

신호 검파 문제에 쓰는 모수와 비모수 검정 통계량 사이의 상관계수

정회원 박 소 령*, 준회원 권 형 문**, 종신회원 배 진 수***,
 준회원 최 상 원**, 준회원 이 주 미** 종신회원 송 익 호**

Correlation Coefficients between Parametric and onparametric Test Statistics for Signal Detection Problems

So Ryoung Park* *Regular Member*, Hyoungmoon Kwon** *Associate Member*,
 Jinsoo Bae*** *Lifelong Member*, Sang Won Choi** *Associate Member*
 Jumi Lee** *Associate Member* and Ickho Song** *Lifelong Member*

요 약

이 논문에서는 모수와 비모수 신호 검파에서 널리 쓰이는 네 통계량 사이의 결합 분포와 상관계수를 얻는다. 아울러, 상관계수들의 상한과 하한을 얻고, 상관계수들 사이의 눈여겨볼 만한 관계를 살펴본다. 또한, 몇 가지 분포에서 얻은 상관계수 값을 표와 그림으로 정리하여 쉽게 참고할 수 있도록 하였다. 이 논문에서 얻은 결과는 여러 검정 통계량을 서로 견주어 볼 때 쓸모 있을 것이다.

Key Words : signal detection, test statistics, correlation coefficient, nonparametric detection, parametric detection.

ABSTRACT

In this paper, we address the derivation of joint distributions and correlation coefficients for four pairs of statistics used commonly in a number of signal detection schemes. The upper and lower bounds of the correlation coefficients are obtained, and interesting relationships between the correlation coefficients are derived. Explicit values of the correlation coefficients 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

이제까지 여러 연구에서 갖가지 통계량을 신호 검파에 알맞도록[1-3] 써 왔다. 아울러, 어떤 통계량을 바탕으로 한 신호 검파 기법이 최적이고 쓸모

있으며 때때로 강인하다는 것이 여러 연구에서 밝혀졌다[4, 5]. 신호 처리와 통신 분야에서 널리 쓰이는 여러 통계량 가운데, 특히 부호, 순위, 크기 순위는 비모수 신호 검파 기법에서 자주 쓰인다[6-10]. 그러나, 서로 다른 두 통계량을 함께 쓰거나 원래 쓴

* 가톨릭대학교 정보통신전자공학부 (srpark@catholic.ac.kr),

** 한국과학기술원 전자전산학과 ({kwon, swchoi, jmlee}@Sejong.kaist.ac.kr, i.song@ieee.org)

*** 세종대학교 정보통신공학과 (baej@sejong.ac.kr)

논문번호 : KICS2004-12-329, 접수일자 : 2004년 12월 22일

※ 이 논문은 2005년 가톨릭대학교 교비연구비를 지원 받아 연구한 결과입니다.

통계량 대신 다른 통계량을 써서 얻은 검파 기법을 분석한 연구가 거의 없었다는 사실은 새겨들 만하다.

신호를 잘 검파하고자 서로 다른 두 통계량을 어떻게 골라 함께 쓸 것인지, 또는 원래 통계량과 바꾸어 쓸 수 있는 다른 통계량을 어떻게 고를 것인지, 두 통계량 사이의 상관계수를 써서 판단할 수 있다. 보기를 들어, 검정 통계량이 각각 $T_1(X) =$

$\sum_{i=1}^n e_i X_i$ 와 $T_2(X) = \sum_{i=1}^n e_i R_i$ 인 선형 검파기와 선형 순위 검파기를 [7] 생각해보자. 여기서, $X = (X_1, X_2, \dots, X_n)$ 는 관측 (또는 자료) 벡터이고, $\{e_i\}_{i=1}^n$ 는 (아는) 신호를 뜻하며, R_i 는 관측벡터 X 에서 X_i 의 순위이다. 이때, 관측벡터의 한 원소 X_i 와 그 순위 R_i 사이의 상관계수는 (꽤 자주) 1에 가깝기 때문에, 아는 신호 $\{e_i\}_{i=1}^n$ 를 검파하는 선형 검파기와 선형 순위 검파기는 그 성능이 거의 비슷하다는 것이 이미 잘 알려져 있다[7, 11].

또 다른 보기로서, 순위 검파기를 쓰고 있을 때, 순위와 부호 통계량, 그리고 순위와 크기 순위 통계량 사이의 상관계수 값이 각각 0.8과 0.3이라 하자. 이때, 순위와 크기 순위 통계량 사이의 상관계수 값이 상대적으로 좀더 작으므로, 원래 쓰고 있던 순위 통계량에 크기 순위 통계량을 같이 잘 어울러 쓴다면 성능이 좀더 좋은 검파 방법을 찾을 수 있을 것이다. 왜냐하면, 두 통계량 사이의 상관계수 값이 작다는 것은 두 통계량이 공통으로 지니고 있는 정보가 거의 없다는 것을 뜻한다고 볼 수 있기 때문이다.

이와 같이, 신호 검파를 비롯한 여러 신호 처리 문제에서 [12] 통계량 사이의 모든 상관 특성을 분석할 수 있다면, 이를 바탕으로 통계량을 좀더 알맞게 보태거나 바꾸어 쓰는 체계적인 방법을 찾을 수 있다. 이 논문에서는 독립이고 분포가 같은 관측벡터에서 통계량 사이의 결합 분포를 유도할 때 쓸모 있는 몇 가지 새로운 결과들을 생각하고, 널리 쓰이는 네 통계량 사이의 상관계수를 유도한다.

II. 확률 벡터와 통계량

독립이고 분포가 같은 확률신호 n 개가 이루는 확률벡터 $X = (X_1, X_2, \dots, X_n)$ 을 생각하자. 여기서, 확률신호 $\{X_j\}_{j=1}^n$ 는 모두 평균이 $E\{X\} = m_x$ 이고 분산이 $V\{X\} = E\{(X - m_x)^2\} = \sigma_x^2$ 이며, 확률밀도 함수와 누적분포함수는 각각 $f_x = f$ 와 $F_x = F$ 이다.

