The Korean Communications in Statistics
Vol. 13 No. 2, 2006, pp. 243-254

Relative Frequency of Order Statistics in Independent and Identically Distributed Random Vectors[1]

So Ryoung Park[2], Hyoungmoon Kwon[3], Sun Yong Kim[4] and Iickho Song[5]

Abstract

The relative frequency of order statistics is investigated for independent and identically distributed (i.i.d.) random variables. Specifically, it is shown that the probability $\Pr\{X_{[s]} = x\}$ is no less than the probability $\Pr\{X_{[r]} = x\}$ at any point $x \geqq x_0$ when $r < s$, where $X_{[r]}$ denotes the r-th order statistic of an i.i.d. discrete random vector and x_0 depends on the population probability distribution. A similar result for i.i.d. continuous random vectors is also presented.

Keywords : Order statistic; Probability mass function; Probability density function; Relative frequency.

1. 서론

잘 알려진 바와 같이 순서 통계량과 순위 통계량은 (Balakrishnan과 Cohen (1991), Hajek, Sidak과 Sen (1999)) 여러 학문 영역에서 자주 쓰인다. 이 두 통계량은 모수에 (parameter) 대한 가정 없이 가설을 검정하는 비모수적 (nonparametric) 방법에서 매우 중요한 통계량으로서, 표본이 적거나 모집단 분포에 대한 가정이 쉽지 않을 때 쓸 수 있다는 장점이 있다. 순서 통계량과 순위 통계량을 바탕으로 하면 측정 도구의 정확성이 상대적으로 낮아도 되고, 계산하기 쉬우며, 모집단 분포에 대한 정확한 가정이 필요하지 않아 그릇된 결론을 내릴 가능성이 적다.

한편, 순서 통계량이나 순위 통계량에 대한 연구는 여러 학문에의 응용 파급 효과

---

1) This work was supported by the Ministry of Science and Technology under the National Research Laboratory Program of Korea Science and Engineering Foundation.

2) Assistant Professor, School of Information, Communications, and Electronics Engineering, The Catholic University of Korea, Bucheon 420-836, Korea.

3) PhD Student, Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea.

4) Associate Professor, Department of Electronics Engineering, Konkuk University, Seoul 143-701, Korea.

5) Professor, Department of Electrical Engineering, KAIST, Daejeon 305-701, Korea.
Correspondence : i.song@ieee.org

가 커 요즈음도 여러 사람이 이를 많이 다루고 있다. 순서 통계량과 순위 통계량은 여러 분야에서 응용이 되고 있는데, 특히 전자공학의 한 영역인 통신과 신호처리에서도 널리 적용되고 있다 (Brown과 Zoubir (2000), Kim 등 (2002)). 통신의 바탕이며 가장 중요한 처리 과정인 신호 검파 (detection) 문제에서 (Blum (1996), Maras (2003), Song과 Kassam (1992)) 순서나 순위 통계량을 바탕으로 하는 방법을 쓰면 관측값의 순서나 순위, 그리고 부호와 같은 부분 정보만으로 신호를 검파할 수 있어, 검파기의 기억용량을 크게 줄일 수 있고 시스템 얼개도 간단해지는 장점이 있다 (Bae와 Song (1997), Kassam (1988), Poor와 Thomas (1993)).

독립이고 분포가 같은 확률벡터 $\underline{X}=(X_1, X_2, \cdots, X_n)$의 r째 순서 통계량을 $X_{[r]}$이라 하면, $r<s$일 때 $\Pr\{X_{[r]} \leqq X_{[s]}\}=1$이다. 이제, $\underline{X}$가 이산 확률벡터일 때, 사건 $E_r(x)=\{X_{[r]}=x\}$와 $E_s(x)=\{X_{[s]}=x\}$ 가운데 어느 사건의 확률이 더 클지 생각해 보자. 먼저, 확률의 합은 1이라는 제약 조건으로 말미암아, 곧, $\sum_x \Pr\{E_r(x)\}=1$이고 $\sum_x \Pr\{E_s(x)\}=1$이므로, $E_r(x)$와 $E_s(x)$ 둘 가운데 어느 것도 '늘' 더 자주 일어날 수는 없고, x가 어떤 값일 때에는 $E_r(x)$가, x가 다른 값일 때에는 $E_s(x)$가 더 자주 일어날 것이다. 직관적으로, "x가 '크면' $E_s(x)$가 $E_r(x)$보다, x가 '작으면' $E_r(x)$가 $E_s(x)$보다 더 자주 일어날 것이다"라고 할 수 있다. 이때, 궁금한 것이 두 가지 자연스럽게 떠오른다. 첫째, x가 '큰지' 아니면 '작은지' 어떻게 알 수 있을까? 둘째, $\Pr\{X_{[s]}=x_0\}>\Pr\{X_{[r]}=x_0\}$일 때 $\Pr\{X_{[s]}=x_1\}<\Pr\{X_{[r]}=x_1\}$이고 $x_0<x_1$인 점 x_1이 있을까? 이 논문에서는 이 두 물음의 답을 찾아본다.

이 논문에서 얻은 순서 통계량의 상대 크기는 최신 통신 시스템에서 사용하는 여러 수신기의 성능을 개선하는데 응용할 수 있다. 보기를 들면, 에너지가 큰 길의 성분을 선택적으로 받아들여 여러 길 감쇄 (multipath fading) 채널을 보상하는 선택적 갈퀴 (selective rake) 수신기를 (Oppermann 등 (2004), Kim 등 (2006)) 설계하고 그 성능을 분석할 때 이론적인 바탕으로 활용할 수 있다.

2. 순서 통계량의 확률함수의 상대 크기

이 장에서는 보조정리와 함께 순서 통계량의 확률함수의 상대크기의 성질을 이산 확률벡터에서는 정리 1로 연속 확률 벡터에서는 정리 2로 나타내었다. 보조정리와 정리의 증명은 부록으로 넣어 두었다.

