Comp. & Maths. with Appls. Vol. 10, No. 4/5, pp. 343–349, 1984
Printed in the U.S.A.
0097-4943/84 $3.00 + .00
Pergamon Press Ltd.

TWO-STAGE SEQUENTIAL TESTS FOR PROPORTIONS AND MEANS WITH FOLLOW-UP PERIOD

S. C. Choi
Department of Biostatistics, Medical College of Virginia, Richmond, VA 23298, U.S.A.

(*Received December* 1983)

Communicated by I. Norman Katz

Abstract—This paper extends a procedure, based on [1], wherein a sequential probability ration test (SPRT) is combined with a likelihood ratio test that incorporate both sequential observations and additional delayed observations. Two two-sample sequential tests are considered: one for binomial parameters and the other for the means of normal distributions. The essential feature of the method enables us to construct a SPRT with the risk probabilities set greater than the desired level, depending on a number of delayed observations. The terminal decision based on a likelihood ratio test can be made with the desired risk level. The adequacy and advantage of the procedure are discussed and a number of computational formulas is presented to facilitate its applications.

1. INTRODUCTION

The primary aim of sequential procedures in experiments is to reduce (on the average) the necessary sample size. In experiments where each observation requires a long follow-up period, however, there is relatively little scope for a sequential method to reduce the sample size. That is, sequential methods are not well suited for applications unless the response on which the stopping rule is based is available soon after or within a period comparatively shorter than the period in which the subjects enter the trial (see [12]). In many biomedical and industrial experiments the response time for observations may be at least several weeks to several months.

Consider an experimental trial in which each subject (or unit) is assigned randomly to one of two treatment groups soon after its arrival and the response time is relatively long. Suppose a sequential test is applied as observations become available. When a decision is made to stop the trial, there will be a number of subjects already on trial but for which the required observation time has not elapsed. The number of such delayed observations is an increasing function of the length of the response time and the arrival rate of the subjects. The essential feature of the proposed two-stage test enables us to construct an initial sequential test with the risk probabilities set larger than the desired level, taking consideration of delayed observations. It should be noted that, although a SPRT provides both the stopping rule and the terminal decision rule, it is used only as a stopping rule in the proposed test. We consider two tests, one for comparing binomial parameters and the other for the means of two normal distributions.

Most sequential methods previously studied for experiments with long follow-up periods are different in philosophy from the method considered here. The main question dealt with by [2, 4, 8, 11], among others, is whether treatments differ in their effect on the time elapsing before a certain response is observed; usual applications of these techniques are for comparing survival time. Here, we are dealing with the traditional type of sequential test in which the observations are independent and identically distributed and the test statistic is based only on data with completed responses.

2. TEST FOR BINOMIAL PROPORTIONS

First, consider the problem of comparing the probabilities of success, p_1 and p_2, of two treatments. A formulation of the null hypothesis and alternative suitable in many experiments is given by

$$H_0: p_1 = p_2 = p, \tag{1}$$

$$H_1: p_1 = p + \Delta/2, \; p_2 = p - \Delta/2, \tag{2}$$

where only Δ is a specified nonnegative constant.

For testing the hypothesis against the alternative under the condition described in Section 1, we propose a two-stage test; the first stage is a SPRT proposed by [9] and the second stage a likelihood ratio test combining the SPRT and the data from delayed observations. The similar procedure for paired observations has been studied by [5]. In some situations, pair-wise allocation is not very practical aside from its possible inefficiency. For the first stage SPRT, we shall require that whenever H_0 is true, the probability of accepting H_0 be at least $1 - \alpha_1$ $(0 < \alpha_1 < 1)$ and that whenever H_1 is true, the probability of accepting H_1 be at least $1 - \beta_1$ $(0 < \beta_1 < 1)$. Of all available observations with known responses at a given time, let n_i, $i = 1, 2$, be the number of subjects in the ith group, and s_i and f_i be the number of successes and failures among n_i observations such that $s_i + f_i = n_i$.

