The Korean Communications
in Statistics Vol. 5, No. 2, 1998
pp. 477-489

위험률의 변화점모형에 대한 추론[1)]

정 광 모[2)], 한 미 혜[3)]

요 약

위험률 변화점모형에 대해 변화점의 최우추정을 고려하였다. 추정량의 점근분포 및 붓스트랩 분포의 성질을 알아보고 변화점의 신뢰구간을 제안한다. 변화점의 위치 및 변화점을 전후하여 위험률의 값에 따라 모의실험을 수행하고 포함확률을 조사하였다. 추정량의 점근분포가 매우 복잡하기 때문에 이를 직접 이용한 변화점의 통계적 추론이 매우 어려운 점을 감안할 때 제안된 방법은 바람직한 대안이 될 수 있다.

1. 서 론

시간에 의존하는 데이타에서 분포함수의 모수가 어떤 시점을 기준으로 변화되는 경우에 미지의 변화점을 추정하거나 변화점 유무에 대한 가설검정을 수행한다. 이러한 통계적추론과정을 변화점문제(change-point problem)라 한다. 변화점문제는 연구대상 및 연구방법에 따라 매우 다각적으로 연구되어왔다. 평균이나 분산의 변화점모형, 회귀모형의 변화점문제, 환자나 제품 수명의 위험률(hazard rate)에 대한 변화점모형등이 있고 연구방법에 따라서는 모수 및 비모수적 방법과 베이지안 방법등이 있다. 본 연구에서는 제품이나 환자의 수명에서 자주 대두되는 위험률에 대한 변화점문제를 논의한다.

확률밀도함수가 $f(t)$이고 분포함수가 $F(t)$일 때, 위험률 $h(t)$는

$$h(t) = f(t)/\{1-F(t)\} \tag{1.1}$$

와 같이 정의된다. 또한, 누적위험함수(cumulative hazard function) 는

$$H(t) = \int_0^t h(u)du = -\log\{1-F(t)\}$$

이다. 누적위험함수 $H(t)$와 분포함수 $F(t)$간의 관계식

$$F(t) = 1 - e^{-H(t)}$$

이 성립한다. 위험률 변화점모형의 특별한 경우로서 어떤 상수 β_1, β_2에 대해 미지의 시점 τ ("변화점"이라 불림)를 기준으로

---

1) 이 연구는 1996년도 부산대학교 학술연구조성비(발전기금) 지원에 의해 부산대학교 기초과학연구소에서 수행되었음(RIBS-PNU-96108)
2)(609-735) 부산시 금정구 장전동 부산대학교 통계학과, 교수
3)(609-735) 부산시 금정구 장전동 부산대학교 통계학과, 시간강사

$$h(t)=\begin{cases} \dfrac{1}{\beta_1}, & t\leq\tau \\ \dfrac{1}{\beta_2}, & t>\tau \end{cases} \tag{1.2}$$

와 같이 표현되는 모형은 τ 이전과 이후의 분포함수가 지수분포와 대응된다.

변화점문제에 대한 일반적인 논의는 Page(1954), Chernoff와 Zacks(1964), Bhattacharyya와 Johnson(1968), Hinkley(1970), Pettitt(1979) 등을 참고할 수 있으며, 본 연구에서는 제품이나 환자의 수명에서 자주 대두되는 위험률에 대한 변화점 문제를 중심으로 그 배경을 살펴본다. 지수분포의 변화점모형 (1.2)에 대한 연구로서 Nguyen, Rogers와 Walker(1984)는 변화점의 최우추정을 제안하고 추정량의 일치성에 대해 논의하였다. 우도함수가 관찰값의 양 끝 부분에서 급격히 증가하는 비유계(unbounded)인 성질을 갖기 때문에 제한된(restricted) 구간에서 최우추정량을 제안하고 시뮬레이션을 통해 그 성질을 검토하였다. Yao(1986)도 변화점 및 위험률에 대한 최우추정을 제안하고 그 근사분포에 관해 논의하였다. Pham과 Nguyen(1990)은 최우추정량의 강일치성(strong consistency)을 보였고, Pham과 Nguyen(1991)은 변화점의 최우추정량에 대한 모수적 붓스트랩(parametric bootstrap) 분포가 극한분포에 수렴하는 것을 증명하였다.

