

신호 검파에 알맞은 비모수 통계량 사이의 상관 계수

준회원 주 현*, 중신회원 송 익 호**, 배 진 수*[†]

Correlation Coefficients between Some Nonparametric Statistics Used for Signal Detection

Hyun Joo* Associate Member, Ickho Song**, Jinsoo Bae*[†] Lifelong Members

요 약

이 논문에서는 신호 검파에서 널리 쓰이는 세 개의 비모수 통계량 사이의 결합 분포와 상관계수를 얻는다. 아울러, 상관계수들의 상한과 하한을 얻고, 상관계수들 사이의 눈여겨 볼만한 관계를 살펴본다. 또한, 몇 가지 의미 있는 분포에서 얻은 구체적인 상관계수 값을 표와 그림으로 정리하였다. 이 논문에서 얻은 결과는 여러 검정 통계량을 서로 견주어 볼 때 쓸모 있을 것이다.

Key Words : signal detection, test statistics, correlation coefficient, nonparametric statistic, statistical signal processing.

ABSTRACT

In this paper, we address the derivation of joint distributions and correlation coefficients for three pairs of statistics used commonly in a number of signal detection schemes. The upper and lower bounds of the correlation coefficients for the three pairs are obtained, and interesting relationships between the correlation coefficients are derived. Explicit values of the correlation coefficients evaluated for some meaningful distributions are given in the form of tables and figures for easy reference. The results in this paper should be useful in comparing various detection statistics.

I. 머리말

신호 검파에 쓰이는 모수 통계량과 일부 비모수 통계량 사이의 상관 계수에 관한 연구 결과가 최근 발표되었다 [1]. 그 이전에는 서로 다른 두 통계량을 함께 쓰거나, 원래 쓰는 통계량 대신 다른 통계량을 써서 얻은 검파 기법의 성능을 분석한 연구가 거의 없었다. 신호 처리와 통신 분야에서 부호, 순위, 크기 순위의 비모수 신호 검파 기법이 많이 쓰이는데, 이 논문에서는 최근 발표된 논문에서는 다루어지지 않았던 비모수 통계량 사이의 상관계수들

을 다룬다. 이 상관계수로부터 서로 다른 두 통계량을 적절히 선택하여 신호를 더 잘 검파할 수 있다. 또한, 이러한 비모수 상관계수들의 상한과 하한을 얻고, 상관계수들 사이의 관계를 살펴본다.

II. 확률 벡터와 통계량

독립이고 분포가 같은(independent and identically distributed: i.i.d.) 확률 변수 n 개로 이루어진 관측 표본 $X = (X_1, X_2, \dots, X_n)$ 를 생각하자. 여기서, 확률 변수 $\{X_j\}_{j=1}^n$ 는 모두 평균이 $E\{X\} =$

* 세종대학교 정보통신공학과 (baej@sejong.ac.kr), † 교신저자
논문번호 : KICS2005-05-208, 접수일자 : 2005년 5월 23일

** 한국과학기술원 전자전산학과 (i.song@ieee.org)

※이 논문은 2004년도 한국학술재단의 지원에 의하여 연구되었음(KRP-2004-041-D00449).

m_X 이고 분산이 $V(X) = E\{(X - m_X)^2\} = \sigma_X^2$ 이며, 확률 밀도 함수와 확률 분포 함수는 각각 $f_X = f$ 와 $F_X = F$ 라고 하자.

이제, $Z_i = \text{sgn}(X_i)$ 는 X_i 의 부호, R_i 는 관측 표본 $\{X\}$ 에서 X_i 의 순위 (곧, X_i 보다 작거나 같은 관측량의 갯수+1), 그리고 Q_i 는 $|X| = (|X_1|, |X_2|, \dots, |X_n|)$ 에서 $|X_i|$ 의 순위로 두자. 여기서, 통계량 Q_i 는 X_i 의 크기 순위라 불리며, 부호 함수 $\text{sgn}(\cdot)$ 는 아래와 같다.

$$\text{sgn}(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases} \quad (1)$$

한편, X 의 r 째 순서 통계량 $U_r = X_{[r]}$ 의 확률 밀도 함수는 아래와 같다 [2].

$$f_{U_r}(x) = n \binom{n-1}{r-1} F^{r-1}(x) \cdot \{1 - F(x)\}^{n-r} f(x) \quad (2)$$

여기서, $\binom{n}{r} = \frac{n!}{(n-r)!r!}$ 는 이항계수이다. 한편,

$$G(y) = F(y) - F(-y), \quad (3)$$

$$g(y) = \frac{dG(y)}{dy}, \quad (4)$$

라 하면 $|X_i|$ 의 확률 분포 함수는 $F_{|X_i|}(y) = G(y)u(y)$ 이므로 $|X_i|$ 의 확률 밀도 함수는 $f_{|X_i|}(y) = g(y)u(y)$ 이다. 여기서, $u(y)$ 는 단위 계단 함수로, $y \geq 0$ 일 때 $u(y) = 1$ 이고 $y < 0$ 일 때 $u(y) = 0$ 이다. 따라서, (2)를 쓰면 $|X|$ 의 q 째 순서 통계량 $V_q = |X|_{[q]}$ 의 확률 밀도 함수를 아래와 같이 구할 수 있다.

$$f_{V_q}(y) = n \binom{n-1}{q-1} G^{q-1}(y) \cdot \{1 - G(y)\}^{n-q} g(y)u(y) \quad (5)$$

정의: 독립인 확률 변수 집합에서 한 확률 변수의 확률 밀도 함수가 h 이고 나머지 확률 변수의 확률 밀도 함수가 모두 f 일 때, 이 확률 변수들을 독립이고 분포가 거의 같은(*independent and semi-identically distributed: i.s.i.d.*) 확률 변수들이라 한다. ■

따라서, $h=f$ 이면, '독립이고 분포가 거의 같은 확률 변수'들은 '독립이고 분포가 같은 확률 변수'들이다. 이 논문에서는, 독립이고 분포가 거의 같은 확률 변수들의 집합에서 확률 밀도 함수가 h 인 확률 변수를 첫째 확률 변수 X_1 로 두겠다.

