Superefficient estimation of power functions in simulation experiments
Kleijnen, J.P.C.

Publication date:
1984

Link to publication

Citation for published version (APA):
RESEARCH MEMORANDUM

SUPEREFFICIENT ESTIMATION OF POWER FUNCTIONS
IN SIMULATION EXPERIMENTS

Dr. Jack P.C. Kleijnen
Professor of Simulation and Information Systems
Department of Business and Economics
Tilburg University (Katholieke Hogeschool Tilburg)
5000 LE Tilburg, Netherlands

November 1983 (Version 1)
February 1984 (Version 2)
ABSTRACT

The power function equals the probability of rejecting the null-hypothesis as a function of the true parameter. This function can be estimated through Monte Carlo simulation, possibly exploiting common random numbers. A superefficient procedure computes the extreme parameter value that would still be accepted under the null-hypothesis, given the sample outcome. This simple idea transforms the estimation of the power function into the estimation of a distribution function. Comparing power functions becomes straightforward. The power function is also contrasted to Schruben's coverage function.

KEY WORDS: power function, Monte Carlo, observed significance level, coverage, distribution function.

1. INTRODUCTION

The clue of our proposal is very simple, as we shall explain in Section 4. Nevertheless to the best of our knowledge there are no publications exploiting our idea. We do know of some publications containing related ideas; see Dempster and Schatzoff (1965), Schruben (1980).

Monte Carlo or simulation experiments are often used in order to evaluate statistical procedures (such as the Student test applied to nonnormal distributions). Estimation of the type I (or $\alpha$) errors of statistical procedures is only a first step. A second step is the estimation of the type II (or $\beta$) errors as a function of the true parameter (the true parameter is known in simulation experiments). The latter function is known as the power function. We discuss computational and statistical procedures for the evaluation and comparison of estimated power functions. In many cases estimation of the power function turns out to be equivalent to estimation of a distribution function.

Schruben (1980, p. 19) proposed the coverage function $P(\eta^* \leq \eta)$ where $\eta^*$ is the smallest confidence coefficient such that a (known) parameter, $\theta$, is within the (confidence) region $R(\eta^*, X)$; "here X denotes the sample data. We propose the estimated (or empirical) power function, when comparing different estimators. We have the following objections against Schruben's
suggestion to choose the estimator with the "greatest coverage frequency": the ideal estimator should show a uniform coverage function, that is, \( P(\eta^* \leq \eta) = n \). Therefore we prefer the estimator that has a coverage function with minimal deviation from the uniform distribution. Among those estimators having an estimated coverage function not significantly different from the uniform distribution, we prefer the estimator with the maximal power; also see Nozari (1983). In other words, if two (or more) estimators do result in the desired coverage (or type I error) then we prefer the estimator with the shortest confidence interval length; a short confidence interval is equivalent to a small type II error in a testing approach. Schruben (1980, p. 26) himself also concluded: "... the coverage function ... does not provide a basis for comparing sizes of different estimator (confidence) regions".

2. THE POWER FUNCTION

The classical derivation of the power function concept is as follows. Suppose we wish to test a parameter \( \theta \). We specify the null-hypothesis \( H_0: \theta = \theta_0 \), say \( \theta_0 = 1 \). And we also specify the alternative hypothesis, for instance, \( H_1: \theta \neq \theta_0 \) (we shall also discuss one-sided alternative hypotheses). We take a sample \( X \) from which we compute a statistic, say \( t \), to test the hypothesis \( H_0 \) (also see Section 1 where we defined \( X \)). For example, when \( \theta \) denotes the mean we take \( n \) observations \( X = (x_1, \ldots, x_n) \) and we might use the classical Student statistic \( t_{n-1} \):

\[
t_{n-1} = \frac{x - \theta}{s_x}
\]