또한, Matthews와 Farewell(1982)은 변화점의 가설에 대한 우도비검정과 이것의 카이제곱 근사분포를 논의하고, 몬테칼로 모의실험을 통해 근사분포의 정확성 및 중도절단 비율에 따른 민감성을 검토하였다. 그 후, Matthews, Farewell과 Pyke(1985)는 스코어 통계량(score statistic)을 써서 변화점에 대한 가설검정을 수행하였다. 한편, Loader(1991)는 경계선교차확률(boundary crossing probability)의 성질을 이용하여 LRT의 유의확률과 변화점의 신뢰영역을 구하고, 실제 임상자료에 적용하였으며 신뢰영역의 실제 포함확률(coverage probability)이 보수적(conservative)임을 발견하였다. 이와같이 지수분포에 대한 변화점모형은 많은 연구가 이루어져 왔으나 이를 일반화한 와이블분포의 변화점모형에 대해서는 연구가 거의 이루어지지 못한 실정이다.

반면에 비모수적 방법에 관한 연구로 Muller와 Wang(1990)은 넬슨추정량(Nelson estimator)을 써서 누적위험함수를 구하고 커넬스무딩(kernel smoothing)에 의한 위험률곡선의 변화가 가장 뚜렷하게 변하는 점을 변화점으로 추정하게 된다. 반면에 Antoniadis와 Gregoire(1993)는 변화점이 알려진 것으로 가정하고 커넬스무딩을 이용하여 변화의 크기(jump size) 및 위험률을 추정하였다. Chang, Chen과 Hsiung(1994)은 누적위험함수를 플롯(plot)하고 어떤 관찰값을 전후하여 기울기의 차가 최대일 때의 변화점추정을 제안하고 마팅게일(martingale)을 써서 추정량의 일치성을 보였다.

지금까지는 대부분 식 (1.2)의 변화점모형을 가정하여 연구되었는데 이를 좀 더 일반화한 모형을 다루어보자. 수명에 관련된 자료는 그 성격에 따라 여러 가지 형태의 위험률을 나타내게 되는데, 특히, 와이블분포(Weibull distribution)의 위험률은

$$h(t)=\left(\frac{\gamma}{\eta}\right)\left(\frac{t}{\eta}\right)^{\gamma-1}$$

와 같이 주어지며 $\beta=\eta^{\gamma}$ 라 놓으면

$$h(t)=\frac{\gamma}{\beta}\,t^{\gamma-1}$$

와 같이 나타낼 수 있다. 변화점 이전에는 모수가 (γ_1, β_1)이고 그 이후에는 (γ_2, β_2)인 와이블분포의 변화점모형

$$h(t)=\begin{cases} \dfrac{\gamma_1}{\beta_1} t^{\gamma_1-1}, & t\le\tau \\ \dfrac{\gamma_2}{\beta_2} t^{\gamma_2-1}, & t>\tau \end{cases} \tag{1.3}$$

에 대해 변화점 τ의 추정 및 점근적 성질에 관해 논의한다.

변화점모형 (1.3)은 고장시간 또는 수명 데이터를 모형화하는 데 매우 유용하다. 예를 들어, 제품이 갑작스런 충격이나 어떤 원인으로 초기에 고장을 일으키는 경우와 제품을 오랫동안 사용하다 보면 자연스런 마모등으로 인해 고장을 일으키는 경우로 구분할 수 있다. 초기 고장의 수명시간은 위험률이 감소형태를 나타내는 반면에 제품의 마모 등에 의한 고장시간의 위험률은 증가하는 양상을 나타내게 된다. 와이블분포에서 형태모수(shape parameter) 값이 1보다 작으면 위험률이 감소하고 1보다 크면 증가한다.

두 가지 또는 그 이상의 복합적인 원인에 의한 수명시간의 위험률도 모형 (1.3)과 같은 변화점모형으로 표현할 수 있다. 신뢰성분야에서는 이러한 모형을 복합분포모형(composite distribution model)이라 부른다. 또한, 변화점모형은 혼합모형(mixture model)의 분포를 설명하는 데도 널리 이용되기 때문에 매우 폭넓은 활용 분야를 갖는다. 지수분포의 위험률은 시간에 관해 일정한 반면 이를 일반화한 와이블분포의 변화점모형은 모형적합도(model goodness-of-fit)를 더욱 향상시킬 수 있는 효과가 있다.

