A TEST OF SIGNIFICANCE OF A DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS WHEN THE PROPORTIONS ARE VERY SMALL

By P. B. PATNAIK

Indian Statistical Institute, Calcutta

1. Introduction

Various methods have been proposed for testing the significance in a 2×2 table. The problem is that if there is a random sample of N_1 individuals from a population, of which x_1 have a character A, and a random sample of N_2 from a second population, of which x_2 have A, then it is desired to test whether the chance of possessing A is the same in the two populations, i.e. whether $p_1(A)=p_2(A)$.

In their discussion of this and similar problems, Barnard (1947), Pearson (1947) and Patnaik (1948) have described the difference in the two methods of regarding this

(a) as a two dimensional problem in which the observational result (x_1, x_2) referred to an experimental probability set consisting of all possible points for which $0 \leqslant x_1 \leqslant N_1$ and $0 < x_2 \leqslant N_2$,

(b) as a one dimensional problem in which the probability set consists of all points for which x_1+x_2=observed n.

In the present paper we shall be concerned with the case where (i) N_1 and N_2 are very large (but of known values), (ii) $p_1(A)$ and $p_2(A)$ are very small and (iii) $m_1=p_1N_1$, $m_2=p_2N_2$ are finite. Thus the sampling distributions of x_1 and x_2 may be taken as following the Poisson limit to the binomial and the null hypothesis $p_1=p_2$ becomes equivalent to the hypothesis that the expectations m_1 and m_2 are proportional to the sample sizes N_1 and N_2. In the special case of $N_1=N_2$, the problem reduces to that considered by Przyborowski and Wilenski (1939), namely that of testing whether the two samples are drawn from a common Poisson population.

The probability of an observed pair (x_1, x_2) is

$$p(x_1, x_2) = \frac{e^{-m_1} m_1^{x_1}}{x_1!} \cdot \frac{e^{-m_2} m_2^{x_2}}{x_2!}$$

$$= \frac{e^{-\mu}\mu^n}{n!} \cdot \frac{n!}{x_1!(n-x_1)!}\left(\frac{m_1}{\mu}\right)^{x_1} \cdot \left(\frac{m_2}{\mu}\right)^{n-x_1}$$

where
$$\left.\begin{aligned}\mu &= m_1+m_2,\\ n &= x_1+x_2.\end{aligned}\right\} \quad \ldots \quad (1)$$

This is the product of two probabilities, the probability of n and the relative probability of x_1 given n. Thus

$$\left.\begin{aligned} p(n|\mu) &= \frac{e^{-\mu}\mu^n}{n!}, \\ p(x_1|n,\rho) &= \frac{n!}{x_1!(n-x_1)!}\,\rho^{x_1}(1-\rho)^{n-x_1}, \end{aligned}\right\} \qquad \dots \quad (2)$$

in which $$\rho = \frac{m_1}{\mu} = \frac{m_1}{m_1+m_2}. \qquad \dots \quad (3)$$

Denoting $N_1/(N_1+N_2)$ by ρ_0, we see that the hypothesis to be tested is that $\rho=\rho_0$.

The problem can be illustrated in a diagram. Fig. 1 represents the sample space in two dimensions and taking $N_1=3N_2$(i.e. $\rho_0=0.75$), the line OK contains points

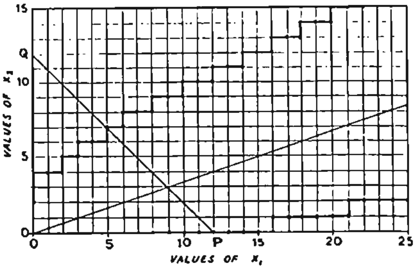

Fig. 1
SHOWING CRITICAL CONTOURS. $N_1=3N_2$ $(\rho=0.75)$ $\alpha=0.05$

for which $x_1/N_1=x_2/N_2$. If the observations are such that $x_1+x_2=n=12$, say, then, regarded as a one-dimensional problem, we shall only be concerned with the 13 points lying along the diagonal line PQ shown. The probabilities associated with these points are terms in the binomial, $(0.25+0.75)^{12}$ and significance will be judged by the appropriate sum of the tail terms of that binomial.

On the other hand, if we regard the problem as a two-dimensional one, we must

(a) provide some method of determining significance contours of the type illustrated running across the whole diagram, one above and one below the line $x_1 = 3x_2$,

(b) associate with each of these the chance or an upper limit of the chance that the sample point (x_1, x_2) will fall on or beyond a contour when the null hypothesis is true.

Przyborowski and Wilenski (1939), for the case $N_1 = N_2$ and Barnard (Unpublished paper, 1944), in the more general case have determined the contours for a nominal significance level α by cutting off on each diagonal n the points corresponding to the terms in the tail of the binomial $[(1-\rho_0)+\rho_0]^n$ such that their sum is less than $\alpha/2$, but as nearly equal to it as possible. If we denote these tail sums by t_n and t'_n and ω_n their sum and the critical region formed by combining the ω_n for all n by ω, then we have the size of the critical region

$$P[E \in \omega | \rho_0] = \sum_{n=0}^{\infty} [p(n|\mu).(t_n + t'_n)] \leqslant \alpha. \qquad \dots \quad (4)$$

It will be seen that the actual or true significance level, namely $P[E \in \omega | \rho_0]$ associated with a pair of contours will depend on (i) the tail sums t_n and t'_n, which are considerably less than $\alpha/2$, when n is small, and (ii) the unknown $\mu(=m_1+m_2)$.