And in regression analysis we test a hypothesis concerning a single regression parameter using \( t_v \) (where \( v = n - q \) while \( q \) denotes the number of regression parameters and \( n \) denotes the number of observations). The power (or the complement of the type II error) is: \( P(H_0 \text{ rejected } \theta = \theta_t) \) where \( \theta_t \) denotes the true value of the parameter \( \theta \) (and \( \theta_0 \) might equal \( \theta_t \) in which case the power equals the type I error). For instance, eq. (2.1) yields:

\[
P\left( t_{n-1} > \frac{\alpha/2}{\theta_t - \theta_0} \bigg| \theta = \theta_t \right) =
\]

\[
P(t_{n-1} > \frac{\alpha/2}{n-1} - (\theta_t - \theta_0)/s_x) + P(t_{n-1} < -\frac{\alpha/2}{n-1} - (\theta_t - \theta_0)/s_x) =
\]

\[
f(\theta_0, \theta_t, \alpha, n)
\]
FIGURE 1. The Power of the t Statistic when Testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

FIGURE 2. The Power of the t Statistic when Testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.
Figure 1 illustrates that the power reaches its minimum value (equal to the type I error rate $\alpha$) if the value $\theta_0$ in $H_0$ happens to equal the true value $\theta_t$; and a test with a larger sample size $n$ has higher power except for $\theta_t = \theta_0$ and for $\theta_t = -\infty$ and $\theta_t = +\infty$ respectively where for all sample sizes the power has the value $\alpha$ and one respectively. In the example of a two-sided test with the (symmetric) t statistic, the power function is symmetric around $\theta_0$, i.e., the power depends only on the absolute value of $\theta_0 - \theta_t$.

Figure 2 illustrates the power function for a one-sided test $H_0: \theta \leq \theta_0$ versus $H_1: \theta > \theta_0$. Also see Appendix 1 for more details.

3. MONTE CARLO ESTIMATION OF THE POWER FUNCTION

In (Monte Carlo) simulation we can estimate the power at the true value $\theta_t$ of the parameter $\theta$, as follows. We sample $X$ a number of times from its distribution with parameter $\theta = \theta_t$. We sample $X$ $r$ times, using $r$ different random number seeds. We compute the $t$ statistic $r$ times, each time observing whether $t$ is significant. For example, when $t$ is the Student statistic $t_{n-1}$ of eq. (2.1) the estimated power is $\hat{p} = \frac{1}{r} \sum_{i=1}^{r} y_i / r$ with

$$y_i = 1 \text{ if } |\bar{x}_i - \theta_0| / s_{x_i} > t_{n-1}^{\alpha/2} \quad (i = 1, \ldots, r)$$

$$= 0 \text{ if } |\bar{x}_i - \theta_0| / s_{x_i} \leq t_{n-1}^{\alpha/2} \quad (3.1)$$

Then $\hat{p}$ is an unbiased estimator of the power at the value $\theta = \theta_t$ ($\theta_t$ remains constant over the $r$ replications and $\theta_t$ is known to the simulation experiment er).

In order to estimate the power function at other values of $\theta$, we might proceed as follows (it is our experience that many simulation practitioners do proceed in the following way). We repeat the experiment with a new value for $\theta_t$ (keeping $\theta_0$ fixed). If we use a new random number seed then we obtain an independent estimator of the power at that new value of $\theta_t$.

However, we may also use the same seed so that we obtain a more accurate estimator of the change in power as $\theta$ changes; we obtain a less accurate estimator of the absolute value of the power at different $\theta$ values; see Dempster and Schatzoff (1965, p. 433), Kleijnen (1985), Schruben (1979).
If we use the same seeds at, say, \( w \) different \( \theta \) values (\( w > 2 \)), then we can simplify the computation of the realizations of the test statistic \( t \). For instance, if we change the true mean from \( \theta_1 \) to \( \theta_2 \) in eq. (2.1) then \( s_x^{-} \) remains unchanged and \( \bar{x} \) increases with \( \theta_2 - \theta_1 \). In general, when we sample \( X \) we can perform \( w \) (quasi) parallel computations corresponding to \( w \) different values for the (true) value of \( \theta \) which take far less time than \( w \) times the computation time needed for a single (true) value of \( \theta \). Another example is provided by simulation experiments with regression analysis: changing \( \beta \) from \( \beta_1 \) to \( \beta_2 \) in \( y = X \beta + e \) changes the realizations of \( \hat{\beta} \) from \( \hat{\beta}_1 \) to \( \hat{\beta}_2 = \hat{\beta}_1 + (\hat{\beta}_2 - \hat{\beta}_1) \), i.e., all we have to do is add a constant to the old \( \beta \) realizations (no sampling, no solving of normal equations). If we have a one-sided test then the computation of the new values of the test statistic \( t \) itself is even simpler. For instance, when testing the mean we add the constant
\[
(\theta_2 - \theta_1)/s_x^{-} \text{ when testing a regression parameter we add } (\beta_2 - \beta_1)/s_\beta^{-} \text{ where } \text{var}(\hat{\beta}_2) = \text{var}(\hat{\beta}_1) = s_\beta^{-} \text{ and when testing the variance we multiply the } \chi^2 \text{ statistic by the constant } \theta_2/\theta_1 = \sigma_2^2/\sigma_1^2. \]

