Efficient estimation of power functions

Kleijnen, J.P.C.; Denis, H.M.M.T.; Kerckhoffs, R.M.G.

Published in:
Communications in statistics: Part A: Theory and methods

Publication date:
1985

COMM. STATIST.-THEOR. METH., 14(8), 1865-1878 (1985)

EFFICIENT ESTIMATION OF POWER FUNCTIONS

J.P.C. Kleijnen, H.M.M.T. Denis, R.M.G. Kerckhoffs

Tilburg University (Katholieke Hogeschool Tilburg)
Department of Business and Economics
Postbus 90153 - 5000 LE Tilburg, The Netherlands

Key Words and Phrases: type II error; Monte Carlo; significance level.

ABSTRACT

A conceptually and computationally simple procedure is presented for the Monte Carlo estimation of a power function. The procedure asks "what would be the extreme parameter values in the null-hypothesis that would not be rejected, given the sample outcome?" Computationally the procedure requires sorting of these extremes values. The resulting estimate covers the whole range of the power function. The procedure is applied to the Student, the sign and the χ^2 statistics. The procedure gives good estimates provided the test statistics have symmetric distributions.

1. INTRODUCTION

The clue of our procedure is very simple (so that at first we felt that our procedure must have been published before; nevertheless a search of the literature gave no results). The procedure aims at efficient Monte Carlo estimation of power functions. The traditional procedure takes a sample x_j ($j = 1,\ldots,n$) from the

0361-0926/85/1408-1865$3.50/0

parent distribution with the (true) parameter θ; computes a test statistic t and observes whether t is significant. This process is repeated r times (using r different random number seeds), resulting in an unbiased estimator of the power at the value θ. Next a different value for θ is selected, keeping θ_0 of the null-hypothesis (H_0) fixed. And the procedure repeats, resulting in an unbiased estimator of the power at the new θ value. Our new procedure takes samples from the parent distribution without changing the true value of the parameter θ, and yet it estimates the power function over the whole range of θ. The clue is that for each sample outcome we ask: what would be the extreme value or values of θ_0 in H_0 that we would still accept? For example, when we use the Student statistic

$$t_{n-1} = \frac{\bar{x}-\theta}{s_{\bar{x}}} \tag{1.1}$$

to test $H_0 : \theta \leq \theta_0$ then the maximum value of θ_0, not leading to rejection of H_0, is $\bar{x} + t^{\alpha}_{n-1} s_{\bar{x}}$. In a two-sided test there are two extreme values, namely $\bar{x} - t^{\alpha/2}_{n-1} s_{\bar{x}}$ and $\bar{x} + t^{\alpha/2}_{n-1} s_{\bar{x}}$. In the following section we formalize our procedure; in section 3 we present computational algorithms; in section 4 we apply the procedure to the Student, the sign and the χ^2 tests; in section 5 we present our conclusions.

2. ANALYSIS OF THE NEW PROCEDURE

We specify a null-hypothesis H_0 (for instance, $H_0 : \theta \leq \theta_0$) and an alternative hypothesis H_1 (for instance, $H_1 : \theta > \theta_0$). We further select a test statistic t (for example, the Student statistic t_{n-1} of eq. 1.1). Then we define the binary (or indicator) random variable y_{θ_0}:

$$\begin{aligned} y_{\theta_0} &= 1 \text{ if } H_0 \text{ rejected for } \theta_0 \\ &= 0 \text{ if } H_0 \text{ not rejected for } \theta_0 \end{aligned} \tag{2.1}$$

