Journal of the Korean
Statistical Society
Vol. 23, No. 1, 1994

# Nonparametric Test for Equality of Survival Distributions Using Probit Scale†

Sang-Un Yun [1] and Chung-Seon Park [2]

**ABSTRACT**

To test the equality of survival distributions in the presence of arbitrary right censorship, the choice of weights which are functions of the number of individuals at risk at the time of each death is very important in increasing the power of the test. In this paper a weight by probit scale is derived and the efficiencies relative to the other weight's are also investigated.

**KEYWORDS:** Distribution-free test, Survival distribution, Failure or uncensored time, Censoring time, Probit scale, Relative efficiency.

## 1. INTRODUCTION

A distribution-free two-sample test is proposed, which is an extension of the Wilcoxon test, to samples with arbitrary censoring on the right. Suppose that the failure times or censoring times are observed for two different populations. For the first population, let the failure times, $T_1, T_2, \cdots, T_m$ be iid each with distribution function $F_1$, and $C_1, C_2, \cdots, C_m$ be iid each with distribution

[1] Department of the Applied Statistics, Yonsei University, Seoul, 120-349, Korea
[2] Agency for Defense Development, Taejon, 305-600, Korea

† This research is supported by Korean Research Foundation, 1988.

function $G_1$ where $C_i$ is the censoring time associated with $T_i$. We can observe $(X_1, I_1), \cdots, (X_m, I_m)$ where $X_i$ is defined as the minimum of $T_i$ and $C_i$, and $I_i$ is a binary random variable equaling 1 if $T_i \le C_i$ and 0 otherwise. For the second population, let, $U_1, U_2, \cdots, U_n$ be iid each with distribution function $F_2$, and $D_1, D_2, \cdots, D_n$ be iid each with distribution function $G_2$ where $D_j$ is the censoring time associated with $U_j$. We can observe $(Y_1, J_1), \cdots, (Y_n, J_n)$ where $Y_j$ is defined as the minimum of $U_j$ and $D_j$, and $J_j$ is a binary random variable equaling 1 if $U_j \le D_j$ and 0 otherwise.

In this study we consider a generalized Wilcoxon procedure by Mantel and Haenszel(1959), and Gehan(1965) to test the equality of survival distributions of two populations. Taron and Ware(1977) demonstrate that the test statistics for these two methods differ only in the choice of weights which are functions of the number of individuals at risk at the time of each failure or censoring. Additionally they proposed a third statistic which is constructed with different weights. Based on the suggestions of Cochran(1954) regarding the combination of $2 \times 2$ contingency tables, a fourth statistic using probit scale is proposed which may be more satisfactory when the form of the alternatives is in question.

## 2. MODIFIED WILCOXON TEST STATISTICS

Let $Z_{(k)}, k = 1, 2, \cdots, m+n$, be the combined ordered sample based on the two samples, and let $K_{(k)}$ be the indicator of uncensoredness associated with $Z_{(k)}$, i.e., $K_{(k)} = 1$ if $Z_{(k)}$ is uncesored and $K_{(k)} = 0$ otherwise. For each failure time point a $2 \times 2$ contingency table is constructed. For the $i$th failure time point (uncensored observation) the contingency table is as follows:

| | # failure | # alive | |
|---|---|---|---|
| X(Population 1) | $a_i$ | | $n_{i_1}$ |
| Y(Population 2) | | | |
| | $m_{i_1}$ | | $n_i$ |

where $a_i$ is the number of uncensored observation in the first population at time $t_i$ where $t_i$ is the time of the $i$th uncensored observation, and $m_{i_1}$ is the number of uncensored observation in the combined ordered sample at time $t_i$. If we assume no ties, $m_{i_1} = 1$ and $a_i = 0$ or 1. $n_{i_1}$ is the number of $X$'s remaining at time $t_i^-$, and $n_i$ is the number of $Z$'s remaining at time $t_i^-$. Given marginals fixed in the $i$th $2 \times 2$ table, the random variable $A_i$, which is the entry in the (1, 1) cell of the $i$th contingency table, has a hypergeometric distribution with mean $(n_{i_1} \cdot m_{i_1})/n_i$ and variance $\{n_{i_1} \cdot m_{i_1} \cdot (n_i - n_{i_1})(n_i - m_{i_1})\}/\{n_i^2 \cdot (n_i - 1)\}$. After construction of a $2 \times 2$ table for each of the $k$ uncensored observations, Tarone and Ware (1977) suggest weighting each table, forming

