Journal of Korean
Data & Information Science Society
2002, Vol. 13, No.1 pp. 165 175

Comparison of Change-point Estimators with Scores[1)]

Jaehee Kim[2)] · Hyunjoo Seo[3)]

Abstract

We consider the problem of estimating the change-point in mean change model with the one change-point. Lombard(1987) suggested change-point estimation based on score functions. Gombay and Huskova(1998) derived a class of change-point estimators with the score function of rank. Various change-point estimators with the log score functions of ranks are suggested and compared via simulation.

Keywords : Change-point, Rank, Score function, Wiener process, Brownian bridge

1.서론

변화점(change-point)에 관한 문제는 품질관리 부문에서 응용, 연구되기 시작했다. 예를 들어 특정 생산라인의 결과에 대한 관측치가 허용범위를 벗어나기 시작하면 품질관리 담당자는 변화가 일어나는 점을 알아내어 원인을 파악하고 문제를 해결하는 과정이 필요하다. 마찬가지로 시간에 따른 실험, 관측치를 분석해야하는 의학, 생명과학, 경제학, 기상학, 환경공학 등의 분야에서도 변화가 일어나는 현상에 대한 통계적 분석의 중요성은 점점 강조되고있다.

변화가 일어나는 상황은 시간에 따라 변화하는 확률과정모형(stochastic process)으로 표현할 수 있어 변화가 일어나는 것을 시간에 따른 확률과정 또는 확률변수의 분포에 대한 변화로 이해할 수 있다. 변화에 대한 기본적인 물음은 관측한 확률변수가 독립적이고 동일한 분포를 따르는가 아니면 적어도 한 점 이상에서 분포의 변화가 일어나느냐에 관한 것으로 볼 수 있다.

변화점 추론에 대한 일반적인 접근방법은 모든 데이터에 대하여, 매 시점에서 테이터를 두 부분으로 나누고 적당한 측도(measure)를 사용하여 두 부분의 비균일성(heterogeneity)을 계산하여 비균일성을 최대로 하는 점을 변화점으로 추정하는 것이

1. 이 연구는 2002년도 덕성여자대학교 교내연구비에 의해 지원 받았음.
2. 서울시 도봉구 쌍문동 419 덕성여자대학교 통계학과 부교수
E-mail : jaehee@duksung.ac.kr
3. 서울시 도봉구 쌍문동 419 덕성여자대학교 통계학과 통계학석사

다. 이러한 단계는 확률과정에 대한 이해와 이론을 통한 통계량의 특성연구를 필요로 하며 때로는 복잡한 수리 통계적 계산이 요구된다.

본 논문에서는 기존의 비모수적 변화점 추정방법과 점수함수를 이용한 Lombard(1987)의 방법과 log 점수함수를 이용한 방법, Gombay 와 Huskova(1998)의 조건에 맞는 점수함수(score function)와 순위점수(rank score)를 이용하여 변형시킨 변화점 추정통계량들의 추정능력을 비교하고자 한다. 기존의 변화점 추정통계량들에 대한 이론을 2장에서 소개하고, 3장에서는 log 점수함수와 순위를 이용한 변화점추정량의 변형을 고려하고 4장에서는 2장과 3장에서 소개된 변화점 추정통계량들에 대해 S-plus를 이용한 모의실험을 통해 비교하고자 한다. 마지막 5장에서는 본 논문의 결론 및 제안을 한다.

2. 기존 변화점 추정통계량

변화점(change-point)란 연속적인 시간에 의해 발생된 확률변수가 자료에서 변화가 발생될 때의 시점을 나타내며 위치모수의 변화, 분산모수의 변화 등이 일어나는 시점이다. 변화점 추정연구에서 고려되어지는 문제로는 변화시간의 추정과 발생된 변화의 크기 추정으로 나눌 수 있으며 본 연구에서는 위치모수의 변화점 추정문제를 다루고자 한다.