정리 1: 독립이고 분포가 거의 같은 확률 변수 집합 $\{X_j\}_{j=1}^n$ 을 생각하자. 이때, $r=1, 2, \dots, n$ 일 때 $\beta_i(r) = \Pr\{R_i = r\}$ 은 아래와 같다 [1,3].

$$\beta_1(r) = \binom{n-1}{r-1} \int_{-\infty}^{\infty} F^{r-1}(x) \cdot \{1 - F(x)\}^{n-r} h(x) \quad (6a)$$

$$\beta_i(r) = \frac{1 - \beta_1(r)}{n-1}, \quad i=2, 3, \dots, n. \quad (6b)$$

한편, 위 (6a)와 (6b)에서 $\sum_{r=1}^n \beta_1(r) = 1$ 임을 알 수 있다. 특히, $h=f$ 일 때는 $\beta_i(r) = \frac{1}{n}$, $i=1, 2, \dots, n$ 인데 이는 독립이고 분포가 같은 때에 얻을 수 있는 결과와 정확히 같다.

정리 2: 독립이고 분포가 거의 같은 n 확률변수에서, $\mathbb{F}_i(q) = \Pr\{Q_i = q\}$ 라 하자. 그러면, $q=1, 2, \dots, n$ 일 때 아래와 같은 식을 얻을 수 있다 [3].

$$\mathbb{F}_1(q) = \binom{n-1}{q-1} \int_{-\infty}^{\infty} G^{q-1}(x) \{1 - G(x)\}^{n-q} \cdot h(x) + h(-x)u(x) dx, \quad (7a)$$

$$\mathbb{F}_i(q) = \frac{1 - \mathbb{F}_1(q)}{n-1}, \quad i=2, 3, \dots, n. \quad (7b)$$

III. 비모수 통계량 사이의 상관 계수

독립이고 분포가 같은 관측 표본에서 상관 계수 ρ 의 표준화 상관 계수를 아래와 같이 정의하자.

$$\tilde{\rho} = \sqrt{\frac{n+1}{n-1}} \rho. \quad (8)$$

보기를 들어, 독립이고 분포가 같으며 크기가 n 인 관측 표본 Z 에서 $\tilde{\rho}_{Z,R_i} = \sqrt{\frac{n+1}{n-1}} \rho_{Z,R_i}$ 는 확률

변수 Z_i 와 그 순위 R_i 사이의 표준화 상관 계수를 뜻한다.

부록에서 보인 (41)과 (44)를 이용하여 아래와 같은 공분산을 얻을 수 있다.

$$\begin{aligned} \text{Cov}(Z_i, R_i) &= \int_{-\infty}^{\infty} \text{sgn}(x) \{1 + (n-1)F(x) \cdot f(x) dx - \frac{n+1}{2} \int_{-\infty}^{\infty} \text{sgn}(x) f(x) dx \\ &= \frac{n-1}{2} \int_{-\infty}^{\infty} \text{sgn}(x) \{2F(x) - 1\} f(x) dx \\ &= (n-1)F(0)\{1 - F(0)\}, \end{aligned} \quad (9)$$

$$\begin{aligned} \text{Cov}(Z_i, Q_i) &= \int_{-\infty}^{\infty} \text{sgn}(x) \{1 + (n-1)G(|x|) \cdot f(x) dx - \frac{n+1}{2} \int_{-\infty}^{\infty} \text{sgn}(x) f(x) dx \\ &= \frac{n-1}{2} \int_{-\infty}^{\infty} \text{sgn}(x) \{2G(|x|) - 1\} f(x) dx \\ &= (n-1)F(0) \left\{ 1 - \int_{-\infty}^{\infty} F(-x) f(x) dx \right\}. \end{aligned} \quad (10)$$

이를 바탕으로 상관계수를 아래와 같이 얻을 수 있다.

$$\tilde{\rho}_{Z,R_i} = \sqrt{3} \sqrt{F(0)\{1 - F(0)\}} \quad (11)$$

$$\tilde{\rho}_{Z,Q_i} = \sqrt{3} \frac{F(0) - \int_{-\infty}^{\infty} F(-x) f(x) dx}{\sqrt{F(0)\{1 - F(0)\}}} \quad (12)$$

한편 f 가 짝함수 일때 $F(0) = \frac{1}{2}$, $\int_{-\infty}^{\infty} F(-x) \cdot f(x) dx = \frac{1}{2}$ 이므로 $\tilde{\rho}_{Z,R_i} = \frac{\sqrt{3}}{2}$ 이고 $\tilde{\rho}_{Z,Q_i} = 0$ 이다. 여기서 $F(0) = 0$, $F(x) = 1$ 일때 $\int_{-\infty}^{\infty} F(-x) f(x) dx = 0$ 이고 $\int_{-\infty}^{\infty} F(-x) f(x) dx = 1$ 이라는 것을 새겨들 만하다. 이것은 $F(0) = 0$ 또는 $F(0) = 1$ 일때 $\text{Cov}(Z_i, R_i) = \text{Cov}(Z_i, Q_i) = 0$ 인 것을 의미한다.