For the initial SPRT for (1) and (2), consider for the moment a simple hypothesis and a simple alternative given by

$$H_0': p_1 = p_2 = 0.5 \tag{3}$$

$$H_1': p_1 = 0.5 + \Delta/2, \; p_2 = 0.5 - \Delta/2. \tag{4}$$

Let

$$T_1 = (0.5 + \Delta/2)^{s_1+f_2}(0.5 - \Delta/2)^{s_2+f_1}/(0.5)^{n_1+n_2}.$$

The SPRT for (3) and (4) is: continue sampling if

$$B < T_1 < A \tag{5}$$

where $B = \ln[\beta_1/(1 - \alpha_1)]$, $A = \ln[(1 - \beta_1)/\alpha_1]$, and accept H_0' or H_1' according to whether the left-hand or the right-hand inequality is the first violated. It follows from [9] that the SPRT given by (5) yields a test for the (1) and (2) with the error probabilities bounded by α_1 and β_1. That is, the error probabilities of the SPRT for (1) and (2) will be not constant over the set H_0 and H_1, but are maximized at H_0' and H_1'.

After the SPRT is used to stop the experiment, let m_i, $i = 1, 2$, be the number of subjects in the ith treatment group whose observations become available after the required follow-up period. Let s_i' and f_i' be the number of successes and failures among m_i such that $s_i' + f_i' = m_i$ and $m_1 + m_2 = M$. Let

$$T_2 = (0.5 + \Delta/2)^{s_1+s_1'+f_2+f_2'}(0.5 - \Delta/2)^{s_2+s_2'+f_1+f_1'}/(0.5)^{N+M},$$

where N denotes the total number of observations required in the SPRT.

The second stage and final test we propose is: accept H_0 if

$$T_2 < k \tag{6}$$

and accept H_1 otherwise, where k is a nonnegative constant.

The overall error probability α of falsely rejecting H_0 by the final test given by (6) can be formulated as

$$\alpha = P(T_1 \geq A, \; T_2 \geq k \mid H_0) + P(T_1 \leq B, \; T_2 \geq k \mid H_0). \tag{7}$$

and an analogous expression for the probability β of falsely rejecting H_1 by the terminal test:

$$\beta = P(T_1 \leq B,\ T_2 \leq k|H_1) + P(T_1 \geq A,\ T_2 \leq k|H_1). \tag{8}$$

The choice of k is somewhat arbitrary as long as α determined from (7) is less than the significance level. In practice, however, k can be fixed at a constant so as to balance α and β at one's desired levels. In this paper for simplicity, we shall let $k = 1$, which is a reasonable choice when $\alpha = \beta$.

The probabilities given by (7) and (8) can be calculated using the usual approximations of SPRT known as "neglecting the excess", and applying the computational method similar to that described by [5]. Let

$$c = (-A - F)/G,\ d = (-B - F)/G,$$

where

$$F - (m_1 + m_2)\ln(2) + m_1 \ln(0.5 - \Delta/2) + m_2 \ln(0.5 + \Delta 2),$$

and

$$G = \ln[(0.5 + \Delta/2)/(0.5 - \Delta/2)].$$

We obtain

$$\alpha = \alpha_1 \sum_{a(c)}^{m_1} b_1(x;p) \sum_{0}^{g(c)} b_2(y;p) + (1 - \alpha_1) \sum_{a(d)}^{m_1} b_1(x;p) \sum_{0}^{g(d)} b_2(y;p) \tag{9}$$

where

$$a(z) = \max\ [z + 1,\ 0],\ g(z) = \min\ [m_2,\ x - z - 1],\ b_i(x;\ p) = \binom{m_i}{x} p^x (1-p)^{m-x},$$

with $[x]$ denoting the greatest integer less or equal to x. Similarly,

$$\beta = \beta_1 \sum_{f(d)}^{m_2} b_2(y;p_2) \sum_{0}^{h(d)} b_1(x;p_1) + (1 - \beta_1) \sum_{f(c)}^{m_2} b_2(y;p_2) \sum_{0}^{h(c)} b_1(x;p_1) \tag{10}$$

where p_i, $i = 1, 2$, is given by (2) and

$$f(z) = \max[1 - z, 0],\ h(z) = \min[m_1, y + z - 1].$$

The error probabilities α and β given by (9) and (10) for a given Δ depend on p_1, p_2 and p. However, it is expected that the α and β are maximized at H_0' and H_1'. Thus it is possible to construct the test with only Δ specified. Table 1 gives the probabilities of errors α and β when $\alpha_1 = \beta_1 = 0.05$ for selected value of m where $m = m_1 = m_2$.