2.1 보조 정리

먼저, 정수 n, s, r이 $1 \leqq r < s \leqq n$을 만족시킬 때, $\kappa_{r,s}$를 아래와 같이 두자.

$$\kappa_{r,s}=\left\{\frac{(n-s)!(s-1)!}{(n-r)!(r-1)!}\right\}^{1/(s-r)} \tag{1}$$

보조 정리 1: 식 (1)에 정의한 $\kappa_{r,s}$는 아래 다섯 가지 성질을 지닌다.

성질 1. $1 \leqq r < s \leqq n$일 때, $r+s \lesseqgtr n+1$이면 $\kappa_{r,s} \lesseqgtr 1$이다.
성질 2. $1 \leqq r < s \leqq n$일 때, $\kappa_{r,s}\kappa_{n-s+1,n-r+1} = 1$이다.
성질 3. $1 \leqq r < s \leqq n-1$일 때, $\kappa_{r,s+1} > \kappa_{r,s}$이다.
성질 4. $2 \leqq r+1 < s \leqq n$일 때, $\kappa_{r+1,s} > \kappa_{r,s}$이다.
성질 5. $3 \leqq r+2 < s \leqq n$일 때, $r+s \lesseqgtr n+1$이면 $\kappa_{r,s} \lesseqgtr \kappa_{r+1,s-1}$이다.

위 성질 2로 말미암아 $r+s < n+1$일 때 $\kappa_{r,s}$를 살펴보면 $r+s > n+1$일 때 $\kappa_{r,s}$의 특성도 알 수 있다. 그러므로, 이 논문에서는 특별한 경우를 제외하고는 $r+s < n+1$이라고 둔다. 이제, 좀더 일반적으로 κ_{r_i,s_i}와 κ_{r_j,s_j}를 견주어 보자. 먼저, $\{r_i \geq r_j, s_i \geq s_j\}$이면 $\kappa_{r_i,s_i} \geq \kappa_{r_j,s_j}$임을 성질 3과 4에서 쉽게 알 수 있다. 한편, $\{r_j > r_i, s_j < s_i, r_j+s_j \geqq r_i+s_i\}$일 때에는, $\Delta = (r_j+s_j)-(r_i+s_i)$라 두면 $\Delta \geqq 0$이고 $r_j - r_i - \Delta = s_i - s_j > 0$이므로, 성질 4와 5에서

$$\begin{aligned}\kappa_{r_i,s_i} &\leqq \kappa_{r_i+\Delta,s_i} \\ &< \kappa_{r_j,s_i-(r_j-r_i-\Delta)} \\ &= \kappa_{r_j,s_j}\end{aligned} \tag{2}$$

이다. 곧, $\{r_j > r_i, s_j < s_i, r_j+s_j < r_i+s_i\}$일 때를 빼고는 κ_{r_i,s_i}와 κ_{r_j,s_j} 가운데 어느 것이 더 큰지 알 수 있다. 이제, $\{r_j > r_i, s_j < s_i, r_j+s_j < r_i+s_i\}$라는 조건은 조건 $\{r_j > r_i, r_j+s_j < r_i+s_i\}$와 같으므로, $\{r_j > r_i,\ r_j+s_j < r_i+s_i\}$일 때 κ_{r_i,s_i}와 κ_{r_j,s_j}의 상대크기는 다음 보조정리 2에서 알 수 있다.

보조 정리 2: 식 (1)에 정의한 $\kappa_{r,s}$에서, $r_j > r_i$이고 $r_j+s_j < r_i+s_i$이면,

$$\frac{1}{s_i-r_i}\ln\left\{\frac{(s_i-1)!}{(r_i-1)!}\right\} > \frac{1}{s_j-r_j}\ln\left\{\frac{(s_j-1)!}{(r_j-1)!}\right\}\text{일 때에는 } \kappa_{r_i,s_i} > \kappa_{r_j,s_j}\text{이고,} \tag{3}$$

$$\frac{1}{s_i-r_i}\ln\left\{\frac{(s_i-1)!}{(r_i-1)!}\right\} < \frac{1}{s_j-r_j}\ln\left\{\frac{(s_j-1)!}{(r_j-1)!}\right\}\text{일 때에는}$$

$$n\text{이 작으면 } \kappa_{r_i,s_i} > \kappa_{r_j,s_j}\text{이고, } n\text{이 크면 } \kappa_{r_i,s_i} < \kappa_{r_j,s_j}\text{이다.} \tag{4}$$

보조 정리 1과 2에서 얻은 바를 좀더 쉽게 보고자, $\kappa_{r,s}$를 아래와 같이 늘어놓았을 때, $r+s=t$이도록 오른쪽 위로 올라가는 줄을 '맞모줄 t' 라고 한다.

보기를 들면, 아래에서 맞모줄 6은 $\kappa_{1,5}$와 $\kappa_{2,4}$를 잇는 줄이고 맞모줄 7은 $\kappa_{1,6}$, $\kappa_{2,5}$, $\kappa_{3,4}$를 잇는 줄이다. 그러면, 맞모줄 $n+1$과 그 위쪽만 생각할 때, κ_{r_i,s_i}는 열 r_i와 그 왼쪽 그러면서 맞모줄 r_i+s_i와 그 위쪽에 있는 것들보다 크고, 열 r_i와 그 오른쪽 그러면서 맞모줄 r_i+s_i와 그 아래쪽에 있는 것들보다 작다. 한편, 나머지 영역인 열 r_i

$$
\begin{array}{llllll}
\kappa_{1,2} & & & & & \\
\kappa_{1,3} & \kappa_{2,3} & & & & \\
\kappa_{1,4} & \kappa_{2,4} & \kappa_{3,4} & & & \\
\kappa_{1,5} & \kappa_{2,5} & \kappa_{3,5} & \kappa_{4,5} & & \\
\kappa_{1,6} & \kappa_{2,6} & \kappa_{3,6} & \kappa_{4,6} & \kappa_{5,6} & \\
\kappa_{1,7} & \kappa_{2,7} & \kappa_{3,7} & \kappa_{4,7} & \kappa_{5,7} & \kappa_{6,7} \\
\vdots & & & & &
\end{array}
$$

의 왼쪽 그러면서 맞모줄 r_i+s_i 아래쪽 또는 열 r_i의 오른쪽 그러면서 맞모줄 r_i+s_i 위쪽에서는 (곧, κ_{r_j,s_j}가 κ_{r_i,s_i}의 오른쪽 위쪽에 있을 때에는) 보조 정리 2에서 밝혔듯이 n을 알아야 어느 것이 큰지 말할 수 있다. 보조 정리 1과 2에서 얻은 것을 모으면 <표 1>과 같다.

