The Korean Communications
in Statistics Vol. 6, No. 2, 1999
pp. 467-477

Change-point Estimators Using Rank Average in Location Change Model[1)]

Jaehee Kim[2)] and Heeyoon Jang[3)]

Abstract

This paper deals with the problem of change-point estimation where there is one level change in location with iid errors. A change-point estimator using rank average is proposed with the proof of its consistency. A comparison study of various change-point estimators is done by simulation on the mean, the proportion, and the variance when the errors are from the normal and the double exponential distributions.

1. 서론

급속화되고 있는 정보과학과 기술의 개발은 통계자료분석의 문제점 개선과 새로운 방법 모색의 필요성을 부각시키고 있다. 여러 학문 분야, 품질관리 분야 등에서는 통계적 모형 설계와 추론을 포함한 통계적 분석에 대한 필요성은 나날이 높아지고 있다. 특히, 데이터에서 변화점의 감지 문제가 중요할 경우 변화에 대한 검정 문제와 변화점 추정 문제가 대두된다.

변화점은 얻어진 관찰 값들이 연속적인 시간에 근거하여 발생되는 경우, 또는 어떤 다른 모양의 패턴 내에서 내부적으로 순서화된 경우 흔히 발생된다. 이런 성격을 지닌 자료를 사용하여 그 자료의 유동성을 고려하고, 변화가 일어나는 시점인 변화점을 추정할 수 있는 통계량에 관한 연구가 활발하다. 변화점 추정 방법은 변화가 발생되어지는 위치나 크기를 고려하여 변화점을 찾아낸다. 본 연구에서는 데이터의 위치 변화가 있는 경우, 변화가 시작되는 변화점을 추정해 내는 문제를 다룬다.

본 논문에서는 변화점 추정에 대한 기존에 연구되어진 방법들을 2장에서 소개하고, 3장에서는 순위 평균을 이용하여 변화점을 추정하는 새로운 방법을 제안하고 그 방법의 일치성을 보인다. 4장에서는 앞장에서 소개된 변화점 추정통계량들과 제안하는 추정통계량을 S-plus를 이용한 모의실험을 통해 비교한다. 마지막 5장에서는 본 연구의 결론을 맺는다.

---

1) This research was supported by Natural Science Research Fund 1999, Duksung Women's University

2) Assistant Professor, Department of Statistics, Duksung Women's University, Ssangmun-Dong 419 Tobong-Gu, Seoul, Korea

3) Master in Statistics, Department of Statistics, Duksung Women's University

2. 기존 변화점 추정통계량 연구

변화점이란 순차적 확률변수가 주어진 경우, 자료 내에서 변화가 발생될 때의 시점을 말한다. 변화점 추정연구에서 고려되어지는 두 가지 문제는 첫 번째로 변화시간의 추정과 두 번째로 발생된 변화의 크기추정으로 크게 나눌 수 있다.

본 연구에서는 변화시간의 추정, 즉 변화점 추정의 문제를 다루고자 하며, 먼저 기존에 소개되어진 변화점 추정 방법을 알아보고자 한다.

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

$$\begin{aligned} &X_1, X_2, \cdots, X_\tau \sim^{iid} F(x), \\ &X_{\tau+1}, X_{\tau+2}, \cdots, X_n \sim^{iid} G(x), \quad \tau \in \{1, \cdots, n-1\} \\ &F(x) \neq G(x). \end{aligned} \tag{2.1}$$

여기서, 자료내의 실제 변화점 τ 는 분포 F에서 G로 변화가 시작될 때의 시점이 된다.

다음에서는 분포의 위치모수의 변화가 존재할 경우 모수적, 비모수적 접근에 의해 연구된 변화점 추정방법을 설명하고자 한다.

