

## 모수적과 비모수적 위험률 변화점 통계량 비교<sup>†</sup>

김재희<sup>1</sup> · 이시은<sup>2</sup>

<sup>1</sup>덕성여자대학교 정보통계학과 · <sup>2</sup>아주대학교 의과대학 의과학연구소

접수 2016년 8월 10일, 수정 2016년 9월 20일, 게재확정 2016년 9월 24일

### 요약

위험률에 변화점이 존재할 경우 위험률 변화점에 대한 추정 정확한 모수 추정을 위해 매우 필요하다. 본 연구에서는 한 개 위험률 변화점이 존재하는 경우 위험률의 변화점 추정량에 대한 비교 연구를 수행하였다. 우도함수에 기반한 모수적 방법인 Matthews와 Farewell (1982) 위험률 변화점 추정량과 Nelson-Aalen 누적 위험률에 기반한 비모수적 방법의 Zhang 등 (2014) 위험률 변화점 통계량을 고찰하여 특성을 파악하였다. 모의실험에서 지수분포를 따르는 생존데이터에 대해 위험률 변화점이 한 개 있는 경우 중도절단이 없는 경우와 중도절단이 있는 경우 위험률 추정량의 능력을 평균제곱오차를 계산하여 비교하였다. 실제 데이터에 대한 적용으로 백혈병 생존데이터와 원발성 담배중 경화 생존데이터에 대해 위험률 변화점을 추정하고 비교해 보았다.

주요용어: 누적위험함수, 생존함수, 위험률 변화점, 중도절단, 지수분포.

### 1. 서론

생존함수 또는 위험함수에 변화점이 존재할 경우에는 변화점 전, 후에 대한 위험률을 다르게 추정해야 하는 문제가 발생한다. 이러한 문제를 위험률 변화점 문제라 한다. 위험률에 변화점이 존재할 경우 위험률 변화점 문제의 해결은 정확한 모수 추정을 위해 매우 필요하다.

우도함수 기반 변화점정에 대한 연구로 Matthews와 Farewell (1982), Matthews 등 (1985), Yao (1986), Worsley (1988), Henderson (1990), Loader (1991) 등이 있다. Nguyen 등 (1984)은 제시된 혼합모형의 밀도의 특성을 조사하여 연속적인 변화점을 추정하였다. Pham과 Nguyen (1990)과 Luo 등 (1997)은 변화점에 대한 우도비 검정을 제안했다. Gijbels와 Gürler (2003)는 자료에 임의 중도절단 (random censoring)이 있는 경우 구분적으로 연속인 (piecewise continuous) 위험률 함수의 변화점을 추정하는 문제를 제안했고, Dupuy (2006)는 위험률과 회귀 모수에 변화점이 있는 모수적 생존회귀모형을 제안했다. Kim (2009)은 상대위험도를 이용해 위험률 변화점 추정법을 제안하였다. Chang 등 (1994)은 Nelson-Aalen 누적위험함수 기반 비모수적 위험률 변화점 추정법을 제안했으며 Goodman 등 (2010)은 위험률이 계단함수를 형성하는 경우 다중 변화점 추정법을 제안했다. Williams과 Kim (2011)은 알려지지 않은 변화점을 가진 두 개의 연속적이고 단조로운 함수로 표현되는 위험률에 대한 우도비 검정을 제안하였다. Zhang 등 (2014)은 Chang 등 (1994) 방법을 기반으로 확장하여 비모수적인 방법으로 한 개 또는 여러 개 변화점 추정량을 제안하였다. Jeong (1998), Jeong과 Han (1998)은 위험률 변화점에 대한 우도함수 기반 검정과 비모수적 변화점 추정을 제안했고 Shin과 Kim (2014)은 불안정한 관측을 가진 재발 사건 자료를 분석하였다. 최근 Lee 등 (2015)는 직장암 (rectal cancer) 데이터에 대해 생존함수 추정과 더불어 위험률 변화점도 추정하였다.

<sup>†</sup> 본 연구는 2015년도 덕성여자대학교 교내 연구비 지원을 받았습니다.

<sup>1</sup> 교신저자: (132-714) 서울시 도봉구 삼양로 144길 33, 덕성여자대학교 정보통계학과, 교수.

E-mail: jaehee@duksung.ac.kr

<sup>2</sup> (16499) 경기도 수원시 영통구 월드컵로 164, 아주대학교 의과대학 의과학연구소, 통계학석사.

## 2. 위험률 변화점 통계량

우도함수 기반 변화점 추정량으로 Matthews와 Farewell (1982)과 Zhang 등 (2014)의 비모수적 변화점 추정량을 설명하고 비교하고자 한다.

각 개체에 대한 생존 시간 데이터  $\{T_1, \dots, T_n\}$ 은 서로 독립이며 모르는 변화점  $\tau$  이전에는 위험률  $\lambda_0$ , 이후에는 위험률  $\lambda_1$ 을 갖는다.  $\{(\lambda_0, \lambda_1, \tau) : \lambda_0 \geq 0, \lambda_1 \geq 0, \tau > 0\}$ 은 모수공간이다. 중도절단된 대상자에 대한 관측중단시점이  $\{C_1, \dots, C_n\}$  일 때, 우중도절단 여부는  $\{\delta_1, \dots, \delta_n\}$ 라고 표기한다.  $\tilde{T}_i$ 는 사건발생까지의 시간이라 할 때 생존 데이터는  $\{(T_i, \delta_i), i = 1, \dots, n\}$ 로 표현되며 여기서  $T_i = \min(\tilde{T}_i, C_i)$ ,  $\delta_i = I(\tilde{T}_i < C_i)$ 이고 우중도절단의 경우  $\delta_i = 0$ 이다.  $\{T_1, \dots, T_n\}$ 에 대한 순서통계량은  $\{t_1, \dots, t_n\}$ 으로 표기한다. 즉,  $t_1 < \dots < t_n$ 이다.

이와 같은 생존데이터에 대한 위험함수는

$$h(t) = \begin{cases} \lambda_0, & 0 \leq t \leq \tau \\ \lambda_1, & t > \tau \end{cases} \quad (2.1)$$

으로 위험률 변화점  $\tau$ 를 갖는다. 지수분포를 따른다고 가정할 경우 확률밀도함수는 다음과 같다.

$$f(t) = \begin{cases} \lambda_0 \exp(-\lambda_0 t), & t \leq \tau \\ \lambda_1 \exp(-\lambda_0 \tau - \lambda_1(t - \tau)), & t > \tau. \end{cases} \quad (2.2)$$

분포함수는

$$F(t) = \begin{cases} 1 - e^{-\lambda_0 t}, & t \leq \tau \\ 1 - e^{-\lambda_1 t + (\lambda_1 - \lambda_0)\tau}, & t > \tau \end{cases}$$

으로 표현된다.