$$\sum_{i=1}^{k} W_i\{a_i - E_0(A_i)\} = \sum_{i=1}^{k} W_i\{a_i - \frac{m_{i_1} n_{i_1}}{n_i}\}, \tag{2.1}$$

where $W_i$ is the weight given to each table and $E_0(A_i)$ is the expectation under the null hypothesis for random variable $A_i$, and $a_i, m_{i_1}, n_{i_1}$ and $n_i$ are taken from the $2 \times 2$ table. For the variance, use

$$\sum_{i=1}^{k} W_i^2 Var_0(A_i) = \sum_{i=1}^{k} W_i^2 \{\frac{m_{i_1}(n_i - m_{i_1})}{n_i - 1}\}\{\frac{n_{i_1}}{n_i}(1 - \frac{n_{i_1}}{n_i})\}, \tag{2.2}$$

where $Var_0(A_i)$ is the variance of $A_i$ calculated under the null hypothesis. And the test statistic

$$S^2 = \frac{[\sum_{i=1}^{k} W_i\{a_i - E_0(A_i)\}]^2}{\sum_{i=1}^{k} W_i^2 Var_0(A_i)} \tag{2.3}$$

is shown to be approximately distributed as $\chi^2$ with 1 degree of freedom. This is an approximation to the exact conditional distribution under the null hypothesis.

Radhakrishna(1965) investigated the choice of efficient weights in the combination of $2 \times 2$ tables for various alternatives of the form $f(p_{1i}) - f(p_{2i}) = c$, where $p_{1i}$ and $p_{2i}$ are the response probabilities in the $i$th table; $f(p)$ is a known monotonic function for $0 < p < 1$, and $c$ is a constant. There are three special cases:

(i) Mantel and Haenszel(MH:1959) consider a logit scale which results in $W_i = 1$,

(ii) Gehan(1965) considers a constant scale which results in $W_i = n_i$, and

(iii) Tarone and Ware(TW:1977) suggest an arcsin scale which results in $W_i = \sqrt{n_i}$.

In this paper a probit scale is studied which can be considered a fourth weight.

## 3. WEIGHT BY PROBIT SCALE

The weight which is optimal for constant differences on the probit scale, $\Phi^{-1}(p)$, is estimated from continuous life time data, where

$$\Phi(p) = \int_{-\infty}^{p} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Radhakrishna(1965) showed that the weight by probit scale is

$$W_i = \frac{\exp[-\{\Phi(\Pi_i)\}^2/2]}{\sqrt{2\pi}\Pi_i(1-\Pi_i)} \tag{3.1}$$

where $\Pi_i$ denotes the expected response of marginals for the $i$th table. Since the estimate of $\Pi_i$ is $1/n_i$, the weight $W_i$ is estimated as

$$\widehat{W_i} = \frac{\exp[-\{\Phi(1/n_i)\}^{-2}/2]}{\sqrt{2\pi}} \cdot \frac{n_i^2}{n_i - 1}. \tag{3.2}$$

But using the Mclaurin series and after some manipulation, it can be shown that

$$\exp[-\{\Phi(1/n_i)\}^{-2}/2] = e^{-2}(1 + \frac{8}{\sqrt{2\pi n_i}}) + O(\frac{1}{n_i^2})$$

where $-0.1865 < O(\frac{1}{n_i^2}) < 0.1723$.

Constant multiplications to weight, $W_i$, do not change the test statistics of equation. Hence the estimate of weight becomes from (3.2)

$$\widehat{W_i} = n_i + \frac{8}{\sqrt{2\pi}}$$

by letting $\frac{n_i}{n_i-1} \cong 1$. Thus the weight by probit scale is shown to be a combination of the weights by logit and constant scale of section 2-(i) and (ii).

