J. Hyg., Camb. (1961), **59**, 449
Printed in Great Britain

A sampling study of the power function of the binomial χ^2 'index of dispersion' test

BY B. M. BENNETT AND P. HSU†

University of Washington, Seattle, U.S.A.

(*Received* 24 *May* 1961)

INTRODUCTION AND STATEMENT OF THE PROBLEM

It is known that in observing the occurrence of a certain event (e.g. the number of successes of a certain medical treatment in a series of clinical trials in which it is used) the frequencies of these occurrences should conform to the binomial distribution. In order to indicate whether the results of a number of independent clinical trials, each of which consists of the number of successes in a certain number of trials, show more variability than can be usually expected from a binomial distribution, it is customary to employ Fisher's χ^2-test of heterogeneity (Fisher, 1954) based on the 'index of dispersion' given in equation (2) below. It is of interest to determine how easily this test will in fact detect heterogeneity in the results of a series of clinical trials. This paper will be concerned with the 'power function' of the χ^2-test, i.e. the probability that the test indicates significant variability in the presence of heterogeneity. A similar investigation of the special case of the Poisson 'index of dispersion' was given by one of the authors (Bennett, 1959). This earlier paper gives a review of other studies on the index of dispersion. Of particular interest will be situations in which the number of trials is small.

Suppose that $x_1, x_2, \ldots, x_t$ represent t independent observations, each being the result of n trials from the separate binomial distribution

$$\frac{n!}{x_i!\,(n-x_i)!}\,\mathbf{p}_i^{x_i}\,(1-\mathbf{p}_i)^{n-x_i}, \tag{1}$$

for $x_i = 0, 1, \ldots, n$ ($i = 1, \ldots, t$), and where $\mathbf{p}_i$ represents the probability that a certain event E (e.g. success) occurs:

Series	No. of successes	No. of failures	Probability of success
1	x_1	$n-x_1$	$\mathbf{p}_1$
2	x_2	$n-x_2$	$\mathbf{p}_2$
.	.	.	.
.	.	.	.
.	.	.	.
t	x_t	$n-x_t$	$\mathbf{p}_t$
Total	$x = \sum_i x_i$	$N-x$	

Thus the above series may denote the evaluation of results of repeated clinical trials involving certain therapeutic treatments, or possibly various levels of the same treatment, on separate samples of equal numbers of patients.

† Now at the Boeing Company, Seattle.

In order to test the hypothesis H_0: $\mathbf{p}_1 = \mathbf{p}_2 = \ldots = \mathbf{p}_t$ of the equality of probability of E from series to series, the 'index of dispersion', or χ^2-test (Fisher, 1954)

$$\chi^2 = \frac{1}{n\bar{p}(1-\bar{p})} \sum_{i=1}^{t} (x_i - n\bar{p})^2 \geqslant \chi^2_\alpha \tag{2}$$

is used, where $N\bar{p} = x$, $N = nt$, and χ^2_α is the critical value of the (uncorrected) χ^2 distribution with $(t-1)$ degrees of freedom and level of significance equal to $\alpha\,\%$. Cochran (1936) and Hoel (1943) have examined the approximate distribution form of χ^2 if H_0 is true.

This paper will be concerned with some 'Monte-Carlo' results on the corresponding power function

$$\beta(\delta) = P\{\chi^2 \geqslant \chi^2_\alpha \mid \delta\}, \tag{3}$$

or the probability of a significant result of the test (2) as a function of the parameter

$$\delta = \frac{n}{\bar{\mathbf{p}}(1-\bar{\mathbf{p}})} \sum_{i=1}^{t} (\mathbf{p}_i - \bar{\mathbf{p}})^2 \tag{4}$$

if we set $\bar{\mathbf{p}} = (\mathbf{p}_1 + \mathbf{p}_2 + \ldots + \mathbf{p}_t)/t$. In particular, emphasis will be placed on alternatives H_t: $\mathbf{p}_1 \leqslant \mathbf{p}_2 \leqslant \ldots \leqslant \mathbf{p}_t$ of a 'trend' in the binomial probabilities for small assigned values of t and n in order to examine the adequacy of the χ^2 distribution for (2). Other tests have been proposed for the alternatives H_t (Cochran, 1954), but these will not be discussed here.

