응용통계연구
제 4 권 제 1 호, 1991년

다중자료를 갖는 변화시점 모형에서의 비모수적인 검정법*

김 경 무**

〈요 약〉

변화시점 모형은 지금까지 한 시점에서 단 한개의 관측자료를 갖는 모형만 생각해 왔다. 이러한 모형을 확장시켜 각 시점에 한개 이상의 관측자료를 갖는 변화시점 모형을 생각한다. 이러한 모형에서 비모수적인 단측 그리고 양측 검정법을 찾았다. 검정 통계량은 지금까지 소개된 검정 통계량 형태를 확장시킨 형태이고 이들의 귀무가설 분포를 구하여 보았다. 또한 Monte Carlo연구를 통해 이들의 검정력을 비교해 보았다.

1. 서 론

시간의 흐름에 따라 연속적으로 얻어지는 서로 독립적인 확률변수 $X_1, X_2, \cdots, X_r, X_{r+1}, \cdots, X_T$를 생각해보자. 각 확률변수 X_i, $i=1, 2, \cdots, r$는 분포함수 $F(x)$, $i=r+1, \cdots, T$는 분포함수 $F(x-\triangle)$, $-\infty<\triangle<\infty$의 분포를 따른다. 여기에서 정수 r은 변화시점(changepoint)이고 $\triangle$는 변화시점 이후 변화된 양을 의미하는 위치모수(location parameter). 여기에서 관심의 대상이 되는 것은 변화시점 r을 모르는 경우 위치모수 $\triangle$의 검정 문제이다. 물론 r을 아는 경우에는 이 문제는 모집단이 두개일 때의 위치모수에 대한 검정문제이다. 변화시점이 여러개 있을 수 있으나 본 연구는 기껏해야 한 개인 경우만 다루기로 한다. 분포함수 F의 가정은 연속 이외의 분포형태에 대한 가정은 없다.

비모수적인 관점에서 이러한 문제를 처음 다룬 이는 Page(1954, 1955)이다. 그 후 여러사람에 의해 변화시점 문제가 다루워져 왔으나 모두 각 시점마다 독립적인 자료가 한 개씩만 얻는 경우이었다. 그러나 각 시점마다 여러개의 독립적인 자료를 얻을 수 있는 경우를 생각할 수 있다. 같은 시간에 서로 다른 지역에서 관측치를 얻는 경우이나, 서로 다른 실험자에

* 이 연구는 88년도 한국과학재단 연구 자원에 의한 결과임
(과제번호 : 883－0105－013－1)
** (713－714) 경북 경산군 진량면 내리동 산 23 대구대학교 자연과학대학 통계학과

의해 얻을 수 있는 여러 개의 자료가 이 경우에 해당된다. 예를 들어보면 어떤 지역에 댐 건설이후 기온의 변화에 관심이 있다 하자. 우리는 기온의 변화에 대한 직접적인 원인이 무엇인지 알 수 없고 어느 시점에서 변화되었는지에 대해서도 알 수 없다. 이러한 기온 변화가 있었으냐에 대한 문제로서, 우리는 과거부터 댐 근처의 여러지역에서 측정된 기온 자료들을 갖을 수 있다. 시간의 흐름에 따라 같은 시점에서 여러개의 관측치를 얻는 변화시점 모형에서, 미지의 변화시점 이후 변화된 양 △의 존재 유무에 대한 검정법을 찾고 그들의 검정력을 비교해 보는 것이 본 연구의 주된 목적이다. 이러한 검정법을 찾기 위해 기존의 변화시점 통계량을 적절히 확장시켜 그 검정 통계량으로 생각하였다.

비모수적인 변화시점 통계량은 크게 합-형태(sum-type)와 최대-형태(max-type)로 나눌 수 있는데 이들의 정확한 귀무가설 분포(null distribution)를 찾기 힘들기때문에 이들의 근사적 분포(asymptotic distribution)를 구하였고 또한 이들의 근사적 접근도와 검정력 비교를 Monte Carlo 모의실험을 통해 알아보았다.