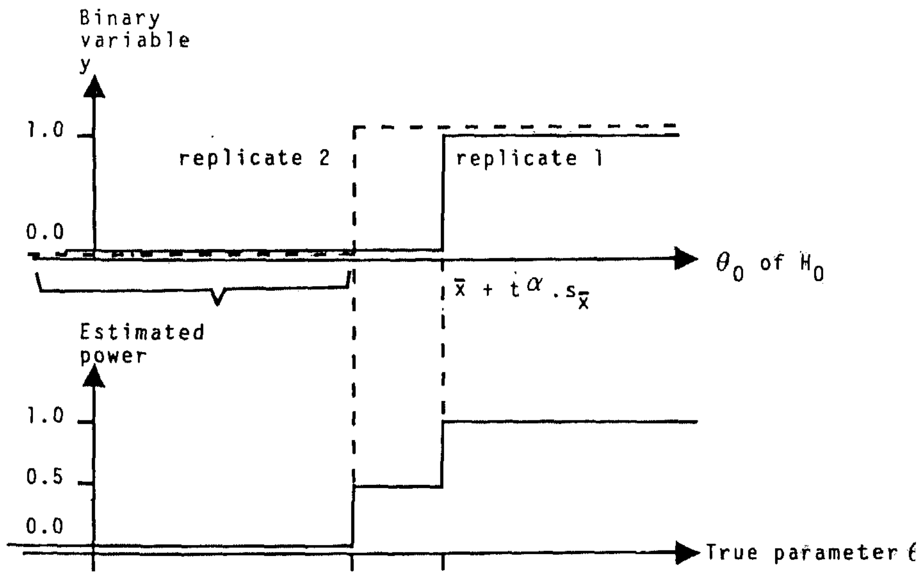

FIG. 1. Binary Variable y and Estimated Power in One-sided Test.

One realization ($y_{\theta_0 1}$) is shown by the solid lines in the upper part of FIG. 1 (we suppress the index n-1 of t^{α}_{n-1} and χ^2_{n-1} in all figures). A second replication ($y_{\theta_0 2}$) results in a different range of rejected θ_0 values; see the dashed lines in FIG. 1. The curly bracket corresponds to the interval of possible θ_0 values accepted in both replications. If we had only these two replicates, then our estimate of the power function (over the whole range of θ) would be shown by the lower part of FIG. 1. For a two-sided test FIG. 2 would result. In general, if we have r replicates then we estimate the power function through the average $\bar{y}_{\theta_0} = \sum_{i=1}^{r} y_{\theta_0 i}/r$. The resulting estimated power function shows r jumps in the one-sided test, and 2r jumps (up and down) in the two-sided test.

We shall first prove that the estimate $\bar{y}_{\theta}$ is unbiased in case we use the Student statistic of eq. (1.1) to test $H_0 : \theta = \theta_0$ ver-

