한국통계학회논문집 제2권 2호
1995년 12월 pp. 213-220

On Combination of Several Weighted Logrank Tests[1)]

Park, Sang-Gue[2)] and Jeong, Gyu-Jin[3)]

Abstract

We consider a class of the weighted logrank tests and 4 types of weights in this class. We propose a test based on the maximum of 4 weighted logrank statistics and suggest a simulation technique to obtain the p-value of proposed test. It is shown through the simulation studies that the proposed test is robust and has reasonably good powers comparing with the well known efficient tests.

1. Introduction

Suppose that we are interested in detecting the differences of two treatment effects when some of data are censored. Among many researches in this area, the logrank test and Gehan(1965) test seem to be the most popular nonparametric tests. The weighted logrank test has been introduced by Tarone and Ware(1977) who showed that the logrank test and Gehan test belong to same class of the weighted logrank tests. Since Tarone and Ware, many modifications of the logrank test have been done by using different weights(See Fleming and Harrington(1991) and Jeong and Park(1994)).

When we test the equality of two distributions, it is well-known that the logrank test is efficient for exponential distributions and Peto-Prentice test(Peto and Peto(1972), Prentice(1978)) is efficient for logistic distributions. More general results on optimal weights of the weighted logrank tests can be found in Harrington and Fleming(1982). Since the optimal weights depend on the underlying distribution, one might fail to detect the existing differences with the improper weights. Because of difficulty to choose the right weights, it seems to be natural to emphasize the robust aspects of the test. In this point of view, Tarone(1981) proposed a test, which combine the logrank test and Gehan test. This test is based on the maximum of those two test statistics. This is the motivation of our research and we generalize the Tarone's test.

We consider a class of the weighted logrank tests and 4 weights in this class. We propose a test based on the maximum of 4 weighted logrank statistics and suggest a method to find p-value of the proposed test by simulation instead of numerical computations. We further

1) This paper was supported by 1994 NONDIRECTED RESEARCH FUND, Korea Research Foundation.
2) Department of Applied Statistics, Chung-Ang University, Seoul, 156-756, KOREA.
3) Department of Applied Statistics, Hannam University, Taejon, 300-791, KOREA.

examine the empirical powers under some population distributions through the simulation studies.

2. Weighted logrank test

The heuristic development of the logrank statistic used a conditioning argument based on the risk set, which consists of the members at risk of failing just prior to each observed failure time. Let $T_1<T_2<...<T_d$ denote the ordered observed distinct failure times in the sample formed by combining the two groups, whose sample sizes are n_1 and n_2 respectively, and let D_{ik} and $\overline{Y_{ik}}$, (i=1,2; k=1,2,...,d) denote the number of observed failures and number at risk, respectively, in sample i at time T_k. The data at T_k can be summarized as in the following table.

Table 1. Numbers of cases failing and not failing at T_k from the risk set

	samples		
Failure	1	2	Total
Yes	D_{1k}	D_{2k}	D_k
No	$\overline{Y_{1k}}-D_{1k}$	$\overline{Y_{2k}}-D_{2k}$	$\overline{Y_k}-D_k$
Total	$\overline{Y_{1k}}$	$\overline{Y_{2k}}$	$\overline{Y_k}$

Given $\overline{Y_{1k}}$, D_{1k} has a binomial distribution with number of trials $\overline{Y}_{1k}$. As in Fisher's exact test, by conditioning further on D_k, D_{1k} has the hypergeometric distribution. This has conditional mean E_{1k} and variance V_{1k} given by

$$E_{1k} = D_k \frac{\overline{Y_{1k}}}{\overline{Y_k}}$$

$$V_{1k} = D_k \frac{\overline{Y_{1k}}\,\overline{Y_{2k}}}{\overline{Y_k}^2} \frac{\overline{Y_k}-D_k}{\overline{Y_k}-1}.$$

Given the margins each of the d tables at the observed failure times, if we assume that observed minus expected number of failures,

$$\{D_{11}-E_{11},...,D_{1d}-E_{1d}\},$$

are independent over time, then the weighted logrank statistic with weights $W(T_k)$

$$Q_{N,w} = \frac{\sum_{k=1}^{d} W(T_k)(D_{1k}-E_{1k})}{\sqrt{\sum_{k=1}^{d} W(T_k)^2 V_{1k}}} \quad (1)$$

should have approximately a standard normal distribution as $N(=n_1+n_2)$ goes to ∞. Now we can use the statistic $Q_{N,w}$ in the usual way to test the hypothesis $H_0 : F_0(x) = F_1(x)$ vs $H_1 : F_0(x) \geq F_1(x)$ with at least one strict inequality for some real number x.