2. 우도함수와 모수추정

수명시간을 T라 할 때 변화점모형 (1.3)의 가정하에서 확률밀도함수는

$$f(t)=\begin{cases} \dfrac{\gamma_1}{\beta_1} t^{\gamma_1-1}\exp\left(-\dfrac{t^{\gamma_1}}{\beta_1}\right), & t\le\tau \\ \dfrac{\gamma_2}{\beta_2} t^{\gamma_2-1}\exp\left(-\dfrac{t^{\gamma_2}}{\beta_2}-\left\{\dfrac{\tau^{\gamma_1}}{\beta_1}-\dfrac{\tau^{\gamma_2}}{\beta_2}\right\}\right), & t>\tau \end{cases} \tag{2.1}$$

와 같이 유도된다.

미지의 변화점을 τ라 가정할 때 확률표본 $T_1, T_2, \cdots, T_n$에 대한 로그우도함수(log-likelihood function)는

$$\begin{aligned} l(\gamma,\beta,\tau) &= X(\tau)\log\left(\frac{\gamma_1}{\beta_1}\right)+(n-X(\tau))\left[\log\left(\frac{\gamma_2}{\beta_2}\right)-\left\{\frac{\tau^{\gamma_1}}{\beta_1}-\frac{\tau^{\gamma_2}}{\beta_2}\right\}\right] \\ &+ \sum_{t_i\le\tau}\left[(\gamma_1-1)\log(t_i)-\frac{t_i^{\gamma_1}}{\beta_1}\right]+\sum_{t_i>\tau}\left[(\gamma_2-1)\log(t_i)-\frac{t_i^{\gamma_2}}{\beta_2}\right] \end{aligned} \tag{2.2}$$

이다. 단, $\gamma=(\gamma_1,\gamma_2)$, $\beta=(\beta_1,\beta_2)$이고 $X(\tau)$는 τ보다 작거나 같은 관찰값 t_i의 개수이다. 모수 γ, β 및 변화점 τ의 추정방법을 알아보자.

우선 변화점 τ를 주어진 값으로 가정하고 γ, β를 추정해보자. $l(\gamma, \beta, \tau)$를 β_i에 대해 각각 미분하여 0으로 놓고 풀면,

$$\beta_1 = \frac{\sum_{t_i \le \tau} t_i^{\gamma_1} + (n - X(\tau))\, \tau^{\gamma_1}}{X(\tau)}, \quad \beta_2 = \frac{\sum_{t_i > \tau} t_i^{\gamma_2} - (n - X(\tau))\, \tau^{\gamma_2}}{n - X(\tau)} \quad . \tag{2.3}$$

을 만족한다. 같은 요령으로 $l(\gamma, \beta, \tau)$를 γ_i에 대해 각각 미분하여 0으로 놓으면 다음 식

$$\frac{X(\tau)}{\gamma_1} - (n - X(\tau)) \frac{1}{\beta_1} \tau^{\gamma_1} \log(\tau) = \frac{1}{\beta_1} \sum_{t_i \le \tau} t_i^{\gamma_1} \log(t_i) - \sum_{t_i \le \tau} \log(t_i) \tag{2.4}$$

$$\frac{n - X(\tau)}{\gamma_2} + (n - X(\tau)) \frac{1}{\beta_2} \tau^{\gamma_2} \log(\tau) = \frac{1}{\beta_2} \sum_{t_i > \tau} t_i^{\gamma_2} \log(t_i) - \sum_{t_i > \tau} \log(t_i) \tag{2.5}$$

을 얻는다. 식 (2.3)을 (2.4), (2.5)에 대입하면 결국 τ 및 γ_i에 관한 식으로 나타낼 수 있다. 와이블분포의 누적위험함수 $H(t)$와 형태모수간의 관계식

$$\log H(t) = \gamma_i \log t - \log \beta_i, \quad i = 1, 2$$

에서 γ_i는 기울기를 나타내므로 $\log H(t)$와 $\log t$의 플롯에서 γ_i의 초기추정치를 구할 수 있다. 초기추정치에서 시작하여 식 (2.3)-(2.5)를 동시에 만족하도록 반복적인 수치계산에 의해 추정치 $\hat{\gamma}_i$, $\hat{\beta}_i$ 를 구한다. 구해진 $\hat{\gamma}_i$, $\hat{\beta}_i$를 (2.2)식의 로그우도함수에 대입하면

