Journal of Korean
Data & Information Science Society
2004, Vol. 15, No. 1 pp. 107~117

Application of Bootstrap Method for Change Point Test based on Kernel Density Estimator

Daehak Kim[1)]

Abstract

Change point testing problem is considered. Kernel density estimators are used for constructing proposed change point test statistics. The proposed method can be used to the hypothesis testing of not only parameter change but also distributional change. Bootstrap method is applied to get the sampling distribution of proposed test statistic. Small sample Monte Carlo Simulation were also conducted in order to show the performance of proposed method.

Keywords : 변화점 검정, 분포의 변화, 붓스트랩방법, 제1종 오류, 커널 확률밀도함수추정량, AMOC

1. 서 론

시간에 따라 미지의 모집단으로부터 연속적으로 얻어진 랜덤표본 $X_1, X_2, \cdots, X_n$에 대하여

$$X_1, \cdots, X_\tau \sim F, \quad X_{\tau+1}, \cdots, X_n \sim G \tag{1.1}$$

로 두자. 즉 τ시점까지의 랜덤표본 $X_1, \cdots, X_\tau$은 알려지지 않은 연속인 분포 F를 따르고 $\tau+1$ 시점부터 끝까지의 랜덤표본 $X_{\tau+1}, \cdots, X_n$은 마찬가지로 알려지지 않은 연속인 분포 G를 따른다고 두자. 이때 F와 G는 독립이고 미지의 시점 τ는 변화점(Change point)이라 불리운다. 이러한 형태의 자료에 대하여 모집단을 대표하는 모수의 변화나 모집단의 분포의 변화를 발견할 수 있는 경우가 자주 발생한다. 예를 들면 10년 전의 초등학생들의 체격과 오늘날의 초등학생들의 체격사이에는 여러 가지 원인으로 인한 변화 즉 모집단의 변화를 가정해 봄은 그리 어려운 일은 아니다. 이와 같이 주어진 자료 속에 모수 혹은 모집단과 관련된 변화가 있었는지에 관심을 두고

1) 경북 경산시 하양읍 금락리 330번지 대구가톨릭대학교 정보통계학전공 교수
Email : dhkim@cu.ac.kr

변화가 존재하는지 또 있다면 언제 발생했는지 등에 관해 통계적 추론을 하는 분야는 변화점 문제(change point problems)로 잘 알려져 있으며 많은 사람들에 의해 연구가 진행되어 왔다. Bhattacharyya와 Jhonson(1968) 은 $G(x)=F(x-\Delta)$와 같은 형태의 대립가설에 대하여 위치모수(location parameter)의 변화에 대한 연구를 하였고 Sen과 Srivastava (1975), Pettitt(1979) 그리고 Wolfe(1984)등에 의해 많은 연구가 이루어져 왔다. 또한 $G(x)=F(x/\Delta)$형태의 대립가설은 측도모수(scale parameter)의 변화에 대한 표현으로서 Hsu(1977)등에 의해 분산의 변화에 대한 연구가 시도되었다. 또한 변화점 위치의 추정문제에 대해서는 Schechtman(1983)등이 연구 한 바 있다. 이러한 변화점 문제의 해결을 위해 대부분의 경우 순위통계량(rank statistics)을 이용한 연구가 이루어져 왔으며 대부분의 경우 현실적 문제 해결에 응용되어 왔다. 특히 지금까지의 연구 결과들은 모평균의 변화나 모분산의 변화등에 초점을 두고 연구가 진행되었으며 평균과 분산 즉 위치모수와 척도모수의 동시변화가능성이나 모집단의 다른 분포로의 변화에 대한 연구가 그리 많지 않은 실정이다.