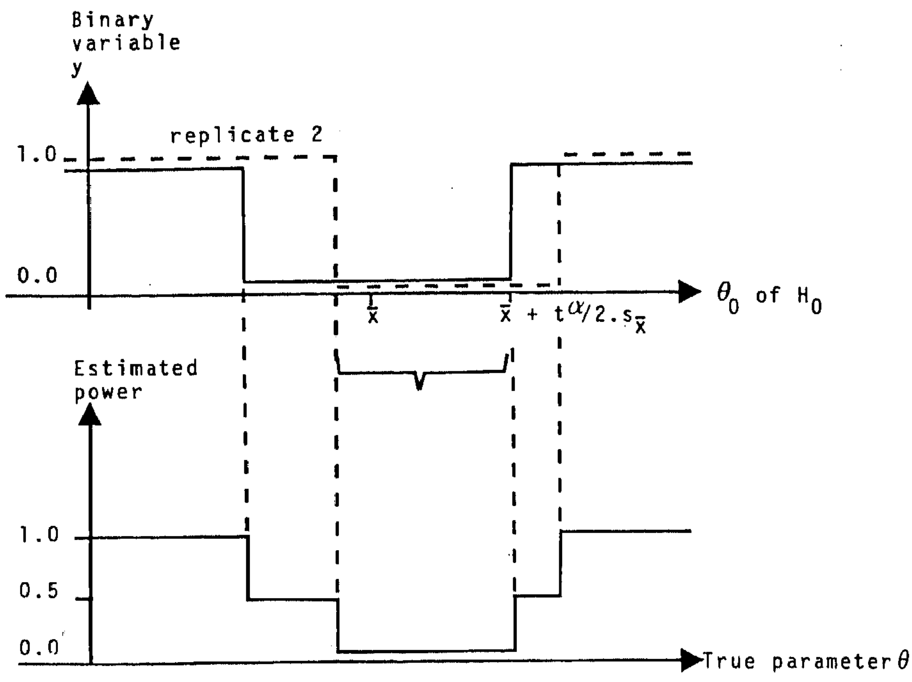

FIG. 2. Binary Variable y and Estimated Power in Two-sided Test.

sus $H_1 : \theta \neq \theta_0$ where θ is the mean of the normally and independently distributed x_j (or $x_j \sim NID(\theta,\sigma^2)$) where $j = 1,\ldots,n$. The classical procedure estimates the power by varying the "true" value θ. Our procedure, however, keeps θ fixed (at, say, $\theta = 0$) and varies θ_0. It is simple to prove that both procedures give identical results, because the Student statistic has a symmetric distribution [als see FIG. 3 where $t(\delta)$ denotes the noncentral t with non-centrality parameter δ; Johnson and Kotz (1970, p. 204) state that the distribution of $t(-\delta)$ is the mirror-image (reflected at $t = 0$) of that of $t(\delta)$]:

$$P(H_0 \text{ rejected} \mid \theta) = P(\frac{\bar{x}-\theta_0}{s_{\bar{x}}} > t_{n-1}^{\alpha/2} \mid \theta) =$$

