

Research Article

Weighted Wilcoxon-Type Rank Test for Interval Censored Data

Ching-fu Shen,¹ Jin-long Huang,² and Chin-san Lee^{1,3}

¹ Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung 80424, Taiwan

² Center for Fundamental Sciences, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

³ Graduate School of Human Sexuality, Shu-Te University, Kaohsiung 82445, Taiwan

Correspondence should be addressed to Chin-san Lee; chinsan@stu.edu.tw

Received 8 November 2012; Accepted 19 December 2012

Academic Editor: Jen Chih Yao

Copyright © 2013 Ching-fu Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Interval censored (IC) failure time data are often observed in medical follow-up studies and clinical trials where subjects can only be followed periodically, and the failure time can only be known to lie in an interval. In this paper, we propose a weighted Wilcoxon-type rank test for the problem of comparing two IC samples. Under a very general sampling technique developed by Fay (1999), the mean and variance of the test statistics under the null hypothesis can be derived. Through simulation studies, we find that the performance of the proposed test is better than that of the two existing Wilcoxon-type rank tests proposed by Mantel (1967) and R. Peto and J. Peto (1972). The proposed test is illustrated by means of an example involving patients in AIDS cohort studies.

1. Introduction

Interval censored (IC) failure time data often arise from medical studies such as AIDS cohort studies and leukemic blood cancer follow-up studies. In these studies, patients were divided into two groups according to different treatments. For example, in leukemic cancer studies, one group of the patients was treated with radiotherapy alone, and the other group of patients was treated with initial radiotherapy along with adjuvant chemotherapy. The two groups of patients were examined every month, and the failure time of interest is the time until the appearance of leukemia retraction; the object is to test the difference of the failure times between the two treatments. Some of the patients missed some successive scheduled examinations and came back later with a changed clinical status, and they contributed IC observations. For our convenience, we assume that in such a medical study, the underlying survival function can be either discrete or continuous, and there are only finitely many scheduled examination times. IC data only provide partial information about the lifetime of the subject, and the data is one kind of incomplete data. To deal with such incomplete data, Turnbull [1] introduced a self-consistent algorithm to compute the maximum likelihood estimate of the survival function for arbitrarily censored and truncated data. For IC data, there

have been some related studies in the literature as well. For example, Mantel [2] extends Gehan's [3, 4] generalized Wilcoxon [5] test to interval censored data, and R. Peto and J. Peto [6] also develop a different version. Sun [7] applied Turnbull's algorithm to estimate the number of failures and risks of IC data and then propose a log-rank type test.

Fay [8], Sun [7], Zhao and Sun [9], Sun et al. [10], and Huang et al. [11] extend the log-rank test to interval censored data. Petroni and Wolfe [12] and Lim and Sun [13] generalize Pepe and Fleming's [14] weighted Kaplan-Meier (WKM) [15] test to interval censored data.

For the purpose of comparing the power of the test statistics, Fay [8] proposed a model for generating interval censored observation. A similar selection scheme can also be seen in the Urn model of Lee [16] and mixed cased model of Schick and Yu [17]. In this paper, we propose a Wilcoxon-type weighted rank test to compare with the existing two Wilcoxon-type rank tests proposed by Mantel [2] and R. Peto and J. Peto [6]. We restrict ourselves to the Wilcoxon-type rank tests because these tests are simple to use and have the robustness property that their powers are fairly stable under different lifetime distributions.

This paper is organized as follows. In Section 2, we review the Turnbull's [1] algorithm and introduce Fay's [8] selection model for generating interval censored data. This

TABLE 1: The probability of selected interval.

True value of X	Selected interval	Probability
1	(0,1]	$p_1 a_1$
	(0,2]	$p_1(1 - a_1)a_2$
	(0,3]	$p_1(1 - a_1)(1 - a_2)$
2	(1,2]	$p_2 a_1 a_2$
	(0,2]	$p_2(1 - a_1)a_2$
	(1,3]	$p_2 a_1(1 - a_2)$
	(0,3]	$p_2(1 - a_1)(1 - a_2)$
3	(2,3]	$p_3 a_2$
	(1,3]	$p_3 a_1(1 - a_2)$
	(0,3]	$p_3 a_1(1 - a_1)(1 - a_2)$

selection model can be extended to a more general one, and the consistency property can be found in Schick and Yu [17]. In Section 3, we introduce Mantel's [2] and R. Peto and J. Peto's [6] generalized Wilcoxon-type rank tests and propose our weighted rank test. In Section 4, a simulation study is conducted to compare the performance of the three tests under different configurations. Finally, an application to AIDS cohort study is presented in Section 5.