The table reveals the intuitively clear fact that the reductions, α_1–α and β_1–β are strictly increasing function of m and of Δ. For a given Δ, results for $p = 0.7$, for example, will be identical to those for $p = 0.3$ due to the symmetry.

In practice it would be desirable to determine α_1 and β_1 for the first stage test so that, for given m_1 *and* m_2, the terminal error probabilities are less than or equal to predetermined values of α and β. A trial and error method using (9) and (10) can be used to determine the desired nominal level of α_1 and β_1. For example, Table 2 gives the approximate values of $\alpha_1 = \beta_1$ which will assure the terminal error probabilities $\alpha \leq 0.05$ and $\beta \leq 0.05$ by the proposed two-stage test. As an illustration, let $\Delta = 0.3$ and $m = 30$. We can construct a SPRT based on $\alpha_1 = \beta_1 = 0.084$ which is substantially greater than the desired level of 0.05.

Table 1. Terminal error probabilities for given m and p when $\alpha_1 = \beta_1 = 0.05$ for testing $p_1 = p_2 = p$ vs $p_1 = p + \Delta/2$ and $p_2 = p - \Delta/2$

Δ	p		m 10	20	30	40
0.2	0.3	α	.049	.047	.043	.039
		β	.049	.048	.043	.038
	0.5	α	.049	.047	.044	.040
		β	.049	.048	.044	.040
0.3	0.3	α	.045	.033	.031	.022
		β	.045	.032	.024	.014
	0.5	α	.045	.036	.034	.027
		β	.045	.035	.028	.018

Table 2. Approximate initial values of error probability $A_1 (= B_1)$ for SPRT to make terminal error probabilities of $\alpha \leqslant 0.05$ and $\beta \leqslant 0.05$

Δ	m 10	15	20	25	30	35	40
0.2	0.051	0.052	0.053	0.056	0.059	0.062	0.067
0.3	0.059	0.061	0.075	0.079	0.084	0.095	0.123

The expected sample size for the proposed method was examined by a computer simulation study. Notice that the expected sample size is the sum of the average sample number (ASN) of the SPRT given by (5) and m. Table 3 presents the empirically determined expected sample size required from each group under H_i, $i = 0, 1$, when $\Delta = 0.2$ and 0.3, and $p_1 = 0.3$ and 0.5. The ASN corresponding to each entry was empirically estimated using a simulation based on 1,000 independent sampling experiments. The

Table 3. Expected sample size required from each group to achieve $\alpha \leqslant 0.05$ and $\beta \leqslant 0.05$ using the proposed method and the corresponding fixed-sample size

Δ	p_1	True State	m 10	20	30	40	Fixed-sample Test
0.2	0.3	H_0	86.0	87.9	97.1	99.4	152
		H_1	87.6	90.0	99.8	101.8	
	0.5	H_0	77.5	84.2	93.2	99.3	
		H_1	78.4	88.5	99.3	100.1	
0.3	0.3	H_0	40.7	47.6	54.7	59.9	71
		H_1	40.9	49.3	57.0	60.2	
	0.5	H_0	38.4	46.4	54.2	59.0	
		H_1	39.9	46.5	54.5	60.1	

standard error of the estimated ASN ranged from 1.3 to 1.7 when $\Delta = 0.2$ and from 0.5 to 0.7 when $\Delta = 0.3$. The results indicate that when H_0 is true, regardless of Δ, the test requires a smaller sample size on the average when $p_1 = p_2 = 0.5$, and when H_1 is true the minimum occurs at $p_1 = 0.5 + \Delta/2$ and $p_2 = 0.5 - \Delta/2$. This finding agrees with the argument made in [9] that the ASN is minimized over H_0 at H_0' and over H_1 at H_1'. The differences for the most cases, however, appear to be small.

For the purpose of comparison, the corresponding sample size required by a fixed sample test based on [10] with the risk levels of $\alpha = \beta = 0.05$ is also shown in Table 3. Substantial saving in the average sample size by the proposed test is indicated. For example, let $\Delta = 0.2$ and $m = 20$. Referring to Table 3, the proposed test requires, on the average, 84 or 90 subjects from each group while the fixed sample test will need 152. As expected, the expected sample size is shown to increase with m. Clearly, the optimal situation is when there is no delayed observation at all.

