 Vol 32: 49–57, 2005

Generalized Log-Rank Tests for Interval-Censored Failure Time Data

JIANGUO SUN and QIANG ZHAO

Department of Statistics, University of Missouri

XINGQIU ZHAO

Department of Mathematics and Statistics, McMaster University

ABSTRACT. Several non-parametric test procedures have been proposed for incomplete survival data: interval-censored failure time data. However, most of them have unknown asymptotic properties with heuristically derived and/or complicated variance estimation. This article presents a class of generalized log-rank tests for this type of survival data and establishes their asymptotics. The methods are evaluated using simulation studies and illustrated by a set of real data from a cancer study.

Key words: asymptotic distribution, clinical trials, interval-censoring, log-rank test, survival comparison

1. Introduction

This paper discusses non-parametric comparison of survival functions based on incomplete survival data: interval-censored failure time data (cf. Li *et al.*, 1997; Sun, 1998; Pan, 2000). By interval-censored data, we mean that the survival time of interest is observed only to belong to an interval instead of being exactly known or right-censored as usually assumed (cf. Li, 2003). One field in which interval-censored data often occur is observational or follow-up studies where patients are not continuously under observation. In this case, only the status about the occurrence of a certain event is observed at observation times, rather than the occurrence time of the event. One such example from a cancer study is provided in Finkelstein (1986) and will be discussed below in more details. Another field that commonly produces interval-censored failure time data is tumorgenicity experiments (cf. Lagakos & Louis, 1988). In this case, it is usually the case that the survival time of interest is either left-censored or right-censored, a special case of interval-censored data.

Survival comparison is usually one of main goals in survival studies. For the problem, when right-censored failure time data are available, a number of well-established methods have been developed (cf. Fleming & Harrington, 1991; Kalbfleisch & Prentice, 2002). For the case of interval-censored failure time data, several authors have discussed the problem. For example, Peto & Peto (1972) considered the two-sample comparison problem under the Lehmann-type alternatives $G_2(t) = G_1^{\theta}(t)$, where G_1 and G_2 are survival functions corresponding to the two different samples and θ is a parameter. In this case, the comparison problem reduces to testing $\theta = 0$ and they suggested using the score test, which they referred to as the log-rank test. Assuming the proportional hazards model, a special case of Lehmann-type alternatives, Finkelstein (1986) investigated the general k-sample comparison problem. For the problem, she also suggested applying the score test for testing regression parameters equal to zero. Following Finkelstein (1986), Sun (1996) studied the same problem without assuming the proportional hazards model and developed a non-parametric test using the idea behind the log-rank test for right-censored data (cf. Kalbfleisch & Prentice, 2002). Other existing test procedures for interval-censored data can be found in Sun (1998).

A main drawback for most of the existing test procedures for survival comparison based on interval-censored failure time data is that they are ad hoc methods and their properties are unknown. In particular, the variance estimation of the test statistics was usually heuristically derived and/or is complicated. In this paper, we propose a class of non-parametric tests for the problem; the proposed tests are generalizations of the log-rank test given in Peto & Peto (1972). The tests are presented in section 2, which also discusses their relationship with some existing tests. In section 3, the asymptotic distributions of the proposed test statistics are derived and section 4 reports some simulation results for evaluating the proposed methodology. They suggest that the approach works well for the situations considered. Also in section 4 we apply the approach to a real set of interval-censored data from a cancer study. Section 5 contains some concluding remarks.

