Journal of the Korean
Data & Information Science Society
2000, Vol. 11, No. 2, pp. 279 ~ 293

위치모수 변화 모형에서 순위함수와 평균함수를 이용한 비모수적 변화점 추정 [1]

김재희[2] · 이경원[3]

요 약

위치 모수에 대해 1개의 변화점이 있는 경우 Carlstein(1988)의 변화점 추정량을 순위함수와 평균함수를 이용하여 변형시킨 변화점 추정통계량을 제안하였다. 모의 실험을 통해 Carlstein(1988) 변화점 추정량과 제안하는 변화점 추정량의 평균, 평균제곱오차와 변화점 추정비율을 계산하여 비교하였다.

주제어: 변화점, 경험누적함수, 순위함수, 평균함수

1. 서론

정보화 사회라고 일컬어지는 현대 사회에서는 기술의 급속한 발달과 정보의 중요성으로 인 해 수많은 자료들이 나오고 정리되고 있다. 이로 인하여 통계자료분석에서 발생하는 문제점에 대 한 개선과 새로운 방법의 모색에 대한 필요성이 더욱 요구되어지고 있다. 이러한 점에서 통계자료 분석의 방법 면에 요구되어지는 점은 수많은 자료들 중에서 의미있는 자료를 선택하여 분석하고, 만약 자료 내에 변화가 있는 경우 변화의 흐름을 파악하고 그 변화의 시점을 찾아내어 변화원인 에 대한 분석과 대안의 필요성을 알리는 것이다. 현대 통계학에서 연구되어져야하고 발전되어져야 할 부분은 단순한 자료의 분석이 아닌 자료 속에서 의미있는 정보를 찾아내는 방법론의 개발이라 할 수 있다.

변화점(change point)이란 관측값이 연속적인 시간에 의해 발생하는 경우 혹은 어떠한 일정 한 패턴에서 순서적으로 관측된 경우, 자료 내에 변화가 발생하는 시점을 의미한다. 변화점 추정 (change-point estimation)은 이러한 변화를 감지하고 변화가 발생하는 시점을 찾아내는 방법으로, 변화점 추정을 통해서 우리는 자료를 균일적인 부분으로 나눌 수 있다.

[1]연구는 한국과학기술 평가원의 2000년도 여자대학교 연구기반 확충사업의 연구비지원에 의한 것입니다.
[2]서울시 도봉구 쌍문동 419 덕성여자대학교 통계학과 부교수
[3]서울시 도봉구 쌍문동 419 덕성여자대학교 통계학 석사

변화점 추정에 대한 방법은 크게 모수적 접근방법과 비모수적 접근방법으로 나눌 수 있는데, 본 논문에서는 모집단의 분포함수에 대하여 모수형 가정을 하지 않은 비모수적 방법을 이용하여, 위치모수의 변화가 있는 경우 변화점 추정통계량을 개발하고자 한다. 특히 Carlstein(1988) 통계량 을 순위함수와 평균함수를 이용하여 변형시킨 변화점 통계량들을 제안하고, 모의실험을 실시하여 변화점 추정통계량들의 움직임을 비교하고자 한다.

본 논문에서는 비모수적 접근으로 연구되어진 기존의 변화점 추정 통계량들에 대해 2절에서 소개하고, Carlstein(1988)의 변화점 추정 통계량을 순위함수와 평균함수를 이용하여 변형시킨 통계량들을 3절에서 제안하고, 4절에서는 Carlstein(1988)의 변화점 추정량과 제안하는 변화점 통계량들을 S-PLUS를 이용한 모의실험을 통해 비교, 분석한다. 마지막 5절에서는 결론 및 제안을 한다.

2. 변화점 추정에 대한 기존연구

변화점(change-point)이란 연속적인 시간에 의해 발생된 확률변수가 주어진 경우, 자료에서 변화가 발생될 때의 시점을 말한다. 변화점 추정연구에서 고려되어지는 문제로는 변화시간의 추정 과 발생된 변화의 크기추정으로 나눌 수 있다. 본 연구에서는 변화시간의 추정, 즉 변화점 추정 문제를 다루고자 한다.

$X_1, X_2, \ldots, X_n$ 은 독립변수로 다음의 모형을 만족한다.

$$X_1, X_2, \ldots, X_\tau \quad \sim \text{iid} \quad F(x),$$

$$X_{\tau+1}, X_{\tau+2}, \ldots, X_n \quad \sim \text{iid} \quad G(x)$$

그리고

$$F(x) \neq G(x).$$

여기서 변화점 τ는 분포 F에서 분포 G로 변화가 시작하는 시점이 된다.

