Journal of the Korean Statistical Society 42 (2013) 227–234

Contents lists available at SciVerse ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Estimation of the treatment effect for the survival time data

Hyo-Il Park *

Department of Statistics, Chongju University, Chongju, Choongbook 360-764, Republic of Korea

ARTICLE INFO

Article history:
Received 23 May 2012
Accepted 24 August 2012
Available online 20 September 2012

AMS 2000 subject classification:
62G05

Keywords:
Asymptotic normality
Convolution
Least square method
Right censored data
Two-sample problem

ABSTRACT

In this paper, we propose a nonparametric estimation procedure for the treatment effect by applying the least square method for the survival time data. We derive an explicit formula for the estimate which is also easy to compute. Then we discuss the consistency and asymptotic normality for the estimate under several assumptions. Then we illustrate our procedure with a numerical example and compare performance by carrying out a simulation study. Finally we discuss some interesting aspects for the estimation procedure as concluding remarks.

1. Introduction

When one takes a comparison study for a newly developed treatment over a control or between two treatments, one may consider comparing survival functions through the hypothesis test assuming a suitable model, which defines a relation between two survival functions. If one can be assured that there must exist any difference between them, then one may be interested in trying to measure the difference. The proportional hazards model is famous and widely used one in the survival analysis (cf. Cox, 1972). Since that defines the proportionality between two hazard functions, it would be inappropriate to be used to measure the difference between two treatments directly. Then a well-known and useful alternative for this purpose would be the location translation model. In short, that assumes that the difference between two quantile points maintains constant over time. Then any difference between two quantile points can be considered as the difference between two treatments and has been called the treatment effect.

The study of the estimation of the treatment effect for the two-sample problem based on the right censored data has been a research topic for a long time, with the estimation problem for the mean-life in the survival analysis. Many authors have reported their research results with some drawback or drawbacks which may be inevitable because of the possibility of censoring of the largest observation. We review some important results in the sequel and note that they are all based on the location translation model. Akritas (1986) proposed an estimate as a median of the convolution (cf. Feller, 1971) of two Kaplan–Meier estimates (cf. Kaplan & Meier, 1958). However a simple median of the convolution may lack consistency since the Kaplan–Meier estimate should be incomplete when the largest observation is right censored. In order to overcome this inconsistency, Bassiakos, Meng, and Lo (1991) and Meng, Bassiakos, and Lo (1991) have strived to modify Akritas' estimate and proposed new estimates by introducing some artificial auxiliary variable, which made the procedures and forms of estimates very complicated and required additional information for the censoring distributions, which are of no interest in our concern. Tsiatis (1990) considered to use the linear rank statistics which may be used for testing the equality between

* Tel.: +82 43 229 8203.
E-mail address: hipark@cju.ac.kr.

1226-3192/$ – see front matter
doi:10.1016/j.jkss.2012.08.003

two distributions or survival functions under the two-sample problem setting as the estimation functions. The resulting estimating procedure requires an iterative computational process and so the estimate cannot be expressed with a closed form. Also Park and Park (1995) considered an estimate by integrating the difference between two quantiles. However the estimate may incur some efficiency loss by deleting or omitting some observations of the data. Zhou and Liang (2005) considered a procedure assuming the distribution of control group is known but that of the treatment group, unknown. However it is difficult to assume a specific distribution or survival function for the survival data. Thus Zhou and Liang's procedure has some intrinsic drawback for real application.

In this study, we consider proposing a new nonparametric estimate procedure which is simple in calculation and easy to use. In the nest section, we obtain a nonparametric estimate by applying the least square method and discuss the asymptotic properties under several assumptions for the estimate. In Section 3, we illustrate our estimate procedure with a numerical example and compare performance by carrying out a simulation study obtaining the empirical confidence interval for the treatment effect. Finally we discuss some interesting aspects for the estimation procedure as concluding remarks.

2. Least square estimate

We consider the following simple linear model.

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_1, \quad i = 1, \ldots, n. \tag{2.1}$$

The covariate x_i takes value 0 or 1 accordingly as the nonnegative response variable Y_i comes from the control or treatment group. Without loss of generality, we assume that the first n_1, $n_1 < n$, number of observations are assigned to the control group and the remaining $n_2 = n - n_1$ number of observations, assigned to the treatment group. Also we assume that the distribution F of the error term ϵ_i is unknown but continuous with mean 0 and finite variance, which is also unknown. Then we have the following results:

$$E(Y_i) = \beta_0, \quad i = 1, \ldots, n_1$$
$$E(Y_i) = \beta_0 + \beta_1, \quad i = n_1 + 1, \ldots, n.$$

Therefore β_1 can be considered as a location translation parameter and should take on the role of the treatment effect. Based on this two-sample problem setting, a lot of research results for the estimation of β_1 have been proposed using the parametric and nonparametric methods when no censoring is involved. However in this study, we will consider the possibility of right censoring for Y_i. For this let $C_1, \ldots, C_n$ be the censoring random variables independent of $Y_1, \ldots, Y_n$ with arbitrary censoring distributions G_1 for $1 \le i \le n_1$ and G_2 for $n_1 + 1 \le i \le n$. Then we can only observe that for each i, $i = 1, \ldots, n$

$$T_i = \min\{Y_i, C_i\} \quad \text{and} \quad \delta_i = I(Y_i \le C_i), \tag{2.2}$$

where $I(\cdot)$ is an indicator function. Thus if T_i is censored then $\delta_i = 0$ and uncensored, $\delta_i = 1$. Under this random censoring scheme, by considering β_0 as a nuisance parameter, one may obtain an estimate of β_1 by finding β_1 which minimizes the following equation

$$Q_M(\beta_0, \beta_1) = \sum_{i=1}^{n} (T_i - \beta_0 - \beta_1 x_i)^2 \Delta F_n(T_i - \beta_0 - \beta_1 x_i). \tag{2.3}$$

In (2.3), F_n is the Kaplan–Meier estimate of F based on

$$T_1 - \beta_0 - \beta_1 x_1, \ldots, T_n - \beta_0 - \beta_1 x_n$$

and $\Delta F_n(t) = F_n(t) - F_n(t-)$ is the jump size of F_n at t. Therefore when $\delta_i = 0, \Delta F_n(T_i - \beta_0 - \beta_1 x_i) = 0$ and $\Delta F_n(T_i - \beta_0 - \beta_1 x_i) > 0$ if $\delta_i = 1$. This version of the application of the least square method was initiated by Miller (1976) and requires an iterative procedure for calculation since the value of $F_n(T_i - \beta_0 - \beta_1 x_i)$ can vary with the values of β_0 and β_1. Also one may estimate β_1 using the method of Buckley and James (1979) by introducing pseudo random variables for the censored observations and minimizing a similar equation with (2.3). However the use of an iterative procedure for calculation has been inevitable.

Now we note the following fact about the model (2.1). Under the model (2.1), for any two observations Y_i and Y_j, $1 \le i \le n_1$ and $n_1 + 1 \le j \le n$, we have that with the notation that $\epsilon_{ij} = \epsilon_j - \epsilon_i$

$$Y_j - Y_i = \beta_1 + \epsilon_{ij}.$$

We note that the distribution H of the difference $Y_j - Y_i$ takes β_1 as its mean or median since the distribution of ϵ_{ij} should be symmetric about 0. In order to proceed our discussions more concretely, let F_1 and F_2 be the distribution functions of Y_i and Y_j, respectively. From (2.1) and the assumptions introduced up to now, the location translation model holds for F_1 and F_2 such as for any real number y,

$$F_2(y) = F_1(y - \beta_1).$$