### 2.1. 우도함수 기반 위험률 변화점 추정량

우도함수 기반으로 한 개 위험률 변화점 추정법을 제안한 기존 연구를 살펴보고자 한다. Henderson (1990)은 위험률 변화 검정을 위해 두 번 로그 취한 우도를 활용하여 우도 기반 위험률 변화점 추정량을 제안하였다. Loader (1991)는 위험률 변화점에 대한 우도비 과정에 기반하여 변화 여부에 대한 검정 통계량을 제안하였다. 임의 중도절단 (random censoring)이 있는 경우에도 확장하여 사용할 수 있도록 유의수준 조절에 대해 언급하였다.

Matthews와 Farewell (1982)은 로그 우도함수 (log likelihood) 기반 변화 검정법을 제안하였다. 고정된 변화점인  $\tau$ 에 대해  $N(\tau)$ 는  $\tau$  시점까지의 관측값 개수로  $N(\tau) = r$ 이다. 우도함수는

$$L_1(\tau) = r \log \lambda_0 + (n - r) \log \lambda_1 - \lambda_0 \sum_{i=1}^r t_i - \lambda_1 \sum_{i=r+1}^n t_i + (n - r)\tau(\lambda_1 - \lambda_0)$$

이다. 고정된  $\tau$ 에 대해  $\lambda_0$ 와  $\lambda_1$ 의 최대우도추정량 (maximum likelihood estimator; MLE)은

$$\hat{\lambda}_0 = \frac{r}{\sum_{i=1}^r t_i + (n - r)\tau}, \quad \hat{\lambda}_1 = \frac{n - r}{\sum_{i=r+1}^n t_i - (n - r)\tau} \quad (2.3)$$

이고, 우도기반 검정통계량은

$$L_{MF} = 2 \times \max_{\tau_l \leq \tau \leq \tau_u} [L_1(\tau) - L_1(0)] \quad (2.4)$$

이다. Matthews와 Farewell (1982)의 우도기반 변화점 추정량은

$$\hat{\tau}_{MF} = 2 \times \operatorname{argmax}_{\tau_1 \leq \tau \leq \tau_u} [L_1(\tau) - L_1(0)] \tag{2.5}$$

이다. 여기서  $\tau_1$ 과  $\tau_u$ 는 각각 변화점의 가능한 구간의 하한과 상한이다.

중도절단이 있는 경우 Matthews와 Farewell (1982) 위험률 변화점 추정량을 얻기위한 우도함수를 계산하는 방법을 정리하여 모수적 추정법에 대해 알아본다.

집합  $J_1 = \{i : \delta_i = 1\}$  사건에 대한 관측값을 얻은 개체 집합,  $J_0 = \{i : \delta_i = 0\}$ 는 중도절단인 경우를 나타낸다. 매 데이터 시점  $\tau$ 에서 다음과 같이 세분화된 집합을 구해 부분집합에 따라 기여하는 바에 따른 우도함수를 계산한다. 시점  $\tau$ 에서 해당 개체수를 구해보면

$$\begin{aligned} n_1 &= \#[(t_i \leq \tau), \delta_i = 1] \text{은 시점 } \tau \text{ 이하 값을 갖는 데이터 개수,} \\ n_2 &= \#[(t_i > \tau), \delta_i = 1] \text{은 시점 } \tau \text{ 보다 큰 값을 갖는 데이터 개수,} \\ c_1 &= \#[(t_i \leq \tau), \delta_i = 0] \text{은 시점 } \tau \text{ 이하 값을 갖는 중도절단 데이터 개수,} \\ c_2 &= \#[(t_i > \tau), \delta_i = 0] \text{은 시점 } \tau \text{ 보다 큰 값을 갖는 중도절단 데이터 개수} \end{aligned}$$

로 표기한다.  $n_1 + n_2 + c_1 + c_2 = n$ 이다.

중도절단이 있는 경우, 우도함수를 다음과 같이 계산하여 Matthews와 Farewell (1982)의 우도함수 기반 위험률 변화점을 추정할 수 있다.

$$\begin{aligned} L(\tau, \lambda_0, \lambda_1) &= \prod_{i \in J_1} f(T_i) \times \prod_{i \in J_0} \{1 - F(T_i)\} \\ &= \prod_{t_i \leq \tau: \text{obs}} \lambda_0 \exp(-\lambda_0 t_i) \prod_{t_i \leq \tau: \text{cen}} \exp(-\lambda_0 t_i) \prod_{t_i > \tau: \text{obs}} \lambda_1 \exp\{-\lambda_1 t_i + (\lambda_1 - \lambda_0)\tau\} \\ &\quad \times \prod_{t_i > \tau: \text{cen}} \exp\{-\lambda_1 t_i + (\lambda_1 - \lambda_0)\tau\} \\ &= \lambda_0^{n_1} \exp\left(-\lambda_0 \sum_{i=1}^{n_1} t_i\right) \exp\left(-\lambda_0 \sum_{i=1}^{c_1} t_i\right) \lambda_1^{n_2} \exp\left\{-\lambda_1 \sum_{i=1}^{n_2} t_i + n_2\tau(\lambda_1 - \lambda_0)\right\} \\ &\quad \times \exp\left\{-\lambda_1 \sum_{i=1}^{c_2} t_i + c_2\tau(\lambda_1 - \lambda_0)\right\}. \end{aligned}$$

로그 우도함수를 구하면

$$\begin{aligned} \log L &= n_1 \log \lambda_0 - \lambda_0 \sum_{i=1}^{n_1} t_i - \lambda_0 \sum_{i=1}^{c_1} t_i + n_2 \log \lambda_1 - \lambda_1 \sum_{i=1}^{n_2} t_i + n_2\tau(\lambda_1 - \lambda_0) \\ &\quad - \lambda_1 \sum_{i=1}^{c_2} t_i + c_2\tau(\lambda_1 - \lambda_0) \\ &= n_1 \log \lambda_0 + n_2 \log \lambda_1 - \lambda_0 \left(\sum_{i=1}^{n_1} t_i + \sum_{i=1}^{c_1} t_i\right) - \lambda_1 \left(\sum_{i=1}^{n_2} t_i + \sum_{i=1}^{c_2} t_i\right) \\ &\quad + (n_2 + c_2)\tau(\lambda_1 - \lambda_0) \\ &= n_1 \log \lambda_0 + n_2 \log \lambda_1 - \lambda_0 \left\{\sum_{i=1}^{n_1} t_i + \sum_{i=1}^{c_1} t_i + (n_2 + c_2)\tau\right\} - \lambda_1 \left\{\sum_{i=1}^{n_2} t_i + \sum_{i=1}^{c_2} t_i - (n_2 + c_2)\tau\right\} \end{aligned}$$

이다. 위험률 변화점 추정량은

$$\hat{\tau} = \operatorname{argmax}_{t_l < \tau < t_u} L(\tau)$$

이다. 중도절단이 있으며 변화점  $\tau$ 를 갖는 경우  $\lambda_0$ 와  $\lambda_1$ 의 최대우도추정량을 구하면

$$\hat{\lambda}_0 = \frac{n_1}{\sum_{i=1}^{n_1} t_i + \sum_{i=1}^{c_1} t_i + (n_2 + c_2)\tau}, \quad \hat{\lambda}_1 = \frac{n_2}{\sum_{i=1}^{n_2} t_i + \sum_{i=1}^{c_2} t_i - (n_2 + c_2)\tau}$$

이다.

