한국통계학회논문집 제2권 1호
1995년 4월 pp. 185-193

확률화 블럭 모형에서 우산형 대립가설에 대한 "Near-Matches"의 갯수를 이용한 순위검정법에 관한 연구

김 동 희[1], 김 영 철[1]

요 약

b개의 블럭과 m개의 처리를 가지는 확률화 블럭 모형에서, 블럭내 순위들 사이의 "near-matches"의 갯수를 이용하여 처리효과의 우산형 대립가설에 대한 순위검정법을 제안하고자 한다.

b가 고정되어 있고 처리 m의 갯수를 무한히 하면, Wilcoxon score에 대한 블럭내 순위들을 근거로 해서 제안된 검정법과 수정된 W-검정법과의 점근상대효율을 구하였고, 또한 소표본에서 Monte Carlo 연구를 통하여 제안된 통계량과 수정된 W-통계량을 비교 연구하였다.

1. 서 론

X_{ijk} , $k = 1, \cdots, n_{ij}$; $j = 1, \cdots, m$; $i = 1, \cdots, b$를 b개의 블럭과 m개의 처리를 갖고 각 칸의 반복수가 n_{ij} 인 확률화 블럭 모형에서의 관측값이라 하자. 그리고 i-번째 블럭의 j-번째 처리에서의 관측값 X_{ijk}, $k=1, \cdots, n_{ij}$은 미지의 누적분포함수 F_{ij}를 가진다고 가정하자. $F_{ij}(t) = F(t-\beta_j-\alpha_i)$ 라 하고 β_j는 처리효과를 나타내고 α_i는 블럭효과를 나타낸다.

본 논문에서는 처리 효과의 우산형 대립가설의 검정문제에 관심이 있다.

$$H_0 \ : \ \beta_j = 0, \quad \text{모든 } j \text{에 대해서}$$

$$vs \ \ H_1 \ : \ \beta_1 \le \beta_2 \le \cdots \le \beta_h \ge \beta_{h+1} \ge \cdots \ge \beta_m \ .$$

(적어도 한 부등식은 만족)

위의 검정문제에서, 본 논문에서는 "Near-Matches"의 방법을 이용하고자 하는데 "Near-Matches"는 $n_{ij} = 1$ 일 때, 블럭내에서의 각 칸의 관측값들의 순위 R_{ij}와 자연순위 $c_j = j$ 와의 부합성을 고려하는 것으로 $\sum_{i=1}^{b} \sum_{j=1}^{m} I(|R_{ij}-c_j| \le k_{j,m})$ 에서 $k_{j,m} \ge 1$인 경우를 말하며

1) (609-735) 부산시 금정구 장전동 부산대학교 통계학과

$k_{j,m}=0$ 인 특수한 경우를 "Perfect Match"라 한다. 이 "Near-Matches"의 방법을 이용한 순서대립가설에 대한 비모수적 검정법으로 Jammalamadaka, Tiwari, Sen, and Ebneshahrashoob(1991)의 검정법을 들 수 있다. 본 논문에서는 $n_{ij}>1$일 때, 블럭내에서 각 칸의 평균값들의 순위와 정점이 주어진 경우의 동위회귀함수의 한 형태와의 부합성을 고려한 우산형대립가설에 대한 검정법을 소개하고자 한다.

2. 제안된 통계량과 점근분포

다루고자 하는 모형은 다음과 같이 표현될 수 있다.

$$X_{ijk} = Y_{ijk} + \beta_j + \alpha_i, \quad k=1,\cdots,n_{ij}\,;\ j=1,\cdots,m\,;\ i=1,\cdots,b.$$

여기서 Y_{ijk}는 서로 독립이고 동일한 분포를 갖는 확률변수이며 분포함수 F(밀도함수 f 존재)를 갖는다.

R_{ij}는 $\{X_{i1\cdot},\cdots,X_{im\cdot}\}$ 에서 $X_{ij\cdot}$의 순위라 하고, $X_{ij\cdot}$는 (i,j)번째 칸의 평균, 즉 $X_{ij\cdot}=\sum_{k=1}^{n_{ij}} X_{ijk}/n_{ij}$, $j=1,\cdots,m,\,;\ i=1,\cdots,b$ 이라 하자.

