Volume: 10 Number: 01 Page: 85-94 ISSN No: 1303-6319 Cilt: 10 Sayı: 01

HAZARD CHANGE POINT PROBLEMS FOR CENSORED AND TRUNCATED DATA

Ülkü GÜRLER* Deniz YENİGÜN**

ABSTRACT

This paper is a summary of the work by Gürler and Yenigün (2011) prepared for the conference in memory of the late Professor Yalçın Tuncer. We consider the estimation of the hazard change point when the observations are subject to random censoring and truncation. The hazard function is assumed to have single jump and otherwise piecewise constant. Full and conditional likelihood approaches are considered and the conditions under which they perform better are discussed.

Keywords: Censoring, Estimation, Hazard Function, Truncation.

1. HAZARD FUNCTION

Hazard function is one of the important functions in reliability, survival, actuarial and other studies, which quantifies the instantaneous risk of failure of an item at a given time point. From statistical point of view, main problems of interest are the estimation of the hazard function, and the estimation of the effects of the covariates. There has been a huge literature on the characteristics and the estimation of hazard functions which are well known, hence the related literature is omitted here. Various functional forms of survival functions that are observed in applications are depicted in Figure 1 below.

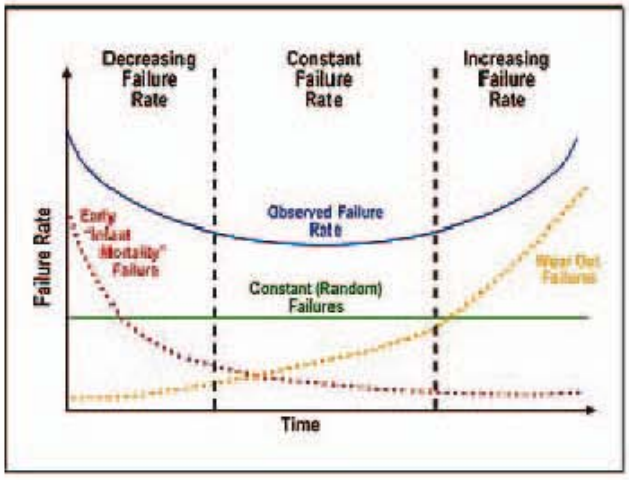

Figure 1: Various failure rate functions

*Prof. Dr., Bilkent University, Department of Industrial Engineering, Ankara, e-mail: ulku@bilkent.edu.tr
**Asst. Prof. Dr., Bilkent University, Faculty of Business Administration, Ankara e-mail: yenigun@bilkent.edu.tr

2. CENSORING AND TRUNCATION

In medical follow-up or reliability studies the random variable of interest (lifetime) may not be fully observable. Common forms of incomplete data are censoring and truncation. There is a vast literature on the analysis of censored and truncated data, including Lynden-Bell (1971), Woodroofe (1985), Kalbeisch and Lawless (1991), Gürler and Wang (1993), Gürler and Prewitt (2000), Frobish and Ebrahimi (2009), Dupuy (2009).

Right Censoring: Right Censoring occurs when event is observed only if it occurs prior to some predetermined time. Let X denote the random variable of interest (lifetime), let C denote the censoring variable. Define $T = \min(X, C)$ and let $\delta = I(T = X)$. Here X and C are independent and nonnegative. Then in the presence of right censoring, one observes the pairs (T, δ). The observed data is denoted by (t_i, δ_i) for i= 1,...,n.

Right Truncation: Right Truncation occurs if time origin of lifetime is after the time origin of the study. Let X denote the random variable of interest (lifetime), let Y denote the truncation variable. Here X and Y are independent and nonnegative. The pair (X,Y) is observable only if $X \leq Y$. In other words, observations come from conditional distribution of (X,Y), given that $X \leq Y$. The observed data is denoted by (x_i, y_i) for i=1,...,n.

A well-known example of right truncation is the Transfusion Related (TR) AIDS Data. Here X is the time from transfusion to diagnosis of AIDS, Y is the time from transfusion to the end of study. Various realizations of X and Y are illustrated in Figure 2.