2.1 모수적 방법

모수적 방법으로 Hinkley(1970)는 최대가능도방법(maximum likelihood method)을 이용하여 변화점 추정통계량을 제안했다. 동일한 모수족에서 평균치의 차이가 존재하는 분포함수 F와 G를 얻어 그 가운데 존재하는 변화점을 최대가능도방법을 이용하여 추정한다. Hinkley(1970)의 모형을 살펴보면, 연속된 확률변수 $\{X_1, X_2, \cdots, X_n\}$에 1개의 변화점이 존재하는 경우, 그 모형은

$$\begin{aligned} X_t &= \theta_0 + \varepsilon_t \qquad t = 1, 2, \cdots, \tau \\ X_t &= \theta_1 + \varepsilon_t \quad t = \tau+1, \cdots, n \end{aligned} \tag{2.2}$$

이며, 여기서 오차항 ε_t 는 서로 독립이고 동일한 분포 $N(0, \sigma^2)$ 을 따른다. θ 는 연속적인 평균함수이고 변화점 τ 는 모르는 모수이다. 가능도함수

$$L(x_1, \cdots, x_n \mid \theta_0, \theta_1, \tau) = \frac{1}{2}\left\{ \sum_{i=1}^{\tau}(x_i - \theta_0)^2 + \sum_{i=\tau+1}^{n}(x_i - \theta_1)^2 \right\} \tag{2.3}$$

를 최대로 하는 변화점 추정량으로 Hinkley(1970)는

$$T_{Hink} = \arg\max_t \{Z_t^2\}$$

을 제안하였다. 여기서

$$Z_t^2 = \frac{t(n-t)(\overline{x_t} - \overline{x_t^*})^2}{n}, \quad t = 1, \cdots, n-1$$

이고 $\overline{x_t} = \sum_{i=1}^{t} x_i / t, \quad \overline{x_t^*} = \sum_{i=t+1}^{n} x_i / (n-t)$ 이다.

이 밖에도 모수적 방법을 통한 변화점 추정법은 Smith(1975)의 경우 베이지안 방법을 사용하여 변화점 추정방법을 제안하였고, Cobb(1978)은 분포함수 F와 G가 알려져 있는 모수족에 속하는

경우에 대하여 조건부적인 방법을 통해 변화점 추정방법을 제시했다.

2.2 비모수적 방법

Hawkins(1986)는 최소제곱법을 이용하여

$$Q_t = \left(\frac{nt}{n-t}\right)(\overline{x_t} - \overline{x_n})^2 / \hat{\sigma}^2,$$

$$\overline{x_t} = \frac{\sum_{k=1}^{t} x_k}{t}, \quad t = 1, 2, \cdots, n \text{ 일 때,}$$

변화점 추정량으로

$$T_{Hawk} = \arg\max_t \{Q_t\} \tag{2.4}$$

을 제안했다.

Pettit(1979)은 부호함수를 이용한 맨-휘트니(Mann-Whitney) 방법으로 변화점 추정통계량을 제안하였다.

$$M_{(t,n-t)} = \frac{1}{2}\left\{\sum_{j=1}^{t}\sum_{k=t+1}^{n} sgn(x_j - x_k) + t(n-t)\right\}$$

여기서, 부호함수의 정의는 다음과 같다.

$$sgn(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{그외.} \end{cases}$$

이때, $W_t = 2M_{(t,n-t)} - t(n-t)$, $t = 1, 2, \cdots, n-1$ 이며, 변화점 추정통계량으로

$$T_{Pt} = \arg\max_t \{W_t\} \tag{2.5}$$

을 고려했다.

Schechtman(1982)의 변화점 추정량은 Pettit(1979)의 방법을 수정한 모형으로 기본 개념은 맨-휘트니 통계량 M을 구하고 변화점 추정 방법으로

$$S_t = \frac{\left\{\frac{M_{(t,n-t)}}{t(n-t)} - 0.5\right\}}{\left\{\frac{(n+1)}{12t(n-t)}\right\}^{0.5}}, \quad t = 1, 2, \cdots, n-1$$

을 계산하고 변화점 추정량으로

$$T_{Schet} = \arg\max_t \{S_t\} \tag{2.6}$$

을 고려하였다.