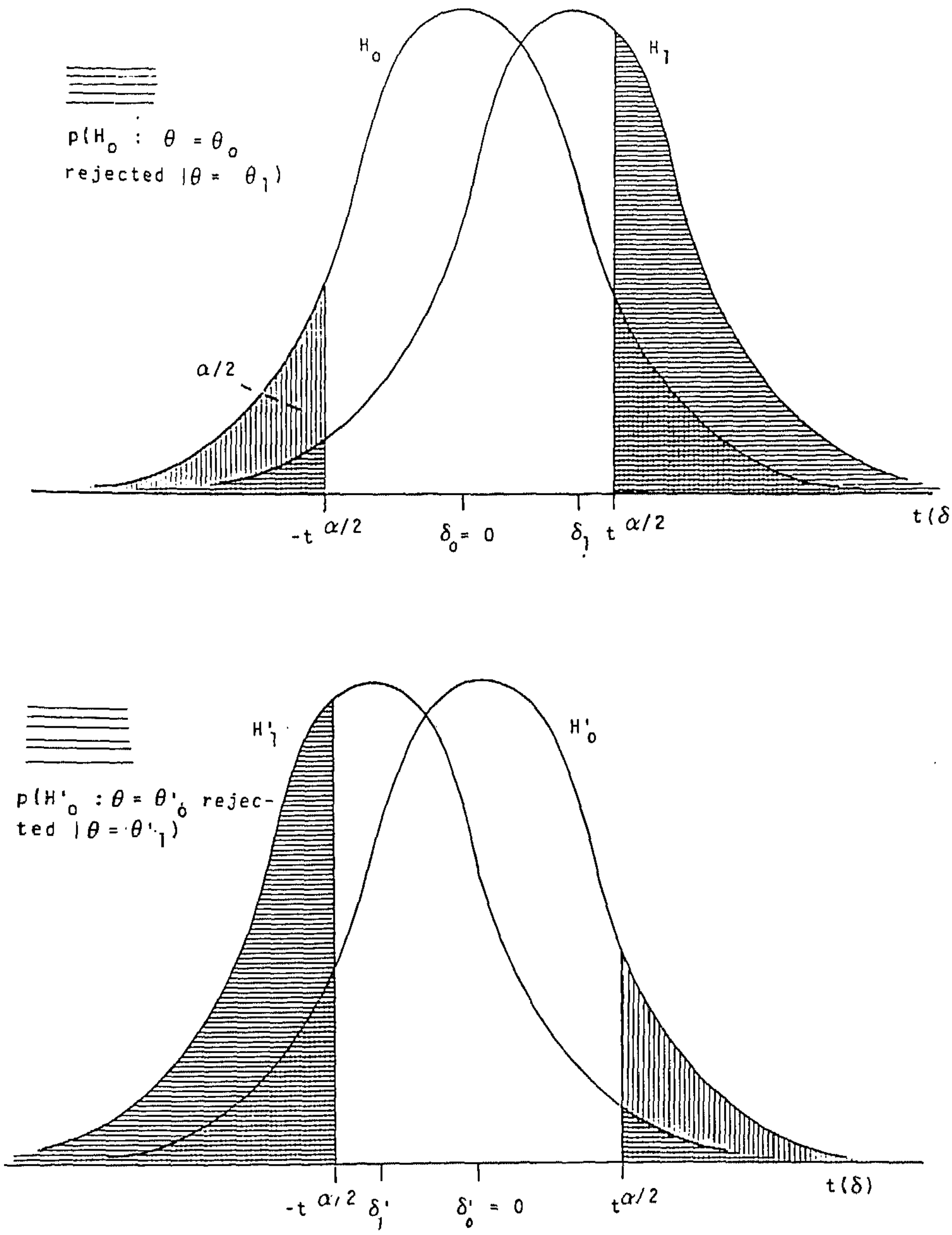

FIG. 3. Switching θ_0 and θ_1 in the Student test.

$$P(\frac{\bar{x}-\theta}{s_{\bar{x}}} > t_{n-1}^{\alpha/2} \mid \theta = \theta_0) = P(t_{n-1}(\delta) > t_{n-1}^{\alpha/2}) \qquad (2.2)$$

In the one-sided test $H_0 : \theta \leq \theta_0$ we have:

$$P(\frac{\bar{x}-\theta_0}{s_{\bar{x}}} > t_{n-1}^{\alpha} \mid \theta) = P(\frac{\bar{x}-\theta}{s_{\bar{x}}} < -t_{n-1}^{\alpha} \mid \theta_0) \qquad (2.3)$$

In general we conjecture that our procedure gives correct results, if we apply the Student statistic to test the location parameter of a symmetric parent distribution. We shall come back to this conjecture at the end of the present section.

Next we derive that our procedure does not work if the test statistic has an asymmetric distribution, for example, a χ^2 distribution. In FIG. 4 the two horizontally shaded areas of the upper part represent:

$$P(H_0 : \sigma^2 = 1 \text{ rejected} \mid \sigma^2 = 0.5) = 0.20 \qquad (2.4)$$

In the lower part two completely different areas represent:

$$P(H_0' : \sigma^2 = 0.5 \text{ rejected} \mid \sigma^2 = 1) = 0.45 \qquad (2.5)$$

More generally, our procedure overestimates the true power in the range $\sigma_1^2 < \sigma_0^2 = 1$.

Finally we consider the sign statistic to test the mean θ of a symmetric distribution, i.e., we define $y_j = 1$ if $x_j < \theta_0$ and $y_j = 0$ if $x_j \geq \theta_0$ ($j = 1,\ldots,n$) and $S = \Sigma\, y_j$. Under the null hypothesis the sign statistic has a symmetric distribution, namely a binomial distribution with parameters $p = 0.5$ and n. Under the alternative hypothesis we have $p \neq 0.5$. FIG. 5 (upper part) shows a situation where $\theta_0 < \theta_1$ so that $p < 0.5$ (the picture ignores the discrete character of S; our computations in the next section do account for this discreteness). Our procedure switches H_0 and H_1, i.e., the lower part of FIG. 5 shows that p becomes 1-p, and the