한편 최근 들어 컴퓨터의 급속한 발달에 기인한 초고속의 계산 능력을 이용한 많은 실질적인 연구가 이루어지고 있다. Efron(1979)에 의해 소개된 붓스트랩 방법을 이용하여 복잡하고도 수리적으로 증명하기 힘든 통계량의 표본분포를 구할 수 있다. 붓스트랩 방법은 다방면에로의 응용이 가능하여 많이 활용되고 있다. 통계적 방법으로서의 비모수적 방법에서도 많은 변화가 이루어지고 있다. 도수(frequency)나 부호(sign) 그리고 순위(rank)를 이용하는 통계량에서부터 최근에는 커널함수를 이용한 추정량들이 각광을 받고 있으며 특히 확률밀도함수나 회귀함수의 추정에 커널함수를 이용한 비모수적 방법이 많이 사용되고 있다.

본 논문에서는 변화점 문제에 있어서 변화점의 새로운 통계적 검정 방법을 제안하고자 한다. 평균이나 분산등의 변화 뿐 만 아니라 위치모수와 척도모수의 동시변화가능성이나 모집단의 다른 분포로의 변화에 대한 검정이 가능한 검정방법을 제안하고자 한다. 그 방법으로서 커널 확률밀도함수 추정량으로 구성되는 새로운 통계량을 제안하고 붓스트랩 방법을 활용하여 제안된 통계량의 귀무가설 하에서의 분포를 유도하고자 한다. 또한 소표본의 경우 모의실험을 통하여 제안된 방법의 효용성을 밝히고자 한다.

2. 변화점 검정통계량의 제안과 붓스트랩 분포

본 절에서는 변화점의 존재 여부에 대한 통계적 가설검정을 위하여 검정통계량을 제안하고자 한다. 이를 위하여 확률밀도함수의 커널추정방법을 간단히 소개하고 커널 확률밀도함수 추정량을 이용한 두 분포 사이의 거리를 재는 기준을 소개하고 이를 이용한 검정통계량을 제안한다. 제안된 검정통계량의 붓스트랩 분포에 대하여도 설명하였다.

2.1 변화점의 가설 검정

변화점의 여러 가지 문제 가운데에서 본 논문에서는 가장 근본적이면서도 핵심적인 문제인 변화시점의 존재여부에 초점을 맞추고자 한다. 즉 변화점이 존재하는가에 대

한 통계적 가설검정 문제를 고려하였다. 이때 기껏해야 하나의 변화점이 존재하는 AMOC(At Most One Change)모형을 고려하고 귀무가설과 대립가설을 미지의 변화점 τ를 중심으로 다음과 같이 표현하였다.

$$H_0 : F = G \text{ vs } H_1 : F \neq G \tag{2.1}$$

모수의 변화와 관련된 가설

$$\begin{aligned} H_0 &: \mu_1 = \cdots = \mu_\tau = \mu_{\tau+1} = \cdots = \mu_n \\ H_1 &: \mu_1 = \cdots = \mu_\tau \neq \mu_{\tau+1} = \cdots = \mu_n \end{aligned} \tag{2.2}$$

에 대한 검정 문제는 본 논문에서 고려하는 가설 검정문제에 포함되고 있음을 알 수 있다. 특히 분포의 변화와 관련된 문제의 경우 기존의 방법으로 해결될 수 없음도 가설의 표현에서 명백해 짐을 알 수 있다. τ가 알려져 있으면 이는 이표본 문제(Two sample Problem)가 되고 이 경우에는 이표본 T-검정법 또는 콜모고로프-스미르노프 검정방법 등에 의해 분석되어질 수 있다.

2.2 커널추정량의 활용과 두분포 사이의 거리

랜덤표본 $Y_1, Y_2, \cdots, Y_n$ 에 대하여 미지의 확률밀도함수 $f(y)$의 커널추정량은 식 (2.3)과 같다.