본 논문에서 제안하고자 하는 통계량은 다음과 같고,

$$\begin{aligned} M &= \sum_{i=1}^{b}\sum_{j=1}^{m} I(|R_{ij}-c_j|\le k_{j,m}) \\ &= \sum_{i=1}^{b}\sum_{j=1}^{m} I\left(b_{mj}\le \frac{R_{ij}}{m}\le a_{mj}\right) \end{aligned}$$

a_{mj}와 b_{mj}는 다음과 같이 주어진다.

$$a_{mj} = \frac{1}{m}\min(m,\ c_j+k_{j,m})$$
$$b_{mj} = \frac{1}{m}\max(0,\ c_j-k_{j,m})$$

점근상대효율을 계산하기 위해 다음과 같은 축차 대립가설 H_{1m}을 고려하는데

$$H_{1m}:\ X_{ijk} = Y_{ijk} + \beta_{mj} + \alpha_i, \quad k=1,\cdots,n_{ij}\,;\ j=1,\cdots,m\,;\ i=1,\cdots,b \tag{2.1}$$

이고 $\max\{|\beta_{mj}| : 1 \le j \le m\} = O(m^{-1/2})$ 이다. (2.2)

그리고 $\beta_{mj} = \phi_m \cdot \beta_j$, $\phi_m = \phi / \sqrt{m}$ 이고 ϕ는 임의의 상수이며, (2.1)식의 β_{mj}는 반드시 등간격일 필요는 없다.

$F(y)$를 Y_{11} 의 실제 분포함수라 할 때, H_{1m}하에서 $X_{ij\cdot}$의 분포함수를 $F(y-\beta_{mj})$, $j=1,\cdots,m$ 이라 하고 $F_{mi}(y) = \sum_{j=1}^{m} I(X_{ij\cdot} \le y)/m$ 를 $X_{i1\cdot},\cdots,X_{im\cdot}$의 경험분포함수라 하자. 그러면 통계량 M은 다음과 같이 표현될 수 있다.

$$M = \sum_{i=1}^{b}\sum_{j=1}^{h} I(|m \cdot F_{mi}(X_{ij\cdot}) - c_j| \le k_{j,m})$$

또한, $\overline{F}_{(m)}(y) = \sum_{j=1}^{m} F(y-\beta_{mj})/m$를 평균분포함수라 할 때, 다음과 같은 가정을 한다.

가정 1. F 는 유계인 이차 도함수 f 를 가진다.

Taylor전개와 (2.2)식에 의해, $0<\theta<1$에 대해서, $\overline{F}_m(y)$는 다음과 같이 전개된다.

$$\begin{aligned}
\overline{F}_{(m)}(y) &= \frac{1}{m}\sum_{j=1}^{m} F(y-\beta_{mj}) \\
&= \frac{1}{m}\sum_{j=1}^{m}\left\{ F(y) - \beta_{mj} f(y) + \frac{1}{2}\beta_{mj}^2 f(y-\theta\cdot\beta_{mj})\right\} \\
&= F(y) + \frac{1}{2m}\sum_{j=1}^{m}\beta_{mj}^2 f(y-\theta\cdot\beta_{mj}), \quad 0<\theta<1 .
\end{aligned}$$

따라서,

$$\sup|\overline{F}_{(y)} - F(y)| \le \frac{1}{2m}\{\sup_y |f(y)|\} = O(m^{-1}) . \tag{2.3}$$

$$m^{1/2} \| F_{mi} - F \| = \sup_y\{m^{1/2}|F_{mi}(y) - F(y)|\} = O_p(1) .$$

또한, $k_{j,m}/m = k_m(c_j/(m+1))$, $j=1,\cdots,m$ 으로 정의하고 다음의 두 조건을 만족한다.

(i) $k_m(u)$는 구간 $\left(\frac{c_j-1}{m+1}, \frac{c_j}{m+1}\right]$, $j=1, \cdots, m$ 에서 상수이고,

(ii) $u \in (0,1)$에 대해, $0 < \int 2k(u)\ [1-2k(u)]\ du < \infty$의 가정을 만족하면서 $m \to \infty$ 일 때 $k_m(u) \ \to \ k(u)$ 이다.