$$\begin{aligned} l(\hat{\gamma}, \hat{\beta}, \tau) \propto \quad & X(\tau) \log\left(\frac{\hat{\gamma}_1}{\hat{\beta}_1}\right) + (n - X(\tau)) \log\left(\frac{\hat{\gamma}_2}{\hat{\beta}_2}\right) \\ & + \sum_{t_i \le \tau} \hat{\gamma}_1 \log(t_i) + \sum_{t_i > \tau} \hat{\gamma}_2 \log(t_i) \end{aligned} \tag{2.6}$$

와 같다. 이 때, 식 (2.6)을 최대로 하는 τ를 추정하는 방법이 변화점에 대한 최우추정(maximum likelihood estimation:MLE)법이다. 즉, 변화점의 최우추정량 $\hat{\tau}$은

$$\hat{\tau} = \arg\sup_{\tau_0 \le \tau \le \tau_1} l(\hat{\gamma}, \hat{\beta}, \tau)$$

와 같이 정의된다. 여기서 $[\tau_0, \tau_1]$ 은 자료의 양 끝 부분에서 우도함수가 변칙적으로 커지는 것을 방지하기 위해 택해진 구간이다. 예를들면, $t_{(i)}$를 순서통계치라 할 때 $[\tau_0, \tau_1] = [t_{(1)}, t_{(n-1)}]$ 등으로 가정할 수 있다. 구간 $[\tau_0, \tau_1]$의 선택요령에 대해서는 Worsley(1988), Loader(1991), Chang, et al.(1994) 등에 언급되어 있다.

3. 추정량의 점근적 성질

지수분포의 변화점모형 (1.2)에서 Yao(1986)는 $\hat{\tau}$의 일치성 및 그 근사분포를 논의하였다. 와이

블분포에서는 형태모수 γ_i가 2보다 큰 경우에 최우추정량의 점근적 성질이 그 의미를 갖는다. 앞으로의 논의는 형태모수가 미지이고 동일한 값을 갖는 경우, 즉, $\gamma_1=\gamma_2=\gamma$(공통인 미지의 상수값)인 경우를 가정하자. 이 때, 우도함수는

$$l(\tau,\gamma) \propto X(\tau)\log\{\frac{X(\tau)}{\sum(t_i\wedge\tau)^\gamma}\}$$
$$+ (n-X(\tau))\log\{\frac{n-X(\tau)}{\sum\{t_i^\gamma-(t_i\wedge\tau)^\gamma\}}\}+\gamma\sum\log(t_i)+n\log(\gamma)$$

와 같이 표현되고 다음과 같은 성질이 성립한다.

<u>Proposition 1</u> γ가 알려진 경우에 추정량 $\hat{\tau}$는 τ의 일치추정량이다.

(증명) 변수 Z_i, $i=1,\cdots,n$, 가 서로 독립이고 위험률이 1인 지수분포를 따르는 확률변수라 할 때, 변수변환된

$$T_i=\begin{cases}(\beta_1 Z_i)^{1/\gamma}, & Z_i\le\tau^\gamma/\beta_1\\ (\beta_2 Z_i-(\frac{\beta_2}{\beta_1}-1)\tau^\gamma)^{1/\gamma}, & Z_i>\tau^\gamma/\beta_1\end{cases}$$

는 위험률 (1.3)을 갖는 서로 독립인 확률변수이다. 고정된 γ에 대해 Yao(1986)의 Lemma 3과 Lemma 4와 같은 방법으로 변화점 추정량의 일치성을 보일 수 있다.

추정량 $\hat{\tau}$의 근사분포를 유도하기 위해

$$\omega_1=\frac{\gamma}{\beta_1}\tau^{\gamma-1}e^{-\tau^\gamma/\beta_1},\quad \omega_2=\frac{\gamma}{\beta_2}\tau^{\gamma-1}e^{-\tau^\gamma/\beta_2}$$

라 하고,

$$R_i=\begin{cases}-\omega_1^{-1}\sum_{j=i}^{0}\beta_1 Z_j, & \text{for } i\le 0\\ \omega_2^{-1}\sum_{j=1}^{i}\beta_2 Z_j, & \text{for } i>0\end{cases}$$

라 정의하면 다음 정리가 얻어진다.