이제, $Z_i = \text{sgn}(X_i)$ 는 X_i 의 부호, R_i 는 관측벡터 X 에서 X_i 의 순위 (곧, X_i 보다 작거나 같은 X_j 의 갯수), 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 밀도함수는 아래와 같다 [11].

$$f_{U_r}(x) = n_{n-1} C_{r-1} F^{r-1}(x) \{1 - F(x)\}^{n-r} f(x). \quad (2)$$

여기서, $n C_r = n! / \{(n-r)! r!\}$ 은 이항계수이다. 다음에,

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

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

라 하면 $|X_i|$ 의 누적분포함수는 $F_{|X_i|}(y) = G(y)u(y)$ 이므로 $|X_i|$ 의 확률밀도함수는 $f_{|X_i|}(y) = g(y)u(y)$ 이다. 여기서, $u(y)$ 는 $y \geq 0$ 일 때 $u(y) = 1$ 이고 $y < 0$ 일 때 $u(y) = 0$ 인 단위 계단 함수이다. 따라서, (2)를 쓰면 $|X|$ 의 q 째 순서 통계량 $V_q = |X|_{[q]}$ 의 확률밀도함수는 아래와 같다.

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

정의: 독립인 확률신호 집합에서 한 확률신호의 확률밀도함수가 h 이고 나머지 확률신호의 확률밀도 함수가 모두 f 일 때, 이 확률신호들을 독립이고 분포가 거의 같은(*independent and semi-identically distributed: i.s.i.d.*) 확률신호들이라 한다.

곧, $h = f$ 이면, 독립이고 분포가 거의 같은 확률신호들은 독립이고 분포가 같은 (*independent and identically distributed: i.i.d.*) 확률신호들이다. 이 논문에서는, 독립이고 분포가 거의 같은 확률신호들 집합에서 확률밀도함수가 h 인 확률신호를 첫째 확률신호 X_1 로 두겠다.

정리 1: 독립이고 분포가 거의 같은 확률신호 집합 $\{X_j\}_{j=1}^n$ 을 생각하자. 그러면, $r = 1, 2, \dots, n$ 일

때 $\beta_i(r) = \Pr\{R_i = r\}$ 은 아래와 같다.

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

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

증명: 먼저, 아래와 같이 쓸 수 있다.

$$\Pr\{R_1 = r\}$$

$$\begin{aligned} &= \Pr\{X_1 \text{은 } \{X_j\}_{j=1}^n \text{에서 } r\text{째 순서 통계량이다}\} \\ &= \Pr\{\{X_j\}_{j=1}^n \text{ 가운데에서 } (r-1)\text{개는 } X_1 \text{보다 작거나 같고, 나머지 } (n-r)\text{개는 } X_1 \text{보다 크다}\} \\ &= \int_{-\infty}^{\infty} \Pr\{\{X_j\}_{j=1}^n \text{ 가운데에서 } (r-1)\text{개는 } x \text{보다 작거나 같고, 나머지 } (n-r)\text{개는 } x \text{보다 크다}\} \\ &\quad h(x) dx \\ &= {}_{n-1}C_{r-1} \int_{-\infty}^{\infty} F^{r-1}(x) \{1-F(x)\}^{n-r} \cdot h(x) dx. \end{aligned} \quad (7)$$

다음에, 사건 $\{R_1 = r\}, \{R_2 = r\}, \dots, \{R_n = r\}$ 은 서로 배반이고 $\sum_{i=1}^n \Pr\{R_i = r\} = 1$ 을 만족시킨다. 더욱이, $\{X_2, X_3, \dots, X_n\}$ 은 독립이고 분포가 같으므로 $r=1, 2, \dots, n$ 일 때 $\Pr\{R_2 = r\} = \Pr\{R_3 = r\} = \dots = \Pr\{R_n = r\}$ 이다. 따라서, $1 = \sum_{i=1}^n \Pr\{R_i = r\} = \beta_1(r) + (n-1)\beta_i(r)$ 이다.

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

정리 2: 독립이고 분포가 거의 같은 n 확률신호에서, $\tilde{\beta}_i(q) = \Pr\{Q_i = q\}$ 라 하자. 그러면, $q=1, 2, \dots, n$ 일 때 아래를 얻을 수 있다.

$$\tilde{\beta}_1(q) = {}_{n-1}C_{q-1} \int_{-\infty}^{\infty} G^{q-1}(x) \{1-G(x)\}^{n-q} \cdot \{h(x) + h(-x)\}u(x) dx, \quad (8a)$$

$$\tilde{\beta}_i(q) = \frac{1-\tilde{\beta}_1(q)}{n-1}, \quad i=2,3,\dots,n. \quad (8b)$$

증명: 확률변수 X_1 의 크기인 $|X_1|$ 의 확률밀도함수가 $\{h(x) + h(-x)\}u(x)$ 이고 $|X_j|, j=2, 3, \dots, n$ 의 누적분포함수가 $G(x)u(x)$ 이므로, (6a)와 (6b)에서 (8a)와 (8b)를 얻을 수 있다.

III. 상관계수

독립이고 분포가 같은 확률벡터에서 상관계수 ρ 의 표준화 (scaled) 상관계수를 아래와 같이 정의하자.

$$\tilde{\rho} = \sqrt{\frac{n+1}{n-1}} \rho. \quad (9)$$

보기를 들어, 독립이고 분포가 같으며 크기가 n 인 확률벡터 X 에서 $\tilde{\rho}_{X,R_i} = \sqrt{\frac{n+1}{n-1}} \rho_{X,R_i}$ 는 확률신호 X_i 와 그 순위 R_i 사이의 표준화 상관계수를 뜻한다.

