NOTES

NOTE ON RELATIVE EFFICIENCY OF TESTS

BY COLIN R. BLYTH[1]

University of Illinois and Stanford University

1. Summary. This note is concerned with possible definitions of relative efficiency for two sequences of tests of the same hypothesis. For two examples of one kind of definition, relative efficiencies of the Student test and sign test against normal alternatives are calculated for fixed sample size and asymptotically.

2. Introduction. Consider the following problem of relative efficiency of tests. Experiments $X_1, X_2, \cdots$ and two sequences $\{A_n(X_1, \cdots, X_n)\}$, $\{A_n^*(X_1, \cdots, X_n)\}$ of level α tests are available for testing the same hypothesis. We must decide whether to use an A test or an A^* test. Commonly one sequence, say the A^*'s, gives better power for given sample size, but for some reason such as wider validity we may prefer one of the "less efficient" A tests.

The general decision formulation for this problem would use three loss functions (i) cost of experimentation (ii) loss from wrong decisions (iii) disadvantages of using A^*. The usual kinds of decision problems for three loss functions could then be discussed. In practice (iii) is hard to assess and there is no natural comparability between (i) and (ii). So what is usually done is to consider (i) and (ii) only, and having required a bound on one of them, to decide whether the decrease in the other is enough to compensate for the disadvantages of using A^* instead of A. More specifically, the following two types of problems are of interest.

(a). *Fixed power requirement problems.* For a given power requirement, shall we use A_n or $A_{n^*}^*$? Here n and n^* are the smallest sample sizes for which the respective kinds of tests satisfy the given power requirement. Some function $K(n, n^*)$ such as $C(n) - C(n^*)$ or $1 - C(n^*)/C(n)$ is chosen as measuring our loss (extra experimentation cost) from using A_n instead of $A_{n^*}^*$. If $K(n, n^*)$ is small enough we will prefer to use A_n because of the advantages (iii) of A tests. If the given power requirement is a function of an unknown parameter θ, the loss $K(n, n^*)$ will also be a function of θ and so cannot be used directly for deciding between A_n and $A_{n^*}^*$. Some measure of loss not dependent on θ is needed. One natural choice is the worst possible loss $\sup_\theta K(n, n^*)$. (Weighted averages over θ and limits over particular sequences of θ's have also been used.) Asymptotic behavior of $K(n, n^*)$ and $\sup_\theta K(n, n^*)$ can be investigated for sequences of power requirements forcing $n \to \infty$ and $n^* \to \infty$. The particular choice $K(n, n^*) = 1 - n^*/n$ (with n^*/n being called the efficiency of A relative to A^*) and its asymptotic properties has been of wide interest [1], [2], [3], [4].

Received December 9, 1957; revised April 25, 1958.

[1] Work supported by the Office of Naval Research.

(b). *Fixed sample size problems.* For a given sample size n shall we use A_n or A_n^*? Let β_n be the power of A_n and β_n^* the power of A_n^*. Some function $L(\beta_n, \beta_n^*)$ such as $\beta_n^* - \beta_n$ or $1 - \beta_n/\beta_n^*$ is chosen as measuring our loss (extra wrong decisions) from using A_n instead of A_n^*. If $L(\beta_n, \beta_n^*)$ is small enough we will prefer to use A_n because of the advantages (iii) of A tests. If the powers β_n and β_n^* are functions of an unknown parameter θ, the loss $L(\beta_n, \beta_n^*)$ will also be a function of θ and so cannot be used directly for deciding between A_n and A_n^*. Some measure of loss not dependent on θ is needed. One natural choice is the worst possible loss $\sup_\theta L(\beta_n, \beta_n^*)$. This choice appears, with $L = \beta^* - \beta$, in the definition of stringency. Asymptotic behavior of $L(\beta_n, \beta_n^*)$ and $\sup_\theta L(\beta_n, \beta_n^*)$ as $n \to \infty$ can be investigated.

Though interest has been mostly in type (a) problems, it would seem that type (b) problems should be about equal in interest and applicability. The purpose of the present note is to discuss, as an illustration of type (b) problems, the following simple example.