Pettit(1979)은 평균이 변화하는 모형에서 두 표본일 경우 맨-휘트니 통계량(Mann-Whitney two-sample test statistic)을 기초로 하여 비모수적 변화점 추정량을 제안하였다.

$$U_{(t,n-t)} = \frac{1}{2}\left\{\sum_{i=1}^{t}\sum_{k=t+1}^{n} sgn(X_i - X_k) + t(n-t)\right\}.$$

여기서

$$sgn(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

이다.

$$W_t = 2U_{(t,n-t)} - t(n-t), \quad t = 1, 2, \ldots, n-1$$

를 이용하여 Pettit(1979)은 변화점 추정통계량을

$$T_P = \arg\max_t W_t$$

로 제안하였다.

Schechtman(1982)은 Pettit(1979)의 추정량에서

$$(X_1, X_2, \ldots, X_j), \quad (X_{j+1}, \ldots, X_n)$$

의 두 표본에 대해, 맨-휘트니(Mann-Whitney) 통계량의 형태를 갖는 $U_{(t,n-t)}$를 수정하여

$$V_t = \frac{\frac{U_{(t,n-t)}}{t(n-t)} - 0.5}{\left[\frac{(n+1)}{12t(n-t)}\right]^{0.5}}, \quad t = 1, 2, \ldots, n-1,$$

를 얻는다.

Schechtman(1982)의 변화점 추정 통계량은

$$T_S = \arg\max_t V_t$$

이다.

Hawkins(1977, 1986)는 최소제곱추정법(least squares method)을 이용하여 변화점을 추정했 다. Hawkins(1977, 1986)는 다음의 모형으로부터 나온 자료로부터 최소제곱방법을 이용하여 변화점 τ 의 추정을 제안하였다.

$$X_i = \mu + \delta I(i \geq \tau) + \epsilon_i, \quad i = 1, 2, \ldots, n$$

여기서 μ, δ, τ 는 모르는 상수이고 ϵ_j는 $E[\epsilon_j] = 0$이고 $\sigma^2 = E[\epsilon_j{}^2] < \infty, \quad 1 \leq j \leq n$인 독립이고 동일한 분포를 따르는 오차항이다.

$$\bar{X}_t = \frac{1}{t}\sum_{i=1}^{t} X_i, \quad i = 1, 2, \ldots, n$$

일 때

$$Q_t = \frac{nt}{n-t}\left[\bar{X}_t - \bar{X}_n\right]^2 / \sigma^2$$

가 얻어지고 Hawkins(1986)의 변화점 추정통계량은

$$T_H = \arg\max_t Q_t$$

이다.

Lombard(1987)은 독립적인 관측치로부터 하나 또는 그 이상의 변화점을 검정하기 위해 순위 통계량에 기초를 둔 변화점 통계량을 제안하였다. 데이터를 순위함수로 바꿈으로

써 변화가 없는 귀무가설 하에서 분포무관 검정통계량을 얻을 수 있고 이상점의 영향을 덜 받게 된다. 서로 독립 인 확률변수 $X_1, \ldots, X_n$이 연속인 분포함수 $F(x, \theta_1), \ldots, F(x, \theta_n)$ 을 갖는다고 하면,

$$\theta_1 = \cdots = \theta_\tau = \theta, \qquad \theta_{\tau+1} = \cdots = \theta_n = \theta^*$$

인 경우 시점 τ를 변화점이라 한다. Lombard(1987)는 평활변화모형(smooth change model)과 가파른 변화모형(abrupt change model)을 고려하였다.

단순 가파른 변화(single abrupt change) 모형으로,

$$\theta_i = \begin{cases} \xi_1, & 1 \le i \le \tau \\ \xi_2, & \tau < i \le n, \end{cases}$$

점진적 변화를 고려한 모형으로 다음의 평활변화모형(smooth change model) 을 고려하였다.

$$\theta_i = \begin{cases} \xi_1, & i \le \tau_1 \\ \xi_1 + (i - \tau_1)(\xi_2 - \xi_1)/(\tau_2 - \tau_1), & \tau_1 < i \le \tau_2 \\ \xi_2, & i > \tau_2 \end{cases}$$

여기서 θ_i는 위치모수일 필요는 없고 $\xi_1, \xi_2, \tau_1, \tau_2$는 모르는 모수이다. $\tau_2 = \tau_1 + 1$ 인 경우에는 평활변화모형이 단순 가파른 변화모형이 됨을 알 수 있다.