2. Data Treatment

Assume that X is the lifetime random variable of a survival study, measured in discrete units and taking values $0 = x_0 < x_1 < x_2 < \dots < x_m$. Let $U = \{(x_i, x_j], 0 \leq i < j \leq m\}$ be the collection of all $m(m+1)/2$ admissible intervals, and define $p_j = P(X = x_j)$, where $\sum_{j=1}^m p_j = 1$, so that $F(x) = \sum_{x_j \leq x} p_j$, and $S(x) = \sum_{x_j > x} p_j$. Note that the observed failure time data in a clinical trial can be discretized if the underlying variable is continuous.

2.1. Turnbull's Algorithm. Suppose that there is a sample of n i.i.d. observations $(X_L^i, X_R^i]$ of X , $i = 1, 2, \dots, n$. Here, $(X_L^i, X_R^i]$ is the IC observation of the i th individual in the sample, where $X_L^i, X_R^i \in \{x_0, x_1, x_2, \dots, x_m\}$, and $X_L^i < X_R^i$. The case $X_R^i = x_m$ is to denote that the failure time of the i th subject occurs after the last examination time x_{m-1} . Turnbull [1] proposed an algorithm to estimate the unknown probabilities $p = (p_1, p_2, \dots, p_m)$. The algorithm can be described by the following four steps.

Step 1. Start with initial values $p^{(0)} = (p_1^{(0)}, p_2^{(0)}, \dots, p_m^{(0)})$.

Step 2. Obtain improved estimates $p_j^{(1)}$ by setting

$$p_j^{(1)} = \frac{1}{n} \sum_{i=1}^n \frac{\alpha_j^i p_j^{(0)}}{\sum_{l=1}^m \alpha_l^i p_l^{(0)}}, \quad j = 1, 2, \dots, m, \quad (1)$$

where $\alpha_j^i = I\{x_j \in (X_L^i, X_R^i]\}$.

Step 3. Return to Step 1 with $p^{(1)}$ replacing $p^{(0)}$.

Step 4. Stop when the required accuracy has been achieved.

TABLE 2: Selection probability $Q(I)$ for all admissible intervals.

Interval I	Probability $Q(I)$
(0,1]	$p_1 a_1$
(1,2]	$p_2 a_1 a_2$
(2,3]	$p_3 a_2$
(0,2]	$(p_1 + p_2)(1 - a_1)a_2$
(1,3]	$(p_2 + p_3)a_1(1 - a_2)$
(0,3]	$(p_1 + p_2 + p_3)(1 - a_1)(1 - a_2)$

The algorithm is simple and converges fairly rapidly. The estimate $\hat{p} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_m)$ yielded from the iteration is in fact the unique maximum likelihood estimate of $p = (p_1, p_2, \dots, p_m)$ and is a self-consistent estimate.

2.2. Return Probability Model. To comply with the periodical clinical inspection, Fay [8] proposed a simulation model for generating IC data. He assumed that the probability for a patient to return to the clinic for inspection at time points x_1, x_2, \dots, x_{m-1} are i.i.d. Bernulli random variables A_1, A_2, \dots, A_{m-1} ; that is, $P(A_i = 1) = q$, $P(A_i = 0) = 1 - q$, $0 < q < 1$, $i = 1, 2, \dots, m - 1$. $A_i = 1$ means that the patient returned to the clinic at the inspection time x_i , and $A_i = 0$ means that the patient missed the inspection. In our model, we always assume that $A_m = 1$. The failure time X is independent of $(A_1, A_2, \dots, A_{m-1})$, and the observable random interval is

$$(X_L, X_R] = (x_{s_j}, x_{t_j}], \quad \text{where } s_j = \max\{0 \leq l < j : A_l = 1\},$$

$$t_j = \min\{j \leq l \leq m : A_l = 1\},$$

$$x_{s_j} < X \leq x_{t_j}. \quad (2)$$

2.2.1. Model Consistency. Under Fay's [8] selection model, the consistency property has been proved. This selection model can be generalized to the case that the return probability at each examination time point may be different; say that $P(A_i = 1) = a_i$, $i = 1, 2, \dots, m$. To demonstrate the generalized return model, we set $m = 3$ and $x_1 = 1$, $x_2 = 2$, and $x_3 = 3$. The selection probabilities for all admissible intervals are shown in Tables 1 and 2.