3. Sign Test vs. Student Test. Let $X_1, X_2, \cdots$ be independent, each with Normal (θ, σ^2) distribution. We are to test at level α the one-sided hypothesis $\{\theta \leqq 0\}$ against the alternative $\{\theta > 0\}$. Let $\delta = \theta/\sigma$ and $p = p(\delta) = P(X_i > 0) = F(\delta)$ where F is the Normal (0, 1) cumulative. Then the number R_n of positive observations among $X_1, \cdots, X_n$ has a Binomial (n, p) distribution. And

$$T_n = n^{\frac{1}{2}}\bar{X}/[\textstyle\sum(X_i - \bar{X})^2/(n-1)]^{\frac{1}{2}}$$

has a Student t distribution with $n - 1$ degrees of freedom which is central when $\delta = 0$ and non-central with parameter $n^{\frac{1}{2}}\delta$ in general.

The sign test A_n of $\{\theta \leqq 0\}$ is

$$\begin{cases} \text{Reject when } R_n - n/2 > k_n \\ \text{Reject with prob. } \gamma_n \text{ when } R_n - n/2 = k_n, \end{cases}$$

where k_n, γ_n are constants determined by

$$P(R_n - n/2 > k_n \mid \delta = 0) + \gamma_n P(R_n - n/2 = k_n \mid \delta = 0) = \alpha.$$

The power function of this test is

$$\beta_n(\delta) = P(R_n - n/2 > k_n) + \gamma_n P(R_n - n/2 = k_n).$$

Values of $k_n, \gamma_n, \beta_n(\delta)$ can be obtained from tables such as [5] of the binomial distribution. For large values of n the normal approximation to binomial gives

$$\beta_n(\delta) \cong F\left(\frac{\sqrt{n}\,(2p-1) - c}{2\sqrt{p(1-p)}}\right) \quad \text{where} \quad F(c) = 1 - \alpha. \tag{1}$$

The Student test A_n^* of $\{\theta \leqq 0\}$ is

$$\text{Reject when } T_n > c_n$$

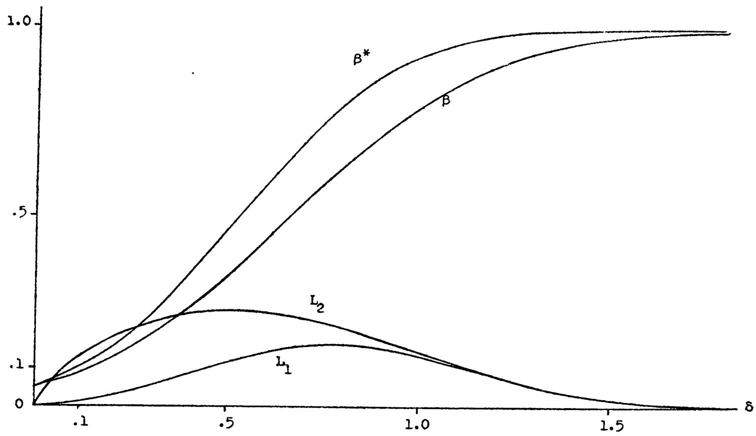

FIG. 1. Power functions β^* of Student test and β of sign test for $\alpha = .05$, $n = 11$; $L_1 = \beta^* - \beta$, $L_2 = 1 - \beta/\beta^*$.

where c_n is a constant determined by

$$P(T_n > c_n \mid \delta = 0) = \alpha.$$

The power function of this test is

$$\beta_n^*(\delta) = P(T_n > c_n) .$$

Values of c_n can be obtained from tables of the Student t distribution, and values of $\beta_n^*(\delta)$ from tables such as [6] of the non-central Student t distribution. For large values of n the normal approximation to non-central Student t gives

$$\beta_n^*(\delta) \cong F(\sqrt{n}\delta - c) \quad \text{where} \quad F(c) = 1 - \alpha. \tag{2}$$

Loss functions such as

$$L_1^n(\delta) = L_1(\beta_n, \beta_n^*) = \beta_n^*(\delta) - \beta_n(\delta)$$
$$L_2^n(\delta) = L_2(\beta_n, \beta_n^*) = 1 - \beta_n(\delta)/\beta_n^*(\delta)$$

can easily be plotted for particular values of n and α. This is done in Figure 1 for $n = 11$, $\alpha = .05$. As δ increases from 0 each function $L_i(\delta)$, $i = 1, 2$ increases from 0 to a maximum and then decreases toward 0.

