응용통계연구 제 **14**권 *1*호, *2001*년, *pp. 191-199*

불완전 순위 자료를 위한 몬테칼로 임의순열 검정

허명회[1] 최원[2]

요 약

본 소고는 n명의 심사자가 k개의 객체를 평가하여 얻어진 불완전 순위자료에서 객체간 선호도에 있어 차이가 없다는 영가설을 검정하는 방법에 관한 연구이다. 주어진 자료에서 결측값들을 다중대체하는 방식을 제안하고 이들을 평균 p-값으로 묶는 몬테칼로 방식의 임의순열 검정을 제안한다.

주요용어: 불완전 순위자료, 결측자료, 다중대체, 몬테칼로 방법, 임의순열 검정.

1. 서론

n명의 심사자(judges)가 k개의 객체(objects)를 평가하여 순위를 매기는 경우를 생각하기로 하자. 이 때, 원칙적으로 심사자들은 가장 좋게 생각하는 객체에 순위 1을, 그 다음으로 좋게 생각하는 객체에 순위 2를, $\cdots$, 그리고 마지막으로 남은 객체에 순위 k를 부여하기로 되어 있다고 하자. 그러나 실제에 있어 심사자들은 좋게 생각하는 처음 몇 객체에 대하여는 순위를 제대로 부여하였지만 나머지 객체들에 대하여는 순위를 부여하지 못하여 불완전한 순위자료(incompletely ranked data)가 얻어졌다고 하자. 표 1.1의 (a)가 한 예이다 ($n = 6, k = 5$). 이 자료 사례에서 예컨대 심사자 J1은 객체 Ob3에 순위 1을, 객체 Ob1에 순위 2를 부여하였지만 나머지 객체들인 Ob2, Ob4, Ob5에 대하여는 순위를 부여하지 못하였다. 이렇게 순위화에 있어 결측이 발생하는 주된 이유는 후순위 개체들에 대하여는 심사자들의 분별력이 떨어지는 경향이 있기 때문이다.

본 소고는 객체들 사이에 선호도 차이가 있는지를 검정하는 방법에 관한 연구이다. 이를 위하여 관측자료값을 다음과 같이 표기한다.

$$R_{ij} = \text{심사자 } i\text{가 객체 } j\text{에 부여한 순위값 } (i = 1, \cdots, n; j = 1, \cdots, k).$$

그리고 "k개 객체간 선호도에 있어 차이가 없다"를 영가설 H_0라고 하고 이에 대한 일반적 부정인 대안가설 H_1이라고 하자.

표본자료 $\{R_{ij} | i = 1, \cdots, n, j = 1, \cdots, k\}$에 결측이 없는 경우, Friedman 검정이 대표적 방법이며 통계량은 다음과 같다.

1) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 정경대학 통계학과, 교수
E-mail: stat420@mail.korea.ac.kr
2) (150-010) 서울시 영등포구 여의도동 35-4, d2K 솔루션, 컨설턴트
E-mail: woniccu@d2ksolutions.com

표 1.1: Lordo and Wolfe (1994)의 불완전 순위자료 (a)*와 평균순위대체 (b)**

(a)		객 체				
		Ob1	Ob2	Ob3	Ob4	Ob5
심사자	J1	2	-	1	-	-
	J2	-	1	-	-	2
	J3	1	-	3	2	-
	J4	1	3	4	2	5
	J5	1	-	-	-	-
	J6	3	2	5	1	4

(b)		객 체				
		Ob1	Ob2	Ob3	Ob4	Ob5
심사자	J1	2	**4**	1	**4**	**4**
	J2	**4**	1	**4**	**4**	2
	J3	1	**4.5**	3	2	**4.5**
	J4	1	3	4	2	5
	J5	1	**3.5**	**3.5**	**3.5**	**3.5**
	J6	3	2	5	1	4

* '-'는 결측을 나타냄.

