Nonparametric Tests for Comparing Two Treatments in Sequential Trials

Edit Gombay and Giseon Heo
Department of Mathematical Sciences
University of Alberta
Edmonton, Canada
egombay@gpu.srv.ualberta.ca
heo@stat.ualberta.ca

1. Introduction

Sequential procedures that compare two populations' (treatments') survival (failure) times in the presence of censoring are needed in medical trials, as ethical and monetary considerations force researchers to evaluate data as they arrive, instead of just waiting for a fixed sample of size n to be realized. The most frequently used procedures of this type are the repeated significance tests. Another approach, presented here, is to use a stochastic process approximation to the arriving data which allows an inspection of results after each new data-point. We propose some easy to perform sequential procedures based on U-statistics.

2. Test Statistics

Our model is a simplified version of the censored data with staggered entry. Assume subject j in treatment group i enters the study at time $V_j^{(i)}$, $i = 1, 2$, $j = 1, 2, \ldots$, and let $T_j^{(i)}$ denote its survival time measured from entry time $V_j^{(i)}$. Variable $T_j^{(i)}$ may be censored on the right by random variable $C_j^{(i)}$, so we observe $X_j^{(i)} = \min(T_j^{(i)}, C_j^{(i)})$ only. The event time for this subject is $V_j^{(i)} + X_j^{(i)}$. As usual, we assume the independence of the T and C random variables, and that the censoring random variables have the same distribution in both treatment groups. With $\delta_j^{(i)} = I(X_j^{(i)} = T_j^{(i)})$, we can summarize the observations as $\{Z_j^{(i)}\} = \{X_j^{(i)}, \delta_j^{(i)}\}$, $i = 1, 2$, $j = 1, 2, \ldots$. We assume a design, where subjects entering the study are randomly assigned to treatment 1 with probability λ, and to treatment 2 with probability $1 - \lambda$. Let $h(Z^{(1)}, Z^{(2)})$ be an anti-symmetric kernel satisfying some general conditions. Assuming continuous distributions, we have distinct event (failure) times $t_1 < t_2 < \cdots < t_k < \cdots < t_n$. The number of events observed by t_k is k, $k = k^{(1)} + k^{(2)}$, where $k^{(i)}$ is the random number of events in treatment group i, $i = 1, 2$, and we use U-statistic, $U_k = \sum_{i=1}^{k^{(1)}} \sum_{j=1}^{k^{(2)}} h(Z_i^{(1)}, Z_j^{(2)})$ to compare the two populations at time t_k, $k \geq 2$. In U_k observations with event time less than or equal to t_k are considered only.

We obtain the following sequential tests, truncated after n observations. Let α denote the level of significance, $T = log(n)$, $a(T) = (2\,log\,T)^{1/2}$, $d(T) = 2\,log\,T + (1/2)\,log\,log\,T - (1/2)log\,\pi$.

Test 1. *For $k = 2, 3, \ldots, n$ calculate*

$$k^{-3/2}\sigma^{-1}\lambda^{-1/2}(1-\lambda)^{-1/2}|U_k|.$$

Stop and reject H_0 the first time it exceeds critical value $CV_1(\alpha, n)$, where $CV_1(\alpha, n) = a^{-1}(T)[-\log(-1/2\,\log(1-\alpha)) + d(T)]$, otherwise do not reject H_0.

Test 2. *For $k = 2, 3, \ldots, n$ calculate*

$$k^{-1}n^{-1/2}\sigma^{-1}\lambda^{-1/2}(1-\lambda)^{-1/2}|U_k|.$$

Stop and reject H_0 the first time it exceeds critical value $CV_2(\alpha)$, for example, $CV_2(.05) = 2.24$.

3. Application

Slud and Wei (1982) published a subset of the data gathered by the Veterans Administration Cooperative Urological Research Group (VACURG). Cancer of the prostrate was treated by two methods, A and B. The main difference was that treatment A had estrogen treatment, whereas treatment B had placebo instead.

In our application we shall use the Gehan-Gilbert score $h(Z^{(1)}, Z^{(2)}) = I(X^{(1)} > X^{(2)}, \delta^{(2)} = 1) - I(X^{(1)} < X^{(2)}, \delta^{(1)} = 1)$. At $\alpha = 0.05$ level of significance and using $n = 100$ as a truncation value, after the $k = 53$ observations have arrived we could have said that there was difference between the two treatments. This corresponds to 160 months after the start of the study in 1960. Test 1 was not significant this is because Test 2 has more power to detect small derivations from H_0.

4. Conclusion

The new type of sequential tests can monitor censored observations from the very beginning of a clinical trial, hence they may facilitate an early decision. This is desirable in many cases for financial and ethical reasons. Our sequential tests are very simple, non-parametric, and the boundaries do not depend on the data, only on the total number of observations. In some studies using repeated significance tests, the choice of group sizes may affect the outcome. This problem does not arise in our procedures where data is monitored continuously as they arrive.