정리 3: 표준화 상관 계수 $\tilde{\rho}_{Z,R_i}$ 과 $\tilde{\rho}_{Z,Q_i}$ 은 다음 조건을 만족시킨다.

$$0 \leq \tilde{\rho}_{Z,R_i} \leq \frac{\sqrt{3}}{2} \quad (13)$$

$$\tilde{\rho}_{Z,Q_i} \leq \tilde{\rho}_{Z,R_i} \quad (14)$$

$$-\frac{\sqrt{3}}{2} \leq \tilde{\rho}_{Z,Q_i} \leq \frac{\sqrt{3}}{2} \quad (15)$$

(13)번 식에서 $F(0) = \frac{1}{2}$ 일 때 위쪽 경계에, $F(0) = 0$ 또는 1일 때 아래쪽 경계에 이르게 된다.

증명: 먼저 (11)로부터 (13)를 얻는다. 다음 $x \geq 0$ 일때 $0 \leq F(-x) \leq F(0) \leq F(x) \leq 1$ 이므로 다음 식을 얻을 수 있다.

$$0 \leq \int_0^{\infty} F(-x) f(x) dx \leq F(0) \int_0^{\infty} f(x) dx \quad (16)$$

$$\begin{aligned} F(0) \int_0^{\infty} f(-x) dx &\leq \int_0^{\infty} F(x) f(-x) dx \\ &\leq \int_0^{\infty} f(-x) dx \end{aligned} \quad (17)$$

$$\text{또한 } \int_{-\infty}^{\infty} F(-x) f(x) dx$$

$$= \int_0^{\infty} F(-x) f(x) + F(x) f(-x) dx$$

이므로 우리는 다음과 같은 결과를 얻는다.

$$\begin{aligned} -F(0)\{1 - F(0)\} &\leq F(0) - \int_{-\infty}^{\infty} F(-x) f(x) dx \\ &\leq F(0)\{1 - F(0)\} \end{aligned} \quad (18)$$

(12), (18)로부터 (14)를 얻고 (11),(12), (18)로부터 (15)를 얻는다.

4절에서는 (13)에 보여지는 위쪽 경계와 아래쪽 경계에 이르는 잘 알려진 분포들의 보기 몇 개가 주어진다. (14)에 보여지는 위쪽 경계와 아래쪽 경계에 도달하는 확률 밀도 함수를 위한 충분 조건은 정리4에 기술되어 있다.

정리 4: 확률 밀도 함수 f 가 구간 $a < b \leq 0 \leq c < d$ 에서 $\int_a^b f(x) dx = \int_c^d f(x) dx = \frac{1}{2}$ 이라 하자.

이때 $-b \geq d$ 라면 $\tilde{\rho}_{Z,Q_i} = -\frac{\sqrt{3}}{2}$, $-a \leq c$ 라면

$\tilde{\rho}_{Z,Q_i} = \frac{\sqrt{3}}{2}$ 이다.

증명: $x \geq -a$ 일 때 $F(-x) = 0$, $-c \leq x \leq -b$

일 때 $F(-x) = \frac{1}{2}$, $x \leq -d$ 일 때 $F(-x) = 1$ 이다. 또한 구간 $(-\infty, a), (b, c), (d, \infty)$ 일 때 $f(x) = 0$ 이다. $-b \geq d$ 일 때 $a < b \leq -d < -c \leq c < d \leq -b < -a$ 이므로 다음 식을 얻을 수 있으며,

$$\begin{aligned} \int_{-\infty}^{\infty} F(-x)f(x)dx &= \int_{-\infty}^{-d} 1 \cdot f(x)dx \\ &+ \int_{-d}^{-c} F(-x)f(x)dx + \int_{-c}^{-b} \frac{1}{2} f(x)dx \\ &+ \int_{-b}^{\infty} F(-x)f(x)dx \\ &= \int_a^b f(x)dx = \int_c^d \frac{1}{2} f(x)dx = \frac{3}{4} \end{aligned} \quad (19)$$

$-a \leq c$ 일 때 $-d < -c \leq a < b \leq 0 \leq -b < -a \leq c < d$ 이므로 다음 식을 얻을 수 있다.

$$\begin{aligned} \int_{-\infty}^{\infty} F(-x)f(x)dx &= \int_{-\infty}^{-c} F(-x)f(x)dx \\ &+ \int_{-c}^{-b} \frac{1}{2} f(x)dx + \int_{-b}^{-a} F(-x)f(x)dx \\ &+ \int_{-a}^{\infty} F(-x)f(x)dx \\ &= \int_a^b \frac{1}{2} f(x)dx = \frac{1}{4} \end{aligned} \quad (20)$$

$f(x) = \{u(x+2) - u(x+1) + u(x-3) - u(x-4)\}/2$ 일 때 $\hat{p}_{Z,Q_i} = \frac{\sqrt{3}}{2}$, $f(x) = \{u(x+4) - u(x+3) + u(x-1) - u(x-2)\}/2$ 일 때 $\hat{p}_{Z,Q_i} = -\frac{\sqrt{3}}{2}$ 라는 것을 정리 4로부터 얻을 수 있다.