Carlstein(1988)은 서로 다른 분포를 갖는 임의의 표본에서 비모수적 방법을 통해 변화점을 추정하고, 추정한 값의 오차 확률의 범위와 수렴 정도의 비교를 통해 강한 일치성을 지닌다는 사실을 보였다. 주어진 독립변수들은 다음 두 분포를 따른다고 모형을 가정한다.

$$X_1, X_2, \cdots, X_\tau \sim^{iid} F(x)$$

$$X_{\tau+1}, X_{\tau+2}, \cdots, X_n \sim^{iid} G(x)$$

여기서 $F(x)$와 $G(x)$는 모르는 임의의 분포함수로

$$\int_{\psi} dF(x) > 0, \quad \int_{\psi} dG(x) > 0$$

$$\psi = \{ x \in \mathbb{R} : | F(x) - G(x) | > 0 \} \qquad \mathbb{R}\text{: 실수의 집합}$$

을 기본 조건으로 한다. 이때, $s \in \Lambda = \{i/n : 1 \le i \le n-1\}$을 만족하는 경우에 대하여 s시점 이전의 경험누적분포함수(empirical CDF)와 s시간 이후의 경험누적분포함수를 다음과 같이 정의한다:

s시점 이전의 경험누적분포함수 (pre-s empirical CDF)

$$_{s}h(x) = \sum_{i=1}^{ns} I\{x_i \le x\}/ns \ , \tag{2.7}$$

s시점 이후의 경험누적분포함수 (post-s empirical CDF)

$$h_s(x) = \sum_{i=ns+1}^{n} I\{x_i \le x\}/n(1-s) \tag{2.8}$$

이때

$$I(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{그외}, \end{cases}$$

로 정의한다. Carlstein(1988)은 위의 경험누적분포함수를 이용한 방법을 통해 다음의 3가지 기준에 의한 통계량들을 제안했다.

(1) 차이의 절대값의 합을 이용한 통계량

$$D_1(s) = s^{0.5}(1-s)^{0.5} \frac{1}{n} \sum_{i=1}^{n} | \ _{s}h(x_i) - h_s(x_i) | \tag{2.9}$$

변화점 추정통계량 $T_{Carl\,1} = \arg\max_s \{D_1(s)\}$.

(2) 차이의 제곱합을 이용한 통계량

$$D_2(s) = s^{0.5}(1-s)^{0.5} \left[\frac{1}{n} \sum_{i=1}^{n} (\ _{s}h(x_i) - h_s(x_i))^2 \right]^{0.5}$$

변화점 추정통계량 $T_{Carl\,2} = \arg\max_s \{D_2(s)\}$.

(3) 차이의 최대값을 이용한 통계량

$$D_3(s) = s^{0.5}(1-s)^{0.5} \operatorname{Sup}_{1 \le i \le n} | \ _{s}h(x_i) - h_s(x_i) |$$

변화점 추정통계량 $T_{Carl\,3} = \arg\max_s \{D_3(s)\}$.

Carlstein(1988)의 변화점 추정통계량은 위치모수의 변화만이 아니라 확률적으로 F와 G가 일치하지 않는 모든 경우에 적용될 수 있다.

3. 순위 평균을 이용한 변화점 추정통계량 제안

본 연구에서 사용되는 순위 평균을 이용한 방법은 발생된 자료에 대해 순위를 주고, 시간 t가 변화할 때마다 t를 중심으로 전후에 발생하는 두 개의 데이터그룹에 대하여 각각의 순위 평균을 계산하고, 그 순위 평균들의 차를 제곱하여 그 값이 최대가 되는 시점을 변화점으로 추정한다.

