

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

SP 109

(S 402)

R.D. Boschloo

Raised conditional level of significance for the
2x2-table when testing the equality of two probabilities.

(Statistica Neerlandica 24 (1970), p. 1-35)



Raised conditional level of significance for the 2×2 -table when testing the equality of two probabilities

by R. D. BOSCHLOO *

Summary In this paper it is to be shown that FISHER's non-randomizing exact test for 2×2 -tables, which is a conditional test, can by simple means be changed into an unconditional test using raised levels of significance; not seldom, especially for not too large samples, the level of significance can be doubled. This leads in many cases to a considerable increase of power of the test. A table with raised levels has been prepared up to sample sizes of 50 and a rule of thumb, which can be used if this table is not available, has been developed.

1. Introduction

The usual test for the hypothesis $H_0 : p_1 = p_2$, given the results of two independent sequences of experiments, is the exact test of FISHER (1925) in the 2×2 -table.

p_1	A	C	m
p_2	B	D	n
	R	S	t

Denoting the number of successes in the first sequence by A^{**} and the number of successes in the second sequence by B , the test is based on the conditional distribution of A under the condition $R = r$, where $R = A + B$ and where r is the value actually found for R . The conditional distribution mentioned is, under H_0 , a hypergeometric distribution.

The level of significance used for this non-randomized conditional test is further to be called the "conditional level of significance" and often shortly written as "conditional level". If α is this conditional level then the unconditional level of significance (being the true probability of unjustified rejection of H_0 , in this paper often written as "unconditional level") depends on the common value p of p_1 and p_2 .

This unconditional level is denoted by $\alpha(p)$.

The use of the conditional test derives its justification from the fact that the inequality $\alpha(p) < \alpha$ holds for every value of p . Numerical calculations, however, show that not only for most values of

to make the conditional size equal to α , for every r separately. There are, however, well known objections of practical and fundamental character against randomization of this kind.

Another method was advocated for some time by BARNARD (1945). He abandoned his proposal later after a discussion with FISHER (1945). The idea of BARNARD is to use an unconditional test for H_0 instead of a conditional one. This idea was taken up by GARSIDE and MACK (1967) in a slightly modified form. They give tables for sample sizes up to 15.

The test proposed in the present paper is another modification, more closely linked up with FISHER's original test and suitable for samples of all sizes: tables, with only two entrances for each α , up to samples of size 50 and an approximation for larger samples.

2. The unconditional level of significance of the fisher test

The possible outcomes for the binomially distributed number of successes A in the first sequence of experiments are $a = 0, 1, \dots, m$. Likewise B in the second sequence has the possible outcomes $b = 0, 1, \dots, n$. Each of the possible results ($A = a$,