3. TEST FOR MEANS OF NORMAL DISTRIBUTIONS

Next, consider the problem of testing the hypothesis and the alternative regarding the means of two normal distributions:

$$H_0: \mu_1 = \mu_2 = \mu,$$

$$H_1: \mu_1 = \mu + \Delta\sigma, \ \mu_2 = \mu,$$

where σ^2 is the common but unknown variance and Δ is a specified constant which, without loss of generality, is assumed to be a positive constant. The form of H_1 is set slightly different from (2) for mathematical simplicity.

Let $x_1, x_2, \ldots$ be observations from $N(\mu_1, \sigma^2)$ and $y_1, y_2, \ldots$ be those from $N(\mu_2, \sigma^2)$. As the first stage of the proposed test, we wish to consider the asymptotic SPRT investigated by [2, 6, 7]. If we assume $n_1 = n_2 = n$ for simplicity, using Ghosh's approach the SPRT becomes: continue sampling if

$$(2B + n\Delta^2/2)(s/\Delta) < \sum^n x_i - \sum^n y_i < (2A + n\Delta^2/2)(s/\Delta) \tag{11}$$

where A and B are defined as in Section 2 and s is the pooled sample standard deviation based on $2n$ observations. Decision is made to accept H_0 or H_1 according to whether the left-hand or the right-hand inequality is the first violated. As we shall see later, this SPRT appears to be satisfactory for testing the above hypotheses.

Again, let m_i, $i = 1, 2$, be the lagged observations expected after the first stage decision based on (11) is made. Assume $m_1 = m_2 = m$ for simplicity and now let s denote the pooled standard deviation based on $2(n+m)$ observations. The second stage test based on approximate likelihood ratio test, using the substitution $\mu_2 = \Sigma^{n+m} y_i/(n+m)$ and $\sigma = s$, reduces to the following: accept H_0 if

$$\sum^{n+m} x_i - \sum^{n+m} y_i < (n+m)\Delta s/2, \tag{12}$$

and accept H_1 otherwise.

The overall error probabilities α and β for the final test given by (12) using the approximation of "neglecting the excess" can be obtained from (7) and (8) as follows:

$$\alpha = \alpha_1\left[1 - \phi\left(\frac{m\Delta^2/2 - 2A}{\Delta\sqrt{2m}}\right)\right] + (1-\alpha_1)\left[1 - \phi\left(\frac{m\Delta^2/2 - 2B}{\Delta\sqrt{2m}}\right)\right] \tag{13}$$

$$\beta = \beta_1\phi\left(\frac{m\Delta^2/2 - 2B - m\Delta^2}{\Delta\sqrt{2m}}\right) + (1-\beta_1)\phi\left(\frac{m\Delta^2/2 - 2A - m\Delta^2}{\Delta\sqrt{2m}}\right) \tag{14}$$

where $\phi(x)$ represents the standard normal distribution function.

Table 4. Terminal error probability α for given m with $\alpha_1 = \beta_1 = 0.05$ for SPRT based on theoretical approximation and simulation

			m			
Δ			10	20	30	40
0.04	α	Approximation	0.050	0.049	0.047	0.044
		Simulation	0.043	0.036	0.034	0.033
0.6	α	Approximation	0.049	0.042	0.034	0.027
		Simulation	0.037	0.036	0.031	0.027
0.8	α	Approximation	0.044	0.030	0.020	0.013
		Simulation	0.034	0.025	0.017	0.013

The test procedure proposed by (11) and (12) involves a number of approximations, but it is not practical to theoretically assess their effect or to examine the adequacy of the approximations given by (13) and (14). For these reasons a computer simulation study was performed in order to examine α and β empirically. The nominal risk probabilities of the initial SPRT are set to $\alpha_1 = \beta_1 = 0.05$, and without loss of generality it is assumed that $\mu = 0$ and $\sigma = 1$. For (11) the simulation is based on 2000 independent tests for each Δ. This part is combined with another 2000 independent tests using (12) for the given Δ and for each m. The result of the simulation study along with the value calculated using (13) is presented in Table 4.

