On Best Precedence Test when Data are subject to Unequal Patterns of Censorship[+]

Tai-Kyoo Kim · Sang-Gue Park
Dept. of Applied Statistics, Hannam University

Abstract

Nonparametric tests for comparing two treatments when data are subject to unequal patterns of censorship are discussed. Best precedence test proposed by Slud can be viewed as a nice alternative test comparing with weighted log-rank tests, not to mention the advantage of short experimental period. This research revises some missing parts of Slud's test and examines the asymptotic power of it under the nonproportional hazard alternatives through the simulation. The simulation studies show best precedence test has reasonable power in the sense of robustness under nonproportional hazard alternatives and could be recommended at such situation.

1. Introduction

Researchers are frequently interested in testing the equality of differences of two treatments from the properly controlled experiment. When one collects the data from the experiments(especially coding lifetimes of experimental units), it is often unable to observe the data fully. It's so-called "censored data". The way to analyze the collected data for the case of censored data is little different from the usual way.

The well-known testing procedures for two-sample problem with censored data are log-rank test and Gehan(1965) test. Fleming and Harrington(1991) treated excellently about weighted log-rank tests including the above two tests. Precedence test based on preselected quantiles, which seems to be little different from weighted log-rank test, also used widely in lifetime experiments since it

+ 이 연구는 1993년 학술진흥재단 지방대육성과제 연구기금으로 연구되었음.

could shorten the experimental duration if data are obtained sequentially. Recently Slud(1992) considered best precedence test, which is optimizing power with respect to local Lehmann alternatives. Slud mentioned the advantage of his test over the log-rank test by examining asymptotic efficiencey. The present paper is to discuss the Slud's test procedure and compare the power of it with the well-known two-sample rank tests such as log-rank test and Gehan test under various probability models through the simulation studies.

2. Assumptions and Notation

In order to avoid notational difficulties, we use similar notations given at Slud (1992)'s. Let $X_1, X_2, \cdots, X_m$ denote independent and identically distributed sample of the control group latent survival times, with common continuous distribution function $F_0(t)$, and let $C_1, C_2, \cdots, C_m$ denote independent sample of latent censoring times for the same experimental units, with distribution function $L_0(t)$. The control group data consist of $T_i = \min(X_i, C_i)$ and $\delta_i = I_{[X_i \le C_i]}$ for $i=1, 2, \cdots, m$. The corresponding observations for the experimental group, assumed to be independent of the control group data, are $U_j = \min(Y_j, D_j)$ and $\varepsilon_j = I_{[Y_j \le D_j]}$ for $j=1, 2, \cdots, n$, where the latent survival times Y_j's have continuous distribution function $F_1(t)$, and the latent control group censoring times D_j's have distribution function $L_1(t)$. We assume the censoring distributions are different in samples, which is more reasonable in practical cases comparing equal censoring model of Koziol-Green model(1975). Let the cumulative hazard function associated with $F_k(t)$ for $k=0, 1$ be denoted by $\Lambda_k(t)$. Denote by $N_k(t)$ the number of observed group-k failures up to time-on study t; by $r_k(t)$ the corresponding number at risk at time t in group k; and let

$$\bar{F}_k(t) = 1 - \prod_{s \le t}\left(1 - \frac{\Delta N_k(s)}{r_k(s)}\right),\ k=0, 1, \qquad (1)$$

so-called Kaplan-Meier estimator of $F_k(t)$. For any right-continuous nondecriasing function G, define

$$G^{-1}(x) = \inf\{\, t \in R : G(t) > t \,\}, \qquad (2)$$

we call the quantities $\bar{F}_k^{-1}(t)$ "Kaplan-Meier quantile".

3. Best Precedence test

A precedence test is a two-sample test based on order statistics with rejection region of the form $X_{(j)} < Y_{(k)}$ for the test of $H_0: F_0 = F_1$ versus the stochastically ordered alternative $H_1: F_0 \geq F_1$. Such test has been studied and proved the usefulness since Nelson(1963), Eilbott and Nadler(1965), Shorack(1967). Recently Lin and Sukhatme(1992) discussed the choice of the proper precedence test.

