The Sequential Test of Two Treatments When Subjects are Paired in Many-to-One Ratio

S. C. Choi*

1. Introduction

In the designs previously developed for the sequential clinical trial of two treatments it was intended that subjects should be paired in one-to-one ratio. See Armitage [1960], Bross [1952] and Choi [1968] for example. Consider a sequential test of a new treatment against the standard treatment. In many instances it is conceivable that standard and new treatments are given in the ratio of m-to-one, ($m \geq 2$). Not only can such ratios give more efficient results, but it is likely that ethical requirements would also point in this direction. For example, the situation in which an illness is severe and effectiveness of the standard treatment is known to be at worst non-negative would suggest such ratios. Let the response variables of the standard and the new treatments be represented by S and T respectively. A set of observations consists of m independent observations S from the standard and one observation T from the new treatment. It is assumed that S and T have unknown but continuous distributions and ranking is accomplished within each set of observations in combined array. Although the technique has been described with reference to clinical trials, it will be apparent that the results are applicable in other areas.

In the subsequent discussion two constants A and B, ($B < A$) represent the usual upper and lower boundaries of the sequential probability ratio test (SPRT). In this paper A and B are approximated by the usual Wald bounds given by

$$A = (1-\beta)/\alpha, \qquad B = \beta/(1-\alpha)$$

where α and β are the probabilities of error of the first and the second types respectively.

---

* The author is Associate Professor at the Washington University School of Medicine, St. Louis, Missouri, U.S.A. This work was supported in part by the Public Health Service Grant PHT 6-8B-67.

2. Mathematical Formulation

Let $F(t)$ and $G(s)$ denote the distribution functions of T and S respectively. A comparative treatment characteristic of interest is the probability of possible superiority of a new treatment over a standard treatment. This probability can be represented by $p=P\ (T<S)$. Then we wish to test the hypotheses of the forms,

$$H_0 : p=\frac{1}{2} \quad vs. \quad H_1 : p=p'\left(p'>\frac{1}{2}\right) \tag{1}$$

A special case of (1) is a formulation due to Lehmann [1953]. The set of hypotheses is

$$H_0 : G(t)=F(t) \ \ vs. \ \ H_1 : G(t)=F^k(t), \quad k>1 \tag{2}$$

An implication of (2) in terms of (1) was given by Wilcoxon, Rhodes and Bradley [1963]. Like Wilcoxon et al., we will develop the tests based on (2). Bradley, Martin and Wilcoxon [1965] have demonstrated that the use of the sequential methods based on (2) is satisfactory for most practical applications for data from normal populations differing only in location.

The one-sidedness of the test is not a serious disadvantage in clinical experiments when a new treatment is compared with a standard since there may be no interest in showing whether the new treatment is worse or not.

3. The Binomial Test

Suppose that subjects are paired in m-to-1 ratio in a sequential test of the standard against a new treatment. Let r represent the rank of T. Clearly, we have $1\leq r\leq m+1$. Define a new random variable Z_i $(2\leq i\leq m+1)$ as follows.

$$\begin{aligned} Z_i&=1 \text{ if } r\geq i \\ &=0 \text{ otherwise.} \end{aligned} \tag{3}$$

Then Z_i has a Bernoulli distribution. For example; if $m=3$, then Z_3 has a Bernoulli distribution with the parameter $\frac{1}{2}$ under H_0. If we can determine the parameters of Z_i under H_0 and H_1, then m different SPRT can be constructed to test the same decision problem given by (2).

First, the parameter p_0 of Z_i under $H_0 : p=\frac{1}{2}$ is simply

$$p_0=1-(i-1)/(m+1) \tag{4}$$

Next, the probability distribution of r, $p(r/p)$, for given p and m can be obtained using the results given by Lehmann and Wilcoxon et al.,

$$p(r|p)=\frac{k\Gamma(r+k-1)\Gamma(m+1)}{\Gamma(m+k+1)\Gamma(r)}, \quad r=1,2,\cdots\cdots, \quad m+1 \tag{5}$$

where $k=p/q$ with $q=1-p$.

Using (5) the parameter p_1 under H_1 for the binomial test based on Z_i can be determined as follows

$$p_1=\sum_{r=i}^{m+1} p(r|p') \tag{6}$$

As an example, let $m=3$ and $p'=0.8$ under H_1. Then p_1 of Z_4 corresponding to $p'=0.8$ is

$$p(4|0.8)=\frac{4\Gamma(7)\Gamma(4)}{\Gamma(8)\Gamma(4)}=4/7.$$

Hence, a SPRT of $p_0=\frac{1}{4}$ against $p_1=\frac{4}{7}$ is one of three binomial tests which can be used to test (2) when $m=3$. The SPRT of $p=p_0$ versus $p=p_1$ is described in Wald [1947]. For $m=2$ (1) 6 and $p'=.6(.1).9$ the corresponding values of p_1 for the binomial test of Z_i are calculated to facilitate the SPRT of (2). These appear in Table 1. The appropriate value of p_0 is not given since it is simply obtained from (4).