먼저, 부록에서 보인 (36)과 (39)를 써서 공분산 $\text{Cov}(X_i, R_i)$ 와 $\text{Cov}(X_i, Q_i)$ 를 얻은 뒤, 이를 바탕으로 상관계수를 얻으면 아래와 같다.

$$\tilde{\rho}_{X,R_i} = \frac{\sqrt{3}}{\sigma_X} \int_{-\infty}^{\infty} x(2F(x)-1)f(x) dx, \quad (10)$$

$$\tilde{\rho}_{X,Q_i} = \frac{\sqrt{3}}{\sigma_X} \int_{-\infty}^{\infty} x(2G(|x|)-1)f(x) dx. \quad (11)$$

여기서, (10)은 다른 곳에서도 (보기: [7]) 찾아볼 수 있다. 한편, f 가 짝 함수일 때는 $\tilde{\rho}_{X,Q_i} = 0$ 이다.

정리 3: 표준화 상관계수 $\tilde{\rho}_{X,R_i}$ 와 $\tilde{\rho}_{X,Q_i}$ 는 아래를 만족시킨다.

$$0 \leq \tilde{\rho}_{X,R_i} \leq 1, \quad (12)$$

$$\tilde{\rho}_{X,R_i} \geq \tilde{\rho}_{X,Q_i}. \quad (13)$$

증명: 먼저, X_i 의 중앙값을 a 라 하자. 곧, $F(a) = \frac{1}{2}$ 이다. 그러면, $\int_{-\infty}^{\infty} (2F(t+a)-1) \cdot f(t+a) dt = \int_0^1 (2u-1) du = 0$ 이다. 이제, (10)에 있는 적분을 아래와 같이 다시 쓸 수 있다.

$$\begin{aligned}
 & \int_{-\infty}^{\infty} x\{2F(x)-1\}f(x) dx \\
 &= \int_{-\infty}^{\infty} (t+a)\{2F(t+a)-1\}f(t+a)dt \\
 &= a \int_{-\infty}^{\infty} \{2F(t+a)-1\}f(t+a)dt \\
 & \quad + \int_{-\infty}^0 t\{2F(t+a)-1\}f(t+a)dt \quad (14) \\
 & \quad + \int_0^{\infty} t\{2F(t+a)-1\}f(t+a)dt \\
 &= \int_{-\infty}^0 t\{2F(t+a)-1\}f(t+a)dt \\
 & \quad + \int_0^{\infty} t\{2F(t+a)-1\}f(t+a)dt.
 \end{aligned}$$

여기서, $t \leq 0$ 이고 $f(t+a) \geq 0$ 일 때 $2F(t+a) - 1 \leq 0$ 이므로, (14)의 마지막 줄에서 첫째 항은 0보다 크거나 같다. 비슷한 방법으로, $t \geq 0$ 이고 $f(t+a) \geq 0$ 일 때 $2F(t+a) - 1 \geq 0$ 이므로, (14)의 마지막 줄에서 둘째 항도 0보다 크거나 같다. 그러므로, $\tilde{\rho}_{|X_i, R_i} \geq 0$ 이다. 한편, (13)은 아래와 같이 보일 수 있다.

$$\begin{aligned}
 & \tilde{\rho}_{|X_i, R_i} - \tilde{\rho}_{|X_i, Q_i} \\
 &= \frac{2\sqrt{3}}{\sigma_X} \int_{-\infty}^{\infty} x f(x) \{F(x) - G(|x|)\} dx \\
 &= \frac{2\sqrt{3}}{\sigma_X} \int_{-\infty}^0 x f(x) \{2F(x) - F(-x)\} dx \\
 & \quad + \int_0^{\infty} x f(x) F(-x) dx \quad (15) \\
 &= \frac{2\sqrt{3}}{\sigma_X} \int_0^{\infty} x \{f(-x)F(x) + f(x)F(-x) \\
 & \quad - 2f(-x)F(-x)\} dx \\
 & \geq \frac{2\sqrt{3}}{\sigma_X} \int_0^{\infty} x \{f(-x)F(x) - f(x)F(-x) \\
 & \quad - 2f(-x)F(-x)\} dx \\
 & \geq 0.
 \end{aligned}$$

여기서, $x \geq 0$ 일 때 $F(x) - F(-x) \geq 0$ 이고 $\lim_{x \rightarrow \infty} x F(-x) = 0$ 임을 썼다.

다음에, 부록에서 보인 (42)와 (44)를 써서 공분산 $\text{Cov}(|X_i|, R_i)$ 와 $\text{Cov}(|X_i|, Q_i)$ 를 얻은 다음, 이를 바탕으로 상관계수를 얻으면 아래와 같다.

$$\begin{aligned}
 \tilde{\rho}_{|X_i, R_i} &= \frac{\sqrt{3}}{\sqrt{\sigma_X^2 + 4m_X^+ m_X^-}} \\
 & \cdot \int_{-\infty}^{\infty} |x| \{2F(x) - 1\} f(x) dx, \quad (16)
 \end{aligned}$$

$$\begin{aligned}
 \tilde{\rho}_{|X_i, Q_i} &= \frac{\sqrt{3}}{\sqrt{\sigma_X^2 + 4m_X^+ m_X^-}} \\
 & \cdot \int_{-\infty}^{\infty} |y| \{2G(|y|) - 1\} f(y) dy. \quad (17)
 \end{aligned}$$

여기서,

$$m_X^{\pm} = \int_{-\infty}^{\infty} x f(x) u(\pm x) dx \quad (18)$$

는 반 평균인데 (half mean), $m_X^- \leq 0$ 이고 $m_X^+ \geq 0$ 이므로 $\sigma_X^2 + 4m_X^+ m_X^- \leq \sigma_X^2$ 이라는 것을 새겨들 만하다. 확률밀도함수 $f(x)$ 가 x 의 짝 함수일 때는 $2F(x) - 1$ 이 홀 함수이므로 $\tilde{\rho}_{|X_i, R_i} = 0$ 이다.