2. 다중자료를 갖는 변화시점 모형

확률변수 $X_{11}, \cdots, X_{1n_1}; X_{21}, \cdots, X_{2n_2}; \cdots; X_{r1}, \cdots, X_{rn_r}; \cdots; X_{T1}, \cdots, X_{Tn_T}$는 연속 분포에서 추출된 N_T개의 서로 독립인 확률변수이다. 여기에서 r은 미지의 변화시점이고 $n_i(i=1, 2, \cdots, T)$는 시점 i에서의 표본 크기이다. $X_{ij}(i=1, 2, \cdots, T; j=1, 2, \cdots, n_i)$는 시점 i에서 j번째 얻어지는 관측자료이다. $i=1, 2, \cdots, r; \quad j=1, 2, \cdots, n_i$에서의 X_{ij}는 연속분포 $F(x)$를 따르고 $i=r+1, \cdots, T; j=1, 2, \cdots, n_i$에서의 X_{ij}는 연속 분포 $F(x-\triangle)$를 따른다. 여기에서 △는 위치 모수이다. 이러한 변화시점 모형을 다중자료를 갖는 변화시점 모형이라 부르기로 하자. 본 연구는 변화시점이 기껏해야 한 개인 경우만 다루기로 한다.

본 연구에서 관심의 대상이 되는 귀무가설은

$$H_0 : \triangle = 0$$

즉, 변화시점이 없다는 가정이다. 변화시점이 한 개 있다는 단측 대립가설로서

$$H_1 : \triangle > 0$$

그리고 양측 대립가설

$$H_2 : \triangle \neq 0$$

이다. 본 연구에서는 모집단의 분포 F는 연속 분포라는 가정 이외에 분포형태에 대한 가정이 없는 분포무관검정(distribution-free test)을 생각하였다.

3. 확장된 비모수 변화시점 통계량

다중자료를 갖고 변화시점이 한 개인 변화시점 모형에서 위치모수에 관한 검정 통계량을

찾기 위해 우리는 기존의 변화시점 통계량을 적절히 확장시키려 한다. 변화시점 통계량은 크게 두가지 형태의 통계량으로 나눌 수 있는 데 합-형태의 통계량으로

$$S_T = \sum_{k=1}^{T-1} C_{k,\ T-k} A_{k,\ T-k}$$

그리고 최대-형태의 통계량으로

$$M_T = \underset{1 \leq k \leq T-1}{\mathrm{Max}}\, C_{k,\ T-k} A_{k,\ T-k}$$

여기에서 $C_{k,\ T-k}$는 양의 가중계수이고, $A_{k,\ T-k}$는 두 표본 $X_1, \cdots, X_k$ 그리고 $X_{k+1}, \cdots, X_T$에 대한 통계량이다.

3. 1 합-형태 통계량

Bhattacharyya와 Johnson(1968)은 변화시점 모형에서 최초의 수준 $E(X_1)$을 모를 경우, 귀무가설 $H_0 : \triangle=0$ 대 대립가설 $H_1=\triangle>0$에 대한 검정 통계량으로 다음을 제시하였다.

$$J = \sum_{i=1}^{T} Q_i E[-f'(V^{(Ri)})/f(V^{(Ri)})],$$

여기에서

$Q_i = \sum_{t=1}^{i} q_t$는 누적 가중치이고 $q_1=0$이다. 그리고 $R=(R_1, R_2, \cdots, R_n)$은 연속형의 확률밀도 함수 f, 분포함수 F로부터 추출된 확률변수 $X_1, \cdots, X_n$의 순위 벡터이고 f'은 f의 미분함수이고 그리고 $V^{(1)} < \cdots < V^{(T)}$는 순서 통계량이다.

베이지안 관점에서 고려할 때 가중치 q_t는 사전 확률로 생각할 수 있다. 예를 들어 균일한 가중치(uniform weight)를 생각하면 아래와 같이 둘 수 있다.

$$q_t = (T-1)^{-1},\ t=2,\ \cdots,\ n$$

즉

$$Q_i = \sum_{t=1}^{i} q_t = (i-1)/(T-1).$$

통계량 J에 Wilcoxon 득점을 갖는 선형 순위 통계량을 생각한다면 J와 유사한 J'을 생각할 수 있다.

$$J' = \sum_{i=1}^{T} \sum_{j=1}^{T} (i-1)\phi(X_i - X_j)$$

여기에서

$$\phi(x)=1,\ x\geq 0 \text{ 그리고 } 0,\ x<0.$$

통계량 J'에서 각 시점마다 여러개의 독립자료를 얻을 수 있는 변화시점 모형의 통계량으로 확장시킨다면 다음을 얻을 수 있다.

$$B=\sum_{i=1}^{T}\sum_{j=1}^{n_i}\sum_{p=1}^{T}\sum_{q=1}^{n_p}(i-1)\ \phi(X_{ij}-X_{pq}).$$