Naturally, it is of interest to examine the relative efficiency of each of m binomial tests. For fixed α and β values, ($\alpha=\beta=0.01$ and $\alpha=\beta=0.05$) the average sample number (ASN), $E_j(n)$, under $H_j\,(j=0,1)$ is computed by the well known approximation formula given by Wald [1947].

$$E_j(n)\approx\frac{(1-L(p_j))\log A+L(p_j)\log B}{p_j\log(p_1/p_0)+(1-p_j)\log\{(1-p_1)/(1-p_0)\}} \tag{7}$$

where $L(p_0)=1-\alpha$ and $L(p_1)=\beta$.

The ASN functions corresponding to the SPRT for $m=2$, 3 and 4 are given in Table 2.

Consider the binomial test Z_d where $d=1+m/2$ if m is an even and $d=(m+1)/2$ or $d=(m+3)/2$ if m is an odd number. The test Z_d can be considered as the sequential median test. For example, Z_2 when $m=2$ and Z_2 or Z_3 for $m=3$ are the corresponding median tests. Examination of Table 2 and further computations indicate that within the class of binomial tests, the median test appears to be the most efficient in terms of the ASN function in general.

4. The Rank Test

It can be conjectured that the efficiency of the binomial test would decrease as m increases. A more powerful test can be constructed using the probability distribution of the rank. The SPRT when observations are taken in groups of m-to-m ratio of subjects was proposed by Bradley, et al. [1963, 1966]. The problem considered in this paper can be looked on as a special case and the test reduces to a simple sequential test.

Let $L_n(r|p)$ denote the likelihood ratio of the distribution of r under H_1 to that of H_0 when n sets of observations are made. From (5) we have

$$L_n(r|p)=\left\{\frac{k\Gamma(m+1)}{\Gamma(m+k+1)}\right\}^n \prod_{l=1}^{n}\frac{\Gamma(r_l+k-1)}{\Gamma(r_l)} \tag{8}$$

where r_l represents the rank of T in the l th sample. It follows that the continuation region of the SPRT based on the rank of T is given by

$$\log B-n\log\left\{\frac{k(m+1)!}{(m+k)!}\right\}<\sum_{l=1}^{n}\log\frac{(r_l+k-2)!}{(r_l-1)!}<\log A-n\log\left\{\frac{k(m+1)!}{(m+k)!}\right\} \tag{9}$$

TABLE 1 VALUES OF p_1 FOR THE BINOMIAL TEST

m	p'	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7
2	.6	.771	.429				
	.7	.862	.538				
	.8	.933	.667				
	.9	.982	.818				
3	.6	.848	.619	.333			
	.7	.922	.740	.437			
	.8	.971	.857	.571			
	.9	.995	.955	.750			
4	.6	.889	.723	.515	.273		
	.7	.951	.836	.645	.368		
	.8	.986	.929	.786	.500		
	.9	.999	.986	.923	.692		
5	.6	.915	.787	.627	.441	.231	
	.7	.966	.888	.758	.569	.318	
	.8	.992	.960	.881	.722	.444	
	.9	.998	.994	.972	.890	.643	
6	.6	.932	.830	.702	.552	.385	.200
	.7	.976	.920	.826	.690	.509	.280
	.8	.995	.976	.929	.833	.667	.400
	.9	1.000	.998	.989	.956	.857	.600