정리 2. 1. (Jammalamadaka, Tiwari, Sen and Ebneshahrashoob, 1991)
H_{1m}, (2.2)식과 가정 1 에서 $m \to \infty$ 일 때,

$$\frac{M - b(\mu_m^0 + \Delta_m)}{\sqrt{bm} \cdot \sigma} \xrightarrow{d} N(0,1)\ .$$

위 식에서 μ_m^0, Δ_m, 그리고 σ^2은 다음과 같다.

$$\begin{aligned}
\mu_m^0 &= \sum_{j=1}^{m} (a_{mj} - b_{mj})\ , \\
\Delta_m &= -\sum_{j=1}^{m} d_{mj}\ [f(F^{-1}(a_{mj})) - f(F^{-1}(b_{mj}))]\ , \\
d_{mj} &= \frac{1}{\sqrt{m}}\, d\left(\frac{c_j}{m+1}\right), \quad j=1, \cdots, m\ , \\
\sigma^2 &= \int_0^1 2k(u)\ [1-2k(u)]\ du\ .
\end{aligned} \tag{2.4}$$

점수함수 $\psi(u)$ 를 다음과 같이 두자.

$$\Psi(u) = -\frac{f'\left(F^{-1}\left(\frac{c_j}{m+1}\right)\right)}{f\left(F^{-1}\left(\frac{c_j}{m+1}\right)\right)}\ .$$

가정 2. $\psi(u)$는 연속이고 적분가능함수에 의해 절대값이 유계이다.

$m \to \infty$ 일 때, $\Delta_m / \sqrt{m}\,\sigma$ 는 다음과 같이 주어진다.

$$\frac{\Delta_m}{\sqrt{m}\,\sigma} \to \frac{\sqrt{2}\int_0^1 d(u)\,k(u)\,\Psi(u)\,du}{\left[\int_0^1 k(u)\,(1-2k(u))\,d(u)\right]^{1/2}} .$$

위의 수식에서 다음의 보조정리를 얻는다.

따름정리 2. 2. (2.2)식과 가정 1 , 2 하에서, $m \to \infty$ 일 때,

$$\frac{M - b\mu_m^0}{\sqrt{bm}\,\sigma} \xrightarrow{d} N\left(\frac{\sqrt{2b}\int_0^1 d(u)\,k(u)\,\Psi(u)\,du}{\left[\int_0^1 k(u)(1-2k(u))\,du\right]^{1/2}} , 1 \right) .$$

$c_j - k \le R_{ij} \le c_j + k$로 *match*를 정의한 $k = k_{j,m}$의 특별한 경우를 고려해 보자. 그러면, a_{mj}와 b_{mj}는 다음과 같다.

$$a_{mj} = \frac{1}{m} \min(m, c_j + k)$$
$$b_{mj} = \frac{1}{m} \max(0, c_j - k)$$

만약에 $k/m \to p(0 < p < 1/2)$ 이면, 극한 윈도우 함수 $\lim_{m\to\infty}(a_{mj} - b_{mj})/2$를 다음과 같이 대응시킨다.

$$k(u) = \begin{cases} \frac{1}{2}(u+p) , & 0 < u \le p \\ p , & p \le u \le 1-p \\ \frac{1}{2}(1-u+p) , & 1-p \le u < 1 \end{cases}$$

$$\sigma_p^2 = \int_0^1 2k(u)\,[1-2k(u)]\,du = p(10p^2 - 15p + 6)/3$$ 이다.

따라서, 다음의 보조정리를 얻는다.

따름정리 2. 3. 정리 2.1 의 가설하에서, $k_{m,j} = k_m = [mp]$, $o < p < 1/2$ 에 대해, $m \to \infty$ 일 때,

$$\frac{M - b\,[(2k_m+1)+\Delta_m]}{\sqrt{bm}\,\sigma_p} \xrightarrow{d} N(0\,,1)$$

이고, f 가 가정 2를 만족한다면, $m \to \infty$일 때,

$$\frac{M - b\ [2k_m+1]}{\sqrt{bm}\,\sigma_p} \xrightarrow{d} N\left(\frac{2\sqrt{b}\int_{-1}^{1} d(u)\,k(u)\,\Psi(u)\,d(u)}{\sigma_p}\,,\,1\right)$$

이다.