It is not difficult to see that the selection probability of the interval $I = (x_u, x_v]$ is

$$Q\{I\} = P(I) \left[a_u \times \prod_{i=u+1}^{v-1} (1 - a_i) \times a_v \right], \quad (3)$$

$$u = 0, 1, 2, \dots, v - 1, \quad v = 1, 2, \dots, m - 1,$$

$$Q\{(x_u, x_m]\} = P((x_u, x_m]) \left[a_u \times \prod_{i=u+1}^{m-1} (1 - a_i) \right], \quad 0 \leq u \leq m - 1, \quad (4)$$

where $P(I) = (p_{u+1} + p_{u+2} + \dots + p_v)$, $x_0 = 0$, and $a_0 = 1$. For instance, the interval $(0, 2]$ may be selected under two

possibilities. First, the true value of X is $X = 1$, and the patient who missed the inspection at $x_1 = 1$ then goes to inspection at $x_2 = 2$; in this case, the interval is selected with probability $p_1(1 - a_1)a_2$. Second, the true value of X is $X = 2$, and the patient missed the inspection at $x_1 = 1$ then goes to inspection at $x_2 = 2$; in this case, the interval is selected with probability $p_2(1 - a_1)a_2$, and therefore $Q\{(0, 2]\} = (p_1 + p_2)(1 - a_1)a_2$.

The generalized return probability model can be viewed as a special case of the mixed case model in Schick and Yu [17]; under very mild conditions, the estimate of $p = (p_1, p_2, \dots, p_m)$ computed by Turnbull's algorithm is still consistent.

3. Wilcoxon-Type Rank Tests for Interval Censored Data

Two-sample Wilcoxon rank test is a well-known method to test whether two samples of exact data come from the same population. The method is constructed by ranking the pooled samples and giving an appropriate rank to each observation. However, this ranking technique is in general not admissible for intervals. In this section, we will discuss how to generalize the ranking technique and then propose a Wilcoxon-type rank test for IC data to compare with two existing rank tests proposed by Mantel [2] and R. Peto and J. Peto [6]. Suppose that two samples of IC data for X and Y are, respectively, $(X_L^i, X_R^i), i = 1, 2, \dots, n_1$ and $(Y_L^i, Y_R^i), i = 1, 2, \dots, n_2$. To test whether these two samples come from the same population is equivalent to testing the equality of survival functions $S_X(t)$ and $S_Y(t)$, for all $t \geq 0$; that is,

$$H_0 : S_X(t) = S_Y(t), \quad \forall t \geq 0. \tag{5}$$

3.1. Mantel's Test. Mantel [2] extended Gehan's [3, 4] generalized Wilcoxon test to interval censored data by defining the score of the k th observation as the number of observations that are definitely greater than the k th observation minus the number of observations that are definitely less than the k th observation. He proposed the test statistic

$$W = \sum_{k=1}^{n_1} V_k, \quad \text{where } V_k = \sum_{h=1}^{n_1+n_2} V_{kh},$$

$$V_{kh} = \begin{cases} 1 & \text{if we know for sure obs-}k > \text{obs-}h, \\ -1 & \text{if we know for sure obs-}k < \text{obs-}h, \\ 0 & \text{if not sure.} \end{cases} \tag{6}$$

Under H_0 , the test statistic is approximately normal distributed with mean 0 and variance

$$\text{Var}(W) = n_1 n_2 \sum_{k=1}^{n_1+n_2} \frac{V_k^2}{(n_1 + n_2)(n_1 + n_2 - 1)}. \tag{7}$$

3.2. R. Peto and J. Peto's Test. Different from the Mantel's generalized version, R. Peto and J. Peto [6] defined the score of the i th observation as

$$U_i = \frac{f(\widehat{S}(X_L^i)) - f(\widehat{S}(X_R^i))}{\widehat{S}(X_L^i) - \widehat{S}(X_R^i)}, \tag{8}$$

where \widehat{S} is the estimated survival function, $f(y) = y^2 - y$; hence, $U_i = \widehat{S}(X_L^i) + \widehat{S}(X_R^i) - 1$. They proposed the test statistic

$$Z^2 = \frac{(Y_1^2/n_1 + Y_2^2/n_2)}{s^2}, \quad \text{where } Y_1 = \sum_{i=1}^{n_1} U_i,$$

$$Y_2 = \sum_{i=n_1+1}^{n_1+n_2} U_i, \tag{9}$$

$$s^2 = \frac{\sum_{i=1}^{n_1+n_2} U_i^2}{(n_1 + n_2 - 1)}.$$

Under H_0 , the test statistic Z^2 is approximately distributed as χ_1^2 .