## 2.2. 비모수적 위험률 변화점 통계량 비교

비모수적 방법으로는 Zhang 등 (2014) 추정량을 구하기 위해 누적위험함수의 기울기를 이용함에 대해 설명한다. Change 등 (1994)는 우중도절단 (right censoring)이 있는 경우에 대한 변화점 추정을 위해 최대우도와 마팅계일 (martingale)에 기반하여 다중 변화점에 대한 검정통계량을 제안하였다. Zhang 등 (2014) 추정량은 Chang 등 (1994) 방법을 기반으로 다중변화점 추정까지 확장한 것이다. 특히 장기간 생존 데이터가 있는 경우 우도함수만 고려하는 것이 충분하다고 할 수 없으므로 우도함수와 마팅계일을 결합한 방법을 고려하였다. Zhao (2009) 또한 장기간 생존데이터가 있는 경우 변화점 추정법을 다루었다. 다음과 같이 조각 선형 (piecewise constant) 함수를 가지며 2개 위험률 변화점을 가진 모형을 고려한다:

$$\lambda(t) = \begin{cases} a_1, & 0 \leq t \leq \tau_1 \\ a_2, & \tau_1 \leq t < \tau_2 \\ a_3, & t \geq \tau_2 \end{cases}$$

여기서  $a_1, a_2, a_3$ 는 위험률,  $\tau_1$ 과  $\tau_2$ 는 변화점이다.

지수분포를 따른다고 가정할 경우, 확률밀도함수는

$$f(t) = a_1 e^{-b_1(t)} I(0 \leq t < \tau_1) + a_2 e^{-b_2(t)} I(\tau_1 \leq t < \tau_2) + a_3 e^{-b_3(t)} I(t \geq \tau_2)$$

이다. 여기서  $b_1(t) = a_1(t)$ ,  $b_2(t) = a_1 \tau_1 + a_2(t - \tau_1)$ ,  $b_3(t) = a_1 \tau_1 + a_2(\tau_2 - \tau_1) + a_3(t - \tau_2)$ .

생존함수는

$$S(t) = e^{-b_1(t)} I(0 \leq t < \tau_1) + e^{-b_2(t)} I(\tau_1 \leq t < \tau_2) + e^{-b_3(t)} I(t \geq \tau_2)$$

이다.

이러한 위험률함수를 갖는 경우 일반적으로 두 가지 방법으로 접근한다. 첫 번째, 우도함수를 이용한 연구로는 Matthews와 Farewell (1982), Nguyen 등 (1984), Loader (1991), Henderson (1990)과 Williams와 Kim (2011) 등이 있다.

두 번째로는 비모수적 접근을 들 수 있으며 Chang 등 (1994)과 Zao 등 (2009)이 있다.

Chang 등 (1994) 통계량은 개수과정 (counting process) 기반으로  $0 < t < t_n$

$$Y(t) = \left\{ \frac{\Lambda(t_n) - \Lambda(t)}{t_n - t} - \frac{\Lambda(t) - \Lambda(0)}{t} \right\} g(t(t_n - t)) \quad (2.6)$$

제안하였다. Zhang 등 (2014)의 검정통계량은 Chang et al. (1994) 통계량의 다중변화점 문제로 확장하여

$$X(t) = \left\{ \frac{\Lambda(q_2) - \Lambda(t)}{q_2 - t} - \frac{\Lambda(t) - \Lambda(q_1)}{t - q_1} \right\} g((t - q_1)(q_2 - t)) \quad (2.7)$$

를 제안하였다. 여기서  $\Lambda_n$ 은 누적위험함수  $\Lambda$ 의 Nelson-Aalen 추정량이고  $g(x) = x^\alpha$  ( $0 \leq \alpha \leq 1$ ),  $t$ 의 범위는  $q_1 < t < q_2$ 이다. 변화점 추정량은

$$\hat{\tau}_Z = \inf \left\{ \operatorname{argmax}_{u \in (q_1, q_2)} |X_n(u)| \right\} \quad (2.8)$$

이다. 첫 번째 변화점은 가장 큰  $|X(t)|$  값을 갖는 시점이다. 두 번째 변화점 추정량은  $(q_1, \hat{\tau}_1)$ 과  $(\hat{\tau}_1, q_2)$  구간에서 각각 구한다. 점프 크기를 결정하기 위해서는 알고리즘적인 계산이 필요하다. 본 연구에서는 한 개 변화점 경우만 고려하기로 한다.

### 3. 모의실험

변화점 추정량에 대한 비교를 위해 모의실험을 수행하였다. 확률표본  $\{T_1, \dots, T_n\}$ 은 다음의 위험률 함수

$$h(t) = \begin{cases} \lambda_0, & 0 \leq t \leq \tau \\ \lambda_1, & t > \tau \end{cases}$$

를 갖는다. 지수분포를 따른다고 가정할 경우 다음의 확률밀도함수

$$f(t) = \begin{cases} \lambda_0 \exp(-\lambda_0 t), & t \leq \tau \\ \lambda_1 \exp(-\lambda_0 \tau - \lambda_1(t - \tau)), & t > \tau \end{cases}$$

를 따르도록 랜덤 데이터가 발생되었다. 분포함수가 (0,1)에서 균일분포를 따르므로 균일분포로부터 난수를 발생하여 해당 분포함수를 갖는 데이터를 구하였다. 표본의 크기 (sample size)는 100개로 정하고 각 실험에 대한 반복은 1,000번으로 하였다. 생존 데이터의 특성을 반영하여 중도절단이 없는 경우와 중도절단이 있는 경우를 고려하였다. 중도절단이 없는 경우에는 변화점 이후로 위험률이 감소하는 경우 ( $\lambda_0 = 2, \lambda_1 = 1$ )와 증가하는 경우 ( $\lambda_0 = 1, \lambda_1 = 2$ )에 대해 변화점이 0.25, 0.5, 0.7, 1.0일 때를 고려하였다. 실제 데이터 상황은 중도절단을 포함하며 위험률이 증가하는 경향이 있는 경우가 많이 발생한다. 이러한 면을 반영하여 중도절단이 있는 경우로는 랜덤 우중도절단으로 중도절단 비율이 각각 20%, 30%인 경우 변화점 이후로 위험률 증가하며 위험률 변화량이 다른 두 가지 상황을 고려하였다. 변화점이 각각 0.25, 0.5, 0.7, 1.0일 때  $\lambda_0 = 1, \lambda_1 = 1.5$ 와  $\lambda_0 = 1, \lambda_1 = 2$ 인 경우 위험률 변화점을 추정하였다.

