A NEW METHOD FOR TESTING TWO BINOMIAL SAMPLES WITH ITS APPLICATION IN STATISTICAL QUALITY CONTROL

by P. V. Krishna Iyer and M. N. Bhattacharyya

Defence Science Laboratory, Ministry of Defence, New Delhi

ABSTRACT

A new method for comparing the proportions observed in two binomial samples involving the characters, say A and B with probabilities p & q has been given in this paper. This method provides the basis for a more reliable comparison than the usual 2×2 χ^2—contingency test. Taking the samples in order of occurrence to be $x_1, x_2 \ldots\ldots x_n$; $y_1, y_2 \ldots\ldots\ldots\ldots\ldots\ldots y_n$, the tests depend on the statistics X_s and Y_s which are based on a system of scoring $\Sigma\,(x_k—y_l)$ or $\Sigma\,|\,x_k—y_l\,|$ where $|\,k—l\,|\leqslant s$, a constant. The score for $(x_k—y_l)$ is zero for combination AA or BB and is +1 or —1 if the combination is AB or BA. Similarly $|\;x_k\;—y_l\;|$ is zero in the first case and +1 in the latter two cases. The tests have been illustrated for examining the percentage of defectives produced in two machines. The applications of the tests in psychological investigations have also been explained with reference to the data discussed by McNemar.

Introduction

The χ^2 test with one degree of freedom is usually used for testing the significance of the difference between the observed proportion or percentage of defectives in two independent samples. The samples are arranged in a 2×2 contingency table and the χ^2 is calculated as follows:

TABLE I

2×2 contingency table

	Defective	Non-Defective	Total
Sample I	f_1 $(e_1=N_1N_1'/N)$	f_2 $(e_2=N_1N_2'/N)$	N_1
Sample II	f_3 $(e_3=N_2N_1'/N)$	f_4 $(e_4=N_2N_2'/N)$	N_2
Total ..	N_1'	N_2'	N

In the above table f represents the observed frequency while e stands for the expected frequency calculated on the assumption that the two samples are homogeneous and the proportions of defectives and non-defectives are N_1'/N and N_2'/N respectively.

Then

$$\chi^2 = \sum_{i=1}^{4} \frac{(f_i - e_i)^2}{e_i} = \frac{(f_1 f_4 - f_2 f_3)^2 N}{N_1 N_2 \, N_1' N_2'} ; \; \text{d.f.}=1$$

The difference between the two proportions is considered to be significant or not according as χ^2 is greater than 3·841, the 5% value (6·635 if 1% level is desired), or less than 3·841.

The two samples can also be compared by calculating the tandardised deviate of the difference between the proportion of defectives,

$$Z = \frac{f_1/N_1 - f_3/N_2}{\sqrt{\frac{N_1' N_2'}{N^2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

If $|Z| > 1{\cdot}96$ the sample proportions are significantly different at 5% level.

The above two tests, namely χ^2 and Z are, however, equivalent for comparing two binomial samples.

It may be noted that these tests make only an overall comparison of the two samples ignoring completely the order of occurrence of the individual observations. Theoretically, these tests are valid only foe random samples. Hence the correct application of the tests calls for prior testing of randomness of the samples which obviously necessitates the maintenance of the order of occurrence of the individuals at the time of collection of the data. From practical point of view also the order of occurrence of the observations and examination of its randomness seem to be of importance, especially in Statistical Quality Control work (S.Q.C.) while comparing two processes or work in two shifts or two machines etc. It is reasonable that a test which takes the order of occurrence also into consideration will provide a more reliable comparison between two samples. A class of such tests developed by us are briefly discussed below.

