + n + 2)}.$$

Their ratio is $(s + m + 1)/(2s + n + 2)$, and condition (2) becomes

$$(12) \qquad\qquad \frac{m + s + 1}{n + 2s + 2} = \frac{m + \sqrt{n}/2}{n + \sqrt{n}}.$$

To satisfy (12), we set $s = \sqrt{n}/2 - 1$ and get

$$(13) \qquad f(p) \equiv (pq)^{\sqrt{n}/2-1} \cdot \frac{\Gamma(\sqrt{n})}{(\Gamma(\sqrt{n}/2))^2}$$

as the density sought to close the game. Thus the method G defined by (10) is the solution of our problem. It is the best estimate of p because it minimizes the mean square error. The case $n = 4$ is obviously contained as a special case in (10); it is the only case in which the hypothesis of uniform distribution a priori is consistent, in a certain sense at least, with our theory.

It is interesting to note that the classical rules—Bayes' hypothesis of uniform prior distribution and the estimate m/n—can both be saved if $(p - p')^2/pq$ is the amount to be paid by S to D. This is easily verified by writing $a = 0, b = 0$ in (7) and dividing (9) by pq to compute the expected gain of D in the new game; it is $1/n$ and thus independent of p. On the other hand, if we put $f(p) \equiv 1$ in (1) and replace $(p - x)^2$ by $(p - x)^2/pq$ to satisfy the new rules of the game, we are led by the same reasoning as employed for the old rule to the formula

$$x_m = \int_0^1 p^m q^{n-m-1} \, dp \Big/ \int_0^1 p^{m-1} q^{n-m-1} \, dp = m/n,$$

instead of (2).

5. The method of S given by G holds the devil tightly by his horns: he can neither increase nor diminish his mean gain which is equal to Λ as defined by (11). Λ is the value of the game and, consequently, the indemnity which, when paid in advance by D to S in every round, makes the whole game fair: this signifies that the mean balance of the sequence of rounds tends to zero with probability 1 if this indemnity is agreed upon and both partners play as well as possible. If S adheres to G the devil's lack of skill does not alter the limit. There are, however, bad methods for D (i.e., bad densities $f(p)$); if S knew that D has chosen such a method, he could apply a method securing himself a positive mean balance.

The expression

$$(14) \qquad \sqrt{\Lambda} = \tfrac{1}{2}(\sqrt{n} + 1)$$

may be called the mean error of the estimate (10). We emphasize that it is dependent exclusively on n. This enables the scientist to plan his experiments with a desired accuracy: he gets by an adequate choice of n the accuracy desired, neither more nor less, which is very important (e.g., the estimation of a lot by counting good and defective items in a sample of n has to avoid samples too great and too small: the first cost too much, the second give insufficient information). To appreciate this feature of the estimate (10) we may compare it with the "confidence interval" method where the length of the interval, given the confidence level, is known only after the experiment, because it depends on both n and m. Finally, our formulae are extremely simple, and can easily be tabulated for practical purposes.

6. Our estimate G is free from the disadvantage of "mixed strategies" antici-
pated at the end of Sec. 3. We may put the question of its uniqueness. Stanisław
Trybuła has found the following simple proof [11] that there is no other estimate
equivalent to G. It has already been proved here that there is no *better* estimate;
we have the author's consent to repeat here his proof of uniqueness. The proof
that there is no other estimate *equivalent* to G is simple. It is stated, in more
general form in [4], where the estimate was first announced. I happened to hear
it first from Stanisław Trybuła who rediscovered it in connection with [11]. You
have already seen that there is no better *estimate*; and now here is the proof of
uniqueness.

We have already stated in Sec. 2 that the solution (2) of the minimum problem
for the integral $I(x, m)$ defined in (1) is unique:

(15) $I(x, m) > I(x_m, m)$ for every x different from x_m .