3. 2 최대－형태 통계량

Pettitt(1979)는 귀무가설 H_0 : $\triangle=0$대 대립가설 H_1 : $\triangle>0$에 대한 검정 통계량 K를 제시하였다.

$$K=-\min_{1\leq k\leq T-1}\left(\sum_{i=1}^{k}\ \sum_{j=k+1}^{T}\text{sign}(X_i-X_j)\right),$$

여기에서

$$\text{sign}\,(X_i-X_j)=\begin{cases}1, & X_i>X_j\\ 0, & X_i=X_j\\ -1, & X_i<X_j\ .\end{cases}$$

Wolfe와 Schechtman(1984)은 Pettitt 통계량을 재 표현하여 K'을 생각하였다.

$$K'=2\max_{1\leq k\leq T-1}\ (V_{k,\ T-k}-E_0(V_{k,\ T-k})),$$

여기에서

$V_{k,\ T-k}=\sum_{i=k+1}^{T}\sum_{j=1}^{T}\phi(X_i-X_j)$이고 E_0은 귀무가설 하의 기대값을 의미한다.

위 통계량 K'을 본 다중자료를 갖는 변화시점 모형에 적절하게 변형하여 다음의 통계량 P_1을 생각하고 대립가설 H_2 : $\triangle\neq 0$에 대한 검정 통계량으로는 절대값을 이용하여 P_2를 설정하였다.

$$P_1=2\max_{1\leq k\leq T-1}(U_{k,\ T-k}-E_0(U_{k,\ T-k})),$$

$$P_2=2\max_{1\leq k\leq T-1}\left|\,U_{k,\ T-k}-E_0(U_{k,\ T-k})\,\right|$$

여기에서

$$U_{k,\ T-k}=\sum_{i=k+1}^{T}\sum_{j=1}^{n_i}\sum_{p=1}^{T}\sum_{q=1}^{n_p}\phi(X_{ij}-X_{pq}). \qquad (3.\ 1)$$

4. 근사적 귀무가설 분포

귀무가설 하에서 확장된 변화시점 통계량의 정확한 분포를 찾기 힘들기 때문에 이들의 근사적 분포(asymptotic distribution)를 구하여 보았다. 근사적 분포를 구하기 위해 필요되는 가정은 시점 T가 무한대로 커질 때 n_i/N_T는 상수 $\lambda_i \in (0, 1)$로 접근하는 것이다. 여기에서 $N_T = \sum_{i=1}^{T} n_i$는 누적된 표본수이다.

4. 1 확장된 Bhattacharyya와 Johnson 통계량

전 장에서 제시된 Bhattacharyya와 Johnson통계량의 근사적분포를 구하기 위해 다음과 같은 상수를 정의하자.

$$c(i) = \begin{cases} 0, & i=1, 2, \cdots, N_1 \\ 1, & i=N_1+1, \cdots, N_2 \\ \cdots & \\ T-1, & i=N_{T-1}+1, \cdots, N_T. \end{cases} \qquad (4.\ 1)$$

다음의 보조정리는 Hajek(1961) 정리의 특별한 경우이다.

보조정리 4. 1 A_T는 (4. 1)에서 정의된 회귀계수 $c_T(i)$를 갖는 선형 순위 통계량이다. ϕ는 (0, 1)에서 제곱 적분 가능한 함수이고 순위득점(rank score) $a_T(i)=\phi(i/(N_T+1))$이다. 그리고

$$\bar{c}_T = N_T^{-1} \sum_{i=1}^{N_T} c_T(i) \quad \text{그리고} \quad \bar{a}_T = N_T^{-1} \sum_{i=1}^{N_T} a_T(i)$$

라 한다면 귀무가설 H_0에서,

$$(A_T - \mu_T)/\sigma_T \xrightarrow{d} N(0, 1), \ T \longrightarrow \infty$$

여기에서

$$\sigma_T^2 = (N_T-1)^{-1} \sum_{i=1}^{N_T} (c_T(i) - \bar{c}_T)^2 \sum_{j=1}^{N_T} (a_T(j) - \bar{a}_T)^2.$$

B_1과 $(N_T+1)^{-1}B_2$는 보조정리 4. 1을 만족하는 선형 순위통계량이므로 다음 정리를 쉽게 얻을 수 있다.