Among the weighted logrank statistics given by (1), we are specially interested in the class of statistics with weight function $W(s) = \{\hat{S}(s-)\}^{\rho}\{1-\hat{S}(s-)\}^{r}$ for $\rho \geq 0$, $r \geq 0$, where $\hat{S}(t)$ is Kaplan-Meier estimator of survival function based on the combined sample. This class is denoted by $G_N^{\rho r}$. With $r=0$, this $G_N^{\rho r}$ reduces to the G_N^{ρ} class introduced by Harrington and Fleming(1982). Particularly, if we set $\rho=0$ and $r=0$, then we get the logrank test and if we set $\rho=1$ and $r=0$, then we get Peto-Prentice test.

We consider 4 types of test statistics among $G_N^{\rho r}$ class; i.e., $(\rho,r)=\{(0,0), (1,0), (0,1), (1,1)\}$. Let $T_{N,1}$, $T_{N,2}$, $T_{N,3}$, $T_{N,4}$ be the weighted logrank test statistics based on the above 4 weights. We propose a test statistic for testing H_0 against H_1

$$T_{\max} = \max_{1 \leq i \leq 4} T_{N,i}.$$

In order to obtain a powerful test we should use the proper weight function. As we mentioned previously, the logrank test based on $T_{N,1}$ is efficient for exponential distribution and Peto-Prentice test based on $T_{N,2}$ is efficient for logistic distribution. Tests based on $T_{N,3}$ and $T_{N,4}$ are sensitive to detect late and middle differences, respectively. We never know, however, the proper weights unless we know the population distributions and the improper weights cause inefficient tests in some situations(Jeong and Park(1994)). This is the motivation of this research and we want to construct a robust test. We expect that 4 weights considered in $G_N^{\rho r}$ class are sufficient for our purpose.

We now describe the distribution theory needed to implement the distribution of $T_{\max}$. It is well known that the asymptotic distribution of $T_{N,i}$ is normal distribution, but the distribution of $T_{\max}$ is very complicated. One might use Bonferroni inequality or Sidak's inequality, but it is too conservative. Tarone(1981) used the maximum of $T_{N,1}$ and Gehan statistic. Tarone considered the asymptotic distribution of it and provided critical values with the corresponding correlation coefficients. One can use numerical integration techniques for multivariate normal

distribution such as Tarone used the bivariate normal distribution. But it is very boring and complicated. We suggest to use simulation technique to obtain the p-value of test based on $T_{\max}$. This method was used in solving different problems by Hettmansperger and Norton(1987).

Let Σ be the covariance matrix of $\underline{T}_N=(T_{N,1},T_{N,2},T_{N,3},T_{N,4})'$. Since Σ is positive definite, it can be written as

$$\Sigma = BB'$$

where B is a upper triangular matrix and can be calculated by the square root method(See Graybill(1969), p.299). Letting $\underline{Z} \sim$ MVN($\underline{0}, I$) and if N goes to ∞ in such a way that $\frac{n_1}{N}$ remains a constant, we get

$$\underline{T}_N \longrightarrow B\underline{Z},$$

where $\longrightarrow$ means convergence in distribution. Then $\underline{T}_N$ has a limiting MVN($\underline{0}$, BB') distribution. It further follows that, under H_0,

$$\max \underline{T}_N \longrightarrow \max B\underline{Z}.$$

The distribution of $\max B\underline{Z}$ is intractable and can be computed or approximated only in a very few special cases(See Johnson and Kotz(1972), Ch.3). It is quite simple, however, to simulate probabilities of $\max B\underline{Z}$. If the observed value of $\underline{T}_N$ is $\underline{t}$, then the appropriated p-value of a test based on $\max \underline{T}_N$ is given by

$$\Pr(\max B\underline{Z} \geq \max \underline{t}).$$

We now can obtain the p-value of $T_{\max}$ by calculating B with given data and generating standard normal random vector $\underline{Z}$'s.