한편 $F(0) = 1$ 일 때 $|X_i| = -X_i$ 이고, $Q_i = n+1 - R_i$ 라는 것을 알 수 있다. 이때 (46)으로부터 $Cov(R_i, Q_i) = E\{R_i Q_i\} - E\{R_i\}E\{Q_i\} = \frac{(n+1)(n+2)}{6} - \frac{(n+1)^2}{4} = -\frac{n^2-1}{12}$ 임을 알 수 있고, 다음을 얻을 수 있다.

$$\rho_{R,Q_i} = -1 \quad (21)$$

또한 $F(0) = 0$ 일 때 $|X_i| = X_i$, $Q_i = R_i$ 을 알 수 있고, (48)의 몇몇 과정을 사용하여 다음을 얻을 수 있다.

$$\rho_{R,Q_i} = 1 \quad (22)$$

이제부터 $0 < F(0) < 1$ 인 경우를 고려해보자.

$$\begin{aligned} \int_0^{\infty} \{F(x)f(-x) - F(x)f(x)\} dx &= -F(x) \cdot F(-x) \Big|_0^{\infty} = F^2(0) \text{ 와 } \int_{-\infty}^{\infty} F(x)f(x)dx = \frac{1}{2} \text{ 로부터} \\ \int_{-\infty}^{\infty} G(|x|)f(x)dx &= \int_{-\infty}^0 F(-x)f(x)dx \\ &- \int_0^{\infty} F(-x)f(x)dx + \int_0^{\infty} F(x)f(x)dx \\ &+ \int_{-\infty}^0 F(x)f(x)dx \\ &= \int_0^{\infty} \{F(x)f(-x) - F(-x)f(x)\} dx \\ &- F^2(0) + \frac{1}{2} = \frac{1}{2} \end{aligned} \quad (23)$$

임을 알 수 있고, (50)를 써서 다음을 얻는다.

$$\begin{aligned} Cov(R_i, Q_i) &= \int_{-\infty}^{\infty} \{1(n-1)G(|x|)\} \{1 \\ &+ (n-1)F(x)\} f(x)dx - \frac{(n+1)^2}{4} \\ &= (n-1)^2 \left\{ \int_{-\infty}^{\infty} G(|x|)F(x)f(x)dx \right. \\ &\left. - \frac{1}{4} \right\} \end{aligned} \quad (24)$$

$0 < F(0) < 1$ 일 때 식 (24)는 $\hat{\rho}_{R,Q_i} = \frac{n+1}{n-1} \rho_{R,Q_i}$ 임을 쓰면, 아래와 같다.

$$\hat{\rho}_{R,Q_i} = 12 \left\{ \int_{-\infty}^{\infty} G(|x|)F(x)f(x)dx - \frac{1}{4} \right\} \quad (25)$$

$0 < F(0) < 1$ 이고, 확률 분포 함수 f 가 짝함수일 때 $\int_{-\infty}^{\infty} G(|x|)F(x)f(x)dx = \int_{-\infty}^{\infty} G(x)\{1 - F(x)\} \cdot f(x)dx + \int_0^{\infty} G(x)F(x)f(x)dx = \int_0^{\infty} G(x) = \int_0^{\infty} G(x)f(x)dx = \int_0^{\infty} \{2F(x) - 1\}f(x)dx = F(0)\{1 - F(0)\} = \frac{1}{4} (0) = \frac{1}{4}$ 이므로 $\hat{\rho}_{R,Q_i} = 0$ 임을 알 수 있다.

IV. 몇 개의 보기들

4.1 고른 분포

아래와 같은 조건일 때

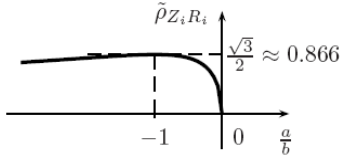


그림 1. 고른 분포 $U(a, b)$, $a < 0 < b$ 에서의 표준화 상관계수 $\tilde{\rho}_{Z_i R_i}$

$$F(0) = \begin{cases} 1, & a < b \leq 0, \\ \frac{-a}{b-a}, & a < 0 < b, \\ 0, & 0 \leq a < b. \end{cases} \quad (26)$$

(11)의 식으로부터 구간 $a < b \leq 0$ 또는 $0 \leq a < b$ 에서 값이 0인 다음 결과를 얻을 수 있다.

$$\tilde{\rho}_{Z_i R_i} = \frac{\sqrt{3}}{2} \frac{\sqrt{-ab}}{b-a} \cdot \{ \text{sgn}(b) - \text{sgn}(a) \} \quad (27)$$

(27)에서 $a = -b < 0$ 일 경우 최고값은 $\frac{\sqrt{3}}{2}$ 인 것을 그림 1에서 확인할 수 있다. 또한 계속해서 다음을 얻을 수 있다.

$$\int_{-\infty}^{\infty} F(-x)f(x)dx = \begin{cases} 1, & a < b \leq 0, \\ 1 - \frac{2b^2}{(b-a)^2}, & a < -b \leq 0, \\ \frac{2a^2}{(b-a)^2}, & -b < a \leq 0, \\ 0, & 0 \leq a < b. \end{cases} \quad (28)$$

(12)의 식에 (26), (28)를 이용하여 아래와 같이 얻을 수 있고,

$$\begin{aligned} \tilde{\rho}_{Z_i Q_i} &= \frac{\sqrt{3}}{2} \frac{b+a}{b-a} \sqrt{\frac{\max(a, -b)}{\min(a, -b)}} \\ &\quad \cdot \{ \text{sgn}(b) - \text{sgn}(a) \} \end{aligned} \quad (29)$$

$a < b \leq 0$ 혹은 $0 \leq a < b$ 일 경우 0이다. 여기서 $a < 0 < b$ 일때 $\max(a, -b)$ 와 $\min(a, -b)$ 모두 음수인 것과 $a < b$ 이므로 $\min(a, -b)$ 가 항상 음수인 것에 주목할 만하다. 그림 2에 $\tilde{\rho}_{Z_i Q_i}$ 와 $\frac{a}{b}$ 의 관계를 나타내었다. $\frac{a}{b} = 2 - \sqrt{5} \approx -0.236$ 일때 최고값 $\frac{\sqrt{3}}{2} \sqrt{\sqrt{5}-2}(\sqrt{5}-1) \approx 0.5201$, $\frac{a}{b} = -2 - \sqrt{5} \approx -4.236$ 일때 최소값 $-\frac{\sqrt{3}}{2} \sqrt{\sqrt{5}-2}(\sqrt{5}-1)$