To illustrate, consider the case $\rho_0 = 0.6$(i.e.$N_1 = 3N_2/2$), $\alpha = 0.05$. As described above, the tail sums are calculated for different values of n and shown in Table 1.

TABLE 1. TAIL SUMS $(t_n + t'_n)$ OF THE BINOMIAL $(0.4+0.6)^n$

n	$t_n+t'_n$	n	$t_n+t'_n$	n	$t_n+t'_n$
4	0	11	0.0095	18	0.0285
5	0.0102	12	0.0340	19	0.0346
6	0.0041	13	0.0204	20	0.0370
7	0.0188	14	0.0256	21	0.0233
8	0.0253	15	0.0145	22	0.0201
9	0.0130	16	0.0374	23	0.0318
10	0.0183	17	0.0229	24	0.0351

This table shows that the tail sums fall much below 0.05 when n has low values, due to the fact that the corresponding binomial has few terms. The probability mass of ω for different values of μ is given in Table 2.

TABLE 2. VALUES OF $P\{E\epsilon\omega|\rho_0=0.0\}$

$\alpha=0.05$

μ	prob.	μ	prob.
2	0.0005	8	0.0151
3	0.0019	9	0.0172
4	0.0042	10	0.0187
5	0.0071	15	0.0254
6	0.0091	20	0.0291
7	0.0128		

These figures which are the true significance levels are considerably below the nominal level. Although the two levels tend to come closer as μ increases, it is only the range of μ that is relevant in the problem we are considering. Hence by applying the test of Przyborowski and Wilenski. we shall be rejecting the hypothesis of equality of proportions actually at one or two percent level although, nominally, at five percent.

We will obtain in the next section a method of forming significance contours which (a) will bring the true $P\{E\,\epsilon\omega|\rho_0\}$ nearer to the nominal level than is possible by the P. W. (Przyborowski and Wilenski) method, (b) will make this value less dependent on the unknown μ and (c) will be associated with a simple rule that could easily be applied for testing significance, using existing tables.

The question of bias in this test for equal proportions will be considered in section 3 and a general method of constructing unbiased tests of a special type when the distributions involved are discrete will be developed.

2. A TEST FOR EQUALITY OF PROPORTIONS WITH PROPER CONTROL OF THE FIRST KIND OF ERROR

2.1. *Formation of Significance Contours*

In the two dimensional lattice of points (x_1, x_2), consider the diagonal line $x_1+x_2=n$. We shall find two suitable points on it with x_1 co-ordinates equal to k and $k'(k>k')$, through which the two significance contours pass. The sum of the conditional probabilities, $p(x_1|n,\rho)$ at the points $x_1=k, k+1,\ldots, n$, is from (2), the tail sum of a binomial expansion,

$$t_n = \sum_{x_1=k}^{n} [n!/(x_1!(n-x_1)!).\, \rho^{x_1}(1-\rho)^{n-x_1}]$$

which can be expressed as a Beta-Function ratio

$$t_n = I_\rho\ (k, n-k+1). \qquad \ldots\ (5)$$

Similarly the total probability at the points $x_1 = 0, 1, \ldots, k'$ on the same diagonal is given by

$$t'_n = I_{1-\rho}(n-k', k'+1). \qquad \ldots (6)$$

Now suppose we choose k so as to satisfy the conditions

$$\left.\begin{aligned} & I_\rho(k+\tfrac{1}{2}, n-k+\tfrac{1}{2}) < \frac{\alpha}{2}, \\ \text{and} \quad & I_\rho(k-\tfrac{1}{2}, n-k+\tfrac{3}{2}) \geqslant \frac{\alpha}{2}. \end{aligned}\right\} \qquad \ldots (7)$$

and the integer k' to satisfy the conditions

$$\left.\begin{aligned} & I_{1-\rho}(n-k'+\tfrac{1}{2}, k'+\tfrac{1}{2}) < \frac{\alpha}{2}, \\ \text{and} \quad & I_{1-\rho}(n-k'-\tfrac{1}{2}, k'+\tfrac{3}{2}) \geqslant \frac{\alpha}{2}. \end{aligned}\right\} \qquad \ldots (8)$$

By this procedure, the tail sums (5) and (6) will, for some values of n, exceed $\alpha/2$ and the total $(t_n+t'_n)$ may then exceed α. However, the probability mass of the critical region beyond the contours will be close to α. Tables 3 and 4 illustrate this for the case considered before, viz. $\rho_0 = 0.6$ and $\alpha = 0.05$.