독립적인 확률변수 $\{X_1, X_2, \cdots, X_n\}$에 평균의 변화가 한번 발생되는 경우 즉, 자료 내에서 하나의 변화점이 존재하는 경우로 다음 변화점 모형을 고려한다 :

$$X_t = \mu + \varepsilon_t \qquad t = 1, 2, \cdots, \tau$$
$$X_t = \mu + \delta + \varepsilon_t \qquad t = \tau + 1, \cdots, n \tag{3.1}$$

δ는 0보다 큰 임의의 상수이고, 오차항 ε_t 는 서로 독립이며 평균이 0이고 분산이 σ^2인 동일한 분포를 따른다. 이때 $\{X_1, \cdots, X_n\}$의 순위값 $\{R_1, \cdots, R_n\}$을 얻는다. t 시점에서 순위 평균의 차이를 이용한 통계량으로

$$R_t^2 = (\overline{R_{1t}} - \overline{R_{2t}})^2 \tag{3.2}$$

을 고려한다. 여기서,

$$\overline{R_{1t}} = \frac{1}{t} \sum_{i=1}^{t} R_i \ ,$$

$$\overline{R_{2t}} = \frac{1}{n-t} \sum_{i=t+1}^{n} R_i \ .$$

변화점인 경우 순위 평균의 차이가 최대가 된다는 점을 이용하여 변화점 추정통계량으로

$$T_1 = \ \arg\max_t \{R_t^2\} \tag{3.3}$$

을 제안한다. 위의 순위평균 모형을 분산을 고려한 모형으로 수정하여 적용하면

$$R_t^{*2} = \frac{(\overline{R_{1t}} - \overline{R_{2t}})^2}{\widehat{Var}(\overline{R_{1t}}) + \widehat{Var}(\overline{R_{2t}})}$$

을 얻을 수 있고 변화점 추정통계량으로

$$T_2 = \arg\max_t \{ R_t^{*2} \} \tag{3.4}$$

을 제안한다.

보조정리 3.1. $\phi(x)$가 표준 정규분포의 확률밀도함수이고 $\Phi(x)$는 표준정규분포의 누적분포함수일때 다음이 성립한다.

$$1 - \Phi(x) \le \frac{\phi(x)}{x}, \ \ x > 0$$

정리 3.1. 순위평균을 이용한 변화점 추정방법의 일치성

식(3.1)의 변화점 모형에서 식(3.3)의 순위평균을 이용한 변화점추정 통계량 T_1 은 일치성을 갖는다. 즉,

$$\lim_{r \to \infty} \lim_{n \to \infty} \sup P\{ \ | \ T_1 - \tau \ | > r\} \ = \ 0, \ \ r > 0 \tag{3.5}$$

증명:

변하는 시점 t 를 중심으로 양쪽으로 두 그룹으로 나누어 다음과 같이 정의하자.

$$\overline{R_{1t}} = \frac{1}{t} \sum_{1}^{t} R_j \ , \ \ \overline{R_{2t}} = \frac{1}{n-t} \sum_{t+1}^{n} R_j \ ,$$

$$\overline{R} = \frac{1}{n} \left(t\overline{R_{1t}} + (n-t)\overline{R_{2t}} \right)$$

이므로 $\overline{R_{2t}} = \dfrac{n\overline{R} - t\overline{R_{1t}}}{n-t}$ 을 얻을 수 있고 식(3.2)는 다음과 같다.

$$\begin{aligned} R_t^2 &= (\overline{R_{1t}} - \overline{R_{2t}})^2 \\ &= \left(\overline{R_{1t}} - \frac{n}{n-t}\overline{R} + \frac{t}{n-t}\overline{R_{1t}}\right)^2 \\ &= \left(\frac{n}{n-t}\overline{R_{1t}} - \frac{n}{n-t}\overline{R}\right)^2, \end{aligned}$$

여기서, $\overline{R} = \dfrac{1}{n}\sum_{j=1}^{n} j = \dfrac{n(n+1)}{2n} = \dfrac{n+1}{2}$ 이다.