Suppose that S adopts any method M, of the form $q(m, n)$ different from G.
Let the devil choose the method defined by (13). Since this density has been
shown to give G when substituted in (2), we see by (15) that there is at least one
subsequence G_m giving to D a greater gain in the mean than the gain he would
obtain if S applied the method M: this subsequence is generated by the subscript
m for which M differs from $G(m, n)$; there is at least one such value m among
$0, 1, \cdots, n$. For any values m, for which M equals $G(m, n)$ the corresponding
subsequences G_m give the same gain to D as he would gather if opposed by $G(m, n)$.
Thus S will suffer in the mean over all rounds played a loss greater than L, if
he applies M and is opposed by (13) as the devil's method. Thus M is not a
minimax solution, which proves the uniqueness of our minimax solution $G(m, n)$.

The question of uniqueness of the devil's best method has a different answer.
It too has been found by Trybuła [11], and it is negative. There are densities
different from (13) that are equivalent to it as gambling methods in every respect.
They are, however, not so simple.

We have already seen that there does not exist a single coin p_0 as a "rigid
method" for D; the reason is the possibility of S's choosing $p' = p_0$ if D's method
is to always use the same coin p_0 . Nevertheless, there are other methods, which
correspond neither to densities nor to constants, to speak more generally, which
have no distributions. We have omitted taking them into account; because our
argument led first to a method G for S which made his mean loss independent of
all methods of D, the distributionless methods included, and we had only to
find one definite method for D which could never be better answered than by G.
Thus we can admit all possible methods of D, having a distribution or not; the
theory expounded remains valid without restrictions. As the devil represents
nature, we have no reasons to jeopardize his freedom; it is an advantage for S,
the human player, that our theory enables him to master the situation in the
general case.

In all cases, when the existence of limits is asserted, the assertions are true
with probability one. There are other points where we have sinned against the

professional vocabulary, e.g., when speaking about the "mean error"; it is rather a metaphor: we always think of the expression

$$\lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} (p^{(j)} - p'^{(j)})^2.$$

For practical purposes this limit is a better measure for the error than any having the disadvantage of being dependent on m and of becoming 0 for $m = 0$ and for $m = n$. These disadvantages have shocked practicians and compelled them to leave the hard soil of sound statistics ([7], [9]).

Our method may be criticized as leading to a biased estimate. This means, in plain words, that in the case of estimating the same quality p over and over, say r times, we get r values m_j ($j = 1, \cdots, r$) and r estimates

$$p'^{(j)} = \frac{m_j + \sqrt{n}/2}{n + \sqrt{n}}$$

whose mean $1/r \sum_{j=1}^{r} p'^{(j)}$ does not tend to p for $r \to \infty$. This is, however, an artificial situation. If we have different lots L_j we are not entitled to suppose $p^{(j)}$ the same for all lots and, to be on the safe side, we must give separate estimates for each lot. If, however, we have to deal with the same lot r times, we will not compute its quality by taking the mean of the estimates but we will employ the formula (10) with $m = \sum_{j=1}^{r} m_j$, writing

$$p'_r = \frac{m + \sqrt{nr}/2}{nr + \sqrt{nr}}.$$

The estimate p'_r has, obviously, the property that $\lim_{r \to \infty} p'_r = p$.

Our method is essentially based on the measure of the error given by $(p - p')^2$. Of course, this measure is by no means the only possible one. In some applications the minimizing of the average error $|p - p'|$ would be more realistic; in some cases even some asymmetric function of the difference $p - p'$ could claim to be the best measure of the error in view of the economic consequences of a bad estimate which ought to be minimized. Notwithstanding this objection, we do not see any possibility of covering all the statistical applications by a single formula and must either abstain from giving any general rule or adopt the principle of least squares as being by far the best known and most applied. Its simplicity is inherited by our formula; no tables are necessary to employ it for small n, and there is no distinction between small and large n to be taken care of.