<u>Proposition 2</u> γ가 알려지고, $n\to\infty$일 때 추정량 $\hat{\tau}$의 근사분포는

$$n(\hat{\tau}-\tau)\xrightarrow{d}R_I$$

이다. 단, 기호 $\xrightarrow{d}$ 는 분포수렴(convergence in distribution)을 나타내고, I는 다음 식

$$V_i = i\log\left(\frac{\beta_2}{\beta_1}\right) + \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)R_i$$

를 최대로 하는 지수(index) i 의 값이다.

(증명) 식 (2.1)의 확률밀도함수에 대해서 Chernoff와 Rubin(1956)에서 요구되는 조건들이 성립함을 확인할 수 있으며, 따라서 Chernoff와 Rubin(1956)의 정리 결과로 부터 앞의 분포수렴에 관한 정리가 성립한다.

4. 붓스트랩 분포와 신뢰구간

4.1 붓스트랩 신뢰구간

변화점의 점추정 뿐만 아니라 신뢰영역에 대한 일반적인 논의는 Siegmund(1988)에 잘 나타나 있다. 변화점의 신뢰영역을 구하는 문제는 추정량의 점근분포와 직결되므로 이에 대한 근사공식을 유도하는 문제로 귀착된다. 그러나 와이블분포의 변화점모형에서 $\hat{\tau}$의 점근분포가 매우 복잡하기 때문에 본 연구에서는 붓스트랩(bootstrap)을 이용한 신뢰영역을 논의한다. Pham과 Nguyen(1991)은 모수적 붓스트랩(parametric bootstrap)을 제안하고 붓스트랩 추정량의 일치성을 논의하였다. 와이블분포의 변화점모형에서 $\hat{\tau}$의 점근분포를 직접 이용해서 신뢰영역을 구하는 문제는 매우 어렵기 때문에 붓스트랩을 이용한 근사적 방법이 바람직하다.

모수적 붓스트랩 절차를 간단히 논의하기 위해 $\theta_n = (\tau_n, \beta_{1n}, \beta_{2n})$라 하고 θ_n에 의해 정해지는 확률밀도함수를 f_{θ_n}와 같이 표현하자. Proposition 2에 주어진 $n(\hat{\tau}_n - \tau_n)$가 $U_n(t_1, \cdots, t_n, \tau_n)$에 의해 $n(\hat{\tau}_n - \tau_n) = U_n(t_1, \cdots, t_n, \tau_n)$와 같이 표현된다고 하자. 관찰값 $t_1, \cdots, t_n$이 주어질 때, 밀도함수 $f_{\hat{\theta}_n}$로 부터의 확률표본을 $t_1^*, \cdots, t_n^*$라 하면 이것이 모수적 붓스트랩 표본이 된다. 따라서 $n(\hat{\tau}_n - \tau_n)$의 붓스트랩 분포는 $U_n(t_1^*, \cdots, t_n^*, \hat{\tau}_n)$의 분포가 된다. 붓스트랩 표본에 근거한 최우추정량을 $\hat{\tau}_n^*$라 하면 Proposition 2와 유사한 방법으로 다음 결과를 얻을 수 있다.

<u>Proposition 3</u> $n(\hat{\tau}_n^* - \tau) \xrightarrow{d} R_I$, 단, R_I 는 Proposition 2에서 정의된 식이다.

붓스트랩 분포를 이용하여 변화점 τ의 $100(1-\alpha)\%$ 신뢰구간을 구하는 알고리듬은 다음과 같다.

단계 1: $\theta=(\tau,\beta_1,\beta_2)$의 MLE $\hat{\theta}$를 구한다.

단계 2: 식 (2.1)의 $f(t;\hat{\theta})$로 부터 크기 n인 붓스트랩 표본 $t_1^*,\cdots,t_n^*$를 얻고 이를 바탕으로 τ^*를 구한다.

단계 3: 단계2를 B번 반복하여 얻어진 $\hat{\tau}^{*(b)}$, $b=1,\cdots,B$, 의 경험적 분포를 이용하여 θ의 $100(1-\alpha)\%$ 신뢰구간 $(\hat{\theta}_L,\ \hat{\theta}_U)$를 구한다. 이 때, $\hat{\theta}_L(\hat{\theta}_L)$은 하위(상위) $\alpha/2$ 분위점이다.