$\hat{\tau} = T_1$ 으로 놓자. $P(\,|\,\hat{\tau} - \tau\,|\,> r) = P(\hat{\tau} > \tau + r) + P(\hat{\tau} < \tau - r)$ 으로 $\hat{\tau}$ 이 구간 $(\tau - r, \tau + r)$ 밖에서 발생될 확률을 의미하며 다음의 확률 부등식을 얻는다.

$$\begin{aligned} P(T_1 < \tau - r) &= P(\max_t \{R_t^2 > R_\tau^2\}) \\ &\le P\Big(\bigcup_{t<\tau-r} \{R_t^2 > R_\tau^2\}\Big) \\ &\le \sum_{t<\tau-r} P(R_t^2 > R_\tau^2). \end{aligned}$$

$t \le \tau$ 인 경우에 대하여 평균과 분산을 계산하면,

$$E(\overline{R_{1t}}) = \frac{\tau(\tau+1)}{2n}$$

$$E(\overline{R_{2t}}) = \frac{1}{2n(n-t)}\{n(n+1)(n-\tau) + \tau(\tau+1)(2\tau - t - n) - t(t+1)(\tau - t)\}$$

$$Var(\overline{R_{1t}}) = \frac{\tau(\tau+1)(2\tau+1)}{6nt} - \frac{\tau^2(\tau+1)^2}{4n^2 t}$$

$$Var(\overline{R_{2t}}) = \frac{(n^2-1)(n-\tau)}{12(n-t)^2} + O(n^{-1}).$$

중심극한 정리와 보조정리 3.1을 이용하면

$$\begin{aligned} P(R_t^2 > R_\tau^2) &= P\left\{\left(\frac{n}{n-t}\right)^2(\overline{R} - \overline{R_{1t}})^2 > \left(\frac{n}{n-\tau}\right)^2(\overline{R} - \overline{R_{1\tau}})^2\right\} \\ &= P\left\{\frac{1}{n-t}\left(\frac{n+1}{2} - \overline{R_{1t}}\right) > \frac{1}{n-\tau}\left(\frac{n+1}{2} - \overline{R_{1\tau}}\right)\right\} \\ &= P\left\{(n-t)\overline{R_{1\tau}} - (n-\tau)\overline{R_{1t}} > \frac{(\tau-t)(n+1)}{2}\right\} \\ &= P\left\{Y > \frac{(n+1)(\tau-t)}{2}\right\} \\ &\approx P\left\{\frac{Y - E(Y)}{\sqrt{Var(Y)}} > \frac{(n+1)(\tau-t)/2 - \tau(\tau+1)(\tau-t)/2n}{A\sqrt{n\tau^2}}\right\} \\ &= P\left\{Z > B\frac{(\tau-t)(n+1)}{2\sqrt{n\tau^2}}\right\} \end{aligned}$$

$$= P\left\{Z > C(\tau - t)\sqrt{\frac{n}{\tau^2}}\right\}$$

$$\leq \frac{1}{\sqrt{2\pi}C(\tau - t)}\sqrt{\frac{\tau^2}{n}}\, e^{-\frac{C^2(\tau-t)^2 n}{2\tau^2}}. \tag{3.6}$$

이 때,

$$Y = (n-t)\overline{R_{1\tau}} - (n-\tau)\overline{R_{1t}}$$

$$= (n-t)\frac{1}{\tau}\sum_1^{\tau} R_i - (n-\tau)\frac{1}{t}\sum_1^{t} R_i$$

$$= \left(\frac{n-t}{\tau} - \frac{n-\tau}{t}\right)\sum_1^{t} R_i + \frac{n-t}{\tau}\sum_{t+1}^{\tau} R_i$$

이고, A, B, C 는 임의의 상수이다. 또한, $t < \tau$ 인 경우, Y의 기대값과 분산을 계산하면 다음과 같다.