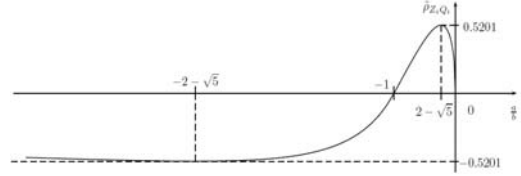


그림 2. 고른 분포 $U(a, b)$, $a < 0 < b$ 에서의 표준화 상관계수 $\tilde{\rho}_{Z_i Q_i}$

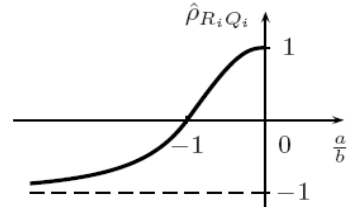


그림 3. 고른 분포 $U(a, b)$, $a < 0 < b$ 에서의 표준화 상관계수 $\hat{\rho}_{R_i Q_i}$

≈ -0.5201 를 갖는다.

마지막으로 (21), (22), (26)로부터 $a < b \leq 0$ 일때는 $\rho_{R_i Q_i} = -1$, $0 \leq a < b$ 일때는 $\rho_{R_i Q_i} = 1$ 인 것을 알 수 있다. $a < 0 < b$ 일때는 다음을 얻을 수 있다.

$$\begin{aligned} \int_{-\infty}^{\infty} G(|x|)F(x)f(x)dx &= \begin{cases} \frac{1}{6} \left\{ 1 + \frac{4b^3}{(b-a)^3} \right\}, & a < -b \leq 0, \\ \frac{1}{3} \left\{ 1 + \frac{2a^3}{(b-a)^3} \right\}, & -b < a \leq 0, \end{cases} \end{aligned} \quad (30)$$

(25)를 써서 아래와 같은 결과를 얻을 수 있으며,

$$\hat{\rho}_{R_i Q_i} = \text{sgn}(a+b) \left[1 + \left\{ \frac{2 \max(a, -b)}{b-a} \right\}^3 \right] \quad (31)$$

이를 그림 3에 나타내었다. □

4.2 평균이 0인 비대칭 분포

확률 밀도 함수가

$$f_a(x) = \begin{cases} \frac{-a}{a+1}, & -1 \leq x < 0, \\ \frac{1}{a(a+1)}, & 0 \leq x < a, \\ 0, & \text{그 밖에는} \end{cases} \quad (32)$$

라 하자. 여기서, $a > 0$ 이다. 확률 밀도 함수 $f_a(x)$ 는 평균이 0이고 $a \neq 1$ 일때는 대칭이 아니다. 먼저, (11)에서 아래 표준화 상관 계수를 쉽게 얻을 수 있다.

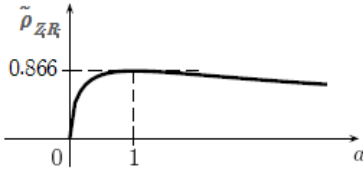


그림 4. 확률밀도함수 f_a 에서의 표준화 상관계수 $\tilde{\rho}_{Z,R_i}$

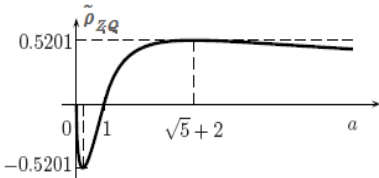


그림 5. 확률밀도함수 f_a 에서의 표준화 상관계수 $\tilde{\rho}_{Z,Q_i}$

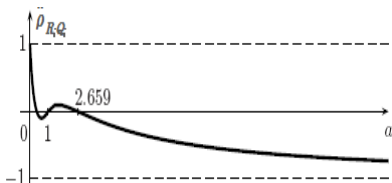


그림 6. 확률밀도함수 f_a 에서의 표준화 상관계수 $\hat{\rho}_{R,Q_i}$

$$\tilde{\rho}_{Z,R_i} = \frac{\sqrt{3a}}{a+1} \quad (33)$$

그 밖의 상관 계수를 얻으면 아래와 같다.

$$\tilde{\rho}_{Z,Q_i} = \frac{\sqrt{3v(v-1)}}{v+1} \quad (34)$$

$$\hat{\rho}_{R,Q_i} = \frac{(v-1)(2v^3+v^2+2v-1)}{(v+1)^3} \quad (35)$$

표 1. 고른분포와 평균이 0인 비대칭 분포의 표준화 상관계수

($B_{ab} = \max(a, -b)$, $S_{ab} = \max(a, b)$, $D_{ab} = \text{sgn}(b) - \text{sgn}(a)$, $v = \min(a, a^{-1})$)

	정규 확률밀도함수	평균이 0인 비대칭 분포
$\tilde{\rho}_{Z,R_i}$	$\frac{\sqrt{3}}{2} \frac{\sqrt{-ab}}{b-a} D_{ab}$ $0 \leq \tilde{\rho}_{Z,R_i} \leq \frac{\sqrt{3}}{2} \approx 0.866$	$\frac{\sqrt{3a}}{a+1}$ $0 \leq \tilde{\rho}_{Z,R_i} \leq \frac{\sqrt{3}}{2} \approx 0.866$
$\tilde{\rho}_{Z,Q_i}$	$\frac{\sqrt{3}}{2} \frac{b+a}{b-a} \sqrt{\frac{B_{ab}}{S_{ab}}} D_{ab}$ $-0.5201 \leq \tilde{\rho}_{Z,Q_i} \leq 0.5201$	$\frac{\sqrt{3v(v-1)}}{v+1}$ $-0.5201 \leq \tilde{\rho}_{Z,Q_i} \leq 0.5201$
$\hat{\rho}_{R,Q_i}$	$\text{sgn}(a+b) \left[1 + \left\{ \frac{2B_{ab}}{b-a} \right\}^3 \right]$ $-1 \leq \hat{\rho}_{R,Q_i} \leq 1$ 만약 $a < b \leq 0, 0 \leq a < b$ 면 $\hat{\rho}_{R,Q_i} = 1$	$\frac{(v-1)(2v^3+v^2+2v-1)}{(v+1)^3}$ $-1 \leq \hat{\rho}_{R,Q_i} \leq 1$