** 굵은 숫자는 대체된 평균순위값.

$$S = \frac{12n}{k(k+1)} \sum_{j=1}^{k} (\overline{R}_{.j} - \overline{R}_{..})^2,$$

여기서 $\overline{R}_{.j} = \frac{1}{n}\sum_{i=1}^{n} R_{ij}$, $\overline{R}_{..} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} R_{ij}$이다.

n과 k가 작은 경우 Friedman 통계량 S의 영분포에 대하여는 Hollander and Wolfe (1973)를 참조할 수 있다. 그리고, n이 큰 경우에는 S가 근사적으로 자유도 $k-1$의 카이제곱 분포를 따른다는 수리적 성질이 알려져 있다. 이론적으로 S의 영분포(null distribution)는 순열검정(permutation test)의 방법을 적용하여 얻어질 수 있으며, 관측된 S의 p-값 정도에 대하여는 몬테칼로(Monte Carlo) 방법으로 임의순열(random permutations)을 발생시켜 훌륭한 근사값을 계산해낼 수 있다 (Good 1994, p.153).

표본자료$\{R_{ij} | \, i = 1, \cdots, n, j = 1, \cdots, k\}$에 부분적 결측이 있는 경우, 단순한 대처방법은 결측값을 잔여 순위들의 평균으로 대체하는 것이다. 예컨대 표 1.1 (a)의 자료에서 심사자 J1이 순위를 부여하지 않은 객체들인 Ob2, Ob4, Ob5에 대하여 남은 순위들인 3, 4, 5의 평균인 4를 주는 것이다. 표 1.1의 (b)를 보라. 이런 결측값 처리를 거쳐 Friedman 통계량 S를 계산함으로써 영가설 H_0를 검정할 수 있다. 그러나 S의 영분포가 이런 결측값 처리에 의하여 영향받는다는 점을 고려해야 한다. 때문에, 몬테칼로 방법에 있어서는 평균순위로 대체된 자료값들을 임의순열화하여 S의 영분포를 산출하여야 한다 (최원, 2000).

이러한 평균순위(Average Ranks) 대체(imputation) 외에 Lordo and Wolfe (1994), Lim (1998)은 Summing Ranks 대체, Counting Ranks 대체와 아울러 상응하는 Friedman 검정법을 제안하였다.

한편, Lim, Lordo and Wolfe (1999)와 임동훈 (1999)은 일부 결측된 순위에 대한 모든 대체로부터의 Friedman 통계량 S 중에서 최대값 S_{max}와 최소값 S_{min}를 활용한 다음의 검정법을 제안하였다.

- $S_{min} \geq s(\alpha, k, n)$ 또는 $(S_{min} + S_{max})/2 \geq s(\alpha, k, n)$이면 H_0를 기각하고,

표 2.1: 심사자 J1과 J3의 완전순위 대체

(a) J1	객 체					
	Ob1	Ob2	Ob3	Ob4	Ob5	확률
경우 1	2	**3**	1	**4**	**5**	1/6
경우 2	2	**3**	1	**5**	**4**	1/6
경우 3	2	**4**	1	**3**	**5**	1/6
경우 4	2	**4**	1	**5**	**3**	1/6
경우 5	2	**5**	1	**3**	**4**	1/6
경우 6	2	**5**	1	**4**	**3**	1/6

(b) J3	객 체					
	Ob1	Ob2	Ob3	Ob4	Ob5	확률
경우 1	1	**4**	3	**2**	**5**	1/2
경우 2	1	**5**	3	**2**	**4**	1/2

- $S_{max} \leq s(\alpha, k, n)$ 또는 $(S_{min} + S_{max})/2 \leq s(\alpha, k, n)$이면 H_0를 기각하지 않는다.

여기서 $s(\alpha, k, n)$은 Friedman 통계량 S의 영분포에서 상위 α 분위수이다. 따라서 이 검정법에서는 결측값 패턴에 관계없이 동일한 임계값을 사용한다.

우리는 이 연구에서 S_{max}와 S_{min}에 의한 임동훈의 검정법이 일종의 다중대체(multiple imputation)를 활용하고 있음에 착안하여 다음 절에서 새로운 검정법을 제안하고자 한다.

