A TEST FOR THE EFFECT OF TREATMENT OR TIME ON A BINOMIAL POPULATION

Haskel Benishay, State University of New York at Buffalo

I. Introduction

The influence exerted on the parameter of a binomial population by a change in underlying conditions during a given time interval is a common topic of research in economics, marketing, finance, opinion polling, medicine, and other fields. It is usually investigated either through a comparison of the means of two different samples, one taken at the beginning and another at the end of the period under consideration, or through the observation of the changes occurring during the period in the binomial attributes of one sample. In the latter approach, the individuals before and the individuals after the test period are the same and the binomial sample is a "matched" one, the "match" being in the sameness of the individuals over time. This paper provides a critical evaluation of the matched sample test currently available for the determination of the statistical significance of an influence exerted on a binomial population and develops an alternative test for the same end. The two tests are discussed and evaluated within the context of two probability models which are believed to cover a wide range of applications and of two major conditions of *a priori* information. The distinguishing characteristic of the two conditions is the existence or absence of *a priori* information about the extent of the changes which are expected to occur over time in the dichotomous characteristics of individuals even when the binomial parameter remains constant over time. It is suggested that the alternative method is more appropriate for the problem at hand for both probability models and for the case of *a priori* knowledge about change. In the absence of such knowledge a procedure based on the notions of the alternative test is proposed for situations described by both models.

II. The Problem

A certain influence or "treatment", like propaganda in a political setting, TV message in a drive to increase consumer demand, the administration of a drug in a medical experiment, or simply the passage of time, is presumed to be capable of increasing the number of "successes" in a binomial population. The presumed effect of such treatment is to be tested statistically on a matched random sample of n observations. Measurements on the binomial characteristics of the sample observations before and after the treatment are to be taken and evaluated in a test of significance.

In the specific context where the observations are individuals and the binomial characteristic is an answer of either yes or no to a particular question, the following procedure would be followed. Each individual is asked the question before and after an exposure to a treatment or an influence. The responses are recorded in four categories: (1) affirmative on both occasions; (2) negative on both occasions; (3) first "yes" and then "no"; (4) first "no" and then "yes". The problem is to determine by means of a test of significance whether attitudes have changed as a result of an intervening influence.

The customary test for this matched sample design compared the number of changes from one to zero against the number of changes from zero to one under the null hypothesis that the number of changes in each direction is one half of the total number of changes. The reasoning underlying this approach can be summarized approximately as follows: If there is no treatment effect in either the direction of increasing, or in the direction of decreasing, the number of ones, we would expect to observe only changes due to "chance". These "chance switches," negative changes from one to zero and positive changes from zero to one, would occur even in the complete absence of a treatment effect. Under the null hypothesis that the treatment exerts no effect at all and that the binomial population retains the same parameter in the absence of a treatment effect, the expectation is that the switches during the interval of time under consideration should be equally numerous in either direction. Or, put differently, the sum of the changes, both positive and negative, should be expected to equal zero.

Whatever it may have in its favor, the customary test is inappropriate for the problem at hand because it is a conditional test which does not take account of the total phenomenon under consideration. The conditional framework of the test is unsuitable because it tests the null hypotheses for a given number of total non-zero changes (sum of positive and negative changes) for any sample size. The conclusions of the test hold for the particular number of non-zero changes which occur not only in samples whose size equals that of the sample used in the test, but also in samples of any other size. As long as two samples of unequal size have the same number of non-zero changes, they are considered, in the framework of the traditional test, as equivalent preceding conditions. The test is defective because it neglects to incorporate an important and substantial amount of the total process under consideration. It does not at all take account of the non-changing individual in the process and thereby neglects altogether to consider the stable part of the process. It is incapable of providing a comprehensive answer to the truly relevant question: What is the statistical significance of a particular sample difference between the positive and negative changes for a given sample size, i.e., for the process as a whole including all the non-zero and zero changes?

The alternative test proposed and developed in this paper attempts to supplement and remedy the defects of the current test. As it turns out, the alternative test produces substantially different answers. For a given sample size it appears to reject the null hypothesis that no net change has occurred more often than the customary test.