정리 4. 1 상수 $c_T(i)$는 (4. 1)식과 같고 순위 득점을 다음과 같이 정의하자.

$$a_T(i) = \begin{cases} 1, & (N_T+1)/2 \leq i \leq N_T \\ 0, & \text{다른 경우} \end{cases}$$

그리고

$$a'_T(i) = \begin{cases} i/(N_T+1), & 1 \le i \le N_T \\ 0, & \text{다른 경우.} \end{cases}$$

그리하면 귀무가설 H_0에서,

$$\sigma_1^{-1}(B_1-\mu_1) \xrightarrow{d} N(0,\ 1),$$

$$\sigma_2^{-1}((N_T+1)^{-1}B_2-\mu_2) \xrightarrow{d} N(0,\ 1),\ T \longrightarrow \infty,$$

여기에서

$$\mu_1 = N_T\,\bar{c}_T\,\bar{a}_T,\ \sigma_1^2 = (N_T-1)^{-1}\sum_{i=1}^{N_T}(c_T(i)-\bar{c}_T)^2\sum_{i=1}^{N_T}(a_T(i)-\bar{a}_T)^2,$$

$$\mu_2 = N_T\,\bar{c}_T\,\bar{a}'_T,\ \sigma_2^2 = (N_T-1)^{-1}\sum_{i=1}^{N_T}(c_T(i)-\bar{c}_T)^2\sum_{i=1}^{N_T}(a'_T(i)-\bar{a}'_T)^2.$$

근사적 접근의 정확도를 알아보기 위해 모집단의 분포는 정규분포, 표본의 크기 $T=8$, $n_i=4$, $i=1, 2, \cdots, 8$을 택하여 Monte Carlo시행, 1,000번을 실시하여 유의수준 10%, 5% 그리고 1%에 해당되는 점을 찾았다. 귀무가설 H_0에 대하여 아래식을 만족하는 상수 a 그리고 b를 구하고 각 식의 좌변과 우변을 경험적으로 비교한 표가 아래 나타나있다.

$$\left.\begin{array}{l} P(\sigma_1^{-1}(B_1-\mu_1)>a) \sim \int_a^\infty (2\pi)^{-1/2}\exp(-x^2/2)dx \\ P(\sigma_2^{-1}\{(N_T+1)^{-1}B_2-\mu_2\}>b) \sim \int_b^\infty (2\pi)^{-1/2}\exp(-x^2/2)dx \end{array}\right\} = 0.1,\ 0.05,\ 0.01.$$

〈표 4. 1〉 상수 a 그리고 b의 이론적 추정치와 경험적 추정치와의 비교

상 수 / 유의수준	a		b	
	좌 변	우 변	좌 변	우 변
1%	2.429	2.325	2.335	2.325
5%	1.670	1.645	1.595	1.645
10%	1.366	1.285	1.315	1.285

그 결과 비교적 작은 시점 T에서 상수 a, b의 경험적 추정치와 이론적 추정치와의 차이가 비교적 작게 됨을 알 수 있다.

4. 2 확장된 Pettitt 통계량

확장된 Pettitt의 변화시점 통계량의 근사적 귀무가설 분포를 찾기 위하여 Lombard(1983)의 정리를 이용한다. 그리고 몇가지 필요한 기호의 정의를 알아보면

$[x] = \inf\{k : k \geq x,\ k \in (0, 1, 2, \cdots\}\}$,
$B = \{B(u),\ 0 \leq u \leq 1\}$는 표준 Brownian bridge과정이다.

정리 4. 2(Lombard(1983)) 확률변수 $X_1, X_2, \cdots, X_T$는 분포함수 F로부터 추출된 서로 독립인 확률변수이다. 분포함수 F는 절대연속이고 유한인 Fisher 정보를 갖는다.

$\{(a(T, i),\ 1 \leq i \leq T,\ i \geq 1\}$는 상수 득점의 배열이고 함수 $\phi : (0, 1) \rightarrow (-\infty, \infty)$는 다음을 만족한다.

$$\text{(i)} \quad \int_0^1 \phi(u)du = 0,\ 0 < D^2 = \int_0^1 \phi^2(u)du < \infty$$

$$\text{(ii)} \quad \sum_{i=1}^{T} a(T, i) = 0,\ T \geq 2$$

$$\text{(iii)} \quad \int_0^1 \{a(T, [uT]) - \phi(u)\}^2 du \rightarrow 0,\ T \rightarrow \infty.$$

확률과정 $B_T = \{B_T(u),\ 0 < u < 1\}$,

$$B_T(u) = T^{-1/2} D^{-1} \sum_{i=[uT]+1}^{T} a(T, R_i),\ B_T(1) = 0.$$

그리하면 $B_T \xrightarrow{d} B,\ T \rightarrow \infty$.