# 4. ASYMPTOTIC RELATIVE EFFICIENCY

Asymptotic relative efficiencies for test statistics by logit, arcsin and constant scales which result in MH, TW, and Gehan statistics, are compared by Tarone and Ware(1977) using Pitman efficiencies at several alternatives under the assumption of no censorship and equal sample sizes. The alternatives considered are

(i) Lehmann alternative; $F_2(x) = F_1^{\theta}(x)$

(ii) Logistic alternative; $F_2(x) = [1 + e^{\theta}\{1 - F_1(x)\}/F_1(x)]^{-1}$

(iii) Scale alternative; $F_2(x) = F_1(\theta\, x)$

(iv) Translation alternative; $F_2(x) = F_1(x + \theta)$.

In addition, we can introduce a probit scale for comparison. In order to obtain the Pitman efficiencies we adapt the Crowley's representation(see [2]), which is used to provide a proof of the asymptotic normality of a nonparametric two-sample test statistic, under the above assumptions. After some manipulation we can obtain the Pitman efficacy for each statistic, then we know that the efficiencies depend upon the censoring distributions in a complicated manner and the total sample sizes of the combined sample. The null density function, $f(t)$, is specified in the table 1, and the total sample sizes given to the left entry in the table represent the examples of a small sample size and an asymptotic case. For each alternative, Table 1 lists the Pitman efficiencies relative to the statistic which, of the four statistics considered, has the highest efficacy at the specified alternative.

# 5. DISCUSSION

One can observe from Table 1 that the test statistic by probit scale($S_P$) tends to be the same with the test statistic by logit scale($S_G$) as $n \rightarrow \infty$.

**Table 1.** Relative efficiencies of the two sample rank tests when there is no censorship.

| n | $S_{MH}$ | $S_G$ | $S_{TW}$ | $S_P$ |
|---|---|---|---|---|
| Lehman alternative; $f(t) = \exp(-t), t > 0$ | | | | |
| 20 | 100 | 75 | 89 | 78 |
| 50 | 100 | 75 | 89 | 76 |
| $\infty$ | 100 | 75 | 89 | 75 |
| Logistic alternative; $f(t) = \exp(-t), t > 0$ | | | | |
| 20 | 68 | 90 | 87 | 100 |
| 50 | 72 | 95 | 92 | 100 |
| $\infty$ | 75 | 100 | 96 | 100 |
| Scale alternative; $f(t) = 2t\exp(-t^2), t > 0$ | | | | |
| 20 | 100 | 75 | 89 | 78 |
| 50 | 100 | 75 | 89 | 78 |
| $\infty$ | 100 | 75 | 89 | 75 |
| Translation alternative; $f(t) = 2t\exp(-t^2), t > 0$ | | | | |
| 20 | 67 | 100 | 89 | 78 |
| 50 | 67 | 100 | 89 | 75 |
| $\infty$ | 67 | 100 | 89 | 75 |

In equation (3.2), it is obvious that the weight by probit scale shows that the term with $n_i$ will dominate as $n_i \to \infty$. But when $n_i$ is small it seems that $S_P$ neutralizes $S_G$. The exact calculation of test statistics and application to different types of censoring distributions may enlighten the usage of probit scale.

We know that the results in Table 1 are obtained from assuming no censorship, and it may be more practical to calculate the test power using the various simulation methodologies such as Monte-Carlo simulation methodology when there exists censorship. To get the results of applying the simulation methodologies, further study is needed later.

## REFERENCES

(1) Cochran, W. G.(1954). Some methods of strengthening the common $\chi^2$ tests. *Biometrics* 10, 417-51.

(2) Crowley, J.(1974). Asymptotic Normality of A New Nonparametric Statistic for Use in Organ Transplant Studies. *Journal of the American Statistical Association*, 69, 1006-1011.

(3) Gehan. E.(1965). A generalized Wilcoxon test for comparing arbitrarily single censored samples. *Biometrika* 52, 203-23.

(4) Lee, E.T., Desu, M.M. & Gehan, E.A.(1975). A Monte Carlo study of the power of some two-sample tests. *Biometrika* 62, 425-32.

(5) Mantel, N.(1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Rep.* 50, 163-70

(6) Mantel, N. & Haenszel, W.(1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22, 719-48.

(7) Miller Jr., R. G.(1981). *Survival Analysis*, John Wiley & Sons. 81 - 104

(8) Radhakrishna, S.(1965). Combination of results from several $2 \times 2$ contingency tables. *Biometrics* 21, 85-98.

(9) Tarone, R. E. & Ware, J.(1977). On distribution-free tests for equality of survival distributions. *Biometrika* 64, 156-160.