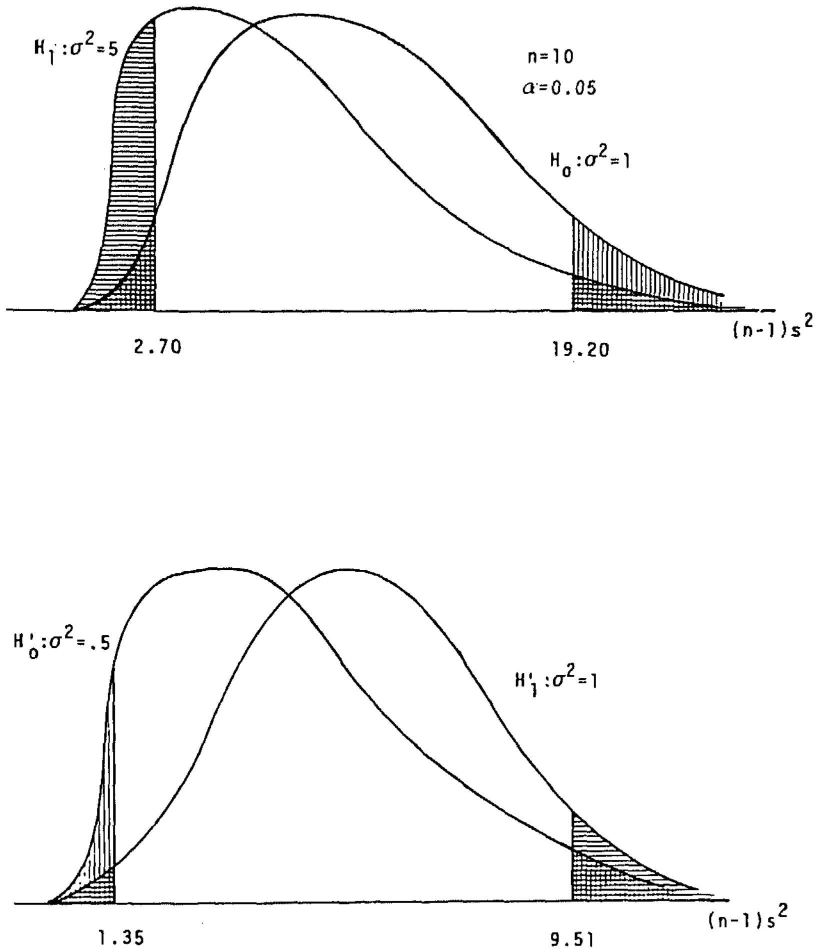

FIG. 4. Switching σ_0^2 and σ^2_1 in the χ^2 test.

distribution of S becomes the mirror image of the upper part of FIG. 5. Consequently the upper and lower parts give identical values for the power.

We emphasize that both the Student and the sign test have symmetric distributions so that their power functions depend on the absolute magnitude of $\theta - \theta_0$; these magnitudes are the same in

