STATISTICAL RESEARCH REPORT
Institute of Mathematics
University of Oslo

No 4
September 1974

TESTS WITHOUT POWER

by

Erling Sverdrup

Tests without power

by

Erling Sverdrup

University of Oslo

For the purpose of the argument in this paper, the principles for determining a critical region for an hypothesis will be considered to be of three kinds.

(i) The "classical" test principle

The density under the null-hypothesis $p_o(x)$ and a statistic $T(x)$, expressing the degree of distrust of the hypothesis, are specified. With $P_o(A) = \int_A p_o(x)dx$ the hypothesis is rejected when $T(X)>K$, where $P_o(T(X)>K)\leq\epsilon$ for any p_o consistent with the null-hypothesis. Of course this is the most commonly used "principle" in practical statistics, even if it hardly deserves to be called a principle.

(ii) The Neyman-Pearson test principle

This expresses a desire to have high power (i.e. probability of rejecting when the hypothesis is wrong) with a given level of significance. In some simple situations this would lead to a rejection region of the form $p_1(X)>c\, p_o(X)$, where p_1 and p_o are some completely spesified densities under the alternative to the hypothesis and the hypothesis respectively and c is adjusted to the level of significance.

(iii) The Per Martin-Løf principle [1]

This consists in rejecting when $p_o(X)<c$, where c is adjusted to the level of significance (and X is a minimal sufficient statistic).

It is hardly necessary to discuss the Neyman-Pearson principle here, its merits are well known. I shall therefore concentrate on throwing som light on the Per Martin-Løf principle. Of course this principle is sweepingly more general than the two first mentioned principles since it disregards the alternatives except for the purpose of constructing the minimal sufficient X under the a priori assumtions.

The principle will be elucidated by two examples.

Example1. Card dealing in bridge

The hypothesis is that all $N = 52!/(13!)^4$ combinations of hands are equally likely. If now the dealer gives himself (say) 13 spades, one might become suspicious. This is not due to this hand being less likely than others, which it is not, but that according to the rules of the game this hand is favourable to a player. There are, in other words, other circumstances than those which follow from the density under the hypothesis that are taken into account.

To be more precise let

$$G_1 , G_2 , \ldots G_r$$

be r groups of four-hands-combinations which are such that all combinations x belonging to a fixed G_i are equally favourable to the dealer, whereas if $x \in G_i$ and $x' \in G_j$ with $j>i$ then x' is more favourable than x to the dealer. It certainly would be a formidable (really prohibitive) task to determine $G_1, \ldots ,G_r$, but in principle they are given. Let

$$N_1,N_2, \ldots ,N_r, \qquad (\sum_1^r N_i=N)$$

be the number of four-hands-combinations in $G_1 \ldots ,G_r$ respectively. We now let the test statistic be $T(x)$, where $T(x)$ is defined by $T(x)=t$ if $x \in G_t$. By the classicel test principle ((i) above) the hypothesis (of no cheating) should be rejected if $T(X)$ is large. This seems rather obvious from the definition of $T(x)$.

On the other hand it seems to be rather irrelevant whether $N_{T(X)}/N$ is large or small, i.e. which G_i contains few or many combinations. Thus it is the rules of the game, and a thorough knowledge of the game of bridge, which are required to determine the test. The test can not be constructed from the hypothesis alone. (It may be of some interest to realize that little can be gained by specifying the class of distributions when the dealer is cheating. Thus we have a test method "without power").

Example 2. Non-paying passengers on tramcars.

It is assumed that at most one out of thousand passengers is non-paying (by cheating). In a certain area at certain times of the day it is however suspected that the ratio is much higher. Hence a crude inspection is made in which the inspector finds the first non-paying passenger after x checks. We have under the hypothesis, that the suspicion is unjustified, that

$$\Pr(X = x) = p_o(x) = p(1-p)^{x-1}; x = 1,2,\ldots$$

where $p = 0.001$. Obviously the hypothesis should be rejected if X is small i.e. a non-paying passenger is quickly discovered. However, by the Per Martin-Løf principle the hypothesis should be rejected if X is large, since $p_o(x)$ is a decreasing function of x. (Per Martin-Løf assumes that $p_o(x)$ is constructed wholy by combinatorial methods starting from equally probable events. If this is an important part of the theory, one could assume that the number of passenger in the suspicious area is N and hence $(a=Np)$

$$p_o(x) = \Pr(X=x) = \binom{N-x}{a-1}/\binom{N}{a} =$$

$$= a\frac{(N-x)(N-x-1)\ldots(N-x-a+2)}{N(N-1)\ldots(N-a+1)}$$

which is still a decreasing function of x).

The present author has been used to think of the idea that the most unlikely events should result in rejection as "naive", arising among non-trained statisticians. He has felt that it is pre-Galileian in the sense that like the idea that the heavier bodies will attain the greatest speed, it is a natural misconception among earlier non-sophisticated physicists.

The present author has still the same feeling and has felt it necessary to voice a dissident opinion, which is of course a reckless undertaking among so many prominent congratulants. Fortunelately it has been possible to keep the argument on an elementary level.

Perhaps it should be mentioned that the idea of replacing likelihood ratios with entropy (or redundancy) has been developed in great detail in a book by Kullback [2] from 1959. His treatment is general in the sense that other alternatives to the hypothesis than the uniform distribution are allowed.

References

[1] Per Martin-Løf: "The notion of redundancy and its use as a quantitative measure of discrepancy between a statistical hypothesis and observational data". Scand. J. Statist.(1974) vol. 1, p.1.

[2] Solomon Kullback: Information theory and statistics, New York 1959.