<표 1> 맞모줄 $n+1$ 위쪽에서 매개변수 κ_{r_i,s_i}와 κ_{r_j,s_j}의 상대 크기

	$r_i+s_i > r_j+s_j$	$r_i+s_i = r_j+s_j$
$r_i > r_j$	$\kappa_{r_i,s_i} > \kappa_{r_j,s_j}$	$\kappa_{r_i,s_i} > \kappa_{r_j,s_j}$
$r_i = r_j$	$\kappa_{r_i,s_i} > \kappa_{r_j,s_j}$	$\kappa_{r_i,s_i} = \kappa_{r_j,s_j}$
$r_i < r_j$	(3),(4)	$\kappa_{r_i,s_i} < \kappa_{r_j,s_j}$

보조 정리 2를 쓰면 $\kappa_{1,5} > \kappa_{2,3}$, $\kappa_{1,6} > \kappa_{2,4}$, $\kappa_{1,8} > \kappa_{2,6}$, $\kappa_{1,10} > \kappa_{3,7}$, $\kappa_{2,6} > \kappa_{3,4}$, $\kappa_{2,7} > \kappa_{3,5}$, $\kappa_{2,8} > \kappa_{4,5}$, $\kappa_{2,9} > \kappa_{4,6}$, $\kappa_{2,10} > \kappa_{4,7}$임을 알 수 있으며, 이 결과와 보조 정리 1을 바탕으로 아래를 얻을 수 있다.

$$
\begin{aligned}
&\kappa_{1,2} < \kappa_{1,3} < \kappa_{1,4} < \kappa_{2,3} < \kappa_{1,5} < \kappa_{2,4} < \kappa_{1,6} < \kappa_{2,5} \\
&< \{\kappa_{1,7}, \kappa_{3,4}\}_{247} < \kappa_{2,6} < \{\kappa_{1,8}, \kappa_{3,5}\}_{27} < \kappa_{2,7} < \kappa_{3,6} \\
&< \{\kappa_{1,9}, \kappa_{4,5}\}_{16} < \kappa_{2,8} < \{\kappa_{1,10}, \kappa_{3,7}, \kappa_{4,6}\}_{15,20} < \cdots .
\end{aligned} \tag{5}
$$

식 (5)에서 $\{\kappa_{r_1,s_1}, \kappa_{r_2,s_2}\}_{n_1}$은 $n \leqq n_1$이면 $\kappa_{r_2,s_2} < \kappa_{r_1,s_1}$이고 $n \geqq n_1+1$이면 $\kappa_{r_1,s_1} < \kappa_{r_2,s_2}$라는 뜻이고, $\{\kappa_{r_1,s_1}, \kappa_{r_2,s_2}, \kappa_{r_3,s_3}\}_{n_1,n_2}$는 $n \leqq n_1$이면 $\kappa_{r_2,s_2} < \kappa_{r_3,s_3} < \kappa_{r_1,s_1}$이고, $n_1+1 \leqq n \leqq n_2-1$이면 $\kappa_{r_2,s_2} < \kappa_{r_1,s_1} < \kappa_{r_3,s_3}$이며, $n \geqq n_2$이면 $\kappa_{r_1,s_1} < \kappa_{r_2,s_2} < \kappa_{r_3,s_3}$임을 의미한다.

2.2 순서 통계량의 확률함수의 상대 크기

이제, 확률벡터 $\underline{X}$가 독립이고 분포가 같은 확률벡터라 (박철훈, 송익호, 남동경 (2004)) 하고, X_i의 누적분포함수를 F라 하면, 순서 통계량 $X_{[r]}$의 누적분포함수 $F_r(x)$는

$$F_r(x) = \Pr\{X_{[r]} \le x\} = \sum_{j=r}^{n} \frac{n!}{(n-j)!j!} F^j(x)\{1-F(x)\}^{n-j} \tag{6}$$

이다 (David와 Nagaraja (2003)). 한편, $\underline{X}$가 연속 확률벡터일 때, X_i의 확률밀도함수를 f라 하면 $X_{[r]}$의 확률밀도함수 $f_r(x) = dF_r(x)/dx$는

$$f_r(x) = \frac{n!}{(n-r)!(r-1)!} F^{r-1}(x)\{1-F(x)\}^{n-r} f(x) \tag{7}$$

이고, $\underline{X}$가 이산 확률변수일 때 $X_{[r]}$의 확률질량함수 $p_r(x) = \Pr\{X_{[r]} = x\} = F_r(x) - F_r(x-1)$는

$$p_r(x) = \frac{n!}{(n-r)!(r-1)!} \int_{F(x-1)}^{F(x)} v^{r-1}(1-v)^{n-r} dv \tag{8}$$

이다.