Our proposal is related to the approaches followed in Dempster and Schatzoff (1969) and Nozari (1984).

4. A SUPEREFFICIENT ESTIMATION PROCEDURE

Now we propose a procedure that is even more efficient. The clue of our proposal is that we ask: what would be the most extreme value (or values) of \( \theta_0 \) in \( H_0 \) that we would still accept, given the sample outcome \( X \)? For example, using eq. (2.1) in a one-sided test (see Figure 2) the maximal value of \( \theta_0 \) not leading to rejection of \( H_0 \) would be: \( \theta_0 = \bar{x} + n a / s_x^{-} \). When using eq. (2.1) in a two-sided test, the two extreme values would be: \( x - n a / s_x^{-} \) and \( x + n a / s_x^{-} \). When testing \( H_0 : \sigma^2 < \sigma_0^2 \) using the \( \chi^2 \) statistic, we reject all values of \( \sigma^2 \) exceeding \( \sum (x_i - \bar{x})^2 / (\chi^2) \) where \( \text{P}(\chi^2 \geq \chi^2) = \alpha \) and \( v = n-1 \).

And so on. Our proposal to look at extreme values is related to the "coverage" function in Schruben (1980) and to the "observed significance level" in Dempster and Schatzoff (1965).

We formalize our proposal as follows. We define the binary random variable \( y_\theta \):

\[
y_\theta = 1 \text{ if } H \text{ rejected for } \theta = \theta_0 \\
= 0 " \text{ not } " \\
(4.1)
\]
See Figure 3, upper part, solid lines. A second replication results in a different range of rejected \( \theta_0 \) values; see the dashed lines in Figure 3. The curly bracket shows the interval of possible \( \theta_0 \) values accepted in both replications. If we took only these two replicates then our estimate of the power function would be shown by the lower part of Figure 3. Figure 4 is the analogue of Figure 3 in case we have a two-sided test. In general, if we have \( r \) replications then we estimate the power function through the average

\[
\bar{y} = \frac{1}{r} \sum_{i=1}^{r} y_{\theta_i}.
\]

Obviously the estimator \( \bar{y} \) is unbiased. So we estimate the power for the whole range of \( \theta \); our estimated power function shows \( r \) jumps in the one-sided test, and \( 2r \) jumps in the two-sided test.

Computationally we obtain the estimated power function by sorting the extreme values, found in the \( r \) replications, in increasing order. How we proceed depends, because we distinguish three situations:

Case (i): One-sided test
We sort the \( r \) extremes. For instance, in Figure 3 we rank the upper limits \( u_i = x_i + t^{a/2} x_i^{-} \). This ranking results in the order statistics \( u_{(i)} \) where \( i = 1, \ldots, r \). Then the estimated power function shows jumps of size \( 1/r \) at the values \( u_{(i)} \).

