응용통계연구 제 7권 1호
1994년 3월 pp. 35 - 46

이변량 변화시점모형에 대한 비모수적인 검정법[1)]

김경무[2)]

요약

이변량 변화시점 모형에서 위치모수에 대한 비모수적 방법인 순위-모양 검정법을 제시하였다. 이를 경험적인 검정력을 통하여 모수적인 검정과 비교한 결과,귀무가설분포가 이변량정규분포일 때를 제외하고는 순위-모양 검정이 월등히 우수함을 알 수 있었다.또한 변화시점에 대한 점추정량들을 경험적인 평균제곱오차를 이용하여 비교 분석하였다.

1. 서론

사회현상은 복잡한 요인들에 의하여 변화되어 가고있다. 이러한 시간의 흐름에 따라 얻어지는 관측자료가 있을 때, 이들 자료가 언제부터인가 변화된 값으로 나타나게 되는 사회현상을 볼 수 있게 되는데, 이러한 변화가 실제 있었는 지를 알아보는 문제가 변화시점 문제이다.

시간의 흐름에 따라 연속적으로 얻어지는 서로 독립적인 확률변수가 각각 아래의 분포를 따른다 하자.

$$\begin{aligned} &X_1, X_2, \cdots, X_c && \sim F(x) \\ &X_{c+1}, X_{c+2}, \cdots, X_n && \sim F((x-\delta)/\eta), \end{aligned}$$

여기에서 양의 정수 c는 미지의 변화시점(changepoint)이고, δ는 위치모수, η는 척도모수를 나타내고, 그리고 F는 분포함수를 의미한다.위 모형은 미지의 변화시점 이후 분포함수가 변화 되었음을 뜻한다. 변화시점을 알 수 없을때,변화된 양 그리고 변화시점에 대한 추정 및 검정문제를 다루는 모형을 변화시점 모형이라고 한다. 물론 변화시점을 알고 있다면 위치모수에 대한 검정은 모수적방법인 이표본 t-검정을 실시하던지 혹은 비모수적인 Wilcoxon 순위합 검정을 이용하면 될 것이다.변화시점을 알 수 없을때 위치모수 혹은 척도모수에 관한 연구는 Page(1954)이후 최근까지 많이 연구되어 왔다.변화시점 모형에 관한 연구 결과를 비교적 잘 정리해 놓은 논문으로는 Shaban(1980) 그리고 Wolf-Schechtman (1984)을 들 수 있다.

그러나 지금까지의 변화시점 문제는 위치모수,척도모수 혹은 기타 다른 문제에서도 대부분 일변량(univariate)인 경우에만 국한되어 연구되어왔다. 미지의 변화시점에서 분포변화의 유,무에 대한 문제는 어느 한 변수보다는 그 변수와 상관관계가 있는 다른 변수를 동시에 생각하여 본다면 더욱 효율적일 것이다.이러한 문제를 처음으로 생각한 이들은 Sen-Srivastava(1973)이다. 이들은 일변량 변화시점 모형을 다변량으로 확장하여 위치모수에 대한 모수적인 검정법을

1) 이 연구는 1992년도 교육부지원 한국학술 진흥재단의 지방대학 육성과제 학술연구 조성비에 의하여 연구되었음.

2) (713-714) 경북 경산군 진량면 내리리 대구대학교 자연과학대학 통계학과

제시하였다.

위치모수에 대한 이변량 변화시점모형(bivariate changepoint model)으로 다음과 같은 독립적인 n개의 이변량 확률변수를 생각하여 보기로 하자.

$$\begin{aligned} &(X_1,Y_1),(X_2,Y_2),\ldots,(X_c,Y_c) &&\sim F(x,y), \\ &(X_{c+1},Y_{c+1}),\ldots,(X_n,Y_n) &&\sim F(x-\delta_1,y-\delta_2) \end{aligned}$$

여기에서 F(x,y)는 이변량 분포함수이고, δ_1와 δ_2는 위치모수 그리고 양의 정수 c는 미지의 변화시점이다.