$X_1, X_2, \ldots, X_n$ 은 독립변수로 $E(X_i) = \mu_i$, $Var(X_i) = \sigma^2$ 을 가지며 평균값의 수준의 변화가 변화점 τ 에서 1번 발생하는 다음의 모형을 갖는다.

$$X_i = \mu + \delta I_{(i \geq \tau)} + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

여기서, μ, δ, τ 는 모르는 상수이고 ε_i는 $E[\varepsilon_i] = 0$이고 $\sigma^2 = E[\varepsilon_i^2] < \infty$, $1 \leq i \leq n$인 독립이고 동일한 분포를 따르는 오차항이다.

Hinkley(1970)는 최대가능도법(maximum likelihood method)을 이용하여 변화점 추정통계량을 제안했다. 동일한 모수족에서 평균치의 차이가 존재하는 분포함수 F와 G를 얻어 그 가운데 존재하는 변화점을 최대가능도법을 이용하여 추정한다. Hinkley(1970)는 오차항이 정규분포를 따를 때 우도함수와 θ_0 에 대한 최대우도추정량을 얻고

$$\overline{X}_t = \sum_{j=1}^{t} X_j / t \ , \quad \overline{X}_t^* = \sum_{j=t+1}^{n} X_j / (n - t)$$

일 때

$$Z_t^2 = t(n - t)(\overline{X}_t - \overline{X}_t^*)^2 / n, \quad t = 1, \ldots, n - 1$$

를 얻고 변화점 추정량으로

$$T_H = \arg\max_{1 < t < n} \{Z_t^2\}$$

을 제안했다. 즉, 우도함수(likelihood function)를 최대화하는 시점이 주어진 모형 내에서 발생하는 변화점으로 추정된다.

Darkhovsh(1976)은 맨-휘트니 통계량(Mann-Whitney statistic)에 기초를 둔 변화점 통계량을 제안했고, Bhattachryya 와 Johnson(1978)은 변화점의 유무에 관한 비모수적 검정법을 제안하였다. Pettit(1979)는 평균이 변화하는 모형에서 두 표본일 경우

맨-휘트니 통계량을 기초로 하여 비모수적 변화점 추정량들을 제안하였다. Pettit(1979)의 변화점 추정량은다음과 같다.

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

여기서, 부호함수의 정의는 다음과 같다.

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

이다.

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

를 이용하여 Pettit(1979)은 변화점 추정통계량으로

$$T_P = \arg\max_{1<t<n} W_t$$

을 제안하였다.

Schechtman(1982)은 $U_{(t,n-t)}$이 $(X_1, X_2, \ldots, X_j)$, $(X_{j+1}, \ldots, X_n)$ 와 같이 두 표본인 경우, 맨-휘트니 윌콕슨 통계량(Mann-Whitney-Wilcoxon statistic)으로부터 맨-휘트니 형태를 갖는다고 가정하였다. 이로부터 다음과 같은 형태

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

를 갖는다. Schechtman(1982)의 평균 변화 모형에서 변화점 추정량으로

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

일 때

$$T_S = \arg\max_{1<t<n}\{V_t\}$$

를 제안하였다.

Hawkins(1977, 1986)는 최소제곱추정법을 이용하여 평균의 차이가 존재하는 임의의 표본에 대한 변화점 추정방법을 제안하였다.

$$\overline{X_t} = \sum_{i=1}^{t} X_i / t, \quad t = 1,2,\ldots,n \text{ 이고,}$$

$$Q_t = \frac{nt}{n-t}\left[\overline{X_t} - \overline{X_n}\right]^2 / \sigma^2$$

일 때 Hawkins(1986)는 변화점 추정통계량으로

$$T_{Hawk} = \arg\max_{1<t<n}\{Q_t\}$$

를 제안하였다. 특히 오차항이 정규분포를 따를 경우 Hinkley(1970)의 변화점 추정량 Hawkins(1986)의 변화점 추정 통계량과 같게 된다. Lombard(1987)은 독립적인 관측치로부터 하나 또는 그 이상의 변화점을 검정하기 위해 순위 통계량에 기초를 둔 변화점통계량을 제안하였다. 데이터를 순위함수로 바꿈으로써 변화가 없는 귀무가설 하에서 분포무관 검정통계량을 얻을 수 있고 이상점의 영향을 적게 받게 된다. 서로 독립인 확률변수 $X_1, X_2, \ldots, X_n$ 이 연속인 분포함수 $F(x,\theta_1), \ldots, F(x,\theta_n)$ 을 가질때

