Interval Estimation for a Change in the Hazard Rate with Staggered Entry

Dong–Yun Kim and Yanhong Wu
Illinois State University, University of the Pacific

ABSTRACT

We construct a confidence interval for a change in hazard rate of the patient's survival distribution when the patients enter the trial at random times. We show that the local-likelihood ratio process converges weakly to a certain process and obtain the maximum distribution of the process which does not depend on the change point. We illustrate the method using the Stanford Heart Transplant data. Using the Monte Carlo simulation we compare the limiting distribution to the empirical density function and discuss the empirical coverage probability of the confidence interval.

Key Words and Phrases: change point; local likelihood ratio process; hazard rate; staggered entry.

1. INTRODUCTION

Suppose that the patients arrive for treatment at times $0 < \tau_1 < \tau_2 < \cdots,$ following a Poisson process with known rate γ. Let N be the total number of patients who arrived in the time interval $[0, T]$ and for $i = 1, \ldots, N$, let Y_i be the survival time of the ith patient, and suppose that the density of Y_i is of the form

$$f_Y(y) = \begin{cases} \lambda_1 e^{-\lambda_1 y} & \text{if } y < \nu; \\ \lambda_2 e^{-\lambda_1 \nu - \lambda_2 (y-\nu)} & \text{if } y \geq \nu, \end{cases} \tag{1}$$

where $0 < \lambda_1 < \lambda_2,\ \nu$ are unknown parameters. That is, the failure rate may change at an unknown time ν.

The purpose of this paper is to construct a confidence interval for the change point ν in the hazard rate of the survival distribution based on the observations

$$\begin{aligned} X_i &= \min(Y_i, T - \tau_i), \\ \delta_i &= \mathbf{1}\{Y_i \leq T - \tau_i\}, \end{aligned} \tag{2}$$

where $\mathbf{1}\{A\}$ denotes the indicator of an event A. The log-likelihood function for this data is

$$\ell(\lambda_1, \lambda_2, \nu) = K_1(\nu)\log(\lambda_1) - \lambda_1 T_1(\nu) + K_2(\nu)\log(\lambda_2) - \lambda_2 T_2(\nu)$$

where

$$
\begin{aligned}
K_1(\nu) &= \sum_{i=1}^{N} \mathbf{1}\left\{X_i < \nu,\ \delta_i = 1\right\}, \\
K_2(\nu) &= \sum_{i=1}^{N} \mathbf{1}\left\{X_i \geq \nu,\ \delta_i = 1\right\}, \\
T_1(\nu) &= \sum_{i=1}^{N} \left(X_i \wedge \nu\right), \\
T_2(\nu) &= \sum_{i=1}^{N} \left(X_i - \nu\right)^{+},
\end{aligned}
\tag{3}
$$

$x \wedge \nu = \min(x,\, \nu)$, and $x^{+} = \max(x,\, 0)$. Given ν, the maximum likelihood estimator (MLE) of $\lambda_i(\nu)$ is $\hat{\lambda}_i = K_i(\nu)/T_i(\nu)$ for $i = 1, 2$, respectively. So the profile log-likelihood is

$$
\begin{aligned}
\ell(\nu) &= \ell(\hat{\lambda}_1(\nu), \hat{\lambda}_2(\nu), \nu) \\
&= K_1(\nu)\left\{\log\left(\frac{K_1(\nu)}{T_1(\nu)}\right) - 1\right\} + K_2(\nu)\left\{\log\left(\frac{K_2(\nu)}{T_2(\nu)}\right) - 1\right\}.
\end{aligned}
\tag{4}
$$

The MLE of ν is obtained by maximizing (4) with respect to ν. This must be done numerically.

Several authors have considered the problem of change point in the hazard function. Yao (1986) studied the model in (1) and he derived the MLE of ν under some constraints. He also obtained the limiting distribution of MLE, and used a related quantity to construct a confidence interval for the change point.

On the other hand, Siegmund (1988) advocated the use of alternatives to the maximum likelihood estimator for detecting a change point in the context of drift of Brownian motion, citing his own and others' work. He considered the change point problem in discrete time setting and concluded that there is practically no difference between the confidence set consisting of (often) disconnected intervals obtained by inversion from the log-likelihood process and a confidence interval by joining all the disconnected intervals by taking the minimum of those intervals as the left-end point and the maximum as the right-end point.

Extending the work of Siegmund to the continuous time frame, Loader (1991) also considered the model in (1) in a non-staggered entry and instead assumed the exponential random censoring mechanism. Applying the large deviation technique, he also used the inversion method to obtain the confidence region.

The current work employs the likelihood ratio based approach to construct the confidence interval, by encompassing all the disconnected confidence sets with a single confidence interval, just as described in Siegmund. It is different from the work of Siegmund, however, in that the model is considered in a continuous time interval. It is also different from Loader's work in that the current model allows the staggered entry of data at random times and it considers Type I censoring.

In the next section, we find the limiting distribution of

$$\ell(\hat{\nu}) - \ell(\nu) = \sup_{-\infty<u<\infty} \ell\left(\nu + \frac{u}{\gamma T}\right) - \ell(\nu) \tag{5}$$

by first establishing the weak convergence of the approximate local likelihood ratio process and then obtaining the explicit form of the distribution of the supremum of the limiting process, which proves to be independent of the change point ν.

Based on these asymptotic results, we consider the confidence region of the form $\{\nu | \ell(\nu) \geq \ell(\hat{\nu}) - c\}$ for some $c > 0$. In Section 3, we use the Stanford Heart Transplant data to illustrate our method. We show that the data set fits the model assumptions quite well. Using the result by Kim et al. (2004), we note that we can formally test whether a change point exists in a given interval, and show that in fact the test proves to be significant at 1%. We obtain the confidence interval for the change point.

In Section 4, we use Monte Carlo simulation to compare the limiting density function to the empirical density function constructed from the simulated data. We also discuss how to construct the confidence interval and comment on the empirical coverage probability of the confidence interval with nominal confidence level.