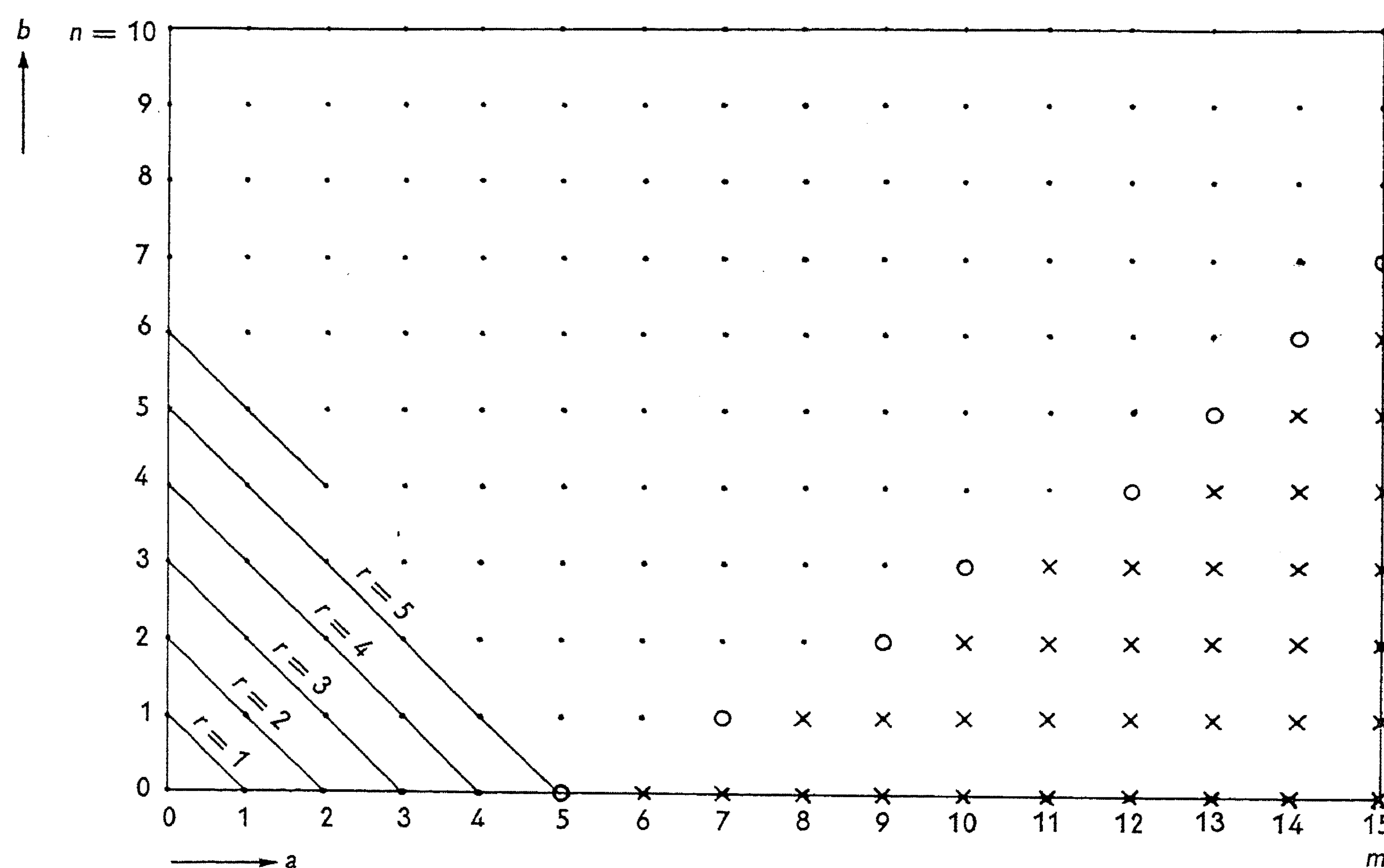


Fig. 1. The sample space of the 2×2 -table with $m = 15$ and $n = 10$. The points marked $\times$ are part of an one sided critical region with conditional level of significance $\alpha = .05$. The points marked $\circ$ are part of an one sided critical region with conditional level of significance $\alpha = .09$.

$B = b$) of the 2×2 -table corresponds with the point (a, b) of the sample space representable by a rectangle as drawn in figure 1.

In figure 1 the sample space of the original FISHER test when $R = r$ consists of the points (a, b) with $a + b = r$ being the points $(a, r - a)$ on the line $a + b = r$.

The critical region K_r in this conditional sample space is to be found in the usual way by means of the hypergeometric distribution. One finds the overall critical region K by taking all regions K_r together.

In figure 1, where $m = 15$, $n = 10$, $\alpha = .05$ and $H_0 : p_1 \leq p_2$, the points marked $\times$ are in the critical region K .

It is interesting to know the true probability, under H_0 , of finding an outcome that corresponds with a point of K . This probability, the unconditional level of significance, is

$$\alpha(p) = P[(A, B) \in K | H_0] = \sum_r \alpha_r \cdot p(r | H_0) \quad (1)$$