귀무가설 H_0에대하여 식 (3. 1)의 $U_{k,\ T-k}$의 평균과 공분산은 다음과 같다.

보조정리 4. 2

$$E_0(U_{k,\ T-k}) = (N_T+1)(N_T-N_k)/2,$$
$$\mathrm{Cov}_0(U_{k,\ T-k},\ U_{m,\ T-m}) = N_k(N_T-N_m)(N_T+1)/12,\ k \leq m.$$

보조정리 4. 3

확률과정 $B_T = \{B_T(u),\ 0 \leq u \leq 1\}$,

여기에서

$$B_T(u) = N_T^{-1}(12/(N_T+1))^{1/2}\{U_{[uT],\ T-[uT]} - E_0(U_{[uT],\ T-[uT]})\}\ ,$$

$$B_T(0) = B_T(1) = 0,\ U_{[uT],\ T-[uT]} = \sum_{i=[uT]+1}^{T} \sum_{j=1}^{n_i} R_{ij}.$$

그리하면 귀무가설 H_0에 대하여 B_T와 표준 Brownian bridge $\{B(N_k/N_T),\ 1 \leq k \leq T\}$은 동일한 평균과 공분산을 갖는다.

증명 $0 < u = N_k/N_T < N_m/N_T = v < 1$에 대하여

$$E_0(B_T(u)) = 0$$

$$\begin{aligned}\mathrm{Cov}_0(B_T(u),\ B_T(v)) &= 12N_T^{-2}(N_T+1)^{-1}\mathrm{Cov}_0(U_{[uT],\ T-[uT]},\ U_{[vT],\ T-[vT]}) \\ &= N_k(N_T-N_m)/N_T^2 \quad \text{(보조정리 4. 2)} \\ &= u(1-v),\ u \leq v.\end{aligned}$$

위 보조정리에 의하면 확률과정 B_T와 표준 Brownian bridge과정은 같은 적률구조(moment structure)를 가지고 있다. Lombard 정리에 Wilcoxon 득점함수를 택하면 조건 (i) 그리고 (iii)을 만족하고 $a(N_T, R_i) = R_i/(N_T+1) - 1/2$라고 한다면 조건 (ii)를 만족하여 아래의 정리를 얻을 수 있다.

정리 4. 3 $B = \{B(u), 0 \leq u \leq 1\}$를 표준 Browian bridge라 하자. 그리하면 귀무가설 H_0: $\triangle = 0$하에서

$$N_T^{-1} = (3/(N_T+1)^{1/2} P_1 \xrightarrow{d} \sup_{0\leq u\leq 1}\{B(u)\}$$

$$N_T^{-1}(3/N_T+1))^{1/2} P_2 \xrightarrow{d} \sup_{0\leq u\leq 1} |B(u)|, \; T \to \infty.$$

Doob(1949)는 아래와 같이 Kolmogorov-Smirnov의 적합도 검정 통계량의 극한 분포와 같은 수렴을 찾아냈다.

$$P[\sup_{0\leq u\leq 1}\{B(u)\} \geq x] = \exp(-2x^2)$$

$$P[\sup_{0\leq u\leq 1} |B(u)| \geq x] = -2 \sum_{r=1}^{\infty} (-1)^r \exp(2r^2x^2)$$
$$\approx 2\exp(-2x^2)$$

이를 이용해 아래 식을 만족하는 상수 a 그리고 b를 조사함으로 근사적 접근도를 조사하였다.

$$\left.\begin{array}{l} P[N_T^{-1}(3/(N_T+1)^{1/2} P_1 > a] \sim 2\exp\ (-2a^2) \\ P[N_T^{-1}(3/(N_T+1)^{1/2} P_2 > b] \sim \exp\ \ (-2b^2) \end{array}\right\} = 0.1,\ 0.05,\ 0.01.$$

다음 표는 시점 $T=16$, $n_i=4$ 그리고 모집단이 표분정규분포에서 알아보았으나 상수 a, b의 경험적 추정치와 이론적 추정치와의 차이가 많음을 보여준다.