2. 다중대체방식과 임의순열검정법

순위자료에서의 결측은 해당 객체들에 대한 심사자의 평가가 균등하기 때문에 발생하는 것으로 생각할 수 있다. 예컨대 표 1.1의 자료에서 심사자 J1이 객체 Ob3과 Ob1에 순위 1과 2를 주었으나 객체 Ob2, Ob4, Ob5에는 순위를 부여하지 못하였다. 이것은 심사자 J1이 객체 Ob3과 Ob1에 비해 객체 Ob2, Ob4, Ob5를 덜 선호하지만 이들 3개 비선호 객체들에 대한 선호도가 같음을 드러낸 것이다. 따라서, 만약 심사자 J1에게 5개 객체를 완전순위화하도록 강요한다면 그의 완전순위 반응은 표 2.1의 (a)와 같이 6개 경우중 임의의 1개가 될 것이다. 심사자 J2는 3개 객체에 대한 평가를 유보하였으므로 그의 완전순위 반응은 6(=3!)개 경우중 임의의 1개가 될 것이다. 2개 객체에 대한 평가를 유보한 심사자 J3의 경우엔 2(=2!)개 경우만이 대상이 되고 (표 2.1의 (b) 참조), 4개 객체에 대한 평가를 유보한 심사자 J5의 경우는 24 (=4!)개 경우중 임의의 1개 경우가 될 것이다.

이와 같이 보면 주어진 불완전순위자료로부터 등확률(equal probability)의 N개의 완전순위 대체세트를 구성할 수 있다. 여기서, 심사자 i가 순위를 부여하지 않은 객체의 수를 $f_i(\leq k-1)$라고 할 때, N은 다음과 같다.

$$N = \prod_{i=1}^{n} f_i\,! \,.$$

표 1.1 (a) 자료의 경우에 $N = 3! \times 3! \times 2! \times 0! \times 4! \times 0! = 1,728$이다.

N개 완전순위 대체세트 각각으로부터 Friedman 통계량 S와 p-값을 구함으로써 영가설 H_0를 검정할 수 있지 않을까? 이를 위하여 다음 알고리즘을 제안한다.

알고리즘 I :

1) 불완전순위자료로부터 N개 완전순위 대체세트를 구성하고 각 세트에서 Friedman 통계량 S_r을 산출한다 ($r = 1, \cdots, N$).

2) 각완전순위 대체세트로부터 p-값을 산출한다. 개별 세트에서 구한 p-값을 p_r이라고 하자 ($r = 1, \cdots, N$).

3) 주어진 원 자료에 대한 p-값을 다음과 같이 계산한다.

$$\overline{p} = \sum_{r=1}^{N} p_r / N.$$

이와 같은 관점에서, 임동훈 (1999)의 방법은, 대략적으로 $p_{min} = min(p_1, \cdots, p_N)$과 $p_{max} = max(p_1, \cdots, p_N)$을 조합하는 방법이라고 볼 수 있다. 왜냐하면 p_{min}은 S_{max}에 해당하는 p-값이고 p_{max}는 S_{min}에 해당하는 p-값이기 때문이다.

표 1.1 (a) 자료에 알고리즘 I을 적용한 결과 $\overline{p} = 0.296$이 산출되었다 (참고: $p_{min} = 0.031$, $p_{max} = 0.729$, $s.d. = 0.146$).

실제에 있어 알고리즘 I은 많은 계산을 필요로 할 것이다. 표 1.1 자료와 같이 비교적 간단한 경우에서도 $N = 1,728$번의 p-값 계산을 요구하기 때문이다. 그러나, p-값 계산이 주목적이라면 반드시 N개의 경우를 모두 고려할 필요는 없을 것이다. 대신, 그 중의 임의적 일부인 N_0개의 경우만을 다루는 것으로 충분할 수 있다. 따라서 N이 큰 경우 우리는 다음 알고리즘을 제안한다.