정리 1 (이산 확률벡터에서 순서 통계량의 확률질량함수의 상대 크기)

독립이고 분포가 같은 이산 확률벡터 $\underline{X}$에서 X_i의 누적분포함수를 F, X_i의 확률질량함수를 p, r째 순서 통계량의 확률질량함수를 p_r로 정의한다. 한편, $1 \leqq r < s \leqq n$일 때

$$\overline{\kappa}_{r,s} = \frac{\kappa_{r,s}}{1+\kappa_{r,s}}, \tag{9}$$

$$h_{r,s}(v) = v^{r-1}(1-v)^{n-s}[v^{s-r} - \{\kappa_{r,s}(1-v)\}^{s-r}], \quad 0 < v < 1, \tag{10}$$

이라 두고, $\varnothing_p = \{x|p(x) = 0\}$이라 하며, $F(x) \leqq \overline{\kappa}_{r,s}$인 가장 큰 정수 x를 $\overline{x}_L$로 둔다. 그러면, $x \in \varnothing_p$이면

$$p_r(x) = p_s(x) = 0 \tag{11}$$

이고, $x \notin \varnothing_p$이면

$$x \in \{\cdots, \overline{x}_L - 1, \overline{x}_L\} \text{일 때}, p_r(x) > p_s(x) \text{이고}, \tag{12}$$

$$x \in \{\overline{x}_L + 2, \overline{x}_L + 3, \cdots\} \text{일 때}, \ p_r(x) < p_s(x) \text{이며}, \tag{13}$$

$$x = \overline{x}_L + 1 \text{이고 } F(\overline{x}_L) = \overline{\kappa}_{r,s} \text{일 때 } p_r(\overline{x}_L+1) < p_s(\overline{x}_L+1) \text{이고}, \tag{14}$$

$$x = \overline{x}_L + 1 \text{이고 } F(\overline{x}_L) \neq \overline{\kappa}_{r,s} \text{일 때}$$

$$\int_{F(\overline{x}_L)}^{F(\overline{x}_L+1)} h_{r,s}(v)dv \gtreqless 0 \text{이면 } p_r(\overline{x}_L+1) \lesseqgtr p_s(\overline{x}_L+1) \text{이다}. \tag{15}$$

정리 1은 $r < s$일 때 r째 순서 통계량보다 s째 순서 통계량이 확률적으로 더 클 것이라는 직관적인 상황을 수학적으로 좀더 명확히 나타내는 것이다. 이 정리가 뜻하는 바는 아래와 같다. 첫째, $r < s$일 때 어떤 점 x_0에서 $p_r(x_0) < p_s(x_0)$이면 $x \geq x_0$인 점 x를 어떻게 고르더라도 $p_r(x) \leqq p_s(x)$이지 $p_r(x) > p_s(x)$일 수 없다는 것이다. 둘째로는, $r < s$일 때 $p_r(x_0) > p_s(x_0)$이면 $x \leqq x_0$인 점 x를 어떻게 고르더라도 $p_r(x) \geqq p_s(x)$이지 $p_r(x) < p_s(x)$일 수 없다는 것이다. 이 결과는, 보기를 들자면, 어떤 점

에서 여러 순서 통계량의 확률질량함수 가운데 어느 것이 더 큰지를 확률함수를 얻지 않고 미리 알아볼 때 유용하다.

정리 2 (연속 확률벡터에서 순서 통계량의 확률밀도함수의 상대 크기)

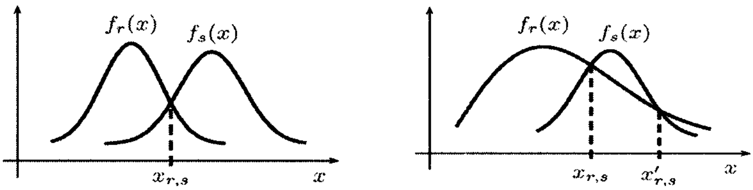
독립이고 분포가 같은 연속 확률벡터 $\underline{X}$에서 X_i의 누적분포함수를 F, X_i의 확률밀도함수를 f, r째 순서 통계량의 확률밀도함수를 f_r로 정의한다. 그러면, $1 \leq r < s \leq n$일 때,

$$x \not\in \varnothing_f \text{이고 } x \gtreqless x_{r,s} \text{이면 } f_s(x) \gtreqless f_r(x) \text{이고,} \tag{16}$$

$$x \in \varnothing_f \text{이면 } f_s(x) = f_r(x) = 0 \text{이다.} \tag{17}$$

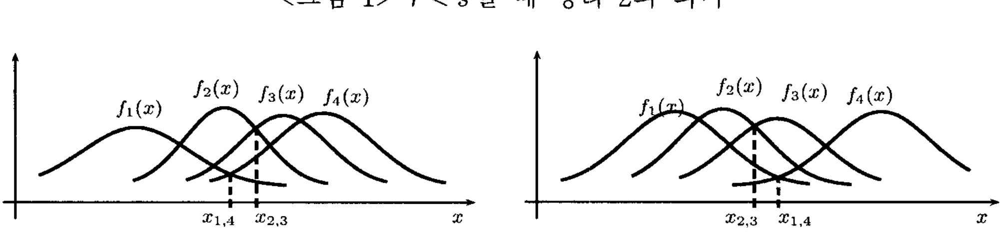
여기서, $\varnothing_f = \{x | f(x) = 0\}$이고, $x_{r,s}$는 $F(x) = \overline{\kappa}_{r,s}$인 실수 x 가운데 하나이다.

정리 2가 뜻하는 바 $r < s$일 때 어떤 점 x_0에서 $f_r(x_0) < f_s(x_0)$이면 $x \geq x_0$일 때 늘 $f_r(x) \leq f_s(x)$이고, $f_r(x_0) > f_s(x_0)$이면 $x \leq x_0$일 때 늘 $f_r(x) \geq f_s(x)$라는 것이다. 보기를 들면, $r < s$일 때, <그림 1> (가)와 같이 만나는 점이 하나인 때는 있으나, <그림 1> (나)와 같이 만나는 점이 하나보다 많은 때는 없다는 것이다. 한편, $x/1+x$는 x의 증가 함수이므로, $\kappa_{r,s}$는 두 순서 통계량의 확률밀도함수를 그릴 때 (또는, 확률질량함수에 '덮개'를 씌워 그릴 때), 만나는 점의 상대적인 위치를 나타낸다. 보기를 들면, <그림 2> (가)와 같이 $f_2(x)$와 $f_3(x)$가 만나는 점 $x_{2,3}$이 $f_1(x)$와 $f_4(x)$가 만나는 점 $x_{1,4}$보다 큰 때는 있으나 <그림 2> (나)와 같이 $x_{2,3}$이 $x_{1,4}$보다 작은 때는 없다는 것을 (5)에서 알 수 있다.