위와 같은 모형의 예를 의학적인 경우에서 찾아 보기로 하자. 유방암 환자는 수술 후에도 암세포가 다른 곳으로 전이 되었는지를 알아 볼 수 있는 여러가지의 생화학적인 검사를 받게된다. 예를 들면 세포조직검사를 한다던지 혹은 약물투입으로 인한 환자의 반응상태를 조사하게 된다.이러한 검사결과는 일반적으로 암세포의 전이를 알아 볼 수 있는 수치형태로 나타나게 된다. 암세포가 다른 신체 부위로 전이되었는 지를 조기에 알아보는 일은 환자에게나 의학적인 측면에서 매우 중요한 문제이다. 만약 한 가지의 검사결과를 가지고 암세포의 전이를 알아 본다면 이는 일변량의 변화시점 문제인 경우이나, 서로 상관관계가 있는 두 가지의 검사 결과를 이용한다면 이는 이변량 변화시점 모형으로 생각할 수 있고 변화의 유,무를 효율적으로 알 수 있는 방법이 될 것이다.

본 연구에서는 이변량 변화시점 문제에 대한 검정 방법으로 모집단이 이변량 정규분포를 따를때, Sen-Srivastava(1973)의 모수적인 검정방법과 모집단의 분포형태에 대한 가정이 없을 때의 비모수적인 순위-모양 검정을 2절에서 제시하려고 한다. 또한 3절에서 이들 두 검정방법을 경험적인 검정력을 통하여 비교한다. 4절에서는 변화시점에 대한 점추정으로 분포함수와 검정통계량을 이용한 점추정량을 소개하고, 이들의 경험적인 평균제곱오차를 계산하여 이들의 효율성을 알아보기로 한다.

2. 이변량 변화시점 모형

위치모수에 대한 이변량 변화시점 모형으로,서로 상관관계가 있는 한 쌍의 확률변수 n개가 서로 독립적으로 각각 아래의 분포를 따른다 하자.

$$\begin{aligned} &(X_i,Y_i),i=1,2,\ldots,c, &&\sim F(x,y), \\ &(X_i,Y_i),i=c+1,\ldots,n, &&\sim F(x-\delta_1,y-\delta_2), \end{aligned}$$

여기에서 양의 정수 c는 미지의 변화시점이고, (δ_1,δ_2)는 변화시점 이후 (X,Y)의 변화된 양, 그리고 F는 연속형의 분포함수를 의미한다. 서로 상관관계가 있는 한 쌍의 확률변수가 미지의 변화시점 이후,실제 변화가 있었는지를 알아보는 모형이 이변량 변화시점 모형이다. 위 모형을 이변량에서 다변량으로 확장하여 생각할 수도 있을 것이다. 위의 이변량 변화시점 모형에서 위치모수에 대한 통계적 가설 형태로, 변화가 없다는 귀무가설은

$$H_0:\ \delta_1=\delta_2=0, \tag{2.1}$$

이고, 그리고 변화시점 이후 변화가 있다는 대립가설은

$$H_1:\ \delta_1 > 0 \text{ 혹은 } \delta_2 > 0\ , \qquad (2.2)$$

으로 둘중 적어도 한 개는 변화가 있다는 가설이다. 본 연구의 주된 목적은 모집단의 분포형태에 대한 가정없이 위 가설에 대한 검정법을 찾는 것이다.

2.1 Sen-Srivastava 검정

Sen-Srivastava(1973)은 위치모수에 대한 일변량 변화시점모형을 다변량으로 확장하여 모수적인 검정법을 제시하였다. 이차원 확률벡터(random vector)가 서로 독립적으로 아래의 이변량 정규분포를 따른다고 가정하자.

$$\begin{aligned} X_1, X_2, \cdots, X_c &\sim BN(\mu, \Lambda), \\ X_{c+1}, \cdots, X_n &\sim BN(\mu+\delta, \Lambda), \end{aligned}$$

여기에서 확률벡터 $X^T=(X,Y)$,미지의 평균벡터 $\mu^T=(\mu_1,\mu_2)=(E(X),E(Y))$, 변화된 양 $\delta^T=(\delta_1,\delta_2)$,그리고 분산행렬인 Λ는 단위행렬이다.