In order to evaluate the traditional test and the proposed alternative in the context of clearly defined underlying processes the binomial process under consideration is described below in two alternative ways. Constructed are two probability models which purport to consist of the essential and relevant elements of the process as these elements appear in most empirical settings of actual problems. These models state explicitly the assumptions underlying the null hypothesis, assumptions which are made implicitly about the binomial process in the usual experimental context. Two sets of underlying assumptions characterize the null hypothesis:

1. The parameter of the binomial population remains stable over time. All zeros remain zeros and all ones remain ones. "Chance" in the form of observational error brings about not true but observed change in particular observations.

2. The parameter of the binomial population remains stable over time, although, due to "chance" conceived as the unsystematic effect of many variables, some ones become zeros and some zeros become ones.

III. Model I

Model of Errors in Measurement or Mirage Switches

In this model no individual in the binomial population really changes over time. Some individuals, however, appear to change because of the mirage effect of observational error.

Let P represent the probability of the existence of some dichotomous attribute (success, yes). Let the existence of this attribute be designated as usual by <u>one</u>. Let (1-P), to be referred to also as Q, represent the probability of its absence (failure, no) and let its absence be designated by <u>zero</u>.

The individuals in the population are assumed to be observed with an error. Specifically, some individuals with true values of <u>one</u> are observed as <u>zeros</u>. Some individuals with true values of <u>zero</u> are observed as <u>ones</u>. Let B represent the probability of observing or measuring an individual with an error and let the probability of observing its true value correctly be G=1-B.

The previous assumption provides a distinction between the true, <u>one</u>, <u>zero</u>, values and the observed values of individuals in a binomial population. It introduces, in principle and on the level of the population as a whole, the existence of observation or measurement errors. It creates a link between the need to deal with observed values subject to error and the fact that most discussions of problems associated with binomial populations omit the consideration of inevitable errors in the measurement of values.

Let t be a point in time and let t+z be a point in time some time later where z is the length of the period under consideration. Let the probability of a true <u>one</u> at t be P_t, at t+z be P_{t+z}, and let $P=P_t=P_{t+z}$. Let the probability of observing an individual at time t, with an error, be represented by B_t and at t+z, by B_{t+z} and let $B=B_t=B_{t+z}$.

Let the probability of observing a change of <u>minus</u> <u>one</u>, -1, i.e., a switch from <u>one</u> to <u>zero</u>, between t and t+z be P_{10}. In terms of previously defined probabilities P_{10} turns out to be:

(1.01)

$$PGB + QBG = BG$$

The term PGB represents the probability of being actually <u>one</u> and being observed correctly as one at t and being mistaken for <u>zero</u> at t+z. The term QBG represents the probability of being really <u>zero</u> and being observed erroneously as <u>one</u> at t and being observed correctly as <u>zero</u> at t+z.

Let the probability of observing a change of <u>plus</u> <u>one</u>, +1, i.e., a switch from zero to one between t and t+z, be P_{01}. In terms of previously defined probabilities P_{01} emerges as:

(1.02)

$$QGB + PBG = BG$$

The first term above, QGB, represents the probability of a true <u>zero</u>, first observed correctly, then observed erroneously; and the second term PBG is the probability of a true value of <u>one</u> observed first incorrectly, later with no error.

With similar reasoning P_{11}, the probability of observing a value of <u>one</u> both at t and t+z, i.e., a change of <u>zero</u> is:

(1.03)

$$PG^2 + QB^2$$

and P_{00}, the probability of observing a value of zero both at t and t+z, i.e., again, a change of zero, is:

(1.04)

$$QG^2 + PB^2$$

As becomes evident upon inspection, the probability BG is the variance of the error distribution. Thus the variance of the error distribution becomes the probability of observing a change of a <u>plus</u> <u>one</u> +1, between t and t+z, as well as the probability of observing a change of <u>minus</u> <u>one</u>, -1, between the same two points in time. Also, 2BG represents the probability of observing a change, i.e., negative and positive unity changes and (1-2BG) represents the probability of observing stability or a change of zero. The probability (1-2BG) is, of course, equal to the sum of (1.03) and (1.04) above.