2. Generalized log-rank tests

Consider a survival study that involves n independent subjects from k different populations. Let T_i denote the survival time of interest for subject i and n_l the number of subjects from population l with survival function $G_l(t)$ and distribution function $F_l(t) = 1 - G_l(t), i = 1, \ldots, n, l = 1, 2, \ldots, k$, where $n_1 + \cdots + n_k = n$. Also let x_i be the $k \times 1$ vector of treatment indicators associated with subject i whose lth element is equal to 1 if it is from population l, and zero otherwise. Suppose that for subject i, we observe $\{x_i, U_i, V_i, \Delta_i = I(T_i \le U_i), \Gamma_i = I(U_i < T_i \le V_i)\}$, where U_i and V_i are non-negative random variables independent of T_i such that $U_i < V_i$ with probability 1, $i = 1, \ldots, n$. Define

$$(L_i, R_i] = \begin{cases} (0, U_i], & T_i \le U_i\ , \\ (U_i, V_i], & U_i < T_i \le V_i, \\ (V_i, \infty), & T_i > V_i \end{cases}$$

to be the interval to which T_i is observed to belong. Our goal is to test the hypothesis $H_0 : G_1(t) = \cdots = G_k(t)$.

Let $G_0(t)$ denote the common survival function under H_0 and $\hat{G}_n(t)$ the non-parametric maximum likelihood estimator of it, whose determination will be discussed below. To test H_0, we propose the following test statistic

$$U_\xi = \sum_{i=1}^{n} x_i \frac{\xi\{\hat{G}_n(L_i)\} - \xi\{\hat{G}_n(R_i)\}}{\hat{G}_n(L_i) - \hat{G}_n(R_i)},$$

where ξ is a known function over (0,1) and will be defined more formally in the next section. Obviously, different ξ can be used and will yield different test statistics in practice. The above statistics were motivated by Peto & Peto (1972), who studied U_ξ with $\xi(x) = x \log x$ for the case of $k = 2$ and referred it the log-rank test statistic. Rabinowitz *et al.* (1995) considered similar statistics for regression analysis of interval-censored data under the accelerated failure time model. In addition to complex covariance estimation, they did not give the proof of the asymptotic distribution of the statistics.

To see the relationship between U_ξ and some existing test statistics, let $0 = s_0 < s_1 < \cdots < s_m = \infty$ denote the ordered distinct time points in $\{L_i, R_i;\ \ i = 1, \ldots, n\}$ and define $\alpha_{ij} = I((s_{j-1}, s_j] \subseteq (L_i, R_i])$, where I is the indicator function. Also define $\hat{p}_j = \hat{G}_n(s_j)/\hat{G}_n(s_{j-1})$ and $\hat{g}_j = \hat{G}_n(s_{j-1}) - \hat{G}_n(s_j)$, $j = 1, \ldots, m$. Then the score test statistic for H_0 proposed by Finkelstein (1986) under the proportional hazards model has the form

$$U_F = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} \left\{ \frac{\log \hat{p}_j \sum_{k=j}^{m} \alpha_{ik} \hat{g}_k}{\sum_{r=1}^{n} \alpha_{ir} \hat{g}_r} - \left(\frac{\log \hat{p}_j}{1 - \hat{p}_j} \right) \frac{\alpha_{ij} \hat{g}_j}{\sum_{r=1}^{n} \alpha_{ir} \hat{g}_r} \right\}.$$

It can be shown that the inside term in U_F can be rewritten as

$$\frac{\hat{G}_n(L_i) \log \hat{G}_n(L_i) - \hat{G}_n(R_i) \log \hat{G}_n(R_i)}{\hat{G}_n(L_i) - \hat{G}_n(R_i)} .$$

Thus U_F is equal to U_ξ with $\xi(x) = x \log x$. It can also be proved that the test statistic given in Sun (1996), which has the form

$$U_S = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} \left(\frac{d_{ij} - n_{ij} d_j}{n_j} \right),$$

is asymptotically equivalent to U_ξ with $\xi(x) = x \log x$, where

$$d_{ij} = \frac{\alpha_{ij} \hat{g}_j}{\sum_u \alpha_{iu} \hat{g}_u}, \quad n_{ij} = \sum_{r=j}^{m} \frac{\alpha_{ir} \hat{g}_r}{\sum_u \alpha_{iu} \hat{g}_u}, \quad d_j = \sum_{i=1}^{n} \frac{\alpha_{ij} \hat{g}_j}{\sum_r \alpha_{ir} \hat{g}_r}, \quad n_j = \sum_{u=j}^{m} \sum_{i=1}^{n} \frac{\alpha_{iu} \hat{g}_u}{\sum_r \alpha_{ir} \hat{g}_r}.$$