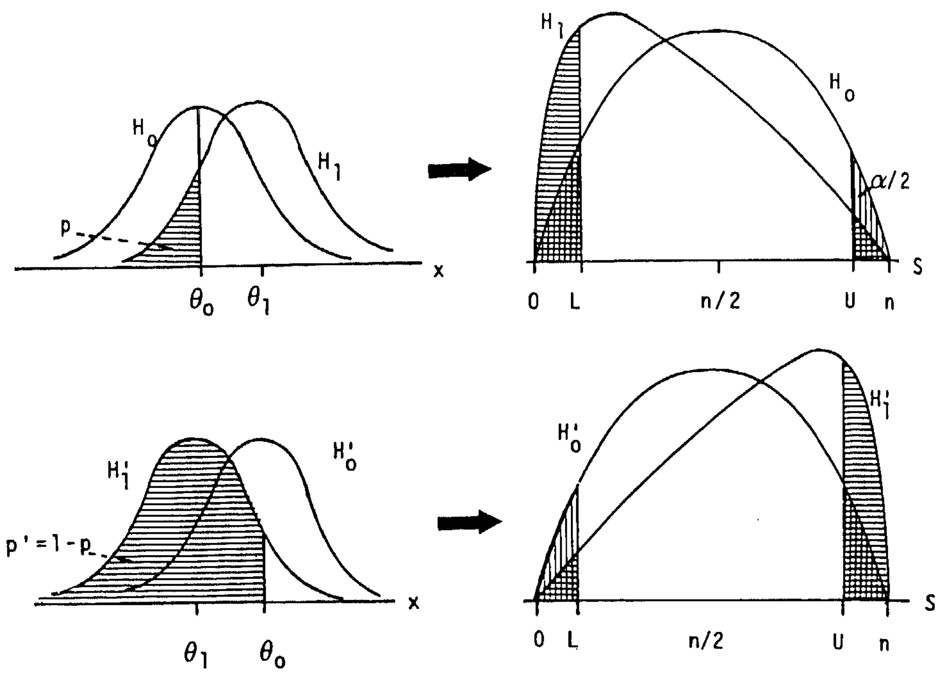

FIG. 5. Switching θ_0 and θ_1 in the sign test.

the upper and lower parts of FIG. 3 and FIG. 5. The χ^2 test has a non-symmetric distribution and its power function depends on θ_0/θ which equals 2 in the upper part of FIG. 4 and 0.5 in the lower part. In other words, suppose we have a test function $\Phi(x,\theta_0)$ depending on an observation x (which may be vector-valued) and depending on a known point θ_0 in the region Ω of the parameter space θ specified by the null hypothesis. Let

$$\beta(\theta) = P[\Phi(x,\theta_0) = 1|\theta) \qquad (2.6)$$

Then our procedure is applicable if

$$\beta(\theta) = P[\Phi(x,\theta_0) = 1|\theta)$$

$$= P[\Phi(x,\theta) = 1|\theta_0] \qquad (2.7)$$

3. COMPUTATION PROCEDURE

Computationally we obtain the estimated power function as follows.

Case (i): One-sided Test

The r replications yield r extreme values, for instance, the upper limits $u_i = \bar{x}_i + t^{\alpha}_{n-1}s_{\bar{x}_i}$ (see *FIG.* 1). Sorting these r extremes, results in the order statistics $u_{(i)}$ with $i = 1,\ldots,r$. Then the estimated power function has jumps of size $1/r$ at the values $u_{(i)}$.

Our procedure implies that our estimated power function is identical to an estimated distribution function $F(u)$. Consequently we can use standard tests when comparing $F(u)$ to a known power function (see the χ^2 goodness-of-fit test in the next section) or when comparing different estimated power functions corresponding to different test statistics (to test if the variable u_I is stochastically smaller than another variable u_{II}, or $F(u_I) > F(u_{II})$, we can use the sign test).

Case(ii): Two-sided Test

With a two-sided test it may happen that the upper limit in replicate 1 is smaller than the lower limit in replicate 2 ($u_1 < \ell_2$) so that the estimated power function jumps up and down (FIG. 2 did not illustrate this possibility). We now sort ℓ_i and u_i separately. The estimated function decreases with $1/r$ when $\theta = \ell_{(i)}$ and increases with $1/r$ when $\theta = u_{(i)}$; see FIG. 6. We can again treat the estimated power function as a distribution function: the right-hand side satisfies the conditions $0 \leq \hat{F}(u) \leq 1$ and $F(u) \leq F(u+\Delta)$ provided we take Δ large enough so that the jumps up and down within the interval Δ result in a net increase; in practice this condition is not restrictive. The left-hand side of $\hat{F}(u)$ can be reflected: $\hat{F}(u)$ becomes $\hat{F}(-u)$ where $u < \theta_0$. (If we know that the power function is symmetric then we might estimate the right-hand side also using the left-hand estimate; we did not apply this idea.)