이와같이 구해지는 신뢰구간은 단순히 붓스트랩 분포의 백분위수를 이용하기 때문에 추정량 $\hat{\tau}$의 분포가 편의된 경우에는 사용하는 데 문제가 발생한다. 따라서 편의수정(bias corrected)-가속(accelerated) 신뢰구간을 구하는 것이 바람직하다. 이에 대한 일반적 논의는 Efron과 Tibshirani(1993)를 참고한다. 편의수정-가속 신뢰구간을 구하는 절차를 간단히 설명해보자.

수정계수 z_0와 가속계수 a는 다음과 같이 정의된다.

$$z_0=\Phi^{-1}\left(\frac{no.(\hat{\tau}^{*(b)}\le\hat{\tau})}{B}\right),\qquad a=\frac{\sum(\hat{\tau}_{(\cdot)}-\hat{\tau}_{(i)})^2}{6\{\sum(\hat{\tau}_{(\cdot)}-\hat{\tau}_{(i)})^2\}^{3/2}}.$$

단, Φ는 표준정규분포의 누적분포이고, 기호 $no.(\hat{\tau}^{*(b)}\le\hat{\tau})$는 $\hat{\tau}$보다 작거나 같은 붓스트랩 추정치의 개수이다. 또한, $\hat{\tau}_{(i)}$는 i번째 관찰값을 빼고 계산된 변화점의 추정량이며 $\hat{\tau}_{(\cdot)}=\sum\hat{\tau}_{(i)}/n$ 이다.

단계 4: 명목수준 $100(1-\alpha)\%$인 편의수정-가속 신뢰구간은 붓스트랩 분포의 하위 α_1 분위점 및 상위 α_2 분위점을 택한다. 이 때

$$\alpha_1=\Phi\left(z_0+\frac{z_0-z_{\alpha/2}}{1-a(z_0-z_{\alpha/2})}\right),\qquad 1-\alpha_2=\Phi\left(z_0+\frac{z_0+z_{\alpha/2}}{1-a(z_0+z_{\alpha/2})}\right).$$

4.2 모의실험

몬테칼로 모의실험(Monte Carlo simulation)을 통해 제안된 신뢰구간의 포함확률(coverage probability)을 검토하기 위해 표본크기, 변화점의 상대적 위치 및 모수값에 따라 다음과 같은 모의실험을 계획한다.

① 표본크기: $n=80,\ 100,\ 120$
② 변화점의 위치: 40 백분위점 및 70 백분위점
③ 모수값의 비 β_2/β_1: 0.7, 1.2, 1.5, 2.0 (단, $\beta_2=1.0$, $\gamma=1.0$)

④ 변화점 위치에 관한 구간 $[\tau_1, \tau_2]$: τ_1는 하위 10분위점, τ_2는 상위 10분위점

위의 실험계획하에서 변화점을 추정하고 추정량의 오차 및 신뢰구간의 포함확률을 구하였다. 이때 붓스트랩 반복은 $B=2,000$이고 포함확률을 계산할 때의 반복은 200번으로 하였다. 포함확률은 변화점 τ를 포함하는 횟수를 200으로 나누어 구해진다. <표4.1>-<표4.6>은 표본크기, 변화점 위치에 다른 포함확률과 평균제곱오차(MSE) 및 신뢰구간 길이의 평균값을 나타낸 것이다. 포함확률과 명목신뢰수준간의 오차는 변화점의 상대적 위치와 위험률의 값에 따라 약간 큰 변이성을 나타내는 반면 표본크기에 따라서는 별로 영향을 받지 않는다. MSE는 위험률의 상대적 비가 클수록 작고 신뢰구간의 길이는 표본크기 및 위험률의 상대적 비가 클수록 짧아진다.