Definition and Distribution of Statistics X_s and Y_s

Let two binomial samples of size n, arranged in order of occurrence be as follows

Sample I x_1 x_2, ... x^n

Sample II y_1 y_2, ... y_n

where x_r and y_r, the r^{th} observations in samples I and II are either A or B with probabilities p & q respectively for both the samples. Then a class of *statistics* X_s and Y_s which will take into account the order of occurrence of the observations in the two samples, is defined by

$$X_s = \sum_{r=1}^{n} (x_r - y_r) + \sum_{i=1}^{s} \sum_{r=1}^{n-i} (x_r - y_{r+i}) + \sum_{i=1}^{s} \sum_{r=1}^{n-i} (x_{r+i} - y_r)$$

$$Y_s = \sum_{r=1}^{n} \left| x_r - y_r \right| + \sum_{i=1}^{s} \sum_{r=1}^{n-i} \left| x_r - y_{r+i} \right| + \sum_{i=1}^{s} \sum_{r=1}^{n-i} \left| x_{r+i} - y_r \right|$$

$$(s = o, 1, 2, \ldots\ldots\ldots k \,;\, k \leqslant n-1)$$

where $(x_r - y_{r+i})$ is zero, $+1$ and -1 according as (*i*) both x_r & y_{r+i} are either A or B, (*ii*) x_r is A & y_{r+i} is B and (*iii*) *vice versa*, while $|x_r - y_{r+i}|$ is zero in case (*i*) and $+1$ in both the cases (*ii*) & (*iii*). It may be noted that X_o, Y_o, X_1 and Y_1 reduce to

$$X_o = \sum_{r=1}^{n} (x_r - y_r), \quad Y_o = \sum_{r=1}^{n} \left| x_r - y_r \right|,$$

$$X_1 = \sum_{r=1}^{n} (x_r - y_r) + \sum_{r=1}^{n-1} (x_r - y_{r+1}) + \sum_{r=1}^{n-1} (x_{r+1} - y_r),$$

$$Y_1 = \sum_{r=1}^{n} \left| x_r - y_r \right| + \sum_{r=1}^{n-1} \left| x_r - y_{r+1} \right| + \sum_{r=1}^{n-1} \left| x_{r+1} - y_r \right|$$

The first two cumulants of X_o, X_1, Y_o and Y_1 are as follows

Statistic X_0

$$\kappa_1 = 0$$
$$\kappa_2 = 2npq$$

Statistic X_1

$$\kappa_1 = 0$$
$$\kappa_2 = 2(9n-10)pq$$

Statistic Y_o

$$\kappa_1 = 2npq$$
$$\kappa_2 = 2npq\,(1-2pq)$$

Statistic Y_1

$$\kappa_1 = 2(3n-2)\,pq$$
$$\kappa_2 = 6npq\,(3-10\,pq)-4pq\,(5-18pq)$$

The expectation and variance for the general ease when the probabilities for A and B are p_1, q_1 and p_2, q_2 for the first and second samples respectively are given below

$$E(X_s)=W_1(p_1-p_2), \quad E(Y_s)=W_1(p_1q_2+q_1p_2)$$

For $s \leqslant n/2$

$V(X_s)=W_1Z2+W_2Z_2$, $V(Y_s)=W_1Z_3+W_2Z_4$ and

For $s\geqslant n/_1$

$V(X_s)=W_1Z_1+W_3Z_2$, $V(Y_s)=W_1Z_3+W_3Z_4$

where

$$W_1=(n+2ns)-s(s+1),$$

$$W_2=2\{(2ns^2+ns)-\tfrac{1}{3}(5s^3+6s^2+s)\},$$

$$W_3=\tfrac{1}{3}\ (n-1)\,(6ns-n^2+2n)-2s\,(s^2-1)\},$$

$$Z_1=(p_1q_2+q_1p_2)-(p_1-p_2)^2,$$

$$Z_2=\{p_1q_2(p_1+q_2)+q_1p_2(q_1+p_2)-2(p_1-p_2)^2\}$$

$$Z_3=(p_1q_2+q_1p_2)-(p_1q_2+q_1p_2)^2$$

$$Z_4=\left\{p_1q_2(p_1+q_2)+q_1p_2(q_1+p_2)-2(p_1q_2+q_1p_2\right\}$$

It has been shown on the basis of the first four cumulants that the distributions of X_s and Y_s for finite values of p tend to the normal form as n tends to infinity. However, when p is small the distribution of Y_0 tends to the Poisson form.