이들을 각각 그림 4-6에 보였다. 한편 $\tilde{\rho}_{Z,R_i}$ 일 경우 $a=1$ 일 때 최댓값 $\frac{\sqrt{3}}{2} \approx 0.866$ 을 갖고, $a \rightarrow 0$ 혹은 $a \rightarrow \infty$ 이면 0으로 수렴한다. $\tilde{\rho}_{Z,Q_i}$ 일 경우 $a=\sqrt{5}-2 \approx 0.236$ 일 때 최솟값 $-\frac{\sqrt{3}}{2} \cdot \sqrt{\sqrt{5}-2}$ ($\sqrt{5}-1$) ≈ -0.5201 , $a=\sqrt{5}+2 \approx 4.236$ 일 때 최댓값 $\frac{\sqrt{3}}{2} \sqrt{\sqrt{5}-2}$ ($\sqrt{5}-1$) ≈ 0.5201 을 갖는다. $\hat{\rho}_{R,Q_i}$ 는 $a=\sqrt{7}-2 \approx 0.646$ 일 때, $\hat{\rho}_{R,Q_i} \approx -0.0991$, $a=\frac{\sqrt{7}+2}{3} \approx 1.5486$ 일 때 $\hat{\rho}_{R,Q_i} \approx 0.0991$ 의 최댓값과 최솟값을 갖는다. 또한 $a=0.3761, 1, 2.659$ 일 때 영교차점들을 갖는다. 4절에서 얻은 몇몇 결과를 표 1에 정리한다.

V. 맺음말

이 논문에서는 독립이고 분포가 같은 관측표본에서 대표적인 비모수 통계량 셋, 곧, 관측량의 부호, 순위, 그리고 크기 순위 통계량 사이의 상관 계수를 닫힌 꼴로 이끌어 내었다. 상관 계수의 몇 가지 특징들을 얻었으며 이를 분석한 결과, 직관적인 예상과 일치하였다. 더불어, 상관 계수의 상한값과 하한값을 얻은 뒤, 의미 있는 몇몇 분포에서 상관 계수 값을 구체적으로 얻었다. 또한, 상한값과 하한값이 나올 수 있는 분포들을 찾아내어 소개하였다. 이 논문에서 얻은 결과는 신호 처리에서 비모수 통계량을 쓸 때 각 검파기들의 성능을 예상할 수 있는 이

론적인 것대로 쓸 수 있을 것이다.

부록. 비모수 통계량들의 결합 확률질량함수의 유도

결합 확률 질량 함수 f_{Z,R_i} 와 f_{Z,Q_i}

먼저, $(1+x)^{n-1} = \sum_{i=0}^{n-1} \binom{n-1}{i} x^i$ 와 $(n-1) \cdot (1+x)^{n-2} = \sum_{i=1}^{n-1} i \binom{n-1}{i} x^{i-1}$ 을 쓰면, $\sum_{i=1}^n \binom{n}{i-1} C_{k-1} w^{k-1} (1-w)^{n-k} = 1$ 과 $\sum_{k=1}^n k \binom{n}{k-1} C_{k-1} w^{k-1} (1-w)^{n-k} = 1 + (n-1)w$ 을 바로 얻는다. 따라서, 아래를 얻을 수 있다.

$$\frac{1}{n} \sum_{r=1}^n r f_{U_r}(x) = (n-1)F(x)f(x), \quad (36)$$

$$\frac{1}{n} \sum_{q=1}^n q f_{V_q}(y) = 1 + (n-1) \cdot G(y)g(y)u(y). \quad (37)$$

한편 (6)을 통하여 다음을 알 수 있고,

$\Pr\{R_i=r|Z_i=1\} = \Pr\{\{X_{j,j=1}^n\}$ 이 독립이고 분포가 거의 같으며 확률 밀도 함수 짝이

$$\left(\frac{f(x)u(x)}{1-F(0)}, f(x)\right) \text{일때 } R_1=r \\ = \int_{-\infty}^{\infty} \binom{n-1}{r-1} F^{r-1}(x) \{1-F(x)\}^{n-r} \cdot \frac{f(x)u(x)}{1-F(0)} dx \\ = \frac{1}{n(1-F(0))} \int_0^{\infty} f_{U_r}(x) dx \quad (38)$$

$\Pr\{R_i=r|Z_i=-1\} = \Pr\{\{X_{j,j=1}^n\}$ 이 독립이고 분포가 거의 같으며 확률밀도함수 짝이

$$\left(\frac{f(x)u(-x)}{F(0)}, f(x)\right) \text{일때 } R_1=r \\ = \frac{1}{nF(0)} \int_{-\infty}^0 f_{U_r}(x) dx. \quad (39)$$

(38), (39)로부터 다음을 얻을 수 있다.

$$\Pr\{Z_i=\pm 1, R_i=r\} = \int_{\pm} f_{U_r}(x) dx \cdot \Pr\{R_i=r\} = \frac{1}{n} \int_{\pm} f_{U_r}(x) dx. \quad (40)$$