For the case of censored data, one might be interested in generalizing it. However, we could not use the rejection region of the above form since we could not observe the X's and Y's fully. In rigt censored data, one can use the rejection region $[\, \overline{F}_0^{-1}(r) < \overline{F}_1^{-1}(s)\,]$, for specified quantiles r and $s \in (0, 1)$. For given α, one might have many α-level precedence tests.

If researchers have pre-chosen r value and α-level, one can easily perform the test with the following theorem. The proof of the theorem would help understand the generalization of precedence test.

Theorem 1. The test of $H_0: F_0 = F_1$ against $H_1: F_0 \geq F_1$, which rejects when $[\overline{F}_0^{-1}(r) < \overline{F}_1^{-1}(s)]$, where

$$s = r + z_\alpha \sigma^*(r), \tag{3}$$

with $z_\alpha = \Phi^{-1}(\alpha)$, ($\Phi(x)$: cumulative standard normal distribution) and

$$\sigma^*(r) = [1-r]\,[\int_0^{F_0^{-1}(r)} \frac{dN_1(t)}{r_1(t)^2} + \int_0^{F_0^{-1}(r)} \frac{dN_0(t)}{r_0(t)^2}\,]^{1/2}. \tag{4}$$

Proof:

From Gastwirth and Wang(1988)'s theorem 1, we can obtain the asymptotic distribution of $\overline{F}_1(\overline{F}_0^{-1}(r))$, which could be viewed as an asymptotically equivalent test of precedence one ; i.e.,

$$\begin{aligned} &P[\overline{F}_1(\overline{F}_0^{-1}(r)) < s] \\ &= P[\overline{F}_1(\overline{F}_0^{-1}(r)) < \overline{F}_1(\overline{F}_1^{-1}(s))] \\ &= P[\overline{F}_0^{-1}(r) < \overline{F}_1^{-1}(s)]. \end{aligned}$$

The asymptotic null variance of $\overline{F}_1(\overline{F}_0^{-1}(r))$ is given by

$$\sigma(r)^2 = [1-r]^2 \int_0^{F_0^{-1}(r)} ((1-L_0(x))^{-1} + \frac{1-\lambda}{\lambda}(1-L_1(x))^{-1})\, \frac{d\Lambda_0(x)}{1-F_0(x)}, \tag{5}$$

where $N=m+n$ and $\frac{m}{N}$ tends to $\lambda(0<\lambda<1)$ as n, m tend to infinity. When the censoring patterns are different in two samples, $\sigma(r)^2$ can be estimated with replacing Λ_0 and $(1-L_i(x))(1-F_i(x))$ (for $i=0, 1$) by Nelson(1969)'s estimator and risk set $r_i(x)$. They showed $\bar{F}_1(\bar{F}_0^{-1}(r))$'s asymptotic normality and it completes the theorem.

From this theorem one might have many tests depending on r for a given α level. Gastwirth and Wang(1988) calculated the optimal r values for the case of exponential and double-exponential population assuming Koziol-Green model (1975). For example, if one choose $r=0.5$ and a proper type-I error level, one can obtain the corresponding s values by using theorem 1.
Then one can be ready to perform the precedence test with the rejection region $[\bar{F}_0^{-1}(r)<\bar{F}_1^{-1}(s)]$.

Slud(1992) further developed the choice of proper r value without specifying any parametric models. Slud proposed a criterion to choose r and s, whose test has greatest asymptotic power against contiguous Lehmann alternatives. Under the contiguous alternatives

$$H_{1A}:\Lambda_1(t)=\int_0^t(1+\frac{c(s)}{\sqrt{n}})dN_0(s)$$

(under usual regularity conditions), this test has asymptotic power $\Phi(z_\alpha-Eff)$, with efficacy Eff given by the formula

$$\int_0^{-\ln(1-r^*)}c(\Lambda_0^{-1}(u))\frac{du}{\sigma(r^*)}, \tag{6}$$

where $c(\ )$ is a bounded function. Slud's idea is to choose r^* to maximize Eff.