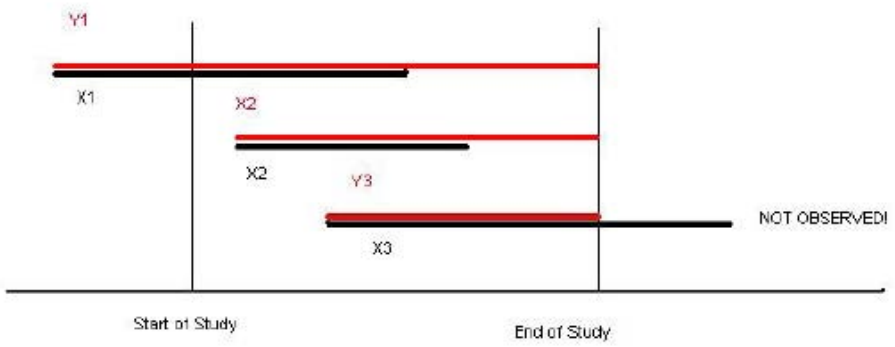

Figure 2: TR-AIDS Data

Left Truncation Right Censoring: Consider a random variable of interest X, representing the time until an event occurs, which may correspond to the survival time of a patient after a treatment or the time until failure of a component. Let Y and C be the truncation and censoring variables respectively, which prevent the complete observation of the variable X. We assume that X, Y, C are independent and

nonnegative. Let $T = min(X, C)$, and $\delta = I(T = X)$, where I is the indicator function. In the presence of left truncation and right censoring, instead of observing independent and identically distributed (i.i.d.) samples of X, we observe triplets (T, Y, δ) only if $Y \le T$, otherwise nothing is observed. Thus the observations come from the conditional distribution of (T, Y, δ), given that $Y \le T$. The observed data are given by a set of i.i.d. observations (t_i, y_i, δ_i) for $i = 1, ..., n$.

3. HAZARD CHANGE-POINT MODELS

In some cases, abrupt changes in the hazard function may be observed. For example, lag for effectiveness of a treatment may change over time. It is of interest to detect the location and the size of the change. One of the earliest works that consider changes in the hazard function is by Matthews and Farewell (1982) which studied a piecewise constant hazard model with a single change-point given by

$$\lambda(t) = \begin{cases} \beta & 0 \le t < \tau \\ \beta + \theta & t \ge \tau \end{cases}, \tag{1}$$

where β and $\beta + \theta > 0$. Here β represents the initial constant hazard rate, θ represents the size of the change in the hazard rate, and τ is the location of the change-point, all of which are unknown.

4. PRELIMINARIES

Suppose X has the hazard function λ as given in (1). Then, the p.d.f. f, the c.d.f. F, the survival function S, and the cumulative hazard function Λ of X are as given below, which are all piecewise functions.

$$f(x) = \begin{cases} \beta e^{-\beta x} \equiv f_1(x) & 0 \le x < \tau \\ (\beta + \theta) e^{-\beta x - \theta(x-\tau)} \equiv f_2(x) & x \ge \tau \end{cases}, \tag{2}$$

$$F(x) = \begin{cases} 1 - e^{-\beta x} \equiv F_1(x) & 0 \le x < \tau \\ 1 - e^{-\beta x - \theta(x-\tau)} \equiv F_2(x) & x \ge \tau \end{cases}, \tag{3}$$

$$S(x) = \begin{cases} e^{-\beta x} \equiv S_1(x) & 0 \le x < \tau \\ e^{-\beta x - \theta(x-\tau)} \equiv S_2(x) & x \ge \tau \end{cases} \tag{4}$$

and

$$\Lambda(x) = \begin{cases} \beta x & 0 \le x < \tau \\ \beta x + \theta(x - \tau) & x \ge \tau \end{cases}. \tag{5}$$

5. CONSTRUCTING THE LIKELIHOOD FUNCTION WITH THE FULL LIKELIHOOD APPROACH

As described above, in the left truncation and right censoring model one observes triplets (T,Y,δ) only if $Y \le T$, otherwise nothing is observed. Hence the observed variables belong to the following conditional distribution:

$F_1 \equiv F_1(t,y,\delta \mid Y \le T) = P(T \le t; Y \le y; \delta \mid Y \le T)$. Let $\alpha = P(Y \le T)$ be the probability that a (Y,T) pair is observed without truncation. We can write α more explicitly as follows:

$$\alpha = P(Y \le T) = P(Y \le \min(X,C) = \int_0^\infty \int_y^\infty \int_y^\infty f(x)h(c)g(y)dcdxdy$$

$$= \int_0^\infty \overline{F}(y)\overline{H}(y)dG(y). \tag{6}$$

We decompose F_1 into two parts, the sub-distribution function of uncensored observations, F_u, and the sub-distribution function of censored observations, F_c. These distributions can be expressed as follows:

$$F_u \equiv F_u(t,y,\delta = 1 \mid Y \le T) = P(T \le t, Y \le y, \delta = 1 \mid Y \le T)$$

$$= P(T \le t, Y \le y, \delta = 1, Y \le T)\alpha^{-1}$$

$$= \alpha^{-1} \int_0^t \int_u^y \overline{H}(x)dF(x)dG(u). \tag{7}$$

The corresponding sub-density of censored observations is

$$f_u(t,y) = \frac{\partial F_u}{\partial y \partial t} = \alpha^{-1} g(y)\overline{H}(t) f(t). \tag{8}$$

Similarly, the sub-distribution function of censored observations is

$$F_c \equiv F_c(t,y,\delta = 0 \mid Y \le T) = P(T \le t, Y \le y, \delta = 0 \mid Y \le T)$$

$$= P(T \le t, Y \le y, \delta = 0, Y \le T)\alpha^{-1}$$

$$= \alpha^{-1} \int_0^y \int_u^t \overline{F}(c)dH(c)dG(u), \tag{9}$$

and the corresponding sub-density function is

$$f_c(t,y) = \frac{\partial F_c}{\partial y \partial t} = \alpha^{-1} g(y)\overline{F}(t)h(t). \tag{10}$$

Now consider the observed sample (t_i, y_i, δ_i) for $i = 1,...,n$. The likelihood contribution of an observed uncensored triplet (t_j, y_j, δ_j) for some $j \in \{1,...,n\}$ is $f_u(t_j, y_j)$, and the likelihood contribution for an observed censored triplet (t_k, y_k, δ_k) for some $k \in \{1,...,n\}$, $k \neq j$, is $f_c(t_k, y_k)$. Then the likelihood function can be written as follows:

$$L = \prod_{i=1}^{n} \alpha^{-1} g(y_i)[\bar{H}(t_i) f(t_i)]^{\delta_i}[\bar{F}(t_i) h(t_i)]^{1-\delta_i}. \quad (11)$$

6. CONSTRUCTING THE LIKELIHOOD FUNCTION WITH THE CONDITIONAL LIKELIHOOD APPROACH

Klein and Moeschberger (2003) summarized the likelihood construction techniques frequently used in survival analysis literature. According to this construction, various types of censoring and truncation schemes have different contributions to the likelihood function. For example, if X is a random variable of interest with p.d.f. f and survival function S, and if X is subject to right censoring, then the contribution of an observed exact lifetime x to the likelihood function is given by $f(x)$, and the contribution of an observed censoring time c to the likelihood function is given by $S(c)$. When we generalize this approach to the left truncation and right censoring model, we have the following.

Recall that in the left truncation and right censoring model, one observes the triplets (T, Y, δ) only if $Y \leq T$, otherwise nothing is observed. Consider an observed random sample (t_i, y_i, δ_i) for $i = 1,...,n$. In this case, the contribution of an exact lifetime $(t_i = x_i)$ to the likelihood function is $f(x_i)/S(y_i)$, and the contribution of an observed censoring time $(t_i = c_i)$ to the likelihood function is $S(c_i)/S(y_i)$. Putting together all the components, one may write the conditional likelihood function as

$$L \propto \prod_{i \in D} \frac{f(x_i)}{S(y_i)} \prod_{i \in R} \frac{S(c_i)}{S(y_i)}, \quad (12)$$

where D is the set of observations where the real lifetimes are observed and R is the set observations where the censoring times are observed only.

7. MAXIMUM LIKELIHOOD ESTIMATION

When we construct the likelihood function for the piecewise constant hazard model (1), is not differentiable with respect to τ, hence it is not possible to find the M.L.E.'s for Ψ using standard methods. Therefore, we take the same approach as in Section 3, where we first fix the value of τ and find the M.L.E.'s for the remaining parameters as a

function of τ. Then we search for the value of τ as our estimator, which maximizes the likelihood function over a number of grid points on a specific interval $[\tau_0, \tau_1]$.