For fixed α the change in appearance of these curves with increasing n differs only slightly from a simple horizontal compression. The curve $L_i^n(\delta)$ rises more quickly to its maximum and then falls more quickly toward 0, with increasing n. The position of the maximum tends to 0 at the rate $1/n^{\frac{1}{2}}$ but the maximum value changes very little and has a limit. Table 1 gives values of $\sup_\delta L_i^n(\delta)$ for $\alpha = .05$ and $n = 2, 3, \cdots, 13$. These values are computed from tables [5], [6] using interpolation and should be in error by not more than one or two units in the third decimal place. The cases $n = 2, 3, 4$ are special because for these the sign test does not reject with probability 1 even when $R_n = 0$ and so the

TABLE 1

Maxima of L_1 = power loss, L_2 = 1 − power ratio, of sign test relative to Student test, $\alpha = .05$

n	sup L_1	sup L_2
2	.800	.800
3	.600	.600
4	.200	.200
5	.130	.197
6	.189	.263
7	.150	.212
8	.153	.238
9	.180	.261
10	.142	.213
11	.167	.252
12	.171	.260
13	.151	.227
∞	.1686	.2610

power of the sign test does not $\rightarrow 1$ as $\sigma \rightarrow \infty$. For $n = 5, 6, \cdots \sup_\delta L_i^n(\delta)$ tends to be smaller if there is a non-randomized sign test with size close to.05 [$n = 5, 8, 10, 13$] and larger if there is no such sign test [$n = 6, 9, 12$]. Even for the smallest of these n the differences from the asymptotic values $\lim_{n\rightarrow\infty} \sup_\delta L_i^n(\delta)$ are not large.

Discussion of this example is concluded with the calculation of these asymptotic values. The following easily proved result is used:

LEMMA.

$$\lim_{n\rightarrow\infty} \sup_\delta L_n(\delta) = \sup_\Delta \lim_{n\rightarrow\infty} L_n(\delta_n)$$

if the former exists, where Δ is the set of all sequences $\{\delta_n\}$ for which $\lim_{n\rightarrow\infty} L_n(\delta_n)$ exists. [If lim be replaced throughout by lim inf or lim sup the same result holds, with existence provisos unnecessary.]

Writing $\delta_n = a_n/n^{\frac{1}{2}}$ it easily follows from (1) and (2) that if $a_n \rightarrow a$ then

$$\beta_n(\delta_n) \rightarrow F(a\sqrt{2/\pi} - c), \qquad \beta_n^*(\delta_n) \rightarrow F(a - c)$$

where $F(c) = 1 - \alpha$. This gives

$$\lim_{n\rightarrow\infty} L_1^n(\delta_n) = F(a - c) - F(a\sqrt{2/\pi} - c) \tag{3}$$

$$\lim_{n\rightarrow\infty} L_2^n(\delta_n) = 1 - F(a\sqrt{2/\pi} - c)/F(a - c). \tag{4}$$

Because of the lemma we can find $\lim_{n\rightarrow\infty} \sup_\delta L_i^n(\delta)$, $n = 1, 2$ by finding the value of a giving a maximum in (3), (4). Differentiating (3) with respect to a and equating the result to zero gives

TABLE 2

Asymptotic maximum power loss R_1 and proportionate power loss R_2 for sign test relative to Student test

α	a'	a''	R_1	R_2
.25	1.5514	1.1784	.0963	.1268
.10	1.6245	1.4086	.1405	.2056
.05	2.3570	1.5593	.1686	.2610
.01	3.0019	1.8574	.2229	.3765
.001	3.7676	2.2087	.2844	.5128
0	∞	∞	1	1

$$(3') \qquad \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(a-c)^2\} = \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(a\sqrt{2/\pi}-c)^2\}$$

which reduces to

$$(a-c)^2 = (a\sqrt{2/\pi}-c)^2 + \log(\pi/2).$$

The root of this quadratic at which the maximum of (3) occurs is

$$a' = \frac{c}{1+\sqrt{2/\pi}}\{1+\sqrt{1+(\log \pi/2)(1+\sqrt{2/\pi})/(1-\sqrt{2/\pi})c^2}\}.$$