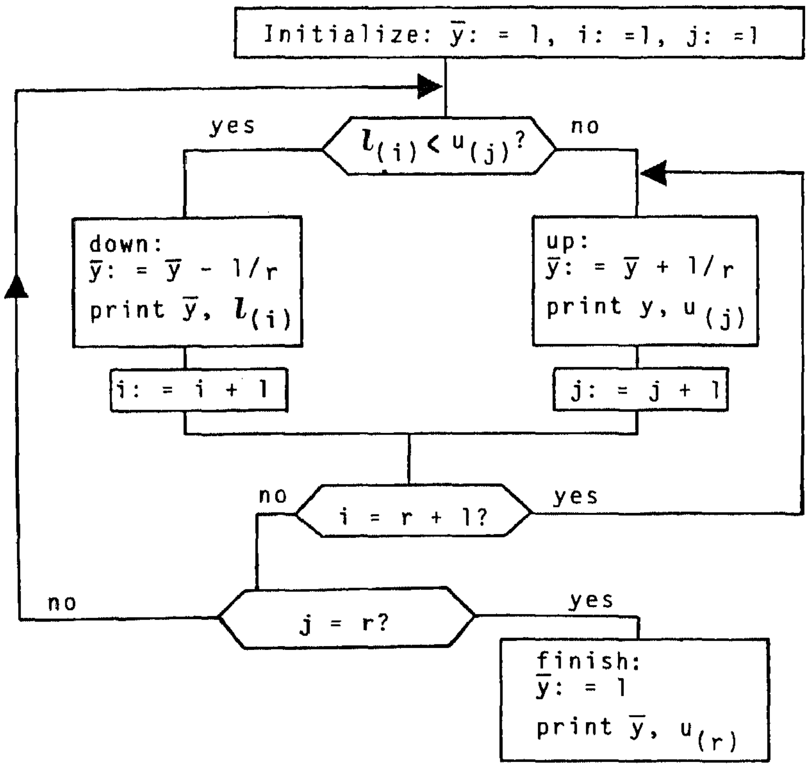

FIG. 6. Estimation of y_{θ_0}.

4. APPLICATIONS

Next we apply the new procedure to a number of situations. In all situations the parent distribution is normal. The multiplicative random number generator has multiplier 13^{13} and modulus 2^{59} (the computer is an ICL 2960). The number of replications (r) is 1000, i.e., the estimated power function has steps of height 1/1000. One-sided tests use $\alpha = 0.025$; two-sided test use $\alpha = 0.05$.

Student Statistic for Testing θ in $x \sim NID(\theta,\sigma^2)$

We use n = 10 observations from N(0,1) to compute the Student statistic. The true power function was evaluated by Owen (1965) at eleven values of θ so that we distinguish twelve classes (for other θ values see Hill, 1978). We compare the theoretical number of observations per class to the empirical number, using the χ^2 goodness-of-fit test (with eleven degrees of freedom and a 2.5% significance level). We repeat this process ten times. None of the ten χ^2_{11} values is significant. For the two-sided test we test the left-hand side and the right-hand side of the power function separately. Again the empirical power function does not deviate significantly from the true function, in all ten repetitions.

The χ^2 Statistic for Testing σ^2 when $x \sim NID(\mu,\sigma^2)$

$H_0 : \sigma^2 < \sigma_0^2$ means that the extreme value u is $\sum_{j=1}^{n} (x_j-\bar{x})^2/(\chi_v^{2\alpha})$ where $P(\chi_v^2 > \chi_v^{2\alpha}) = \alpha$ and $v = n-1$. First we use only n = 10 observations from N(0,1). We compute the true value of the power function (our computer has a subroutine for evaluating $P(\chi_v^2 > a)$ for given v and a). We note that the power function does not reach its minimum at $\sigma^2 = \sigma_0^2 = 1$; it does reach the value α at $\sigma^2 = \sigma_0^2 = 1$. All ten repetitions result in significant deviations between the true and the estimated power functions. Next we use n = 30 observations to estimate σ^2. All ten repetitions still yield significant deviations. However, these deviations are smaller because when n increases, then the χ_n^2 distribution becomes more symmetric. Finally we repeat the computations for a two-sided test, and we find completely analogous results.