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h}\right) \tag{2.3}$$

이때 h는 추정된 곡선에 영향을 가장 많이 미치는 평활계수(smoothing parameter)이고 $K(\cdot)$는 확률밀도함수 조건을 만족하는 커널함수이다. 확률밀도함수의 커널추정에 있어서 중요한 사실은 $\hat{f}$의 편의와 분산은 평활계수 h의 함수라는 것이다. 즉 평활계수 h를 작게 하면 편의는 줄어드나 분산이 증가하게 된다. 평활계수 h는 커널확률밀도함수 추정의 전체적인 모양을 결정하는 중요한 역할을 하는데 평활계수 h가 작고 큼에 따라 확률밀도함수 추정치의 모양이 거칠어지고 평평하게 된다. Silverman(1989)은 확률밀도함수의 커널추정에 관한 다양한 방법과 자료에 근거한 평활계수의 선택을 소개 하였다.

이제 커널추정량을 이용하여 추정된 확률밀도함수들을 이용하여 두 분포의 차이를 고려하자. 두 분포가 동일하면 즉 $F = G$가 의미하는 것은 임의의 공동 표본 공간의 어떤 부분 집합 A에 대하여 $P_F\{A\} = P_G\{A\}$임을 나타낸다. 이것은 두 분포 F와 G사이의 거리의 추정량으로서 $\sup_A | P_F\{A\} - P_G\{A\}|$을 사용할 수 있음을 암시한다. 다행히 Devroye와 Györfi(1985)에 의하면 임의의 확률밀도함수 f와 g에 대하여 다음과 같이 쓸 수 있음을 알 수 있다.

$$L_1(f,g) = \int |f - g| = 2 \sup_A \left| \int_A f - \int_A g \right| = 2 \sup_A |P_F\{A\} - P_G\{A\}| \tag{2.4}$$

식 (2.4)의 $L_1(f,g)$을 두 확률밀도함수 사이의 거리(distance)로 고려할 수 있고 두

분포 F와 G 사이의 거리의 개념으로 자연스럽게 해석할 수 있다. 서로 독립인 랜덤 표본 $Y_1, Y_2, \cdots, Y_n$과 $Z_1, Z_2, \cdots, Z_m$에 대하여서도 각각의 커널확률밀도함수 추정량 $\hat{f}$ 와 $\hat{g}$을 사용하여 두 분포사이의 거리를 추정할 수 있다. 다음 식 (2.5)와 같이 추정된 거리 $\hat{L}_1$가 큰 값을 가지면 이는 두 분포가 동일하지 않음을 나타내는 증거가 될 것이다.

$$\hat{L}_1 = L_1(\hat{f}, \hat{g}) = \int |\hat{f} - \hat{g}| \tag{2.5}$$

그러나 이 통계량의 표본분포를 폐쇄형으로 표현할 수 없다. 다행히 붓스트랩 방법을 사용하여 $\hat{L}_1$의 표본분포를 추정할수 있게 되어 다음과 같은 몬테칼로 근사를 통하여 두 독립표본에 대한 분포의 동일성에 관한 가설을 검정할 수 있다. 먼저 주어진 표본을 합하고 합쳐진 표본으로부터 n개와 m개로 구성된 재표본을 형성한다. 즉 $Y^*_1, Y^*_2, \cdots, Y^*_n$과 $Z^*_1, Z^*_2, \cdots, Z^*_m$를 얻는다. 그 다음 재표본된 각 자료로부터 커널확률밀도함수 추정량을 구한다. 이를 $\hat{f}^*$와 $\hat{g}^*$로 두자. 그 다음 $\hat{L}_1$ 통계량을 $\hat{f}^*$와 $\hat{g}^*$를 이용하여 구한다. 이를

$$\hat{L}_1^* = L_1(\hat{f}^*, \hat{g}^*) = \int |\hat{f}^* - \hat{g}^*| \tag{2.6}$$

로 표기하자. 위의 과정을 독립적으로 B회 독립적으로 반복하여 $\hat{L}_1$의 표본분포의 백분위수를 얻는다. 검정의 절차는 위의 과정을 통하여 얻어진 백분위수의 값이 원래 표본으로부터 계산된 $\hat{L}_1$값보다 작으면 두 분포가 동일하다는 귀무가설을 기각한다.