정리 4: 표준화 상관계수 $\tilde{\rho}_{|X_i, R_i}$ 와 $\tilde{\rho}_{|X_i, Q_i}$ 는 다음을 만족시킨다.

$$0 \leq \tilde{\rho}_{|X_i, Q_i} \leq 1, \quad (19)$$

$$\tilde{\rho}_{|X_i, Q_i} \geq \tilde{\rho}_{|X_i, R_i}. \quad (20)$$

증명: 정리 3의 증명에서, 단조증가 함수 H 가 $0 \leq H(x) \leq 1$, $H(a) = 0$, $H(b) = 1$ 을 만족시키고 H 의 도함수를 h 라 할 때, $\int_a^b x \{2H(x) - 1\} \cdot h(x) dx = \int_0^1 (2u-1) H^{-1}(u) du \geq 0$ 임을 알 수 있다. 곧, $a=0$ 이고 $b=\infty$ 라 두면 $\int_{-\infty}^{\infty} |y| \{2G(|y|) - 1\} f(y) dy = \int_0^{\infty} y \{2G(y) - 1\} f(y) dy + \int_{-\infty}^0 y \{1 - 2G(-y)\} f(y) dy = \int_0^1 (2u-1) G^{-1}(u) du \geq 0$ 을 얻는다. 한편,

$$\begin{aligned}
 \tilde{\rho}_{|X_i, Q_i} - \tilde{\rho}_{|X_i, R_i} &= \frac{2\sqrt{3}}{\sqrt{\sigma_X^2 + 4m_X^+ m_X^-}} \\
 & \cdot \int_{-\infty}^{\infty} |x| f(x) \{G(|x|) - F(x)\} dx
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{2\sqrt{3}}{\sqrt{\sigma_X^2 + 4m_X^+ m_X^-}} \int_0^\infty x \{f(-x)F(x) \\
 &\quad - f(x)F(-x) - 2f(x)F(-x)\} dx \\
 &= \frac{2\sqrt{3}}{\sqrt{\sigma_X^2 + 4m_X^+ m_X^-}} \\
 &\quad \cdot \int_0^\infty F(-x) \{F(x) - F(-x)\} dx \quad (21)
 \end{aligned}$$

이므로 (20)을 얻는다.

한편, $F(0)=1$ 이면, $x \geq 0$ 일 때 $m_X^+ = 0$ 이고 $f(x) = 0$ 이며, $x \leq 0$ 일 때 $F(-x) = 1$ 이므로 $G(|x|) = 1 - F(x)$ 라는 것을 새겨두면, (10), (11), (16), (17)에서 $\tilde{\rho}_{X,R_i} = -\tilde{\rho}_{X,Q_i} = -\tilde{\rho}_{|X|,R_i} = \tilde{\rho}_{|X|,Q_i}$ 임을 알 수 있다. 이와 비슷하게, $F(0)=0$ 이면, $x \leq 0$ 일 때 $m_X^- = 0$ 이고 $f(x) = 0$ 인 것과 $x \geq 0$ 일 때 $F(-x) = 0$ 이므로 $G(|x|) = F(x)$ 이라는 것을 쓰면, (10), (11), (16), (17)에서 $\tilde{\rho}_{X,R_i} = \tilde{\rho}_{X,Q_i} = \tilde{\rho}_{|X|,R_i} = \tilde{\rho}_{|X|,Q_i}$ 임을 알 수 있다.

IV. 상관계수 보기

4.1 고른 분포

고른 확률밀도함수 $f(x) = \frac{1}{b-a} \{u(x-a) - u(x-b)\}$, $b > a$ 를 생각하자. 먼저, $F(x) = \frac{x-a}{b-a} \{u(x-a) - u(x-b)\}$, $\sigma_X = \frac{b-a}{\sqrt{12}}$, 그리고 $\int_{-\infty}^\infty x \{2F(x) - 1\} f(x) dx = \frac{b-a}{6}$ 이므로, (10)에서 $\tilde{\rho}_{X,R_i} = 1$ 을 얻는다. 한편, $b > 0$ 일 때 $t = a/b$ 라 하면 (11)에서 다음을 얻는다.

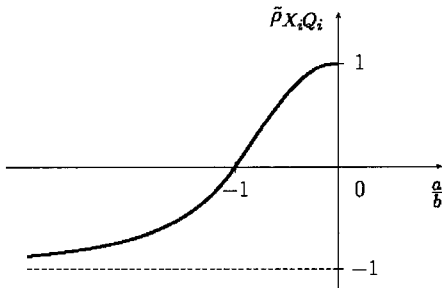


그림 1. 고른 분포 $U(a, b)$, $a < 0 < b$ 에서의 표준화 상관계수 $\tilde{\rho}_{X,Q_i}$, (22)

$$\tilde{\rho}_{X,Q_i} = \begin{cases} -1, & a < b \leq 0, \\ -1 + \left(\frac{2}{1-t}\right)^3, & t \leq -1, \\ 1 + \left(\frac{2t}{1-t}\right)^3, & -1 \leq t \leq 0, \\ 1, & 0 \leq a < b. \end{cases} \quad (22)$$

식 (22)는 $a < b \leq 0$ 일 때 -1 이고 $0 \leq a < b$ 일 때 1 이다. 그림 1에 $a < 0 < b$ 일 때 표준화 상관계수 (22)를 나타내었다.