7. The comparison of our estimate (10) and that called the "confidence interval" [6] leads to the following remarks.

The confidence interval is defined by a "confidence level" and by the postulate of being the shortest one, given m, n and the confidence level. It is not a direct estimate of the unknown p. In many cases, however, such a direct estimate is necessary; let us think, for instance, of selling two kinds of some produce on the basis of a first experience which gave m and $n - m$ buyers respectively for kinds

A and B; the producer is compelled to choose a definite p' so that he may include in the next shipment the proper numbers of items A and B. Nevertheless we could consider our estimate as the middle point of an interval of the length $1 / (\sqrt{n} + 1)$. This length being the double of the expression (14), we may risk the conjecture that the confidence level for it is $0.6827 \cdots$ for n greater than, let us guess, 30. Thus our estimate would appear as a special case of the classical confidence interval: the case of confidence level 0.6827. This is, however, an illusion, because of the classical confidence interval's having a length depending on m which is not true for the length (14). We have already pointed out in Sec. 5 the importance of this independence.

The remarks above lead to the problem of modifying the so-called confidence interval by making its length independent of m. To speak again in the language of the theory of games, it would be necessary to let the devil and the statistician fix n and an arbitrary t ($0 \leqq t \leqq 1$) and let them gamble, the devil choosing the coin to be tossed and the statistician placing an interval J of length t on the real axis after having seen the result of n tossings. The stake would be \$1 paid by D to S if p is covered by J and by S to D in the opposite case. The problem for S would consist in placing the middle-point P of J as well as possible. He would have to find a formula $P(m, n; t)$ for the best location of P. This formula would be the statistician's best method of playing and would lead eventually to a value of the game, $V(n, t)$, independent of p. Solving the equation $\frac{1}{2} V(n, t) = v$ we would get $n = h(t, v)$; the function $h(t, v)$ would answer the following problem: given the confidence level v and the length t, to determine the number n of experiments needed to include the unknown p in an interval of length t with a confidence v. We have not tried to solve this problem; we mention it here to show distinctly the difference between the theory of Neyman's confidence intervals and our line of approach.

We pass now to another question; it is the last of the three examples spoken of in the introduction.

Instead of the simple alternative, or dichotomy, for which the solution has already been described, we have here four possible results of tossing. Let us more generally suppose that there are k possible outcomes, $k > 2$. The coin must be replaced by a k-hedron, every face being painted differently from the others; D knows the respective probabilities $p_i (i = 1, 2, \cdots, k)$ for all faces; he has in his collection a model for every set $\{p_i\}$; obviously $\sum_{c=1}^{k} p_i = 1$. The result of n tossings with a definite k-hedron is a set $\{m_i\}$; m_i ($i = 1, 2, \cdots, k$) are the respective numbers of tossings resulting in the ith face up; obviously $\sum_{i=1}^{k} m_i = n$. S has to estimate $\{p_i\}$ knowing $\{m_i\}$; his estimate $\{p_i\}$ has to minimize the expected value of

(16) $$\sum_{i=1}^{k} (p_i' - p_i)^2.$$

He tries to do it by putting

(17) $$p_i' = \frac{m_i + a}{n + b} \qquad (i = 1, 2, \cdots, k).$$

We now have, instead of (5), the expression

$$
(18) \quad E = \sum_{\{m\}} \frac{n!}{m_1! \, m_2! \cdots m_k!} \, p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}
$$
$$
\cdot \left\{ \left(p_1 - \frac{m_1 + a}{n + b} \right)^2 + \cdots + \left(p_k - \frac{m_k + a}{n + b} \right)^2 \right\},
$$

the summation being extended over all systems of nonnegative integers m_i with $\sum_{i=1}^{k} m_i = n$. Let us denote by E_1 the part of E which is left after cancelling all squares in $\{\ \}$ in (18) but the first. If we denote m_1 by m, p_1 by p, and $\sum_{i=2}^{k} p_i$ by q we get immediately the same expression for E_1 that we have called E in (8). This gives, in view of (9)