Case (ii): Two-sided test with known symmetry
We assume that the (true) power function is symmetric around the value \( \theta = \theta_0 \); see Figure 4. Then we might merge the "low" and "high" extremes and estimate the power function from the resulting \( 2r \) values. For example, we define the lower limits \( l_i = x_i - t^{a/2} x_i^{-} \) and the upper limits \( u_i = x_i + t^{a/2} x_i^{-} \); if \( l_i < \theta_t \) (where \( \theta_t \) denotes the true value of \( \theta \) known to the simulation experimenter) then we replace \( l_i \) by \( l'_i = \theta_t + (\theta_t - l_i) = 2\theta_t - l_i \); if \( u_i < \theta_t \) then we replace \( u_i \) by \( u'_i = 2\theta_t - u_i \); we merge \( l, l', u \) and \( u' \) resulting in, say, \( v_j \) where \( j = 1, \ldots, 2r \); next we sort \( v_j \); the resulting estimate of the power function makes jumps of size \( 1/(2r) \) at the values \( v_{(j)} \). To estimate the power function at values \( \theta < \theta_t \), say \( \theta = \theta_t - c \), we use the estimate at \( \theta_t + c \) (where \( c > 0 \)).

Our solution for case (ii) has one drawback: the resulting estimate for the power at the value \( \theta_t = \theta_0 \) equals zero (because the smallest \( v \) value exceeds \( \theta_t = \theta_0 \) with probability one). We may remedy this drawback by estimat-

Initialize: \( y_\ell = 1, i = 1, j = 1 \)

down:
- \( y_\ell = y - 1/r \)
- print \( y_\ell, l(i) \)
- \( i = i + 1 \)

no:
- \( i = r + 1? \)

yes:
- \( j = r? \)

no:
- \( j = r + 1? \)

yes:
- finish:
  - \( y_\ell = 1 \)
  - print \( y_\ell, u(r) \)

no:
- \( l(i) < u(j)? \)

no:
- \( j = j + 1 \)

yes:
- \( y_\ell = y + 1/r \)
- print \( y_\ell, u(j) \)

FIGURE 5. Estimation of \( y_\ell \).
ing the power at $\theta_0$ using the standard method of eq. (3.1); alternatively we may proceed as in case (iii).

Case (iii): Two-sided test without known symmetry

Now we are confronted with a more complicated computational problem. For example, it may happen that $u_1 < \ell_2$ so that the estimated power function jumps up and down. (We observe in passing that the true power function may indeed be non-monotonic for $\theta < \theta_0$ depending on the statistic, see Johnson (1978) who adapted the Student statistic for asymmetry of the $x$ distribution.) To compute $\bar{y}_0$ we sort the lower limits $\ell_i$ and the upper limits $u_i$ separately. This sorting results in the two vectors $\{\ell_i\}$ and $\{u_i\}$. The estimate $\bar{y}$ decreases with $1/r$ when $\theta = \ell_i$ and $\bar{y}$ increases with $1/r$ when $\theta = u_i$; see Figure 5 and the following Algol-68 procedure:

```algol
procedures initialize, down, up, finish
initialize
while $j < r$
    do
        if $\ell_i < u_j$ then down, $i := i + 1$
        else up, $j := j + 1$
    od
for $k$ from $i$ to $r$ to up
finish
```

We emphasize that in the cases (i) and (ii) we have transformed the problem of estimating a power function into the problem of estimating a distribution function. In case (i) the power function meets the mathematical requirements of a distribution function: $0 \leq F(u) \leq 1$ for $-\infty < u < \infty$, and $F(u) \leq F(u+c)$. In case (ii) we estimate the power function only for $v \geq \theta_0$ and then the requirements are again satisfied. Also see Appendix 2.

5. ILLUSTRATION

We verify our superefficient procedure for power estimation, applying the procedure to a situation with known solution. We test $H_0 : \theta = 0$ in case $x \sim N(\theta,1)$ using the Student statistic. We perform a one-sided and a two-sided test with $\alpha = 0.025$ and $\alpha = 0.05$ respectively, and sample size $n = 10$. We verify the estimated power function applying the $\chi^2$ goodness-of-fit test;
for \( \chi^2 \) we use twelve classes with class limits taken from Owen (1965). Case (iii) shows a complication: the estimated power function is not equivalent to a distribution function. We solve this problem by splitting the \( \theta \) interval into two parts (namely \( \theta < \theta_0 = 0 \) and \( \theta > 0 \)) and applying the \( \chi^2 \) measure per part. We always estimate the power function from 1000 observations \( (r = 1000) \); and this we repeat ten times so that we obtain ten \( \chi^2 \) values. Altogether we sample from \( N(0,1) \times 10 \times 1000 \times 10 \) times.