한편, (16)과 (17)을 쓰면 아래를 얻을 수 있다.

$$\tilde{\rho}_{|X|,R_i} = \begin{cases} -1, & a < b \leq 0, \\ \frac{1+t}{1-t} \frac{1-4t+t^2}{\sqrt{(1-t)^4 - 12t^2}}, & t \leq 0, \\ 1, & 0 \leq a < b, \end{cases} \quad (23)$$

$$\tilde{\rho}_{|X|,Q_i} = \begin{cases} 1, & a < b \leq 0 \text{ 또는 } 0 \leq a < b, \\ \frac{1}{1-t} \frac{3+3t+3t^2-t^3}{\sqrt{(1-t)^4 - 12t^2}}, & t \leq -1, \\ \frac{1}{1-t} \frac{1-3t-3t^2-3t^3}{\sqrt{(1-t)^4 - 12t^2}}, & -1 \leq t \leq 0. \end{cases} \quad (24)$$

그림 2는 $a < 0 < b$ 일 때 상관계수 (23)을 나타낸 것이다. 한편, (24)의 최소값은 $\frac{a}{b} = \pm 1 - \sqrt{2} \approx -0.414$, -2.414 에서 $2^{\frac{7}{4}} - 2^{\frac{5}{4}} \approx 0.9852$ 이며, 최대값은 $\frac{a}{b} = 0, -1$, 또는 $-\infty$ 에서 1 이다. 곧, $\tilde{\rho}_{|X|,Q_i} \approx 1$ 이다.

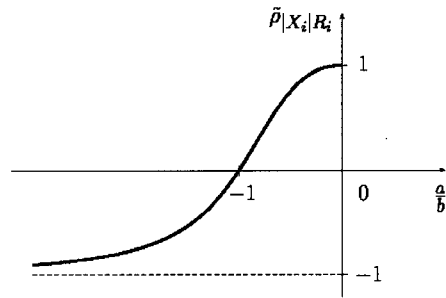


그림 2. 고른 분포 $U(a, b)$, $a < 0 < b$ 에서의 표준화 상관계수 $\tilde{\rho}_{|X|,R_i}$, (23)

4.2 그 밖의 분포

대칭 분포에서는 $\tilde{\rho}_{X,Q_i} = \tilde{\rho}_{|X|,R_i} = 0$ 이므로 $\tilde{\rho}_{X,R_i}$ 와 $\tilde{\rho}_{|X|,Q_i}$ 만 얻는다. 확률밀도함수가

$$f_{DE}(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \quad \lambda > 0, \quad (25)$$

인 겹지수 (라플라스) 분포,

$$f_L(x) = \frac{ke^{-kx}}{(1+e^{-kx})^2}, \quad k > 0, \quad (26)$$

인 로지스틱 분포,

$$f_T(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \quad (27)$$

인 t 분포와 정규 분포에서 $\tilde{\rho}_{X,R_i}$ 와 $\tilde{\rho}_{|X|,Q_i}$ 를 얻어 표 1과 2에 정리하였다. 여기서, $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$ 는 감마함수이고 (27)에 보인 매개변수 n 은 자유도라 (degree of freedom) 부른다. 표 2에서 볼 수 있듯이, t 분포에서 $\tilde{\rho}_{X,R_i}$ 와 $\tilde{\rho}_{|X|,Q_i}$ 는 각각

$\sqrt{\frac{3}{\pi}} \approx 0.9772$ 과 $\frac{\sqrt{3}(2-\sqrt{2})}{\sqrt{\pi-2}} \approx 0.9496$ 까지 커진다. 분포가 양의 방향으로 한 방향일 때는 $\tilde{\rho}_{X,R_i} = \tilde{\rho}_{X,Q_i} = \tilde{\rho}_{|X|,R_i} = \tilde{\rho}_{|X|,Q_i} = 0$ 이므로 $\tilde{\rho}_{X,R_i}$ 만 알아보면 된다. 먼저, 지수 확률밀도함수 $f_E(x) = \lambda e^{-\lambda x} u(x)$ 에서 $\tilde{\rho}_{X,R_i} = \frac{\sqrt{3}}{2}$ 이고, 레일리 확률밀도함수 $f_R(x) = \frac{x}{a^2} \exp\left\{-\frac{x^2}{2a^2}\right\} u(x)$ 에서 $\tilde{\rho}_{X,R_i}$

$= \frac{\sqrt{3\pi}(\sqrt{2}-1)}{\sqrt{8-2\pi}} \approx 0.9705$ 임을 쉽게 보일 수 있다. 다음에, α 와 β 가 0보다 큰 수일 때 감마 확률밀도함수

$$f_G(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left\{-\frac{x}{\beta}\right\} u(x) \quad (28)$$

에서의 결과를 표 3에 보였다.

표 3. 감마 확률밀도함수에서의 표준화 상관계수 $\tilde{\rho}_{X,R_i}$

α	0.1	0.3	0.7	1	2	4
$\tilde{\rho}_{X,R_i}$	0.4837	0.6942	0.8261	0.8660	0.9186	0.9472

V. 맺음말

먼저, 이 논문에서는 새로운 개념으로서 ‘독립이고 분포가 거의 같은’ 신호를 소개하였다. 이 개념은 여러 응용에서 꽤 쓸모 있는 개념인데, 이를 바탕으로 확률벡터에서 관측의 순위와 크기 순위의 확률질량함수를 얻었다. 아울러, 신호 검파 문제에서 많이 쓰이는 순위와 크기 순위를 비롯한 모수, 비모수 통계량 넷 사이의 상관계수를 얻은 뒤 그 값을 구체적으로 얻어보았다.

독립이고 분포가 같은 확률벡터에서 대표적인 통계량 넷, 곧, 관측, 관측의 크기, 순위, 그리고 크기 순위 사이의 상관계수를 닫힌 꼴로 이끌어내었다. 상관계수의 몇 가지 특징들을 얻었고 설명하였는데, 이는 직관적인 예상과 일치하는 결과였다. 더불어, 상관계수의 상한값과 하한값을 얻은 뒤, 잘 알려진 몇몇 분포에서 상관계수 값을 구체적으로 얻었다. 또한, 상한값과 하한값에 다다른 분포들을 보기로

표 1. 몇가지 대칭 확률밀도함수에서의 표준화 상관계수 $\tilde{\rho}_{X,R_i}$ 와 $\tilde{\rho}_{|X|,Q_i}$.