$$
E_i = \frac{1}{(n+b)^2} \cdot (np_i(1 - p_i) + (bp_i - a)^2).
$$

The same applies to each of the k parts E_i of (18) and leads, eventually, to

$$
(19) \quad E = \sum_{i=1}^{k} E_i = \frac{1}{(n+b)^2} \sum_{i=1}^{k} \{ np_i(1 - p_i) + (bp_i - a)^2 \}
$$
$$
= \frac{1}{(n+b)^2} \left\{ n + (b^2 - n) \sum_{i=1}^{k} p_i^2 - 2ab + ka^2 \right\}.
$$

This expression becomes independent of the variables p_i if and only if we put $b = \sqrt{n}$; it then takes the form

$$
(20) \quad E = \frac{1}{(n + \sqrt{n})^2} (n - 2a\sqrt{n} + ka^2).
$$

To determine a we introduce the condition

$$
(21) \quad \sum_{i=1}^{k} p_i' = 1,
$$

obviously needed in most applications. Thus we get from (17) with the value $b = \sqrt{n}$

$$
(22) \quad 1 = \sum_{i=1}^{k} \frac{m_i + a}{n + b} = \frac{n + ka}{n + \sqrt{n}},
$$

and, from (22), $a = \sqrt{n}/k$. We note with gratification that this value of a necessary to insure (21) is also just the value needed in (20) to minimize E without regard for (21). The formulae (17) now become

$$
(23) \quad p_i' = \frac{m_i + \sqrt{n}/k}{n + \sqrt{n}} \qquad (i = 1, 2, \cdots, k).
$$

Putting $a = \sqrt{n}/k$ into (20) we get

$$
(24) \quad E = \frac{n}{(n + \sqrt{n})^2} \cdot \frac{k - 1}{k},
$$

an expression independent of $\{p_i\}$.

To attain our objective by pursuing the analogy with the case $k = 2$, we must find a density on the hyperplane

$$(25) \qquad \sum_{i=1}^{k} p_i = 1 \quad (0 \le p_i \le 1; i = 1, 2, \cdots, k),$$

for which the estimates (23) minimize the expected value of $\sum_{i=1}^{k} (p_i' - p_i)^2$. Let us try the density $f(p_1 \, p_2 \cdots p_k) = C(p_1 \, p_2 \cdots p_k)^s$, where the constants C and s will be determined later. Supposing that the statistician knows that the frequencies of k-hedra employed successively by the devil are distributed according to this density (each point on the hyperplane (25) defines a k-hedron and the converse is also true), he will have to determine his guesses $\{x_i\}$, after having observed the result $\{m_i\}$, so as to minimize the expected loss that is given by the positive quadratic form

$$
\begin{aligned}
(26) \quad & F(x_1, x_2, \cdots, x_k) \\
& = C \frac{n!}{m_1! \cdots m_k!} \int_{p_1+\cdots+p_k=1} \cdots \int (p_1 \cdots p_k)^s p_1^{m_1} \cdots p_k^{m_k} \sum (p_i - x_i)^2 \, dw.
\end{aligned}
$$

The integral in (26) is extended over the domain defined by (25) and dw is the element of the hyperplane (25). The minimum problem is equivalent to the set of equations $\partial F/\partial x_i = 0$ $(i = 1, 2, \cdots, k)$ which are obviously

$$
\begin{aligned}
(27) \quad & \int_{p_1+\cdots+p_k=1} \cdots \int \\
& \cdot (p_1 \, p_2 \cdots p_k)^s p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k} (p_i - x_i) \, dw = 0 \qquad (i = 1, 2, \cdots, k)
\end{aligned}
$$

and the method of guessing defined by (23) will be proved best in the minimax sense if the set $\{x_i\}$ solving (27) can be identified with the set $\{p_i'\}$ given by (23). This amounts to k conditions of which the first is

$$(28) \qquad \frac{\int \cdots \int p_1^{m_1+s+1} p_2^{m_2+s} \cdots p_k^{m_k+s} \, dw}{\int \cdots \int p_1^{m_1+s} p_2^{m_2+s} \cdots p_k^{m_k+s} \, dw} = \frac{m_1 + \sqrt{n}/k}{n + \sqrt{n}}.$$