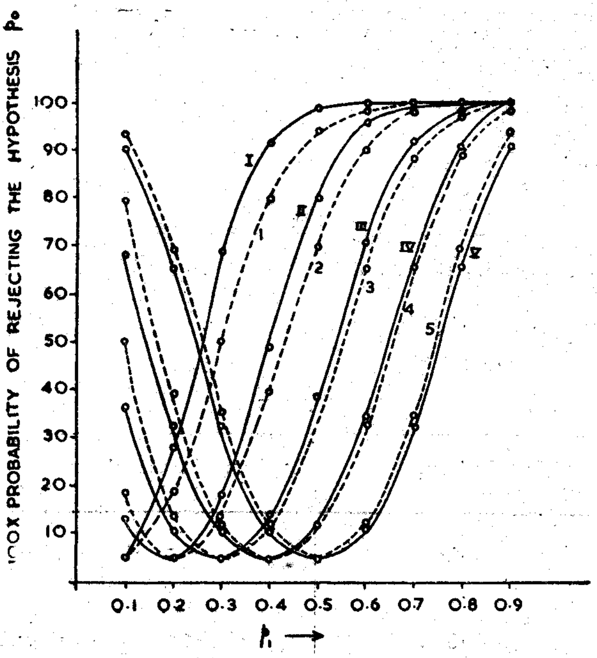
The above statistics X_s and Y_s can be used for testing the difference between two binomial samples. For large values of n, this is done by calculating the standardized deviate of these statistics by assuming p to be the same for both the samples. This value may be taken to be any of the observed proportions for the two samples separately or the proportion obtained by pooling the two samples together. Among these, that value of p which gives the maximum power for the alternatives arising from samples I and II is to be preferred. For small samples the probabilities that $X_s \geqslant a$ or $Y_s \geqslant b$ can be determined from the exact distributions.

Power of the Different Tests

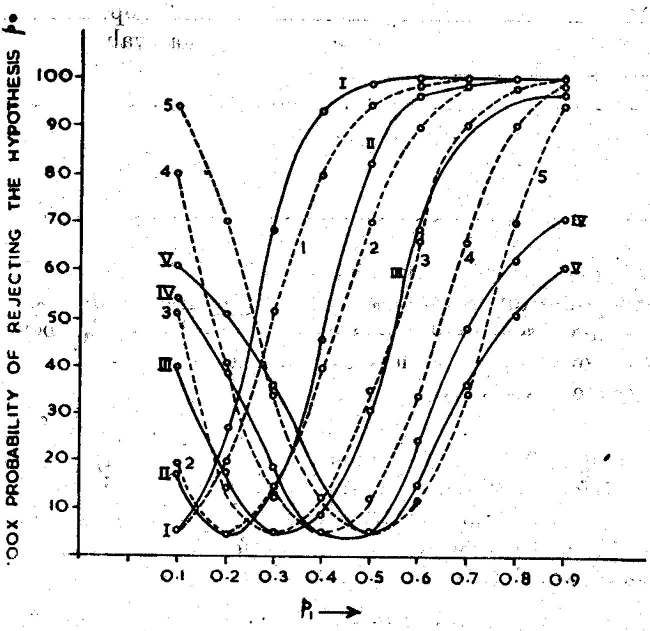
The power of the statistics, *i.e.* the chance that the test shows to be significant, when the samples are not homogeneous, was examined for the hypothesis p_0 and the alternative p_1 varying from 0·1 to 0·9 for n=30 and it was found that the power of the statistics X_s was maximum for X_0. As regards Y_s the power increased rapidly as s increased up to $s={}^n/_2$, beyond which the increase in power was much less and the information on the order included in the test became less and less.

The power curves for the tests based on X_0 and $Y_{n/2}$ for the hypotheses p_0=0·1 (0·1) 0·5 and alternative p_1 = 0·1 (0·1) 0·9 are given in graphs I and II for n=30 when the probability for the rejection of the hypothesis $H_0 = p_0$, when true, is ·05. As the power curves for the hypotheses p_0 = ·9, ·8, ·7 and ·6 are the reverse of those for ·1, ·2, ·3 and ·4 respectively they have not been given.

In fact the standardised deviate of X_0 and χ^2 are algebraically identical. Hence their power curves should be the same. In graph I there is some difference between the power curves for p = ·1 & ·2. This discrepency is due to the fact that the value of n=30 is not big enough for assuming the distribution to be normal.