평활변화모형(smooth change model)에서 Lombard(1987)의 변화점 추정통계량은 다음과 같다. rank $X_i = r_i$ 라고 할 때, 점수 함수(score function) ϕ는 $0 < \int_0^1 \phi^2(u)du < \infty$ 을 만족한다. 순위점수(rank score)를 $s(r_i)$ 라고 하면,

$$s(r_i) = \left[\phi\left(\frac{r_i}{n+1}\right) - \bar{\phi}\right] / A$$

여기서

$$\overline{\phi} = \frac{1}{n}\sum_{i=1}^{n} \phi\left(\frac{i}{n+1}\right),$$

$$A^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\phi\left(\frac{i}{n+1}\right) - \bar{\phi}\right]^2$$

이다. 순위통계량으로

$$v_{t_1,t_2} = \sum_{j=t_1+1}^{t_2} \sum_{i=1}^{t_1} s(r_i)$$

를 계산하고 표준화하여

$$\tilde{v}_{t_1,t_2} = v_{t_1,t_2} / \sigma(t_1/n, t_2/n)$$

를 얻는다. 여기서,

$$\sigma^2(u, v) = (1-u)^3(1+3u)/12 - (1-v)^3(1+3v)/12 - (1-v)^2(v^2 - u^2)/2$$

이다. Lombard(1987)는 변화점 추정 통계량으로

$$T_L = \arg\max_{t_1,t_2} \left|\tilde{v}_{t_1,t_2}\right|$$

을 제안하였다.

Carlstein(1988)은 한 시점을 중심으로 두 분포간의 거리를 최대로 하는 시점을 변화점으로 추정하였다.

$$X_1, X_2, \ldots, X_\tau \sim \text{iid} \quad F(x),$$

$$X_{\tau+1}, X_{\tau+2}, \ldots, X_n \sim \text{iid} \quad G(x)$$

라고 할 때, $F(x)$ 와 $G(x)$ 에 대한 분포함수의 특별한 가정 없이 단지 $\psi = \{x \in R : |F(x) - G(x)| > 0\}$ 에서 $\int_\psi dF(x) > 0$ 또는 $\int_\psi dG(x) > 0$ 가정만을 필요로 한다. $t \in \Lambda = \{i/n : 1 \le i \le n-1\}$ 에 대해 t시점 이전의 경험누적함수(pre-t empirical cdf)와 t시점 이후의 경험누적함수(post-t empirical cdf)는 다음과 같다.

$$_t h(x) = \sum_{i=1}^{nt} I\left\{X_i \le x\right\}/nt,$$

$$h_t(x) = \sum_{i=nt+1}^{n} I\left\{X_i \le x\right\}/n(1-t)$$

여기서

$$I(X \le a) = \begin{cases} 1, & \text{if} \quad X \le a \\ 0, & \text{if} \quad X > a \end{cases}$$

이다. t시점 이전, 이후의 경험누적 함수를 이용하여 Carlstein(1988)은 다음의 3가지의 거리 기준

$$S_1(\cdot) = \frac{1}{n}\sum_{i=1}^{n} |x_i|,$$

$$S_2(\cdot) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{0.5},$$

$$S_3(\cdot) = \sup_{1 \le i \le n} |x_i|.$$

을 고려하여 변화점 통계량을 제안하였다.

(1) 차이의 절대값의 합을 이용한 통계량

$$D_1(t) = t^{0.5}(1-t)^{0.5} n^{-1} \sum_{i=1}^{n} \left|_t h(x_i) - h_t(x_i)\right|,$$

변화점 추정통계량은

$$T_1 = \arg\max_{t} D_1(t)$$

이다.

(2) 차이의 제곱합을 이용한 통계량

$$D_2(t) = t^{0.5}(1-t)^{0.5}\left[\frac{1}{n}\sum_{i=1}^{n}({}_th(x_i) - h_t(x_i))^2\right]^{0.5},$$

변화점 추정통계량은

$$T_2 = \arg\max_t D_2(t)$$

이다.

(3) 차이의 최대값을 이용한 통계량

$$D_3(t) = t^{0.5}(1-t)^{0.5}\sup_{1\le i\le n}|{}_th(x_i) - h_t(x_i)|,$$

변화점 추정통계량은

$$T_3 = \arg\max_t D_3(t)$$

이다. Carlstein(1988)은 정리 1에서 Carlstein의 통계량 $\hat{\tau}$ 에 대하여, $\delta \in [0, 1/2)$에 고정되어 있다고 하면, $|\hat{\tau} - \tau|n^{\delta}$ 는 n이 커짐에 따라 0 에 수렴함을 보였다. 정리 2에서는 $\epsilon > 0$에 대하여 $c_1 > 0$이고 $c_2 > 0$인 상수와 모든 $n \ge n(\epsilon)$에 대하여 $P\left(|\hat{\tau} - \tau| > \epsilon\right) \le c_1 n \exp\left\{-c_2\epsilon^2 n\right\}$이 성립함을 보였다.