(가) 일어날 수 있는 보기 (나) 일어날 수 없는 보기

<그림 1> $r < s$일 때 정리 2의 의미



(가) 일어날 수 있는 보기 (나) 일어날 수 없는 보기

<그림 2> 정리 2와 (5)에서 유추할 수 있는 보기

3. 보기

보기 1: 확률벡터 $\underline{X}$가 이산 확률벡터이고 $n=3$, $s=2$, $r=1$이라 두자. 그러면, $\kappa_{1,2} = \{(3-2)!(2-1)!/(3-1)!(1-1)!\}^{1/(2-1)} = 1/2$이고 $\overline{\kappa}_{1,2} = 1/3$이다. 이제, $F(1) = 1/6$, $F(2) = 1/3$, $F(3) = 1$이라면, $\overline{x}_L = \overline{x}_U - 1 = 2$이다. 따라서, $x = 1, 2$이면 $p_1(x) > p_2(x)$이고 $x = 3$이면 $p_1(x) < p_2(x)$이다.

보기 2: 독립이고 분포가 같은 이산 확률벡터에서 $F(1) = 0.1$, $F(2) = F(3) = F(4) = 0.3$, $F(5) = F(6) = F(7) = F(8) = 0.5$, $F(9) = 0.6$, $F(10) = 0.8$, $F(11) = 1$이라 하고, $\overline{\kappa}_{r,s} = 0.3$이라 하자. 그러면, $\{3, 4, 6, 7, 8\} \subset \varnothing_p$임을 쉽게 알 수 있고, $\overline{x}_L = 4$이며, $\overline{x}_U = 5$이다. 따라서, $x = 1, 2$이면 $p_r(x) > p_s(x)$이고, $x = 5, 9, 10, 11$이면 $p_r(x) < p_s(x)$이며, $x = 3, 4, 6, 7, 8$이면 $p_r(x) = p_s(x) = 0$이다.

보기 1과 2에서 알 수 있듯이, 어떤 정수 x_0에서 $F(x_0) = \overline{\kappa}_{r,s}$이면, $x_0 = \overline{x}_L = \overline{x}_U - 1$이고, 정리 1에 보인 (11)-(14)를 쓰면 두 순서 통계량의 확률질량함수 가운데 어느 것이 더 큰지 어느 점에서도 알 수 있다. 한편, $F(x) = \overline{\kappa}_{r,s}$인 정수 x가 없으면 $\overline{x}_U = \overline{x}_L + 2$이고, 이때에는 점 $x = \overline{x}_L + 1 = \overline{x}_U - 1$에서 어느 순서 통계량의 확률질량함수가 더 큰지 알려면 정리 1에 보인 (15)를 써야 한다.

보기 3: 보기 2에서 $n = 5$, $s = 3$, $r = 2$이라면 $\overline{\kappa}_{r,s} = 0.4$이고, $\overline{x}_L = 4$이며, $\overline{x}_U = 6$이다. 따라서, $x = 1, 2$이면 $p_r(x) > p_3(x)$이고, $x = 9, 10, 11$이면 $p_2(x) < p_3(x)$이며, $x = 3, 4, 6, 7, 8$이면 $p_2(x) = p_3(x) = 0$이다. 한편, $\int_{F(4)}^{F(5)} h_{2,3}(v)dv = \frac{1}{3}\int_{0.3}^{0.5} v(1-v)^2(5v-2)\,dv = -\frac{19}{150000} < 0$이므로 $p_2(5) > p_3(5)$이다.

이산 확률벡터에서 $F(x) = \overline{\kappa}_{r,s}$인 정수 x가 없고 따라서 $\overline{x}_U = \overline{x}_L + 2$일 때에는, $x = \overline{x}_L + 1 = \overline{x}_U - 1$에서 $p_r(\overline{x}_L + 1) < p_s(\overline{x}_L + 1)$일 때도, $p_r(\overline{x}_L + 1) > p_s(\overline{x}_L + 1)$일 때도, 또 $p(\overline{x}_L + 1) > 0$이더라도 $p_r(\overline{x}_L + 1) = p_s(\overline{x}_L + 1)$일 때도 있다. 아래 세 보기에서 이를 각각 보인다.

보기 4: ($\overline{x}_L = \overline{x}_U - 2$이고 $p_r(\overline{x}_L + 1) > p_s(\overline{x}_L + 1)$인 보기): 보기 1에서 $F(1) = 1/10$, $F(2) = 2/5$, $F(3) = 1$이라면, $\overline{x}_L = 1$이고 $\overline{x}_U = 3$이므로 $p_1(1) > p_2(1)$이고 $p_1(3) < p_2(3)$이다. 한편, $\int_{F(1)}^{F(2)} h_{1,2}(v)dv = \frac{1}{2}\int_{0.1}^{0.4} (1-v)(3v-1)\,dv = -63/2000 < 0$이므로

$p_1(2) > p_2(2)$이다.

보기 5: ($\bar{x}_L = \bar{x}_U - 2$이고 $p_r(\bar{x}_L+1) < p_s(\bar{x}_L+1)$인 보기): 보기 1에서 $F(1) = 3/10$, $F(2) = 2/5$, $F(3) = 1$이라면, $\bar{x}_L = 1$이고 $\bar{x}_U = 3$이므로 $p_1(1) > p_2(1)$이고 $p_1(3) < p_2(3)$이다. 한편, $\int_{F(1)}^{F(2)} h_{1,2}(v)dv = \frac{1}{2}\int_{0.3}^{0.4}(1-v)(3v-1)dv = \frac{3}{2000} > 0$이므로 $p_1(2) < p_2(2)$이다.