Sen-Srivastava는 μ 와 δ 에 대한 사전확률을 다변량 정규분포라 가정하고 베이지안 방법으로 귀무가설 H_0:δ = $\mathbf{0}$ 에 대한 검정통계량을 유도하였다. 이를 이변량으로 고쳐서 생각하면 아래와 같다.

$$\begin{aligned} S_n &= n^{-2}\sum_{i=1}^{n-1}\Big(\sum_{j=1}^{n-1}(X_{j+1}-\overline{X})\Big)^T\Big(\sum_{j=1}^{n-1}(X_{j+1}-\overline{X})\Big) \\ &= n^{-2}\sum_{k=1}^{2}\sum_{j=1}^{n-1}\Big(\sum_{i=j}^{n-1}X_{i+1}^{(k)}-\overline{X}\Big)^2, \end{aligned}$$

여기에서 $X_i^{(k)}$는 이차원 확률벡터 X_i의 k번째 요소이고 $\overline{X}=n^{-1}\sum_{i=1}^{n}X_i$ 이다.

위 통계량 S_n의 값이 크면 귀무가설을 기각시키게 된다.한편 이들은 위 검정통계량의 근사적인 귀무가설분포를 아래와 같이 유도하였고 근사적인 수치도 계산해 놓았다.

$$F(z)=\sum_{j=0}^{\infty}2^{(3/2)}\pi^{-1/2}z^{-1/2}\exp(-(1+2j)^2/2z)$$

2.2 순위-모양 검정(Rank-like test)

Moses(1963)에 의해 소개된 순위-모양 검정은 원 자료의 순위를 이용하는 것이 아니라 교환가능한(exchangeable) 확률변수들의 순위를 이용하는 검정법이다.이변량 변화시점 모형에 적용하기 위하여 공간 중앙값(spatial median)과 확률벡터들에 대해 대칭인 용어의 정의 그리고 Randles-Wolfe (1979,pp.356-357)의 정리를 인용하여 보기로 한다.

정의 2.2.1 (Brown,1983) (X_i, Y_i), $i=1,2,\ldots,n$,을 2차원 좌표 표본점이라 할때,

$$T(\gamma_x,\gamma_y)=\sum_{i=1}^{n}\{(X_i-\gamma_x)^2+(Y_i-\gamma_y)^2\}^{1/2}$$

을 최소로 하는 (γ_x,γ_y)의 추정값 (α,β)를 공간 중앙값(spatial median)이라 한다.

정의 2.2.2 p-차원 확률벡터 $(X_1,X_2,\ldots,X_n)$에 대한 어떤 함수 $g(.)$가 n개의 양의 정수 (1,2,...,n)의 임의의 순열 $(d_1,d_2,\ldots,d_n)$에 대해서도

$$g(X_1,X_2,\ldots,X_n)=g(X_{d1},X_{d2},\ldots,X_{dn})$$

을 만족할 때 g(.)는 p차원의 n개 확률벡터들에 대해 대칭인(symmetric in its arguments)함수라고 한다.

정리 2.2.3 (Randles-Wolfe,1979) 연속형의 분포로 부터 추출된 p-차원 확률 벡터를 $X_i=(X_{i1},\ldots,X_{ip})$,$i=1,2,\ldots,n$,라 하자. 그리고 g(.)는 n개의 확률 벡터들에 대해 대칭인 함수이다. 이때 새로운 임의의 실변수함수 h(.,.)에 의해서 정의되는 일변량 확률변수

$$W_i=h(X_i,g(X_1,\ldots,X_n))$$

는 다음의 두가지 성질을 만족한다.

첫째, $W_1,W_2,\ldots,W_n$는 서로 교환가능한(exchangeable) 확률변수이다.

둘째, $R_i=\text{rank}(W_i)$라 하고 $\Pr(W_i=W_j)=0$,$i\neq j$라고 가정하면, 임의의 순열 $(d_1,d_2,\ldots,d_n)$에 대해서 $\Pr[(R_1,R_2,\ldots,R_n)=(r_{d1},r_{d2},\ldots,r_{dn})]$ = 1/n! .

위 정리는 교환가능한 확률변수들의 순위를 이용한 검정은 분포-무관(distribution-free)인 비모수적 방법임을 의미한다. 위 정리를 이용하여 확률벡터들에 대해 대칭인 함수 g(.)를 찾아 보자.