$\theta_1 = \cdots = \theta_\tau = \theta$, $\theta_{\tau+1} = \cdots = \theta_n = \theta^*$ 인 경우 시점 τ를 변화점이라 한다. Lombard(1987)는 평활변화모형(smooth change model)과 가파른 변화모형(abrupt change model) 2가지를 모두 고려하였다.

단순 가파른 변화(single abrupt change)는 다음과 같다.

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

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 ξ_1, ξ_2, τ_1, τ_2는 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. $\text{rank}(X_i) = r_i$ 라고 할 때, 점수함수(score function) ϕ는 $0 < \int_0^1 \phi^2(u)du < \infty$ 을 만족한다. 순위점수(rank score)는 $s(r_i)$,

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

이고

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

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 $T_L = \arg\max_{t_1,t_2}\{|\tilde{v}_{t_1,t_2}|\}$ 을 제안하였다. 특히, 변화점이 1개일 경우,

$$T_L = \arg\max_{1<t<n}\{|\tilde{v}_{t,t+1}|\}$$

로 표현된다.

Carlstein(1988)은 한 시점을 중심으로 두 분포간의 거리를 최대로 하는 시점을 변화점으로 추정하였다. 주어진 독립변수들은 다음 두 분포를 따른다고 모형을 가정한다.

$$X_1, X_2, \cdots, X_\tau \sim \text{iid } F(x),$$
$$X_{\tau+1}, X_{\tau+2}, \cdots, X_n \sim \text{iid } G(x)$$

여기서, $F(x)$와 $G(x)$ 에 대한 분포함수의 특별한 가정 없이 단지

$\phi = \{x \in R : |F(x) - G(x)| > 0$ 에서 $\int_{\phi} dF(x) > 0$ 또는 $\int_{\phi} dG(x) > 0$ 가정만을 기본조건으로 한다. $t \in \Lambda = \{i/n : 1 \le i \le n-1\}$에 대해 t시점 이전의 경험누적함수(pre-t empirical cdf) ${}_t h(x)$ 와 t시점 이후의 경험누적함수(post-t empirical cdf) $h_t(x)$는 다음과 같다.

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

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

여기서 지시함수(indicator function)

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

이다. t시점 이전, 이후의 경험누적 함수를 이용하여 Carlstein(1988)은 3가지 거리 기준을 고려하여 다음의 변화점 통계량을 제안하였다.

(1) 차이의 절대값의 합을 이용한 통계량

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

변화점 추정통계량은 $T_{C1} = \arg\max_{1<t<n} \{D_1(t)\}$이다.

(2) 차이의 제곱합을 이용한 통계량

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

변화점 추정통계량은 $T_{C2} = \arg\max_{1<t<n} \{D_2(t)\}$이다.

(3) 차이의 최대값을 이용한 통계량

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

변화점 추정통계량 $T_{C3} = \arg\max_{1<t<n} \{D_{3(t)}\}$으로 제안하였다.

3. 점수함수를 이용한 변화점 추정통계량 제안

독립적인 확률변수 $\{X_1, X_2, \ldots, X_n\}$는 평균의 변화가 한 번 발생하는 경우 다음의
변화점모형을 갖는다.