	정규 확률밀도함수	겹지수 확률밀도함수	로지스틱 확률밀도함수
$\tilde{\rho}_{X,R_i}$	$\sqrt{\frac{3}{\pi}} \approx 0.9772$	$\frac{3\sqrt{6}}{8} \approx 0.9186$	$\frac{3}{\pi} \approx 0.9549$
$\tilde{\rho}_{ X ,Q_i}$	$\frac{\sqrt{3}(2-\sqrt{2})}{\sqrt{\pi-2}} \approx 0.9496$	$\frac{\sqrt{3}}{2} \approx 0.8660$	$\frac{6(1-\ln 2)}{\sqrt{\pi^2-12(\ln 2)^2}} \approx 0.9088$

표 2. t 확률밀도함수에서의 표준화 상관계수 $\tilde{\rho}_{X,R_i}$ 와 $\tilde{\rho}_{|X|,Q_i}$.

	$n=3$	$n=5$	$n=7$	$n=9$	$n \rightarrow \infty$
$\tilde{\rho}_{X,R_i}$	0.8270	0.9284	0.9493	0.9578	$\sqrt{\frac{3}{\pi}} \approx 0.9772$
$\tilde{\rho}_{ X ,Q_i}$	0.7149	0.8608	0.8966	0.9121	$\frac{\sqrt{3}(2-\sqrt{2})}{\sqrt{\pi-2}} \approx 0.9496$

보였다. 이 논문에서 얻은 결과는 신호 처리에서 모수 또는 비모수 통계량을 쓸 때 그들의 성능을 예상할 수 있는 이론적인 잣대로 쓸 수 있을 것이다.

부록. 확률 함수와 결합 적률

여기서는, 확률신호 $\{X_i\}_{i=1}^n$ 을 연속 확률변수라고 두고, W 와 M 이 각각 연속 확률신호와 이산 확률신호일 때 $\frac{d}{dw} \Pr\{W \leq w, M = m\}$ 을 $\tilde{f}_{w,M}(w, m)$ 으로 쓴다. 이제, $(1+x)^{n-1} = \sum_{i=0}^{n-1} \binom{n-1}{i} x^i$ 와 $(n-1)(1+x)^{n-2} = \sum_{i=1}^{n-1} i \binom{n-1}{i} x^{i-1}$ 을 쓰면, $\sum_{k=1}^n \binom{n-1}{k-1} w^{k-1} (1-w)^{n-k} = 1$ 과 $\sum_{k=1}^n k \cdot \binom{n-1}{k-1} w^{k-1} (1-w)^{n-k} = 1 + (n-1)w$ 을 바로 얻는다. 따라서, 아래를 얻을 수 있다.

$$\frac{1}{n} \sum_{r=1}^n r f_{U_r}(x) = \{1 + (n-1)F(x)\}f(x), \quad (29)$$

$$\frac{1}{n} \sum_{q=1}^n q f_{V_q}(y) = \{1 + (n-1)G(y)\}g(y)u(y). \quad (30)$$

1. X_i 와 $|X_i|$ 사이의 확률 함수

X_i 와 $|X_i|$ 의 결합 누적분포함수 $F_{X_i, |X_i|}$ 는 아래와 같다.

$$\begin{aligned} F_{X_i, |X_i|}(x, y) &= \begin{cases} \Pr\{-y \leq X_i \leq x\}, & -y \leq x \leq y, y \geq 0, \\ \Pr\{-y \leq X_i \leq y\}, & x \geq y, y \geq 0, \\ 0, & \text{그 밖에는,} \end{cases} \quad (31) \\ &= u(y+x)u(y-x)G_1(x, y) \\ &+ u(y)u(x-y)G_1(y, y). \end{aligned}$$

여기서, $G_1(x, y) = F(x) - F(-y)$ 이고, $u(y+x)u(y-x)u(y) = u(y-x)u(y+x)$ 를 써서 정리했다. 이제, $\delta(y-a)f(y) = \delta(y-a)f(a)$, $G_1(x, -x) = 0$, $\delta(x-y) = \delta(y-x)$, 그리고 $u(ax) = u(x)$, $a > 0$ 임을 새기면, (31)에서 X_i 와 $|X_i|$ 의 결합 확률밀도함수를 아래와 같이 얻을 수 있다.

$$\begin{aligned} f_{X_i, |X_i|}(x, y) &= \frac{\partial^2}{\partial x \partial y} F_{X_i, |X_i|}(x, y) \\ &= \{\delta(x+y)f(-y) + \delta(x-y)f(y)\}u(y). \quad (32) \end{aligned}$$

식 (32)를 바탕으로 $|X_i| = y$ 가 주어졌을 때 X_i 의 조건부 확률밀도함수를 얻으면 아래와 같다.

$$\begin{aligned} f_{X_i | |X_i|}(x|y) &= \frac{f_{X_i, |X_i|}(x, y)}{f_{|X_i|}(y)} \\ &= \frac{f(-y)u(y)\delta(x+y)}{f(y) + f(-y)} + \frac{f(y)u(y)\delta(x-y)}{f(y) + f(-y)} \quad (33) \\ &= \frac{f(x)u(-x)\delta(x+y)}{f(x) + f(-x)} + \frac{f(x)u(x)\delta(x-y)}{f(x) + f(-x)} \end{aligned}$$

말할 것도 없이, (33)은 $f_{|X_i|}(y) = \{f(y) + f(-y)\}u(y) > 0$ 일 때 뜻이 있다.