일차원 자료의 중앙값이 자료들의 거리를 최소로 하듯이 공간좌표에서도 이들의 거리를 최소로 하는 것이 공간 중앙값이고 이는 자료를 주어진 좌표축에 대해 회전을 시켜도 공간 중앙값은 변하지 않는다. 즉,확률벡터들의 임이의 순열 $(d_1,d_2,\ldots,d_n)$ 들에 대해서도 같은 공간 중앙값을 갖게 된다.그러므로 함수 g(.)를 n개의 표본벡터로 얻어지는 공간 중앙값 (α,β)를 함수 g(.)로 선택하자.즉,

$$g(X_1,X_2,\ldots,X_n)=(\alpha,\beta)=g(X_{d1},X_{d2},\ldots,X_{dn}).$$

공간 중앙값 M=(α,β)와 i번째 표본점 $P_i=(X_i,Y_i)$를 연결하는 선분 $\overline{MP}$와 수평축 X축과의 양의 각도를 W_i 라 한다면,

$$W_i = \text{angle}\{\overline{MP}, X\text{축}\},\ (0 \le W_i \le 2\pi).$$

이다. 따라서 위 정리에 의하여 일변량 확률변수 W_i, $i=1,2,\ldots,n$,는 교환가능한 확률변수가 될 수 있다.공간 중앙값 M과 표본점 P_i의 거리를 d_i, $i=1,2,\ldots,n$,라 둔다면,

$$X_i = \alpha + d_i \cos W_i$$
$$Y_i = \beta + d_i \sin W_i$$

와 같이 될 것이고,이를 정의 2.2.1의 T에 대입해서 α 와 β에 관하여 각각 미분한다면

$$\sum_{i=1}^{n} \cos(W_i) = \sum_{i=1}^{n} \sin(W_i) = 0$$

이 된다. 또한 각도에 대한 순위를 $r_i = \text{rank}(W_i)$라 두기로 한다면, 이를 이용한 검정통계량은 정리2.2.3에 의하여 분포-무관의 검정법이 될 수 있다. 귀무가설 H_0: $\delta = 0$ 이 참이라면 $(\Sigma \cos(W_i))^2 + (\Sigma \sin(W_i))^2$의 값은 작게 나타날 것이다.

Lombard(1986)는 Mardia(1972)의 검정통계량을 방향 데이터(directional data)에 대한 변화시점 모형에 적용시켰다. 따라서 우리는 2차원 표본점들의 순위 r_i를 Mardia(1972)가 제시한 통계량,

$$u^2_{c,n-c} = \frac{\left[\sum_{i=1}^{c} \cos(2\pi r_i/n)\right]^2 + \left[\sum_{i=1}^{c} \sin(2\pi r_i/n)\right]^2}{c(n-c)} \tag{2.2.1}$$

에 적용하고, 가설 (2.1)과 (2.2)에 대한 미지의 변화시점에 대한 합-형태와 최대-형태를 검정통계량으로 제시한다. 즉, 그 통계량은 아래와 같다.

$$R_n = 2\sum_{c=1}^{n-1} u^2_{c,n-c},$$
$$M_n = 2\,Max_{1 \le c \le n-1}\, u^2_{c,n-c}.$$

한편 $\{B_1(u), 0 \le u \le 1\}$ 그리고 $\{B_2(u), 0 \le u \le 1\}$를 서로 독립인 표준 Brownian Bridge 확률과정이라고 한다면, 이들의 근사적인 귀무가설 분포는 Lombard(1986)에 의해 아래와 같게 된다.

$$R_n \approx \int_0^1 \{B_1^2(u) + B_2^2(u)\}(u(1-u))^{-1} du,$$
$$M_n \approx \sup_u [\{B_1^2(u) + B_2^2(u)\}(u(1-u))^{-1}],\ n \to \infty.$$

또한 Lombard(1986)는 위 검정통계량 R_n의 우측 확률에 대한 근사식을 아래와 같이 유도하

였다.

$$\Pr(R_n > x) \approx 3e^{-x},\ n \to \infty.$$

귀무가설 분포가 이변량 정규분포일때, 상관계수와 유의수준 α = 0.05 에 따른 R_n,그리고 M_n 검정통계량의 임계점을 모의실험을 통하여 경험적으로 구하고 그 값을 <표 2.2.1>에 정리하였다.