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

여기서 $\delta = \mu_2 - \mu_1$ 으로 0이 아닌 상수이고 $(\mu_1 \neq \mu_2)$, 오차항 ε_i는 서로 독립이며 평균이 0,분산이 σ^2인 동일한 연속분포를 따르며 τ는 평균의 변화가 일어나는 변화점이다. $\{X_1, X_2, \ldots, X_n\}$은 순위 $\{r_1, r_2, \ldots, r_n\}$을 가지며 r_i는 $\{X_1, X_2, \ldots, X_n\}$ 중에서의 순위이다. 순위를 이용한 점수함수(score function)로서 다음의 로그함수 $\phi(t) = \log(t+1)$, $t = r_i/(n+1)$ 이용을 제안하고자 하며 순위함수 이용시 등간격인

점수차를 log함수를 통해 점수(score)간의 차이를 등간격이 아니면서 점수간의 차이를 더 크게 할 수 있게된다. $a(r_i)=\phi\left(\frac{r_i}{(n+1)}\right)$로 놓으면 $\boldsymbol{a}=(a(r_1),\ldots,a(r_n))$은 순위를 이용한 점수벡터(a vector of score)이며 부분합(partial sum)은 $S_k=\sum_{j=1}^{k}(a(r_i)-\overline{a_n}),\quad k=1,\ldots,n$ 으로 표현하며 여기서 $\overline{a_n}=1/n\sum_{j=1}^{n}a(j)$ 이다.

Kim과 Seo(2001)에서는 순위에 기초하여 log 점수함수를 이용한 변화점 추정통계량으로

$$T_1=\arg\max_{1<k<n}|S_k|$$

를 제안하였다. 또한, 가중함수를 이용하면

$$T_2=\arg\max_{1<k<n}\sqrt{\frac{n}{k(n-k)}}|S_k|$$

를 제안할 수 있고 Gombay 와 Huskova(1998)의 정리에 의하면 T_1과 T_2는 같은 분포로 근사됨을 알 수 있다.

위의 log함수가 점수함수의 조건을 만족시키므로 log 점수함수를 이용해 Lombard(1987)의 추정방법을 적용시키고자 한다. 다음의

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

를 계산하고 순위통계량으로

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

를 계산하고 표준화하여

$$\widetilde{v_{i,i+1}}=v_{i,i+1}/\sigma(i/n,(i+1/n))$$

를 얻는다. $\sigma^2(u,v)=(1-u)^3(1+3u)/12-(1-v)^{3(1+3v)/1}2-(1-v)^2(v^2-u^2)/2$ 이다.

변화점 추정 통계량으로 log 점수함수를 이용해 다음의

$$T_{LL}=\arg\max_{1<i<n}\{|\widetilde{v_{i,i+1}}|\}$$

을 비교해보고자 한다. 또한, 분산추정량을 사용하지 않고

$$T_3=\arg\max_{1<i<n}\{|v_{i,i+1}|\}$$

로 변화점을 추정하고자 한다.

제안하는 통계량은 Gombay 와 Huskova(1998)의 가정을 만족하므로 Gombay와 Huskova(1998)의 정리가 마찬가지로 성립하게되며 다음의 정리들로 표현할 수 있다.

정리1. $(X_1,\cdots X_\tau)$과 $(X_{\tau+1},\cdots X_n)$는 각각 독립변수로 F와 G인 연속분포함수를 갖는다고 하자. $d_n=n^{-\alpha}$, $\alpha<1/2$ 와 $\frac{n}{\tau(n-\tau)d_n}S_\tau\to b$, $b\neq 0$ 이고 $n\to\infty$ 일 때, $i=1,2,3$

$$\frac{b^2 d_n^2}{\sigma_n^2(\boldsymbol{a})}(T_i - \tau) \rightarrow \min\{z \in R^1; \max\{W(t) - |t|g(t),\ t \in R^1\} = W(z) - |z|g(z)\},$$

여기서

$$g_1(t) = \begin{cases} 1-\theta, & t<0 \\ \theta, & t>0, \end{cases}$$

$$g_2(t) = \frac{1}{2}, \quad t \in R^1$$

그리고

$$W(t) = \begin{cases} W_1(-t), & t<0 \\ W_2(t), & t>0, \end{cases}$$

여기서 $\{W_1(t), , t>0\}$, $\{W_2(t),\ t>0\}$은 서로 독립인 Wiener Process이다.