POWER CURVES FOR THE STATISTIC X_0
GRAPH I



POWER CURVES FOR THE STATISTIC $Y_{n/2}$
GRAPH II

In the above graphs the continuous curves I to V represent the powers of the statistics X_0 and $Y_{n/2}$ for the hypotheses $p_0 = \cdot 1, \cdot 2, \cdot 3, \cdot 4$ and $\cdot 5$ respectively while the dotted curves 1 to 5 show the corresponding powers for the standardized deviate of $(\mathrm{Sin}^{-1}\sqrt{f_1} - \mathrm{Sin}^{-1}\sqrt{f_2})$, where f_1 and f_2 are the observed proportion of defectives in the two samples.

As seen from the above graphs, the standardized deviate of Xo and the contingency χ^2 test (almost equivalent to square of the standardized deviate of $(\mathrm{Sin}^{-1}\sqrt{f_1} - \mathrm{Sin}^{-1}\sqrt{f_2})$ appear to be equally powerful for testing two binomial samples. It may, however, be noted that while the existing test is an overall test ignoring the order of occurrence of the individual observations, the statistic X_0 takes that also into account. For $Y_{n/2}$ the test is slightly more powerful than th: usual χ^2 test for hypothesis $p_0 = q_0 = \cdot 5$ and alternatives near about $p_1 = \cdot 3$ to $\cdot 7$. For other values of p_0, the more powerful regions are as shown below.

TABLE 2

Hypothesis, p_o	Approximate region w_1
·1	$p_1 > \cdot 1$
·2	$p_1 > \cdot 3$
·3	$\cdot 2 < p_1 \leq \cdot 3$ and near about ·6
·4	$\cdot 2 < p_1 < \cdot 4$

Application of the Tests

Example I—Suppose thirty articles are collected from each of two machines according as the order of production. Let the inspection results be as follows:—

Machine I BBABBBABBABBBABBBBBBBBABBABBBB,

Machine II BBBBBBBBBBBBBABAABBBBBBBBBBBBB,

Where A stands for defective and B for non-defective article.

The object is to examine whether the two samples are homogeneous *i.e.* the percentage of defectives in the two machines are same and are produced at random.

The observed proportion of defectives in the first sample is ·2 and that in the second sample is ·1. The hypothesis to be tested in this case is that the two samples are homogeneous. If this be true, either of the sample proportions can be taken to be the estimate of the population proportion p. But the object of the test is to find out the real difference between the two samples, if there be any. In the present case this will be done in a more striking manner if the population value of p is taken to be ·1 instead of ·2 because the power for $p_0 = \cdot 1$ is more than that for $p_0 = \cdot 2$ for all alternatives $(p_1 = \cdot 1$ to $\cdot 9)$. Thus assuming $p = \cdot 1$ and $q = \cdot 9$ and using the formulae (7) a:d (8) the standardized deviates of the *statistics* Y_s for s=0, 1, 5, 10, 15, 20, 25 and 29 and of the *statistic* X_0 along with the usual large sample binomial test are calculated and given in the following table.

TABLE 3

Standardized deviates of the statistics Y_s, X_0 and the binomial test

Value of s	Observed value of Statistic	Expected value	Stand ard deviation	Co-efficient of variation	Standardized deviate Z
Statistic Y_s					
0	7	5·40	2·10	38·88	·76 not significant
1	25	15·84	5·82	36·74	1·57 Do.
5	87	54·00	19·09	35·35	1·72 Do.
10	139	93·60	33·13	35·40	1·37 Do.
15	189	124·20	43·76	35·23	1·48 Do.
20	217	145·80	50·80	34·84	1·40 Do.
25	232	158·40	54·84	34·62	1·34 Do.
29	234	162·00	56·03	34·59	1·28 Do.
Statistic **Xo**	3	0	2·32		1·29 Do.
Usual large sample binomial test					1·08 Do.

It may be noted that Z for Y_s (s=1, 5,..................25) is greater than that for X_0 and the usual large sample binomial test. Z_s is considerably more for values of s between 1 and 15 and is maximum for s=5. This is partly due to the fact that the reduction in the coefficient of variation is sharp up to s=5. So for practical purposes the value of s may be taken to be 5.

None of the tests in the present case has shown any significant difference between the two samples. The best decision that can be made in the present case is to confirm the results by collecting bigger samples.