TABLE 3. TAIL SUMS $(t_n + t'_n)$ OF THE BINOMIAL $(0.4+0.6)^n$ BY THE NEW METHOD ($\alpha = 0.05$)

n	$t_n+t'_n$	n	$t_n+t'_n$	n	$t_n+t'_n$
4	0.0236	11	0.0505	18	0.0531
5	0.0880	12	0.0349	19	0.0582
6	0.0877	13	0.0447	20	0.0370
7	0.0408	14	0.0573	21	0.0722
8	0.0066	15	0.0000	22	0.0483
9	0.0351	16	0.0374	23	0.0530
10	0.0587	17	0.0471	24	0.0617

TABLE 4. $P[E\ \omega \mid \rho_0 = 0.6]$ BY THE NEW METHOD ($\alpha = 0.05$)

μ	prob.	μ	prob.
2	0.0183	8	0.0559
3	0.0336	9	0.0565
4	0.0501	10	0.0539
5	0.0540	15	0.0507
6	0.0578	20	0.0524
7	0.0583		

Comparison of Table 4 with Table 2 shows the considerable improvement effected by the new method of choosing the significance contours. The sizes of the critical regions obtained by the present and the P.W. method are compared for $\rho_0 = 0.6$ and $\alpha = 0.10, 0.01$ and for $\rho_0 = 0.5, 0.75$ and $\alpha = 0.10, 0.01$ in Table 5.

TABLE 5. VALUES OF TRUE SIGNIFCANCE LEVELS CORRESPONDING TO NOMINAL LEVELS α UNDER THE NEW AND P.W. METHODS

(a) $\rho_0=0.6$ $(N_1=1.5\,N_2)$

μ	$\alpha=0.10$ new	$\alpha=0.10$ P.W.	$\alpha=0.01$ new	$\alpha=0.01$ P.W.	μ
2	0.0214	0.0040	0.0004	0.0001	2
3	.0460	.0115	.0016	.0003	3
4	.1153	.0214	.0032	.0006	4
5	.1210	.0328	.0052	.0010	5
6	.1219	.0395	.0073	.0015	6
7	.1198	.0451	.0092	.0020	7
8	.1162	.0484	.0107	.0022	8
9	.1119	.0501	.0117	.0026	9
10	.1079	.0501	.0122	.0027	10
15	.0970	.0535	.0115	.0037	15
20	.0077	.0556	.0106	.0045	20

(b) $\rho_0=0.5$ AND 0.75, $(N_1=N_2$ AND $3N_2)$

μ	$\rho_0=0.5$, $\alpha=0.05$ new	P.W.	$\rho_0=0.5$, $\alpha=0.01$ new	P.W.	$\rho_0=0.75$, $\alpha=0.05$ new	P.W.	$\rho_0=0.75$, $\alpha=0.01$ new	P.W.	μ
2	0.0140	0.0004	0.0004	0.0000	0.0250	0.0038	0.0038	0.0005	2
3	.0200	.0020	.0020	.0001	.0305	.0063	.0063	.0011	3
4	.0413	.0051	.0046	.0003	.0323	.0070	.0078	.0021	4
5	.0463	.0088	.0074	.0007	.0346	.0088	.0072	.0024	5
6	.0474	.0124	.0096	.0013	.0385	.0090	.0066	.0025	6
7	.0470	.0147	.0111	.0019	.0428	.0105	.0084	.0025	7
8	.0408	.0182	.0110	.0024	.0404	.0118	.0083	.0025	8
9	.0473	.0202	.0124	.0028	.0539	.0136	.0683	.0025	9
10	.0480	.0218	.0127	.0031	.0564	.0156	.0087	.0024	10
15	.0532	.0272	.0133	.0042	.0537	.0245	.0121	.0029	15
20	.0536	.0293	.0110	.0048	.0503	.0266	.0117	.0041	20

These tables indicate the following :

(a) The true chance associated with the present critical region is nearer to the nominal significance level α than in the P.W. test.

(b) This chance is fairly stable for all values of the unknown μ, except when it is as small as 2 or 3.

(c) The discrepancy between the true and nominal levels is less when α is high.

(d) For some values of μ the true chance slightly exceeds α. Thus, the critical boundaries passing through the points $x_1 = k_n$ and k'_n on each diagonal n, which were chosen in the manner described, would enable us to form a satisfactory test of the composite hypothesis that the two sample proportions are equal. The critical boundaries for the case $N_1 = 3N_2$ are shown in Fig. 1.

When we are concerned with the one-sided test of whether one proportion is greater than the other, we take only the lower or the upper part of the critical region of the above two-sided test and associate with it the appropriate level. The true chance of a sample point falling in either of the two parts will be close to the nominal chance, except when ρ_0 is very different from 0.5 and μ is small.

2.2. *Formulation of the Test*

If the point (x_1, x_2) falls on or below the lower contour (see Fig. 1), then $x_1 \geqslant k$ and hence

$$I_\rho(x_1+\tfrac{1}{2}, x_2+\tfrac{1}{2}) < I_\rho(k+\tfrac{1}{2}, n-k+\tfrac{1}{2}) < \alpha/2, \text{ from (7)}.$$

Similarly, if the point falls above the upper contour

$$I_{1-\rho}(x_2+\tfrac{1}{2}, x_1+\tfrac{1}{2}) < I_{1-\rho}(n-k'+\tfrac{1}{2}, k'+\tfrac{1}{2}) < \alpha/2, \text{ from (8)}$$

that is,

$$I_\rho(x_1+\tfrac{1}{2}, x_2+\tfrac{1}{2}) \geqslant 1-\alpha/2.$$

Hence to test the hypothesis of equal proportions find $\rho = \dfrac{N_1}{N_1+N_2}$ and read the value of $I_\rho(x_1+\tfrac{1}{2}, x_2+\tfrac{1}{2})$ from the *Incomplete Beta Function Tables* (1934). If it is $\alpha/2$ or $\geqslant 1-\alpha/2$, reject the hypothesis.