$$E(Y) = \frac{t(n-t) - \tau(n-\tau)}{\tau t} \cdot \frac{t\tau(\tau+1)}{2n} + \frac{(n-t)(\tau-t)}{\tau} \cdot \frac{\tau(\tau+1)}{2n}$$

$$= \frac{\tau(\tau+1)(\tau-t)}{2n},$$

$$Var(Y) = \left[\left\{\frac{t(n-t) - \tau(n-\tau)}{\tau t}\right\}^2 t + \frac{(\tau-t)(n-t)^2}{\tau^2}\right]\sigma_1^2$$

$$= O\left(\frac{n^2}{\tau}\right) \cdot O\left(\frac{\tau^3}{n}\right)$$

$$= O(n\tau^2),$$

여기서

$$\sigma_1^2 = Var(R_i) = O\left(\frac{\tau^3}{n}\right), \qquad i < \tau.$$

그러므로 식(3.6)을 이용하면

$$\sum_{t<\tau-r} P(R_t^2 > R_\tau^2) \leq \frac{1}{\sqrt{2\pi}C}\,\frac{\tau}{(\tau-t)}\sqrt{\frac{\tau^2}{n}}\, e^{-\frac{C^2(\tau-t)^2 n}{2\tau^2}} \tag{3.7}$$

이다. $r \to \infty, n \to \infty$ 일 때, $\tau = O(n)$ 이므로 식 (3.7)의 우변은 0으로 수렴한다. 그러므로 $\lim_{r\to\infty}\lim_{n\to\infty}\sup P(|\hat{\tau} - \tau| > r) = 0$ 을 얻는다. 반대의 경우 $P(\hat{\tau} > \tau + r)$역시 마찬가지 방법으로 계산된다 ▪

4. 모의 실험

본 연구에서는 S-plus를 사용한 모의실험을 통해 통계량의 움직임을 비교한다. 모의실험에서 사용되는 모형은

$$X_t = \begin{cases} \mu + \varepsilon_t & t \leq \tau \\ \mu + \delta + \varepsilon_t & t > \tau \end{cases} \tag{4.1}$$

으로 $t=1,\cdots,n$ 이고, δ는 0보다 큰 상수이다. 오차항 ε_t의 분포는 정규분포와 이중지수분포에서 발생되는 2가지 경우를 고려하였다. 또한 오차항은 평균과 분산이 각각 0과 1을 따르도록 실험환경을 설정하였고, 자료의 크기는 n=100에 대하여 $\tau=30,50$인 경우, $\mu=0$에 대하여 위치모수의 차이를 $\delta=1,2,3,4$로 변화시켜가며, 500번의 반복실험을 통해 변화점 추정량들의 움직임을 관찰한다.

제안된 2가지 순위 평균에 의한 변화점 추정통계량(T_1, T_2)과 Hinkley (T_{Hink}), Schechtman (T_{Scht}), Pettit (T_{Pt})이 제안했던 변화점 추정통계량들과 식(2.9)의 Carlstein (T_{Carl})의 변화점 추정 통계량을, 모의실험을 통해 비교한다. 모의 실험에서는 변화점 추정치들의 평균, 분산, 주어진 변화점의 정확한 추정비율을 구하여, 비교한다. 변화점을 추정할 때 자료의 앞부분이나 뒷부분에서 변화점의 발생의 경우를 제외한 $t=10,11,\ldots,90$ 인 경우만을 고려하였다.

표 1은 오차항이 평균이 0이고 분산이 1인 정규분포를 따를 때 변화점이 $\tau=50$, $\tau=30$ 에서 발생할 때의 모의실험 결과이다. 평균차이가 2 이상으로 커지면 순위평균을 이용한 방법은 기존의 다른 추정량의 결과와 비슷한 결과를 나타내고 분산을 고려한 순위평균의 추정량의 경우 추정치에 대한 분산도 안정되게 나타난다.

표 2는 오차항이 평균이 0이고 분산이 1인 이중지수분포를 따르고 변화점이 $\tau=50$ 과 $\tau=30$ 에서 발생할 때의 모의실험 결과이다. 순위평균을 이용한 방법은 평균의 차이가 2 이상인 경우부터 Pettit(1979)의 추정값보다 분산은 줄고 추정비율은 증가한다. 또한 평균 차이가 3 이상이 되면 Carlstein(1988)이나 Schechtman(1982)의 방법보다도 분산이나 추정 비율면에서 더 좋은 결과를 보인다. 그 결과 평균이 2 이상 차이가 나게되면 제안된 순위평균의 방법은 기존의 방법과 비교시 정확한 변화점 추정비율이 더 크다. 제안하는 순위평균의 추정량 T_1, T_2는 평균 차이가 클수록 추정 능력이 향상되고 특히 변화점이 데이터의 중간이 아닌 부분에서 발생하고 오차항이 이중지수분포를 따를 때 Pettit(1979), Carlstein(1988)이나 Schechtman(1982)의 추정결과보다도 우수한 변화점 추정비율을 갖는다. 이는 순위 평균을 이용한 방법이 이상점의 영향을 덜 받고, 다른 통계량에 있는 $t(n-t)$ 값의 영향을 받지 않기 때문인 것으로 여겨진다.