정리2. $X_1, \cdots X_n$ 은 임의의 변수로 분포가 F인 연속분포함수를 갖는다. 그리고 점수 $a(1), \cdots, a(n)$이고 $n \rightarrow \infty$ 일 때

$$\frac{T_i}{n} \rightarrow \min\{t \in (0,1);\ |B(t)| = \max_{0 \le v \le 1} |B(t)|\}, \quad i = 1,2,3$$

이고, $\{B(v),\ 0 \le v \le 1\}$은 Brownian bridge를 나타낸다.

4. 모의실험

변화점 추정능력을 S-plus를 이용한 모의실험을 통해 비교하고자 한다. 모의실험에서 변화점이 1개 존재하는 다음의 변화점모형을 사용한다.

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

여기서 $\mu_1 = 0$ 에 대해 평균의 변화량을 고려하여 $\mu_2 = 1,2,3$ 의 경우에 대해 실험한다. 오차항 ε_i 는 평균이 0, 분산이 1인 정규분포, 이중지수분포를 따르며, 균일분포의 경우에는 (-1.7, 1.7) 구간에서 동일한 확률값을 갖는 경우가 된다. 표본의 크기 $n = 100$에 대해 변화점 $\tau = 50, 30$인 경우, $\mu_1 = 0$, $\mu_2 = 1,2,3$ 에 대하여 1000번의 반복실험으로 변화점 추정통계량의 움직임을 관찰하고자한다. 변화점 추정시 자료의 매우 앞부분이나 매우 뒷부분에서 변화점에 대한 발생가능성을 제외하고자 $\tau = 10, \cdots, 90$ 범위에서의 변화점 추정을 고려하였다. 변화점 추정능력을 비교하기위한 통계량으로 반복실험에서의 평균, 평균제곱오차(MSE), 변화점추정비율, 변화점에 대한 95 % 신뢰구간을 계산한 것이다. 여기서 변화점 추정비율은 주어진 변화점을 ± 1이내의 오차로 추정한 비율, 즉 $P(|\hat{\tau} - \tau| < 1)$에 대해 모의실험을 통한 경험적 확률을 계산한 것이다.

표 1, 2, 3에서 보면 $\tau = 50$ 일 때, 즉 변화점이 자료의 중간에 발생할 때, 변화점 추정통계량 T_L 과 T_{LL}의 평균제곱오차와 변화점추정비율은 비슷하나 T_{LL}의 평

균제곱오차가 큰 편이지만 변화점이 자료의 중간보다 앞부분에서 발생할 때 즉 $\tau = 30$ 일 때는 T_{LL}의 평균제곱오차가 더 적게 나타난다. 또한 $\tau = 50$ 일 때, T_1과 T_3의 평균제곱오차가 비슷하며 T_2 보다 월등히 적다. 그러나 $\tau = 30$ 일 때는 T_2의 평균제곱오차가 T_1이나 T_3 보다 현저히 적은 것을 알 수 있다. 정규분포, 이중지수분포, 균일분포의 경우 비슷한 경향을 보여주며 변화점의 발생위치에 따라 추정능력에 차이가 남을 알 수 있으며 변화점의 위치에 대한 가중값이 고려된 T_2는 변화점이 중간이 아닌 부분에서 발생했을 때 더 좋은 추정능력을 보여주고 있다.

5. 결론 및 제안

변화점 추정문제는 최근 여러 분야에서 대두되고 있는 문제로, 데이터가 동일한 분포에서 나오지 않을 경우 분포의 변화가 일어나는 시점을 추정한다. 이번 연구에서는 자료내에 위치모수의 변화가 한 번 일어날 경우 순위와 log 점수함수를 이용하여 Lombard(1987)의 통계량과 Gombay 와 Huskova(1998)의 통계량의 변형된 형태를 고려하고 기존의 통계량과 비교를 한 것이다. 특히, Gombay 와 Huskova(1998)의 변화점 추정량집합에 속한 추정량을 찾으면 극한분포를 밝힐 수 있어 실제 추정시에 편리하다. 다른 모형에서의 Lombard(1987)의 통계량과 Gombay 와 Huskova(1998)의 통계량의 변형된 형태가 가능하다고 여겨지며 점수함수를 이용한 방법이 변화점 추정에 기여함을 보여준다.