3.3. Our Proposed Wilcoxon-Type Weighted Rank Test. To transform an IC data to exact, we first assign each inspection time x_i a primary rank R_i ; for instance, $R_i = i$. Rewrite any observation, say (X_L^j, X_R^j) , as $(X_L^j, X_R^j) = (x_{u^{(j)}}, x_{v^{(j)}})$, where $x_{u^{(j)}}, x_{v^{(j)}} \in \{0, x_1, x_2, \dots, x_m\}$, and $x_{u^{(j)}} < x_{v^{(j)}}$. Then, we associate the observation (X_L^j, X_R^j) with the weighted rank

$$\text{rank}((X_L^j, X_R^j)) = \sum_{l=u^{(j)}+1}^{v^{(j)}} \frac{p_l}{p_{u^{(j)}+1} + \dots + p_{v^{(j)}}} R_l. \tag{10}$$

Let W_1, W_2 be, respectively, the average weighted rank of the X and Y samples, so that

$$W_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \text{rank}((X_L^i, X_R^i))$$

$$= \frac{1}{n_1} \sum_{i=1}^{n_1} \left(\sum_{l=u^{(i)}+1}^{v^{(i)}} \frac{p_l}{p_{u^{(i)}+1} + \dots + p_{v^{(i)}}} R_l \right), \tag{11}$$

$$W_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} \text{rank}((Y_L^j, Y_R^j))$$

$$= \frac{1}{n_2} \sum_{j=1}^{n_2} \left(\sum_{l=u^{(j)}+1}^{v^{(j)}} \frac{p_l}{p_{u^{(j)}+1} + \dots + p_{v^{(j)}}} R_l \right).$$

To test whether two IC samples come from the same population, we propose the test statistic

$$\text{W.R.T} = \frac{W_1 - W_2}{\sqrt{\text{Var}(W_1) + \text{Var}(W_2)}}. \tag{12}$$

Under H_0 , the central limit theorem implies that W.R.T is approximately distributed as a standard normal random

variable. However, the mean and variance of W_1 and W_2 may depend on the probability space where they are defined; it means, different selection probability for IC intervals in (4) leads to different mean and variance of W_1 and W_2 . We therefore only consider the selection model of Fay defined in Section 2.2. In this model, the selection probability of an IC interval is in one of the following categories:

$$(i) Q\{(0, x_r)\} = \sum_{j=1}^r p_j q(1-q)^{r-1}, \quad 1 \leq r < m, \quad (13)$$

$$(ii) Q\{(0, x_m)\} = \sum_{j=1}^m p_j(1-q)^{m-1}, \quad (14)$$

$$(iii) Q\{(x_u, x_v)\} = \sum_{j=u+1}^v p_j q^2(1-q)^{r-1}, \quad 1 \leq u < v < m, \quad (15)$$

$$(iv) Q\{(x_u, x_m)\} = \sum_{j=u+1}^m p_j q(1-q)^{m-u-1}, \quad 1 \leq u < m. \quad (16)$$

Consider the probability space $(U, 2^U, Q)$, where the probability measure Q is defined in Section 2. To compute the variance of W_1 and W_2 , we define a random variable Z on this space by assigning value $Z\{(x_u, x_v)\}$ to the interval $(x_u, x_v]$ in U , where

$$Z\{(x_u, x_v)\} = \sum_{l=u+1}^v \frac{p_l}{p_{u+1} + \dots + p_v} R_l, \quad 0 \leq u < v \leq m. \quad (17)$$

The value $Z\{(x_u, x_v)\}$ can be viewed as the weighted rank of $(x_u, x_v]$. If $R_l, l = 1, 2, \dots, m$ are chosen as in the Wilcoxon test for exact data, then our proposed test statistic W.R.T is a Wilcoxon-type weighted rank test. Under this probability space, the expectation $E(Z)$ can be simplified as in the following theorem.

Theorem 1. Suppose that Z is the random variable defined on the probability space $(U, 2^U, Q)$ according to (17). Then, the expectation of $Z, E(Z)$, can be simplified as

$$E(Z) = \sum_{l=1}^m p_l R_l, \quad (18)$$

which is independent of the choice of q .

Proof. It is obvious that $E(Z)$ can be written as $E(Z) = \sum_{l=1}^m b_l p_l R_l$, where the coefficients $b_l, l = 1, 2, \dots, m$ are to be determined. The theorem is, hence, proved if we can show that all the coefficients b_l are ones.

Consider b_1 first. An interval $(x_u, x_v]$ contributes $p_1 R_1$ in $E(Z)$ if and only if it contains the point x_1 . Therefore, it must be of the form $(0, x_v], v = 1, 2, \dots, m$. For intervals $(0, x_v], 1 \leq v \leq m - 1$, the probabilities $Q\{(0, x_v)\}$ are defined in (13).