각 모의실험에서는 위험률 변화점, 변화점 이전과 이후의 위험률을 추정하고 이러한 모수 추정량에 대한 평균제곱오차 (mean squared error; MSE)를 구하였다.

중도절단이 없는 경우 Table 3.1은  $n = 100$ 일 때 위험률 변화점 추정 결과이다. 위험률이 감소하는 경우보다 증가하는 경우 MSE 값이 작은 편이다. 위험률 변화가 데이터 발생과정에서 데이터 값 자체에 영향을 미치므로 결과에 영향을 미친듯하다. Zhang 등 (2014) 방법에서 함수의 승수로서  $\alpha = 0, 1/2, 1$ 을 고려한다.

**Table 3.1** Comparison of hazard change-point estimators without censoring (Repetition=1,000,  $n=100$ )

| $\tau$ | Stat              | $\lambda_0 = 1, \lambda_1 = 2$ |                            |                            | $\lambda_0 = 2, \lambda_1 = 1$ |                            |                            |
|--------|-------------------|--------------------------------|----------------------------|----------------------------|--------------------------------|----------------------------|----------------------------|
|        |                   | $\hat{\tau}$<br>(MSE)          | $\hat{\lambda}_0$<br>(MSE) | $\hat{\lambda}_1$<br>(MSE) | $\hat{\tau}$<br>(MSE)          | $\hat{\lambda}_0$<br>(MSE) | $\hat{\lambda}_1$<br>(MSE) |
| 0.25   | $\hat{\tau}_{MF}$ | 0.1915<br>(0.0115)             | 1.1496<br>(0.1268)         | 1.8107<br>(0.0690)         | 0.2660<br>(0.0039)             | 2.1002<br>(0.1338)         | 0.9864<br>(0.0167)         |
|        | $\hat{\tau}_Z$    | 0.1661<br>(0.0103)             | 0.9069<br>(0.0632)         | 1.8985<br>(0.0748)         | 0.1964<br>(0.0132)             | 1.7825<br>(0.1500)         | 1.1375<br>(0.0488)         |
|        | $\alpha = 0$      | 0.2881<br>(0.0044)             | 1.0234<br>(0.0632)         | 2.0785<br>(0.0748)         | 0.1574<br>(0.0151)             | 1.8603<br>(0.1500)         | 1.1597<br>(0.0488)         |
|        | $\hat{\tau}_Z$    | 0.3289<br>(0.0085)             | 1.1225<br>(0.0557)         | 2.0860<br>(0.0748)         | 0.1426<br>(0.0164)             | 1.9257<br>(0.1942)         | 1.1633<br>(0.0488)         |
|        | $\alpha = 1$      | 0.5139<br>(0.0070)             | 0.9898<br>(0.0259)         | 2.1101<br>(0.0990)         | 0.5063<br>(0.0068)             | 2.0631<br>(0.0686)         | 0.9777<br>(0.0320)         |
|        | $\hat{\tau}_Z$    | 0.4631<br>(0.0096)             | 0.9674<br>(0.0283)         | 2.0297<br>(0.0112)         | 0.4707<br>(0.0182)             | 1.8794<br>(0.0767)         | 1.1620<br>(0.0849)         |
| 0.5    | $\hat{\tau}_{MF}$ | 0.5396<br>(0.0074)             | 1.0163<br>(0.0283)         | 2.1181<br>(0.0112)         | 0.4110<br>(0.0229)             | 1.9091<br>(0.0767)         | 1.1970<br>(0.0849)         |
|        | $\hat{\tau}_Z$    | 0.5773<br>(0.0107)             | 1.0568<br>(0.0267)         | 2.1361<br>(0.1112)         | 0.3752<br>(0.0265)             | 1.9442<br>(0.0818)         | 1.2068<br>(0.0849)         |
|        | $\alpha = 0$      | 0.7206<br>(0.0080)             | 0.9968<br>(0.0202)         | 2.1161<br>(0.1287)         | 0.6866<br>(0.0065)             | 2.0559<br>(0.0578)         | 0.9880<br>(0.0404)         |
|        | $\hat{\tau}_Z$    | 0.6959<br>(0.0079)             | 0.9913<br>(0.0209)         | 2.0803<br>(0.1392)         | 0.6538<br>(0.0159)             | 1.9458<br>(0.0599)         | 1.1870<br>(0.1043)         |
|        | $\alpha = 0.5$    | 0.7376<br>(0.0071)             | 1.0104<br>(0.0209)         | 2.1253<br>(0.1392)         | 0.6220<br>(0.0188)             | 1.9548<br>(0.0599)         | 1.2049<br>(0.1043)         |
|        | $\hat{\tau}_Z$    | 0.7645<br>(0.0089)             | 1.0274<br>(0.0193)         | 2.1414<br>(0.1392)         | 0.5928<br>(0.0217)             | 1.9693<br>(0.0616)         | 1.2141<br>(0.1043)         |
| 0.7    | $\hat{\tau}_{MF}$ | 1.0029<br>(0.0105)             | 0.9898<br>(0.0160)         | 2.1170<br>(0.1679)         | 0.9538<br>(0.0090)             | 2.0541<br>(0.0477)         | 1.0177<br>(0.0828)         |
|        | $\hat{\tau}_Z$    | 0.9953<br>(0.0078)             | 0.9927<br>(0.0157)         | 2.0931<br>(0.1494)         | 0.8633<br>(0.0254)             | 1.9946<br>(0.0482)         | 1.3009<br>(0.2560)         |
|        | $\alpha = 0$      | 1.0127<br>(0.0054)             | 0.9976<br>(0.0157)         | 2.1086<br>(0.1494)         | 0.8551<br>(0.0275)             | 1.9961<br>(0.0482)         | 1.3051<br>(0.2560)         |
|        | $\hat{\tau}_Z$    | 1.0244<br>(0.0045)             | 1.0016<br>(0.0145)         | 2.1165<br>(0.1494)         | 0.8474<br>(0.0295)             | 1.9984<br>(0.0486)         | 1.3079<br>(0.2560)         |
|        | $\alpha = 0.5$    | 1.0244<br>(0.0045)             | 1.0016<br>(0.0145)         | 2.1165<br>(0.1494)         | 0.8474<br>(0.0295)             | 1.9984<br>(0.0486)         | 1.3079<br>(0.2560)         |
|        | $\hat{\tau}_Z$    | 1.0244<br>(0.0045)             | 1.0016<br>(0.0145)         | 2.1165<br>(0.1494)         | 0.8474<br>(0.0295)             | 1.9984<br>(0.0486)         | 1.3079<br>(0.2560)         |