8. FULL LIKELIHOOD APPROACH

The censoring variable C and the truncation variable Y are both assumed to have exponential distributions with rate parameters γ and ν respectively. Note that the hazard function of the censoring variable is $\lambda_c(t) = \gamma$, and the hazard function of the truncation variable is $\lambda_y(t) = \nu$.

Consider an observed random sample (t_i, y_i, δ_i) for $i = 1,...,n$. For a fixed τ, let A and B denote the set of observations such that $t_i \le \tau$ and $t_i > \tau$, respectively. Formally, $A = \{i : t_i \le \tau\}$, $B = \{i : t_i > \tau\}$. Let n_{1A} denote the number uncensored observations that are less than or equal to τ, n_{1B} denote the number of uncensored observations that are larger than τ, and n_B denote the number of observations that are larger than τ. Let $\bar{t}$, $\bar{y}$ and $\bar{\delta}$ denote the sample means. Then for a fixed τ, after some steps the log-likelihood function (12) can be written as

$$\log L = n \log\left(\frac{\gamma\nu}{\alpha}\right) - (\gamma + \beta) n\bar{t} - \nu n\bar{y} - \theta\left(\sum_B t_i - n_B\tau\right)$$

$$+ n_{1A} \log\beta + n_{1B} \log(\beta + \theta) - n\bar{\delta} \log\gamma, \qquad (13)$$

where $w = \beta + \gamma + \nu$

and $$\alpha = P(Y \le T) = \frac{\nu}{w} - \frac{\nu\theta e^{-w\tau}}{w(w+\theta)}, \qquad (14)$$

Taking the derivative of $\log L$ with respect to the unknown parameters, the score vector (13) is computed as

$$U(\Psi_\tau) = \begin{bmatrix} \frac{\partial \log L}{\partial \beta} \\ \frac{\partial \log L}{\partial \theta} \\ \frac{\partial \log L}{\partial \gamma} \\ \frac{\partial \log L}{\partial \nu} \end{bmatrix} = \begin{bmatrix} \frac{nE_1}{\alpha} - \sum_{i=1}^{n} t_i + \sum_A \delta_i \frac{1}{\beta} + \sum_B \delta_i \frac{1}{\beta+\theta} \\ \frac{nE_2}{\alpha} - \sum_B (t_i - \tau) + \sum_B \delta_i \frac{1}{\beta+\theta} \\ \frac{n}{\alpha} E_1 + \frac{n}{\gamma} - \sum_{i=1}^{n} t_i - \sum_{i=1}^{n} \delta_i \frac{1}{\gamma} \\ -n\left(\frac{1}{\nu} - \frac{E_1}{\alpha}\right) + \sum_{i=1}^{n}\left(\frac{1}{\nu} - y_i\right) \end{bmatrix}.$$

Here the quantities E_1 and E_2 are given by

$$E_1 = \int_0^\infty y\overline{F}(y)\overline{H}(y)dG(y)$$

$$= \frac{\nu}{w^2} + \nu \exp - w\tau\left[\frac{\tau(w+\theta)+1}{(w+\theta)^2} - \frac{\tau w+1}{w^2}\right]$$

and

$$E_2 = \int_\tau^\infty (y-\tau)\overline{F}(y)\overline{H}(y)dG(y) \ = \frac{\nu \exp - w\tau}{(w+\theta)^2}.$$

For the fixed τ, the M.L.E. $\hat{\Psi}_\tau = (\hat{\beta}, \hat{\theta}, \hat{\gamma}, \hat{\nu})_\tau$ for Ψ_τ is obtained as the solution to the system of equations $U(\Psi_\tau) = 0$. This solution is obtained by numerical methods since closed form expressions cannot be obtained.

