STATISTICAL RESEARCH REPORT No 8
Institute of Mathematics 1971
University of Oslo

OPTIMUM TEST STATISTICS FOR CALCULATING SIGNIFICANCE PROBABILITIES

by

Emil Spjøtvoll

ABSTRACT

A test statistic is defined to be best for calculating significance probabilities if it maximizes the probability of getting a given small significance probability. It is shown that a test statistic which is best from a power function point of view is also best for calculating significance probabilities.

1. INTRODUCTION

Let X be a random variable with distribution belonging to a family $\{P_\theta : \theta \in \Omega\}$ of distributions. Consider a hypothesis testing problem with a simple null hypothesis P_{θ_0} and a simple alternative P_{θ_1} :

$$H : \theta = \theta_0 \text{ against } \theta = \theta_1 . \tag{1}$$

Let $T(X)$ be a statistic such that small values of $T(X)$ seem reasonable if $\theta = \theta_0$, and large values of $T(X)$ are not reasonable if $\theta = \theta_0$ but more reasonable if $\theta = \theta_1$.
The <u>significance probability</u> if the value x of X is observed is

$$\alpha_T(x) = P_{\theta_0} [T(X) \geq t(x)] . \tag{2}$$

We call $T(X)$ a <u>test statistic</u>. A small value of $\alpha_T(x)$ indicates that the null hypothesis is not true. The smaller $\alpha_T(x)$ is, the more we tend to disbelieve the null hypothesis. The use of the significance probability as a measure of how much the data contradicts the null hypothesis has been mentioned by several statisticians. In Fisher [2, p. 80] one can read "The actual value of P obtainable from the table by interpolation indicates the strength of the evidence against the null hypothesis". (P here corresponds to $\alpha_T(x)$.) Hodges and Lehmann [3, p. 283] state that "The significance probability may be thought of as giving, in a single convenient number, a measure of the degree of surprise which the experiment should cause a believer of the null hypothesis". Takeuchi [4, p. 1056] writes "Should the "size" of the test be always preassigned or can the "level of significance" be determined from the sample? This problem is not purely an academic one, although no satisfactory mathematical theory for the latter approach has yet been established, in most real applications, it is usually considered to be, I think, the more appropriate approach". Bahadur [1] used significance probabilities in his "stochastic comparison" of tests. It is the purpose of this paper to formulate a statistical theory for significance probabilities.

2. OPTIMUM STATISTICS FOR CALCULATING SIGNIFICANCE PROBABILITIES

If we have several test statistics for calculating significance probabilities in the problem (1), we would prefer the statistic which gives the smallest significance probability when the alternative is true. We formalize this in the following

Definition. $T(X)$ is a best test statistic at significance probability α if for any other statistic $U(X)$ we have

$$P_{\theta_1}[\alpha_T(X) \leq \alpha] \geq P_{\theta_1}[\alpha_U(X) \leq \alpha] . \tag{3}$$

$T(X)$ is a uniformly best test statistic if for any other statistic $U(X)$ (3) holds for all α, $0 \leq \alpha \leq 1$.

The following theorem gives us a connection between best level α tests in the ordinary sense and best test statistics for calculating significance probabilities.

Theorem 1. Given the problem (1). Suppose that there exists a random variable $T(X)$ and a UMP level α test which rejects when $T(X) \geq c$. Then $T(X)$ is a best test statistic at significance probability α for the problem (1).

Proof: Let

$$F_T(t) = P_{\theta_0}[T(X) < t] . \tag{4}$$

It is easily verified that for any random variable $T(X)$ we have

$$P_{\theta_0}[F_T(T(X)) \geq 1-t] \leq t . \tag{5}$$

Since $F(c) = 1 - \alpha$, we have

$$\{t : F_T(t) \geq 1-\alpha\} \supseteq \{t : t \geq c\} . \tag{6}$$

By (2) and (4)

$$\alpha_T(x) = 1 - P_{\theta_0}[T(X) < T(x)] = 1 - F_T(T(x)) , \tag{7}$$

from which we obtain

$$P_{\theta_i}[\alpha_T(X) \leq \alpha] = P_{\theta_i}[F_T(T(X)) \geq 1 - \alpha] \quad i = 0,1 . \tag{8}$$