참고문헌

1. Carlstein, E. (1988). Nonparametric Changepoint Estimation. *The Annals of Statistics*, 16, 188-197.
2. Darkhovsh, B. S. (1976). A Non-parametric Method for the a Posteriori Detection of the "Disorder" Time of a Sequence of Independent Random Variables. *Theory of Probability and Its Applications*, 21, 178-183
3. Gombay, E. and Huskova, M. (1998). Rank based Estimators of the Change-point. *Journal of Statistical Planning and Inference*, 67, 137-154.
4. Hawkins, D. L. (1986). A simple Least Squared Method for Estimating a Change Point in Mean. *Communications in Statistics*, 15, 655-679.
5. Hinkley, D. V. (1970). Inference about the Change-point in a Sequence of Random Variables. *Biometrika*, 57, 1-16.
6. Lombard, F. (1987). Rank Tests for Changepoint Problems. *Biometrika*, 74, 615-624.
7. Pettit, A. N. (1979). A Nonparametic Approach to the Change Problem. *Applied Statistics*, 28, 126-135.
8. Schechtman, E. (1982). A Nonparametric Test for Detecting Changes in Location. *Communications in Statistics*, 11, 1475-1482.

표 1. $n = 100$, **변화점** $\tau = 50$, $\tau = 30$**일 때 오차항이** $N(0,1)$ **을 따르는 경우 1000번 반복실험에 의한 변화점 추정량 비교**

		$\tau = 50$				$\tau = 30$			
		평균	평균 제곱 오차	추정 비율	95% 신뢰 구간	평균	평균 제곱 오차	추정 비율	95% 신뢰 구간
$\mu_0 = 0$ $\mu_1 = 1$	T_H	50.54	37.724	0.483	(37,66)	30.70	55.706	0.486	(19,52)
	T_S	50.58	39.294	0.474	(36,66)	31.33	66.087	0.484	(19,54)
	T_{C1}	49.37	56.051	0.462	(19,63)	29.47	52.394	0.465	(14,44)
	T_{C2}	49.46	57.079	0.456	(29,63)	29.59	54.050	0.464	(14,46)
	T_L	49.46	41.777	0.425	(34,65)	30.17	62.731	0.441	(18,53)
	T_{LL}	48.95	45.516	0.424	(32,64)	29.61	55.150	0.431	(17,50)
	T_1	50.15	18.568	0.522	(41,60)	33.21	55.674	0.459	(26,53)
	T_2	50.08	43.320	0.483	(33,66)	30.73	54.982	0.485	(19,51)
	T_3	50.46	18.146	0.515	(41,60)	33.66	61.846	0.449	(26,54)
$\mu_0 = 0$ $\mu_1 = 2$	T_H	49.98	1.955	0.839	(47,53)	30.02	2.001	0.863	(27,33)
	T_S	49.99	1.821	0.851	(47,53)	30.26	2.108	0.833	(27,34)
	T_{C1}	49.89	2.083	0.847	(47,53)	30.09	2.004	0.848	(27,33)
	T_{C2}	49.92	2.142	0.842	(47,53)	30.05	2.060	0.844	(27,33)
	T_L	48.96	2.981	0.767	(46,52)	29.27	2.064	0.829	(26,33)
	T_{LL}	48.77	3.521	0.737	(45,51)	29.09	2.768	0.803	(26,32)
	T_1	49.81	1.547	0.861	(47,52)	31.08	6.672	0.746	(29,37)
	T_2	49.79	2.151	0.839	(46,52)	30.13	1.954	0.852	(27,33)
	T_3	50.00	1.414	0.869	(47,53)	31.34	8.717	0.721	(29,39)
$\mu_0 = 0$ $\mu_1 = 3$	T_H	49.99	0.271	0.974	(49,51)	29.97	0.320	0.974	(29,31)
	T_S	49.99	0.322	0.968	(49,51)	30.24	0.640	0.940	(29,32)
	T_{C1}	49.95	0.332	0.964	(49,51)	30.17	0.510	0.954	(29,32)
	T_{C2}	49.97	0.331	0.964	(49,51)	30.11	0.443	0.961	(29,32)
	T_L	48.98	1.373	0.907	(48,50)	29.23	1.166	0.944	(28,31)
	T_{LL}	48.87	1.616	0.875	(47,50)	29.11	1.163	0.941	(28,31)
	T_1	49.89	0.299	0.968	(48,51)	30.64	2.102	0.847	(30,35)
	T_2	49.89	0.359	0.962	(48,51)	30.14	0.461	0.959	(29,32)
	T_3	49.89	0.285	0.972	(49,51)	30.81	3.034	0.817	(30,35)