Scariano and Watkins(1988)는 최소제곱형태 추정치(least-squares type estimators), 면적 추정치(area estimators), 기울기 차이 추정치(slope differences estimators)를 이용하여 변화점 추정 통계량을 제안하였다. Scariano and Watkins(1988)은 다음을 만족하는 확률과정(stochastic process) 을 고려하였다.

(1) $X(t), \quad t \ge 0$ 는 독립적으로 증가한다.

(2) $X(t), \quad t \ge 0$ 는 단계함수(step function)이다.

(3) $X(t)$는 평균함수 $\mu(t)$를 갖는다.

$$\mu(t) = \mu(t; \tau, \nu_1, \nu_2) = \nu_1 t + (\nu_1 - \nu_2)(t - \tau)I(t - \tau)$$

여기서 τ는 변화점이고 τ, ν_1, ν_2 는 모르는 모수이다. Scariano and Watkins(1988)은

$$S^2 = \int_0^n (X(t) - m(t))^2 dt$$

를 최소화하는 τ, ν_1, ν_2의 추정량을 구해 최소제곱형태의 추정량을 제안하였다. 여기서,

$$m(t) = \begin{cases} \nu_1 t, & t = \tau \\ \nu_2(t-\tau) + X(n), & t > \tau \end{cases}$$

이다. $E(m(t)) = \mu(t)$ 라면, $\hat{\nu}_1$, $\hat{\nu}_2$, $\hat{\tau}$은 다음과 같다.

$$\hat{\nu}_1 = \int_0^{\hat{\tau}} 3tX(t)dt/\hat{\tau}^3$$

$$\hat{\nu}_2 = \int_{\hat{\tau}}^{n} 3(n-t)(X(n)-X(t))dt/(n-\hat{\tau})^3.$$

이때 변화점 추정 통계량 $\hat{\tau}$은 다음을 만족하는 해이다.

$$\tau^2(n-\tau)^2X(n) = 3(n-\tau)^2\sum_{i=1}^{c}(\tau)J_i t_i^2 + 3\tau^2 \sum_{i=c(\tau)+1}^{c(n)} J_i(n-t_i)^2.$$

여기서

$$c(t) = \text{no}(t_i \le t),$$

$$J_i = X(t_i) - X(t_i^-)$$

이다.

3. 제안하는 변화점 추정 통계량

독립적인 확률변수 $\{X_1, X_2, \ldots, X_n\}$는 평균의 변화가 1회 발생하는 다 음의 변화점 모형을 갖는다:

$$X_i = \begin{cases} \mu + \epsilon_i, & i = 1, 2, \ldots, \tau \\ \mu + \delta + \epsilon_i, & i = \tau+1, \ldots, n, \end{cases} \tag{1}$$

여기서 δ는 0이 아닌 임의의 상수이고, 오차항 ϵ_i는 서로 독립이며 평균이 0, 분산이 σ^2인 동일한 분포를 따른다. Carlstein(1988)의 변화점 통계량에 사용한 경험누적함수는 다음과 같다.

t시점 이전의 경험누적함수:

$${}_th(x) = \sum_{i=1}^{nt} I\{X_i \le x\}/nt, \tag{2}$$

t시점 이후의 경험누적함수:

$$h_t(x) = \sum_{i=nt+1}^{n} I\{X_i \le x\}/n(1-t). \tag{3}$$

Carlstein(1988) 통계량의 변형으로 경험누적함수 대신 순위함수를 이용하여 변화점 추정량을 제안하고자 한다. $\{X_1, X_2, \ldots, X_n\}$ 을 순서대로 배열하여 순위값 $\{R_1, R_2, \ldots, R_n\}$ 을 얻고 t시점 이전, t 시점 이후의 순위함수를 다음과 같이 정의한다.