By (6)

$$P_{\theta_i}[F_T(T(X)) \geq 1-\alpha] \geq P_{\theta_i}[T(X) \geq c] \quad i = 0,1 . \tag{9}$$

By (5) we have

$$P_{\theta_0}[F_T(T(X)) \geq 1 - \alpha] \leq \alpha. \tag{10}$$

The test of (1) which rejects when $F_T(T(X)) \geq 1-\alpha$ is therefore a level α test. By (9) its power at θ_1 is greater or equal than the power of the test based upon $T(X)$. Since the latter is most powerful, we must have equality sign in (9). Combining (8) and (9) we then get

$$P_{\theta_1}[\alpha_T(X) \leq \alpha] = P_{\theta_1}[T(X) \geq c] . \tag{11}$$

Let $U(X)$ be any other statistic which is used to calculate significance probablities

$$\alpha_U(x) = P_{\theta_0}[U(X) \geq U(x)] .$$

Let

$$F_U(u) = P_{\theta_0}[U(X) < u]$$

As in (8) we get

$$P_{\theta_i}[\alpha_U(X) \leq \alpha] = P_{\theta_i}[F_U(U(X)) \geq 1 - \alpha]. \tag{12}$$

By (5)

$$P_{\theta_0}[F_U(U(X)) \geq 1 - \alpha] \leq \alpha . \tag{13}$$

Hence the test which rejects when $F_U(U(X)) \geq 1-\alpha$ is a level test of (1). The power at θ_1 of this test is

$$P_{\theta_1}[F_U(U(X)) \geq 1-\alpha]. \tag{14}$$

Since the test based on $T(X)$ is as least as powerful we have that (14) is less than (11). Hence by (11) and (12)

$$P_{\theta_1}[\alpha_U(X) \leq \alpha] \leq P_{\theta_1}[\alpha_T(X) \leq \alpha] .$$

The theorem is proved.

Next, let both the null hypothesis and the alternative be composite:

$$H : \theta \in \omega \quad \text{against} \quad \theta \in \Omega - \omega, \tag{15}$$

Given a statistic $T(X)$, let

$$F_{T,\theta}(t) = P_\theta[T(X) < t] \tag{16}$$

$$\alpha_{T,\theta}(x) = P_\theta[T(X) \geq T(x)] = 1-F_{T,\theta}(T(x)) \tag{17}$$

$$\alpha_T(x) = \sup_{\theta\in\omega} \alpha_{T,\theta}(x) . \tag{18}$$

We shall call $\alpha_T(x)$ the significance probability when x is observed.

<u>Definition.</u> $T(X)$ is as least as good a test statistic as $U(X)$ at significance probability α and at the alternative $\theta \in \Omega-\omega$ if

$$P_\theta[\alpha_T(X) \leq \alpha] \geq P_\theta[\alpha_U(X) \leq \alpha] . \tag{19}$$

$T(X)$ is uniformly as good as $U(X)$ at the alternative θ if (19) holds for all α, $0 \leq \alpha \leq 1$.

$T(X)$ is uniformly as good as $U(X)$ at significance probability α if (19) holds for all $\theta \in \Omega - \omega$.

$T(X)$ is uniformly as good as $U(X)$ if (19) holds for all α, $0 \leq \alpha \leq 1$, and all $\theta \in \Omega-\omega$.

Theorem 2. Given the problem (15). Suppose that there exist a random variable $T(X)$ and a most powerful level α test against the alternative $\theta_1 \in \Omega - \omega$ which rejects H when $T(X) \geq c$. Then $T(X)$ is a best test statistic at significance probability α and at the alternative θ_1 for the problem (15).

