

두 표본 순위 검파에서 기준 표본 크기가 검파기 성능에 미치는 영향

배진수*

Effects of the Reference Sample Size on the Performance of the Two-Sample Rank Detector

Jinsoo Bae*

요약

이 논문에서는 두 표본을 쓰는 순위 검파기의 성능과 기준 표본 크기의 상관 관계를 모의실험을 통해 밝힌다. 기준 표본의 크기가 클수록 검파기의 성능은 더 우수하다. 다만 기준 표본 크기가 커질수록 성능 향상은 포화된다.

Key Words : Singal Detection, Rank Detection, Two-Sample Rank, Reference Sample, Nonparametric

ABSTRACT

The effects of the reference sample size on the detection probability of the two-sample rank detector is investigated in this paper. The larger reference sample size shows the better performance of the detector. The effect is also shown to be saturated as the reference sample size becomes larger.

I. 머리말

신호를 검파하기 위해 다양한 기법들이 사용되고 있다¹⁾. 그 가운데에서 관측 표본의 순위통계량을 쓰는 검파기는 비모수 특성을 가지면서도 성능이 비교적 우수하여 널리 쓰인다^{2,3)}. 부호통계량과 관측량의 절대값의 순위를 쓰는 순위 검파기가 한 표본 순위 검

파기이다²⁾. 부호통계량과 절대값 순위통계량을 쓰는 대신, 잡음만으로 이루어진 기준 표본을 덧붙여 쓰는 순위 검파기가 두 표본 순위 검파기이다³⁾.

이 논문에서는 두 표본 순위 검파기에서 기준 표본의 크기가 성능에 미치는 영향에 대해서 살펴본다. 기준 표본의 크기가 검파 확률에 미치는 영향을 정량적으로 보이고, 같은 크기의 관측 표본을 쓰는 한 표본 순위 검파기의 성능과도 비교해 본다.

기존 연구에서는 어느 정도의 크기를 가진 기준 표본을 사용해야 하는지에 대해서는 다루어지지 않았다³⁾.

II. 검정통계량

알려진 신호 검파를 위한 관측모형을

$$X_i = \theta e_i + W_i \tag{1}$$

이라고 할 때 (여기서 $i = 1, \dots, n$ 이다.) 두 표본 순위 검파기의 검정통계량은

$$T^t = \sum_{i=1}^n e_i R_i \tag{2}$$

이다. 여기서 e_i 는 알려진 신호 성분, W_i 는 잡음 성분, θ 는 신호 세기를 나타내는 매개변수이다. R_i 는 X_i 의 순위통계량으로, 식 (1)의 관측모형으로 얻어지는 크기가 n 인 표본에 추가로 m 개의 잡음 관측량을 덧붙여 만든 크기가 $n+m$ 인 새로운 표본 벡터

$$\vec{X}^t = (X_1^t, X_2^t, \dots, X_n^t, X_{n+1}^t, \dots, X_{n+m}^t) \tag{3}$$

에서 얻어진다. 따라서 R_i 는 X_i^t 의 순위통계량이고 1과 $n+m$ 사이의 정수 값을 갖게 된다. 가장 작은 관측량의 순위통계량은 1, 가장 큰 관측량의 순위통계량은 $n+m$ 이다. 이 관측 벡터의 앞 n 개의 관측량은 식 (1)의 관측모형으로부터 얻어지고, 그 뒤 m 개의 관측량은 순수한 잡음성분이고 식 (1)의 관측 모형에서 $\theta = 0$ 으로 두어 얻을 수 있다.