**Case (i):** For the one-sided test none of the ten \( \chi^2 \) values is significant at the 5\% level.

**Case (ii):** For the two-sided test our estimator of the power using symmetry results in two significant \( \chi^2 \) values (out of ten values). We can explain one significant value by the large deviation between expected and realized values of the power function close to \( \theta = \theta_0 = 0 \); see Section 4, Case (ii).

**Case (iii):** When not using symmetry our estimator results in ten \( \chi^2 \) values for the right-hand side of the power function \( (\theta > \theta_0) \) and ten \( \chi^2 \) values for the left-hand side \( (\theta < \theta_0) \), and none is significant.

We prefer the procedure of case (iii) over that of case (ii): the latter procedure gives a bad estimate of the power function close to \( \theta = \theta_0 \). The former procedure may result in an estimated power function that jumps up and down locally; however, over the whole \( \theta \) area the function behaves correctly as demonstrated by the \( \chi^2 \) criterion.

We also measure the computer time for the following situation: Case (i) (one-sided test), \( n = 10, r = 100 \). The "superefficient" procedure requires a total time of 8,600 milliseconds; common random numbers requires 11,000 msec.; independent seeds take 203,000 msec. The last two procedures estimate the power function at 100 values of \( \theta \), whereas our procedure estimates the function over the whole range of \( \theta \).

**APPENDIX 1: PROPERTIES OF THE STUDENT POWER FUNCTION**

We define the non-central t statistic with parameter \( \delta \):

\[
\bar{x} - (\theta_0 + \delta) \\
\frac{n-1}{\sum_{x}}
\]

(A.1)

where \( \delta = \bar{\theta} - \theta_0 \). Then we can rewrite eq. (2.2) as:

\[
P\left( \left| \frac{\bar{x} - \theta_0}{\sum_{x}} \right| > \frac{\alpha/2}{\sqrt{n-1}} \mid \theta = \theta_0 \right) = P\left( \left| t_{n-1}(\delta) \right| > \frac{\alpha/2}{n-1} \right)
\]

(A.2)
The probability in eq. (A.2) is tabulated for different values of \( \alpha, n, \) and \( \delta' = (\theta_0 - \theta_t)/(\sigma/\sqrt{n}) \), see Owen (1965) — where we observe that in a simulation experiment we do know \( \theta_t \) and \( \sigma \) and hence we know \( \delta' \). Actually Owen (1965) gives values of \( \theta_t \) (or \( \delta' \)) such that the probability in eq. (A.2) equals the following twelve values: 0.01, 0.05, 0.10 (0.10) 0.90, 0.95. We refer to the algorithm in Hill (1978) for the computation of eq. (A.2) for other values of \( \theta_t \).

Using the symmetry of the \( t_{n-1} \) distribution, it is simple to prove that we can interchange \( \theta_0 \) and \( \theta_t \) in the two-sided test, or

\[
P(H_0 : \theta = \theta_0 \text{ rejected } | \theta = \theta_t) = \]

\[
P(H_0 : \theta = \theta_t \text{ rejected } | \theta = \theta_0)
\]

and that in the one-sided test we have

\[
P(H_0 : \theta < \theta_0 \text{ rejected } | \theta = \theta_t) =
\]

\[
P(-t_{n-1}^\alpha | \theta = \theta_t) = \]

\[
P(-t_{n-1}^\alpha < -t_{n-1}^\alpha | \theta = \theta_0).
\]