보기 6: ($\bar{x}_L = \bar{x}_U - 2$이고 $p_r(\bar{x}_L+1) = p_s(\bar{x}_L+1)$인 보기): 이제, 보기 1에서 $p(1) = 1/6$, $p(2) = (9-\sqrt{21})/12$, $p(3) = (1+\sqrt{21})/12$이라 두자. 그러면, $F(1) = 1/6$, $F(2) = (11-\sqrt{21})/12$, $F(3) = 1$이다. 따라서, $\bar{x}_L = 1$이고 $\bar{x}_U = 3$이므로 $p_1(1) > p_2(1)$이고 $p_1(3) < p_2(3)$이다. 한편, $\int_{F(1)}^{F(2)} h_{1,2}(v)dv = \frac{1}{2}\int_{\frac{1}{6}}^{\frac{11-\sqrt{21}}{12}}(1-v)(3v-1)\,dv = 0$이므로, $p_1(2) = p_2(2)$이다.

보기 7: 식 (5)는 매우 큰 두 수를 견주어 볼 때에도 쓸 수 있다. 보기를 들면, $n = 2005$라 하면 $\kappa_{3,6} > \kappa_{2,7}$에서 $(2003!1999!5!)^6(1999!5!)^4 > (2002!1998!6!2!)^6(2002!2!)^4$임을 알 수 있다.

보기 8: 한편, (5)에 보인 n_1과 n_2 근처에서 κ_{r_i,s_i}와 κ_{r_j,s_j}를 구체적으로 얻어 보면 다음과 같다. 먼저, $n = 247$이면 $\kappa_{1,7} \approx 0.0122952 > \kappa_{3,4} \approx 0.0122951$이고, $n = 248$이면 $\kappa_{1,7} \approx 0.01224486 < \kappa_{3,4} \approx 0.01224490$이다. 다음에, $n = 27$이면 $\kappa_{1,8} \approx 0.1475 > \kappa_{3,5} \approx 0.1474$이고, $n = 28$이면 $\kappa_{1,8} \approx 0.1413 < \kappa_{3,5} \approx 0.1414$이다. 한편, $n = 16$이면 $\kappa_{1,9} \approx 0.334 > \kappa_{4,5} \approx 0.333$이고, $n = 17$이면 $\kappa_{1,9} \approx 0.306 < \kappa_{4,5} \approx 0.307$이다. 그리고, $n = 19$이면 $\kappa_{1,10} \approx 0.302 > \kappa_{3,7} \approx 0.301$이고, $n = 20$이면 $\kappa_{1,10} \approx 0.281 < \kappa_{3,7} \approx 0.282$이다. 끝으로, $n = 15$이면 $\kappa_{1,10} \approx 0.430 > \kappa_{4,6} \approx 0.426$이고, $n = 16$이면 $\kappa_{1,10} \approx 0.388 < \kappa_{4,6} \approx 0.389$이다.

4. 요약과 결론

이 논문에서는 독립이고 분포가 같은 이산 확률벡터와 연속 확률벡터에서 순서 통계량의 상대 빈도를 살펴보았다. 좀더 구체적으로는, (1)에서 정의한 $\kappa_{r,s}$의 성질을 몇 가지 알아본 뒤, 이를 바탕으로 순서 통계량의 확률함수가 어떤 특성을 보이는지 살펴보았다. 이산 확률벡터에서 순서 통계량의 확률질량함수의 상대 크기와 연속 확률벡터에서 순서 통계량의 확률밀도함수의 상대 크기를 살펴본 바, x가 어떤 값일 때

사건 $E_r(x)=\{X_{[r]}=x\}$의 확률이 사건 $E_s(x)=\{X_{[s]}=x\}$의 확률보다 더 큰지를 밝혔다. 특히, $r<s$일 때 $x \ge x_0$인 점 x를 어떻게 고르더라도 $\Pr\{X_{[s]}=x\}$가 $\Pr\{X_{[r]}=x\}$보다 작지 않음을 보였다. 이 논문에서 얻은 순서 통계량의 확률 분포 특성과 상대 크기는 여러 순서 통계량의 확률질량함수 또는 확률밀도함수 가운데 어느 것이 더 큰지를 확률함수를 얻지 않고 미리 알아볼 때 쓸 수 있고, 통신 시스템에서 비모수 검파기나 선택적 갈퀴 수신기와 같이 순서 통계량을 바탕으로 신호를 검파하는 여러 가지 수신기를 준최적으로 구현하고, 성능 특성을 분석하여 개선하는 데에 쓸 수 있다.

부록: 보조 정리와 정리의 증명

1. 보조 정리 1의 증명:

성질 1. 식 (1)에서

$$\kappa_{r,s}^{s-r}=\frac{(s-1)(s-2)\cdots r}{(n-r)(n-r-1)\cdots(n-s+1)} \tag{18}$$

이다. 식 (18)의 분자와 분모는 둘 다 이어지는 자연수를 $s-r$개 곱한 것이므로, 같은 자리에 있는 수를 견주어 보면 $s-i \leqq n-r-i+1$일 때 $\kappa_{r,s} \leqq 1$임을 쉽게 알 수 있다.