중도절단이 있는 경우 중도절단 비율이 각각 20%, 30%일 때 위험률 증가량을 다르게 하여 모의실험한 결과 Table 3.2는 위험률이  $\lambda_0 = 1, \lambda_1 = 1.5$ 로 증가하는 경우이고 Table 3.3은 위험률이  $\lambda_0 = 1, \lambda_1 = 2$ 로 좀 더 많이 증가하는 경우의 결과이다. 중도절단이 없는 경우에 비해 중도절단이 있는 경우 변화점 추정량의 MSE가 더 큰 편이다. 중도절단이 있는 경우 변화점은 관측값에서만 고려했다. 중도절단된 경우는 실제 사건이 그 시점에서 발생한 것이 아니므로 가능한 변화점 대상에서 제외하고자한다. 이러한 상황이 중도절단이 있는 경우 변화점 추정량의 오차를 크게 할 수 있다.

**Table 3.2** Comparison of hazard change-point estimators with censoring  
(Repetition=1,000,  $n=100, \lambda_0=1.0, \lambda_1=1.5$ )

| $\tau$ | Stat              | $\lambda_0 = 1, \lambda_1 = 1.5$ (censoring 20%) |                            |                            | $\lambda_0 = 1, \lambda_1 = 1.5$ (censoring 30%) |                            |                            |
|--------|-------------------|--|----------------------------|----------------------------|--|----------------------------|----------------------------|
|        |                   | $\hat{\tau}$<br>(MSE)                            | $\hat{\lambda}_0$<br>(MSE) | $\hat{\lambda}_1$<br>(MSE) | $\hat{\tau}$<br>(MSE)                            | $\hat{\lambda}_0$<br>(MSE) | $\hat{\lambda}_1$<br>(MSE) |
| 0.25   | $\hat{\tau}_{MF}$ | 0.2545<br>(0.0067)                               | 1.0548<br>(0.0910)         | 1.5096<br>(0.0598)         | 0.2366<br>(0.0087)                               | 1.1648<br>(0.1550)         | 1.2741<br>(0.3224)         |
|        | $\hat{\tau}_Z$    | 0.1691<br>(0.0113)                               | 0.9909<br>(0.0537)         | 1.4557<br>(0.0524)         | 0.2067<br>(0.0089)                               | 1.0513<br>(0.0690)         | 1.3898<br>(0.1697)         |
|        | $\alpha = 0$      | 0.2754<br>(0.0060)                               | 1.0361<br>(0.0537)         | 1.5360<br>(0.0524)         | 0.2422<br>(0.0059)                               | 1.0762<br>(0.0690)         | 1.4283<br>(0.1697)         |
|        | $\alpha = 0.5$    | 0.3126<br>(0.0074)                               | 1.0885<br>(0.0429)         | 1.5347<br>(0.0524)         | 0.2581<br>(0.0049)                               | 1.0950<br>(0.0626)         | 1.4261<br>(0.1697)         |
|        | $\hat{\tau}_{MF}$ | 0.5096<br>(0.0125)                               | 1.0411<br>(0.0404)         | 1.4942<br>(0.0865)         | 0.5099<br>(0.0130)                               | 1.0607<br>(0.0450)         | 1.4457<br>(0.1838)         |
|        | $\alpha = 0$      | 0.4633<br>(0.0138)                               | 1.0125<br>(0.0298)         | 1.4891<br>(0.0741)         | 0.4866<br>(0.0118)                               | 1.0307<br>(0.0317)         | 1.4861<br>(0.1269)         |
| 0.5    | $\hat{\tau}_Z$    | 0.5217<br>(0.0106)                               | 1.0333<br>(0.0298)         | 1.5194<br>(0.0741)         | 0.5103<br>(0.0097)                               | 1.0375<br>(0.0317)         | 1.4998<br>(0.1269)         |
|        | $\alpha = 0.5$    | 0.5499<br>(0.0113)                               | 1.0503<br>(0.0279)         | 1.5232<br>(0.0741)         | 0.5242<br>(0.0089)                               | 1.0438<br>(0.0308)         | 1.5023<br>(0.1269)         |
|        | $\hat{\tau}_{MF}$ | 0.7108<br>(0.0122)                               | 1.0354<br>(0.0291)         | 1.4788<br>(0.1309)         | 0.7063<br>(0.0123)                               | 1.0481<br>(0.0328)         | 1.4112<br>(0.2037)         |
|        | $\hat{\tau}_Z$    | 0.6974<br>(0.0106)                               | 1.0228<br>(0.0233)         | 1.4981<br>(0.1012)         | 0.6990<br>(0.0097)                               | 1.0320<br>(0.0251)         | 1.4608<br>(0.1321)         |
|        | $\alpha = 0$      | 0.7159<br>(0.0094)                               | 1.0267<br>(0.0233)         | 1.5052<br>(0.1012)         | 0.7034<br>(0.0091)                               | 1.0326<br>(0.0251)         | 1.4646<br>(0.1321)         |
|        | $\alpha = 0.5$    | 0.7272<br>(0.0091)                               | 1.0305<br>(0.0225)         | 1.5070<br>(0.1012)         | 0.7063<br>(0.0084)                               | 1.0334<br>(0.0246)         | 1.4661<br>(0.1321)         |
| 0.7    | $\hat{\tau}_Z$    | 0.9779<br>(0.0122)                               | 1.0177<br>(0.0195)         | 1.5245<br>(0.1625)         | 0.9785<br>(0.0119)                               | 1.0246<br>(0.0223)         | 1.4973<br>(0.1895)         |
|        | $\alpha = 0$      | 0.9690<br>(0.0110)                               | 1.0135<br>(0.0176)         | 1.5306<br>(0.1415)         | 0.9621<br>(0.0113)                               | 1.0186<br>(0.0194)         | 1.5084<br>(0.1490)         |
|        | $\hat{\tau}_{MF}$ | 0.9766<br>(0.0096)                               | 1.0144<br>(0.0176)         | 1.5327<br>(0.1415)         | 0.9632<br>(0.0106)                               | 1.0185<br>(0.0194)         | 1.5104<br>(0.1490)         |
|        | $\hat{\tau}_Z$    | 0.9820<br>(0.0084)                               | 1.0152<br>(0.0172)         | 1.5333<br>(0.1415)         | 0.9648<br>(0.0100)                               | 1.0186<br>(0.0190)         | 1.5097<br>(0.1490)         |
|        | $\alpha = 0.5$    | 0.9766<br>(0.0096)                               | 1.0144<br>(0.0176)         | 1.5327<br>(0.1415)         | 0.9632<br>(0.0106)                               | 1.0185<br>(0.0194)         | 1.5104<br>(0.1490)         |
|        | $\alpha = 1$      | 0.9820<br>(0.0084)                               | 1.0152<br>(0.0172)         | 1.5333<br>(0.1415)         | 0.9648<br>(0.0100)                               | 1.0186<br>(0.0190)         | 1.5097<br>(0.1490)         |