t시점 이전의 순위함수:

$$ {}_tR(x) = \sum_{i=1}^{nt} R_i I\left\{X_i \le x\right\}/nt, \tag{4} $$

t시점 이후의 순위함수:

$$ R_t(x) = \sum_{i=nt+1}^{n} R_i I\left\{X_i \le x\right\}/n(1-t). \tag{5} $$

거리측도 S_1을 이용하여 D_{R1} 을 구하면

$$ D_{R1}(t) = t^{0.5}(1-t)^{0.5} n^{-1} \sum_{i=1}^{n} \left| {}_tR(x_i) - R_t(x_i) \right| $$

이다. 변화점 추정통계량으로

$$ T_{R1} = \arg\max_t \left\{D_{R1}(t)\right\} \tag{6} $$

를 제안한다. 거리측도 S_2를 이용하여 D_{R2}를 구하면

$$ D_{R2}(t) = t^{0.5}(1-t)^{0.5} \left[\frac{1}{n}\sum_{i=1}^{n}({}_tR(x_i) - R_t(x_i))^2\right]^{0.5} \tag{7} $$

이다. 제안하는 변화점 추정통계량은

$$ T_{R2} = \arg\max_t \left\{D_{R2}(t)\right\} \tag{8} $$

이다.

또한, 순위함수 대신 평균함수를 이용하여 변화점 추정통계량을 제안하고자 한다. t시점 이전, t 시점 이후의 평균함수를 다음과 같이 정의한다.
t시점 이전의 평균함수:

$$ {}_tA(x) = \sum_{i=1}^{nt} X_i I\left(X_i \le x\right)/nt, \tag{9} $$

t시점 이후의 평균함수:

$$ A_t(x) = \sum_{i=nt+1}^{n} X_i I\left(X_i \le x\right)/n(1-t). \tag{10} $$

거리측도 S_1를 이용하여 D_{A1}을 구하면

$$ D_{A1}(t) = t^{0.5}(1-t)^{0.5} n^{-1} \sum_{i=1}^{n} \left| {}_tA(x_i) - A_t(x_i) \right| \tag{11} $$

이다. 변화점 추정통계량으로

$$T_{A1} = \arg\max_t \{D_{A1}(t)\} \tag{12}$$

을 제안한다. 거리측도 S_2를 이용하여 D_{A2}를 구하면

$$D_{A2}(t) = t^{0.5}(1-t)^{0.5}\left[\frac{1}{n}\sum_{i=1}^{n}({}_tA(x_i) - A_t(x_i))^2\right]^{0.5} \tag{13}$$

이다. 변화점 추정통계량으로

$$T_{A2} = \arg\max_t \{D_{A2}(t)\} \tag{14}$$

을 제안하고자 한다.

제안하는 변화점 추정통계량에서 순위함수를 이용할 경우에는 순위값을, 평균함수를 이용할 경우에는 데이터 고유의 값을 변화점 추정통계량에 반영시키게되므로 사용하고자 하는 정보에 따라 변화점 추정통계량을 선택할 수 있다.

4. 모의실험

변화점 추정통계량들의 추정 능력을 비교하기위하여 S-PLUS를 이용한 모의실험을 행하고자 한다. 모의실험에서 사용한 변화점 모형은 변화점이 1개 존재하는 모형으로 다음과 같다.

$$X_i = \begin{cases} \mu_1 + \epsilon_i, & i = 1, 2, \ldots, \tau \\ \mu_2 + \epsilon_i, & i = \tau+1, \ldots, n \end{cases}$$

여기서 $\mu_2 = \mu_1 + \delta$ 로 표현될 수 있고 δ는 0이 아닌 상수이고 τ는 변화점이다. 오차항 ϵ_i은 평균이 0, 분산이 1인 정규분포, 이중지수분포를 따르며, 균일분포의 경우는 (-1,1)사이에서의 균일확률을 갖는 경우를 고려하였으며, 평균이 0, 분산이 1/3을 갖는다. 자료의 크기는 $n = 100$의 경우로 변화점 이 $\tau = 30$, 50인 경우, $\mu = 0$에 대하여 위치모수의 차이를 $\delta = 1, 2, 3, 4, 5$로 변화시키며 $1,000$번 반복실험으로 변화점 추정통계량의 움직임을 관찰하고자 한다. 변화점 추정 통계량의 추정 능력비교를 위해 반복 실험에서의 평균, MSE값과 변화점 추정비율을 계산한다. 여기서 변화점 추정비율은 주어진 변화점을 정확히 추정한 경우의 수를 모의실험 반복횟수로 나눈 값이다.