2. 확률 함수 \tilde{f}_{X_i, R_i} , \tilde{f}_{X_i, Q_i} , $\tilde{f}_{|X_i|, R_i}$, $\tilde{f}_{|X_i|, Q_i}$

먼저, (6)을 쓰면 $\Pr\{R_i = r | X_i = x\} = \Pr\{X_j\}_{j=1}^n$ 이 독립이고 분포가 거의 같으며 확률밀도함수 $\tilde{f}(t-x)$ 일 때

$$\begin{aligned} R_i = r & \\ &= \binom{n-1}{r-1} C_{r-1} \int_{-\infty}^{\infty} F^{r-1}(t) \{1 - F(t)\}^{n-r} \\ &\quad \cdot \delta(t-x) dt \quad (34) \\ &= \binom{n-1}{r-1} C_{r-1} F^{r-1}(x) \{1 - F(x)\}^{n-r} \end{aligned}$$

를 얻고, 이 (34)에서 아래를 이끌어낼 수 있다.

$$\begin{aligned} \tilde{f}_{X_i, R_i}(x, r) &= \Pr\{R_i = r | X_i = x\} f_{X_i}(x) \\ &= \frac{1}{n} f_{U_r}(x). \quad (35) \end{aligned}$$

이제, (29)와 (35)에서 아래와 같이 결합 적률을 얻는다.

$$\begin{aligned} E\{X_i R_i\} &= \frac{1}{n} \sum_{r=1}^n \int_{-\infty}^{\infty} r x f_{U_r}(x) dx \\ &= \int_{-\infty}^{\infty} x \{1 + (n-1)F(x)\}f(x) dx. \quad (36) \end{aligned}$$

다음에, (8)에서 아래 식을 얻는다.

$\Pr\{Q_i = q | X_i = x\} = \Pr\{\{X_j\}_{j=1}^n$ 이 독립이고 분포가 거의 같으며 확률밀도함수 $\tilde{f}(t-x)$ 일 때

$$\begin{aligned} Q_i = q &= \binom{n-1}{q-1} C_{q-1} \int_{-\infty}^{\infty} G^{q-1}(t) \{1 - G(t)\}^{n-q} \\ &\quad \cdot \{\delta(t-x) + \delta(-t-x)\}u(t) dt \end{aligned}$$

$$= {}_{n-1}C_{q-1} G^{q-1}(|x|) \{1 - G(|x|)\}^{n-q}. \quad (37)$$

이 식을 쓰면

$$\begin{aligned} \tilde{f}_{X_i, Q_i}(x, q) &= \Pr\{Q_i = q | X_i = x\} f_{X_i}(x) \\ &= \frac{1}{n} f_{V_i}(|x|) \frac{f(x)}{g(|x|)} \end{aligned} \quad (38)$$

인데, 이 식과 (30)에서 아래 결합 적분을 얻을 수 있다.

$$\begin{aligned} E\{X_i | Q_i\} &= \int_{-\infty}^{\infty} x \sum_{q=1}^n q \frac{1}{n} f_{V_i}(|x|) \frac{f(x)}{g(|x|)} dx \\ &= \int_{-\infty}^{\infty} x \{1 + (n-1)G(|x|)\} f(x) dx. \end{aligned} \quad (39)$$

다음에, (33)에서 보였듯이 $|X_i| = y$ 가 주어졌을 때 X_i 의 확률밀도함수는 $f_{X_i|X_i}(x|y)$

$$= \frac{f(x)u(-x)\delta(x+y)}{f(x)+f(-x)} + \frac{f(x)u(x)\delta(x-y)}{f(x)+f(-x)} \text{ 이다.}$$

따라서, (6)을 쓰면 다음을 얻는다.

$$\begin{aligned} \Pr\{R_i = r | |X_i| = y\} \\ = \Pr\{\{X_j\}_{j=1}^n \text{이 독립이고 분포가 거의 같으며 확률밀도함수 } f_{X_i|X_i}(t|y), f(t) \text{ 일 때} \end{aligned}$$

$$\begin{aligned} R_1 = r \\ = {}_{n-1}C_{r-1} \int_{-\infty}^{\infty} F^{r-1}(t) \{1 - F(t)\}^{n-r} \\ \cdot \left\{ \frac{f(t)u(-t)\delta(t+y)}{f(t)+f(-t)} + \frac{f(t)u(t)\delta(t-y)}{f(t)+f(-t)} \right\} dt \\ = \frac{\{f_{U_i}(y) + f_{U_i}(-y)\} u(y)}{ng(y)}. \end{aligned} \quad (40)$$

그러므로, $y \geq 0$ 일 때

$$\tilde{f}_{|X_i|, R_i}(y, r) = \frac{1}{n} \{f_{U_i}(y) + f_{U_i}(-y)\}. \quad (41)$$

식 (29)와 (41)을 쓰면 $|X_i|$ 와 R_i 사이의 결합 적분을 아래와 같이 얻을 수 있다.

$$\begin{aligned} E\{|X_i| R_i\} \\ = \sum_{r=1}^n \int_{-\infty}^{\infty} r y \frac{1}{n} \{f_{U_i}(y) + f_{U_i}(-y)\} u(y) dy \\ = \int_0^{\infty} y \{1 + (n-1)F(y)\} f(y) \\ + \{1 + (n-1)F(-y)\} f(-y) dy \end{aligned}$$

$$= \int_{-\infty}^{\infty} |y| \{1 + (n-1)F(y)\} f(y) dy. \quad (42)$$

끝으로, (8)에서 아래를 쉽게 알 수 있다.