참고로 한 표본 순위 검파기의 검정 통계량은

* First Author : Department of Infomation and Communication Engineering, Sejong University, baej@sejong.ac.kr, 중신희원
 논문번호 : KICS2015-07-224, Received July 16, 2015; Revised July 22, 2015; Accepted July 22, 2015

$$T^o = \sum_{i=1}^n e_i Z_i Q_i \quad (4)$$

이고, 여기서 Z_i 는 관측벡터

$$\vec{X}^o = (X_1^o, X_2^o, \dots, X_n^o) \quad (5)$$

의 i 번째 관측량 X_i^o 의 부호통계량으로 X_i^o 의 값이 음수가 아니면 +1, 음수이면 -1의 값을 갖는다. Q_i 는

$$|\vec{X}^o| = (|X_1^o|, |X_2^o|, \dots, |X_n^o|) \quad (6)$$

에서 $|X_i^o|$ 의 순위통계량으로 1에서 n 사이의 정수 값을 갖는다.

한 표본 순위 검파기는 부호통계량과 관측량의 절대값을 쓰기 때문에 구조가 복잡하고, 두 표본 순위 검파기는 보다 큰 크기의 표본을 다루어야 하기 때문에 계산량이 많다.

두 표본 순위 검파기에 사용되는 식 (3)의 관측벡터 \vec{X}^t 의 뒤쪽 m 개의 관측량들로 이루어진 벡터를 기준 표본이라고 하고, m 은 기준 표본의 크기라 한다.

이 논문에서는 m 값에 따른 검파기의 성능을 모의 실험을 통해 알아본다. 모의실험에서 사용된 매개변수의 값은 $n = 10, m = 3, 5, 10, 20, 30, 40, 50$ 이다. 잡음 성분 W_i 는 정규분포를 갖는 백색 잡음이고, e_i 값은 일반성의 훼손없이 모두 1로 정한다. 모의실험에서 오경보확률은 0.005이고, 검파확률은 1000번의 실험을 반복하여 얻는다.

III. 모의 실험 결과

그림 1은 m 값이 각각 5, 10, 20, 30, 40, 50 일 때 두 표본 순위 검파기의 성능을 신호세기를 나타내는 매개변수 θ 에 따라 보여준다.

그래프가 왼쪽에 있을수록 보다 작은 신호 세기에서 보다 높은 검파확률을 가지므로 성능이 우수하다. 왼쪽의 그래프부터 m 값이 각각 50, 40, 30, 20, 10, 5 일 때의 검파확률을 나타낸다. m 값이 증가할수록 더 좋은 성능을 보여주지만, m 값이 매우 큰 경우에는 성능 향상 폭이 감소하는 포화현상이 있다.

그림 2는 $n = 10$ 인 한 표본 순위 검파기와

$(n, m) = (10, 20)$ 인 두 표본 순위 검파기의 성능을 비교한 것이다. 이 모의 실험에서 부호통계량과 절대값 순위통계량을 쓰는 (복잡한 구조를 갖는) 한 표본 순위 검파기와 비슷한 성능을 보이기 위해서 필요한 기준 표본 크기 m 은 (관측 표본 크기 $n = 10$ 의 2배인) 20이다. 순위통계량을 얻는 과정의 계산량은 표본 크기에 따라 지수적으로 증가한다. 따라서 간단한 구조를 갖는 두 표본 순위 검파기를 쓰기 위해서는 매우 많은 계산량을 더 필요로 한다.

그림 3은 m 값과 두 표본 순위 검파기의 성능의 양의 상관 관계가 m 값이 증가함에 따라 포화됨을 다른 방법으로 보여준다. 각각의 m 값을 가진 두 표본 순위 검파기의 검파확률이 0.999보다 크거나 같게되는 신호세기 매개변수의 θ 의 최소값을 나타낸다.

m 값이 증가함에 따라 검파확률 0.999을 얻기 위해 필요한 신호 세기가 감소하지만, 그 감소폭은 m 값이 증가함에 따라 감소하여 그 효과가 작아짐을 볼 수 있다.