<표 2.2.1>

		통계량	
		R_n	M_n
상관계수 ρ	0.5	0.3187	3.3429
	0.6	0.2963	3.5296
	0.7	0.2999	3.5770
	0.8	0.3149	4.2365

3.검정력 비교

귀무가설 분포가 이변량 정규분포들의 혼합 분포인 경우,모수적인 Sen-Srivastava의 검정통계량 S_n, 비모수적인 순위-모양 검정통계량 R_n,그리고 M_n의 검정력을 모의실험을 통하여 경험적으로 구하여 보았다.

대립가설 H_1:δ > 0 에 대한 검정력은 모집단의 분포함수, 유의수준, 표본의 크기, 변화시점, 변화된 양, 반복수, 그리고 검정통계량에 따라 다르게 나타나기 때문에 이들 각 경우의 검정력을 구한다는 것은 많은 시간이 필요하게 된다.그래서 본 연구는 비모수적인 경우에도 동일하게 아래의 조건으로 모의실험하였고 그 결과가 <표3.1>에 나타나 있다.

* 프로그램 언어 : FORTRAN Ver.5.0,
* 난수(Random number) :IMSL(International Mathematical & Statistical Libraries) Ver.1.1, 함수 부프로그램 RNUN, RNMVN,
* 공간 중앙값 계산 : Gower(1974)의 함수 부프로그램 MEDIAN,
* 귀무가설 분포 : $f(x,y) = p\Phi(x,y;0,0,1,1,\rho) + (1-p)\Phi(x,y,0,0,\sigma^2,\sigma^2,\rho)$,
  여기에서 Φ는 이변량정규분포의 확률밀도함수이고 확률 p, 모분산 σ^2 그리고 상관계수 ρ의 값은 각각 다음과 같이 실험하였다.
  p = 0.5, 0.75, 1.0,
  σ^2= 25, 100, σ^2= 25, 100,
  ρ = 0.8,
* 표본 크기 : n = 30,
* 변화시점 : c = 5, 15,
* 검정통계량 : S_n, R_n, M_n,
* 반복수 : S_n ; 1000회, R_n, M_n,;100회,
* 유의수준 : α = 0.05,

* 변화된 양 : δ_1, δ_2 = 1.0, 2.0.

<표 3.1> 대립가설 $H0:\delta > 0$ 에 따른 검정력비교

c=5

σ^2	p / δ_2 \ δ_1	0.5: 1	0.5: 2	0.75: 1	0.75: 2	1.0: 1	1.0: 2
25	1	.158	.392	.093	.188	.272	.625
		.33	.46	.14	.36	.18	.34
		.13	.19	.14	.24	.17	.27
	2	.384	.589	.189	.294	.622	.844
		.50	.59	.39	.59	.34	.46
		.13	.20	.26	.30	.27	.32
100	1	.156	.391	.085	.169	.272	.625
		.43	.54	.53	.71	.18	.34
		.25	.34	.32	.28	.17	.27
	2	.387	.596	.165	.253	.622	.844
		.53	.63	.72	.78	.34	.46
		.31	.35	.31	.37	.27	.32

c=15

25	1	.476	.885	.254	.579	.710	.988
		.75	.90	.50	.87	.23	.67
		.65	.85	.52	.81	.23	.72
	2	.887	.969	.577	.807	.986	.999
		.87	.95	.87	.96	.63	.93
		.83	.89	.85	.94	.69	.90
100	1	.471	.889	.220	.510	.710	.988
		.85	.95	.93	.97	.23	.67
		.89	.93	.91	.98	.23	.72
	2	.882	.967	.510	.748	.986	.999
		.93	.96	.97	.98	.63	.93
		.90	.96	.98	.97	.69	.90

* 각 칸에 첫번째, 두번째, 그리고 세번째 행은 각각 $S_n, M_n,$그리고 R_n의 검정력을 의미한다.
* 검정력 추정치 $(\hat{p})$의오차 = $\sqrt{\hat{p}(1-\hat{p})/\text{반복수}}$