성질 2. 식 (1)에서

$$\kappa_{r,s}\kappa_{n-s+1,n-r+1}=\left\{\frac{(n-s)!(s-1)!}{(n-r)!(r-1)!}\right\}^{1/(s-r)}\left\{\frac{(r-1)!(n-r)!}{(s-1)!(n-s)!}\right\}^{1/(s-r)}$$
$$=1. \tag{19}$$

성질 3. 식 (1)에서

$$\left\{\frac{\kappa_{r,s+1}}{\kappa_{r,s}}\right\}^{(s-r+1)(s-r)}=\left\{\frac{s}{n-s}\right\}^{s-r}\frac{1}{\kappa_{r,s}^{s-r}}$$
$$=\frac{s(n-r)}{(n-s)(s-1)}\frac{s(n-r-1)}{(n-s)(s-2)}\cdots\frac{s(n-s+1)}{(n-s)r} \tag{20}$$

인데, 이 식에서 곱하는 분수는 모두 $\frac{ab}{cd}$인 꼴이고 $a>d$, $b>c$이므로, 그 곱이 1보다 크다는 것을 쉽게 알 수 있다.

성질 4. 위 성질 3의 증명과 비슷하게

$$\left\{\frac{\kappa_{r+1,s}}{\kappa_{r,s}}\right\}^{(s-r-1)(s-r)}=\frac{(n-r)(s-1)}{r(n-r)}\frac{(n-r)(s-2)}{r(n-r-1)}\cdots\frac{(n-r)r}{r(n-s+1)} \tag{21}$$

이다. 이 식에서 곱하는 분수는 모두 $\frac{ab}{cd}$인 꼴이고, $s>r+1$이므로 $\{a \ge d, b>c\}$ 또는 $\{a>d,\ b \ge c\}$이다. 따라서, 이 분수들을 모두 곱하면 1보다 크다.

성질 5. 먼저, 성질 1에서

$$r+s=n+1\text{이면 } \kappa_{r+1,s-1}=\kappa_{r,s}=1 \tag{22}$$

를 바로 알 수 있다. 다음에, $r+s<n+1$이면, $n-s+1>r$이고, $n-r>s-1$이며, 성질 1에서 $1/\kappa_{r,s}>1$이므로

$$\left\{\frac{\kappa_{r+1,s-1}}{\kappa_{r,s}}\right\}^{s-r-2} = \left\{\frac{(n-r)(n-s+1)}{(s-1)r}\right\}^{s-r}\frac{1}{\kappa_{r,s}^2} > 1 \tag{23}$$

이다. 끝으로, (23), 성질 1, 성질 2에서

$$r+s>n+1 \text{이면 } \kappa_{r,s}>\kappa_{r+1,s-1} \tag{24}$$

임을 바로 알 수 있다.

2. 보조 정리 2의 증명: 먼저, $r_j>r_i$이고 $r_j+s_j<r_i+s_i$이면 $s_j<s_i-r_j+r_i<s_i$, $s_i-r_i>s_j-r_j$, $r_i<r_j<s_j<s_i$, $n-r_i>n-r_j>n-s_j>n-s_i$, $(s_i-r_i)-(s_j-r_j)=s_i-s_j+r_j-r_i>0$임을 생각하면 (1)에서

$$\left\{\frac{\kappa_{r_i,s_i}}{\kappa_{r_j,s_j}}\right\}^{(s_i-r_i)(s_j-r_j)} = M_{r_i,s_i,r_j,s_j}\frac{N_{r_i,s_i,r_j,s_j}(n)}{D_{r_i,s_i,r_j,s_j}(n)} \tag{25}$$

를 얻는다. 이 식에서,

$$M_{r_i,s_i,r_j,s_j} = \left\{\frac{(s_i-1)!}{(r_i-1)!}\right\}^{s_j-r_j}\left\{\frac{(r_j-1)!}{(s_j-1)!}\right\}^{s_i-r_i}, \tag{26}$$

$$N_{r_i,s_i,r_j,s_j}(n) = (n-r_j)(n-r_j-1)\cdots(n-s_j+1)^{s_i-s_j+r_j-r_i}, \tag{27}$$

$$\begin{aligned} D_{r_i,s_i,r_j,s_j}(n) = {} & (n-r_i)(n-r_i-1)\cdots(n-r_j+1)^{s_j-r_j} \\ & \cdot (n-s_j)(n-s_j-1)\cdots(n-s_i+1)^{s_j-r_j} \end{aligned} \tag{28}$$

이다. 그런데, $N_{r_i,s_i,r_j,s_j}(n)$과 $D_{r_i,s_i,r_j,s_j}(n)$은 둘 다 n의 $(s_i-s_j+r_j-r_i)(s_j-r_j)$차 함수이고, $N_{r_i,s_i,r_j,s_j}(n)$의 가장 큰 영 s_j-1은 $D_{r_i,s_i,r_j,s_j}(n)$의 가장 큰 영 s_i-1보다 작으므로, $n>s_i-1$일 때 $N_{r_i,s_i,r_j,s_j}(n)/D_{r_i,s_i,r_j,s_j}(n)$은 1보다 작지 않고 감소 함수이며

$$\lim_{n\to\infty}\frac{N_{r_i,s_i,r_j,s_j}(n)}{D_{r_i,s_i,r_j,s_j}(n)} = 1 \tag{29}$$

이다. 따라서, $M_{r_i,s_i,r_j,s_j}>1$이면 $\kappa_{r_i,s_i}>\kappa_{r_j,s_j}$이고, $M_{r_i,s_i,r_j,s_j}<1$이면 n이 작을 때에는 $\kappa_{r_i,s_i}>\kappa_{r_j,s_j}$이고, n이 클 때에는 $\kappa_{r_i,s_i}<\kappa_{r_j,s_j}$이다. 이를 자연 대수를 써서 나타내면 보조 정리 2와 같다.