where

$$p(r | H_0) = P[R = r | H_0] = \binom{m+n}{r} p^r (1-p)^{m+n-r} \quad (2)$$

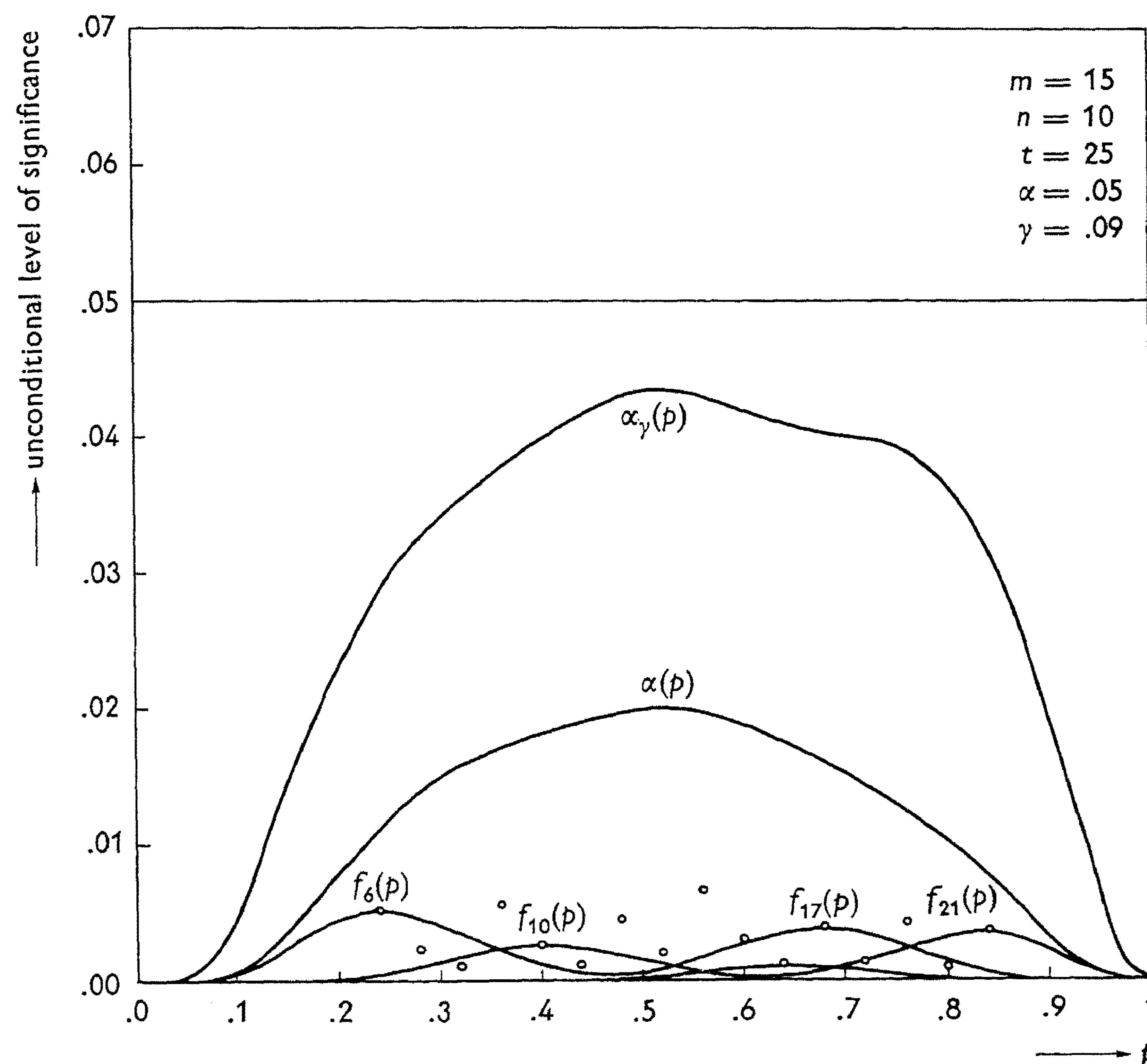


Fig. 2. The unconditional level of significance $\alpha(p) = \sum f_r(p)$ where $f_r(p) = \alpha$

and where

$$\alpha_r = P[(A, B) \in K_r | R = r, H_0] \quad (3)$$

is the sum of some hypergeometric point probabilities which is usually smaller than α .

In figure 2 this function $\alpha(p)$ and some of the functions $f_r(p) = \alpha_r \cdot p(r|H_0)$ are drawn; m , n and α here have the same values as in figure 1.