The numerator L and the denominator M in (28) can be brought into the forms

$$
\begin{aligned}
(29) \quad & L = \frac{\Gamma(m_1 + s + 2)\Gamma(m_2 + s + 1) \cdots \Gamma(m_k + s + 1)}{\Gamma[n + k(s + 1) + 1]} \\
& M = \frac{\Gamma(m_1 + s + 1)\Gamma(m_2 + s + 1) \cdots \Gamma(m_k + s + 1)}{\Gamma[n + k(s + 1)]}
\end{aligned}
$$

respectively, if the classical identity [13]

$$(30) \qquad \iint \cdots \int \prod_{i=1}^{k} p_i^{\alpha_i - 1} \, dw = \frac{\prod \Gamma(\alpha_i)}{\Gamma(\sum \alpha_i)}$$

is applied to both of them. (29) immediately yields

$$(31) \qquad \frac{L}{M} = \frac{m_1 + s + 1}{n + k(s + 1)}$$

and the condition (28) leads, through (31), to

$$(32) \qquad \frac{m_1 + s + 1}{n + k(s + 1)} = \frac{m_1 + \sqrt{n}/k}{n + \sqrt{n}},$$

which can be satisfied by setting $s + 1 = \sqrt{n}/k$; since the same reasoning applies to all k conditions of which (28) is the first, all will be satisfied by the value chosen for s. As to the constant C it can now be found from the condition

$$(33) \qquad C \int_{p_i \geq 0;\, p_1 + \cdots + p_k = 1} \cdots \int (p_1 p_2 \cdots p_k)^{\sqrt{n}/k - 1}\, dw = 1 \qquad (k > 2)$$

which gives

$$(34) \qquad C = \frac{\Gamma(\sqrt{n})}{\Gamma^k(\sqrt{n}/k)}$$

and, finally,

$$(35) \qquad \frac{\Gamma(\sqrt{n})}{\Gamma^k(\sqrt{n}/k)}\, (p_1 p_2 \cdots p_k)^{\sqrt{n}/k - 1}$$

is the density sought. If it is employed by the devil in an infinite game where the payment (16) in every round is agreed upon, the statistician's best answer is defined by (23), whether or not S is bound to respect the condition (21). This statement has just been proved; it enables us to apply once more the argument employed in the case $k = 2$ and to recognize the estimates (24) as the best in the minimax sense. The uniqueness of our solution follows by the reasoning already applied (for $k = 2$) in Sec. 6: as F in (26) shares with I in (1) the property of attaining its minimum at a single point, no modifications of the former reasoning are necessary.

8. If we drop the condition (21) we can write p for p_1, q for $1 - p$ and seek with S the best estimate of p_1 alone by the theory of the dichotomous case. This leads immediately to the solution (10) which may be written adequately

$$p_1' = \frac{(m_1 + \sqrt{n}/2)}{(n + \sqrt{n})};$$

we can, of course, replace the subscript 1 on both sides by $i(i = 1, 2, \cdots, k)$. We will have for the total loss (16) the mean $k\Lambda$, where Λ is given by (11). On the other hand, the loss has been computed in (24) under condition (21) as being equal to $4[(k - 1)/k]\Lambda$; since $4[(k - 1)/k]$ is less than k for $k > 2$, we arrive at the paradox that S can reduce his loss in the general case $k > 2$ by accepting the obligation (21). The explanation lies in the circumstance that the

loss Λ has been computed as minimax, taking into account the possibility of D's playing his best method, which has been found to be the density $c(pq)^s$. In the general case he can not apply the same method for every face of the k-hedron because he would have to choose a density $f(p_1, p_2, \cdots, p_k)$ that would have the form $c_i p_i(1 - q_i)$ for $i = 1, 2, \cdots, k$ simultaneously, which is impossible. Thus the plurality of faces hinders the devil from simultaneously playing all the alternative games contained in the general case as well as possible. This obstacle for D opens the way for S to play every alternative game in the general case better than he could do it in the simple alternative $k = 2$; the inequality $4[(k - 1)/k] < k$ is the proof that the advantage thus gained by S is greater than the disadvantage that results from the condition (21). In fact we remarked just after (22) that this disadvantage is purely illusory.