$\Pr\{Z_i=1\} \Pr\{R_i=r\} \neq \Pr\{Z_i=1, R_i=r\}$ 이고 $\Pr\{Z_i=-1\} \Pr\{R_i=r\} \neq \Pr\{Z_i=-1, R_i=r\}$ 이기 때문에 일반적으로 R_i 와 Z_i 는 서로 독립이 아니라는 것에 주목할 만하다. (36)과 (40)으로부터 결합 적률을 다음과 같이 얻을 수 있다.

$$E\{Z_i R_i\} = \sum_{r=1}^n r \int_{-\infty}^{\infty} \text{sgn}(x) f_{U_r}(x) dx \\ = \int_{-\infty}^{\infty} \text{sgn}(x) \{1+(n-1)F(x)\} \cdot f(x) dx. \quad (41)$$

X_i 의 조건부 확률 밀도 함수는 $\{Z_i=1\}$ 일 때 $f(x)u(x)/(1-F(0))$ 이다. 결과적으로 $\{X_{j,j=1}^n\}$ 는 독립이고 분포가 거의 같은 확률 밀도 함수 짝이 $(f(x)u(x)/(1-F(0)), f(x))$ 인 확률 변수이다. 따라서 (7)과 $x \neq 0$ 일때 $\{f(x)u(x)+f(-x) \cdot u(-x)\} \{f(x)u(x)+f(-x)u(-x)\}u(x) = f(x)u(x)$ 로 부터 다음을 얻는다 [3].

$$\Pr\{Z_i=1, Q_i=q\} = \Pr\{Q_i=q|Z_i=1\} \Pr\{Z_i=1\} \\ = \{1-F(0)\} \Pr\{\{X_{j,j=1}^n\} \text{이 독립이고 분포가 거의 같으며 확률 밀도 함수 짝이 } \\ \left(\frac{f(x)u(x)}{1-F(0)}, f(x)\right) \text{일때 } Q_1=q\} \\ = \{1-F(0)\} \binom{n-1}{q-1} \int_{-\infty}^{\infty} G^{q-1}(x) \{1-G(x)\}^{n-q} \cdot \frac{\{f(x)u(x)+f(-x)u(-x)\}u(x)}{1-F(0)} dx \\ = \binom{n-1}{q-1} \int_0^{\infty} G^{q-1}(x) \{1-G(x)\}^{n-q} f(x) dx \\ = \frac{1}{n} \int_0^{\infty} f_{V_q}(x) \frac{f(x)}{g(x)} dx. \quad (42)$$

비슷하게 $\{Z_i=-1\}$ 일 때 (7)로부터 다음을 얻는다.

$$\Pr\{Z_i=-1, Q_i=q\} \\ = F(0) \Pr\{\{X_{j,j=1}^n\} \text{이 독립이고 분포가 거의 같으며 확률 밀도 함수 짝이 } (f(x)u(-x)/F(0), f(x)) \text{일 때 } Q_1=q\} \\ = \binom{n-1}{q-1} \int_0^{\infty} G^{q-1}(x) \{1-G(x)\}^{n-q} f(-x) dx \\ = \frac{1}{n} \int_0^{\infty} f_{V_q}(x) \frac{f(-x)}{g(x)} dx. \quad (43)$$

일반적으로 Q_i 와 Z_i 는 서로 독립이 아니라는 것

에 주목할 만하다. 반면 f 가 짝함수일때 $\Pr\{Z_i=1, Q_i=q\} = {}_{n-1}C_{q-1} \cdot \int_0^\infty \{2F(x)-1\}^{q-1}\{2F(x)\}^{n-q}f(x)dx$
 $= {}_{n-1}C_{q-1} \int_0^1 t^{q-1}(1-t)^{n-q} \frac{1}{2} dt = \frac{1}{2n}$
 $= \Pr\{Z_i=1\}\Pr\{Q_i=q\}$ 이고 $\Pr\{Z_i=-1, Q_i=q\} = \Pr\{Q_i=q\}\Pr\{Z_i=-1\}$ 이므로 Z_i 와 Q_i 가 독립이라는 것을 알 수 있다. Z_i 와 Q_i 의 결합 적률은 (42)와 (43)으로부터 다음과 같이 구할 수 있다.

$$E\{Z_i Q_i\} = \sum_{z_i=1}^n \sum_{q_i} q z_i \Pr\{Q_i=q, Z_i=z_i\} \\ = \int_0^\infty \{1+(n-1)G(x)\}f(x)dx - \int_0^\infty \{1+(n-1)G(x)\}f(-x)dx \\ = \int_{-\infty}^\infty \text{sgn}(x)\{1+(n-1)G(|x|)\} \cdot f(x)dx. \quad (44)$$

결합 확률 질량 함수 f_{R_i, Q_i}

R_i 와 Q_i 의 결합 확률 질량 함수를 유도하기 위해서 먼저 R_i 와 Q_i 가 선형 독립인지(경우 1과 2) 아닌지를(경우 3) 알 필요가 있다.