It is striking in this diagram that even the maximum $\alpha_{\max}$ of $\alpha(p)$, being the size of the test, remains considerably under the conditional level of significance α ($= .05$) and in this case amounts to about .02.

3. A raised conditional level of significance

The improvement of the test suggested in this paper is based on the following idea:

Seeing that the use of the conditional level α leads to a too small unconditional level $\alpha(p)$, a raised conditional level γ is selected in such a way that the unconditional level $\alpha_\gamma(p)$ does not exceed for any value of p the desired α . Under this condition γ is chosen as high as possible.

In the case of figure 1 and 2 the use of a raised level $\gamma = .09$ is allowed. Using $\gamma = .09$ instead of .05 as a conditional level, the critical region in figure 1 is extended with the points marked $\bigcirc$ and the unconditional level $\alpha_\gamma(p)$ drawn in figure 2 gives a much better approximation of the desired level $\alpha = .05$.

It is stressed once more that the calculations for Fisher's test remain unchanged; only the level of significance used is raised.

In the two sided test the critical region K consists of two separate parts of the sample space lying diametrical with regard to each other. First only one part is taken into account: Suppose the conditional level $\gamma'/2$ corresponding to this part is chosen as high as possible but in such a way that the unconditional level does not exceed $\alpha/2$.

If $\alpha_{\gamma'/2}(p)$ is the unconditional level corresponding to this part of the critical region than $\alpha_{\gamma'/2}(1-p)$ is the unconditional level corresponding

References

- BARNARD, G. A. (1945), A new test for 2×2 -tables. *Nature* **156**, 177.
- BARNARD, G. A. (1945), A new test for 2×2 -tables. *Nature* **156**, 783–784.
- BARNARD, G. A. (1947), Significance test for 2×2 -tables. *Biometrika* **34**, 123–138.
- BENNETT, B. M. and P. HSU (1960), On the power function of the exact test for the 2×2 contingency table. *Biometrika* **47**, 393–396.
- FISHER, R. A. (1925), *Statistical methods for research workers*. Edinburgh and London: Oliver and Boyd.
- FISHER, R. A. (1945), A new test for 2×2 -tables. *Nature* **156**, 388.
- GARSDIE, G. R. and C. MACK (1967), Correct confidence limits for the 2×2 homogeneity contingency table with small frequencies. *The new journal of statistics and operational research*, Vol. III, part II, 1–25.
- KLERK-GROBBEN, G. and H. J. PRINS (1954), Toets voor gelijkheid van twee kleine kansen met behulp van even grote steekproeven en het onderscheidingsvermogen van deze toets. *Statistica* **8**, 7–20.
- LEHMANN, E. L. (1950), *Testing statistical hypotheses*. New York: Wiley and Sons.
- PATNAIK, P. B. (1948), The power function of the test for the difference between two proportions in a 2×2 -table. *Biometrika* **35**, 157–175.
- SILLITTO, G. P. (1949), Note on approximation to the power function of a 2×2 comparison trial. *Biometrika* **36**, 347–352.
- TOCHER, K. D. (1950), Extensions of the Neyman-Pearson theory of test to discontinuous variates. *Biometrika* **37**, 130–144.

Raised conditional level of significance for 2x2-tables (Fisher's test).

		Unconditional level of significance (one or two sided) :							
Sample sizes		one s. .005	two s. .01	one s. .01	one s. .025	two s. .05	one s. .05	two s. .10	one s. .10
2	1	-	-	-	-	-	-	-	.3
3	1	-	-	-	-	-	-	-	.2
3	2	-	-	-	.1	.2	.1	.2	.10
4	1	-	-	-	-	-	-	.6	.3
4	2	-	-	-	.1	.2	.1	.2	.3
4	3	.04	.08	.04	.025	.050	.13	.26	.2
5	1	-	-	-	-	-	-	.6	.2
5	2	-	.10	.05	.1	.2	.05	.4	.2
5	3	.02	.04	.02	.1	.2	.14	.28	.2
5	4	.01	.02	.01	.05	.10	.09	.18	.24
6	1	-	-	-	-	-	-	.4	.2
6	2	-	.06						