5. 결론

변화점에 관한 문제는 데이터가 동일한 분포에서 얻어지지 않는 경우에 해당하는 여러 가지 모형에서 고려될 수 있다. 본 연구에서는 특히 위치모수의 변화가 있을 때 데이터 내에 존재하는 변화점을 추정하기 위해 비모수적 접근으로 순위 평균을 이용한 방법을 제안하였고 모의 실험 결과를 비교하여 기존에 제안된 변화점 추정방법들과 마찬가지로 순위 평균을 이용한 방법이 유의하게 사용될 수 있음을 알았다. 그리고 변화점이 자료의 중간점에서 발생하지 않는 경우 순위 평균을 이용한 추정방법은 Pettit(1979)의 방법보다도 우수하게 나타났고 평균값의 차이가 좀더 커지게 되면 Carlstein(1988)이나 Schechtman(1982)의 추정법과 비교했을 때도 더 좋은 결과를 얻을 수 있어 순위 평균 방법의 유효성을 알 수 있었다. 본 논문에서 소개한 방법 이외에도 여러 가지 변화점 추정 방법을 모색할 수 있고 효율이 더 좋은 변화점 추정방법을 기대한다.

참 고 문 헌

[1] 송문섭, 박창순 (1994). 비모수통계학개론, 자유아카데미.

[2] Bhattacharya, P. K. (1994). Some aspects of change-point analysis. *IMS lecture notes. 23*

[3] Carlstein, E. (1988). Nonparametric change-point estimation. *Annals of Statistics. 16, 188-197.*

[4] Cobb, G. W. (1978) The problem of the Nile : Conditional solution to a change-point problem. *Biometrika 65, 243-251.*

[5] Hawkins, D. L. (1986). A simple least squares method for estimating a change in mean. *Communications in statistics. 15(3), 655-679.*

[6] Hinkley, D. V. (1970). Inference about the change-point in a sequence of random-variables. *Biometrika. 57, 1-16.*

[7] Lombard, F. and Hart, J. D. (1994). The analysis of change-point data with dependent errors. *IMS Lecture Notes-Monograph Series. Vol 23*

[8] Scariano, S. M. and Watkins, T. A. (1990). Comparisons of change-point estimators. *Communications in Statistics. 19(2), 619-636.*

[9] Schechtman, E. (1982). A nonparametric test for detecting changes in location. *Communications in statistics-Theory and Methods. 11(13), 1475-1482.*

[10] Smith, A. F. M. (1975) A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika 62, 407-416.*

[11] Page. E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika 42, 523-527.*

[12] Pettit, A. N. (1979). A Nonparametric approach to the change-point problem. *Applied Statistics, 28, 126-135.*