An alternative procedure using tables of percentage points of the Beta distribution (Thompson, 1941), is to find r satisfying $I_r(x_1+\tfrac{1}{2}, x_2+\tfrac{1}{2}) = \alpha$, when $\dfrac{x_1}{N_1} > \dfrac{x_2}{N_2}$ and reject the hypothesis if $\rho < r$. When $\dfrac{x_1}{N_1} < \dfrac{x_2}{N_2}$ interchange x_1, x_2 and N_1, N_2 and apply the above rule. In the notation of Thompson's tables $\nu_1 = 2x_2+1$, $\nu_2 = 2x_1+1$ and $x = r$. These tables cover $\nu_2 = 1(1)\ 30, 40, 60, 120, \infty$ and $\nu_1 = 1(1)\ 10, 12, 15, 20, 24, 30, 40, 60, 120, \infty$. We can read r directly for observed x_1 and x_2 values up to 14 and 4 respectively. For higher values it is, in general, possible to get a decision by inspection of the tables.

In the one-sided case, where admissible alternatives are that $p_1 > p_2$, we take $I_\rho(x_1+\frac{1}{2}, x_2+\frac{1}{2})$ and see if it is $\leqslant \alpha$. Alternatively, we compare ρ with the 100α point of the Beta distribution with parameters $x_1+\frac{1}{2}$ and $x_2+\frac{1}{2}$.

Illustration 1: Two samples of 150 and 100 machine parts produced on two types of automatic machines contain respectively 12 and 2 defectives. Is there any significant difference between the two machines?

Here, $\rho_0=0.6$, $\nu_1=5$, $\nu_2=25$. Choosing $\alpha/2=0.025$, we find from Thompson's tables that the 2.5 per cent point is 0.615. This being greater than 0.6, we establish significance at the 5 per cent level.

Illustration 2: The following data relate to the numberof first cousin marriages amongst the parents of mentally defective patients at the National Hospital, London and of the normal inpatients of University College Hospital, London. (Julia Bell, *Annals of Eugenics*, 10, 1940).

source of data	no. of parents	first cousins amongst parents	percentage of first cousins
National Hospital (mental defective) ..	1280 (N_1)	24 (x_1)	1.875
University College Hospital (normal) ..	1257 (N_1)	11 (x_2)	0.875

Is the proportion of first cousin marriages amongst parents of mental defectives higher than that amongst parents of normal persons?

Here $\rho_0=0.505$. Entering Thompson's tables with $\nu_1=23$, $\nu_2=49$, we see by inspection that the 5 per cent point is greater than 0.505 and so the result is significant at that level.

3. CONSTRUCTION OF AN 'UNBIASED' TEST FOR EQUALITY OF PROPORTIONS

3.1. *Bias in Tests with Discrete Criteria*

We shall first consider the general case of tests based on discrete distributions and then derive an unbiased test for the equality of two proportions when they are small. Suppose $p(x|\theta)$ is the probability function of x, where x is a discrete variate taking values 0, 1, 2,...,k and θ is a continuous parameter. If it is desired to test the hypothesis $\theta=\theta_0$ on the basis of a sample of observations X: $x_1, x_2, \ldots, x_n$ controlling the first kind of error at a level α, then we have to find a region in the lattice of points X such that

$$\left.\begin{aligned} &\Sigma p(X|\theta_0) < \alpha, \\ &\sum_{w+\Delta w} p(X|\theta_0) > \alpha, \end{aligned}\right\} \qquad \ldots \quad (9)$$

where $\Delta\omega$ is the smallest possible addition to ω. The region ω will be unbiased in the Neyman-Pearson sense if for any $\theta \neq \theta_0$,

$$\sum_{\omega} p(X|\theta) > \sum_{\omega} p(X|\theta_0). \qquad \dots \quad (10)$$

Owing to the discrete character of X, the condition (9) and (10) are rarely satisfied together; that is, it is not generally possible to construct an unbiased test. A different approach is suggested here giving an extension of the idea of unbiasedness. We shall also extend the concept of a locally most powerful test.

3.2. *Locally Powerful Unbiased Tests*

First suppose it is possible to have unbiased tests satisfying (9) and (10). Although the first kind of error is generally less than α (the nominal level), and is equal to α_1 say, the power with respect to any alternative $\theta > \theta_0$ or $< \theta_0$ is greater than α_1. Amongst the class of such tests we can choose the one that has, in the close neighbourhood of θ_0, the greatest power relative to the first kind of error, that is, the smallest ratio of the second and first kinds of errors.