**Table 3.3** Comparison of hazard change-point estimators with censoring  
(Repetition=1,000,  $n=100, \lambda_0=1.0, \lambda_1=2.0$ )

| $\tau$ | Stat              | $\lambda_0 = 1, \lambda_1 = 2$ (censoring 20%) |                            |                            | $\lambda_0 = 1, \lambda_1 = 2$ (censoring 30%) |                            |                            |
|--------|-------------------|--|----------------------------|----------------------------|--|----------------------------|----------------------------|
|        |                   | $\hat{\tau}$<br>(MSE)                          | $\hat{\lambda}_0$<br>(MSE) | $\hat{\lambda}_1$<br>(MSE) | $\hat{\tau}$<br>(MSE)                          | $\hat{\lambda}_0$<br>(MSE) | $\hat{\lambda}_1$<br>(MSE) |
| 0.25   | $\hat{\tau}_{MF}$ | 0.2553<br>(0.0045)                             | 1.0759<br>(0.0724)         | 1.9998<br>(0.0877)         | 0.2432<br>(0.0085)                             | 1.2552<br>(0.1941)         | 1.6498<br>(0.4990)         |
|        | $\hat{\tau}_Z$    | 0.1716<br>(0.0102)                             | 1.0248<br>(0.0598)         | 1.8619<br>(0.0903)         | 0.2183<br>(0.0082)                             | 1.1585<br>(0.0934)         | 1.7078<br>(0.2929)         |
|        | $\alpha = 0$      | 0.2795<br>(0.0047)                             | 1.1156<br>(0.0598)         | 2.0239<br>(0.0903)         | 0.2564<br>(0.0052)                             | 1.1873<br>(0.0934)         | 1.7653<br>(0.2929)         |
|        | $\hat{\tau}_Z$    | 0.3180<br>(0.0073)                             | 1.1956<br>(0.0782)         | 2.0325<br>(0.0903)         | 0.2682<br>(0.0045)                             | 1.2036<br>(0.0953)         | 1.7712<br>(0.2929)         |
|        | $\alpha = 0.5$    | 0.3180<br>(0.0073)                             | 1.1956<br>(0.0782)         | 2.0325<br>(0.0903)         | 0.2682<br>(0.0045)                             | 1.2036<br>(0.0953)         | 1.7712<br>(0.2929)         |
|        | $\alpha = 1$      | 0.3180<br>(0.0073)                             | 1.1956<br>(0.0782)         | 2.0325<br>(0.0903)         | 0.2682<br>(0.0045)                             | 1.2036<br>(0.0953)         | 1.7712<br>(0.2929)         |
| 0.5    | $\hat{\tau}_{MF}$ | 0.5124<br>(0.0104)                             | 1.0631<br>(0.0361)         | 1.9788<br>(0.1380)         | 0.5049<br>(0.0146)                             | 1.1045<br>(0.0519)         | 1.8420<br>(0.3033)         |
|        | $\hat{\tau}_Z$    | 0.4945<br>(0.0116)                             | 1.0556<br>(0.0340)         | 1.9554<br>(0.1405)         | 0.4976<br>(0.0116)                             | 1.0972<br>(0.0427)         | 1.8501<br>(0.2317)         |
|        | $\alpha = 0$      | 0.5378<br>(0.0097)                             | 1.0806<br>(0.0340)         | 2.0043<br>(0.1405)         | 0.5169<br>(0.0095)                             | 1.1052<br>(0.0427)         | 1.8694<br>(0.2317)         |
|        | $\hat{\tau}_Z$    | 0.5643<br>(0.0110)                             | 1.0997<br>(0.0356)         | 2.0207<br>(0.1405)         | 0.5289<br>(0.0084)                             | 1.1125<br>(0.0424)         | 1.8734<br>(0.2317)         |
|        | $\alpha = 0.5$    | 0.5643<br>(0.0110)                             | 1.0997<br>(0.0356)         | 2.0207<br>(0.1405)         | 0.5289<br>(0.0084)                             | 1.1125<br>(0.0424)         | 1.8734<br>(0.2317)         |
|        | $\alpha = 1$      | 0.5643<br>(0.0110)                             | 1.0997<br>(0.0356)         | 2.0207<br>(0.1405)         | 0.5289<br>(0.0084)                             | 1.1125<br>(0.0424)         | 1.8734<br>(0.2317)         |
| 0.7    | $\hat{\tau}_{MF}$ | 0.7185<br>(0.0131)                             | 1.0595<br>(0.0282)         | 1.9458<br>(0.2181)         | 0.7108<br>(0.0150)                             | 1.0782<br>(0.0369)         | 1.8448<br>(0.4327)         |
|        | $\hat{\tau}_Z$    | 0.7370<br>(0.0125)                             | 1.0664<br>(0.0273)         | 1.9648<br>(0.1866)         | 0.7127<br>(0.0094)                             | 1.0774<br>(0.0321)         | 1.8411<br>(0.2977)         |
|        | $\alpha = 0$      | 0.7408<br>(0.0110)                             | 1.0663<br>(0.0273)         | 1.9694<br>(0.1866)         | 0.7133<br>(0.0081)                             | 1.0760<br>(0.0321)         | 1.8480<br>(0.2977)         |
|        | $\hat{\tau}_Z$    | 0.7416<br>(0.0099)                             | 1.0663<br>(0.0253)         | 1.9656<br>(0.1866)         | 0.7136<br>(0.0072)                             | 1.0756<br>(0.0307)         | 1.8454<br>(0.2977)         |
|        | $\alpha = 0.5$    | 0.7416<br>(0.0099)                             | 1.0663<br>(0.0253)         | 1.9656<br>(0.1866)         | 0.7136<br>(0.0072)                             | 1.0756<br>(0.0307)         | 1.8454<br>(0.2977)         |
|        | $\alpha = 1$      | 0.7416<br>(0.0099)                             | 1.0663<br>(0.0253)         | 1.9656<br>(0.1866)         | 0.7136<br>(0.0072)                             | 1.0756<br>(0.0307)         | 1.8454<br>(0.2977)         |
| 1.0    | $\hat{\tau}_{MF}$ | 0.9943<br>(0.0143)                             | 1.0280<br>(0.0199)         | 1.9818<br>(0.3105)         | 1.0042<br>(0.0142)                             | 1.0478<br>(0.0223)         | 1.9361<br>(0.2781)         |
|        | $\hat{\tau}_Z$    | 1.0169<br>(0.0104)                             | 1.0356<br>(0.0190)         | 1.9878<br>(0.2496)         | 1.0241<br>(0.0096)                             | 1.0545<br>(0.0217)         | 1.9395<br>(0.2336)         |
|        | $\alpha = 0$      | 1.0073<br>(0.0096)                             | 1.0317<br>(0.0189)         | 1.9828<br>(0.2496)         | 1.0140<br>(0.0090)                             | 1.0508<br>(0.0217)         | 1.9364<br>(0.2336)         |
|        | $\hat{\tau}_Z$    | 1.0002<br>(0.0086)                             | 1.0295<br>(0.0183)         | 1.9689<br>(0.2496)         | 0.9988<br>(0.0084)                             | 1.0470<br>(0.0207)         | 1.9132<br>(0.2337)         |
|        | $\alpha = 0.5$    | 1.0002<br>(0.0086)                             | 1.0295<br>(0.0183)         | 1.9689<br>(0.2496)         | 0.9988<br>(0.0084)                             | 1.0470<br>(0.0207)         | 1.9132<br>(0.2337)         |
|        | $\alpha = 1$      | 1.0002<br>(0.0086)                             | 1.0295<br>(0.0183)         | 1.9689<br>(0.2496)         | 0.9988<br>(0.0084)                             | 1.0470<br>(0.0207)         | 1.9132<br>(0.2337)         |