PROCEDURE

For this purpose a number (= 375) of 'Monte-Carlo' trials, each consisting of 100 successive samples $\{x_{ik}\}$ from t different binomial populations ($i = 1, \ldots, t$; $k = 1, 2, \ldots, 100$) were prepared on the IBM-709. The agreement of the successive samples $\{x_{ik}\}$ with the binomial distributions (1) were separately tested by means of the 'goodness-of-fit' test.

For each of the 100 samples, the quantities $N\bar{p}_k = \sum_{i=1}^{t} x_{ik}$ and

$$\chi^2_k = \frac{1}{n\bar{p}_k(1-\bar{p}_k)} \sum_{i=1}^{t} (x_{ik} - np_k)^2$$
$$= \frac{1}{n\bar{p}_k(1-\bar{p}_k)} \left[\left(\sum_{i=1}^{t} x^2_{ik}\right) - \frac{1}{t}\left(\sum_{i=1}^{t} x_{ik}\right)^2\right] \tag{5}$$

were computed for $k = 1, 2, \ldots, 100$ and χ^2_k tested whether it exceeds the critical value χ^2_α for levels of significance α equal to $5\,\%$ and $1\,\%$, respectively. Because of computational checks at various stages, it is believed that the present realizations are essentially error-free.

Table 1 indicates the number of empirical realizations of the power function $\beta^*(\delta)$, where

$$\beta^*(\delta) = \text{relative frequency of event: } \chi^2_k \geqslant \chi^2_\alpha, \tag{6}$$

i.e. the proportion or percentage of 'exceedances' in samples of a hundred realizations each. These have been computed for various combinations of values of the sequence $\{\mathbf{p}_i\}$ $(i = 1, \ldots, t)$ under H_t such that $\bar{\mathbf{p}} = 0{\cdot}20, 0{\cdot}25, 0{\cdot}30, 0{\cdot}40$ is constant. For a given t and n, each of the combinations listed represents a different monotone increasing sequence of the $\mathbf{p}$'s.

It will be noted from the range of computational values of the sequence $\mathbf{p}_i$ that this is operationally equivalent to the use of the χ^2 variance test with small expected values in a variety of experiments. Since the χ^2 test is also one of association in a $2 \times t$ contingency table, concerning which the following approximations are known (e.g. Cochran, 1954),

$$\left.\begin{aligned} E(\chi^2) &\cong (t-1)\left(1+\frac{1}{N}\right), \\ \text{var. } (\chi^2) &\cong 2(t-1)\left(\frac{n-1}{n}\right)\left(1-\frac{1-7\bar{\mathbf{p}}\bar{\mathbf{q}}}{N\bar{\mathbf{p}}\bar{\mathbf{q}}}\right), \end{aligned}\right\} \tag{7}$$

where $\bar{\mathbf{q}} = 1-\bar{\mathbf{p}}$, if the hypothesis H_0 is true, the variance of χ^2 is less than that of the approximating distribution of χ^2, especially if n is small.

Table 1. *Number of realizations of empirical power function* $= \beta^*(\delta)$

(Samples of 100 each)

t	n	$\bar{\mathbf{p}} = 0{\cdot}20$	$\bar{\mathbf{p}} = 0{\cdot}25$	$\bar{\mathbf{p}} = 0{\cdot}30$	$\bar{\mathbf{p}} = 0{\cdot}40$	Total
3	4	9	14	7	16	46
3	8	9	14	7	16	46
3	10	9	14	7	16	46
4	6	11	10	18	13	52
5	8	11	16	14	12	53
5	12	11	16	14	12	53
7	10	10	7	12	15	44
9	15	9	7	10	9	35
					Total	375

RESULTS

These realizations are presented in the figures below in the form of graphs of values of $\beta^*(\delta)$ for various assigned δ's. Separate graphs are given for the values of $\bar{\mathbf{p}}$ as indicated by the various dotted lines. The assigned level of significance is 5 %; the results for the corresponding 1 % levels are not reproduced because of limitations of space.

While the power function $\beta(\delta)$ of the test (2) will not coincide exactly with the non-central χ^2 distribution under the alternatives H_t, this has been used for comparison purposes in the present series of results.