In the above, we need to determine $\hat{G}_n(t)$. The simplest method for this, which is used below in simulation studies and the example, is perhaps the direct application of the Turnbull's self-consistency algorithm (cf. Turnbull, 1976). An alternative is to use, for example, the approach given by Gentleman & Geyer (1998); Sun (1998) gave a brief review of other available algorithms.

3. Asymptotic distributions

In this section, we will establish the asymptotic distribution of U_ξ. Let $\eta(x) = 1 - \xi(1 - x)$ and assume that $\lim_{x \to 0} \eta(x) = \lim_{x \to 1} \eta(x) = c_0$, where c_0 is a constant. Also let H and h denote the distribution and density functions of (U_i, V_i), respectively, $F_0(t) = 1 - G_0(t)$ and $\hat{F}_n(t) = 1 - \hat{G}_n(t)$. Then we can rewrite U_ξ as

$$U_\eta = \sum_{i=1}^{n} x_i \left[\Delta_i \frac{\eta\{\hat{F}_n(U_i)\} - c_0}{\hat{F}_n(U_i)} + \Gamma_i \frac{\eta\{\hat{F}_n(V_i)\} - \eta\{\hat{F}_n(U_i)\}}{\hat{F}_n(V_i) - \hat{F}_n(U_i)} + (1 - \Delta_i - \Gamma_i) \frac{c_0 - \eta\{\hat{F}_n(V_i)\}}{1 - \hat{F}_n(V_i)} \right].$$

Let λ_2 and v_2 denote the Lebesgue measure on R^2 and counting measure on the set $\{(0,1),(1,0),(0,0)\}$, respectively. Define

$$q_{F_0,H}(u, v, \delta, \gamma) = h(u, v)\{F_0(u)\}^{\delta}\{F_0(v) - F_0(u)\}^{\gamma}\{1 - F_0(v)\}^{1-\delta-\gamma}$$

with respect to $\lambda_2 \otimes v_2$, which is the density function of $(U_i, V_i, \Delta_i, \Gamma_i)$. Also define $dQ_0 = q_{F_0,H} d(\lambda_2 \otimes v_2)$,

$$Q_n(u, v, \delta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{(U_i, V_i) \le (u,v), (\Delta_i, \Gamma_i) = (\delta, \gamma)\}}$$

and

$$K_0(u, v, \delta, \gamma) = \delta \frac{\eta\{F_0(u)\} - c_0}{F_0(u)} + \gamma \frac{\eta\{F_0(v)\} - \eta\{F_0(u)\}}{F_0(v) - F_0(u)} + (1 - \delta - \gamma) \frac{c_0 - \eta\{F_0(v)\}}{1 - F_0(v)}.$$

We assume that the regularity conditions given in Groeneboom and Wellner (1992) for the strong consistency of $\hat{F}_n$ hold. Also we assume that $F_0(t)$ has a support in $[0,M]$ with a continuous density function and that there exist $0 < \delta_0, \varepsilon_0 < M/2$ and $M_0 < M$ such that $\Pr(U < \delta_0) = 0$, $\Pr(U + \varepsilon_0 \le V \le M_0) = 1$, $0 < F_0(\delta_0) < F_0(M_0) < 1$ and $\min_{\delta_0 \le t \le M_0} - \varepsilon_0[F_0(t + \varepsilon_0) - F_0(t)] \neq 0$, where M is a constant. The asymptotic distribution of U_ξ is given in the following theorem.