For interval $(0, x_m]$, the probability $Q\{(0, x_m)\}$ is defined in (14). Therefore, the coefficient b_1 is

$$b_1 = \sum_{v=1}^{m-1} q(1-q)^{v-1} + (1-q)^{m-1} \quad (19)$$

$$= q \frac{1 - (1-q)^{m-1}}{1 - (1-q)} + (1-q)^{m-1} = 1.$$

Next, consider the coefficient b_j for $1 < j \leq m - 1$. An interval contributes $p_j R_j$ if and only if it contains the point x_j . Therefore, it must be of the form $(x_u, x_v]$, where $0 \leq u < j \leq v \leq m$. It is necessary to study the contribution of the interval $(x_u, x_v]$ to b_j in four different categories.

(i) $u = 0, v \leq m - 1$.

By (13), this category contributes $\sum_{v=j}^{m-1} q(1-q)^{v-1}$.

(ii) $u = 0, v = m$.

By (14), the interval $(0, x_m]$ contributes $(1-q)^{m-1}$.

(iii) $1 \leq u < v \leq m - 1$.

By (15), this category contributes $\sum_{u=1}^{j-1} \sum_{v=j}^{m-1} q^2(1-q)^{v-u-1}$.

(iv) $u \geq 1, v = m$.

By (16), this category contributes $\sum_{u=1}^{j-1} q(1-q)^{m-u-1}$. Consequently, the coefficient of b_j is

$$b_j = \sum_{v=j}^{m-1} q(1-q)^{v-1} + (1-q)^{m-1}$$

$$+ \sum_{u=1}^{j-1} \sum_{v=j}^{m-1} q^2(1-q)^{v-u-1} + \sum_{u=1}^{j-1} q(1-q)^{m-u-1}$$

$$= q \frac{(1-q)^{j-1} [1 - (1-q)^{m-j}]}{1 - (1-q)} + (1-q)^{m-1}$$

$$+ \sum_{u=1}^{j-1} q \left(q \frac{(1-q)^{j-u-1} [1 - (1-q)^{m-j}]}{1 - (1-q)} \right)$$

$$+ \sum_{u=1}^{j-1} q(1-q)^{m-u-1}$$

$$= (1-q)^{j-1} - (1-q)^{m-1} + (1-q)^{m-1}$$

$$+ \sum_{u=1}^{j-1} q(1-q)^{j-u-1}$$

$$- \sum_{u=1}^{j-1} q(1-q)^{m-u-1} + \sum_{u=1}^{j-1} q(1-q)^{m-u-1}$$

TABLE 3: The mean, sample variance, and sample deviation of \hat{q} .

n	q	0.8	0.5	0.3
50	Estimate	0.8001	0.5029	0.3024
	Variance	0.0020	0.0021	0.0012
	Std.	0.0448	0.0461	0.0341
100	Estimate	0.8039	0.5036	0.3012
	Variance	0.0010	0.001	0.0005
	Std.	0.0320	0.0312	0.0233
150	Estimate	0.8009	0.4977	0.3033
	Variance	0.0005	0.0008	0.0004
	Std.	0.0225	0.0277	0.0207

$$\begin{aligned}
 &= (1 - q)^{j-1} + (1 - (1 - q)^{j-1}) \\
 &= 1.
 \end{aligned}
 \tag{20}$$

Finally, the proof for the case $j = m$ is

$$\begin{aligned}
 b_m &= \sum_{u=1}^{m-1} q(1 - q)^{m-u-1} + (1 - q)^{m-1} \\
 &= 1 - (1 - q)^{m-1} + (1 - q)^{m-1} = 1.
 \end{aligned}
 \tag{21}$$

□

The variance of Z , $\text{Var}(Z)$, is

$$\begin{aligned}
 \text{Var}(Z) &= E(Z^2) - E^2(Z) \\
 &= \sum_{i=1}^{m(m+1)/2} Q(I_i) R^2(I_i) - E^2(Z),
 \end{aligned}
 \tag{22}$$

where $Q(I_i)$ and $R(I_i)$ are the selected probability and the weighted rank of the i th admissible interval of I_i , respectively, $I_i \in U$.

Consider the formulas (13)–(16), the selection probability $Q(I)$ depends on $p = (p_1, p_2, \dots, p_m)$ and q ; therefore, the likelihood function can be written as

$$L(p_1, p_2, \dots, p_m, q) = P(p_1, p_2, \dots, p_m) G(q), \tag{23}$$

where $G(q) = q^{k_1}(1 - q)^{k_2}$, k_1 and k_2 are positive integers determined by the sample. Since the probability $p = (p_1, p_2, \dots, p_m)$ can be estimated by Turnbull's [1] algorithm discussed in Section 2.2, and q can also be estimated by $k_1/(k_1 + k_2)$ trivially.