Proof: Since by assumption

$$P_\theta[T(X) \geq c] = 1 - F_{T,\theta}(c) \leq \alpha, \ \theta \in \omega ,$$

we have

$$\inf_{\theta \in \omega} F_{T,\theta}(c) \geq 1-\alpha.$$

It follows that

$$\{t : \inf_{\theta \in \omega} F_{T,\theta}(t) \geq 1-\alpha\} \supseteq \{t : t \geq c\} . \tag{20}$$

By (17) and (18) for all θ

$$P_\theta[\alpha_T(X) \leq \alpha] = P_\theta[\inf_{\theta' \in \omega} F_{T,\theta'}(T(X) \geq 1-\alpha] . \tag{21}$$

From (20) and (21) we have for all θ

$$P_\theta[\inf_{\theta' \in \omega} F_{T,\theta'}(T(X)) \geq 1-\alpha] \geq P_\theta[T(X) \geq c] . \tag{22}$$

Furthermore, when $\theta \in \omega$, we have by (5)

$$P_\theta[\inf_{\theta' \in \omega} F_{T,\theta'}(T(X)) \geq 1-\alpha] \leq P_\theta[F_{T,\theta}(T(X) \geq 1-\alpha] \leq \alpha . \tag{23}$$

Therefore, the test which rejects H in (15) when $\inf_{\theta \in \omega} F_{T,\theta}(T(X)) \geq 1-\alpha$ is a level α test of H. By (22) its power at the alternative θ_1 is greater or equal to the power of the test based on $T(X)$. Since the latter is most powerful, we must have equality in (22) when $\theta = \theta_1$. Combining (21) and (22) we get

$$P_{\theta_1}[\alpha_T(X) \leq \alpha] = P_{\theta_1}[T(X) \geq c]. \tag{24}$$

Let $U(X)$ be any other statistic used to calculate significance probabilities, and let $F_{U,\theta}$ and α_U be defined in the same way as (16) and (18). We have as in (21) and (23)

$$P_\theta[\alpha_U(X) \leq \alpha] = P_\theta[\inf_{\theta' \in \omega} F_{U,\theta}(U(X)) \geq 1-\alpha] \quad \text{for all } \theta. \tag{25}$$

and

$$P_\theta[\inf_{\theta' \in \omega} F_{U,\theta}(U(X)) \geq 1-\alpha] \leq \alpha, \text{ when } \theta \in \omega .$$

Hence the test which rejects when $\inf_{\theta' \in \omega} F_{U,\theta}(U(X)) \geq 1-\alpha$ is a level test of (15). The power at the point θ_1 is

$$P_{\theta_1}[\inf_{\theta' \in \omega} F_{U,\theta}(U(X)) \geq 1-\alpha]$$

which must be equal or less than the power of the most powerful test which is $P_{\theta_1}[T(X) \geq c]$. Hence by (24) and (25)

$$P_{\theta_1}[\alpha_U(X) \leq \alpha] \leq P_{\theta_1}[\alpha_T(X) \leq \alpha], \tag{26}$$

which proves the theorem.

It is seen from the proofs of Theorems 1-2 that if we formulate the results of an experiment in terms of significance probabilities, then a standard question like: "How many observations do we need to get a preassigned power β at an alternative θ_1 for a level α test?" is equivalent to "How many observations do we need to get a significance probability $\leq \alpha$ with probability β when the alternative is θ_1 ?".

The results of Theorem 1 and 2 also hold if we are comparing test statistics in a restricted class of test statistics as, for example, invariant test statistics or unbiased test statistics.

That two test statistics which are equally good when comparing power functions, may not be equally good when calculating significance probabilities is seen from the following example.

Let X_1 and X_2 be independent random variables each with probability distribution

$$P[X=0] = P[X=1] = \tfrac{1}{2}(1-p), \ P[X=2] = p .$$

Consider the problem

$$H : p = \alpha \qquad \text{against} \qquad p > \alpha. \tag{27}$$

A level α test is given by rejecting the hypothesis $p = \alpha$ if $X_1 \geq 2$. The power of this test is p at the alternative p.

Let another test be based upon the sum $Y = X_1+X_2$. We have $P[Y = 0] = \frac{1}{4}(1-p)$, $P[Y = 1] = \frac{1}{2}(1-p)^2$, $P[Y = 2] = \frac{1}{4}(1-p)^2 + p(1-p)$, $P[Y = 3] = p(1-p)$, $P[Y = 4] = p^2$.

The test which rejects when $Y \geq 3$ is a level α test. Also this test has power p at the alternative p. Hence the two tests are equally good as tests of (27).