2.3 검정통계량의 제안과 붓스트랩 분포

이제 변화점의 가설검정을 위하여 커널 추정량을 이용한 통계량의 제안과 붓스트랩 방법을 이용하여 제안된 통계량의 분포를 구하고 이를 토대로 변화점의 가설검정 문제를 시도하고자 한다. 랜덤표본 $X_1, X_2, \cdots, X_n$에서 만약 위치 t 에서 변화가 있었다면 통계량

$$D_t = \int |\hat{f}_t - \hat{g}_t| \tag{2.7}$$

의 값은 다른 어느 위치에서의 값보다 클 것이다. 이때 $\hat{f}_t$ 는 처음부터 $t-1$시점까지의 자료로부터 계산된 커널추정량이고 $\hat{g}_t$ 는 t 시점부터 끝까지의 자료로부터 계산된 커널추정량이다. 이런 중심생각에 착안하여 다음과 같은 검정통계량을 제안한다.

$$\text{검정 통계량 } \hat{D} = \max_{2 \le t \le n} D_t \tag{2.8}$$

이때 $\widehat{D}$ 의 값이 클수록 귀무가설을 기각하게 된다. 그러나 이 통계량 $\widehat{D}$의 귀무가설 하에서의 분포를 폐쇄형으로 표현할 수 없게 되어 얼마나 큰 값을 가져야 기각할 수 있는지를 알 수 없다. 다행히 붓스트랩 방법을 사용하여 $\widehat{D}$의 표본분포를 추정할 수 있게되어 다음과 같은 몬테칼로 근사를 통하여 표본분포를 추정할수 있다.

변화점의 존재에 관련된 가설검정을 위하여 다음과 같이 붓스트랩 방법을 적용하여 귀무가설 하에서의 검정통계량 $\widehat{D}$의 붓스트랩 분포를 유도할 수 있다. 이를 위한 알고리즘은 다음과 같다.

[단계1] $X_1, X_2, \cdots X_n$ 로부터 복원추출을 통한 붓스트랩 표본 $X_1^*, X_2^*, \cdots X_n^*$ 을 얻고 붓스트랩 표본으로부터

$$\widehat{D}^* = \max_{2 \le t \le n} D_t^* = \max_t \int |\hat{f}_t^* - \hat{g}_t^*| \qquad (2.10)$$

를 계산한다.

[단계2] [단계1]을 B번의 붓스트랩 반복을 통한 B 개의 $\widehat{D}^*$ 를 얻고 이를 통해 $\widehat{D} = \max_{2 \le t \le n} D_t$ 의 붓스트랩 분포를 유도한다.

[단계3] 붓스트랩 분포의 백분위수를 이용하여 표본에서의 $\widehat{D} = \max_{2 \le t \le n} D_t$ 값이 붓스트랩 분포의 기각역에 속하면 귀무가설을 기각한다. 즉 $\widehat{D}^* \ge \widehat{D} + 1 \le \alpha \times (B+1)$이 성립되면 귀무가설을 기각한다.

3. 모의실험

제안된 검정통계량의 활용을 위해 본 절에서는 모의실험의 설계와 그 결과를 제시하고자 한다. $\widehat{D} = \max_{2 \le t \le n} D_t$을 이용하여 변화점의 존재에 관한 가설 $H_0 : F = G$ vs $H_1 : F \ne G$ 에 대한 검정을 여러 가지 상황을 설정하여 모의실험을 실시하였다.

3.1 모의실험 설계

평균의 변화, 표준편차의 변화 그리고 평균과 표준편차의 동시변화에 따른 모의실험을 위하여 표본을 추출한 모집단은 정규분포를 고려하였다. 또한 분포의 변화에 따른 제안된 방법의 경험적 검정력을 구하기 위하여 지수분포, 정규분포 그리고 균일분포등을 고려하였다. 이들 세 분포는 평균과 분산이 동일한 값을 가지도록 하기 위하여 각각 지수분포 $E(1)$, 정규분포 $N(1,1)$ 그리고 균일분포 $U(-.732, 2.732)$를 사용하였다. 평균과 분산은 같지만 분포가 다른 경우의 결과는 주목할 만 할 것이다.