알고리즘 II :

1) 불완전순위자료로부터 구성가능한 N개 완전순위 대체세트 중에서 임의로 N_0개 세트를 복원추출한다. 그리고, 각 세트에서 Friedman 통계량 S_r을 산출한다 ($r = 1, \cdots, N_0$).

2) 각 완전순위 대체세트로부터 p-값을 산출한다. 개별 세트에서 구한 p-값을 p_r이라고 하자 ($r = 1, \cdots, N_0$).

3) 주어진 원 자료에 대한 p-값을 다음과 같이 계산한다.

$$\overline{p} = \sum_{r=1}^{N_0} p_r / N_0.$$

표 1.1 (a) 자료에 $N_0 = 400$개의 완전순위자료 세트를 구성하여 알고리즘 II를 적용한 결과 $\overline{p} = 0.302(s.e. = 0.007)$가 산출되었다 ($p_{min} = 0.041$, $p_{max} = 0.675$, $s.d. = 0.144$).

n이 작은 경우에는 알고리즘 I과 II의 단계 2에서 p-값의 계산시 Friedman 통계량의 영분포 표를 활용할 수 있을 것이고 n이 충분히 큰 경우에는 카이제곱 분포를 쓸 수 있을 것이다. 그러나 n의 대소에 대한 명확한 기준이 있는 것이 아닐 뿐만 아니라 k 값에 따라 상이한 통계표를 일일이 대조하는 작업은 번거로운 일이 아닐 수 없다. 이에 대한 대안으로 몬테칼로 방법으로 Friedman 통계량 S의 근사적 영분포를 산출·활용하는 다음 알고리즘을 제안한다.

알고리즘 III :

1) 주어진 불완전순위자료로부터 구성가능한 N개 완전순위 대체세트 중에서 임의로 N_0개 세트를 복원추출하고 각 세트에서 Friedman 통계량 S_r을 산출한다 ($r = 1, \cdots, N_0$).

2) 각 완전순위 대체세트로부터 p-값을 산출한다. 이 때, 개별 세트에서 p-값 p_r을 구하는 요령은 다음과 같다 ($r = 1, \cdots, N_0$).

   (a) 임의로 $n \times k$ 완전순위자료를 생성시킨다. 여기서 각 행은 $1, \cdots, k$의 임의순열이며 n개 행은 독립적으로 생성된다. Friedman 통계량을 산출하여 S^*로 놓는다.

   (b) 앞 단계를 T번 반복한다. 예컨대 $T = 10,000$이다.

   (c) 이렇게 하여 얻어진 T개의 Friedman 통계값을 $S_1^*, \cdots, S_T^*$라고 할 때,

$$p_r = \sum_{t=1}^{T} 1\{S_t^* \geq S_r\}/T.$$

3) 주어진 원 자료에 대한 p-값을 다음과 같이 계산한다.

$$\overline{p} = \sum_{r=1}^{N_0} p_r/N_0.$$

표 1.1 (a) 자료에 알고리즘 III를 적용한 결과 ($N_0 = 400$, $T = 10,000$), $\overline{p} = 0.308(s.e. = 0.007)$이 산출되었다 ($p_{min} = 0.043$, $p_{max} = 0.693$, $s.d. = 0.147$).

3. 또 다른 사례와 응용

이 절에서는 우리의 방법론이 후순위 객체에 대한 결측 자료의 경우에 뿐만 아니라 선(先)순위와 후(後)순위 객체에 대한 혼합 결측 자료의 경우에도 적용가능함을 한 사례를 통하여 보이고자 한다. Conover (1980, p.301)에서 인용한 표 3.1의 자료를 보라. $n = 12$명의 심사자가 $k = 4$개의 객체에 대한 평가를 순위로 보여주고 있다. 그런데 심사자 J3와 J6의 경우 후순위 객체에 대하여 결측을 남긴 반면 심사자 J9와 J12는 선순위 객체에 대하

표 3.1: Conover (1980)의 불완전 순위자료*

		객 체			
		Ob1	Ob2	Ob3	Ob4
심사자	J1	1	2	3	4
	J2	1	3	2	4
	J3	2	-	-	1
	J4	2	4	3	1
	J5	1	3	4	2
	J6	-	-	-	1
	J7	4	2	3	1
	J8	3	1	4	2
	J9	-	4	3	-
	J10	1	4	2	3
	J11	1	3	2	4
	J12	-	4	3	-

* 원자료에서는 순위 값이 클수록 우선적으로 선호함을 뜻하나
여기서는 일관성을 위하여 역으로 재코딩하였음.