〈표 4. 2〉 상수 a, b의 경험적 추정치와 이론적 추정치

상수 / 유의수준	a		b	
	좌 변	우 변	좌 변	우 변
10%	.917	1.072	1.065	1.223
5%	1.041	1.223	1.216	1.358
1%	1.350	1.517	1.485	1.627

5. Monte Carlo 검정력 비교

본 장에서는 다중자료를 갖는 변화시점 검정 통계량의 경험적 검정력을 Monte Carlo시행을 통해 알아보았다. 시점 수 $T=8$ 그리고 총 표본 수 $N_T=32$ 각 시점마다 자료수는 같

을 때와 서로 다를 때 조사하였다. Monte Carlo시행 수는1,000번 실시하였다. 귀무가설 H_0 : $\triangle=0$에 대하여 그리고 여러형태의 위치모수 그리고 변화시점에 따른 대립가설을 생각하였다. 주어진 모집단 분포는 정규, 균일, 지수, Cauchy 그리고 이중지수분포를 생각하였다.

이에 따른 난수(random number)는 I.M.S.L. 패키지를 이용해 추출해냈다. 특별히 이중지수분포난수는 자유도 2인 카이 제곱 분포를 하는 확률변수를 변수 변환하여 구하였다. 이러한 분포에 대하여 변화시점이 25%, 50% 그리고 75% 즉 $r=2$, 4 그리고 6에서 생각하였다. 또한 변화된 양 $\triangle$는 다음 식 $P[X_1>X_2]=0.6$, 0.7, 0.8 그리고 0.9 단, X_1 그리고 X_2는 각각 $F(x)$ 그리고 $F(x-\triangle)$의 분포를 따른다. 이 식을 따르는 $\triangle$의 값은 아래 표와 같다.

〈표 5. 1〉 변화된 양 △의 값

$P[X_2>X_1]=$ / 분 포	0.6	0.7	0.8	0.9
균일분포	.1060	.2250	.3680	.5530
정규분포	.3606	.7425	1.1995	1.8219
지수분포	.2231	.5108	.9163	1.6094
Cauchy 분포	.6498	1.4531	2.7528	6.1553
이중지수 분포	.4094	.8731	1.4661	2.3973

유의수준은 0.05에서 생각하였다. 그 임계점은 확장된 변화시점 검정 통계량 값을 경험적으로 1,000번 시행하여 그 크기가 50번째 되는 값으로 택하였다. 모든 Monte Carlo시행 결과는 VAX－11/750 컴퓨터를 사용하였다.

5. 1 확장된 변화시점 검정통계량들의 검정력 비교

본 절에서는 단측 혹은 양측 대립가설하에서 확장된 변화시점 통계량의 검정력을 비교하여 보았다. 표본수는 $T=8$, $n_i=4$, $i=1$, 2, $\cdots$, 8에서 실시하였다. 모든 Monte Carlo시행에 따른 검정력 함수는 변화시점 $r=4$에서 서로 대칭을 이룬다. 다시 말하면 $r=2$에서의 검정력은 $r=6$에서의 검정력과 같다.

다음의 표는 변화시점 25% 그리고 50%에서 대립가설의 H_1 : $\triangle>0$ 그리고 H_2 : $\triangle\neq0$에서의 검정력을 나타낸다. 표 5. 1 그리고 표 5. 2의 모든 검정력은 단측 혹은 양측 대립가설에 대해서 유의수준 0.05보다 큰 값을 나타낸다. 그리고 모든 검정력은 △값이 증가함에 따라 높아짐을 알 수 있다. 또한 모든 검정력은 변화시점이 50%일 때 최대이고 25%에서 최소를 나타낸다.

5. 2 확장된 변화시점 통계량과 기존 통계량과의 검정력 비교

본 절에서는 매 시점마다 여러개의 관측자료를 갖는 확장된 변화시점 통계량과 한 시점에 단 한개의 관측자료를 갖는 기존의 변화시점 통계량의 검정력을 비교한다는 것은 문제가 있지만 다중 자료를 갖는 변화시점 통계량이 없기 때문에 총 표본 수가 같을 때 이들의 검정력을 비교하여 보았다. 여기에서 기존의 변화시점 통계량이라는 것은 $n_i=1$, $i=1$, 2, $\cdots$, T인

〈표 5. 1〉 확장된 변화시점 통계량의 검정력 비교
($\alpha=0.05,\ T=8,\ n_i=4$)