표본의 크기 n은 모수가 변화하는 경우는 20, 30, 40 그리고 50의 네 가지를, 분포

의 변화 경우에는 30 ,40, 50, 80, 100 그리고 200의 여섯가지를 고려하였다. 커널확률밀도 함수추정의 계산의 효용성을 위해 Epanechnikov(1969) 커널

$$K(x)=\frac{3}{4}(1-x^2) \quad \text{if} \quad |x|\le 1 \tag{3.1}$$

을 활용하였다. 평활계수의 선택문제는 확률밀도의 함수의 커널추정에 결정적인 영향을 초래하지만 모의실험의 방대한 계산을 효율적으로 하기 위해 사전 모의실험의 결과로부터 $h=2n^{-1/5}\hat{\sigma}$를 이용하였다. $\hat{\sigma}$는 주어진 표본에서의 추정된 표준편차를 의미한다. 붓스트랩 반복의 횟수는 충분하다고 고려되는 200을 고려하였고 모의실험을 위한 전체 반복횟수는 1000번을 실시하였다. 가설검정의 유의수준은 일반적으로 가장 많이 활용되는 α = 0.05를 고려하였다.

제안된 방법의 타당성을 조사하기 위하여 귀무가설 하에서의 제1종 오류도 계산되었다. 또한 변화점의 위치도 여러 가지를 고려하였다.

3.2 모의실험 결과

모의실험의 결과를 본 절에서 설명하고자 한다. 먼저 제안된 통계량의 타당성을 조사하였다. <표 1>은 표준정규분포로부터 표본의 크기가 40인 랜덤표본을 얻고 각 변화점의 위치(18에서부터 22까지)에서 1000번을 반복하여 제안된 통계량의 값이 최대값을 취한 위치의 빈도를 조사한 결과이다. <표 1>의 왼쪽의 경우는 변화가 없을 경우이고 오른쪽의 경우는 평균이 1로 변화되었을 때의 빈도를 나타내고 있다.

<표 1> 최대값의 빈도

변화여부	변화가 없는경우					평균이 변한 경우				
위치 / 참변화점	18	19	20	21	22	18	19	20	21	22
18	135	93	84	110	140	**373**	121	85	59	55
19	125	103	94	108	133	171	**337**	102	97	79
20	119	131	106	87	120	114	144	**313**	127	89
21	128	118	99	101	131	87	83	125	**333**	145
22	125	103	100	112	126	60	69	109	131	**331**

<표 1>에서 나타나듯이 변화가 없을 때에는 최대값을 취한 도수가 모든 위치에서 균일하게 분포되어 있음을 알 수 있다. 또한 변화가 있을 때에는 변화가 발생한 위치에서 빈도가 높아짐을 나타내고 있다. 변화량이 클수록 변화위치를 잘 찾고 있음도 나타난다.

이제 모의실험의 결과를 살펴보고자 한다. <표 2>부터 <표 4>까지는 4가지의 표본의 크기에 대하여 제안된 검정통계량의 경험적 검정력을 구한 결과이다.

<표 2>는 평균의 변화만을, <표 3>은 표준편차만의 변화가 있을 때이고 <표 4>는

평균과 표준편차의 동시변화가 있을 때의 결과이다. 변화점의 위치는 표본의 크기 n에 따라 바뀌도록 계획되었지만 표현의 제한으로 인하여 가운데를 중심으로 앞뒤로 표기하였다. 즉 20일 때는 7에서 14까지 또 표본의 크기가 40일 때 37에서 44 까지와 같이 나타내었다. <표 2>를 살펴보면 평균에 대한 변화량이 없을 때 즉 변화점이 없을 때의 경험적 검정력을 살펴보면 유의수준을 대체로 만족하고 있다. 즉 제 1종 오류가 주어진 유의수준의 근방에서 형성되고 있음을 알 수 있다. 변화량이 커질수록 또 표본의 크기가 커질수록 경험적 검정력은 증가하고 있으며 변화점의 위치에는 검정력의 변화가 거의 없음을 알 수 있다. <그림 1>은 변화점의 위치가 가운데에 있을 때의 평균의 변화량에 대한 표본의 크기에 따른 검정력을 도시한 결과이다.