The Sign Test for Testing θ in $x \sim NID(\theta,\sigma^2)$

We obtain a $1-\alpha$ confidence interval for θ by sorting the observations x; the L-th and (U+1)th ordered observation form the extreme points of the confidence interval; L and U are the $\alpha/2$ critical points of the binomial distribution with p = 0.5; see FIG. 5 and Conover (1971). We use n = 20 observations from N(0,1) so that

$L = 4$ and $U = 17$. The power function was computed by Dixon (1953). No χ^2 goodness-of-fit value is significant (as we expect; see section 2).

We also measure the computer time, but only for the following situation (computer time heavily depends on the programmer's skill): a one-sided test using the Student statistic with $n = 10$ observations and only $r = 100$ replications. Our procedure takes 8,600 milli-seconds whereas independent seeds take 203,000 msec.; our procedure estimates the power function over the whole range of θ whereas the traditional procedure estimates the function at $r = 100$ values of θ.

5. CONCLUSIONS

Our new procedure estimates the power function not at only a few prespecified values of θ (the parameter to be tested). Instead each sampled value of the test statistic creates a step in the estimated power function. So our procedure is both efficient (less computer time) and effective (the whole range is covered; the estimate can be tested because it behaves as an estimated distribution function). Yet the conceptual and computational details are simple.

Unfortunately the procedure is not always applicable. This restriction is shown by the experiment with the χ^2 statistic for testing the variance of a Gaussian variable. Good estimates result for the Student statistic and the sign statistic for testing the mean of a symmetric distribution. We explained this behavior, observing that a non-symmetrically distributed test statistic does not permit exchanging θ_0 and θ in the power function. In other words, if the power function depends on $\theta_0-\theta$ only, then our procedure works.

Future research may further clarify the performance of our new efficient procedure for estimating the power function. Such research might also reveal relationships with other ideas in the literature, such as observed significance levels and common random numbers; see Dempster and Schatzoff (1965), Nozari (1984), Kleijnen (1984).

ACKNOWLEDGEMENT

We thank H. Moors (Tilburg University) for his many helpful comments on earlier versions. We thank the referee for his comments that helped to improve our presentation.

BIBLIOGRAPHY

Conover, W.J., (1971). Practical nonparametric statistics. John Wiley & Sons, Inc., New York.

Dempster, A.P. & Schatzoff, M., (1965). Computed significance level as a sensitivity index for test statistics. Journal American Statistical Association, 60, no. 310, 420-436.

Dixon, W.J., (1953). Power functions of the sign test and power efficiency for normal alternatives. Annals Mathematical Statistics, 24, 467-473.

Hill, I.D., (1978). A remarkt on algorithm AS76: an integral useful in calculating non-central t and bivariate normal probabilities. Applied Statistics, 27, 379.

Johnson, N.L. & Kotz, S., (1970). Continuous univariate distributions-2. Houghton Mifflin Company, Boston.

Kleijnen, J.P.C., (1984). Superefficient estimation of power functions in simulation experiments. Tilburg University, Tilburg (Netherlands), Version 3.

Nozari, A., (1984). Generalized and ordinary least squares with estimated and unequal variances. Communications in Statistics, Simulation and Computation, 13, no. 4, 521-537.

Owen, D.B., (1965). Power of Student's t-test. Journal American Statistical Association, 60, 1965, 320-333.

Received by Editorial Board member November, 1984; Revised February, 1985.

Recommended by A. M. Kshirsagar, University of Michigan, Ann Arbor, MI

Refereed by S. R. Adke, University of Michigan, Ann Arbor, MI