3. 점근상대효율

i-번째 블럭내의 각 칸의 평균들의 순위의 부합성을 비교하는 것으로 Page(1963)의 통계량을 수정한 통계량은 다음과 같다.

$$W = \sum_{i=1}^{b}\sum_{j=1}^{m}(c_j - \bar{c})\left(R_{ij} - \frac{m+1}{2}\right).$$

여기서, c_j는 동위회귀함수의 한 형태이고 $\bar{c}$ 는 다음과 같다.

$$\bar{c} = \frac{1}{m}\sum_{j=1}^{m} c_j = \frac{m+1}{2}\ .$$

$c_j = j$ 인 경우에 통계량 W는 Page(1963) 통계량이 되고, Pirie(1974)는 b를 고정시키고 m을 무한히 했을 때, Hollander(1966) 검정법에 대한 Page(1963) 검정법의 점근상대효율을 구하였다. 또한, Pirie and Hollander(1972)와 Pirie(1985)의 검정법들은 Page(1963) 검정법의 정규점수화를 고려했다.

$d(u) = \sqrt{12}\,(u - 1/2)$ 또는 $d_j = c_j\,/\,m$ 에 대응하는

$$d_{mj} = \frac{\sqrt{12}\,(\,c_j - (\,m+1)/2)}{\sqrt{m(m^2-1)}}$$

를 고려하면, H_{1m}에서 W의 기대값은 다음과 같다.

$$\begin{aligned} E\{W|H_{1m}\} &= \sum_{i=1}^{b}\sum_{j=1}^{m}(c_j-\overline{c})E(R_{ij}|H_{1m}) \\ &= b\sum_{j=1}^{m}(c_j-\overline{c})\left[1+\sum_{l=1,\neq j}^{m}E\{I(X_{il\cdot}<X_{ij\cdot})|H_{1m}\}\right] \\ &= b\sum_{j=1}^{m}(c_j-\overline{c})\left[\frac{m+1}{2}+md_{mj}\int f^2(x)\,dx+O(m^{-1/2})\right] \\ &= bm\left(\int f^2(x)\,dx\right)\sqrt{\frac{m(m^2-1)}{12}}+O(m^{-1/2}) \end{aligned}$$

또한, 귀무가설에서 W의 기대값과 분산(Pirie(1974))은

$$E(W|H_0)=0$$

$$Var(W|H_0)=\frac{bm^2(m-1)(m+1)^2}{144}$$

이므로, W의 "*efficacy*" 는 다음과 같다.

$$\begin{aligned} eff(W) &= \lim_{m\to\infty}\frac{b^2m^2\left(\int f^2(x)\,dx\right)^2\frac{m(m^2-1)}{12}}{\frac{bm^2(m-1)(m+1)^2}{144}} \\ &= 12b\left(\int f^2(x)\,dx\right)^2 \end{aligned}$$

또한, M의 "*efficacy*" 는

$$eff(M)=\frac{2b\left\{\int_0^1 d(u)\Psi(u)k(u)\,du\right\}^2}{\int_0^1 k(u)\,[1-2k(u)]\,du}$$

이므로, W에 대한 M의 점근상대효율은 다음과 같이 주어진다.

$$e(M, W) = \frac{\left[\int_0^1 d(u) k(u) \Psi(u)\, du \right]^2}{6\,(f^2(x)\, dx)^2 \left(\int_0^1 k(u)\ [1-2k(u)]\ du \right)}$$

$$= \frac{2\ \left[\int_0^1 \left(u - \frac{1}{2} \right) \Psi(u) k(u)\, du \right]^2}{\left(\int f^2(x)\, dx \right)^2 \left(\int_0^1 k(u)\ [1-2k(u)]\ du \right)}$$

$$= \frac{2\ \left[\frac{1}{2}\left(\frac{1}{6} p^3 - \frac{1}{4} P^2 \right) \frac{f'(F^{-1}(p))}{f(F^{-1}(p))} + \frac{1}{2}\left(-\frac{1}{6} p^3 + \frac{1}{4} p^2 - \frac{1}{12} \right) \frac{f'(F^{-1}(1-p))}{f(F^{-1}(1-p))} \right]^2}{\left[\int f^2(x)\, dx \right]^2 \left(\frac{1}{6}\, p(10p^2 - 15p + 6) \right)}$$

4. 소표본 Monte Carlo 연구

소표본에서, 제안된 통계량 M을 통계량 W와 비교하기 위하여 세가지 분포들(정규, 이중지수, 코시)을 사용해서 Monte Carlo 연구를 하였다. 모의 실험에서는 4개의 블럭과 5개의 처리를 가지는 확률화 블럭 모형을 고려했다. 각 칸내의 반복횟수의 형태는 다음과 같다.