Slud(1992) provided the asymptotic results of best precedence test inhis theorem 2. But the theorem 2 turns out to be insufficient because his notations were so confused and Slud made mistake in transforming $\sigma(r)$ in (5) to a convenient form by hiring counting process theory(Fleming and Harrington(1991); i.e., an estimator of $\sigma(r)^2$ should be changed like

$$\sigma^*(r)^2=[1-\hat{r}^*]^2[\int_0^{\bar{F}_0^{-1}(\hat{r}^*)}\{\frac{dN_1(t)}{r_1(t)^2}+\frac{dN_0(t)}{r_0(t)^2}\}]$$

instead of $[1-\hat{r}^*]^2[\int_0^{\bar{F}_0^{-1}(\hat{r}^*)}\{\frac{dN_1(t)}{r_1(t)}+\frac{dN_0(t)}{r_0(t)}\}]$. We do not give the proof of the following theorem 2, since other parts of the proof are correct.

Theorem 2. If there is a value r^* uniquely maximizing the *Eff* on [0, 1], then this r^* is consistently estimated by the maximzer $\hat{r}^*$ on [$\hat{r}_*$, 0.797] of

$$\ln^2(1-x) \,/\, \int_0^{F_1^{-1}(x)} \{ \frac{dN_1(t)}{r_1(t)^2} + \frac{dN_0(t)}{r_0(t)^2} \}, \tag{7}$$

where

$$\hat{r}_* = \inf\{ x : \frac{mn}{(m+n)} [\frac{1}{r_1(\bar{F}_0^{-1}(x))} + \frac{1}{r_0(\bar{F}_0^{-1}(x))}] \geq 2 \}. \tag{8}$$

The best precedence test with r replaced by r^* and s replaced by

$$r^* - z_\alpha [1-\hat{r}^*] [\int_0^{\bar{F}_0^{-1}(\hat{r}_*)} \{ \frac{dN_1(t)}{r_1(t)^2} + \frac{dN_0(t)}{r_0(t)^2} \}]^{1/2}. \tag{9}$$

From this theorem one can feel that the power of best precedence test improves a lot comparing with any precedence test based on arbitrarily selected quantiles under Lehmann alternatives. But it should be noted that best precedence test might be longer experimental duration since one have to observe values on [$\hat{r}_*$, 0.797].

Best precedence test obtained from theorem 2 is definitely less powerful than the log-rank test under the proportional hazard alternatives. However, this expects to be more reasonable(robust) test under nonproportional hazard alternatives(log-rank test is known as less efficient under nonproportional hazard alternatives). We examine this in section V.

4. Illustrated example

This section provides an examlpe to help understand the best precedence test. This example shows the best precedence testing procedures by using theorem 2 in section III and compares the well-known two-sample rank tests such as log-rank test and Gehan test.

The following data set is generated from exponential random numbers since we could not find a suitable example. Suppose we want to compare two devices' lifetimes under random censorship with the significance level of 0.05. The hypothesis is $H_0 : F_0 = F_1$ against $H_1 : F_0 \geq F_1$.

Control sample(X's are from F_0)

1.56 $^{+}$, 10.4, 17.33, 22.35, 23.17, 34.24, 35.31, 45.46, 46.84, 48.42, 53.05, 56.9, 60.11, 64.41, 70.66, 70.88, 71.71, 87.21, 92.80, 98.01 $^{+}$, 109.71, 111.46, 120.59, 137.69, 140.94, 183.16, 258.77, 269.47, 274.56, 294.7

Treatment sample(Y's are from F_1)