9. CONDITIONAL LIKELIHOOD APPROACH

Let us start with the problem of finding the M.L.E.'s of β and θ for a fixed τ. Consider an observed random sample (t_i, y_i, δ_i) for $i = 1,...,n$. Note that the p.d.f. and the survival function of X are piecewise functions as described in (2) and (4). Then for a fixed τ, there are six possible types of observations that have different contributions to the likelihood function. Let

$$A = \{(t_i, y_i, \delta_i) : \delta_i = 1,\ y_i < x_i \le \tau\},\ B = \{(t_i, y_i, \delta_i) : \delta_i = 1,\ y_i \le \tau < x_i\},$$

$$C = \{(t_i, y_i, \delta_i) : \delta_i = 1,\ \tau < y_i < x_i\},\ D = \{(t_i, y_i, \delta_i) : \delta_i = 0,\ y_i < c_i \le \tau\},$$

$$E = \{(t_i, y_i, \delta_i) : \delta_i = 0,\ y_i \le \tau < c_i\},\ F = \{(t_i, y_i, \delta_i) : \delta_i = 0,\ \tau < y_i < c_i\}.$$

For example, A denotes the set of observed triplets for which t is an actual lifetime x, and both x and the observed truncation variable y are less than τ. The contribution of such (x, y) pairs to the likelihood function is $f_1(x)/S_1(y)$, where f_1 and S_1 are as described in (2) and (4). We define the sets B, C, D, E, F, and their likelihood contributions similarly. Let n_A denote the number of observed triplets in set A, let $\prod_A$ denote the product over set A, let $n_{B,C}$ denote the total number of observed triplets in sets B and C, and let $\sum_{B,C}$ denote the sum over sets B and C. Define all the other related subscripts similarly. Then for a fixed τ, we can write the likelihood function as follows:

$$L(\beta, \theta \mid y, t, \delta) = \prod_A \frac{f_1(x_i)}{S_1(y_i)} \prod_B \frac{f_2(x_i)}{S_1(y_i)} \prod_C \frac{f_2(x_i)}{S_2(y_i)} \prod_D \frac{S_1(c_i)}{S_1(y_i)} \prod_E \frac{S_2(c_i)}{S_1(y_i)} \prod_F \frac{S_2(c_i)}{S_2(y_i)}$$

$$= \beta^{n_A}(\beta+\theta)^{n_{B,C}}$$

$$\exp\left\{n\beta(\bar{y}-\bar{t})+\theta\left[\sum_{C,F} y_i - \sum_{B,C} x_i - \sum_{E,F} c_i + n_{B,E}\tau\right]\right\}.$$

Following the notation above, for a fixed τ, let $\Psi_\tau = \{\beta,\theta\}_\tau$ be the parameter set to be estimated, and let $U(\Psi_\tau)$ be the corresponding score vector composed of the first derivatives of the log-likelihood function, which is given by

$$U(\Psi_\tau) = \begin{bmatrix} \frac{\partial \ln L}{\partial \beta} \\ \frac{\partial \ln L}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \frac{n_A}{\beta} + \frac{n_{B,C}}{\beta+\theta} + n(\bar{y}-\bar{t}) \\ \frac{n_{B,C}}{\beta+\theta} + \sum_{C,F} y_i - \sum_{B,C} x_i - \sum_{E,F} c_i + n_{B,E}\tau \end{bmatrix}. \quad (15)$$

Then, the M.L.E. $\hat{\Psi}_\tau = (\hat{\beta},\hat{\theta})_\tau$ for Ψ_τ is obtained as the solution of the system of equations $U(\Psi_\tau) = 0$, which results in the estimators:

$$\hat{\beta} = \frac{n_A}{\sum_{C,F} y_i - \sum_{B,C} x_i - \sum_{E,F} c_i + \tau n_{B,E} - n(\bar{y}-\bar{t})} \quad (16)$$

and $$\hat{\theta} = \frac{-n_{B,C}}{\sum_{C,F} y_i - \sum_{B,C} x_i - \sum_{E,F} c_i + \tau n_{BE}} - \hat{\beta}. \quad (17)$$

Let $\tau_i \in [\tau_0,\tau_1]$, $i = 1,...,m$ denote the fixed grid points in the search interval and let L_{τ_i} denote the maximum of the likelihood function for $\tau = \tau_i$. That is

$$L_{\tau_i} = L(\{\hat{\beta},\hat{\theta}\}_{\tau_i},\tau_i).$$

Then the proposed estimators for the change-point τ and the rest of the parameters are given by

$$\hat{\tau} = argmax_{\tau_i} L_{\tau_i} \quad (18)$$

and $$\hat{\Psi} = \{(\hat{\beta},\hat{\theta})_{\hat{\tau}},\hat{\tau}\}. \quad (19)$$

