

# 임펄스성 잡음의 유무를 결정하는 Kolmogorov-Smirnov 검증의 수치적 접근의 효율성

오형국\*, 남해운<sup>o</sup>

## Numerical Approach with Kolmogorov-Smirnov Test for Detection of Impulsive Noise

Hyungkook Oh\*, Haewoon Nam<sup>o</sup>

요약

본 논문에서 임펄스성 잡음의 유무를 검증하는 알고리즘을 제안한다. 본 알고리즘을 제안하는 이유는 기존의 Kolmogorov-Smirnov 검증의 단점으로 낮은 분류 성공률 및 높은 복잡도가 있기 때문이다. 이는 이론적으로 문제가 없으나 실제로 구현함에 있어 많은 문제를 야기한다. 먼저 기존의 검증 방법을 설명 후 제안하는 알고리즘을 설명한다. 이 알고리즘은 기존의 Kolmogorov-Smirnov 검증 방법의 이론적 배경으로부터 제안된다. 알고리즘의 효율성을 증명하기 위해 임펄스성 잡음의 샘플을 이용하여 실험 후, 검증 실패 확률을 조사한다. 검증 실패 확률에 기반한 실험 결과는 제안한 알고리즘의 효율성을 증명한다.

**Key Words** : Impulsive Noise, Kolmogorov-Smirnov Test, Chi-Square Test, Noise separation

ABSTRACT

This paper proposes an efficient algorithm based on Kolmogorov-Smirnov test to determine the presence of impulsive noise in the given environment. Kolmogorov-Smirnov and Chi-Square tests are known in the literature to serve as a goodness-of-fit test especially for a testing for normality of the distribution. But these algorithms are difficult to implement in practice due to high complexity. The proposed algorithm gives a significant reduction of the computational complexity while decreasing the error probability of hypothesis test, which is shown in the simulation results. Also, it is worth noting that the proposed algorithm is not dependent on the noise environment.

### I. 서론

현재에 이르러, 사람들의 편의를 위해 많은 전자기기들이 만들어 졌다. 이 기기들은 예상치 못한 전자기파를 생성하고 그 전자기파는 통신의 성능을 현저히

저하시킨다<sup>[1,2]</sup>. 사회에서 쓰이는 대부분의 수신기들은 가우시안 잡음에 최적화되었기 때문에, 그 전자기파들은 그 가정의 효율성을 떨어뜨린다. 이 성능 저하를 극복하기 위해 많은 연구자들이 잡음을 관찰한 결과, 어느 일정시간 갑작스럽게 임펄스가 발생하는 임펄스

\* 본 연구는 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. 2014-000000057210)입니다.

• First Author : Hanyang University Department of Electronics & Communication Engineering, ohhk@hanyang.ac.kr, 학생회원

◦ Corresponding Author : Hanyang University Department of Electronics & Communication Engineering, hnam@hanyang.ac.kr, 정회원  
논문번호 : KICS2014-05-192, Received May 24, 2014; Revised July 8, 2014; Accepted August 19, 2014

성 특징을 지닌다는 것을 발견했다<sup>3-5)</sup>. 이에 따라, 임펄스성 잡음을 잡음 환경으로 모델링했으며 특정장소에서는 이 모델이 유효하다. 지금까지 임펄스성 잡음 환경에서 성능 저하를 극복하기 위한 방법들이 많이 연구되고 여러 개의 임펄스성 잡음 모델마다 각각의 파라미터들을 추정하는 연구 결과도 있었다<sup>6,7,8)</sup>. 하지만 임펄스성 잡음 환경에 관한 연구는 거의 없었다. 이 연구가 중요한 이유는 최적화된 에러율을 보여주기 위해 잡음 환경에 따라 사용되는 수신기 모델이 달라지기 때문이다. 흔히 임펄스성 잡음의 경우, 비선형 블락을 사용하여 임펄스성 잡음의 영향을 줄여 급격하게 오류율을 줄이는 효과를 가져올 수 있다<sup>9)</sup>. 기존의 가우시안 잡음의 경우, 기존의 수신기를 사용하면 되기 때문에 비선형 블락의 사용은 의미가 없다. 그러므로 잡음 환경에 따라 수신기에 사용되는 전략이 달라지므로 통신 환경이 임펄스성 성격을 갖는지에 관한 연구가 필요하다. 또한 이 분야의 연구가 실제 사회에 적용이 가능할 정도로 낮은 복잡도 및 오류율을 낮추는 방향으로 진행되어야 한다. 많은 연구자들은 가우시안 잡음과 임펄스성 잡음은 그들이 갖는 확률 분포가 틀리다는 사실을 이용하여 기존의 적합도 테스트 (Goodness-of-Fit test)를 이용한다<sup>10,11)</sup>. 하지만 확률 밀도 함수를 이용한 적합도 테스트는 복잡도가 너무 커서 구현이 어렵다<sup>12)</sup>. 그러므로 본 논문에서 확률 밀도 함수를 통한 접근이 아닌 다른 방법으로 접근하여 임펄스성 잡음 환경의 유무를 판단하는 방법을 제안한다. 이 방법은 기존의 Kolmogorov-Smirnov 테스트 방법의 단점인 높은 복잡도와 분류 에러율 (error rate)을 낮춘다.