<표 2> 평균의 변화에 따른 경험적 검정력
(H_0 : $N(0,1)$ *vs* H_1 : $N(\mu,1)$, $n_1 = n/2$)

변화량(μ)	n	변화점의 위치							
		n_1-3	n_1-2	n_1-1	n_1	n_1+1	n_1+2	n_1+3	n_1+4
0	20	0.047	0.050	0.041	0.043	0.040	0.040	0.041	0.051
	30	0.048	0.040	0.042	0.046	0.038	0.034	0.042	0.058
	40	0.047	0.051	0.047	0.046	0.042	0.046	0.048	0.044
	50	0.052	0.054	0.046	0.047	0.047	0.045	0.041	0.053
0.5	20	0.074	0.075	0.080	0.068	0.700	0.082	0.066	0.062
	30	0.120	0.106	0.096	0.094	0.098	0.104	0.108	0.090
	40	0.130	0.098	0.122	0.124	0.110	0.108	0.130	0.134
	50	0.156	0.148	0.156	0.160	0.166	0.134	0.122	0.130
1.0	20	0.210	0.228	0.188	0.172	0.212	0.206	0.224	0.198
	30	0.330	0.340	0.340	0.342	0.308	0.296	0.320	0.296
	40	0.420	0.452	0.450	0.438	0.434	0.446	0.444	0.452
	50	0.558	0.540	0.576	0.596	0.552	0.574	0.534	0.518
1.5	20	0.462	0.492	0.482	0.454	0.442	0.476	0.460	0.454
	30	0.712	0.688	0.716	0.694	0.708	0.708	0.690	0.678
	40	0.858	0.848	0.850	0.852	0.844	0.852	0.842	0.862
	50	0.946	0.926	0.912	0.912	0.928	0.938	0.938	0.962
2.0	20	0.720	0.7540	0.720	0.754	0.714	0.728	0.740	0.744
	30	0.946	0.936	0.912	0.934	0.920	0.938	0.912	0.926
	40	0.982	0.990	0.984	0.996	0.980	0.990	0.992	0.990
	50	0.996	1.000	1.000	0.996	0.998	0.998	0.998	0.998

<표 3> 표준편차의 변화에 따른 경험적 검정력

(H_0 : $N(0,1)$ *vs* H_1 : $N(0,\sigma^2)$, $n_1 = n/2$)

변화량(σ)	n	변화점의 위치							
		n_1-3	n_1-2	n_1-1	n_1	n_1+1	n_1+2	n_1+3	n_1+4
1.5	20	0.048	0.044	0.048	0.054	0.040	0.050	0.050	0.046
	30	0.051	0.053	0.048	0.059	0.062	0.054	0.071	0.057
	40	0.064	0.055	0.082	0.064	0.096	0.063	0.066	0.095
	50	0.068	0.068	0.084	0.098	0.094	0.094	0.074	0.092
2.0	20	0.042	0.045	0.048	0.064	0.074	0.086	0.066	0.106
	30	0.093	0.108	0.094	0.094	0.119	0.133	0.178	0.172
	40	0.112	0.143	0.184	0.196	0.181	0.242	0.207	0.223
	50	0.250	0.214	0.212	0.222	0.222	0.250	0.274	0.306
2.5	20	0.056	0.063	0.100	0.108	0.134	0.122	0.142	0.186
	30	0.161	0.174	0.184	0.238	0.253	0.252	0.310	0.299
	40	0.254	0.291	0.362	0.382	0.378	0.352	0.412	0.442
	50	0.420	0.426	0.472	0.492	0.465	0.521	0.460	0.528