5. 결론 및 제언

본 연구는 중도절단 관찰값을 갖지 않는 완전자료를 가정하였으나 다음과 같이 중도절단자료의 변화점문제에도 확장될 수 있다. 중도절단 자료의 경우 X_i를 i번째 개체의 수명, C_i를 임의중도절단(random censoring) 시간이라 하면 관찰시간 $T_i = \min(X_i, C_i)$이고, 임의중도절단 데이타는 (T_i, δ_i), $i=1,2,\cdots,n$, 와 같이 표현된다. δ_i는 중도절단된 경우 0, 그렇지 않은 경우 1의 값을 갖는다. 일반적으로 우도함수는 다음 식

$$\prod_{\delta_i=1} f(t_i) \prod_{\delta_i=0} [1-F(t_i)]$$

과 같이 정의된다. 이를 변화점모형에 적합한 형태로 표현하면 중도절단이 없는 자료와 유사한 방법으로 추정치를 구할 수 있다. 앞으로의 연구에서 중도절단 비율이 추정치의 점근적 성질에 미치는 영향을 검토할 필요가 있다. 임의중도절단 자료에서 붓스트랩표본을 얻는 방법에 대해서 Efron(1981)은 카플란-마이어(Kaplan-Meier) 생존곡선(survival curve)으로부터 분포함수를 추정한 후 여기서 확률표본을 얻는 방법에 대해 논의하였다.

앞의 논의에서는 추정량의 점근적 성질을 유도하기 위해 $\gamma_1 = \gamma_2 = \gamma$(공통인 상수값) 경우를 가정하였는 데, 이것을 $\gamma_1 \neq \gamma_2$ 이고 미지인 일반적인 상황에서 추정량의 일치성과 점근분포에 대한 이론적 연구는 앞으로의 과제로 남겨둔다.

<표4.1> 경험적 포함확률(n=80)
변화점 위치=0.40 (40백분위점)

β_1	τ	MSE	$1-\alpha$	포함확률	평균길이
0.7	.7298	.3982	.90	.895	1.1791
			.95	.965	1.4647
			.99	.995	1.8773
1.2	.4257	.4591	.90	.970	1.3039
			.95	.995	1.6190
			.99	1.000	1.9610
1.5	.3406	.2689	.90	.950	.9858
			.95	.990	1.3597
			.99	1.000	1.8312
2.0	.2554	.1160	.90	.875	.5380
			.95	.960	.8598
			.99	.990	1.4530

<표4.2> 경험적 포함확률(n=80)
변화점 위치=0.70 (70 백분위점)

β_1	τ	MSE	$1-\alpha$	포함확률	평균길이
0.7	1.7200	.5972	.90	.835	1.8161
			.95	.935	2.1180
			.99	.990	2.4404
1.2	1.0033	.4088	.90	.935	1.4487
			.95	.980	1.6796
			.99	.995	1.9110
1.5	.8026	.2499	.90	.895	1.1687
			.95	.960	1.4090
			.99	.990	1.6892
2.0	.6020	.1117	.90	.905	.7516
			.95	.950	.9859
			.99	.985	1.3554

<표4.3> 경험적 포함확률(n=100)
변화점 위치=0.40 (40 백분위점)

β_1	τ	MSE	$1-\alpha$	포함확률	평균길이
0.7	.7298	.3955	.90	.905	1.1537
			.95	.975	1.4506
			.99	1.000	1.8777
1.2	.4257	.4550	.90	.945	1.2825
			.95	.980	1.6205
			.99	1.000	1.9898
1.5	.3406	.2165	.90	.925	.8742
			.95	.965	1.2530
			.99	1.000	1.7733
2.0	.2554	.0712	.90	.865	.4071
			.95	.970	.6796
			.99	.995	1.2860

<표4.4> 경험적 포함확률(n=100)
변화점 위치=0.70 (70 백분위점)

β_1	τ	MSE	$1-\alpha$	포함확률	평균길이
0.7	1.7200	.5588	.90	.850	1.8046
			.95	.930	2.1252
			.99	.990	2.4655
1.2	1.0033	.4036	.90	.920	1.4392
			.95	.970	1.6812
			.99	.990	1.9240
1.5	.8026	.2202	.90	.915	1.1417
			.95	.955	1.4033
			.99	.995	1.7055
2.0	.6020	.0885	.90	.855	.6394
			.95	.935	.8739
			.99	.970	1.2912

<표4.5> 경험적 포함확률(n=120)
변화점 위치=0.40 (40 백분위점)

β_1	τ	MSE	$1-\alpha$	포함확률	평균길이
0.7	.7298	.3679	.90	.905	1.1176
			.95	.960	1.3989
			.99	.980	1.8073
1.2	.4257	.4047	.90	.940	1.2624
			.95	.990	1.6262
			.99	1.000	2.0120
1.5	.3406	.2061	.90	.885	.8405
			.95	.985	1.2349
			.99	1.000	1.8022
2.0	.2554	.0755	.90	.840	.3651
			.95	.935	.6067
			.99	.985	1.1952