From Patnaik (1949) it is known that the power function of the corresponding non-central χ^2 $(= \chi_\nu'^2)$ with $\nu = t-1$ degrees of freedom and parameter equal to δ is given by

$$\begin{aligned} \beta(\delta) &= P\{\chi_\nu'^2 \geqslant \chi_\alpha^2\} \\ &= e^{-\frac{1}{2}\delta} \sum_{j=0}^{\infty} \frac{1}{j\,!} (\tfrac{1}{2}\delta)^j\, Q(\chi_\alpha^2 \mid \nu+2j), \end{aligned}$$

where
$$Q(\chi_\alpha^2 \mid r) = \frac{1}{\Gamma(\frac{1}{2}r)} \int_{\frac{1}{2}\chi_\alpha^2}^{\infty} v^{\frac{1}{2}r-1} e^{-v}\, dv. \tag{8}$$

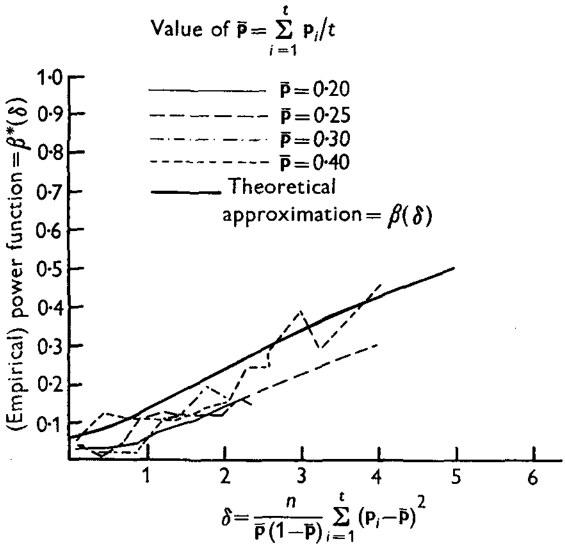

Fig. 1*a*. $t = 3$, $n = 4$.

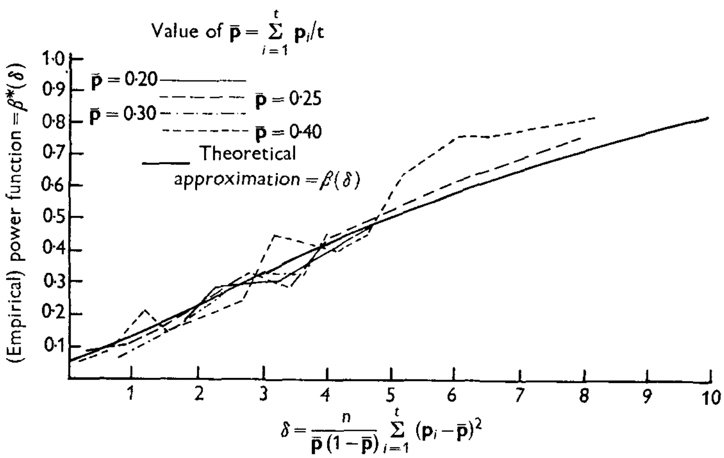

Fig. 1*b*. $t = 3$, $n = 8$.

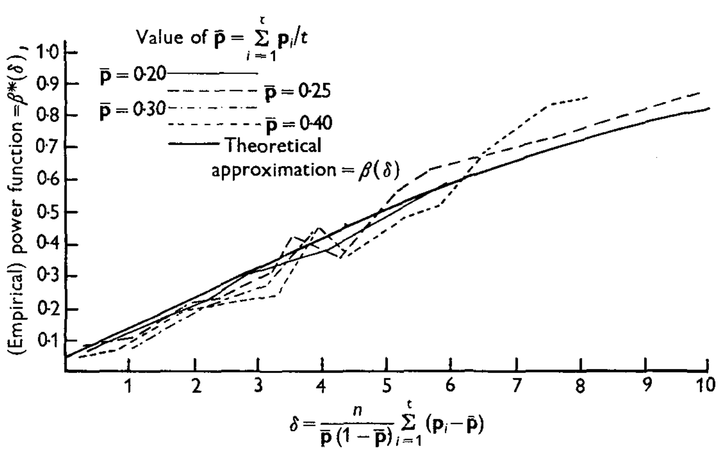

Fig. 1*c*. $t = 3$, $n = 10$.