분포	$P[X_{8,1}>X_{1,1}]$	25%($r=2$) B	25%($r=2$) P_1	25%($r=2$) P_2	50%($r=4$) B	50%($r=4$) P_1	50%($r=4$) P_2
균 일 분 포	0.6	.172	.147	.078	.238	.215	.129
	0.7	.376	.361	.214	.557	.563	.,408
	0.8	.664	.709	.514	.877	.908	.832
정 규 분 포	0.6	.175	.142	.080	.229	.214	.125
	0.7	.377	.365	.219	.555	.568	.411
	0.8	.661	.718	.535	.874	.905	.827
지 수 분 포	0.6	.164	.153	.097	.209	.208	.119
	0.7	.357	.366	.224	.534	.551	.404
	0.8	.629	.679	.520	.864	.892	.817
Cauchy 분 포	0.6	.165	.166	.090	.216	.230	.150
	0.7	.379	.422	.258	.566	.593	.465
	0.8	.656	.721	.579	.850	.904	.838
이중지수분포	0.6	.175	.147	.088	.239	.217	.141
	0.7	.166	.361	.263	.564	.553	.437
	0.8	.668	.690	.560	.865	.886	.838

경우이다.

기존의 변화시점 통계량과 확장된 변화서점 통계량과의 검정력을 비교하기 위해 총 표본 $N_T=\sum_{i=1}^{T} n_i=32$를 택하고 다음과 같은 방법을 사용하였다.

방법 Ⅰ : $T=32,\ n_i=1,\ i=1,\ 2,\ \cdots,\ 32$
방법 Ⅱ : $T=8,\ n_i=4,\ i=1,\ 2,\ \cdots,\ 8$
방법 Ⅲ : $T=8,\ n_i=n_8=1,\ n_2=n_7=2,\ n_3=n_6=6,\ n_4=n_5=7$
방법 Ⅳ : $T=8,\ n_1=n_8=7,\ n_2=n_7=6,\ n_3=n_6=2,\ n_4=n_5=1$

위 방법은 총 표본수는 같지만 특히 방법 Ⅰ은 기존의 변화시점 통계량인 경우이고, 방법 Ⅱ는 각 시점마다 관측자료가 같은 경우이고 방법 Ⅲ은 시점 중앙이 가장자리보다 많은 경우이고 방법 Ⅳ는 방법 Ⅲ의 반대인 경우이다.

한편 변화된 양 Δ의 값은 $P[X_2>X_1]=0.7$(X_1은 $F(x)$ 그리고 X_2는 $F(x-\triangle)$의 분포를 따르고 서로 독립인 확률 변수이다.)을 만족하는 Δ의 값을 택하였다.

본 절에서는 여러형태의 분포에서 단측 혹은 양측 대립가설하에서 경험적인 검정력을 구하였다. 그 결과 방법 Ⅰ과 방법 Ⅱ는 검정력의 큰 차이를 볼 수 없다. 그러나 방법 Ⅳ는 다른 어떤 방법보다 그 검정력이 높음을 알 수 있다. 그 이유는 두 표본 위치모수 문제에서

두 표본의 수가 일치할 때 검정 통계량의 검정력이 최대가 되는 이유가 같다. 그러므로 미지의 모수 r에서 $\sum_{i=1}^{r} n_i = \sum_{i=r+1}^{T} n_i$일 때 최대의 검정력을 갖는다. 그러나 우리는 변화시점 r을 모르기 때문에 시점 가장자리가 중앙보다 많은 관측자료를 갖게하여야 할 것이다.

〈표 5. 2〉 각 시점에서의 독립자료 수에 따른 검정력 비교

분포 \ 통계량 \ 변화시점	25%($r=2$)			50%($r=4$)		
	B	P_1	P_2	B	P_1	P_2
균 일 분 포	.357	.389	.221	.539	.570	.409
	.376	.361	.214	.557	.563	.408
	.182	.115	.078	.519	.606	.460
	.535	.514	.375	.613	.568	.438
정 규 분 포	.358	.385	.219	.552	.575	.407
	.377	.365	.219	.555	.568	.411
	.182	.118	.068	.524	.594	.456
	.542	.522	.396	.602	.576	.435
지 수 분 포	.369	.344	.231	.523	.524	.416
	.357	.366	.224	.534	.551	.404
	.176	.100	.074	.508	.547	.457
	.550	.467	.374	.609	.524	.446
Cauchy 분 포	.387	.392	.233	.568	.569	.432
	.379	.422	.258	.566	.593	.465
	.145	.128	.068	.518	.592	.459
	.555	.563	.421	.615	.609	.453
이 중 지 수 분 포	.357	.386	.226	.540	.573	.439
	.371	.361	.263	.564	.553	.437
	.166	.110	.063	.522	.590	.454
	.558	.576	.396	.623	.636	.448

첫번째부터 네번째 값은 각각 방법 Ⅰ, Ⅱ, Ⅲ 그리고 Ⅳ의 검정력을 의미한다.