Suppose ω is a region in the X-space for which we have, using an obvious notation,

$$\sum_{\omega} p_0 < \alpha, \text{ say} = \alpha_1 = c\alpha, \ (c < 1) \qquad \dots \quad (11)$$

and

$$\left[\frac{\partial}{\partial\theta} \sum_{\omega} p_0\right]_{\theta=\theta_0} = \sum_{\omega} p_0' = 0, \qquad \dots \quad (12)$$

assuming that the boundary of ω is independent of θ. Let ω_1 be any other region satisfying (12) whose size is greater than that of ω when $\theta = \theta_0$ that is,

$$\sum_{\omega_1} p_0 = c_1\alpha, \quad \text{where} \quad 1 > c_1 > c \qquad \dots \quad (13)$$

and

$$\sum_{\omega_1} p_0' = 0. \qquad \dots \quad (14)$$

We shall compare ω with ω_1. We do not here consider regions whose sizes are smaller than that of ω, for then the true significance levels are more divergent from α than the level associated with ω.

We will define the region ω as locally better than ω_1 if

$$\frac{\sum\limits_{\omega} p_0''}{\sum\limits_{\omega} p_0} > \frac{\sum\limits_{\omega_1} p_0''}{\sum\limits_{\omega_1} p_0}. \qquad \dots \quad (15)$$

This condition can be written as

$$\left[\frac{\partial^2}{\partial\theta^2}\cdot\frac{1}{c}\,\beta(\omega)\right]_{\theta=\theta_0} > \left[\frac{\partial^2}{\partial\theta^2}\,\frac{1}{c_1}\,\beta(\omega_1)\right]_{\theta=\theta_0}$$

where $\beta(\omega)$ and $\beta(\omega_1)$ denote the power functions of the tests based on ω and ω_1. This means that if the two power curves Γ and Γ_1 (see Fig. 2) are moved up by expand-

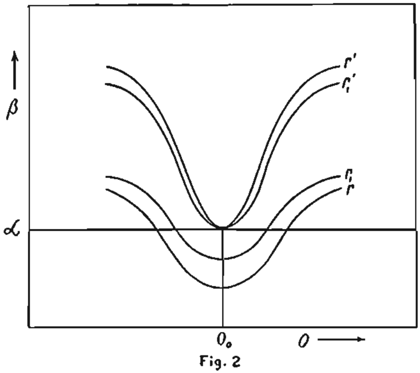

Fig. 2

ing the β-scale by $1/c$ and $1/c_1$ respectively, so that they touch the line $\beta=\alpha$ at $\theta=\theta_0$, the tangent to Γ' turns faster than the tangent to Γ'_1 at the point of contact. This condition therefore is a generalisation of the Neyman and Pearson condition of maximum local power in the case $c = c_1$.

It can be seen that a region ω satisfying (12) and (15) is such that within it,

$$\left.\begin{aligned} &p_0''-l_1p_0'-l_2p_0 \geqslant 0 \\ \text{and outside it,}\quad &p_0''-l_1p_0'-l_2p_0 < 0, \end{aligned}\right\} \qquad \dots \quad (16)$$

where l_1 and l_2 are suitably chosen. For, if $\omega\omega_1$ is the region common to ω and ω_1,

$$\sum_{\omega-\omega\omega_1} \frac{1}{c}\,(p_0''-l_1p_0'-l_2p_0) - \sum_{\omega_1-\omega\omega_1} \frac{1}{c_1}(p_0''-l_1p_0'-l_2p_0) \geqslant 0,$$

from (16). Adding to both sides

$$\sum_{\omega\omega_1} \left(\frac{1}{c}-\frac{1}{c_1}\right)(p_0''-l_1p_0'-l_2p_0)$$

and transferring the terms in p_0 and p'_0 to the right side, we get

$$\sum_{\omega} \frac{1}{c}\,p_0 - \sum_{\omega_1} \frac{1}{c_1}\,p_0'' \geqslant l_1\Big[\sum_{\omega}\frac{1}{c}\,p_0' - \sum_{\omega_1}\frac{1}{c_1}\,p_0'\Big] +$$
$$+\,l_2\Big[\sum_{\omega}\frac{1}{c}\,p_0 - \sum_{\omega_1}\frac{1}{c_1}\,p_0\Big] +$$
$$+\sum_{\omega\omega_1}\left(\frac{1}{c}-\frac{1}{c_1}\right)\left(p_0''-l_1\,p_0'-l_2p_0''\right).$$

The first and second bracketed expressions on the right vanish in virtue of conditions (11) to (14). Since $\frac{1}{c}-\frac{1}{c_1}>0$ and $p_0''-l_1p_0'-l_2p_0 \geqslant 0$ in $\omega\omega_1$, the last expression is $\geqslant 0$. Thus (15) is satisfied.

The constants l_1 and l_2 should satisfy the condition (12) of unbiasedness and the requirement that α_1 in (11) should be a subject to the condition $\alpha_1 \leqslant \alpha$. maximum. It may be remarked here that a region satisfying (16) and having a small value of α_1 would be locally better in the sense of inequality (15) than another region also satisfying (16) but having a larger α_1. We would, however, prefer the latter as the true level is closer to α. Thus we take as our locally best critical region that member of this class for which α_1 is closest to α.