10. COMPARISON OF FULL AND CONDITIONAL LIKELIHOOD METHODS

The full likelihood model specifies parametric families of distributions for the censoring and truncation variables and it is expected to give more accurate results when the model

specification is correct. Its drawbacks are the risk of model misspecification, the large number of parameters to be estimated, and the lack of closed form estimators which forces one to use numerical methods for estimation. Sometimes the numerical methods may not lead to the maximum likelihood estimators, especially for small sample sizes. The conditional likelihood model on the other hand, does not assume any parametric families of distributions for censoring and truncation variables, and focuses only on estimating the model parameters of the hazard function. This simpler approach provides closed form estimators for the model parameters and does not have the risk of model misspecification. The conditional likelihood approach, however, emphasizes more the observed values of the censoring and truncation variables and somewhat overlooks their random nature. This results in increased bias and variance especially for small samples, which disappears as the sample size increases.

For a large set of parameter seetings, numerical experiments are done in Gürler and Yenigün (2001) and the following conclusions are drawn: (i) Both methods can easily be implemented and their performances are comparable; (ii) When distributional assumptions of full likelihood method are correct, two methods are close for estimating location of change-point and initial hazard rate; (iii) Full likelihood performs better for estimating size of change and the difference tends to vanish as the sample size increases; (iv) Full likelihood is not robust to model misspecification and in some cases it is outperformed by conditional likelihood model.

11. REFERENCES

Dupuy, J. F., 2009. Detecting Change in a Hazard Regression Model with Right-Censoring. Journal of Statistical Planning and Inference 139, 1578 -1586.

Frobish, D., Ebrahimi, N., 2009. Parametric Estimation of Change-Points for Actual Event Data in Recurrent Events Models. Computational Statistics and Data Analysis 53, 671-682.

Gürler, Ü., Prewitt, C., 2000. Bivariate Density Estimation with Randomly Truncated Data. Journal of Multivariate Analysis, Vol. 74, 88-115.

Gürler, Ü., Wang, J. L., 1993. Nonparametric Estimation of Hazard Functions and Their Derivatives Under Truncation Model. The Annals of the Institute of Statistical Mathematics, Vol. 45, 249-264.

Gürler, Ü., Yenigün, D., 2001. Full and Conditional Likelihood Approaches for Hazard Change-Point Estimation with Truncated and Censored Data. Computational Statistics & Data Analysis, Vol. 55, 2856-2870.

Kalbfleisch, J. D., Lawless, J. F., 1989. Inference Based on Retrospective Ascertainment: An Analysis of The Data on Transfusion-Related AIDS. Journal of the American Statistical Association 84, 360-372.

Klein, J. P., Moeschberger, M.L., 2003. Survival Analysis, Second Ed., Springer, New York.

Lynden-Bell, D., 1971. A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars. Monthly Notices of the Royal Astronomical Society, Vol. 155, p. 95.

Matthews, D. E., Farewell, V. T., 1982. On Testing for a Constant Hazard Against a Change-Point Alternative. Biometrics 38, 463-468.

Woodroofe, M., 1986. Estimating a Distribution Function with Truncated Data. The Annals of Statistics. Vol. 13, No. 1., 163-177.

SANSÜRLÜ VE BUDANMIŞ VERİLER İÇİN DEĞİŞİM NOKTASI PROBLEMLERİ

ÖZET

Bu makale, Gürler ve Yenigün'ün (2011) merhum Profesör Yalçın TUNCER'in anısına düzenlenmiş olan konferans için hazırlanmış çalışmasının bir özetidir. Gözlemler rastgele sansür ve budanmaya tabi olduğunda tehlike değişim noktasının tahmini dikkate alınmıştır. Tehlike fonksiyonunun tek bir sıçrama yaptığı ve bunun dışında parçalı sabit olduğu varsayılmıştır. Tam ve koşullu olabilirlik yaklaşımları düşünülmüş ve daha iyi performans gösterdikleri koşullar tartışılmıştır.

Anahtar Kelimeler: Budanmış veri, Hazard fonksiyonu, Kesikli veri, tahmin.