6. 결 론

본 연구에서는 다중자료를 갖는 변화시점 모형에서 위치모수에 대한 비모수적 검정법을 생각하여 보았다. 여기에서 우리는 검정 통계량을 기존의 변화시점 통계량을 적절히 확장시켰다. 지금까지는 각 시점에 여러개의 관측자료를 얻을 수 있는 경우일지라도 시점에 단 한개의 관측자료를 갖는 기존의 변화시점 통계량을 사용하여왔다. 그러나 이 문제는 본 연구에서 제시된 확장된 변화시점 통계량을 사용하면 될 것이다.

여러 형태의 확장된 변화시점 통계량의 검정력에서 우리는 시점의 중앙보다 가장자리에 많은 관측자료를 가질 때 높은 검정력을 가진다는 것을 알았다. 여기에서 각시점마다 갖는 관측자료의 수는 각 시점에서의 가중치를 의미하게 된다. 변화시점 통계량의 형태가 유사한

이유는 이들의 가중치에 차이를 둔 것이다.

본 연구에서 위치모수에 대해 생각하였지만 이를 다른 형태의 모수에도 적용시킬 수 있을 것이다. 미지의 변화시점 r과 변화된 양 Δ의 점추정 혹은 구간추정에 대한 문제는 별로 다루어지지 않았다. 다만 Pettitt(1980)와 Lombard(1985)는 변화시점 r에 대한 점추정 문제를 Schechtman(1983)은 변화된 양 Δ에 대한 구간추정 문제를 다루었다. 이와같이 변화시점 문제에는 아직까지 밝혀지지 않은 많은 문제를 갖고 있다.

◇ 참 고 문 헌 ◇

(1) Bhattacharyya, G. K. and Johnson, R. A. (1986). "Nonparametric tests for shift at an unknown time point." *Ann. Math. Statist.*, **39**. 1731 – 1743.

(2) Doob, J. L. (1949), "Heuristic approch to the Kolmogorov Smirnov theorems." *Ann. Math. Statist.*, **20**, 393 – 403.

(3) Hajek, J. (1961). "Some extensions of the Wald-Wolfowitz-Noether theorem." *Ann. Math. Statist.*, **32**, 506 – 523.

(4) Lewis, P. A. W. (1961). "Distribution of the Anderson-Darling statistics." *Ann. Math. Statist.*, **32**, 1118 – 1123.

(5) Lombard, F. (1983). "Asymptotic distribtuions of rank statistics in the change point problem." *South Afr. Statist.* J. **17**, 83 – 105.

(6) Page, E. S. (1954). "Continous inspection schemes." *Biometrics*, **41**, 100 – 115.

(7) Page, E. S. (1955). "A test for change in a parameter occuring at an unknown point." *Biometrika*, **42**, 523 – 526.

(8) Pettitt, A. N. (1979). "A non-parametric approch to the changepoint problem." *Appl. Statist.*, **28**, 126 – 135.

(9) Schechtman, E. (1983). "A conservative nonparametric distribution-free confidence bound for the shift in the change point problem." *Comm. Statist. – Theor Math. A* **12**(21), 2455 – 2464.

(10) Wolfe, D. A. and Schechtman, E. (1984). "Nonparametric statistical procedures for changepoint problem." *Journal of Statistical Planning and Inference*, **9**, 389 – 396.

Nonparametric Test Procedures for the Changepoint Problem with Multiple Observations

Kyung Moo Kim*

〈Abstract〉

In the analysis of changepoint model the situation where single observation is taken at each time point has been considered. In an effort to extend this to the general situation, we may consider the changepoint model with more than one observation at each time point.

These tests are developed without assuming any particular form for the underlying distribution, we propose the one-sided and two-sided nonparametric tests by extending the tests that have been considered in the changepoint model with single observation at each time point and obtain their asymptotic null distributions. We compare the empirical powers among the extended changepoint tests under one-sided or two-sided alternatives. We also compare the powers of the extended changepoint tests with those of the original test via the Monte Carlo simulation.

* Dept. of Statistics, Taegu Univ, 23 Naeri-dong, Jinryang-myun, Kyungsan-gun, kyungpook, 713－714, Korea.