Theorem 1

Suppose that the above assumptions hold and η is a bounded Lipschitz function on $[a,1]$ for any finite positive number a. Also suppose that as $n \to \infty$, $n_l/n \to p_l$, where $0 < p_l < 1$ and $p_1 + p_2 + \cdots + p_k = 1$. Then under H_0 and as $n \to \infty$, $U_\eta/\sqrt{n}$ has an asymptotic normal distribution with mean zero and covariance matrix $\Sigma = (\sigma_{lr})_{k\times k}$, where

$$\sigma_{lr} = \begin{cases} p_l(1-p_l)Q_0(K_0^2), & \text{if } l = r, \\ -p_l p_r Q_0(K_0^2), & \text{otherwise.} \end{cases}$$

The proof of the above theorem is sketched in the appendix. Let $\hat{K}_n$ denote K_0 with F_0 replaced by $\hat{F}_n$. Then it can be easily seen that the covariance matrix Σ can be consistently estimated by $\hat{\Sigma} = (\hat{\sigma}_{lr})_{k\times k}$, where

$$\hat{\sigma}_{lr} = \begin{cases} \dfrac{n_l(n-n_l)}{n^2} Q_n(\hat{K}_n^2), & \text{if } l = r, \\ \dfrac{-n_l n_r}{n^2} Q_n(\hat{K}_n^2), & \text{otherwise.} \end{cases}$$

Let U_0 denote the first $k-1$ components of U_η and $\hat{\Sigma}_0$ the matrix by deleting the last row and column of $\hat{\Sigma}$. Then the hypothesis H_0 can be tested by using the statistic $\chi_0 = U_0^t \hat{\Sigma}_0^{-1} U_0/n$, which has asymptotically the χ^2 distribution with $(k-1)$ degrees of freedom. This is because the sum of the components of U_η is equal to zero.

4. Numerical results

To assess the finite sample performance of the proposed approach, simulation studies were conducted with a focus on the size and power of the test procedure and the normal approximation to the distribution of the test statistic U_ξ. In the simulation, we considered the two-sample comparison problem and generated the survival times T_i's from the exponential distribution with mean $\exp(\alpha + \beta x_i)$, where α and β are constants and $x_i = 0$ or 1. For censoring intervals, we first generated U_1 and U_2 independently from the uniform distributions $U(0,\theta_1)$ and $U(0,\theta_2)$, respectively. Here θ_1 and θ_2 are positive constants chosen to give the proper percentages of left-censored, interval-censored and right-censored observations in simulated data. Then U and V were defined as the nearest integer to U_1 and the maximum of the nearest integer to $U_1 + U_2$ and $U + 1$, respectively. The results reported below are based on $n_1 = n_2 = 100$, $\alpha = 2$ and 5000 replications.

For function ξ in the simulation, we used the class of functions $\xi(x) = (x \log x)x^\rho(1-x)^\gamma$ motivated by the weight functions commonly used for weighted log-rank test statistics for right-censored data (Fleming & Harrington, 1991), where ρ and γ are some constants. Table 1 presents the empirical sizes and powers of the proposed test based on simulated interval-censored data for different values of β. In the table, we considered four different situations in terms of the percentages of left-censored, interval-censored and right-censored observations in the data, which are given in the first column of the table. The second and third columns give the values of parameters used in $\xi(x)$. For comparison, we also calculated and included in the table the empirical sizes and powers of the parametric score test for $\beta = 0$ assuming that