여 결측을 남겼다. 이런 경우, 후순위가 결측된 자료 행에서 남은 순위들 $k, k-1, \cdots$을 임의 배열함으로써 결측값들을 대체시켰듯이, 선순위가 결측된 자료 행에서도 남은 순위들인 $1, 2, \cdots$를 임의배열하여 결측값들을 대체시킴으로써 완전순위 대체세트를 구성할 수 있다.

객체간 차이에 관한 Friedman 검정을 위하여, 2절의 알고리즘 III를 적용하여 본 결과 ($T = 10,000$, $N_0 = 400$), $\overline{p} = 0.054$ (s.e.=0.002), $p_{min} = 0.016$, $p_{max} = 0.106$, $s.d. = 0.030$이 산출되었다.

이 사례에서는 5% 수준에서 객체간 차이가 가까스로 유의하지는 않았으나 다소 느슨한 10% 수준에서는 객체간 차이가 인정된다. 따라서 어느 객체와 어느 객체의 차이가 유의한가를 생각하게 된다. 그것을 튜키(J.W. Tukey)형의 다중비교(multiple comparison)로써 접근하기로 하자. 이를 위하여, 통계량 W를

$$W = max_{p \neq q}|\overline{R}_{.p} - \overline{R}_{.q}|$$

로 정의하자 ($p, q = 1, \cdots, k$). 그러면, 특정 객체 쌍 p와 q의 관측된 절대값 차이 $|\overline{R}_{.p} - \overline{R}_{.q}|$에 관한 p-값은 심사자들의 선호도에서 모든 객체간 차이가 없다는 영가설 H_0 하에서 W의 분포로부터 산출될 수 있다. 표 3.1과 같은 불완전순위자료에 이 방법을 구현하기 위하여는 2장의 알고리즘 III에서 통계량 S를 W로 대치시키기만 하면 될 것이다. 그 결과는 다음과 같다 ($T = 10,000, N_0 = 400$) .

	$\overline{\mathrm{p}}$ (s.e.) p_{min}, p_{max}, $s.d.$	객체 Ob2	 Ob3	 Ob4
객체	Ob1	**.097** (.003) .014, .263, .061	**.142** (.004) .024, .349, .079	**.875** (.005) .643, .999 .104
	Ob2	-	**.974** (.002) .891, .999, .034	**.314** (.005) .142, .540, .094
	Ob3	-	-	**.410** (.006) .197, .643, .111

따라서 10% 수준에서 객체간 차이에 관한 다중비교에서 유의한 객체 쌍은 Ob1 · Ob2이다. 물론 5% 수준에서 유의한 객체 쌍은 없다.

4. 다른 검정법과의 비교 및 결언

평균순위(Average Ranks) 대체에서는 Friedman 통계량의 영분포가 결측값 패턴에 따라 달라지게 된다. 또한 대체값을 왜 잔여순위들의 평균으로 해야 하는지가 명확하지 않다. 이에 비하여 2절의 다중대체와 임의순열검정에서는 Friedman 통계량의 영분포를 결측값 패턴에 관계없이 기준분포(reference distribution)로 한다는 점과 대체값을 지정해 놓을 필요가 없다는 데 상대적 강점이 있다.