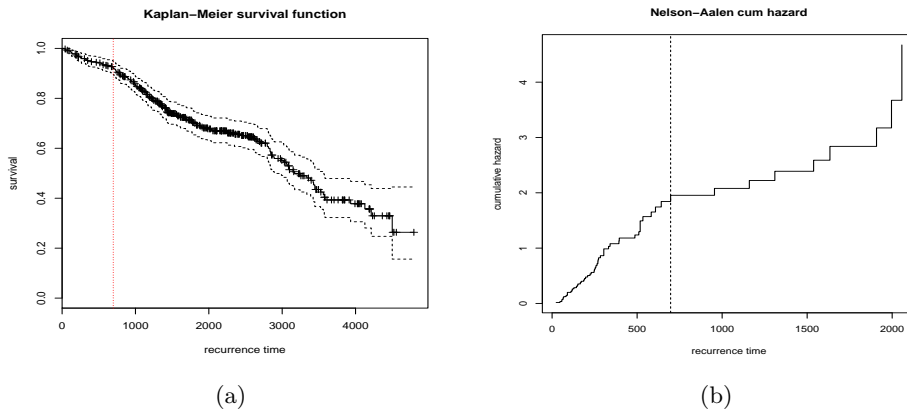
### 4. 실제 생존데이터에 대한 위험률 변화점 추정

실제로 두 개의 데이터셋에 대해 위험률 변화점을 추정해보고 추정법에 따른 결과를 비교하고자 한다.

첫 번째는 Matthews와 Farewell (1982)도 분석한 데이터로 84명의 백혈병 환자들의 재발까지 걸린 시간(일)이다. 51명의 재발시간은 관측되었고 33명은 중도절단 되었다. 이 데이터는 장기 생존값이 우 중도절단으로 여러 개 포함된 데이터이다. 이상 패턴을 보이는 중도절단 값 182인 경우로 24명은 제외 하고 전체 데이터 60명에 대해 분석하고자 한다. Zhang 등 (2014) 통계량 계산시 위험률 변화점 가능 범위로  $q_1 = 304, q_2 = 1,160$ 을 고려하였다. Matthews와 Farewell (1982) 방법과 Zhang 등 (2014) 방법에 의한 위험률 변화점 추정량이 같음을 Table 4.1에서 알 수 있다. Figure 4.1은 Kaplan-Meier 생존 함수와 누적위험함수 그라프와 함께 위험률 변화점 추정값을 보여준다.

**Table 4.1** Hazard change-point estimates for Leukemia recurrence data

| Stat                          | $\hat{\tau}$ | $\hat{\lambda}_0$ | $\hat{\lambda}_1$ |
|-------------------------------|--------------|-------------------|-------------------|
| $\hat{\tau}_{MF}$             | 697          | 0.002559          | 0.00028           |
| $\hat{\tau}_Z (\alpha = 1)$   | 697          | 0.002559          | 0.00028           |
| $\hat{\tau}_Z (\alpha = 1/2)$ | 697          | 0.002559          | 0.00028           |
| $\hat{\tau}_Z (\alpha = 0)$   | 697          | 0.002559          | 0.00028           |

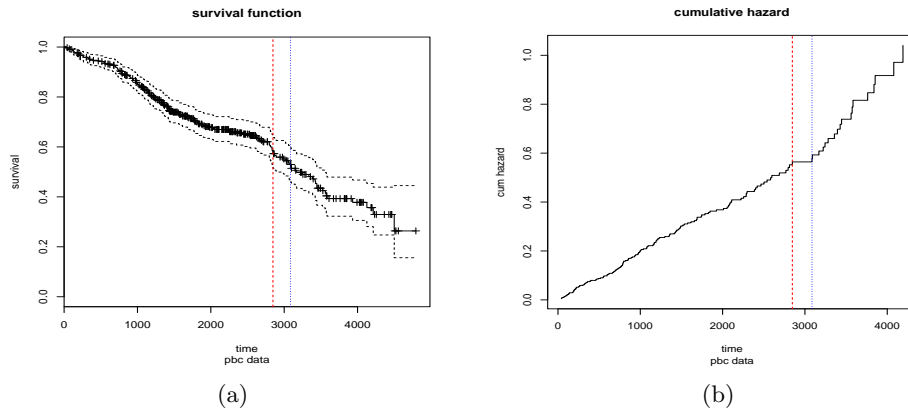


**Figure 4.1** (a) Survival function (b) Cumulative hazard with change-point for Leukemia recurrence data ( $\hat{\tau}_{MF} = \hat{\tau}_Z = 697$ )

두 번째 데이터는 1974년부터 1984년까지 미국 Mayo clinic에서 치료받은 원발성 담즙성 경화 418명 환자들의 생존데이터이다. 257명은 간이식 등으로 중도절단 되었다. Table 4.2에서 보면 Matthews와 Farewell (1982) 방법과 Zhang 등 (2014) 방법에서  $\alpha = 0$ 인 경우 추정량이 같고  $\alpha = 1/2$  과  $\alpha = 1$  경우의 변화점 추정량이 동일하다.  $\alpha = 0$ 인 경우는 시점의 위치에 대한 함수가 고려되지않은 추정량으로 이 데이터에서는  $\alpha$ 값에 따라 변화점 추정이 달라지는 것으로 보아 시점의 위치가 좀 더 영향을 미친다고 볼 수 있다. Figure 4.2은 Kaplan-Meier 생존함수와 누적위험함수 그라프와 함께 서로 다른 위험률 변화점 추정값을 보여준다.