Table 1. *Estimated powers and sizes with independent censoring intervals*

			β						
Percentages of censoring	ρ	γ	−1.0	−0.8	−0.4	0.0	0.4	0.8	1.0
1/3∼1/3∼1/3	0	0	1.000	0.997	0.645	0.053	0.580	0.982	0.998
		1	1.000	0.993	0.590	0.051	0.512	0.960	0.994
	1	0	0.994	0.950	0.462	0.054	0.444	0.946	0.992
		1	0.992	0.988	0.578	0.054	0.552	0.978	0.997
Score test			0.998	0.987	0.571	0.048	0.641	0.993	1.000
1/4∼1/2∼1/4	0	0	1.000	0.996	0.617	0.054	0.606	0.987	0.999
		1	1.000	0.988	0.547	0.053	0.522	0.967	0.996
	1	0	0.991	0.930	0.403	0.056	0.385	0.923	0.988
		1	0.999	0.977	0.516	0.053	0.542	0.978	0.998
Score test			0.996	0.975	0.529	0.050	0.687	0.998	1.000
1/2∼1/4∼1/4	0	0	1.000	0.997	0.653	0.050	0.625	0.991	1.000
		1	1.000	0.994	0.613	0.048	0.578	0.982	0.999
	1	0	0.986	0.927	0.440	0.052	0.464	0.961	0.996
		1	0.999	0.977	0.558	0.052	0.570	0.985	0.999
Score test			0.996	0.976	0.565	0.044	0.703	0.999	1.000
1/4∼1/4∼1/2	0	0	1.000	0.994	0.579	0.053	0.477	0.946	0.992
		1	1.000	0.985	0.500	0.049	0.401	0.893	0.969
	1	0	0.995	0.962	0.471	0.056	0.415	0.917	0.985
		1	1.000	0.989	0.554	0.056	0.455	0.932	0.986
Score test			1.000	0.988	0.517	0.048	0.530	0.961	0.993

we know the underlying distribution. It can be seen from the table that the proposed test procedure seems to have correct sizes and its power is close to that of the parametric score test, suggesting that it performs well under these situations. We noticed that for several situations, the proposed test gave slightly larger powers than the score test and one possible reason for this is that the convergence of the score test is slower than that of the presented test.

In the simulation study, suggested by a referee, we also considered the set-up that yields interval-censored data analogous to those arisen from periodic follow-up studies. Specifically, the T_i's were generated in the same way as above. For censoring intervals, we started by generating a sequence of observation times $W_1 < W_2 < \cdots < W_k$ by first generating k from a Poisson distribution with mean $\lambda_0 K$ and defining the W_j's as the order statistics of a random sample of size k from the uniform distribution $U(0,K)$, where λ_0 and K are some constants. Then U and V are defined as W_j and W_{j+1} if $T \in (W_j, W_{j+1}]$. If $T \leq W_1$, define U and V to be W_1 and W_2 and if $T > W_k$, define U and V to be W_{k-1} and W_k. Note that here U and V are not completely independent of T. Table 2 presents the estimated sizes and powers of the proposed test based on simulated interval-censored data with all other set-ups the same as for Table 1 and with $\lambda_0 = 0.4$ and $K = 10$. For the data here, the percentages of left-censored,

Table 2. *Estimated powers and sizes with dependent censoring intervals*

			β						
Percentages of censoring	ρ	γ	−1.0	−0.8	−0.4	0.0	0.4	0.8	1.0
1/4∼1/2∼1/4	0	0	1.000	0.998	0.690	0.049	0.627	0.992	1.000
		1	1.000	0.997	0.620	0.054	0.563	0.977	0.998
	1	0	0.988	0.929	0.430	0.049	0.460	0.959	0.998
		1	0.999	0.988	0.603	0.051	0.592	0.987	0.999
Score test			1.000	0.998	0.700	0.053	0.620	0.992	0.999

interval-censored and right-censored observations were approximately 25, 50 and 25 percent, respectively. It can be seen that the results are similar to those given in Table 1 and the power for the current set-up is similar to or a little higher than that given in Table 1. This could be because the set-up here gives more information than the one used for Table 1. We also considered other percentages of left-censored, interval-censored and right-censored observations and obtained similar results.

To evaluate the normal approximation given in theorem 1 to the finite distribution of the proposed test statistic, we studied the probability plot of the standardized test statistic against the standard normal distribution under different set-ups. They all suggest that the normal approximation seems reasonable.