Carlstein(1988) 변화점 추정통계량 T_1, T_2와 순위 함수를 이용한 변화점 통계량 T_{R1}, T_{R2}, 평균함수를 이용한 변화점 통계량 T_{A1}, T_{A2}를 계산하여 변화점 추정치의 평균, 분산, 변화점의 정확한 추정비율을 계산하였다. 모의실험의 결과 같은 경험누적함수, 순위함수, 평균함수를 사용하더라도 거리 측도 기준에 따라 추정능력이 다름을 보여준다.

표 1은 오차항이 평균이 0이고 분산이 1인 정규분포를 따를 때 변화점이 $\tau = 30$, $\tau = 50$인 경우의 모의실험 결과이다. $\delta = 1$인 경우 T_{R1}, T_{R2}의 MSE값 이 크며 T_{A1}, T_{A2}의 추정비율

이 T_{R1}, T_{R2}에 비해 높으나 δ가 커질수록 순위함수를 이용한 T_{R1}, T_{R2}의 추정비율이 높아짐을 볼 수 있으나 Carlstein(1988) 통계량 T_1, T_2보다 $T_{R1}, T_{R2}, T_{A1}, T_{A2}$의 추정비율이 약간 떨어지는 것을 볼 수 있다.

표 2는 오차항이 $(-1,1)$에서 균일분포를 따를 때 변화점이 $\tau = 30$, $\tau = 50$인 경우의 모의실험 결과이다. T_{R1} 보다 T_{R2}가, T_{A1}보다 T_{A2}가 높은 추정비율을 보여준다. $\tau = 50$이고 평균의 차이가 2이상인 경우 Carlstein(1988) 통계량 T_1, T_2가 변화점에 대한 추정능력이 우수함을 볼 수 있다.

표 3은 오차항이 평균이 0이고 분산이 1인 이중지수분포를 따를 때 변화점이 $\tau = 30$, $\tau = 50$인 경우의 모의실험 결과이다. δ가 커질수록 T_{A1}, T_{A2}보다 T_{R1}, T_{R2}의 변화점 추정비율이 높음을 알 수 있다.

δ가 커질수록 제안하는 통계량 중 순위를 이용한 통계량 T_{R1}, T_{R2}가 평균함수 를 이용한 통계량 T_{A1}, T_{A2}보다 변화점 추정비율이 높음을 알 수 있고 경험누적분포함수를 이용한 Carlstein(1988) 통계량에 비해 변화점 추정비율이 크게 떨어지지 않음을 볼 수 있다. 같은 함수를 이용하더라도 절대값을 이용한 S_1측도보다는 유클리드 거리를 이용한 S_2측도를 사용한 경우 변화점 추정비율이 높게 나타났다.

5. 결론 및 제안

변화점에 대한 문제는 자료가 같은 분포에서 얻어지는 경우 뿐만 아니라 그렇지 않은 경우에도 고려되어질 수 있다. 본 연구에서는 특히 위치모수의 변화가 있을 경우, 자료 내에 존재하는 변화점을 추정하기 위해 기존의 비모수적 접근 방법들 중 Carlstein(1988)의 통계량으로부터 순위함수와 평균함수를 이용하여 변형시킨 방법을 제안하였다. 모의실험 결과를 보면 제안하는 변화점 통계량 중 순위함수를 이용한 통계량이 평균함수를 이용한 경우보다 변화의 크기가 커질수록 추정비율이 높음을 알 수 있다. 또한 제안하는 변화점 추정량이 Carlstein(1988) 변화점 추정량보다 추정비율면에서 약간 떨어지나 추정량의 평균은 크게 차이가 나지 않음을 볼 수 있다. 다양한 접근에 의한 여러 가지 변화점 추정 방법을 모색할 수 있고 효율이 더 좋은 변화점 추정방법을 기대한다.

참 고 문 헌

1. 송문섭, 박창순 (1994). 비모수통계학 개론, 자유아카데미

2. 장희윤 (1999). 위치모수의 변화가 있는 경우 비모수적 변화점 추정 통계량에 관한 연구, 덕성여자대학교.

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

4. Carlstein E. (1988). Nonparametric Changepoint Estimation, *The Annals of Statistics,* 16, 188-197.

5. Chernoff, H. and Zacks, S. (1964). Estimating The Current Mean of a Normal Distribution wich is Subjected to Changes in Time, *The Annals of Mathematical Statistics,* 35, 116-126.

6. Hawkins, D. L. (1986). A simple Least Squared Method for Estimating a Change Point in Mean, *Communications in Statistics,* 15(3), 655-679.

7. Lombard. F (1987). Rank Tests for Changepoint Problems, *Biometrika,* 74, 615-624.

8. Ritov, Y. (1990). Asymtotic Efficient Estimation of the Change Point with Unknown Distributions, *The Annals of Mathematical Statistics,* 18,1829-1839.