순위-모양 검정의 검정력은 공간중앙값을 구하는데 많은 시간이 필요하기 때문에 반복수를 100회로 하였다. 변화된 양은 두 확률변수 X,Y에 대한 표준편차의 1.0배, 그리고 2.0배 만큼 이동하였다. 모의실험한 결과를 살펴보면, 모든 검정력은 변화된 양, 두 모분산 그리고 상관계수 크기에 비례하고, 변화시점이 중앙에 있을 때가 가장 높게 나타남을 알 수 있다. 이는 이표본

검정에서 두 표본의 표본 크기가 같게 됐을 때가 검정력이 높게 나타나는 경우와 동일하다. 귀무가설 분포에서 $p=1.0$일 경우, 즉 이변량 표준정규분포인 경우에는 모수적인 S_n 검정이 가장 높게 나타난다. 그러나 나머지의 혼합 분포인 경우에는 비모수적인 두 순위-모양 검정들이 모수적방법보다 검정력이 월등히 높게 나타남을 알 수 있다. 또한 순위-모양 검정의 최대-형태 M_n이 합-형태 R_n보다 검정력이 약간 높게 나타나게 된다. 일반적으로 일변량의 변화시점 검정통계량의 검정력은 비교적 낮게 나타나게 되는데, 본 이변량 변화시점모형에서 제시된 순위-모양 검정법은 검정력이 높게 나타남을 알 수 있다.

4.변화시점에 대한 점추정량

본 절에서는 변화시점에 대한 점추정량을 생각하고 이들의 경험적인 평균제곱오차(MSE)를 이용하여 효율적인 점추정량을 알아보려 한다.분포함수와 검정통계량을 이용하여 세가지의 점추정량을 들 수 있겠다.

4.1 Carlstein 추정량

Carlstein(1988)는 일변량 변화시점모형에서 변화시점에 대한 점추정을 경험적인 분포함수를 이용하였다. 변화시점 이전과 이후의 분포함수의 차이가 최대가 되는 시점을 점추정량으로 생각하였다. 이를 이변량인 경우로 확장하여 보기로 한다.

변화시점 이전의 표본들의 분포함수를 $\widehat{F_c}$,변화시점 이후의 표본들의 분포함수를 $\hat{F}_{n-c}$ 라 두기로 하자.

$$D_c=\sqrt{(c/n)(1-c/n)}\,\underset{1\le i\le n}{Max}\,|\,\widehat{F_c}(x_i,y_i)-\hat{F}_{n-c}(x_i,y_i)|$$

라 한다면 변화시점에 대한 점추정량을 다음과 같이 생각할 수 있다.

$$\widehat{c_1}=Min\{j:D_j=Max_{1\le c\le n-1}D_c\}.$$

위 점추정량과 다음으로 나오는 점추정량의 Min{.}의 표현은 D_c의 최대값이 같은 경우가 여러개 나타날 수 있는데 이들중 c가 가장 작은값을 추정치로 택한다.

4.2 우도비 추정량

다음으로 모집단이 이변량정규분포를 따르고 변화시점을 알고 있을 때, 귀무가설 H_0: δ = 0에대한 우도비 검정은 c개의 표본과 나머지 (n-c)개의 이표본 Hotelling T^2-검정과 동일하게 된다.그러므로 Hotelling T^2 통계량에 변화시점에 대한 최대값을 이용하여 보기로 한다.

변화시점 이전 c개의 표본 확률벡터들의 평균벡터를 $\overline{X}_c$, 변화시점 이후 (n-c)개의 평균벡터를 $\overline{X}_{n-c}$,그리고 합친 두 표본의 분산(pooled variance)을

$$S_p^2 = \frac{\sum_{j=1}^{c}(X_j - \overline{X}_c)(X_j - \overline{X}_c)^T + \sum_{j=c+1}^{n}(X_j - \overline{X}_{n-c})(X_j - \overline{X}_{n-c})^T}{n-2}$$

라 한다면, 그때 변화시점에 대한 우도비 추정량은

$$\widehat{c_2} = Min\{j: T^2_{j,n-j} = Max_{1\le c\le n-1} T^2_{c,n-c}\},$$

으로 정의된다. 여기에서

$$T^2_{c,n-c} = (\overline{X}_{n-c} - \overline{X}_c)^T \left[\left(\frac{1}{c} + \frac{1}{n-c}\right)S_p\right]^{-1}(\overline{X}_{n-c} - \overline{X}_c).$$

4.3 순위-모양 추정량

2.2절의 순위-모양 검정통계량을 이용한다면 다음의 점추정량을 생각할 수 있다.

$$\widehat{c_3} = Min\{j: u^2_{j,n-j} = Max_{1\le c\le n-1} u^2_{c,n-c}\},$$

여기에서 $u^2_{c,n-c}$는 (2.2.1)에 나타나 있다.