The maximum value $R_1 = \lim_{n\to\infty} \sup_\delta L_1^n(\sigma)$ can now be found by substituting a' for a in (3). For example $\alpha = .05$ gives $c = 1.6449$, $a' = 2.3750$, and $R_1 = .1686$ for the asymptotic maximum loss. Differentiating (4) with respect to a and equating the result to zero gives

$$(4') \qquad F(a-c)\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(a\sqrt{2/\pi}-c)^2\} = F(a\sqrt{2/\pi}-c)\frac{1}{\sqrt{2\pi}}\exp\{-\tfrac{1}{2}(a-c)^2\}.$$

For given α the solution a'' of (4′) can be found numerically and shown to maximize (4). The maximum value $R_2 = \lim_{n\to\infty} \sup_\delta L_2^n(\delta)$ can now be found by substituting a'' for a in (4). For example $\alpha = .05$ gives $c = 1.6449$, $a'' = 1.5593$, and $R_2 = .2610$ for the asymptotic maximum loss.

Table 2 gives a', a'', R_1 (the asymptotic maximum power loss), R_2 (the asymptotic maximum amount by which the power ratio falls below 1) for several values of α. The most noticeable feature of this table is the strong dependence of R_1 and R_2 on the value of α. For small α use of sign test instead of Student test results in very severe loss of power at some alternatives. For example when $\alpha = .001$ there is an alternative where 51% of the power is lost by using sign test instead of Student test, and an alternative where the amount of power lost is .28.

REFERENCES

[1] E. J. G. PITMAN, *Lecture notes on nonparametric statistical inference*, Columbia University, 1949.
[2] H. CHERNOFF, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.
[3] W. J. DIXON, "Power functions of the sign test and power efficiency for normal alternatives," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 467–473.
[4] J. L. HODGES, JR. AND E. L. LEHMANN, "The efficiency of some nonparametric competitors of the *t*-test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 324–335.
[5] HARVARD COMPUTATION LABORATORY, *Tables of the cumulative binomial probability distribution*, Harvard University Press, 1955.
[6] J. NEYMAN AND B. TOKARSKA, "Errors of the second kind in testing Student's hypothesis," *Jour. Am. Stat. Assoc.*, Vol. 31 (1936), pp. 318–326.

A NOTE ON CONFIDENCE INTERVALS IN REGRESSION PROBLEMS

BY JOHN MANDEL

National Bureau of Standards

This note deals with the construction of confidence intervals for arbitrary real functions of multiple regression coefficients.

Consider the usual model

$$y_\alpha = \sum_i \beta_i x_{i\alpha} + \epsilon_\alpha \qquad \begin{array}{l} i = 1, \cdots, k \\ \alpha = 1, \cdots, N \end{array} \tag{1}$$

in which the ϵ_α are independently and normally distributed with mean zero, and common variance σ^2.

It is customary to construct confidence intervals for the β_i, using Student's t distribution. Alternatively, a joint confidence region can be constructed for the β_i using critical values of the F distribution. In both cases the usual statistic s^2, based on $N - k$ degrees of freedom, is used as an estimate of σ^2.

Durand [1] has discussed the use of the joint confidence region of the β_i, an ellipsoid in a k-dimensional space, for the construction of confidence intervals for linear functions, $Q = \sum_i h_i\beta_i$ of the regression coefficients. He points out that the chosen confidence coefficient (corresponding to the ellipsoid) is a lower bound for the joint confidence of any set of intervals thus derived.

Our first objective is to generalize this procedure by removing the restriction of linearity. Let

$$z = f(\beta_1, \beta_2, \cdots, \beta_k) \tag{2}$$

be any real function of the coefficients β_i. The form of the function is arbitrary but known.

For any arbitrarily selected value of z, say z_0, equation (2) represents a

Received February 15, 1957; revised March 1, 1958.