9. Scariano, S. M. and Walkins, T. A.(1988). Nonparametric Point Estimators for the Changepoint Problem, *Communications in Statistics,* 17(1), 3645-3675.

10. Scariano, S. M. and Walkins, T. A. (1988), An Algorithm for Computing Nonparametric Estimators of Changepoints, *Communications in Statistics,* 17(2), 1511-1532.

11. Scariano, S. M. and Walkins, T. A. (1990). Comparison of changepoint estimators, *Communications in Statistics,* 19(2), 619-636.

12. Schechtman, E. (1982). A Nonparametric Test for Detecting Changes in Location, *Communications in Statistics*, 11, 1475-1482.

13. Sen, A. and Srivastava, M. S. (1975). On Tests for Detecting Change in Mean, *The Annals of Statistics,* 3, 98-108.

표 1. $n = 100$일 때 오차항이 $N(0,1)$인 경우

		$\tau = 30$			$\tau = 50$		
		평균	MSE	추정비율	평균	MSE	추정비율
$\mu_1 = 0$	T_1	31.448	65.854	0.256	50.083	44.237	0.284
$\mu_2 = 1$	T_{R1}	35.904	288.286	0.142	52.091	155.201	0.152
	T_{A1}	27.850	80.040	0.170	44.698	140.338	0.162
	T_2	31.547	73.059	0.245	50.105	48.135	0.277
	T_{R2}	34.626	228.178	0.169	51.532	133.89	0.166
	T_{A2}	28.597	74.709	0.207	45.831	112.681	0.198
$\mu_1 = 0$	T_1	30.385	2.757	0.613	49.980	1.634	0.618
$\mu_2 = 2$	T_{R1}	30.873	13.923	0.447	50.360	6.566	0.467
	T_{A1}	27.889	19.391	0.401	46.796	47.166	0.345
	T_2	30.313	2.681	0.617	49.979	1.671	0.608
	T_{R2}	30.531	9.053	0.490	50.270	5.256	0.471
	T_{A2}	28.870	10.292	0.483	48.456	17.022	0.461
$\mu_1 = 0$	T_1	30.270	0.648	0.790	50.015	0.251	0.859
$\mu_2 = 3$	T_{R1}	30.492	2.134	0.661	50.230	1.042	0.736
	T_A1	28.820	9.100	0.512	48.066	20.098	0.440
	T_2	30.202	0.562	0.818	50.016	0.256	0.856
	T_{R2}	30.414	1.822	0.703	50.255	1.045	0.727
	T_{A2}	29.725	1.429	0.670	49.856	2.346	0.657
$\mu_1 = 0$	T_1	30.242	0.486	0.835	49.993	0.063	0.952
$\mu_2 = 4$	T_{R1}	30.463	1.351	0.743	50.199	0.425	0.812
	T_{A1}	29.646	2.238	0.605	49.261	5.057	0.531
	T_2	30.170	0.348	0.881	49.994	0.072	0.949
	T_{R2}	30.342	0.944	0.792	50.206	0.546	0.813
	T_{A2}	29.983	0.547	0.768	49.995	0.549	0.760
$\mu_1 = 0$	T_1	30.241	0.485	0.836	50.000	0.010	0.990
$\mu_2 = 5$	T_{R1}	30.453	1.261	0.750	50.202	0.328	0.837
	T_{A1}	30.011	0.951	0.675	49.820	1.754	0.616
	T_2	30.154	0.316	0.891	50.000	0.010	0.990
	T_{R2}	30.328	0.864	0.807	50.214	0.458	0.841
	T_{A2}	30.112	0.410	0.806	50.098	0.434	0.794