표2. $n = 100$, **변화점** $\tau = 50, \tau = 30$**일 때 오차항이** $DE(0,1)$ **을 따르는 경우 1000번 반복실험에 의한 변화점 추정량 비교**

		$\tau = 50$				$\tau = 30$			
		평균	평균 제곱 오차	추정 비율	95% 신뢰 구간	평균	평균 제곱 오차	추정 비율	95% 신뢰 구간
$\mu_0 = 0$ $\mu_1 = 1$	T_H	49.96	40.266	0.512	(34,65)	30.50	43.503	0.533	(18,46)
	T_S	49.90	21.149	0.573	(41,61)	30.50	27.747	0.564	(21,42)
	T_{C1}	49.17	31.123	0.560	(38,58)	29.32	28.238	0.546	(15,39)
	T_{C2}	49.17	31.123	0.560	(38,58)	29.32	28.238	0.546	(15,39)
	T_L	48.81	21.906	0.524	(39,58)	29.38	25.774	0.537	(19,41)
	T_{LL}	48.62	24.186	0.504	(39,58)	29.13	24.871	0.543	(18,39)
	T_1	49.77	12.546	0.600	(42,57)	32.26	31.832	0.532	(26,48)
	T_2	49.75	22.649	0.563	(40,61)	30.30	24.069	0.580	(21,41)
	T_3	49.92	12.755	0.601	(42,58)	32.58	36.834	0.510	(26,48)
$\mu_0 = 0$ $\mu_1 = 2$	T_H	50.06	1.702	0.884	(48,53)	29.99	1.471	0.866	(27,33)
	T_S	50.06	1.528	0.888	(48,52)	30.24	1.854	0.857	(28,33)
	T_{C1}	49.99	1.353	0.890	(48,52)	30.05	1.627	0.865	(27,33)
	T_{C2}	49.98	1.228	0.886	(47,52)	30.03	1.228	0.872	(28,33)
	T_L	49.05	2.417	0.836	(47,51)	29.21	2.268	0.836	(27,32)
	T_{LL}	48.95	2.631	0.803	(46,51)	29.07	2.369	0.833	(27,32)
	T_1	49.96	1.298	0.888	(48,52)	31.01	7.006	0.766	(29,38)
	T_2	49.96	1.521	0.880	(47,52)	30.12	1.299	0.872	(28,33)
	T_3	50.02	1.174	0.896	(48,52)	31.12	8.359	0.731	(29,39)
$\mu_0 = 0$ $\mu_1 = 3$	T_H	49.99	0.273	0.970	(49,51)	30.01	0.268	0.968	(29,31)
	T_S	49.99	0.310	0.964	(49,51)	30.21	0.519	0.947	(29,32)
	T_{C1}	49.95	0.326	0.964	(49,51)	30.15	0.463	0.952	(29,32)
	T_{C2}	49.96	0.307	0.966	(49,51)	30.10	0.354	0.958	(29,31)
	T_L	49.98	1.304	0.919	(48,50)	29.20	1.112	0.959	(28,31)
	T_{LL}	48.89	0.508	0.882	(47,50)	29.13	1.128	0.956	(28,31)
	T_1	49.90	0.333	0.964	(48,51)	30.65	2.544	0.853	(30,34)
	T_2	49.90	0.360	0.958	(48,51)	30.14	0.405	0.955	(29,32)
	T_3	49.99	0.296	0.967	(49,51)	30.83	3.638	0.812	(30,36)