Next we applied the proposed test procedure to the set of interval-censored failure time data discussed in Finkelstein (1986). The data arose from a breast cancer study and involve 94 early breast cancer patients. The objective of the study was to compare the patients who had been treated with radiotherapy alone (treatment 1, 46 patients) with those treated with primary radiation therapy and adjuvant chemotherapy (treatment 2, 48 patients). The survival time of interest is the time until the appearance of breast retraction and, in the study, the patients were monitored for breast retraction every 4–6 months. However, they often missed visits as their recovery progressed and returned in a changed status. Thus only interval-censored data on the survival time were observed.

To compare the two treatments, define $x_i = 0$ for the patients with treatment 1 and 1 otherwise. Then by using the function ξ used above and taking $\rho = \gamma = 0$, the application of the presented method yielded $U_1 = -9.9443$ (the first component of U_ξ) with the estimated standard error of 3.6854. This corresponds to a p-value of 0.007 according to the standard normal distribution and suggests that the patients with treatment 1 survived significantly longer than those with treatment 2. In other words, the adjuvant chemotherapy added to the radiation therapy increased the hazards of breast retraction compared with radiation therapy alone. If using $\rho = \gamma = 1$, we obtained $U_1 = -3.0266$ with its estimated standard error was 0.8548, resulting in a p-value of 0.0004. Finkelstein (1986) gave a p-value of 0.004 and obtained a similar result.

5. Concluding remarks

This paper discussed the non-parametric comparison of survival functions when only interval-censored failure time data are available. For the problem, a class of non-parametric tests was proposed and both finite sample and asymptotic properties of the presented approach were established. The proposed test statistics are generalizations of the log-rank test statistic discussed in Peto & Peto (1972). In comparison with the test procedures given in Finkelstein (1986) and Sun (1996), in addition to the given asymptotic distribution, the proposed procedure has the advantage that the calculation of its variance estimate is straightforward. In contrast, the determinations of the variance estimates of both U_F and U_S involve dealing with high dimension matrices. Note that although the asymptotic result given in theorem 1 requires the independence between (U,V) and T, the simulation suggests that the approach works well when the data arise from periodic follow-up studies, where (U,V) and T may not be independent.

In comparison with right-censored failure time data, only limited research exists for interval-censored failure time data although they frequently occur in public health and medical studies such as clinic trials. One obstacle to this is that interval-censoring is much harder to deal with than right-censoring. One consequence resulting from interval-censoring is that the counting process and martingale theory that make the study of right-censored data relatively easy are no longer available for interval-censored data. Instead, the empirical process theory and others seem be to needed to study interval-censored data (Wellner, 1992; Groeneboom & Groeneboom, 1996).

In the above, we have assumed that covariates do not exist. In general, this may not be true and in this case some regression models and related inference procedures would be needed. Also in the proposed method, it is assumed that no exact observation of survival time is observed. This assumption is often needed to study asymptotic properties of the methods for interval-censored data and is required here to guarantee that the statistic U_ξ is valid. Otherwise U_ξ could approach infinity since the denominator term in it may approach zero. As mentioned above, it holds for many periodic follow-up studies and, in particular, the results presented above hold if F_0 has only finite support points. In spite of this, it would still be useful to generalize the proposed approach to situations where observed data include both exact and interval-censored observations on the survival time of interest.

Another direction for future research would be to generalize the proposed approach to situations where the underlying censoring distribution H may be different for different treatment groups. This could occur, for example, if subjects in different treatment groups have different follow-up patterns in a periodic follow-up study. One such example is given by a clinical trial in which patients receiving placebo treatment may feel worse compared with other patients and thus visit doctors more often. Among others, Sun (1999) discussed this problem for current status data, a special case of interval-censored data.

Acknowledgements

The authors wish to thank Dr Hongbin Fang and Dr Liuquan Sun for their discussion, comments and suggestions during the preparation of this paper. They are also grateful to the Editor, Dr Scheike, the Associate Editor and two referees for their many helpful comments and suggestions. The research of the first author was in part supported by a US National Institutes of Health grant.