APPENDIX 2: EVALUATING ESTIMATED POWER FUNCTIONS: TEST PROCEDURES

In general we perform simulation experiments in order to compare two or more statistical procedures. For example, we compare the Student statistic to the Wilcoxon rank statistic, or ordinary to generalized least squares. How can we choose between two procedures? We propose to select procedure I over procedure II if procedure I has a dominant power function, or \( P(H_0 \text{ rejected } | \theta, \text{ procedure I}) \geq P(H_0 \text{ rejected } | \theta, \text{ procedure II}) \) for all values of \( \theta \), provided the type I error is acceptable, or \( P(H_0 \text{ rejected } | \theta = \theta_0, \text{ procedure I}) \leq \alpha \) (if such a dominant power function does not exist, then the selection is problematic). To test the type I error we can apply the familiar binomial test: let \( y_0 \) denote the estimated power at \( \theta = \theta_0 = \theta_t \); then we reject a procedure if \( y_0 - z^\gamma [a(1-a)/r]^{1/2} \) exceeds \( \alpha \) where \( z^\gamma \) is found in the table for the standard Gaussian variable: \( P(z > z^\gamma) = \gamma \). We propose to perform this binomial test for several \( \alpha \) values, say, the classical \( \alpha \) values of 0.10, 0.05,
0.01 plus a very small \( \alpha \) value like 0.001 (where this small value may be relevant when applying the Bonferroni inequality to ensure a specific experimentwise error; see Miller, 1981). We observe that Schruben (1980) evaluated the estimated \( \alpha \) error at all possible values \( 0 < \alpha < 1 \); however, practitioners are not interested in high \( \alpha \) values.

If several procedures meet the condition of acceptable \( \alpha \) errors, then we may compare their power functions by transforming the problem into the comparison of distribution functions, as we noticed at the end of the Section 4. There are standard procedures for testing whether \( u_I \) is stochastically smaller than \( u_{II} \), that is, is \( F(u_I) \geq F(u_{II}) \)? For example, the sign statistic applied to a one-sided test (case i) means that we score a one if the lower limit \( u_I \) is smaller for procedure I than it is for procedure II: \( s_i = 1 \) if \( u_{II} < u_{III} \) (\( i = 1, \ldots, r \)), etc.

We notice that for each \( \alpha \) value we estimate one power function. Consequently from a single replicate \( X \) we estimate one point on the power function, given a specific \( \alpha \) value and given a specific test statistic. Other (dependent) estimators result if we vary \( \alpha \) and/or the test statistic.

**NOTES**

1. The Monte Carlo experiment of Section 5 was programmed and executed by K. Baken and F. van Lent, students in the Department of Econometrics, Tilburg University.

**REFERENCES**


Nozari, A., Dealing with Unequal Variances in Analysis of Experimental Design Applications. School of Industrial Engineering, University of Oklahoma, Norman (Oklahoma), 1983. (Accepted for publication in Communications in Statistics.)


<table>
<thead>
<tr>
<th>No.</th>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>126</td>
<td>H.H. Tigelaar</td>
<td>Identification of noisy linear systems with multiple arma inputs.</td>
</tr>
<tr>
<td>128</td>
<td>A.J. de Zeeuw</td>
<td>Two notes on Nash and Information.</td>
</tr>
<tr>
<td>131</td>
<td>R. van Montfort, R. Schippers, R. Heuts</td>
<td>Johnson $S_0$-transformations for parameter estimation in arma-models when data are non-gaussian.</td>
</tr>
<tr>
<td>132</td>
<td>J. Glombowski en M. Krüger</td>
<td>On the Rôle of Distribution in Different Theories of Cyclical Growth.</td>
</tr>
<tr>
<td>134</td>
<td>W.J. Oomens</td>
<td>The economic interpretation of the advertising effect of Lydia Pinkham.</td>
</tr>
<tr>
<td>135</td>
<td>J.P.C. Kleijnen</td>
<td>Regression analysis: assumptions, alternatives, applications.</td>
</tr>
<tr>
<td>136</td>
<td>J.P.C. Kleijnen</td>
<td>On the interpretation of variables.</td>
</tr>
</tbody>
</table>
138 G.J. Cuypers, J.P.C. Kleijnen en J.W.M. van Rooyen
Testing the Mean of an Asymmetric Population:
Four Procedures Evaluated

139 T. Wansbeek en A. Kapteyn
Estimation in a linear model with serially correlated errors when
observations are missing

140 A. Kapteyn, S. van de Geer, H. van de Stadt, T. Wansbeek
Interdependent preferences: an econometric analysis

141 W.J.H. van Groenendaal
Discrete and continuous univariate modelling

142 J.P.C. Kleijnen, P. Cremers, F. van Belle
The power of weighted and ordinary least squares with estimated
unequal variances in experimental design