Let us apply the above method to construct the best test of the hypothesis of equality of proportions, considered in section 1 when the samples are equal, i.e. $N_1=N_2$ or $\rho=\frac{1}{2}$. The conditional distribution $p(x_1|n)$ in (2) is now a symmetric binomial and we may construct unbiased critical regions of size $\leqslant \alpha$ by taking groups of terms in the expansion of $(\frac{1}{2}+\frac{1}{2})^n$ symmetrically placed about its centre. Amongst the class of such regions, the best one is that for which the requirements of (16) are satisfied.

Now
$$p=\frac{n!}{x_1!\,(n-x_1)!}\,\rho^{x_1}(1-\rho)^{n-x_1}.$$

It can be easily seen that the condition $[p''-l_1p'-l_2p]_{\rho=\frac{1}{2}} \geqslant 0$ is equivalent to

$$16x_1^2-4x_1(4n+l_1)+4n^2-2n(2-l_1)-l_2 \geqslant 0$$

which yields $x_1 \geqslant k$ and $\leqslant k'$. This shows that the best region consists of the tails of the binomial distribution $\frac{n!}{x_1!(n-x_1)!}\left(\frac{1}{2}\right)^n$. The condition of unbiasedness (14) require that the tails should be equal, i.e. $k'=n-k$. Then the integer k has to be chosen such that the size of each tail is as near to $\alpha/2$ as possible.

3.3. *Uniformly More Powerful Unbiased Tests*

As in the Type A_1 region in the continuous case, we can conceive of a critical region ω satisfying conditions (11) and (12) and in addition, a third condition,

$$\sum_{\omega} p(\theta) \geqslant \sum_{\omega_1} p(\theta)$$

for all θ where ω_1 is any other region satisfying (12) and $\sum_{\omega_1} p_0(\theta) = c_1\alpha$, $c_1 \leqslant c \leqslant 1$. This ensures that the power curve associated with ω lies always above that for ω_1. It can be seen easily that such a region is defined by

$$p - lp_0' \geqslant 0 \quad \text{within it,}$$

and

$$p - lp_0' < 0 \quad \text{outside it.}$$

The constant l must satisfy the requirement that $\sum_{\omega} p_0$ is as close to α as possible.

Applying this to the present test for $\rho = \frac{1}{2}$, within ω we should have

$$\frac{n!}{x_1!(n-x_1)!}\rho^{x_1}(1-\rho)^{n-x_1} - l\frac{n!}{x_1!(n-x_1)!}\left(\frac{1}{2}\right)^n(4x_1-2n) \geqslant 0$$

that is,

$$(2\rho)^{x_1}(2-2\rho)^{n-x_1} - 2l(x_1+\overline{x_1-n}) \geqslant 0.$$

This will yield $x_1 \geqslant k$ and $\leqslant n-k$. The integer k is found by making $P\{x_1 \geqslant k\}$ as near to $\frac{\alpha}{2}$ as possible.

This is the same region as derived earlier from consideration of maximum power in the neighbourhood of $\rho = \frac{1}{2}$. Thus we see that the unbiased test which is locally most powerful is uniformly more powerful than any other test in the whole range of ρ.

3.4. *Tests with Minimal Bias*

Next let us consider the general case where an unbiased test in the Neyman and Pearson sense is not possible. The power curve while passing through a point $\beta = \alpha_1 \leqslant \alpha$ at $\theta = \theta_0$ dips below α_1 when $\theta < \theta_0$ or $> \theta_0$. In such a situation we shall replace the condition $\beta'(\theta_0) = 0$ by another which requires the interval (θ_1, θ_2) between which the power curve dips below the α-line to be the shortest. Since θ_1 and θ_2 lie on either side of θ_0, we can interpret this interval as the range of alternative hypotheses which cannot be detected with a power exceeding α, the level at which the first kind of error is nominally controlled. It may also be noted that the same critical region for testing for $\theta = \theta_0$ at the level α will serve for testing any one of the hypotheses, $\theta_2 \leqslant \theta \leqslant \theta_1$ at the same level. It appears therefore reasonable to minimise the range $\theta_1 - \theta_2$ by a proper choice of the critical region. A test with this property will be said to have minimal bias.

Let ω be such that

$$\sum_{\omega} p_0 \leqslant \alpha, \text{ say} = \alpha_1 = c\alpha \ (c \leqslant 1) \quad \dots \quad (17)$$

and ω_1 be any other region for which

$$\sum_{\omega_1} p_0 = c_1\alpha \quad \text{when} \quad 1 \geqslant c_1 \geqslant c. \quad \dots \quad (18)$$

As in the case of unbiased regions discussed earlier, we do not consider here alternative regions where size is $<\alpha$ for they would make the true level even smaller. We require that

$$\frac{\sum_{\omega} p_0''}{\sum_{\omega} p_0} \geqslant \frac{\sum_{\omega_1} p_0''}{\sum_{\omega_1} p_0}. \quad \dots \quad (19)$$

Amongst the class of regions ω satisfying (18) and (19), we define that one as the best for which $(\theta_1-\theta_2)$ is a minimum, θ_1, θ_2 being the values of θ satisfying

$$\sum_{\omega} p(\theta) = \alpha. \quad \dots \quad (20)$$

The condition (19) which is the same as

$$\left[\frac{\partial^2}{\partial\theta^2}\frac{1}{c}\beta(\omega)\right]_{\theta=\theta_0} \geqslant \left[\frac{\partial^2}{\partial\theta^2}\frac{1}{c_1}\beta(\omega_1)\right]_{\theta=\theta_0}$$

means that if the two power curves associated with ω and ω_1 are respectively moved up by increasing the β-scale so as to intersect the line $\beta=\alpha$ at θ_0, then the tangent to the first curve turns faster than that to the second at the point θ_0.