For demonstration, we set $m = 6$, inspection times $x_i = i$, $i = 1, 2, \dots, 6$, and the true lifetime X is exponentially distributed with $\lambda = 1/3$. For different sample sizes $n = 50, 100$, and 150 , different return probabilities of inspection $q = 0.8, 0.5$, and 0.3 , and simulation with 100 replications, Table 3 presents the estimates of q and sample variance and sample deviation of \hat{q} . To show the normality of W.R.T, we assume that the two populations (sample size $n_1 = n_2 = 30$) are coming from the same distribution exponential (1/5).

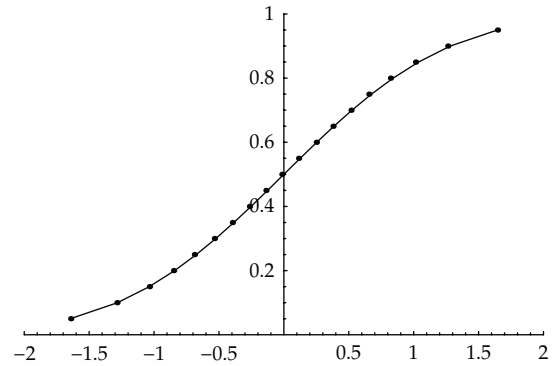


FIGURE 1: CDF of standard normal and simulation result of W.R.T. Line: standard normal. Point: simulation result of W.R.T ($q = 0.5$).

By simulation with 10000 replications and different return probabilities of inspection $q = 0.8, 0.5$, and 0.3 , Table 4 presents the quantiles of W.R.T and $N(0, 1)$. Figure 1 shows the CDF plots of $N(0, 1)$ and W.R.T with $q = 0.5$.

4. Simulation Study

In this section, we carry out simulation studies to compare the performance of W.R.T test with Mantel's [2] and Peto's [6] tests. In the study, we assume that the failure time random variable is distributed as exponential, total sample sizes are $n = 100$ and 200 , and each sample has $(n/2)$ subjects. The interval censored data are generated by the following four steps.

- Step 1. Generate a failure time t_j from some distribution.
- Step 2. Create a 0, 1 sequence $A = \{A_0, A_1, A_2, \dots, A_m\}$ with probabilities $P(A_i = 1) = q$, $i = 1, 2, \dots, m - 1$, and $P(A_0 = 1) = P(A_m = 1) = 1$.
- Step 3. The observation is $(a, b]$, if $a < t_j \leq b$, $A_a = A_b = 1$, and $A_{a+1} = A_{a+2} = \dots = A_{b-1} = 0$.
- Step 4. Repeat Step 1 to Step 3 for n times.

TABLE 4: The quantiles of W.R.T and $N(0, 1)$.

Quantile	Normal (0,1)	W.R.T		
		0.8	0.5	0.3
0.05	-1.6449	-1.6757	-1.6421	-1.6786
0.10	-1.2816	-1.3083	-1.2855	-1.3064
0.15	-1.0364	-1.0543	-1.0354	-1.0700
0.20	-0.8416	-0.8647	-0.8494	-0.8649
0.25	-0.6745	-0.6874	-0.6892	-0.6877
0.30	-0.5244	-0.5326	-0.5351	-0.5338
0.35	-0.3853	-0.3883	-0.3966	-0.3946
0.40	-0.2533	-0.2623	-0.2651	-0.2663
0.45	-0.1257	-0.1247	-0.1379	-0.1314
0.50	0	-0.0007	-0.0152	0.0002
0.55	0.1257	0.1296	0.1136	0.1306
0.60	0.2533	0.2582	0.2503	0.2604
0.65	0.3853	0.3879	0.3789	0.4012
0.70	0.5244	0.5336	0.5176	0.5501
0.75	0.6745	0.6814	0.6549	0.6954
0.80	0.8416	0.8535	0.8215	0.8611
0.85	1.0364	1.0508	1.0146	1.0734
0.90	1.2816	1.2758	1.2628	1.3346
0.95	1.6449	1.6368	1.6458	1.6747

We consider three return probabilities, $q = 0.8, 0.5,$ and $0.3,$ two sets of inspection time points, $m = 6, 10,$ and 1000 replications at significance level $0.05.$

In the case of $m = 6,$ 6 return points, we set the hazards $1/3$ for population 1 and $1/3e^\beta$ for population 2. Figure 2 shows the density plot of exponential distribution with $\beta = -0.4, -0.2, 0, 0.2, 0.4.$ In the case of $m = 10,$ 10 return points, we set the hazards $1/4$ for population 1 and $1/4e^\beta$ for population 2. Figure 3 shows the density plot of exponential distribution with $\beta = -0.6, -0.3, 0, 0.3, 0.6.$ Tables 5 and 6 present the powers of the three tests with sample size $n = 100$ and $200.$ Simulation result shows that when the failure times come from the exponential distribution, our proposed test W.R.T is the most powerful.