위 세가지 변화시점에 대한 점추정량의 경험적인 평균제곱오차를 구하기 위하여 변화시점 이후 변화된 양은 표준편차의 0.5배,1.0배 만큼 변화를 주었다. 나머지의 실험조건은 3절의 실험조건과 동일하며 그 결과가 표<4.1>에 나와 있다. <표4.1>에서 괄호 안의 값은 추정치와 참값과의 편의(bias)에 대한 절대값이다.각 칸의 첫번째 값은 $\widehat{c_1}$,두번째 값은 $\widehat{c_2}$ 마지막 값은 $\widehat{c_3}$의 평균제곱오차를 나타낸다. 변화된 양이 표준편차 의 0.5배,즉 변화가 적을 경우는 $\widehat{c_3}$이 바람직하다. 그러나 대부분의 경우, 효율적인 점추정량은 경험적인 분포함수를 이용한 Carlstein의 $\widehat{c_1}$임을 알 수 있다.

5.토 의

본 연구에서 제시된 이변량 변화시점 모형의 위치모수에 대한 검정법은 공간 중앙값을 이용한 비모수적 방법인 순위-모양 검정이 모수적 방법보다 비교적 우수하게 나타남을 알 수 있었다.앞으로 더 연구되어야 할 과제로서, 미지의 변화시점에 대한 점 추정량의 성질을 생각하고 위치모수에 대한 구간 추정을 Bootstrap 방법 혹은 다른 방법을 이용해 추정해 보아야 할 것이다.또한 척도모수에 대한 추론과 변화시점이 여러개 있는 모형도 아울러 연구되어야 할 과제로 사료된다.

<표4.1> 변화시점에 대한 점추정량의 MSE

c=5

	p	0.5		0.75		1.0	
σ^2	δ_2 \ δ_1	0.5	1.0	0.5	1.0	0.5	1.0
25	0.5	113.0(6.5)	76.1(4.6)	139.0(7.9)	92.0 (5.4)	124.4 (7.5)	100.5 (5.9)
		282.8(11.7)	244.0(9.6)	260.0(10.8)	234.3 (9.4)	246.9 (8.7)	208.5 (7.0)
		95.6(5.7)	81.9(4.8)	128.1(7.5)	94.7 (6.0)	155.1 (8.5)	128.1 (7.2)
	1.0	70.3(4.4)	39.2(2.8)	92.8(5.5)	65.0 (4.1)	97.8 (6.0)	82.0 (5.0)
		245.3(9.7)	209.6(7.8)	231.6(9.3)	208.4 (7.9)	210.1 (7.0)	163.1 (4.8)
		80.9(4.7)	44.6(2.9)	94.8(6.0)	62.5 (4.4)	144.6 (8.1)	117.1 (6.8)
100	0.5	101.5(5.8)	74.6(4.5)	94.0(5.5)	62.3 (3.9)	124.4 (7.5)	100.5 (5.9)
		283.0(11.9)	250.5(10.0)	251.6(10.9)	225.1 (9.5)	246.9 (8.7)	208.5 (7.0)
		59.1(3.8)	49.5(3.3)	73.8(5.0)	37.5 (3.2)	155.1 (8.5)	128.1 (7.2)
	1.0	61.2(3.9)	65.5(4.1)	63.7(3.9)	40.4 (2.9)	97.8 (6.0)	82.0 (5.0)
		249.6(10.0)	211.4(8.0)	227.6(9.6)	208.4 (8.4)	210.1 (7.0)	163.1 (4.8)
		37.1(2.7)	33.8(2.5)	41.1(3.2)	20.7 (2.0)	144.6 (8.1)	117.1 (6.8)