To summarize, the observed change, a random variable henceforth referred to as D_1, can take on values -1, 0 and +1, with associated probabilities BG, (1-2BG) and BG respectively. And more generally, if BG is redefined as K, then associated probabilities of -1, 0 and +1, are K, 1-2K and K respectively. In a typical situation the probability of error is small and the bulk of the probability of the distribution of the changes is centered at zero as is shown in Figure I. The expectation of this distribution of changes, henceforth referred to as $E(D_1)$ will equal zero as can be easily demonstrated by a straightforward use of the definition of expectation:

(1.05)

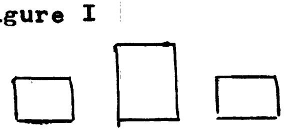
$$E(D_1)=(-1)BG+(0)(1-2BG)+(1)BG=0$$

Obtained in a similar fashion is the variance of the distribution of observed changes. The values of D_1, and their associated probabilities are employed in the definition of variance to obtain:

(1.06)

$$V(D_1) = (-1)^2BG + (1)^2BG = 2BG$$

Figure I



probabilities	BG	1-2BG	BG
probabilities	K	1-2K	K
D_1 values	-1	0	+1

Under the assumptions of Model I and particularly under the assumption that no change occurs in P between t and t+z, the expectation, E, of the distribution of the <u>sample sum</u>, S, of n independent observations randomly drawn from the distribution of changes is equal to zero since:

(1.07)

$$E(S) = \sum_{i=1}^{n} d_{1i} = nE(D_1) = 0.$$

The variance of the distribution of the sum, V(S), is equal to n times $V(D_1)$ which in turn equals n2BG.

(1.08)

$$V(S) = nV(D_1) = n2BG$$

Thus, in a Model I world, the null hypothesis implies an expectation of zero, variance of 2nBG or 2nK for the sample sum.

Under the alternative hypotheses associated with Model I, the parameter P can go either up or down in the following two clearly defined ways: (1) some true <u>zeros</u> becomes true <u>ones</u> and no change occurs in any true <u>ones</u>; or (2) some true <u>ones</u> become true <u>zeros</u> and no change occurs in the true <u>zeros</u>. A one right [left] tail test would imply for the alternative hypothesis the first way, (1), but not the second, (2) [the second way (2), but not the first, (1)]. A two tail test would make the alternative hypothesis either the first way, (1), or the second way, (2), of chaging P.

IV. Model II
Real Switches Model: No Error

In this model the parameter of the binomial population remains stable over time, but some true ones become true zeros and some true zeros become true ones due to a variety of "chance" influences not related to the particular factor whose effect is under investigation. No observational error is assumed to intervene in any way between the true values and their perception and recording.

Let P and Q be defined as they have been in Model I, such that P_t equals P_{t+z}. Let C_{01} be the probability that a true zero becomes a true one, C_{10} the probability that a true one becomes a true zero between t and t+z.

Given these assumptions, P at t+z can be viewed as:

(2.02)

$$P_{t+z} = P_t(1-C_{10}) + Q_tC_{01}$$

Since $P_{t+z} = P_t = P$,

(2.03)

$$P = P(1-C_{10}) + QC_{01}.$$

Expanded (2.03) indicates the nature of the addition to and subtraction from P.

(2.04)

$$P = P - PC_{10} + QC_{01}.$$

It becomes evident that as a consequence of preceding assumptions it is definitely, although implicitly, assumed that:

(2.05)

$$PC_{10} = QC_{01}$$

which in turn implies that:

(2.06)

$$P/Q = C_{01}/C_{10}$$

One more assumption is required to render Model II internally consistent. The probability of a switch away from an individual's own value must necessarily be constrained as follows:

(2.07)

$$C_{10} \overline{\underline{<}} Q/P$$

for $P \overline{\underline{>}} 1/2$. And similarly:

(2.08)

$$C_{01} \overline{\underline{<}} P/Q$$

for $Q \overline{\underline{>}} 1/2$. Without these constraints, it is impossible to maintain that the probability of positive changes is equal to the probability of negative changes without colliding with the model's logic as it now stands. The constraint simply acknowledges the fact that when more than half of the individuals are ones, it cannot be assumed that all ones become zeros, $C_{10} = 1$, and maintain the equality between negative and positive changes even if C_{01} is also assumed to equal unity.