5. An Application to AIDS Cohort Study

Consider the data of 262 hemophilia patients in De Gruttola and Lagakos [18], among them, 105 patients received at least $1,000 \mu\text{g}/\text{kg}$ of blood factor for at least one year between 1982 and 1985, and the other 157 patients received less than $1,000 \mu\text{g}/\text{kg}$ in each year. In this medical study, patients were treated between 1978 and 1988, the observations (X_L, X_R) for the 262 patients, based on a discretization of the time axis into 6-month intervals. The failure time of interest is the time of HIV seroconversion. The object is to test the difference of the failure times between the two treatments. Applying our proposed test, namely, W.R.T, Mantel's [2] and Peto's [6] tests to this data set, the values of the three test statistics are $-7.815, -7.352,$ and $56.476,$ respectively. All the three P values are less than 0.001 and have the same conclusion that the HIV seroconversion appeared in the two groups of patients being significantly different.

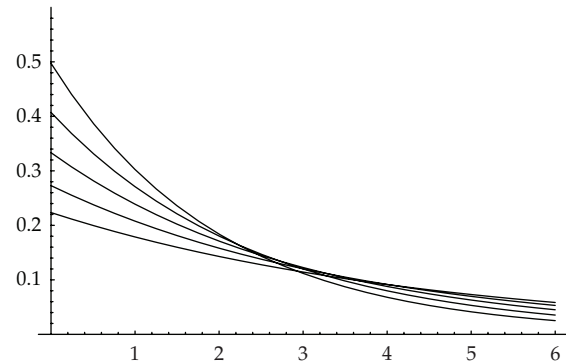


FIGURE 2: Density plot of exponential distribution with hazards $1/3e^\beta.$

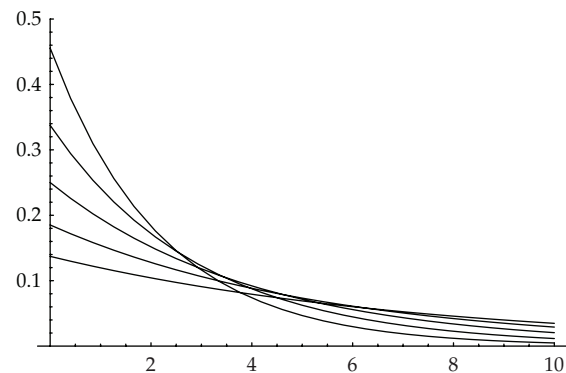


FIGURE 3: Density plot of exponential distribution with hazards $1/4e^\beta.$

TABLE 5: Power comparison of tests under exponential distribution with sample $n = 100$.

m	q	Test	β				
			-0.4	-0.2	0	0.2	0.4
6	0.8	W.R.T	0.419	0.131	0.050	0.150	0.371
		Mantel	0.391	0.120	0.047	0.143	0.362
		Peto	0.385	0.122	0.050	0.140	0.361
	0.5	W.R.T	0.383	0.123	0.045	0.132	0.345
		Mantel	0.360	0.121	0.041	0.124	0.344
		Peto	0.345	0.109	0.045	0.124	0.336
	0.3	W.R.T	0.313	0.102	0.042	0.103	0.254
		Mantel	0.307	0.101	0.040	0.096	0.255
		Peto	0.294	0.099	0.040	0.101	0.248
m	q	Test	β				
			-0.6	-0.3	0	0.3	0.6
10	0.8	W.R.T	0.801	0.289	0.047	0.264	0.779
		Mantel	0.736	0.246	0.051	0.236	0.737
		Peto	0.717	0.242	0.050	0.237	0.740
	0.5	W.R.T	0.793	0.278	0.048	0.275	0.712
		Mantel	0.754	0.247	0.045	0.262	0.678
		Peto	0.718	0.240	0.052	0.256	0.663
	0.3	W.R.T	0.680	0.238	0.052	0.239	0.662
		Mantel	0.667	0.215	0.048	0.223	0.640
		Peto	0.624	0.216	0.049	0.224	0.632

TABLE 6: Power comparison of tests under exponential distribution with sample $n = 200$.