표3. $n = 100$, **변화점** $\tau = 50$, $\tau = 30$**일 때 오차항이** $U(-1.7, 1.7)$ **를 따르는 경우 1000번 반복실험에 의한 변화점 추정량 비교**

		$\tau = 50$				$\tau = 30$			
		평균	평균 제곱 오차	추정 비율	95% 신뢰 구간	평균	평균 제곱 오차	추정 비율	95% 신뢰 구간
$\mu_0 = 0$ $\mu_1 = 1$	T_H	49.90	40.024	0.472	(34,64)	30.86	52.957	0.483	(18,50)
	T_S	49.58	52.255	0.448	(31,64)	31.78	85.383	0.448	(18,59)
	T_{C1}	48.14	73.069	0.438	(22,62)	29.64	70.954	0.440	(13,46)
	T_{C2}	48.16	87.275	0.415	(20,64)	29.70	77.935	0.427	(12,48)
	T_L	48.38	57.455	0.401	(30,63)	30.59	80.540	0.415	(17,55)
	T_{LL}	47.69	58.925	0.385	(28,61)	29.81	68.555	0.402	(16,53)
	T_1	49.37	22.553	0.491	(38,59)	33.24	67.895	0.426	(24,53)
	T_2	48.87	54.018	0.447	(29,63)	30.95	61.048	0.459	(17,54)
	T_3	49.79	21.181	0.484	(39,60)	33.80	70.896	0.413	(25,55)
$\mu_0 = 0$ $\mu_1 = 2$	T_H	50.02	1.768	0.854	(47,53)	30.00	1.969	0.837	(27,33)
	T_S	50.04	1.967	0.855	(47,53)	30.37	3.064	0.824	(28,35)
	T_{C1}	49.92	2.191	0.853	(47,53)	30.12	2.332	0.817	(27,34)
	T_{C2}	49.92	2.391	0.846	(47,53)	30.11	2.851	0.810	(27,34)
	T_L	49.03	2.893	0.763	(46,52)	29.32	3.217	0.771	(26,33)
	T_{LL}	48.80	3.566	0.734	(46,51)	29.12	3.260	0.742	(26,33)
	T_1	49.85	1.457	0.867	(47,52)	31.13	7.312	0.752	(29,38)
	T_2	49.82	2.110	0.850	(47,52)	30.15	2.516	0.891	(27,34)
	T_3	50.03	1.342	0.866	(48,52)	31.42	9.360	0.713	(29,39)
$\mu_0 = 0$ $\mu_1 = 3$	T_H	50.00	0.146	0.992	(49,51)	29.99	0.193	0.985	(29,31)
	T_S	50.00	0.145	0.991	(49,51)	30.22	0.424	0.960	(30,32)
	T_{C1}	49.98	0.156	0.991	(49,51)	30.15	0.252	0.975	(30,31)
	T_{C2}	49.98	0.164	0.990	(49,51)	30.11	0.245	0.979	(30,31)
	T_L	48.99	1.166	0.940	(48,50)	29.20	0.953	0.983	(29,31)
	T_{LL}	48.91	1.353	0.913	(48,49)	29.12	0.911	0.974	(29,30)
	T_1	49.92	0.165	0.986	(49,50)	30.57	1.597	0.861	(30,34)
	T_2	49.91	0.187	0.981	(49,50)	30.13	0.242	0.978	(30,31)
	T_3	49.99	0.139	0.993	(49,51)	30.78	2.459	0.812	(30,35)