stichting
mathematisch
centrum

MC

DEPARTMENT OF MATHEMATICAL STATISTICS SW 45/76 MAY

J.M. BUHRMAN

TESTS AND CONFIDENCE INTERVALS FOR THE DIFFERENCE AND RATIO OF TWO PROBABILITIES

Prepublication

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS(MOS) subject classification scheme (1970): 62F05, 62F25, 62J99, 62K99

Tests and confidence intervals for the difference and ratio of two probabilities *)

by

J.M. Buhrman

SUMMARY

FISHER's (1925) test can deal with the hypothesis of equal probabilities of success of two experiments A and B. The test requires a fixed number of experiments of both kinds. If the experimental design is changed into randomizing the numbers of experiments A and B, other kinds of hypotheses can be tested and exact confidence intervals can be constructed.

KEY WORDS & PHRASES: *Bernoulli trials, confidence intervals, experimental design, linear hypotheses*

---

*) This paper is not for review; it is meant for publication elsewhere.

1. INTRODUCTION. THE EXPERIMENTAL DESIGN

Let the experiments A and B have probabilities of success p_1 and p_2 respectively. If A is performed n_1 times, and B n_2 times, and X_1 and X_2 represent the numbers of successes, then the conditional distribution of X_1, given $X_1 + X_2 = r$, can be used for testing any hypothesis concerning $\theta = p_1(1-p_2)/(p_2(1-p_1))$, because for a fixed value of θ this conditional distribution is independent of p_1 and p_2. In particular, if $\theta = 1$, which is equivalent to $p_1 = p_2$, this conditional distribution is the hypergeometric distribution

$$P(X_1=x \mid X_1+X_2=r) = \binom{n_1}{x}\binom{n_2}{r-x}/\binom{n_1+n_2}{r}.$$

This result was given first by FISHER (1925), and thoroughly discussed by BARNARD (1947).

We now turn to the randomized design. This consists of letting the number of experiments of type A be binomially distributed with parameters n and π. So let us define N_1 as a random variable with a binomial-(n,π) distribution, and $N_2 = n - N_1$. Then we denote by S_1 and F_1 the number of successes and failures among N_1 A-experiments and with S_2 and F_2 the analogous numbers of N_2 B-experiments. All outcomes are given in table 1.

Table 1. Outcomes with the modified experimental design and four examples to be discussed in section 2 and 3.

			(a)			(b)			(c)			(d)		
S_1	F_1	N_1	12	30	42	20	36	56	22	26	48	2	39	41
S_2	F_2	N_2	30	28	58	25	19	44	38	14	52	10	49	59
S	F	n	42	58	100	45	55	100	60	40	100	12	88	100

It should be noted that (S_1,S_2,F_1,F_2) has a multinomial distribution with parameters n, πp_1, $(1-\pi)p_2$, $\pi(1-p_1)$, $(1-\pi)(1-p_2)$. S_1, S_2, F_1 and F_2 can be thought of to be obtained as follows. A total of n times one gambles with probability π whether experiment A must be performed, if not B is performed. Every time one of the four outcomes occurs and S_1, S_2, F_1 and F_2 are the numbers these outcomes occur.

2. TESTS AND CONFIDENCE INTERVALS WITH RESPECT TO $p_1 \pm cp_2$

The marginal distribution of $S_1 + F_2$ is binomial with parameters n and $\pi p_1 + (1-\pi)(1-p_2) = \pi p_1 - (1-\pi)p_2 + 1 - \pi$. The marginal distribution of $S_1 + F_1$ is binomial with parameters n and $\pi p_1 + (1-\pi)p_2$. This means that any test concerning $\pi p_1 \pm (1-\pi)p_2$ can be performed. For a certain hypothesis $p_1 \pm cp_2 = d$ the experimental design must be determined by means of the choice of $\pi = 1/(1+c)$. Then

$$\pi p_1 \pm (1-\pi)p_2 = (p_1 \pm cp_2)/(1+c).$$

Suppose H_0: $p_1 - cp_2 = d$. Then under H_0 $S_1 + F_2$ has a bin(n,p^*) distribution in which

$$p^* = \pi p_1 - (1-\pi)p_2 + 1 - \pi = (p_1-cp_2)/(1+c) + c/(1+c) = (d+c)/(1+c).$$

The test for H_0 rejects if $S_1 + F_2$ falls into the critical region of the bin(n,p^*) distribution.