<그림 1> 평균의 변화량에 대한 경험적 검정력

<표 3>은 표준편차의 변화에 대한 경험적 검정력을 나타내고 있다. 표본의 수가 작고 표준편차의 변화가 작은 경우에는 경험적 검정력이 아주 작게 나타나고 있다. 물론 변화량이 커지고 표본의 수가 증가할수록 경험적 검정력은 증가하는 경향을 나타내고 있으나 평균의 변화에 따른 검정력과는 상당한 차이를 보이고 있다. 표준편차의 변화는 자료들을 훨씬 희석시키기 때문에 추정된 분포의 적분값의 변화가 평균의 변화때 보다 덜 나타나기 때문으로 보인다.

<표 4> 평균, 표준편차의 동시변화에 따른 경험적 검정력
(H_0 : $N(0,1)$ *vs* H_1 : $N(\mu,\sigma^2)$, $n_1=n/2$)

변화량 (μ,σ)	n	변화점의 위치 n_1-3	n_1-2	n_1-1	n_1	n_1+1	n_1+2	n_1+3	n_1+4
(0.5, 1.5)	20	0.052	0.050	0.056	0.064	0.050	0.063	0.074	0.054
	30	0.068	0.071	0.070	0.098	0.096	0.102	0.126	0.093
	40	0.108	0.069	0.132	0.109	0.120	0.104	0.154	0.151
	50	0.122	0.132	0.159	0.149	0.158	0.154	0.142	0.156
(1.0, 2.0)	20	0.070	0.104	0.103	0.126	0.157	0.176	0.162	0.179
	30	0.171	0.206	0.241	0.242	0.266	0.276	0.301	0.280
	40	0.329	0.351	0.339	0.362	0.385	0.382	0.407	0.405
	50	0.438	0.413	0.452	0.446	0.442	0.482	0.488	0.508
(1.5, 2.5)	20	0.157	0.186	0.200	0.240	0.280	0.280	0.294	0.320
	30	0.338	0.347	0.388	0.401	0.466	0.479	0.523	0.482
	40	0.538	0.557	0.554	0.621	0.593	0.636	0.602	0.674
	50	0.718	0.711	0.716	0.716	0.758	0.774	0.787	0.778

<표 5> 모분포의 변화에 따른 경험적 검정력
(EXP= $E(1)$,NOR= $N(1,1)$,UNI= $U(-.732, 2.732)$, $n_1=n/2$)

분포의 변화	n	변화점의 위치 n_1-3	n_1-2	n_1-1	n_1	n_1+1	n_1+2	n_1+3	n_1+4
EXP vs EXP	30	0.038	0.052	0.062	0.062	0.042	0.022	0.032	0.044
	40	0.042	0.048	0.048	0.036	0.042	0.064	0.054	0.034
	50	0.052	0.038	0.046	0.046	0.050	0.042	0.042	0.054
	80	0.046	0.050	0.042	0.048	0.040	0.032	0.038	0.042
	100	0.042	0.060	0.070	0.050	0.070	0.054	0.054	0.052
	200	0.054	0.078	0.060	0.066	0.080	0.026	0.060	0.056
EXP vs NOR	30	0.086	0.064	0.078	0.078	0.084	0.076	0.090	0.104
	40	0.084	0.092	0.098	0.124	0.128	0.152	0.096	0.128
	50	0.128	0.152	0.144	0.162	0.172	0.152	0.154	0.154
	80	0.264	0.256	0.340	0.270	0.272	0.308	0.310	0.280
	100	0.352	0.360	0.384	0.374	0.398	0.366	0.404	0.414
	200	0.798	0.786	0.794	0.798	0.810	0.804	0.806	0.788
EXP vs UNI	30	0.092	0.086	0.090	0.084	0.118	0.108	0.112	0.104
	40	0.098	0.118	0.114	0.144	0.172	0.178	0.176	0.172
	50	0.170	0.190	0.184	0.208	0.224	0.214	0.228	0.204
	80	0.422	0.428	0.421	0.420	0.430	0.428	0.476	0.462
	100	0.584	0.600	0.598	0.580	0.610	0.584	0.634	0.618
	200	0.968	0.978	0.960	0.976	0.960	0.966	0.974	0.978

<표 4>는 평균과 표준편차의 동시변화에 대한 검정력을 나타내고 있다. 평균과 분산이 동시에 변화된 경우의 검정력은 표준편차만의 검정력보다 우위에 있지만 마찬가지로 평균만의 변화의 경우보다 상대적으로 약한 결과를 나타내고 있다. 이 또한 표준편차의 변화로 인한 데이터의 혼합에 기인한 것으로 보인다.