The results indicate that the theoretical approximations of α and β are conservative. An important fact, however, is that both the theoretical approximation and the computer simulation results show that the procedure does reduce the probabilities of errors considerably. Again, as in Section 2, one may employ a trial and error method using (13) and (14) to determine the approximate values of α_1 and β_1 for the SPRT to end up with desired level of α and β.

The above simulation study also facilitated the examination of the expected sample size which is given by the sum of the ASN for the SPRT based on the simulation and m. The results are summarized in Table 5. When $\alpha_1 = \beta_1$ computations based on (13) and (14) yield $\alpha = \beta$ because of symmetry. Moreover, the simulation result for β and the ASN under H_1 was very much similar to those under H_0 presented in Table 5. Thus, the result under H_1 is not included in the table. Again, as expected, the expected sample size is shown to increase with m. Also presented in Table 5 is the approximate sample size from each

Table 5. Expected sample size required from each group to achieve the theoretically approximated α and β (Table 4) for testing $\mu_1 = \mu_2$

		m			
Δ	Test	10	20	30	40
0.04	Proposed	80.6	89.8	101.1	111.2
	Fixed-sample	135	138	142	147
0.6	Proposed	42.1	51.7	60.7	72.6
	Fixed-sample	63	68	75	83
0.8	Proposed	28.5	38.2	48.2	58.0
	Fixed-sample	39	47	54	63

population required by the two-sample t-test to achieve the same terminal error probabilities α and β calculated by the theoretical approximation. The saving by the proposed test does not appear overwhelming when Δ is large but, as noted, the comparison is inequitable for the proposed test since the theoretical approximation is conservative.

4. CONCLUSIONS

A two-stage two-sample sequential test for binomial parameters and a similar test for the means of normal distributions are proposed. Use of delayed observations reduces the probability of errors and, this reduction increases with the number of delayed observations. The most important fact is that one can realize substantial saving in the average sample size by the proposed test compared to the fixed sample test.

The procedure is useful for at least two situations in practice. First, it shows what can be done when delayed observations become available after a decision has been made by a SPRT. The second and more useful application is when the response time is relatively long; a first stage SPRT can be constructed with the probabilities of errors set considerably larger than the desired values of α and β depending on the expected number of delayed observations; the second stage of the test can achieve the desired risk probabilities. For some applications, it may be suggested that an interim decision be made by the SPRT until the terminal decision is reached.

Acknowledgement—This work was supported in part by NINCDS Grant NS-12587.

REFERENCES

1. T. W. Anderson, Sequential analysis with delayed observations. *J. Amer. Statis. Assoc.* **59**, 1006–1015 (1964).
2. P. Armitage, P., *Sequential Medical Trials*, 2nd Edn. Wiley, New York (1975).
3. M. S. Bartlett, The large sample theory of sequential tests. *Proc. Camb. Phil. Soc.* **42**, 239–244 (1946).
4. P. L. Canner, Monitoring treatment differences in long-term clinical trials. *Biometrics* **33**, 603–615 (1977).
5. S. C. Choi and V. A. Clark, Sequential decision for a binomial parameter with delayed observations. *Biometrics* **26**, 411–420 (1970).
6. D. R. Cox, Large sample sequential tests for composite hypotheses. *Sankhya* **25**, 5–12 (1963).
7. B. K. Ghosh, *Sequential Tests of Statistical Hypotheses*. Addison–Wesley, Reading, Mass (1970).
8. M. Halperin and J. Ware. Early decision in a censored Wilcoxan two-sample test for accumulating survival data. *J. Am. Statist. Assoc.* **69**, 414–422 (1974).
9. D. G. Hoel G. H. Weiss and R. Simon, Sequential tests for composite hypotheses with two binomial populations. *J. Roy. Statist. Soc.* **B38**, 302–308 (1976).
10. M. Kramer and S. W. Greenhouse, Determination of sample size and selection of cases. *National Research Council Publication* 583, Washington, D.C., 583 (1959).
11. L. R. Muenz, S. B. Green and D. P. Byar, Applications of the Mantel–Haenszel statistic to the comparison of survival distributions. *Biometrics* **33**, 617–626 (1977).