c=15

25	0.5	46.7(0.6)	25.0(0.3)	49.7(0.2)	26.4 (0.1)	55.4 (0.5)	31.1 (0.2)
		139.4(3.2)	135.6(2.9)	134.9(2.0)	132.3 (2.0)	165.6 (0.1)	156.3 (0.2)
		46.3(1.7)	26.3(1.0)	58.4(1.2)	39.5 (1.9)	77.6 (0.9)	62.2 (1.8)
	1.0	28.3(0.6)	21.7(0.0)	31.1(0.4)	20.4 (0.6)	30.7 (0.2)	25.1 (0.5)
		135.6(2.9)	129.4(2.7)	131.7(2.0)	129.4 (2.1)	156.7 (0.1)	150.0 (0.2)
		26.5(1.4)	16.3(1.0)	39.7(1.8)	24.9 (1.1)	55.7 (1.7)	47.3 (1.8)
100	0.5	20.4(0.3)	15.1(0.4)	33.1(0.2)	9.7 (0.3)	55.4 (0.5)	31.1 (0.2)
		133.9(3.7)	130.9(3.3)	125.3(2.3)	123.7 (2.2)	165.6 (0.1)	156.3 (0.2)
		21.8(1.3)	15.1(0.6)	28.9(0.5)	10.2 (0.5)	77.6 (0.9)	62.2 (1.8)
	1.0	15.2(0.1)	10.6(0.2)	9.1(0.1)	6.5 (0.1)	30.7 (0.2)	25.1 (0.5)
		130.4(3.3)	124.6(3.0)	124.7(2.2)	121.3 (2.1)	156.7 (0.1)	150.0 (0.2)
		12.9(0.7)	12.9(0.7)	7.7(0.3)	2.0 (0.1)	55.7 (1.7)	47.3 (1.8)

참 고 문 헌

[1] Brown, B. M. (1983), Statistical Uses of the Spatial Median, *Journal of the Royal Statistical Society*, B, Vol. 45, No.1, 25-30.

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

[3] Gower, J. C. (1974), The mediancentre, *Applied Statistics*, Vol. 23, 466-470.

[4] Johnson, R. A. & Wichern, D. W. (1982). *Applied Multivariate Statistical Analysis*, Prentice-Hall Inc..

[5] Lombard, F. (1983), Asymptotic distributions of rank statistics in the change point problem, *South African Statistical Journal*, Vol. 17, 83-105.

[6] Lombard, F. (1986), The Change-Point Problem for Angular Data: A Nonparametric Approach, *Technometrics*, Vol. 28, 391-397.

[7] Mardia, K. V. (1972), *Statistics of Directional Data*, New York, Academic Press.

[8] Moses, L. E. (1963), Rank tests of dispersion, *Annals of Mathematical Statistics*, Vol. 34, 973-983.

[9] Page, E. S. (1954), Continuous inspection schemes, *Biometrika*, Vol. 41, 100-115.

[10]Randles, R. H. & Wolfe, D. A. (1979), *Introduction to the Theory of Nonparametric Statistics*, Wiley, New York, 355-360.

[11]Sen, A. K. & Srivastava, M. S. (1973), On Multivariate Tests for Detecting Change in Mean, *Sankhya*, Vol. 35, 173-186.

[12]Shaban, S. A. (1980), Change-Point Problem and Two-Phase Regression: An Annotated Bibliography, *International Statistical Review*, Vol. 48, 83-93.

[13]Wolfe, D. A. & Schechtman, E. (1984), Nonparametric Statistical Procedures for the Change-Point Problem, *Journal of Statistical Planning and Inference*, Vol. 9, 389-396.

Nonparametric Test Procedure for the Bivariate Changepoint Model[3)]

Kim, Kyung Moo[4)]

Abstract

We propose the nonparametric rank-like test for the location parameter in the bivariate changepoint model. Empirical powers between the parametric test and nonparametric test are compared. These results show that rank-like test is better than parametric method except bivariate normal null distribution. The point estimators for the changepoint are also compared by the empirical mean squared errors.

3) This paper was supported in part by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1992.

4) (713-714) Department of Statistics, Taegu University, Kyungsan, Kyungbuk, Korea.