The probabilities of observing

a switch are as follows:

from <u>one</u> to <u>zero</u>:

(2.09)

$$PC_{10}$$

from <u>zero</u> to <u>one</u>:

(2.10)

$$QC_{01}$$

The probability of remaining a <u>one</u> is:

(2.11)

$$P(1-C_{10})$$

and the probability of remaining a <u>zero</u> is:

(2.12)

$$Q(1-C_{01}).$$

Again, under the hypothesis that the treatment under consideration exerts no effect, we obtain under Model II a distribution of changes, D_2, whose values are -1, 0 and +1, and whose probabilities are PC_{10} [as in (2.09)], $1-PC_{10} - QC_{01}$, [the sum of (2.11) and (2.12)] and QC_{01} [as in (2.10)], respectively. As stated for D_1 in Model I (outlined above) the probabilities of the change, D_2, can be stated more generally as K for -1, (1-2K) for 0, and K for +1, in which case Figure I above may serve as a good description of the distribution of changes predicated on the assumptions of Model II. The null hypothesis under Model II for the sample sum, S, can be stated as it was under Model I: The expectation of the distribution of the sum statistic, S, is equal to zero. In a manner analogous to Model I, the alternative hypothesis can be: the expectation of the distribution of the sum of the changes is larger than zero (one tail), smaller than zero (one tail) and smaller or larger than zero (two tail). The difference is, of course, that any real change is a net change. It is the difference between positive true switches and negative true switches.

V. The Two Models and the Tests

It becomes evident that under the assumptions of both Model I and Model II we obtained distributions with common characteristics of observed changes. Under both models the possible values of the changes are -1, 0 and +1, and their associated probabilities can be expressed as K, (1-2K) and K respectively. We shall refer to this distribution generally as D. The difference between the models rests in the following: Under Model I, 2K must necessarily be equal to or smaller than 1/2 since by definition 2K equals 2BG and BG can be at most 1/4. In Model II, 2K is constrained in a different fashion. It can be as large as, or smaller than, twice the probability of <u>one</u> or <u>zero</u>, whichever is smaller. In most actual cases, however, and under both models, 2K is likely to be well below one-half (1/2) and probably closer to zero than to one-half.

The test of significance proposed in this paper as an alternative to the customary test, for situations depicted by either Model I or Model II, is based on a distribution of a statistic computed from a sample from the distribution of changes $D_1=D_2=D$. The statistic is the sum of n sample changes, d, "drawn" at random:

(3.01)

$$\sum_{i=1}^{n} d_i.$$

The expectation and variance of this statistic has been previously described as zero and n2K respectively.

In terms of the distribution of the sample sum of changes, as a context for the comparison of the customary test and the one proposed as an alternative, it can be said that the customary test ignores totally the zero changes in the sample as if they have never existed, and treats the non-zero changes which appear in the sample, as if they were, ex post, a sample drawn from a distribution with only two possible values, -1 and +1, with associated probability of 1/2 for each.[1] The test proposed in this paper is different from the traditional test in that it takes account of the total process including the zero changes. The distribution for the test statistic proposed as an alternative in this paper can be worked out exactly for the case where previous knowledge of 2K is available. Where such a priori knowledge is not available it is proposed that confidence intervals be computed for the significance level.

VI. Advance Knowledge of Intrinsic Change

Advance a priori knowledge of the magnitude of K (or 2K) in a Model I world may be based simply on actual knowledge of the probability of error B. If B is known independently, K may be obtained by means of the relationship between K and B.

(3.02)

$$K = B(1-B) = B\bar{B}$$

[1]Treating the non-zero d outcomes of the sample as if they came (retroactively) from a population with -1, +1 values and probabilities of 0.5, 0.5 respectively is equivalent to what is done in the traditional test where the -1 values are redefined as 0 and the +1 values redefined as one.

In a Model II world, a priori knowledge of the intrinsic turnover is simply a direct knowledge of 2K or K.

If K is known before the start of the sampling process, then the test statistic, the sample sum, S, can be evaluated against its exact probability distribution which can be derived by analytical methods as is illustrated below. For large samples, the central limit theorem can be relied upon to make the normal (Gaussian) distribution an acceptable approximation for the distribution of S in which case the approximately normal distribution under consideration will be one with expectation of zero and a variance of n2K.