9. The method of this paper applies as well to other distributions as to the binomial. The simplest example is given by the Poisson distribution. The problem is to estimate the mean number of signals per unit time, having observed an arbitrary unit interval and having found k signals in it. To apply the theory of games, we must define the amount to be paid for an erroneous estimate. If, to mention an example of analytical interest, this amount is equal to $(c - c')^2/c$, c being the true mean number and c' the estimate, the problem is easily solved.

Let us try the guess $c' = k$. The expected loss for a given c is

$$(36) \qquad E = \sum_{k=0}^{\infty} \frac{(c - k)^2}{c} e^{-c} \cdot \frac{c^k}{k!} = cM_0 - 2M_1 + M_2/c,$$

the M_i $(i = 0, 1, 2)$ being the moments of order i of the Poisson distribution; this leads to

$$(37) \qquad E = c - 2c + (\omega + c^2)/c = 1,$$

ω being the variance of the random variable k, which in this instance is c. Thus the loss E does not depend on c. We now have to find a prior distribution which leads to the method $c' = k$ as the relatively best. We will show that the uniform distribution does so. To speak correctly, it will be the distribution of constant density $T/(T^2 - 1)$ $(T > 1)$ in $(1/T, T)$ for $T \to \infty$. We have to compute x, for a given k so as to minimize the expected loss given by

$$(38) \qquad \frac{T}{T^2 - 1} \int_{1/T}^{T} e^{-c} \cdot \frac{c^k (c - x)^2}{k!c} \, dc,$$

x being the guess we are seeking. The condition for the minimum is evidently

$$\int_{1/T}^{T} e^{-c} c^k \, dc = x \int_{1/T}^{T} e^{-c} c^{k-1} \, dc,$$

and becomes, for $T \to \infty$, $k! = x(k - 1)!$, $x = k$, Q.E.D.

It is interesting to note that no minimax solution exists if $(c - c')^2$ is fixed as the amount to be paid by S to D. The reason for this insolubility is to be sought in the freedom of the devil to increase c as far as he likes, putting for

instance $c_j = j$ or $c_j = 10^j$ in the jth round. It is easily seen that any guess c' far from k is worse for S than $c' = k$; on the other hand the expected value of $(c - k)^2$ is c, rising thus with c beyond every limit and increasing indefinitely the gain expected by D, whatever S does. The schedule of payments $(c - c')^2/c^2$ is also of practical interest, corresponding as it does to a penalty based on percentage error. Here again it is easy to see that no minimax solution exists.

10. Different objections to the new estimate have been raised:

1^0 The loss-function $(p - p')^2$ has been criticized as arbitrary. We have pointed out in the introduction that it is the result of an extrapolation of the estimate Ep of the old school; such a procedure is suggested by the circumstance that the symbol E loses its direct applicability when the prior distribution of p is either unknown or nonexistent.

2^0 The old estimate m/n has been compared with (10) as giving a better result, especially for n not too small. It has, however, been rejected not by mathematicians but by men of practice [7], who became aware of the paradox of error 0 for $m = 0$, $m = n$.

3^0 To save the minimax method, the Bayes hypothesis $f(p) \equiv 1$, and the estimate m/n, the loss-function $(p - p')^2/pq$ has been proposed, but no examples have been adduced for which this loss function seems realistic. There is also some interest in the loss function $(p - p')^2/(pq)^2$—for which no minimax estimate exists. There are examples in which this function corresponds to the real situation for p near 0 or for p near 1, but it is difficult to find natural problems in which the function measures the damage caused by bad guessing throughout the whole interval $0 < p < 1$.