표 1. $n=100$ 이고, 오차항이 $N(0,\ 1)$을 따를 때 변화점 추정통계량 비교

		$\tau=50$			$\tau=30$		
		평균	분산	추정비율	평균	분산	추정비율
$\mu_0=0, \mu_1=1$	T_1	48.99	716	0.174	26.40	162	0.164
	T_2	48.39	740	0.166	26.08	160	0.164
	T_{Hink}	49.87	46.54	0.268	30.68	42.93	0.258
	T_{Carl}	50.02	44.77	0.274	31.25	56.24	0.262
	T_{Schet}	49.94	47.60	0.274	31.13	49.62	0.262
	T_{Pt}	49.62	16.75	0.294	33.62	48.13	0.238
$\mu_0=0, \mu_1=2$	T_1	49.56	46.35	0.606	28.20	19.48	0.556
	T_2	49.61	49.87	0.606	28.25	20.62	0.558
	T_{Hink}	49.98	2.56	0.612	30.00	2.20	0.636
	T_{Carl}	49.98	2.31	0.628	30.41	3.38	0.620
	T_{Schet}	49.98	2.31	0.628	30.41	3.38	0.620
	T_{Pt}	49.95	1.69	0.632	31.41	7.77	0.496
$\mu_0=0, \mu_1=3$	T_1	49.95	0.37	0.864	29.84	0.68	0.828
	T_2	50.04	0.32	0.860	29.92	0.57	0.834
	T_{Hink}	50.02	0.28	0.858	30.01	0.22	0.866
	T_{Carl}	50.03	0.29	0.870	30.25	0.51	0.814
	T_{Schet}	50.03	0.29	0.870	30.25	0.50	0.814
	T_{Pt}	50.01	0.22	0.878	30.87	2.19	0.602
$\mu_0=0, \mu_1=4$	T_1	49.98	0.06	0.940	29.95	0.12	0.928
	T_2	49.99	0.05	0.948	30.03	0.09	0.938
	T_{Hink}	50.00	0.05	0.948	29.99	0.05	0.952
	T_{Carl}	49.99	0.05	0.948	30.23	0.37	0.824
	T_{Schet}	49.99	0.05	0.948	30.23	0.37	0.824
	T_{Pt}	49.99	0.05	0.948	30.72	1.54	0.624

표 2. $n=100$ 이고, 오차항이 이중지수분포 $(\alpha=0, \beta=1/\sqrt{2})$를 따를 때 변화점 추정통계량 비교

		$\tau=50$			$\tau=30$		
		평균	분산	추정비율	평균	분산	추정비율
$\mu_0=0, \mu_1=1$	T_1	49.54	468.4	0.234	25.28	73.70	0.244
	T_2	49.49	536.4	0.220	25.03	101	0.232
	T_{Hink}	50.04	33.27	0.300	29.88	23.23	0.328
	T_{Carl}	50.01	22.10	0.334	30.21	12.36	0.366
	T_{Schet}	49.99	21.71	0.338	30.19	12.35	0.368
	T_{Pt}	49.97	10.75	0.358	32.39	20.33	0.332
$\mu_0=0, \mu_1=2$	T_1	50.02	26.73	0.674	28.99	7.15	0.614
	T_2	50.20	18.67	0.670	28.56	13.60	0.600
	T_{Hink}	50.03	2.17	0.672	29.91	1.35	0.704
	T_{Carl}	50.09	1.54	0.688	30.13	1.63	0.676
	T_{Schet}	50.09	1.54	0.688	30.13	1.63	0.676
	T_{Pt}	50.06	0.92	0.704	31.11	4.72	0.560
$\mu_0=0, \mu_1=3$	T_1	50.02	0.47	0.844	29.79	1.51	0.830
	T_2	50.02	0.47	0.842	29.83	1.74	0.834
	T_{Hink}	50.00	0.34	0.850	30.01	0.29	0.872
	T_{Carl}	50.03	0.36	0.852	30.23	0.44	0.808
	T_{Schet}	50.03	0.36	0.852	30.23	0.44	0.808
	T_{Pt}	50.02	0.35	0.854	30.89	2.81	0.616
$\mu_0=0, \mu_1=4$	T_1	49.97	0.12	0.942	30.00	0.10	0.944
	T_2	49.98	0.11	0.944	30.04	0.24	0.928
	T_{Hink}	49.98	0.08	0.946	30.03	0.13	0.938
	T_{Carl}	49.98	0.10	0.948	30.22	0.33	0.848
	T_{Schet}	49.98	0.10	0.948	30.22	0.33	0.848
	T_{Pt}	49.98	0.09	0.948	30.78	2.25	0.642