3. Simulation and Conclusion

3.1 Simulation design

We compare the empirical powers of 6 tests; test 1 with weights (0,0), test 2 with (1,0), test 3 with (0,1), test 4 with (1,1) and test 5 = $\max \underline{T}_N$ with critical values given by Bonferroni inequality, test 6 = $\max \underline{T}_N$ with critical values given by simulated probabilities.

We set population distributions to 4 types;

Type 1: exponential distribution with parameter 0.01 for control group and exponential distribution with parameter θ for treatment group,

Type 2: exponential distribution with parameter 0.01 for control group and mixed exponential distributions with parameter θ if $F_1(t) < 0.6$, 0.01 for elsewhere,

Type 3: exponential distribution with parameter 0.01 for control group and mixed exponential distributions with parameter 0.01 if $F_1(t) < 0.7$, θ for elsewhere,

Type 4: double exponential distribution with parameter 100 for control group and double exponential distribution with θ for treatment group.

Type 2 and 3 are considered for early and late differences, respectively. We also set censoring distributions proportional to population distributions and use 10% censoring rate. Given distribution of the failure time, the parameter λ of censoring distribution can be calculated by solving the following equation,

$$0.1 = P\{X > C\},$$

where X is the failure time random variable with given distribution function $F(\theta)$ and C is the censoring random variable with distribution function $G(\lambda)$. Given θ, we can calculate λ. We also tried 25% censoring rate, but the results are quite similar so that we decide not to put it in considering the space.

We use two sample sizes configurations $n_1 = n_2 = 30$ and $n_1 = n_2 = 50$ and repeat 1000 times to obtain empirical powers under level $\alpha = 0.05$. We also use 1000 repetition to simulate the distribution of max $\underline{T}_N$ at each repetition. The program is written by FORTRAN language and we use IMSL subroutine programs to generate random numbers.

3.2 Simulation results and conclusions

For the case of type 1, test 1 has been known as the efficient test. Table 1 proves that the test 1 is the most powerful as we expected, but test 6 also very powerful.

For the cases of type 2 and type 4, test 2 has been known to be very efficient for the location models and early differences. Table 2 and 4 are also showing this, but test 6 also very powerful.

For the case of type 3, we can expect that test 3 is efficient for late differences since test 3 emphasizes later weights comparing to earlier ones. Table 3 showes this, but test 6 also very powerful.

Test 5 based on Bonferroni inequality seems to be quite attractive because it is very simple to execute. But it turned out to be too conservative under H_0 and did not seem to be recommendable in view of powers as well.

If an experimenter has some informations about the population distributions, use it and choose the proper weighted logrank test. Then he can have very powerful test. But if he has no idea about population distributions and use the particular weighted logrank test, he might fail to detect real differences. For some cases powers barely maintain the significant levels. That's why we need the robust test. Carefully looking the tables, we can find out that

proposed test 6 is quite powerful test comparing with the efficient test in each table.

One can use different weights (ρ,r) or more than 4 types of weights. Since testing procedure can be completed easily if we have normal random number generator, one can consider to do it. However, if one uses more types of weights, we have to expect some loss of powers. We think that considered 4 types of weights are reasonable and practically very useful and should be recommended in view of robustness and powerfulness.

Table 1
Population Distributions : Type 1
Censoring Distributions : Exponential(λ)

θ / λ	sample sizes	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
0.01	30	0.053	0.052	0.055	0.057	0.032	0.057
0.0011	50	0.056	0.056	0.056	0.053	0.034	0.056
0.009	30	0.110	0.103	0.100	0.105	0.059	0.107
0.001	50	0.125	0.113	0.105	0.113	0.061	0.122
0.008	30	0.220	0.186	0.204	0.192	0.146	0.210
0.0009	50	0.278	0.254	0.268	0.268	0.162	0.304
0.007	30	0.370	0.317	0.351	0.346	0.268	0.355
0.0008	50	0.430	0.389	0.415	0.403	0.288	0.425
0.006	30	0.560	0.462	0.520	0.508	0.452	0.548
0.0007	50	0.752	0.658	0.698	0.702	0.490	0.742