A confidence interval is constructed as follows. Suppose an observation of $S_1 + F_2$ results in a confidence interval (p_ℓ^*,p_r^*) for p^*. Then

$$p_\ell^* < (p_1-cp_2+c)/(1+c) < p_r^*$$

is equivalent to

$$(1+c)p_\ell^* - c < p_1 - cp_2 < (1+c)p_r^* - c.$$

In particular, a confidence interval for $p_1 - p_2$ is obtained by a choice of $\pi = \frac{1}{2}$ as

$$(2p_\ell^*-1,2p_r^*-1).$$

The following example shows that this procedure may give a confidence interval which contains 0, notwithstanding the fact that FISHER's test applied to the same outcomes, would reject the hypothesis $p_1 = p_2$; the reason is that the confidence interval is based on another test. In example (a) of table 1 $S_1 + F_2 = 40$, which gives as a 95% confidence interval for p^*: (0.31,0.51), as found from the charts in Biometrika Tables (1970). The corresponding interval for $p_1 - p_2$ is (-0.38,0.02).

If we apply Fisher's test, disregarding the fact that the numbers of experiments were obtained as random variables, which can easily be shown to be allowed, we have

$$P(S_1 \le 12 | S_1 + S_2 = 42, N_1 = 42) = 0.0168 < 0.025.$$

The opposite can occur as well. The outcomes of table 1 (b) give for $p_1 - p_2$ the 95% confidence interval (-0.42,-0.02). FISHER's test gives

$$P(S_1 \le 20 | S_1 + S_2 = 45, N_1 = 56) = 0.0284 > 0.025$$

and does not reject $p_1 = p_2$.

3. TESTS AND CONFIDENCE INTERVALS WITH RESPECT TO p_1/p_2

The distribution of S_1 conditioned on $S_1 + S_2 = s$ is binomial with parameters s and $p' = \pi p_1/(\pi p_1 + (1-\pi)p_2)$. Since p' depends on p_1 and p_2 only through p_1/p_2, this gives the opportunity to test $p_1/p_2 = c$ for any positive c. Here the choice of π is free. Numerical investigations (BUHRMAN, 1975) show that $\pi = \frac{1}{2}$ is a reasonable value for most c. This consideration is important since for the construction of a confidence interval a test must be available for all values of c. As in section 2 an example shows that 1 may belong to the confidence interval for p_1/p_2, while FISHER's test rejects $p_1 = p_2$. The outcomes of table 1 (c) give as a 95% confidence interval for p': (0.25,0.51). The corresponding interval for p_1/p_2 is found by

$$0.25 < \frac{\pi p_1}{\pi p_1 + (1-\pi)p_2} < 0.51$$

which is, with $\pi = \frac{1}{2}$, equivalent to

$$0.33 < p_1/p_2 < 1.04.$$

However,

$$P(S_1 \leq 22 | S_1+S_2=60, N_1=48) = 0.0049.$$

The opposite occurs in example (d), where the 95% confidence interval for p_1/p_2, is (0.04,0.96). However,

$$P(S_1 \leq 2 | S_1+S_2=12, N_1=41) = 0.061.$$

Although examples like (d) can be found, it should be stressed that for testing $p_1/p_2 = 1$ FISHER's method should be preferred, since its power is in general a lot better than that of the method described above.

It should be noted, that with inverse sampling hypotheses concerning p_1/p_2 can be tested. Experiment A is repeated until f_1 failures occurred, B is repeated until f_2 failures occurred. Now N_1 and N_2 have negative binomial distributions and the conditional distribution of N_1 given N_1+N_2 depends on p_1 and p_2 only through p_1/p_2, but is slightly more complicated than the distribution used above. Moreover, the total number of experiments that can be done, may be limited, so that the investigator may fail to obtain the prescribed numbers of failures.

ACKNOWLEDGEMENT

This paper was written after valuable discussions with Prof.Dr. J. Hemelrijk.

REFERENCES

BARNARD, G.A., (1947), *Significance tests for 2 × 2 tables*, Biometrika 34, 123-37.

BUHRMAN, J.M., (1975), *Linear hypotheses concerning two probabilities*, Report 75-15, Department of Mathematics, University of Amsterdam.

FISHER, R.A., (1925), *Statistical Methods for Research Workers.*

PEARSON, E.S. & H.O. HARTLEY, (1970), *Biometrika Tables for Statisticians,* Volume I.