https://doi.org/10.1017/S0022172400039140 Published online by Cambridge University Press

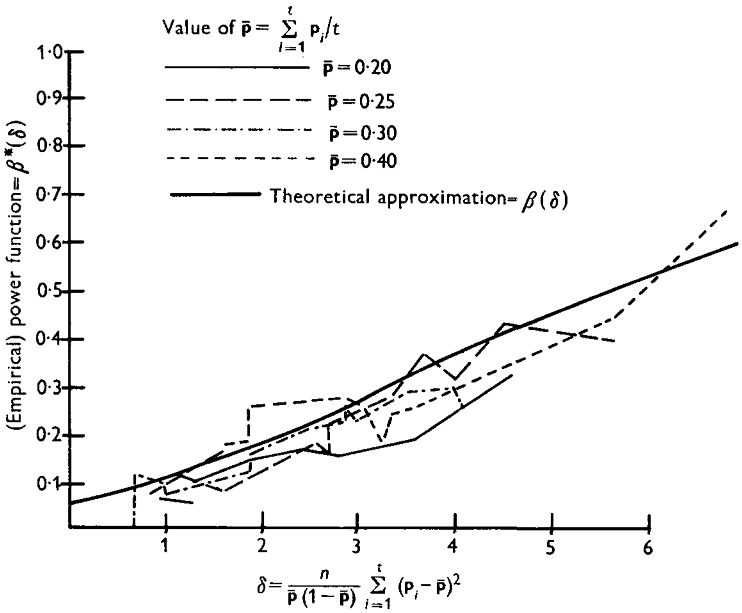

Fig. 2. $t = 4$, $n = 6$.

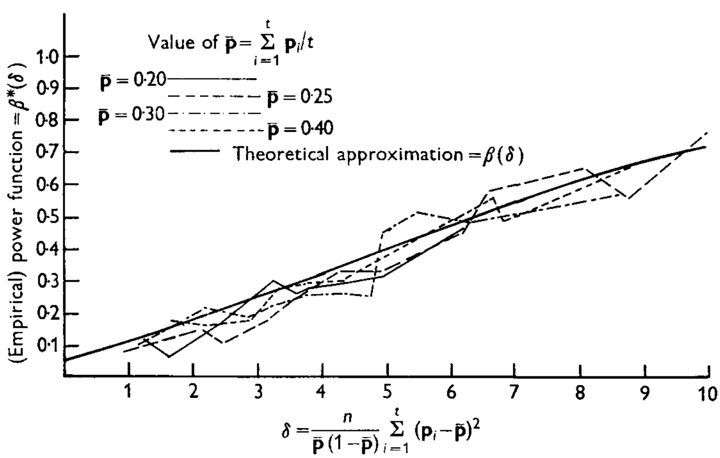

Fig. 3*a*. $t = 5$, $n = 8$.

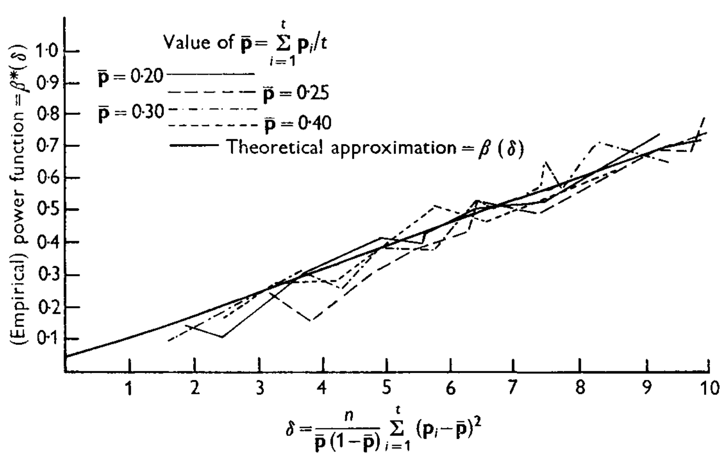

Fig. 3*b*. $t = 5$, $n = 12$.

The values of $\beta(\delta)$ are drawn in the heavy lines on each of the graphs. It is to be noted that the power function of the χ^2-test is almost systematically over-estimated by the values of $\beta(\delta)$, though for a fixed value of t the approximation improves as n becomes larger (Figs. 1a–c; 3a, b). A similar result seems to hold also for the Poisson test (Bennett, 1959).