It can be shown as before that the best region ω is that within which

$$\left.\begin{aligned} &p''_0-lp_0 \geqslant 0 \\ \text{and} \quad &p''_0-lp_0 < 0 \quad \text{without.} \end{aligned}\right\} \quad \dots \quad (21)$$

The constant l is to be determined from the third condition that the length $(\theta_1-\theta_2)$ is the shortest subject to (17). That is, amongst the regions defined by (21), we choose that one whose size is $\leqslant\alpha$ and which minimises $(\theta_1-\theta_2)$.

It may be noticed that when best unbiased regions satisfying (12) and (15) exist and the power curve is symmetrical about $\theta=\theta_0$, the region which has the largest size $\leqslant\alpha$ is also the region for which $\theta_1-\theta_2$ or $\theta_1-\theta_0=\frac{1}{2}(\theta_1-\theta_2)$ is the smallest. Hence this comes as a special case of the general type of best regions discussed above, when the additional condition (12) is fulfilled.

Returning to the test for equality of proportions when $N_1 \neq N_2$, i.e. $\rho \neq \frac{1}{2}$, the best region belongs to the class defined by (21). So, within ω

$$\frac{\partial^2}{\partial\theta^2} p(x_1|n, \rho_0)-lp(x_1|n, \rho_0) \geqslant 0$$

which simplifies to

$$x_1^2-x_1(2n\rho_0-2\rho_0+1)+n(n-1)\rho_0^2-l \geqslant 0$$

i.e. $x_1 \geqslant k$ and $< k'$. Hence the best region is provided by the tails of the binomial series $(\overline{1-\rho_0}+\rho_0)^n$. The integers k' and k are determined from the conditions that the tail sum $< \alpha$ and that $(\rho_1-\rho_2)$ should be a minimum, ρ_1, ρ_2 being the values of ρ satisfying

$$\sum_0^{k'} + \sum_k^n \left[\frac{n!}{x_1!(n-x_1)!} \rho^{x_1} (1-\rho)^{n-x_1} \right] = \alpha.$$

This equation can be written in terms of the Incomplete Beta Functions,

$$I_\rho(k,\, n-k+1) + I_{1-\rho}(n-k',\, k'+1) = \alpha. \quad \ldots \quad (22)$$

We also have

$$I_{\rho_0}(k,\, n-k+1) + I_{1-\rho_0}(n-k',\, k'+1) < \alpha. \quad \ldots \quad (23)$$

So, to find the best k and k' for a given n, we take several pairs of k and k' for which (23) is not violated and find from (22) corresponding sets of ρ_1 and ρ_2 and take the best pair for which $\rho_1-\rho_2$ is least. In practice, this procedure is not laborious. For, we need take a few pairs (k', k) which make each tail nearly $\alpha/2$ and obtain approximately ρ_1, ρ_2 by inspection of (22) with the help of Incomplete Beta Function Tables.

Below are given values of k', k and ρ_1, ρ_2 for a few values of n when $\rho_0 = 0.6$, $\alpha = 0.10$. The best critical limits are indicated by an asterisk.

$\rho_0 = 0.6$; $\alpha = 0.10$						
n	k'	k	ρ_1	ρ_2	$\rho_1-\rho_2$	α_1
8	1	8	.760	.407	.343	.0253
8	2*	8*	.745	.546	.199	.0606
12	3*	10*	.601	.489	.112	.0987
12	3	11	.711	.477	.234	.0349
12	4	11	.706	.569	.137	.0709
16	5*	13	.619	.514	.105	.0843
16	5	14	.699	.505	.194	.0375
16	6	14	.694	.579	.115	.0767

In the above table, α_1 the actual size of the critical region, $x_1 < k'$ and $\geqslant k$ is also given. It will be seen that when $(\rho_1-\rho_2)$ is least, the difference between the nominal and true levels, i.e.$(\alpha-\alpha_1)$ is also smallest. This property is not likely to be general for all discrete distributions.

The test obtained now differs from that formulated in section 2. Our object there was to secure a practically constant size for the critical region in the x_1, x_2 space. For some values of μ (unknown) it might even exceed α slightly.

Following the method described above, values of k' and k are given for $\rho = 0.5, 0.6, 0.6667, 0.75$ (i.e. $\frac{N_1}{N_2} = 1, \frac{3}{2}, 2, 3$), $\alpha = 0.05, 0.10$ and $n = 1$ to 25 in Tables 6(a) and 6(b). By interchanging N_1, N_2 and x_1, x_2 these table can also be used for $\rho = 0.4, 0.3333, 0.25$.