<표 5>는 6가지의 표본의 크기에 대하여 분포의 변화가 있을 때에 제안된 검정통계량의 경험적 검정력을 구한 결과이다. <표 5>로부터 모분포가 변화된 경우의 경험적 검정력을 살펴보면 모수의 변화가 있는 경우와 마찬가지로 제 1종오류는 주어진 유의수준을 대체로 만족하고 있는 것처럼 보인다. 표본의 크기가 작을 경우에는 분포의 변화를 감지할 정보가 적은 관계로 큰 검정력을 보이지는 못하고 있으나 표본의 크기가 커질수록 경험적 검정력은 전체적으로 증가하고 있는 경향을 나타내고 있으며 표본의 수가 80 정도부터는 급격하게 검정력이 증가함을 알수 있다. 대립가설의 분포가 정규분포일 때보다 균일분포일 때가 표본이 커질수록 검정력이 조금 더 높게 나타나고 있다. 변화점의 위치에 대한 검정력의 차이는 거의 없어 보인다.

4. 결론 및 토의

본 논문에서는 변화점의 검정문제를 고려하였다. 특히 평균이나 분산의 변화 뿐만 아니라 평균과 분산의 동시 변화가능성이나 분포 자체의 변화가 있는 경우에 변화점의 검정에 활용할 수 있는 방법을 제안하였다. 그 방법으로서 커널확률밀도 함수 추정량을 이용하여 새로운 검정통계량을 제안하였고 폐쇄형으로 표현하기 어려운 검정통계량의 분포는 붓스트랩 방법을 이용하였다. 소 표본의 모의실험의 결과 제안된 방법의 효용성을 확인할 수 있었다. 그러나 경험적 검정력의 차원에서 보면 좀 더 효율적인 방법이 기대되고 다중 변화점이 있는 경우로의 확장도 필요하다고 사료된다.

참고문헌

1. Bhattacharyya, G.K & Jhonson, R.A. (1968). Nonparametric tests for shift at an unknown time point., *Annals of Statistics.,* 39, 1731-1743
2. Devroye, L. and Györfi, 6ộ6ộ1985) *Nonparametric density estimation in* L_1 *view*, New York: Wiley.
3. Efron, B. (1979) Bootstrap Method; Another look at the jackknife, *Annals of Statistics.,* 7, 1-26
4. Epanechnikov, V.A. (1969). Nonparametric estimation of a multidimensional probability density. *Theory of Probability and its Application*, 14, 153-158
5. Hsu, D. A. (1977). Tests for variance shift at an unknown time point., *Annals of Statistics.,* 26, No. 3, 279-284
6. Pettitt, A.N. (1979). A nonparametric approach to the change point problems. *Applied Statistics,* 28, 126-135
7. Schechtman, E. (1983). A conservative nonparametric distribution free

confidence bound for the shift in the change point problem., *Communication in Statististics., Theory and Method*, A. 12(21), 2455-2464

8. Sen, A. & Srivastava, M.S. (1975). On tests for detecting change in Mean, *Annals of Statistics.,,* 3, No. 1, 98-108
9. Silverman, B.W. (1989). *Density Estimation for statistics and data analysis*., Chapman and Hall, London
10. Wolfe, D.A. (1984). Nonparametric Statistical Procedures for the change point problems., *Journal of Statistical Planning and Inference* 9, 389-396

[2004년 1월 접수, 2004년 2월 채택]