<표4.6> 경험적 포함확률(n=120)
변화점 위치=0.70 (70 백분위점)

β_1	τ	MSE	$1-\alpha$	포함확률	평균길이
0.7	1.7200	.5350	.90	.830	1.7870
			.95	.910	2.1020
			.99	.975	2.4589
1.2	1.0033	.3576	.90	.910	1.4442
			.95	.965	1.6912
			.99	.990	1.9413
1.5	.8026	.1895	.90	.895	1.0920
			.95	.955	1.3551
			.99	.995	1.6776
2.0	.6020	.0715	.90	.885	.5855
			.95	.945	.8126
			.99	.990	1.2125

6. 참 고 문 헌

[1] Antoniadis, A. and Gregoire, G.(1993), Nonparametric Estimation in Change-Point Hazard Rate Models for Censored Data: A Counting Process Approach, *Nonparametric Statistics* 3, 135-154

[2] Bhattacharyya, G. K. and Johnson, R. A.(1968), Nonparametric Tests for Shift at Unknown Time Point, *Annals of Mathematical Statistics* 39, 1731-1743

[3] Chang, I. S., Chen, C. H. and Hsiung, C. A.(1994), Estimation in Change-Point Hazard Rate Models With Random Censorship, Change-Point Problems, 78-92, Institute of Mathematical Statistics, Lecture Notes 23

[4] Chernoff, H. and Zacks, S.(1964), Estimating the Current Mean of a Normal Distribution Which is Subject to Changes in Time, *Annals of Mathematical Statistics* 35, 999-1018

[5] Chernoff, H. and Rubin, H. (1956), The Estimation of the Location of a Discontinuity in Density, Proc. 3rd Berkeley Synposium, *Math. Statist. Prob.* 1, 19-37

[6] Efron, B.(1981), Censored Data and the Bootstrap, *Journal of the American Statistical Association* 76, 312-19

[7] Hinkley, D. V.(1970), Inference About the Change-Point in a Sequence of Random Variables, *Biometrika* 57, 1, 1-17

[8] Loader, C. R.(1991), Inference for a Hazard Rate Change Point, *Biometrika* 78, 4, 749-757

[9] Matthews, D. E. and Farewell, V. T.(1982), On Testing for a Constant Hazard Against a Change-Point Alternative, *Biometrika* 38, 463-468

[10] Matthews, D. E. and Farewell, V. T.(1985), On a Singularity in the Likelihood for a Change-Point Hazard rate Model, *Biometrika* 72, 3, 703-704

[11] Muller, H. G. and Wang, J. L.(1990), Nonparametric Analysis of Changes in Hazard Rates for Censored Survival Data: An Alternative to Change-Point Models, *Biometrika* 77, 2, 305-314

[12] Nguyen, H. T., Rogers, G. S. and Walker, E. A.(1984), Estimation in Change-Point Hazard Rate Models, *Biometrika* 71, 2, 299-304

[13] Page, E. S.(1954), Continuous Inspection Schemes, Biometrika 41, 100-115

[14] Pettitt, A. N.(1979), A Non-parametric Approach to the Change-Point Problem, *Applied Statistics* 28, 126-135

[15] Pham, T. D. and Nguyen, H. T.(1990), Strong Consistency of the Maximum Likelihood Estimates in the Change-Point Hazard Rate Model, *Statisticis* 21, 2, 203-216

[16] Pham, T. D. and Nguyen, H. T.(1991), Bootstrapping the Change-Point of a Hazard Rate, Technical Report URA 397, Laboratoire de Modelisation at Calcul, IMAG, Universite Joseph Fourier Grenoble, 1-14

[17] Siegmund, D.(1988), Confidence Sets in Change-Point Problems, *International Statistical Review*, 56, 1, 31-48

[18] Worsley, K. J.(1988), Exact Percentage Points of the Likelihood-Ratio Test for a

Change-Point Hazard Rate Model, *Biometrics* 44, 259-63

[19] Yao, Yi-Ching(1986), Maximum Likelihood Estimation in Hazard Rate Models with a Change-Point, *Communications in Statistics, Theory and Methods*, 15(8), 2455-2466