References

Fleming, T. R. & Harrington, D. P. (1991). *Counting process and survival analysis*. John Wiley, New York.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42**, 845–854.

Gentleman, R. & Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika* **81**, 618–623.

Groeneboom, P. (1996). Lectures on inverse problems. In *Lecture Notes in Mathematics*, 1648. Springer-Verlag, Berlin.

Groeneboom, P. & Wellner, J. A. (1992). *Information bounds and non-parametric maximum likelihood estimation*. DMV Seminar, Band 19, Birkhauser, New York.

Huang, J. & Wellner, J. A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case I. *Statist. Neerlandica* **49**, 153–163.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The statistical analysis of failure time data,* 2nd edn. John Wiley, New York.

Lagakos, S. W. & Louis, T. A. (1988). Use of tumor lethality to interpret tumorgenicity experiments lacking cause-of-death data. *Appl. Statist*. **37**, 169–179.

Li, G. (2003). Non-parametric likelihood ratio goodness-of-fit tests for survival data. *J. Multivariate Anal.* **86**, 166–182.

Li, L., Watkins, T. & Yu, Q. (1997). An EM algorithm for estimating survival functions with interval-censored data. *Scand. J. Statist*. **24**, 531–542.

Pan, W. (2000). A two-sample test with interval censored data via multiple imputation. *Statist. Med.* **19**, 1–11.

Peto, R. & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *J. Roy. Statist. Soc. Ser. A* **135**, 185–207.
Rabinowitz, D., Tsiatis, A. & Aragon, J. (1995). Regression with interval-censored data. *Biometrika* **82**, 501–513.
Sun, J. (1996). A non-parametric test for interval-censored failure time data with application to AIDS studies. *Statist. Med.* **15**, 1387–1395.
Sun, J. (1998). Interval censoring. *Encyclopedia of Biostatistics*. 2090–2095. John Wiley, New York.
Sun, J. (1999). A non-parametric test for current status data with unequal censoring. *J. Roy. Statist. Soc. Ser. B* **61**, 243–250.
Turnbull, B. W. (1976). The empirical distribution with arbitrarily grouped censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38**, 290–295.
van der Vaart, A. W. & Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer, New York.

Received December 2003, in final form June 2004

Jianguo Sun, Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, USA.
E-mail: sunj@missouri.edu

Appendix: Proof of theorem 1

Let U_l denote the lth element of U_η and define

$$Q_{n_l}(u,v,\delta,\gamma) = \frac{1}{n_l}\sum_i{}' 1_{\{(U_i,V_i)\le(u,v),(\Delta_i,\Gamma_i)=(\delta,\gamma)\}} ,$$

where $\sum_i'$ denotes the summation over subjects i in population l, $l = 1, \ldots, k$. Then we have

$$\frac{1}{\sqrt{n}}U_l = \frac{n_l}{\sqrt{n}}Q_{n_l}(\hat{K}_n) = \frac{n_l}{\sqrt{n}}Q_{n_l}(K_0) - \frac{n_l}{\sqrt{n}}Q_n(K_0)$$
$$+ \sqrt{\frac{n_l}{n}}\sqrt{n_l}(Q_{n_l} - Q_0)(\hat{K}_n - K_0) + \frac{n_l}{n}\sqrt{n}(Q_0 - Q_n)(\hat{K}_n - K_0) + \frac{n_l}{\sqrt{n}}Q_n(\hat{K}_n).$$

It is easy to see that both $Q_{n_l}(K_0)$ and $Q_n(K_0)$ are U-statistics and $\{(n_1/\sqrt{n})Q_{n_1}(K_0) - (n_1/\sqrt{n})Q_n(K_0), \ldots, (n_k/\sqrt{n})Q_{n_k}(K_0) - (n_k/\sqrt{n})Q_n(K_0)\}$ has the asymptotic distribution given in the theorem. Thus, for the proof, it is sufficient to show that the other three terms at the right hand side of the above equation converge to zero in probability.