**Table 4.2** Hazard change-point estimates for Primary biliary cirrhosis data

| Stat                          | $\hat{\tau}$ | $\hat{\lambda}_0$ | $\hat{\lambda}_1$ |
|-------------------------------|--------------|-------------------|-------------------|
| $\hat{\tau}_{MF}$             | 3086         | 0.000190          | 0.000365          |
| $\hat{\tau}_Z (\alpha = 0)$   | 3086         | 0.000190          | 0.000365          |
| $\hat{\tau}_Z (\alpha = 1/2)$ | 2847         | 0.000194          | 0.000276          |
| $\hat{\tau}_Z (\alpha = 1)$   | 2847         | 0.000194          | 0.000276          |

**Figure 4.2** (a) Survival function (b) Cumulative hazard with change-point for Primary biliary cirrhosis data  
( $\hat{\tau}_{MF} = 3086$ ,  $\hat{\tau}_{Z, \alpha=1} = 2847$ )

## 5. 결론

생존데이터에 대해 생존함수 및 위험함수 추정시 위험률 변화점이 존재한다면 변화점을 추정한 후 변화점 이전과 이후에 대해 모수를 추정해야한다. 본 연구에서는 지수분포를 따르는 경우의 위험률 변화점과 비모수적 변화점 추정량의 이론적 배경을 살펴보고 모의실험을 통해 위험률 추정량의 능력을 평균 제곱오차를 계산하여 비교하였다. 생존데이터에 대한 모수적 분포를 알 경우에는 우도함수 기반 추정량을 사용할 수 있으나 그렇지 못한 경우에는 비모수적 추정량이 대안이 될 수 있다. 본 연구에서 비교한 두 통계량은 중도절단을 포함한 데이터에 사용할 수 있으며 중도절단이 없는 경우에 비해서도 추정 능력에 큰 차이가 없는 편이다. 실제 상황에서 위험률 변화점이 여러 개 존재하는 경우가 발생하므로 다중변화점 모형으로의 확장이 필요하다. 또한 중도절단 비율이 큰 경우에 대한 변화점 연구도 필요해 보인다.

## References

- Chang, I. S., Chen, C. H. and Hsiung, C. A. (1994). Estimation in change-point hazard rate models with random censorship. *IMS Lecture Notes - Monograph Series*, **23**, 78-92.
- Dupuy, J. F. (2006). Estimation in a change-point hazard regression model. *Statistics & Probability Letters*, **76**, 182-190.
- Gijbels, I. and Gurler, U. (2003). Estimation of a change point in a hazard function based on censored data. *Lifetime Data Analysis*, **9**, 395-411.
- Goodman, M. S., Li, Y. and Tiwari, R. C. (2010). Detecting multiple change points in piecewise constant hazard functions. *Journal of Applied Statistics*, **38**, 11, 2523-2532.



- Henderson, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika*, **77**, 835-843.
- Jeong, K. M. (1998). Nonparametric estimation of hazard rates change-point. *The Korean Journal of Applied Statistics*, **11**, 163-175.
- Jeong, K. M. and Han, M. H. (1998). Estimation for change-point model of hazard rate. *Communications for Statistical Applications and Methods*, **5**, 477-489.
- Kim, J. (2009). A change-point estimator with the hazard ratio. *Journal of the Korean Statistical Society*, **38**, 377-382.
- Lee, S. and Shim, B. Y. and Kim, J. (2015) Estimation of hazard function and hazard change-point for the rectal cancer data. *Journal of the Korean Data & Information Science Society*, **26**, 1225-1238.
- Loader, C. R. (1991). Inference for a hazard rate change point. *Biometrika*, **78**, 749-757.
- Luo, X., Turnbull, B. W. and Clark, L. C. (1997). Likelihood ratio tests for a change-point with survival data. *Biometrika*, **84**, 555-565.
- Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-term Survivors*, Wiley, New York.
- Matthews, D. E. and Farewell, V. T. (1982). On testing for constant hazard against a change-point alternative. *Biometrics*, **38**, 463-468.
- Matthews, D. E., Farewell, V. T. and Pyke, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. *The Annals of Statistics*, **13**, 583-591.
- Nguyen, H. T., Rogers, G. S. and Walker, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika*, **71**, 299-304.
- Pham, T. D. and Nguyen, H. T. (1990). Strong consistency of the maximum likelihood estimators in the change-point hazard rate model. *Statistics: A Journal of Theoretical and Applied Statistics*, **21**, 203-216.
- Shin, S. B. and Kim, Y. J. (2014). Statistical analysis of recurrent gap time events with incomplete observation gaps. *Journal of the Korean Data & Information Science Society*, **25**, 327-336.
- Williams, M. R. and Kim, D. (2011). Likelihood ratio tests for continuous monotone hazards with an unknown change point. *Statistics & Probability Letters*, **81**, 1599-1603.
- Worsley, K. J. (1988). Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. *Biometrics*, **44**, 259-263.
- Yao, Y. C. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Communications in Statistics - Theory and Methods*, **15**, 2455-2466.
- Zhang, W., Qian, L. and Li, Y. (2014). Semiparametric sequential testing for multiple change points in piecewise constant hazard functions with long-term survivors. *Communications in Statistics - Simulation and Computation*, **43**, 1685-1699.
- Zhao, X. B., Wu, X. Y. and Zhou, X. (2009). A change-point model for survival data with long-term survivors. *Statistica Sinica*, **19**, 377-390.

# Comparison of parametric and nonparametric hazard change-point estimators<sup>†</sup>

Jaehee Kim<sup>1</sup> · Sieun Lee<sup>2</sup>

<sup>1</sup>Department of Information and Statistics, Duksung Women's University

<sup>2</sup>Office of Biostatistics, School of Medicine, Ajou University

Received 10 August 2016, revised 20 September 2016, accepted 24 September 2016

## Abstract

When there exists a change-point in hazard function, it should be estimated for exact parameter or hazard estimation. In this research, we compare the hazard change-point estimators. Matthews and Farewell (1982) parametric change-point estimator is based on the likelihood and Zhang *et al.* (2014) nonparametric estimator is based on the Nelson-Aalen cumulative hazard estimator. Simulation study is done for the data from exponential distribution with one hazard change-point. The simulated data generated without censoring and the data with right censoring are considered. As real data applications, the change-point estimates are computed for leukemia data and primary biliary cirrhosis data.

*Keywords:* Censored, exponential distribution, hazard rate change-point, survival analysis.

---

<sup>†</sup> This research was supported by 2015 Duksung Women's University Research Fund.

<sup>1</sup> Corresponding author: Professor, Department of Information and Statistics, Duksung Women's University, Seoul 132-714, Korea. E-mail: jaehee@duksung.ac.kr

<sup>2</sup> Master, Office of Biostatistics, School of Medicine, Ajou University, Suwon 16499, Korea.