Table 2
Population Distributions : Type 2
Censoring Distributions : Exponential(λ)

θ / λ	sample sizes	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
0.01	30	0.044	0.045	0.059	0.046	0.029	0.055
0.0011	50	0.059	0.06	0.058	0.060	0.037	0.056
0.008	30	0.079	0.108	0.060	0.063	0.051	0.090
0.0009	50	0.103	0.157	0.058	0.081	0.070	0.132
0.006	30	0.159	0.256	0.066	0.137	0.136	0.233
0.0007	50	0.213	0.405	0.068	0.181	0.213	0.387
0.004	30	0.394	0.629	0.120	0.329	0.405	0.599
0.0004	50	0.530	0.832	0.125	0.450	0.647	0.811

Table 3
Population Distributions : Type 3
Censoring Distributions : Exponential(λ)

θ / λ	sample sizes	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
0.01	30	0.053	0.052	0.055	0.057	0.032	0.057
0.0011	50	0.056	0.056	0.056	0.053	0.034	0.056
0.008	30	0.134	0.070	0.182	0.126	0.092	0.152
0.0009	50	0.144	0.068	0.236	0.150	0.110	0.172
0.006	30	0.218	0.096	0.430	0.216	0.256	0.240
0.0007	50	0.368	0.106	0.622	0.328	0.328	0.468
0.004	30	0.368	0.108	0.706	0.329	0.456	0.566
0.0004	50	0.532	0.162	0.868	0.467	0.722	0.699

Table 4
Population Distributions : Type 4
Censoring Distributions : Double Exponential(λ)

θ / λ	sample sizes	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
100.0	30	0.054	0.055	0.060	0.061	0.032	0.057
102.4	50	0.056	0.056	0.056	0.055	0.034	0.057
100.3	30	0.227	0.263	0.162	0.219	0.166	0.238
102.7	50	0.304	0.320	0.166	0.264	0.277	0.290
100.5	30	0.445	0.519	0.238	0.397	0.344	0.480
102.9	50	0.587	0.684	0.304	0.566	0.577	0.644
100.7	30	0.635	0.736	0.338	0.584	0.579	0.689
103.1	50	0.801	0.828	0.576	0.742	0.733	0.803

References

[1] Gehan, E.A.(1965). A generalized Wicoxon test for comparing arbitrarily single censored sample, *Biometrika*, Vol. 52, 203-23.

[2] Fleming, T.R. and Harrington, D.P.(1991). *Counting processes and Survival analysis*, John Wiley and Son, Inc. New York.

[3] Graybill, F.A.(1969). Introduction to matrices with applications in statistics, Wadworth Publishing Company, Inc. Belmont, California.

[4] Harrington, D.P. and Fleming, F.A.(1982). A class of rank test procedures for censored survival data, *Biometrika,* Vol. 69, 133-43.

[5] Hettmansperger,T.P. and Norton, R.M.(1987). Tests for patterned alternatives in k-sample problems, *Journal of the American Statistical Association*, Vol. 82, 292-99.

[6] Jeong, G. and Park, S.(1994). Weighted logrank test for late differences, *Korean Journal of Applied Statistics*, Vol. 7, No. 2, 79-88.

[7] Johnson, N.L. and Kotz, S.(1972). *Distributions in statistics: Continuous multivariate distributions*, John Wiley and Son, Inc. New York.

[8] Peto, R. and Peto, R.(1972). Asymptotically efficient rank invariant test procedures (with discussion), *Journal of the Royal Statistical Society*, A, 135, 185-206.

[9] Prentice, R.L.(1978). Linear rank tests with right censored data, *Biometrika*, Vol. 65, 169-79.

[10] Tarone, R.E.(1981). On the distribution of the maximum of the logrank statistic and the modified Wilcoxon statistic, *Biometrics*, Vol. 37, 79-85.

[11] Tarone, R.E. and Ware, J.(1977). On distribution-free tests for equality of survival distributions, *Biometrika*, Vol. 64, 156-60.