For the last term $Q_n(\hat{K}_n)$, it follows from the proposition 3.2 of Groeneboom (1996) that $Q_n(\hat{K}_n) = 0$. For the other two terms, define

$$\mathcal{F} = \{F : \ F \textit{ is a distribution function defined on } [0,\ M]\},$$

$$\mathcal{G} = \{F : F \in \mathcal{F}, 0 < F(\delta_0) < F(M_0) < 1, \min_{\delta_0 \le t \le M_0 - \varepsilon_0}[F(t+\varepsilon_0) - F(t)] \ne 0\}$$

and

$$\mathcal{H} = \Bigg[\delta\frac{\eta(F(u)) - c_0}{F(u)} + \gamma\frac{\eta(F(v)) - \eta(F(u))}{F(v) - F(u)} + (1-\delta-\gamma)\frac{c_0 - \eta(F(v))}{1 - F(v)}$$
$$- \left\{\delta\frac{\eta(F_0(u)) - c_0}{F_0(u)} + \gamma\frac{\eta(F_0(v)) - \eta(F_0(u))}{F_0(v) - F_0(u)} + (1-\delta-\gamma)\frac{c_0 - \eta(F_0(v))}{1 - F_0(v)}\right\} : (u,v) \in D, F \in \mathcal{G}\Bigg]$$

where $D = \{(u,v): u \geq \delta_0,\ u + \varepsilon_0 \leq v \leq M_0\}$. Because $\mathcal{F}$ is a *P-Donsker* from the proof of corollary 5.1 of Huang & Wellner (1995), $\mathcal{G}$ is a *P-Donsker* by theorem 2.10.1 of van der Vaart & Wellner (1996). Note that for any $F_1, F_2 \in \mathcal{G}$, $(u,v) \in D$,

$$\begin{aligned}&\left|\delta\frac{\eta(F_1(u)) - c_0}{F_1(u)} + \gamma\frac{\eta(F_1(v)) - \eta(F_1(u))}{F_1(v) - F_1(u)} + (1 - \delta - \gamma)\frac{c_0 - \eta(F_1(v))}{1 - F_1(v)}\right.\\&\quad\left. - \delta\frac{\eta(F_2(u)) - c_0}{F_2(u)} + \gamma\frac{\eta(F_2(v)) - \eta(F_2(u))}{F_2(v) - F_2(u)} + (1 - \delta - \gamma)\frac{c_0 - \eta(F_2(v))}{1 - F_2(v)}\right|\\&\leq c[|F_1(u) - F_2(u)| + |F_1(v) - F_2(v)|]\end{aligned}$$

for some constant c. Then it can be shown by using the bracket entropy theorem of van der Vaart & Wellner (1996, pp. 127–159) and the arguments similar to those used in Huang & Wellner (1995) that $\mathcal{H}$ is *P-Donsker*. Also note that $\hat{F}_n \in \mathcal{G}$ for all n sufficiently large and as $n \to \infty$, we have that

$$\int \{|\hat{F}_n(u) - F_0(u)|^2 + |\hat{F}_n(v) - F_0(v)|^2\} \mathrm{d}P \longrightarrow 0$$

in probability from the strong consistency of $\hat{F}_n$ (Groeneboom & Wellner, 1992, p. 85). It thus follows from this and the uniform asymptotic equicontinuity of the empirical process resulting from the Donsker property (van der Vaart & Wellner, 1996, pp. 168–171) that

$$\sqrt{n_l}(Q_{n_l} - Q_0)(\hat{K}_n - K_0) \longrightarrow 0$$

and

$$\sqrt{n}(Q_n - Q_0)(\hat{K}_n - K_0) \longrightarrow 0$$

in probability as $n \to \infty$. This completes the proof.