3.5. *The Over-all Test*

We have defined the best critical region for testing $\rho=\rho_0$ assuming the marginal total n to be fixed. In the notation of section 1, this region is ω_n. The aggregate of ω_n over n gives ω, the critical region for the test when the sample space is in two dimensions. This region ω will have the following properties :

(i) $\sum_{\omega} p_0(x_1, n) = \sum_{n=1}^{\infty} \Big[\sum_{\omega_n} p_0(x|n) \Big] p(n) < \alpha$, from (7),

(ii) $\sum_{n=1}^{\infty} \dfrac{\sum_{\omega_n} p^*_0(x_1|n)}{\sum_{n} p_0(x_1|n)} \cdot p(n)$ is a maximum, from (10)

(iii) $\sum_{n=1}^{\infty} (\rho_{1n}-\rho_{2n})p(n)$ is a minimum, since $(\rho_{1n}-\rho_{2n})$ is a minimum for ω_n.

In the particular case of testing for $\rho = \frac{1}{2}$, if the conditional test is based on ω_n which satisfies (11), (12) and (15) then the over-all test is also unbiased and locally most powerful in the same sense.

The significance contours for the over-all test pass through the points k' and k on each diagonal n. The procedure for applying the test with minimal bias consists in comparing x_1 (the number in the larger sample) with k' and k for observed $n(=x_1+x_2)$ from tables 6(a) and 6(b) with the appropriate $\rho\left(=\dfrac{N_1}{N_1+N_2}\right)$.

TABLE 6(a). VALUES OF k' AND k, THE CRITICAL POINTS OF x_1 IN THE MINIMAL BIAS TEST FOR EQUALITY OF PROPORTIONS

($\alpha = 0.10$)

	$\rho=0.5$ ($N_1=N_2$)		$\rho=0.6$ ($N_1=1.5N_2$)		$\rho=0.6667$ ($N_1=2N_2$)		$\rho=0.75$ ($N_1=3N_2$)	
n	k'	k	k'	k	k'	k	k'	k
6	0	6	1	6	1	6	2	–
7	0	7	1	7	1	7	3	–
8	1	7	2	8	2	8	3	–
9	1	8	2	8	3	9	3	9
10	1	9	2	9	4	10	4	10
11	2	9	3	10	3	10	5	11
12	2	10	3	10	4	11	6	12
13	3	10	4	11	5	12	6	13
14	3	11	4	12	6	13	7	14
15	3	12	5	13	6	14	7	14
16	4	12	5	13	6	14	8	15
17	4	13	6	14	7	15	9	16
18	5	13	7	15	8	16	10	17
19	5	14	7	16	8	17	10	18
20	5	15	7	16	9	17	11	19
21	6	15	8	17	9	18	11	19
22	6	16	8	17	10	19	12	20
23	7	16	9	18	10	19	13	21
24	7	17	10	19	11	20	14	22
25	7	18	10	20	12	21	15	23

TABLE 6(b). VALUES OF k' AND k, THE CRITICAL POINTS OF x_1 IN THE MINIMAL BIAS TEST FOR EQUALITY OF PROPORTIONS

($\alpha = 0.05$)

	$\rho=0.05$ ($N_1=N_2$)		$\rho=0.06$ ($N_1=1.5\,N_2$)		$\rho=0.6007$ ($N_1=2N_2$)		$\rho=0.75$ ($N_1=3N_2$)	
n	k'	k	k'	k	k'	k	k'	k
6	0	6	1	–	1	–	2	–
7	0	7	1	7	2	–	2	–
8	0	8	1	8	1	8	3	–
9	1	8	2	9	2	9	4	–
10	1	9	2	10	3	10	4	–
11	1	10	2	10	3	11	4	11
12	1	11	3	11	4	12	5	12
13	2	11	4	12	4	12	6	13
14	2	12	4	13	5	13	6	14
15	3	12	4	13	5	14	7	15
16	3	13	5	14	6	15	8	16
17	4	13	6	15	7	16	9	17
18	4	14	6	16	7	16	9	18
19	4	15	6	16	8	17	9	18
20	5	15	7	17	8	18	10	19
21	5	16	7	17	9	19	11	20
22	5	17	8	18	9	19	12	21
23	6	17	8	19	10	20	12	22
24	6	18	9	20	11	21	13	23
25	7	18	9	20	11	22	13	23

REFERENCES

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

PATNAIK, P. B. (1948): The power function of the test for the difference between two proportions in a 2×2 table. *Biometrika*, **35**, 157.

PEARSON, E. S. (1947): The choice of statistical tests illustrated on the interpretation of data classified in a 2×2 table. *Biometrika*, **34**, 139.

PEARSON, E. S. and NEYMAN, J. (1936): *Statistical Research Memoirs* I, Issued by the Department of Applied Statistics, University of London.

PEARSON, K. (1934): *Tables of the Incomplete Beta Function*, Biometrika Office, London.

PRZYBOROWSKI, J. and WILENSKI, H. (1939): Homogeneity of results in testing samples from Poisson series. *Biometrika*, **31**, 313.

THOMPSON, C. M. (1941): Tables of percentage points of the incomplete beta function. *Biometrika*, **32**, 151.

Paper Received : March, 1954.