4^0 The concept of a mathematically trained and satanically malicious devil has been denounced as leading to unnecessary caution; the estimate (10) is accused of paying too high a tribute to nature, which is incapable of seeking out such sophisticated devices against the mortals as given by (13). This objection may be answered as follows:

(a) We are compelled to play a game against nature ("the devil") and we do apply the new theory of games. If we are not allowed to do so we must abandon the new theory of games altogether and return to the old phraseology of "taking into account the psychology of the enemy." It is exactly the indefiniteness of such advice which made the new theory of games the only efficient tool in such problems as pursuit at sea; we need not worry whether the helmsman of the other boat is awake or not—he can just as well be an automaton and can be countered by our automaton.

(b) Let us drop the anthropomorphic idea of nature and admit that it acts blindly, without any strategy whatever. Such a view leads us naturally to deprive the devil of his sequence of coins of various frequencies. But even in such a situation it is difficult to imagine a better method for S against D than the new formulae: they do apply—not as minimax solutions but because of their independence of the p_j-sequence; they give a definite mean-square error, the only light in the total darkness.

5^0 Mr. A has been the first to investigate the blood-groups of a certain African tribe: he examined n' men and found m' with Rh-plus blood among them. Mr. B came second and examined another sample; his result was n'' and m'' respectively. Mr. B feels obliged to take into account the information contained in the paper published by A: he considers the result of his predecessor as furnishing a prior distribution and hesitates to apply (10) to his numbers. The sound procedure is, however, for B to publish his own numbers n'' and m'', to quote the numbers of A, and to estimate the Rh fraction of the tribe by setting $n = n' + n''$, $m = m' + m''$ in formula (10) with the remark: "making use of all available observations." This example illustrates the meaning of our arguments 4^0(a) and 4^0(b).

REFERENCES

[1] R. A. FISHER, *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh, 1956.

[2] M. A. GIRSHICK AND D. BLACKWELL, *Theory of Games and Statistical Decisions*. John Wiley and Sons, New York, 1954.

[3] M. A. GIRSHICK AND L. J. SAVAGE, "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951, pp. 53–74.

[4] J. L. HODGES, JR., AND E. L. LEHMANN, "Some problems in minimax point estimation," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 182–197.

[5] J. VON NEUMANN AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, 2nd ed., Princeton University Press, Princeton, N. J., 1947, pp. 143–168.

[6] J. NEYMAN, "On the problem of confidence intervals," *Ann. Math. Stat.*, Vol. 6 (1935), pp. 111–116.

[7] A. M. RITALA, "Zur Berechnung des statistischen mittleren Fehlers (standard error)," *Acta Societatis Medicorum Fennicae "Duodecim,"* Ser. B, 19, fasc. 2, Helsinki, 1933.

[8] L. J. SAVAGE, *The Foundations of Statistics*, John Wiley and Sons, New York, 1954, Sec. 13.4.

[9] H. STEINHAUS, "Sur l'interprétation des résultats statistiques," *Colloquium Mathematicum*, Vol. 1 (1948), pp. 232–238.

[10] H. STEINHAUS, "Uber einige prinzipielle Fragen der mathematischen Statistik," *Tagung über Wahrscheinlichkeitsrechnung und mathematische Statistik, Berlin 19–21. X. 1954*, Deutscher Verlag der Wissenschaften, V: Das statistische Spiel.

[11] S. TRYBUŁA, "Some problems of simultaneous minimax estimation," submitted for publication in *Ann. Math. Stat.*

[12] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.

[13] E. T. WHITTAKER AND G. N. WATSON, *Modern Analysis*, 4th ed., Cambridge University Press, Cambridge, 1940, Sec. 12.5.