$\Pr\{Q_i = q | |X_i| = y\} = \Pr\{\{X_j\}_{j=1}^n \text{이 독립이고 분포가 거의 같으며 확률밀도함수 } f_{X_i|X_i}(t|y), f(t) \text{ 일 때}$

$$\begin{aligned} Q_1 = q \\ = {}_{n-1}C_{q-1} \int_{-\infty}^{\infty} G^{q-1}(t) \{1 - G(t)\}^{n-q} \\ \cdot \left\{ \frac{f(t)u(-t)\delta(t+y)}{f(t)+f(-t)} + \frac{f(t)u(t)\delta(t-y)}{f(t)+f(-t)} \right. \\ + \frac{f(-t)u(t)\delta(-t+y)}{f(t)+f(-t)} \\ \left. + \frac{f(-t)u(-t)\delta(-t-y)}{f(t)+f(-t)} \right\} u(t) dt \\ = {}_{n-1}C_{q-1} G^{q-1}(y) \{1 - G(y)\}^{n-q} u(y). \end{aligned} \quad (43)$$

여기서, $y \neq 0$ 일 때 $u(y)u(-y) = 0$ 이고 $y = 0$ 일 때 $G(-y) = 0$ 이므로 $G^{q-1}(-y) \cdot \{1 - G(-y)\}^{n-q} \frac{f(\pm y)}{f(y) + f(-y)} u(y)u(-y) = 0$ 임을 썼다. 따라서,

$\tilde{f}_{|X_i|, Q_i}(y, q) = \frac{1}{n} f_{V_i}(y)$ 이고, 이 식과 (30)에서 결합 적분을 얻으면 아래와 같다.

$$\begin{aligned} E\{|X_i| Q_i\} &= \frac{1}{n} \sum_{q=1}^n q \int_0^{\infty} y f_{V_i}(y) dy \\ &= \int_{-\infty}^{\infty} |y| \{1 + (n-1)G(|y|)\} f(y) dy. \end{aligned} \quad (44)$$

참고 문헌

- [1] P. Willett and P. F. Swaszek, "On the performance degradation from one-bit quantized detection", *IEEE Trans. Inform. Theory*, vol. 41, pp. 1997-2003, Nov. 1995.
- [2] 박철휘, 조용수, "충격성 잡음 환경에서의 최우 검출기", *한국통신학회 논문지*, 제21권, 522-532쪽, 1996년 6월
- [3] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks", *IEEE Jour. Sel. Areas Comm.*, vol. 22, pp. 1007-1015, Aug. 2004.
- [4] 김홍길, 박소령, 윤석호, 이주식, 송익호, "중양값을 움직인 부호를 바탕으로 한 비모수 검

파기: 점근 성능과 유한 표본 성능”, 한국통신학회 논문지, 제23권, 1454-1460쪽, 1998년 6월

- [5] V. Annampedu, V. V. Roganov, and R. Viswanathan, “Two rank order tests for M-ary detection,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 585-694, Mar. 2000.
- [6] I. Song and S. A. Kassam, “Locally optimum rank detection of correlated random signals in additive noise”, *IEEE Trans. Inform. Theory*, vol. 38, pp. 1311-1322, July 1992.
- [7] H. V. Poor and J. B. Thomas(Ed.), *Advances in Statistical Signal Processing: Vol. 2. Signal Detection*, JAI Press, Greenwich, 1993.
- [8] R. S. Blum, “Locally optimum distributed detection of correlated random signals based on ranks”, *IEEE Trans. Inform. Theory*, vol. 42, pp. 931-942, May 1996.
- [9] 한동석, “비균일 환경에서 표적 검파를 위한 순서통계에 근거한 일정오경보율 검파기의 성능 해석”, 한국통신학회 논문지, 제22권, 1550-1558쪽, 1997년 7월
- [10] H. G. Kim, I. Song, S. Yoon, and S. R. Park, “Nonparametric PN code acquisition using the signed-rank statistic for DS/CDMA systems in frequency-selective Rician fading channels”, *IEEE Trans. Vehic. Techn.*, vol. 51, pp. 1138-1144, Sep. 2002.
- [11] I. Song, J. Bae, and S. Y. Kim, *Advanced Theory of Signal Detection*, Springer-Verlag, Heidelberg, 2002.
- [12] D. Middleton, “Non-Gaussian noise models in signal processing for telecommunications: New methods and results for Class A and Class B noise models”, *IEEE Trans. Inform. Theory*, vol. 45, pp. 1129-1149, May 1999.

박 소 령 (So Ryoung Park)

정회원



1997년 2월 연세대학교 전자공학과(공학사)

1999년 2월 한국과학기술원 전기및전자공학과(공학석사)

2002년 2월 한국과학기술원 전자전산학과(공학박사)

2003년 3월~현재 가톨릭대학교 정보통신전자공학부 전임강사

<관심분야> 이동통신, 다중접속, 통계학적 신호처리

권 형 문 (Hyoungmoon Kwon)

준회원



2000년 2월 연세대학교 기계전자공학부 전기전자전공 졸업

2002년 2월 한국과학기술원 전자전산학과(공학석사)

2002년 3월~현재 한국과학기술원 전자전산학과 박사과정

<관심분야> 이동통신, 통계학적 신호처리, 검파와 추정

배 진 수 (Jinsoo Bae)

종신회원



1990년 2월 경기과학고등학교 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 재원연구원

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

최 상 원 (Sang Won Choi)

준회원



2002년 2월 고려대학교 전기전
자전파공학부(공학사)
2004년 2월 한국과학기술원 전
자전산학과(공학석사)
2004년 3월~현재 한국과학기술
원 전자전산학과 박사과정
<관심분야> 이동통신, 통계학적
신호처리, 검파와 추정

이 주 미 (Jumi Lee)

준회원



1998년 2월 이화여자대학교 수
학과(이학사), 전자공학과(공
학사)
2000년 2월 한국과학기술원 전
자전산학과(공학석사)
2000년 3월~현재 한국과학기술
원 전자전산학과 박사과정
<관심분야> 이동통신, 정보이론

송 익 호 (Ickho Song)

종신회원

한국통신학회 논문지 제30권 5호 참조