THE PROBLEM OF ESTIMATION

By H. Steinhaus

University of Wroclaw

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]

Received July 30, 1956, revised April 22, 1957.