Lim (1998)과 Lim, Lordo and Wolfe (1999), 임동훈 (1999)의 연구에서 제안된 대체 및 검정 방법들, 즉 Summing Ranks나 Counting Ranks, S_{max} 및 S_{min}을 이용한 검정법이 정확한 검정수준을 유지하지 못하는 데 비하여, 본 소고에서 제안된 임의순열검정은 일반적인 순열검정 이론에 따라(Cox and Hinkley 1974, p.196) 항상 정확한 수준을 갖는다는 점이 강점이다. 그러나 본연구에서 제안된 방법은 많은 계산을 필요로 하며 현재까지 일반화된 컴퓨터 모듈로 개발되지 않은 상태이므로 현실적 이용측면에서는 기존 방법론에 비하여 불편하다. 한편, 기존 검정방법들의 수준이 정확하지 않으므로 이들과 검정력을 직접 비교하는 것은 현재로서는 정당하지 않은 것으로 생각된다.

계획적으로 k개의 객체 중 $m(\leq k-2)$개를 순위화를 모든 심사자에게 요구하는 부분순위화(partial ranking)도 일종의 불완전순위자료를 생산하기 때문에 (Critchlow, 1985) 우리의 방법론이 적용가능할 것이다.

본연구에서는 모의생성 대체표본의 수 N_0를 400으로 놓았으나 실제 개별적 응용에 있어서는 산출된 p-값의 정도(精度)를 확인하여 이보다 다소 크게 잡을 수 있을 것이다. 그리고 알고리즘 III에서 Friedman 통계량의 영분포를 산출하기 위하여 임의표본 수 T를 10,000으로 하였는데 이로부터 개별 p-값이 산출되므로 이만큼 충분히 큰 몬테칼로 반복이 필요할 것으로 생각한다.

참고문헌

[1] 임동훈. (1999). "임의 불완전 이원배치 순위계획법에서의 효율적인 검정법", 「응용통계연구」, 12. 191-202쪽.

[2] 최 원. (2000). 「불완전 순위자료의 몬테칼로 임의순열 검정」. 고려대학교 통계학과 석사학위논문.

[3] Conover, W.J. (1980). *Practical Nonparametric Statistics, Second Edition.* New York: Wiley.

[4] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics.* London, Chapman and Hall.

[5] Critchlow, D.E. (1985). *Metric Methods for Analyzing Partially Ranked Data.* New York: Springer-Verlag.

[6] Good, P. (1994). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses.* New York: Springer-Verlag.

[7] Hollander, M. and Wolfe, D.A. (1973). *Nonparametric Statistical Methods.* New York: Wiley.

[8] Lim, D.H. (1998). "Nonparametric methods for analyzing incompletely ranking data," *The Korean Communications in Statistics,* 5. pp. 695-706.

[9] Lim, D.H., Lordo, R.A. and Wolfe, D.A. (1999). "A screeing analysis for incomplete ranking data," *The Statistician,* 48 (Part 1). pp. 95-107.

[10] Lordo, R.A. and Wolfe, D.A. (1994). "Statistical methodology for incompletely ranked data," *Parisankhyan Samikkha* 1. pp. 7-15.

[2000년 6월 접수, 2000년 8월 채택]

Monte Carlo Random Permutation Tests for Incompletely Ranked Data

Myung-Hoe Huh[1)] Won Choi [2)]

ABSTRACT

When k objects are evaluated independently by a panel of n judges, we often meet incompletely ranked data. In this study, we propose a random permutation testing method for the hypothesis that there are no differences among judges in preference for the objects. P-value of the test is obtained by aggregating a number of multiple imputation datasets with the reference distribution which is mathematically derived and/or Monte Carlo generated under the null hypothesis. Friedman statistic was used for the hypothesis in consideration. Also, Tukey-type multiple comparison of differences in objects is numerically illustrated.

Keywords: Incompletely ranked data; Multiple imputation; Monte Carlo method; Random permutation test; Multiple comparison.

1) Professor, Dept. of Statistics, Korea University. E-mail: stat420@mail.korea.ac.kr
2) Consultant, d2K Solutions, Seoul, Korea. E-mail: woniccu@d2ksolutions.com