경우1 : $n_{i1} = 3,\ n_{i2} = 4,\ n_{i3} = 5,\ n_{i4} = 6,\ n_{i5} = 7,\ i = 1,2,3,4$

경우2 : $n_{ij} = 5,\ i = 1,2,3,4\ ;\ j = 1,2,3,4,5$

처리효과의 형태는 $(-\delta\sigma/3,\ \delta\sigma/3,\ \delta\sigma,\ \delta\sigma/3,\ -\delta\sigma/3)$ 이다.

여기서 δ는 0.0에서 1.0까지 0.2씩 증가하며 σ는 각 모집단의 표준편차이다. Cauchy분포의 경우는 이차적률이 존재하지 않기 때문에 표준정규분포의 표준편차에 대응하는 σ를 선택한다. 즉, $\int_{-\sigma}^{\sigma} (\pi(1+x^2))^{-1}\, dx = 0.6827 = \Phi(1) - \Phi(-1)$. Φ는 표준정규분포함수이므로 σ의 값은 1.8326 이다. 각 모형과 분포함수에 대하여 주어진 크기의 표본을 IMSL에 있는 부프로그램을 사용하여 생성하고, 이 표본에 대하여 모든 검정통계량의 값을 계산한 다음에, 검정통계량의 값과 유의수준 5%의 기각값을 비교한다. 정점이 3일 때, $c_1 = 1,\ c_2 = 3,\ c_3 = 5,\ c_4 = 4,\ c_5 = 2$로 주어진 경우의 검정통계량의 유의수준과 검정력들이 표 1 에 수록되어 있다. 표 1 에서 볼 수 있듯이 통계량 M의 검정력은 p가 증가함에 따라 감소하는 경향을 나타내고 있으며, 모든 검정통계량들의 검정력들이 경우 1보다는 경우 2에서 다소 높음을 알 수 있다.

<표 1> 검정통계량들의 실험검정력($\alpha = 0.05$)

	$\delta =$	0.0	0.2	0.4	0.6	0.8	1.0	0.0	0.2	0.4	0.6	0.8	1.0
		경우 1						경우 2					
정 규 분 포	W	55	209	466	726	903	967	47	186	491	790	912	990
	M_1	42	146	342	579	783	915	33	134	350	608	804	920
	M_2	28	126	308	531	732	869	36	104	323	584	763	901
이중지수분포	W	43	186	461	738	910	971	51	209	497	781	932	988
	M_1	36	130	338	604	835	919	31	159	387	639	834	939
	M_2	29	116	321	553	756	876	34	126	313	571	776	904
Cauchy 분 포	W	50	104	181	290	386	478	47	112	184	284	367	473
	M_1	49	78	166	243	316	414	45	82	130	235	327	439
	M_2	24	57	111	158	225	291	20	66	109	173	228	285

M_1 과 M_2는 $p = 0.3$ 과 $p = 0.4$인 경우의 통계량 M을 각각 나타낸다.

참 고 문 헌

[1] Hajek, J. and Sidak, Z. (1967). *Theory of rank tests*, New York : Academic Press.

[2] Hollander, M. (1966). Rank tests for randomized blocks when the alternatives have an a priori ordering, *Annals of Statistics*, 38, 867-877.

[3] Jammalamadaka, S. Rao,Tiwari, Ram C., Sen, P. K., and Ebneshahrashoob, M. (1991). A rank test based on the number of "near-matches" for ordered alternatives in randomized blocks., *The Indian Journal of Statistics*, 53, 183-193.

[4] Page, E. B. (1963). Ordered hypotheses for multiple treatments : a significance test for linear ranks, *Journal of the American Statistical Association*, 58, 226-230.

[5] Pirie, W. R. and Hollander, M. (1972). A distribution-free normal scores test for ordered alternatives in the randomized block design, *Journal of the American Statistical Association*, 67, 855-857.

[6] Pirie, W. R. (1974). Comparing rank tests for ordered alternatives in randomized blocks, *Annals of Statistics*, 2, 374-382.

[7] Pirie, W. R. (1985). Page test for ordered alternatives, In : S. Kotz and Johnson, N. L., eds., *Encyclopedia of Statistical Sciences*, (S. Kotz and N. L. Johnson, eds.) Vol. 6. New York : Wiley.