TABLE 2* THE ASN OF THE BINOMIAL TEST AND THE RANK TEST

m	p'	α	Z_2		Z_3		Z_4		Z_5		Rank Test	
2	.6	.05	93.1	100.5	139.6	135.1					81.3	85.1
		.01	158.2	170.2	237.1	229.5					138.1	144.7
	.7	.05	21.7	26.7	31.0	29.9					18.5	21.9
		.01	36.9	45.4	52.8	50.9					31.4	35.6
	.8	.05	8.5	12.8	11.4	11.4					7.0	9.0
		.01	14.4	21.8	19.5	19.5					11.9	15.2
	.9	.05	3.7	8.0	4.7	5.3					2.9	4.7
		.01	6.3	13.8	7.9	9.0					5.0	8.0
3	.6	.05	82.9	93.7	90.7	92.7	161.4	152.6			64.9	69.5
		.01	140.7	159.3	154.3	157.4	274.3	259.2			110.2	118.1
	.7	.05	19.4	26.7	20.2	22.0	35.0	32.0			14.5	17.2
		.01	33.0	45.2	34.3	37.4	59.4	54.3			24.6	29.2
	.8	.05	7.5	14.0	7.4	9.3	12.4	11.4			5.3	7.4
		.01	12.9	23.8	12.6	15.9	21.1	19.4			9.0	12.5
	.9	.05	3.4	10.1	3.1	5.1	4.8	4.8			2.1	3.9
		.01	5.7	17.1	5.1	8.9	8.2	8.2			3.7	6.5
4	.6	.05	79.0	92.8	75.6	80.3	99.4	98.1	186.5	173.4	56.5	61.7
		.01	134.3	157.8	128.5	136.4	168.9	166.7	316.7	294.9	96.0	104.7
	.7	.05	18.6	28.6	16.9	20.2	21.5	21.8	39.7	35.0	12.4	15.3
		.01	31.6	47.3	28.6	34.3	36.5	37.0	67.3	59.4	21.1	26.0
	.8	.05	7.4	15.7	6.3	9.3	7.5	8.5	13.8	11.9	4.5	6.6
		.01	12.5	26.8	10.5	15.9	13.0	14.5	23.4	20.2	7.6	11.2
	,9	.05	3.2	12.4	2.6	5.9	2.9	4.3	5.1	4.7	1.8	3.4
		.01	5.5	12.0	4.3	10.2	5.0	7.3	8.7	8.0	3.0	5.8

* In each entry the 1st column lists the ASN under H_0 and the 2nd column under H_1.

The decision is to accept $H_0(H_1)$ as soon as the lower (upper) inequality is broken. The ASN function of the rank test is obtained from

$$E_j(n) \approx \frac{(1-L(p_j))\log A + L(p_j)\log B}{E(W_j)}, \quad j=0,1 \tag{10}$$

where

$$E(W_j) = \sum_{r=1}^{m+1} p(r|p) \log \frac{k(m+1)!(r+k-2)!}{(m+k)!(r-1)!}, \quad k=\frac{p'}{1-p'} \tag{11}$$

where $p=\frac{1}{2}$ under H_0 and $p=p'$ under H_1. The ASN function of the rank test appears in the last column of Table 2.

When the ASN functions of the binomial test and the corresponding rank test are compared, it is clear that the later is more efficient than the binomial test as conjectured. When $m=2$, the ASN of the median test is roughly 10% more than that of the rank test, and for $m=3$ about 30% more. When $m=6$ this percentage becomes about 40%.

5. Conclusions

A class of binomial tests and a rank test can be applied for the sequential test of two treatments when subjects are paired in many-to-one ratio. Within the class of binomial tests, the median test appears to be the most efficient in general. However, the rank test is always more efficient than the median test. On the other hand the rank test is not convenient to apply in practice. For example, it would be not practical to employ the usual graphical method for the rank test. Even so, the median test is too inefficient to use when m is large, say $m>4$.

Finally, it is desirable to investigate the adequacy of the mathematical formulation. Questions remain as to the appropriateness of the alternative and robustness of the test under the hypotheses given by (2). Such studies can be carried out at least by Monte Carlo method as Bradley, Martin, and Wilcoxon [1965] have done.

REFERENCES

Armitage, P. [1960], *Sequential Medical Trials,* Blackwell Scientific Publications, Oxford, England.

Bradley, R. A., Martin, D.C., and Wilcoxon, F. [1965], "Sequential Rank Test I, Monte Carlo Studies of the Two-Sample Procedure," *Technometrics* 7, 463—483.

Bradley, R. A., Merchant, S. D., and Wilcoxon, F. [1966], "Sequential Rank Test II, Modified Two-Sample Procedures," *Technometrics* 8, 615—623.

Bross, I. [1952], "Sequential Medical Plans," *Biometrics* 8, 188—205.

Choi, S. C. [1968], "Truncated Sequential Designs for Clinical Trials Based on Markov Chains," *Biometrics* 25.

Lehmann, E. L. [1953], "The Power of Rank Tests," *Ann. Math. Statist.* 24, 23—43.

Wald, A. [1947], *Sequential Analysis*, John Wiley and Sons, Inc., New York, N.Y.

Wilcoxon, F., Rhodes, L.J., and Bradley, R.A. [1963], "Two Sample Grouped Rank Tests with Applications to Screening Experiments," *Biometrics* 19, 58—84.

SUMMARY

A class of sequential binomial tests and a sequential rank test can be applied for testing two treatments when subjects are paired in many-to-one ratio. The efficiency of each test is examined in terms of the average sample number. The binomial tests are much easier and more convenient to apply than the rank test but are not as efficient. Within the class of binomial tests, the median test appears to be the most efficient in general.