As test statistics for calculating significance probabilities, however, X_1 and Y are not equally good.

Let $\alpha_1(x) = P[X_1 \geq x]$ and $\alpha_2(y) = P[Y \geq y]$. Then

$$P_p[\alpha_1(X) \leq \gamma] = \begin{cases} 0 & \text{if} \quad \gamma < \alpha \\ p & \text{if} \quad \gamma \geq \alpha, \end{cases}$$

and

$$P_p[\alpha_2(X) \leq \gamma] = \begin{cases} 0 & \text{if} \quad \gamma < \alpha^2 \\ p^2 & \text{if} \quad \alpha^2 \leq \gamma < \alpha \\ p & \text{if} \quad \gamma \geq \alpha \end{cases}$$

Hence for calculating significance probabilities Y is better than X_1.

3. THE EXPECTED SIGNIFICANCE PROBABILITY

Consider again the problem (15). For a fixed alternative θ, let us consider the expected significance probability. We have the following theorem

<u>Theorem 3.</u> Consider the problem (15). Suppose that $T(X)$ is uniformly as good as $U(X)$ at the alternative θ. Then $E_\theta\ \alpha_U(X) \geq E_\theta\ \alpha_T(X)$.

Proof: The expected significance probability is

$$E_\theta\ \alpha_T(X) = \int_0^1 P_\theta[\alpha_T(X) > \alpha]\, d\alpha = \int_0^1 (1-P_\theta[\alpha_T(X) \leq \alpha])d\alpha$$

$$= 1 - \int_0^1 P_\theta[\alpha_T(X) \leq \alpha]d\alpha \leq 1 - \int_0^1 P_\theta[\alpha_U(X) \leq \alpha]d\alpha$$

$$= E_\theta\ \alpha_U(X)\ ,$$

where the inequality follows from (26).

Typically, the distribution of the significance probability is very skew, and the expectation is not a good measure of the distribution. As an example Let X have an exponential denisty $\lambda e^{-\lambda x}$, and consider the problem

$$H : \lambda = \lambda_o \quad \text{against} \quad \lambda < \lambda_o\ .$$

The UMP test rejects when $X \geq$ constant, hence we use X to calculate significance probabilities. The significance probability when $X = x$ is observed is

$$\alpha(x) = P[X \geq x] = e^{-\lambda_o x}\ .$$

The expected significance probability when λ obtains is

$$E\alpha(X) = \int e^{-\lambda_o x}\ \lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda+\lambda_o}\ .$$

The cdf of $Y = \alpha(X)$ is

$$G(y) = P_\lambda[\alpha(X) \leq y] = P_\lambda[\ e^{-\lambda_o X} \leq y]$$

$$= P[X \geq -(\log y)/\lambda_o] = y^{\lambda/\lambda_o}$$

Hence the density is

$$g(y) = (\lambda/\lambda_o)\ y^{(\lambda/\lambda_o)-1}\ .$$

Since $\lambda < \lambda_o$ this density is very skew.

The probability that $\alpha(X)$ exceeds its expected value is

$$P_\lambda[\alpha(X) \geq \frac{\lambda}{\lambda+\lambda_o}] = 1 - (\frac{\lambda}{\lambda+\lambda_o})^{\frac{\lambda}{\lambda_o}} = 1 - (\frac{1}{1+\lambda_o/\lambda})^{\lambda/\lambda_o}$$

When $\lambda \to 0$ this probability tends to 1 . Hence the expected value of $\alpha(X)$ tells us little about the distribution of $\alpha(X)$.

REFERENCES

[1] Bahadur, R.R., "Stockactic comparison of tests," Annals of Mathematical Statistics, 31 (1960), 276-95.

[2] Fisher, R.A., Statistical Methods for Research Workers, 13th ed., Hafner, New York, 1967.

[3] Hodges, J.L. jr. and Lehmann, E.L., Basic Concepts of Probability and Statistics, Holden-Day, San Francisco, 1964.

[4] Takeuchi, K., "Comments on Blyth's paper", Annals of Mathematical Statistics, 41 (1970), 1054-1058.