경우 1. $F(0)=1$ 일때: $|X_i|=-X_i$ 과 $Q_i=n+1-R_i$ 임을 알 수 있고 따라서

$$\Pr\{R_i=r, Q_i=q\} = \begin{cases} \frac{1}{n}, & 1 \leq q = n+1-r \leq n, \\ 0, & \text{그 밖에는,} \end{cases} \quad (45)$$

이고 다음의 결과를 얻을 수 있다.

$$E\{R_i Q_i\} = \sum_{r=1}^n \sum_{q=1}^n \Pr\{R_i=r, Q_i=q\}qr \\ = \frac{1}{n} \sum_{r=1}^n (n+1-r)r \\ = \frac{(n+1)(n+2)}{6}. \quad (46)$$

경우 2. $F(0)=0$ 일때: $|X_i|=X_i$ 과 $Q_i=R_i$ 임을 알 수 있고 따라서

$$\Pr\{R_i=r, Q_i=q\} = \begin{cases} \frac{1}{n}, & 1 \leq q = r \leq n, \\ 0, & \text{그 밖에는,} \end{cases} \quad (47)$$

이고 다음의 결과를 얻을 수 있다.

$$E\{R_i Q_i\} = \frac{1}{n} \sum_{r=1}^n r^2 = \frac{(n+1)(2n+1)}{6}. \quad (48)$$

경우 3. $0 < F(0) < 1$ 일때: 주어진 $R_i=r$ 은 $\{X_j\}_{j=1}^n$ 가 독립이고 분포가 거의 같은 확률 밀도 함수 짝이 (f_{U_r}, f)인 확률 변수라는 말과 같다. 따라서 (7)로부터 다음을 얻을 수 있다.

$$\Pr\{R_i=r, Q_i=q\} = \Pr\{Q_i=q | R_i=r\} \cdot \Pr\{R_i=r\} = \frac{1}{n} \Pr\{\{X_j\}_{j=1}^n \text{이 독립이고 분포가 거의 같으며 확률밀도함수 짝이 } (f_{U_r}, f) \text{일때 } Q_1=q\} \\ = \frac{1}{n} {}_{n-1}C_{q-1} \int_{-\infty}^\infty G^{q-1}(x)\{1-G(x)\}^{n-q} \cdot \{f_{U_r}(x)+f_{U_r}(-x)\}u(x)dx \\ = \frac{1}{n^2} \int_{-\infty}^\infty \frac{f_{V_r}(x)}{g(x)} \{f_{U_r}(x)+f_{U_r}(-x)\}dx \quad (49)$$

(49)로부터 $\Pr\{Q_i=q | R_i=r\} = \Pr\{R_i=r | Q_i=q\}$
 $= \frac{1}{n} \int_{-\infty}^\infty \frac{f_{V_r}(x)}{g(x)} \{f_{U_r}(x)+f_{U_r}(-x)\}dx$ 을 얻을 수 있다. R_i 와 Q_i 의 결합 적률은 (36), (37), (49)로부터 다음과 같이 구할 수 있다.

$$E\{R_i Q_i\} = \frac{1}{n^2} \sum_{r=1}^n \sum_{q=1}^n \int_{-\infty}^\infty \frac{f_{V_r}(x)}{g(x)} \cdot \{f_{U_r}(x)+f_{U_r}(-x)\}dx \\ = \frac{1}{n} \sum_{r=1}^n r \int_{-\infty}^\infty \{1+(n-1)G(x)\} \cdot \{f_{U_r}(x)+f_{U_r}(-x)\}u(x)dx \\ = \int_0^\infty \{1+(n-1)G(x)\} \cdot [\{1+(n-1)F(x)\}f(x) + \{1+(n-1)F(-x)\}f(-x)]dx \\ = \int_0^\infty \{1+(n-1)G(|x|)\} \cdot \{1+(n-1)F(x)\}f(x)dx. \quad (50)$$

참 고 문 헌

- [1] 권형문, 최상원, 이주미, 박소령, 배진수, 송익호, “여러 가지 신호 검파 문제에 쓰이는 모수와 비모수 검정 통계량 사이의 상관계수”, 한국통신학회 논문지 (게재예정).
- [2] I. Song, J. Bae, and S.Y. Kim, *Advanced Theory of Signal Detection*, Springer-Verlag, Heidelberg, 2002.
- [3] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, 6th Ed., Academic Press, New York, 2000.

주 현 (Hyun Joo)

준회원



2005년 2월 세종대학교 정보통신공학과(공학사)
 2005년 3월~현재 세종대학교 정보통신공학과 석사과정
 <관심분야> 신호처리

배진수 (Jinsoo Bae)

중신회원



1990년 2월 경기과학기술대학교 조기졸업(우등)
 1993년 2월 한국과학기술원 전기전산학과(공학사, 전체차석, 조기졸업, 최우등)
 1995년 2월 한국과학기술원 전기전산학과(공학석사)
 1998년 2월 한국과학기술원 전기전산학과(공학박사)
 1997년 1월~1997년 12월 동경대학 객원연구원
 1998년 1월~1998년 10월 앤더슨컨설팅 컨설턴트
 1998년 11월~1999년 12월 일본모토로라 연구원
 1999년 9월~2000년 2월 LG텔레콤 과장
 2000년 3월~현재 세종대학교 정보통신공학과 전임강사, 조교수
 <관심분야> 신호검파이론

송익호 (Ickho Song)

중신회원



1982년 2월 서울대학교 전자공학과(공학사, 준최우등)
 1984년 2월 서울대학교 전자공학과(공학석사)
 1985년 8월 펜실베이니아대학교 전기공학과(공학석사)
 1987년 3월~1998년 2월 벨 통

신연구소 연구원
 1988년 3월~현재 한국과학기술원 전자전산학과 조교수, 부교수, 교수
 1995년 1월~현재 한국통신학회 논문지 편집위원
 1991년 11월, 1996년 11월 한국통신학회 학술상받음
 1993년 11월 한국음향학회 우수연구상 받음
 1998년 11월 한국통신학회 LG학술상 받음
 1999년 11월 대한 전자공학회 해동논문상 받음
 2000년 3월 젊은 과학자상 받음
 2000년 11월 한국통신학회 모토롤라학술상 받음
 대한전자공학회, 한국음향학회, 한국통신학회 중신회원; IEE 석학회원; IEEE 준석학회원