1.42, 6.13, 7.66, 9.21, 9.36, 11.43, 15.46, 17.61 $^{+}$, 20.82, 22.35, 22.66, 26.09 $^{+}$, 27.91, 39.02, 43.6, 47.16, 49.36, 53.21, 64.45, 80.63, 107.88, 133.68, 136.32, 162.82, 164.72, 189.56 $^{+}$,210.4, 248.47, 401.17, 567.58

(+ means censored data)

From theorem 2, we obtain $\hat{r}_* = 0.379$. Thus we find a value $\hat{r}^*$ which is maximizing $\ln^2(1-x) / \int_0^{\bar{F}_0^{-1}(x)} \{ \frac{dN_1(t)}{r_1(t)^2} + \frac{dN_0(t)}{r_0(t)^2} \}$ on $x \in [0.379, 0.797]$ through numerical method by increasing 0.01. Then we get $\hat{r}^* = 0.734$ and the corresponding $s = 0.56$. Now we are ready to use the best precedence test. That is, if $[\bar{F}_0^{-1}(0.734) < \bar{F}_0^{-1}(0.56)]$, then we reject H_0. We can easily obtain Kaplan-Meier quantiles defined in (2) from Kaplan-Meier estimates defined in (1) so that $\bar{F}_0^{-1}(0.734) = 120.59$, $\bar{F}_0^{-1}(0.56) = 138.68$. The null hypothesis H_0 is rejected at the level of significance 0.05. The log-rank test statistic is obtained -2.2 so that H_0 is also rejected by comparing to -1.645. However, Gehan test statistic is obtained -1.128 so that H_0 is not rejected at the level of siginificance 0.05.

5. Simulation Results and Conclusions

When we consider the two-sample problem with censored data, log-rank test or Gehan test are very popular. In this section we are comparing best precedence test with log-rank test and Gehan test under various probability models through the simulation studies. The study compares the estimated powers for various stochastically ordered alternatives generated by exponential distributions and several types of censoring proportions(10%, 30%, 50%). Censoring patterns are assumed to be followed by exponential law. Sample sizes are set to be equal at 30 in considering safe asymptotic normality use. Random numbers are generated by using IMSL subroutines and the results are based on 5,000 replications using Fortran program and performed at significance level 5% and 1%. The results with 1% are abbreviated on account of space consideration.

〈 Table 1 〉

Control sample : generated by exponential with mean 100
Treatment sample : generated by exponential with mean θ

θ	Censoring rate	Log-rank	Gehan	Precedence
105	10%	0.084	0.068	0.059
	30%	0.072	0.077	0.066
	50%	0.067	0.070	0.068
110	10%	0.091	0.082	0.074
	30%	0.091	0.076	0.080
	50%	0.092	0.095	0.103
120	10%	0.166	0.155	0.133
	30%	0.158	0.143	0.114
	50%	0.147	0.139	0.117

〈 Table 2 〉

Control sample : generated by exponential with mean 100
Treatment sample : generated by mixed exponential with
mean 100 if $F_1(x) \le 0.3$, θ otherwise

θ	Censoring rate	Log-rank	Gehan	Precedence
105	10%	0.072	0.073	0.063
	30%	0.079	0.070	0.066
	50%	0.063	0.059	0.059
110	10%	0.101	0.088	0.085
	30%	0.089	0.080	0.080
	50%	0.063	0.057	0.083
120	10%	0.157	0.125	0.116
	30%	0.116	0.096	0.092
	50%	0.104	0.079	0.068

〈 Table 3 〉

Control sample : generated by exponential with mean 100
Treatment sample : generated by mixed exponential with
mean 100 if $F_1(x) \le 0.5$, θ otherwise

θ	Censoring rate	Log-rank	Gehan	Precedence
105	10%	0.076	0.062	0.077
	30%	0.061	0.051	0.063
	50%	0.063	0.046	0.055
110	10%	0.087	0.065	0.086
	30%	0.081	0.056	0.054
	50%	0.062	0.055	0.063
120	10%	0.136	0.085	0.099
	30%	0.094	0.077	0.074
	50%	0.079	0.050	0.053