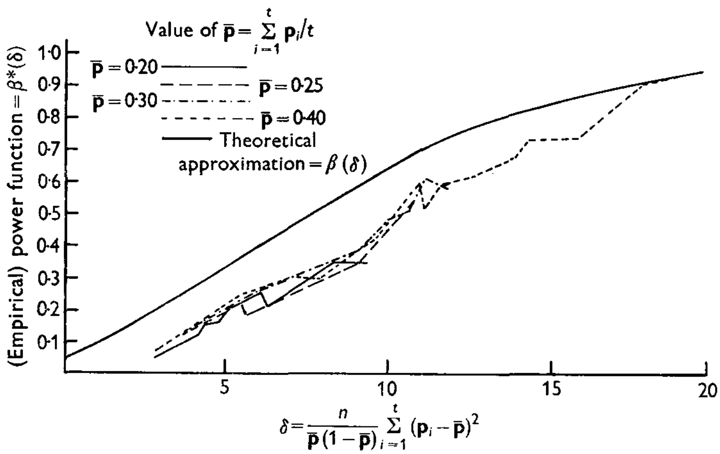

Fig. 4. $t = 7$, $n = 10$.

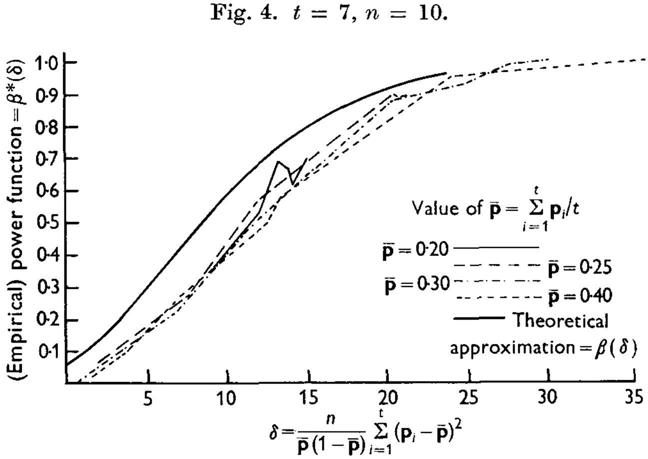

Fig. 5. $t = 9$, $n = 15$.

As regards the graphical interpretations of these results, the sampling errors of $\beta^*(\delta)$ must of course be taken into consideration. Thus if for a given δ_0, $\beta_1(\delta_0)$ represents the value of the true power function of the χ^2-test,

$$N^* \text{ var.}\{\beta^*(\delta_0)\} = \{\beta_1(\delta_0)\}\{1-\beta_1(\delta_0)\}$$

based on samples of size N^*.

SUMMARY

A total of 375 separate sampling experiments were performed on groupings of 100 samples each from distinct binomial populations in order to obtain estimates of the frequency of rejection of the hypotheses of homogeneity against alternatives

https://doi.org/10.1017/S0022172400039140 Published online by Cambridge University Press

of a trend by means of the usual χ^2 or binomial index of dispersion test. These artificial realizations are presented in Figs. 1–5, and compared graphically with corresponding non-central χ^2 distributions.

The authors would like to acknowledge the technical assistance of the Research Computer Laboratory of this University in programming and supervising the computational phases of this investigation, which was partially supported by the State of Washington Initiative 171 Funds for Research in Biology and Medicine and the Boeing Employees' Good Neighbor Fund.

REFERENCES

BENNETT, B. M. (1959). A sampling study on the power function of the χ^2 'index of dispersion' test. *J. Hyg., Camb.*, **57**, 360.

COCHRAN, W. G. (1936). The χ^2 distribution for the binomial and Poisson series, with small expectations. *Ann. Eugen., Lond.*, **7**, 207.

COCHRAN, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, **10**, 417.

FISHER, R. A. (1954). *Statistical Methods for Research Workers*, sect. 19, 12th ed. Edinburgh: Oliver and Boyd.

HOEL, P. (1943). On the indices of dispersion. *Ann. Math. Statist.* **14**, 155.

PATNAIK, P. B. (1949). The non-central χ^2 and F-distributions and their applications. *Biometrika*, **36**, 202.