Example 2—The statistics X_s and Y_s are based on different types of matchings between the individuals of two *independent* samples, on the basis of their order of occurrence. When the two samples are not independent as in the case where every individual of one sample is matched with a corresponding individual of the second sample, the ordinary χ^2 test can not be applied[1]. But the *statistics* X_0 and Y_0 can be applied for comparing the percentages in such matched samples also. An example of such matched samples is provided by McNemar[3]. 205 soldiers were asked before and after a lecture on the difficulties of war against Japan, whether they thought that the war against Japan would last more than or less than a year. Matching occurs because the same soldiers reply before and after the lecture.

The replies in this case may be classified in a 2×2 frequency table as shown in Table 4.

TABLE 4

Comparison of proportions in two samples

		After lecture		Total
		Less	More	
Before Lecture	Less	(*a*) 36	(*b*) 34	70
	More	(*c*) 0	(*d*)135	135
		36	169	205

The problem is to test whether the lecture was effective in convincing the soldiers of the difficulties of war against Japan thereby making a significant change in their opinion.

McNemar's [3] solution is quoted below.

"Before the lecture, 70 men out of 205 thought that the war would last less than a year, whereas after the lecture this number dropped to 36. The comparison which we wish to make is that between the two frequencies 70/205 and 36/205. There are several ways in which the test may be derived. Perhaps the easiest is to note that both numerators, 70 and 36 contain the 36(*a*) men who persisted in thinking that the war would last less than a year. Hence, equality of the numerators would imply that the same number of men changed from 'Less' to 'More' as changed from 'More' to 'Less'. In other words, if the lecture is without effect half the persons who changed their minds to change in one direction and half in the other. Thus the test can be made by testing whether the numbers (*b*) and (*c*) are binomial successes and failures out of n=(b+c) trials, with probability ½. For this

$$X^2 = \frac{(b-\frac{1}{2}n)^2}{\frac{1}{2}n} + \frac{(c-\frac{1}{2}n)^2}{\frac{1}{2}n} = \frac{(b-c)^2}{(b+c)} = 34$$

with 1 degree of freedom. A correction for continuity can be applied by subtracting 1 from the absolute value of the numerator before squaring".

Let us examine the application of the *statistics* X_s and Y_s in the present example. As there is definite matching in the present case, the matched differences $(x_r - y_r)$ or $|x_r - y_r|$ where x_r and y_r are the replies "more" or "less" of the r^{th} soldier before and after lecture, can only be considered. Thus either of the *statistics*. X_0 and Y_0 only is to be used for comparing two samples. Designating the reply "more" by 1 and "less" by 0 the data of Table 4 can be rearranged as in Table 5.

TABLE 5

Replies of Soldiers

Before lecture	After lecture	Number of soldier
1	1	135
0	1	34
0	0	36
1	0	0
		205

The above table shows that the number of soldiers replying (1,1), (0,1) (0,0) and (1,0) are 135, 34, 36 & 0 respectively. Thus the values of X_o and Y_o are —34 and 34 in this case. The proportions of 1 *i.e.* "more" before and after the lecture are 135/205 & 169/205 respectively. Assuming that the proportion of soldiers replying "more" before and after the lecture remained the same and taking that to be 135/205 the standardized deviates of *statistics* X_o and Y_o are calculated and given below

$$Z = \frac{X_o}{\sqrt{2npq}} = -3\cdot54$$

$$\text{and } Z = \frac{Y_o - 2npq}{\sqrt{n[2pq-4p^2q^2]}} = -8\cdot17$$

These values are significant at 1 per cent level indicating thereby that the proportions of soldiers replying "more" after the lecture changed significantly. It is interesting to note the interpretations of *statistics* X_o and Y_o in the present case. The *statistics* X_o represents the difference between the number of soldiers who changed their opinion in one direction, say from "more" to "less" and that of those changing in the other direction *i.e.* from "less" to "more", while Y_o represents the total number of soldiers who changed their opinion after the lecture.

References

1. Cochran, W.G. (1950). "The comparison of percentages in matched samples", *Biometrika, 37,* 256.

2. Krishna Iyer, P.V. and Bhattacharyya, M.N. (1956), " On some *statistics* comparing two binomial sequences". (Sent to the Jour. Ind. Soc. Agric. Stat.)

3. McNemar, Q. (1949), Psychological Statistics. New York, John Wiley and Sons.