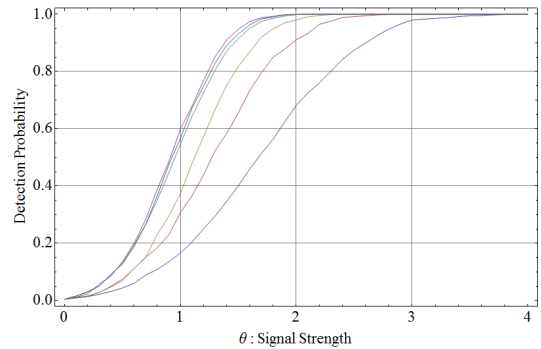


그림 1. Detection probabilities of the two-sample rank detectors with $m = 50, 40, 30, 20, 10, 5$.

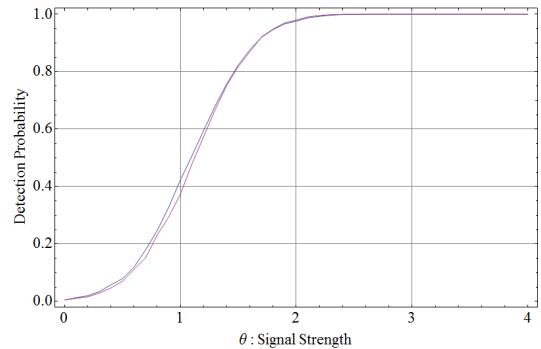


그림 2. Performance comparison of the one-sample rank detector with $n = 10$ and the two-sample rank detector with $(n, m) = (10, 20)$.

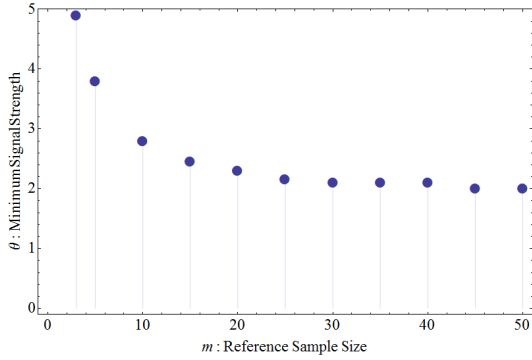


그림 3. The minimum values of θ with detection probability equal or larger than 0.999 for $m = 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$.

IV. 맺음말

두 표본 순위 검파기는 (한 표본 순위검파기와는 다르게) 부호통계량과 절대값 순위통계량을 쓰지 않는 대신 잡음만으로 이루어진 기준 표본을 얻어진 관측벡터에 덧붙여 사용한다.

두 표본 순위 검파기를 설계할 때 기준 표본의 크기를 적절하게 결정하는 일은 (기준 표본 크기가 커질 수록 계산량이 지수적으로 증가하기 때문에) 매우 중요하다.

이 논문의 결과로 기준 표본의 크기는 (같은 크기의 관측 표본을 쓰는 한 표본 순위 검파기와 같은 성능을 내기 위해서는) 생각보다 커야하고 동시에 기준 표본의 크기가 일정한 수준보다 커지면 포화현상이 있으므로 그 사이의 값에서 최적화가 이루어져야 한다. 이후 기준 표본 크기의 최적값을 얻기 위한 알고리즘을 구하는 연구에 관심이 필요하다.

References

- [1] H.-K. Min, T. An, S. Lee, S. R. Lee, and I. Song, "Power signal recognition with high order moment features for non-intrusive load monitoring," *J. KICS*, vol. 39C, no. 7, pp. 608-614, Jul. 2014.
- [2] I. Song and S. A. Kassam, "Locally optimum rank detection of correlated random signals in additive noise," *IEEE Trans. Inf. Theory*, vol. IT-38, no. 4, pp. 1311-1322, Jul. 1992.
- [3] S. Y. Kim and I. Song, "On the score functions of the two-sample locally optimum rank test statistic for random signals in additive noise," *IEEE Trans. Inf. Theory*, vol. IT-41, no. 3, pp. 842-846, May 1995.