3. 정리 1의 증명: 먼저, $v\in(0,1)$이면 $v^{r-1}(1-v)^{n-s}>0$이므로

$$v \gtreqless \overline{\kappa_{r,s}} \text{일 때, } h_{r,s}(v) \gtreqless 0 \tag{30}$$

임을 (10)에서 알 수 있다. 따라서,

$$p_s(x) - p_r(x) = \frac{n!}{(n-s)!(s-1)!} \int_{F(x-1)}^{F(x)} h_{r,s}(v) dv \tag{31}$$

이므로, $p(x) > 0$일 때 $F(x) > F(x-1)$임을 생각하면,

$$F(x) \leqq \bar{\kappa}_{r,s}\text{일 때} \quad p_s(x) - p_r(x) < 0\text{이고,} \tag{32}$$

$$F(x-1) \geqq \bar{\kappa}_{r,s}\text{일 때} \quad p_s(x) - p_r(x) > 0\text{임을} \tag{33}$$

알 수 있다. 이제, $F(x-1) \geqq \bar{\kappa}_{r,s}$인 가장 작은 정수 x를 $\bar{x}_U$라 둔다. 그러면, $F(\bar{x}_L) = \bar{\kappa}_{r,s}$이면 $\bar{x}_U = \bar{x}_L + 1$임을 바로 알 수 있고, x가 어떤 값이더라도 늘 $p(x) > 0$이므로 $F(x)$가 증가 함수라는 것을 생각하면, (32)에서 (12)를 얻을 수 있으며, (33)에서 (13)과 (14)를 얻을 수 있다. 다음에, $F(\bar{x}_L) \neq \bar{\kappa}_{r,s}$라 두면, $\bar{x}_U = \bar{x}_L + 2$이고, (32)와 (33)에서 각각 (12)와 (13)을 얻을 수 있다. 한편, $x = \bar{x}_L + 1$일 때에는 (31)에서 (15)를 얻을 수 있다.

4. 정리 2의 증명: 먼저, $f(x) = 0$인 점 x에서는 늘 $f_r(x) = f_s(x) = 0$임은 (7)에서 바로 알 수 있다. 따라서, $f(x) > 0$이라 하면, $0 < F(x) < 1$일 때 $F(x)/(1-F(x))$는 $F(x)$의 증가 함수이고 0보다 작지 않다는 것을 생각하여

$$\begin{aligned} \frac{f_s(x)}{f_r(x)} &= \frac{(n-r)!(r-1)!}{(n-s)!(s-1)!} \left\{ \frac{F(x)}{1-F(x)} \right\}^{s-r} \\ &\gtreqless 1 \end{aligned} \tag{34}$$

를 풀면

$$F(x) \gtreqless \bar{\kappa}_{r,s} \tag{35}$$

이다. 이제, $\kappa_{r,s} > 0$이므로 $0 < \bar{\kappa}_{r,s} < 1$이라는 것과 $F(x)$가 줄지 않는 함수라는 것을 고려하여, $F(x) = \bar{\kappa}_{r,s}$인 가장 작은 실수 x와 가장 큰 실수 x를 각각 x_L과 x_U라 둔다. 먼저, $x_L \neq x_U$일 때에는 x가 (x_L, x_U)에서 어느 값이더라도 $F(x)$의 값은 바뀌지 않으므로 $f(x) = 0$이다. 곧, $(x_L, x_U) \subset \varnothing_f$이고, 따라서, $x \in (x_L, x_U)$이면 $f_s(x) = f_r(x) = 0$이다. 다음에, $x_L = x_U$일 때에는 $x_L = x_U = F^{-1}(\bar{\kappa}_{r,s}) = x_{r,s}$라 두면, (34)와 (35)에서 (16)을 얻을 수 있다.

감사: 본 논문의 질을 더욱 향상시킬 수 있도록 심사해 주신 심사위원들께 고맙다는 말씀을 올립니다.

참고문헌

[1] 박철훈, 송익호, 남동경 (2004), 「확률과정」, 생능출판사, 서울.

[2] Bae, J. and Song, I. (1997). Rank-based detection of weak random signals in a multiplicative noise model. *Signal Processing*, Vol. 63, 121-131.

[3] Balakrishnan, N. and Cohen, A.C. (1991). *Order Statistics and Inference*, Academic, New York.

[4] Blum, R.S. (1996). Locally optimum distributed detection of correlated random signals based on ranks. *IEEE Transactions on Information Theory*, Vol. 42, 931-942.

[5] Brown, C.L. and Zoubir, A.M. (2000). A nonparametric approach to signal detection in impulsive interference. *IEEE Transactions on Signal Processing*, Vol. 48, 2665-2669.

[6] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics (3rd edn)*, John Wiley and Sons, New York.

[7] Hajek, J., Sidak, Z. and Sen, P.K. (1999). *Theory of Rank Tests (2nd edn)*, Academic, New York.

[8] Kassam, S.A. (1988). *Signal Detection in Non-Gaussian Noise*, Springer-Verlag, New York.

[9] Kim, H.G., Song, I., Yoon, S. and Park, S.R. (2002). Nonparametric PN code acquisition using the signed-rank statistic for DS/CDMA systems in frequency-selective Rician fading channels. *IEEE Transactions on Vehicular Technology*, Vol. 51, 1138-1144.

[10] Kim, B.S., Bae, J., Song, I., Kim, S.Y. and Kwon, H. (2006). A comparative analysis of optimum and suboptimum rake receivers in impulsive environment. *IEEE Transactions on Vehicular Technology*, Vol. 55, (to be published).

[11] Maras, A.M. (2003). Adaptive nonparametric locally optimum Bayes detection in additive non-Gaussian noise. *IEEE Transactions on Information Theory*, Vol. 49, 204-220.

[12] Oppermann, I., Hamalainen, M. and Iinatti, J. (2004). *UWB: Theory and Application*, John Wiley and Sons, New York.

[13] Poor, H.V. and Thomas, J.B. (1993). *Advances in Statistical Signal Processing: Vol. 2. Signal Detection*, JAI Press, Greenwich.

[14] Song, I. and Kassam, S.A. (1992). Locally optimum rank detection of correlated random signals in additive noise. *IEEE Transactions on Information Theory*, Vol. 38, 1311-1322.

[Received December 2005, Accepted April 2006]