표 2. $n = 100$일 때 오차항이 (-1, 1)에서 균일분포를 따르는 경우

		$\tau = 30$			$\tau = 50$		
		평균	MSE	추정비율	평균	MSE	추정비율
$\mu_1 = 0$	T_1	30.390	5.720	0.493	49.964	4.798	0.466
$\mu_2 = 1$	T_{R1}	32.210	83.974	0.283	51.398	49.212	0.325
	T_{A1}	26.853	32.061	0.347	45.780	71.414	0.335
	T_2	30.336	7.142	0.478	49.898	6.400	0.451
	T_{R2}	30.622	27.002	0.362	50.519	29.687	0.339
	T_{A2}	28.064	18.782	0.401	47.518	32.918	0.384
$\mu_1 = 0$	T_1	30.183	0.319	0.864	50.000	0.000	1.000
$\mu_2 = 2$	T_{R1}	30.424	1.170	0.765	50.208	0.336	0.841
	T_{A1}	29.279	5.113	0.525	48.725	10.327	0.486
	T_2	30.121	0.199	0.904	50.000	0.000	1.000
	T_{R2}	30.343	0.881	0.800	50.195	0.325	0.854
	T_{A2}	29.837	1.137	0.702	49.791	0.923	0.710
$\mu_1 = 0$	T_1	30.183	0.319	0.864	50.000	0.000	1.000
$\mu_2 = 3$	T_{R1}	30.424	1.170	0.765	50.208	0.336	0.841
	T_{A1}	30.055	1.567	0.652	49.789	1.801	0.610
	T_2	30.121	0.199	0.904	50.000	0.000	1.000
	T_{R2}	30.343	0.881	0.800	50.195	0.325	0.854
	T_{A2}	30.194	0.518	0.841	50.133	0.311	0.840

표 3. $n = 100$일 때 오차항이 이중지수분포(평균=0, 분산=1)을 따르는 경우

		$\tau = 30$			$\tau = 50$		
		평균	MSE	추정비율	평균	MSE	추정비율
$\mu_1 = 0$	T_1	30.268	6.004	0.491	49.837	4.903	0.500
$\mu_2 = 1$	T_{R1}	30.952	17.100	0.383	50.432	12.716	0.430
	T_{A1}	27.781	51.735	0.244	45.517	107.893	0.227
	T_2	30.183	6.379	0.500	49.891	4.829	0.499
	T_{R2}	30.808	14.540	0.401	50.348	13.258	0.417
	T_{A2}	28.916	26.364	0.327	47.747	49.359	0.317
$\mu_1 = 0$	T_1	30.272	0.912	0.743	49.979	0.455	0.841
$\mu_2 = 2$	T_{R1}	30.530	2.462	0.638	50.242	1.104	0.720
	T_{A1}	29.291	6.103	0.495	48.647	13.971	0.839
	T_2	30.194	0.702	0.782	49.987	0.423	0.839
	T_{R2}	30.378	1.946	0.692	50.241	1.201	0.720
	T_{A2}	29.821	1.861	0.657	49.716	1.968	0.673
$\mu_1 = 0$	T_1	30.226	0.514	0.849	50.005	0.061	0.960
$\mu_2 = 3$	T_{R1}	30.479	1.347	0.715	50.162	0.292	0.850
	T_{A1}	29.958	1.780	0.623	49.672	3.486	0.640
	T_2	30.157	0.315	0.879	50.004	0.066	0.958
	T_{R2}	30374	0.988	0.765	50.151	0.317	0.853
	T_{A2}	30.106	0.578	0.755	50.073	0.447	0.803
$\mu_1 = 0$	T_1	30.189	0.345	0.862	50.001	0.023	0.980
$\mu_2 = 4$	T_{R1}	30.428	1.196	0.763	50.221	0.391	0.834
	T_{A1}	30.292	1.574	0.704	50.063	1.107	0.688
	T_2	30.124	0.202	0.901	50.001	0.023	0.980
	T_{R2}	30.339	0.873	0.800	50.227	0.503	0.838
	T_{A2}	30.193	0.613	0.813	50.155	0.505	0.808
$\mu_1 = 0$	T_1	30.201	0.327	0.841	50.000	0.006	0.997
$\mu_2 = 5$	T_{R1}	30.445	1.235	0.743	50.218	0.386	0.839
	T_{A1}	30.356	1.294	0.682	50.213	0.893	0.727
	T_2	30.134	0.192	0.883	49.999	0.011	0.995
	T_{R2}	30.389	1.071	0.778	50.230	0.498	0.843
	T_{A2}	30.242	0.544	0.791	50.221	0.513	0.823

Nonparametric Change-point Estimation with Rank and Mean Functions in a Location Parameter Change Model [4]

Jaehee Kim [5] · Kyoungwon Lee [6]

Abstract

This article suggests two change-point estimators which are modifications of Carlstein(1988) change-point estimators with rank functions and mean functions where there is one change-point in a mean function. A comparison study of Carlstein(1988) estimators and proposed estimators is done by simulation on the mean, the MSE, and the proportion of matching true change-point.

Key Words and Phrases: change-point, rank function, mean function

---

[4]This research was supported by KISTEP

[5]Associate Professor, Deartment of Statistics, Duksung Women's University, Ssangmun-Dong, Tobong-Ku 419, Seoul, 132-714, Korea