m	q	Test	β				
			-0.4	-0.2	0	0.2	0.4
6	0.8	W.R.T	0.710	0.268	0.049	0.196	0.642
		Mantel	0.678	0.251	0.053	0.192	0.632
		Peto	0.667	0.253	0.054	0.190	0.630
	0.5	W.R.T	0.656	0.201	0.050	0.193	0.573
		Mantel	0.636	0.193	0.046	0.184	0.561
		Peto	0.621	0.188	0.047	0.188	0.558
	0.3	W.R.T	0.549	0.182	0.058	0.171	0.523
		Mantel	0.537	0.182	0.057	0.168	0.506
		Peto	0.523	0.181	0.052	0.164	0.501
m	q	Test	β				
			-0.6	-0.3	0	0.3	0.6
10	0.8	W.R.T	0.984	0.520	0.049	0.473	0.945
		Mantel	0.964	0.472	0.050	0.441	0.930
		Peto	0.957	0.460	0.050	0.439	0.927
	0.5	W.R.T	0.971	0.484	0.046	0.448	0.957
		Mantel	0.961	0.458	0.045	0.424	0.946
		Peto	0.948	0.434	0.039	0.415	0.944
	0.3	W.R.T	0.942	0.429	0.053	0.402	0.901
		Mantel	0.927	0.413	0.050	0.387	0.892
		Peto	0.908	0.385	0.060	0.368	0.889

References

- [1] B. W. Turnbull, "The empirical distribution function with arbitrarily grouped, censored and truncated data," *Journal of the Royal Statistical Society B*, vol. 38, no. 3, pp. 290–295, 1976.
- [2] N. Mantel, "Ranking procedures for arbitrarily restricted observation," *Biometrics*, vol. 23, no. 1, pp. 65–78, 1967.
- [3] E. A. Gehan, "A generalized Wilcoxon test for comparing arbitrarily singly-censored samples," *Biometrika*, vol. 52, no. 1-2, pp. 203–223, 1965.
- [4] E. A. Gehan, "A generalized two-sample Wilcoxon test for doubly censored data," *Biometrika*, vol. 62, no. 3-4, pp. 650–653, 1965.
- [5] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrika*, vol. 1, no. 6, pp. 80–83, 1945.
- [6] R. Peto and J. Peto, "Asymptotically efficient rank invariant test procedures," *Journal of the Royal Statistical Society A*, vol. 135, no. 2, pp. 185–206, 1972.
- [7] J. Sun, "A non-parametric test for interval-censored failure time data with application to AIDS studies," *Statistics in Medicine*, vol. 15, no. 13, pp. 1378–1395, 1996.
- [8] M. P. Fay, "Comparing several score tests for interval censored data," *Statistics in Medicine*, vol. 18, no. 3, pp. 273–285, 1999.
- [9] Q. Zhao and J. Sun, "Generalized log-rank test for mixed interval-censored failure time data," *Statistics in Medicine*, vol. 23, no. 10, pp. 1621–1629, 2004.
- [10] J. Sun, Q. Zhao, and X. Zhao, "Generalized log-rank tests for interval-censored failure time data," *Scandinavian Journal of Statistics*, vol. 32, no. 1, pp. 49–57, 2005.
- [11] J. Huang, C. Lee, and Q. Yu, "A generalized log-rank test for interval-censored failure time data via multiple imputation," *Statistics in Medicine*, vol. 27, no. 17, pp. 3217–3226, 2008.
- [12] G. R. Petroni and R. A. Wolfe, "A two-sample test for stochastic ordering with interval-censored data," *Biometrics*, vol. 50, no. 1, pp. 77–87, 1994.
- [13] H.-J. Lim and J. Sun, "Nonparametric tests for interval-censored failure time data," *Biometrical Journal*, vol. 45, no. 3, pp. 263–276, 2003.
- [14] M. S. Pepe and T. R. Fleming, "Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data," *Biometrics*, vol. 45, no. 2, pp. 497–507, 1989.
- [15] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457–481, 1958.
- [16] C. Lee, "An urn model in the simulation of interval censored failure time data," *Statistics & Probability Letters*, vol. 45, no. 2, pp. 131–139, 1999.
- [17] A. Schick and Q. Yu, "Consistency of the GMLE with mixed case interval-censored data," *Scandinavian Journal of Statistics*, vol. 27, no. 1, pp. 45–55, 2000.
- [18] V. De Gruttola and S. W. Lagakos, "Analysis of doubly-censored survival data, with application to AIDS," *Biometrics*, vol. 45, no. 1, pp. 1–11, 1989.