The probability distribution of the sample sum, S, can be derived by straightforward enumeration or by a method which is conditioned by the view that for a given sample size, the sample sum, S, is an outcome of a two stage process. The cumulative probability table for S, below, was computed by the second method.

The probability distribution of S can be derived by considering a draw from S as a two stage process. (a) A draw to determine the number, C_i, of non-zero outcomes in a sample size n where C_i ranges between 0 and n, i between 1 and n+1. The probabilities of C_i, $P(C_i)$, i.e., the probabilities of $C_1=0$, $C_2=1, \ldots C_{n+1}=n$ are usual binomial probabilities and are determined by 2K and n. (b) C_i draws from a distribution of +1, -1, with associated probabilities 1/2, 1/2 to determine S for a particular C_i. To compute the probabilities of values equal to or larger than S for sample size n, for all possible values of C_i, we compute (1) the probability of values equal to or larger than S for each C_i (2) we multiply the preceding probability by the probability of C_i, $P(C_i)$, and (3) sum this product for all S, for i=1...n+1 (for a given S, all terms from 0 to n). The probability of C_i is given by the binomial parameter 2K, the probability of non-zero change and by the binomial expansion. The probability of values equal to or larger than S for C_i is given by the binomial parameter 1/2 and the binomial expansion.

There are (2n)+1 possible S values ranging from -n to +n in consecutive steps of unity. Since the probability distribution of S turns out to be symmetrical about zero, cumulative probabilities are presented only for values equal to or larger than X where X is zero or <u>positive</u> values of S, i.e., for values of S ranging from zero to n. A table is developed for sample size 1 to 10 and for probability of change 2K, equal to 0.2, [i.e., K=.1, (1-2K)=.8].

Upon comparison of the traditional test probabilities with those produced by our table, it becomes evident that the two methods produce different results. For example, take a case in which the number of non-zero changes equals two, C=2, and sample size is also 2. According to the traditional test the probability that S will be equal to or larger than 2 is the product (0.5)(0.5)=0.25. This result is obtained regardless of the values of 2K. According to the method proposed in this paper and given that 2K is equal to .2, the probability for the same outcome is 0.01 (see Table 1). For C of 2, for 2K of 0.2, but for a sample size of 10, the traditional test produces again the value 0.25 for the probability that S is equal to or larger than 2, the method proposed in this paper produces a probability of 0.13524132 for the same outcome (see Table 1).

VII. No Advance Knowledge of Intrinsic Change

When no knowledge of the magnitude of 2K is available before the sample is taken, it is proposed that the 2K be estimated from the data by the ratio of the observed number of non-zero changes to total observed changes. Once this estimate is available, the usual confidence intervals with Y percent confidence are computed with an upper limit and a lower one. These confidence limits are then used as if they were two known values of 2K to compute two levels of significance for the null hypothesis, a lower and an upper limit. It can then be said with Y percent confidence that the true level of significance falls between the upper and lower significance limits Y percent of the time, given, of course, that the null hypothesis of no net change is true.

TABLE 1

Cumulative Probability Table for Number of Positive (Negative) Changes Over Time in a Sample Size of n* from a Binomial Population. 2K equals 0.2.**

Sample Size	0 6	1 7	2 8	3 9	4 10	5
1	.90000000	.10000000				
2	.83000000	.17000000	.01000000			
3	.78000000	.22000000	.02500000	.00100000		
4	.74350000	.25650000	.04210000	.00330000	.00010000	
5	.71624000	.28376000	.05966000	.00686000	.00041000	.00001000
6	.69540200	.30459800	.07679000	.01149500	.00101500	.00004900
	.00000100					
7	.67910240	.32089760	.09304130	.01697650	.00196640	.00014080
	.00000570	.00000010				
8	.66606755	.33393245	.10822045	.02308197	.00328485	.00030985
	.00001865	.00000065	.00000001			
9	.65542519	.34457473	.12227779	.02961610	.00496706	.00057823
	.00004597	.00000238	.00000007	.00000000		
10	.64656984	.35343008	.13524132	.03641736	.00699308	.00096388
	.00009483	.00000651	.00000029	.00000000	.00000000	

* Sample size from 1 to 10

** The probability of positive and negative changes, 2K, is .02, The probability of +1 is 0.1. The probability of -1 is 0.1.