〈 Table 4 〉

Control sample : generated by exponential with mean 100
Treatment sample : generated by mixed exponential with
mean θ if $F_1(x) \le 0.3$, 100 otherwise

θ	Censoring rate	Log-rank	Gehan	Precedence
105	10%	0.052	0.051	0.048
	30%	0.052	0.047	0.057
	50%	0.059	0.066	0.056
110	10%	0.064	0.055	0.054
	30%	0.071	0.060	0.048
	50%	0.054	0.061	0.063
120	10%	0.072	0.088	0.059
	30%	0.066	0.079	0.067
	50%	0.051	0.078	0.075

〈 Table 5 〉

Control sample : generated by exponential with mean 100
Treatment sample : generated by mixed exponential with mean θ if $F_1(x) \le 0.5$, 100 otherwise

θ	Censoring rate	Log-rank	Gehan	Precedence
105	10%	0.067	0.065	0.060
	30%	0.058	0.052	0.060
	50%	0.063	0.069	0.072
110	10%	0.074	0.073	0.065
	30%	0.063	0.070	0.081
	50%	0.051	0.064	0.068
120	10%	0.070	0.088	0.062
	30%	0.071	0.083	0.072
	50%	0.075	0.083	0.114

From the sumulation results, one can easily see that there are no optimal test. First, we can observe that log-rank test is the most powerful under exponential distributions as known. Secondly, we consider the power comparison under nonproportional hazard alternatives given table 2 through 5. Table 2 and 3 contain power estimates at mixed exponential distribution, which changes hazard rate at later part of the distribution. Since log-rank test has been known as sensitive test at later difference, we can expect log-rank test is still showing good power. However, best precedence test also maintains reasonably good powers. Table 4 and 5 contain power estimates at another mixed exponential distribution, which changes hazard rate at early part of the distribution. Gehan test also has been known as sensitive test at early difference and we expect that Gehan test works better than others. Simulation studies show that best precedence test is very comparable.

It might not be wise that we strongly advocate best precedence test over others from this limited simulation studies. But this simulation studies and Slud's efficiency examination prove that best precedence test is at least considerable not to mention short experimental duration, especially one ecpect nonproportional hazard rates in groups. Furthermore we expect to generalize this best type of precedence test to the location model.

Acknowledgement

Authors wish to thank the referees and the editors for a number of helpful comments which improved the presentations of our results.

References

[1] Eilbott, J. and Nadler.(1965), "On the precedence life testing," *Technometrics*, Vol. 7, pp. 359−377.

[2] Fleming, T. R. and Harrington, D. P.(1991), *Counting processes and survival analysis*, John Wiley & Sons, Inc.

[3] Gastwirth, J. L. and Wang, J.(1988), "Control percentile test procedures for censored data," *J. Statist. Plann. Infer*, Vol. 18, pp. 267−276.

[4] Gehan, E. A (1965), "A generalized Wilcoxon test for comparing arbitraily single-censored sample," *Biometrika*, Vol. 52, pp. 203−223.

[5] Koziol, J. A. and Green, S. B.(1976), "A Cramer-Von Mises statistic for randomly censored data," *Biometrika*, Vol. 63, pp. 465−474.

[6] Lin, C. and Sukhatme, S.(1992), "On the choice of precedence tests," *Comm. Statist.−Theory Meth. A.*, Vol. 21, No. 10, pp. 2949−2968.

[7] Nelson, L. S.(1963), "Tables for precedence test," *Technometrics*, Vol. 5, pp. 154−158.

[8] Nelson, W.(1969), "Hazard plotting for incomplete failure data," *J. Qual. Technol.*, Vol. 1, pp. 27−52.

[9] Shorak, R.(1967), "On the power of precedence life tests," *Technometrics*, Vol. 9, pp. 154−158.

[10] Slud, E. V.(1992), "Best precedence tests for censored data," *J. Statist. Plann. Infer.*, Vol. 31 pp. 283−293.