본 논문은 다음과 같은 구성으로 전개된다. 2장은 미들턴 class A 잡음을 정의하고 특성을 설명한다. 3장인 테스트 방법은 3절로 나뉘며, 1절은 임펄스성 잡음 모델 중 널리 쓰이는 미들턴 class A 잡음 모델을 소개한다. 2절은 기존의 적합도 테스트를 소개하는데, 여러개의 테스트 방법 중 많이 알려진 Chi-square 테스트와 Kolmogorov-Smirnov 테스트를 소개한다. 3절에서 본 논문에서 Kolmogorov-Smirnov 테스트를 이용하여 새로운 알고리즘을 제안한다. 4장에서 제안한 알고리즘을 이용한 실험결과를 보여준다. 이 장에서는 미들턴 class A 잡음 모델의 다양한 파라미터 조합을 사용하고 그에 따라 얻는 실험 결과를 통해, 제안된 테스트가 얼마나 효율적인지 보여준다. 마지막 장은 본 논문에서 설명한 내용들과 제안된 내용들을 정리한다.

## II. 시스템 모델

임펄스성 잡음 중에서 미들턴 class A 노이즈가 대표적으로 시스템에 사용된다<sup>11)</sup>. 주로 미들턴 class A 노이즈가 널리 쓰이는 이유는 간섭이 Poisson 분포에 따라 지역적으로 퍼진 현상을 내포하고 있다<sup>4)</sup>. 또한 2개의 가우시안 분포로 이루어진 잡음 형태가 대부분 이어서 여러 종류의 간섭을 표현하기 힘든 반면, 미들턴 class A 노이즈는 그렇지 않다. 미들턴 class A 노이즈의 진폭 식은 아래와 같이 표현된다.

$$f_X(x) = \sum_{m=0}^{\infty} \frac{e^{-A} A^m}{m!} \frac{x}{\sigma_m^2} e^{-\frac{x^2}{2\sigma_m^2}} \quad (1)$$

위의 식에서  $A$ 는 충격 지수 (impulsive index)이며  $\sigma_m^2 = \frac{m + \Gamma}{1 + \Gamma}$  이고  $\Gamma$ 는 가우시안 잡음과 임펄스성 잡음의 비로 정의된다. 식에서 볼 수 있듯이, Poisson 분포와 Rayleigh 분포의 곱 형태로 표현되었고, 곱해진 텀의 수는 무한개임을 알 수 있다. 무한개가 의미하는 것은 다양한 파워를 지닌 간섭들을 표현한 것이다. 이것은 위에서 언급한 2개의 가우시안 분포로 구성되어 기존의 additive white Gaussian noise (AWGN)과 오직 한 종류의 간섭만을 표현하는 모델과는 다르다는 것을 알 수 있다. 즉, 미들턴 class A 노이즈 모델이 임펄스성 잡음 모델 중에서 많이 쓰이는 이유를 말해 주고 있다.

또한 미들턴 class A 노이즈가 많은 임펄스성 잡음 중에서 대표적으로 쓰이는 또 다른 이유는 파라미터 때문이다. 이 노이즈는 파라미터로  $A$ 와  $\Gamma$ 가 있으며,  $A$ 는 일정 시간 구간동안 평균 노이즈 신호 방사 수와 신호의 평균 방사 길이를 곱한 형태를 의미한다. 예를 들어,  $A$ 는 수치가 커지면 커질수록 관찰하는 시간동안 들어오는 임펄스성 잡음이 많아짐을 의미한다.  $A$ 는 일반적으로  $0.01 \leq A < 1$ 로 범위가 결정된다. 물론 그 이외의 숫자들을 대입하여도 수식은 성립하지만  $0.01$ 보다 작으면 임펄스성 잡음이 관찰 시간동안 드물게 관찰된다.  $1$ 보다 큰 경우는 가우시안 잡음보다 임펄스성 잡음이 오히려 더 많이 관찰되는 효과가 발생한다. 이는 이 잡음이 임펄스성 특성을 잃게 됨을 의미한다.  $\Gamma$ 는 정의에 의해 수치가 작으면 작을수록 임펄스성 특성을 더욱 크게 지니게 된다.  $\Gamma$ 가  $A$ 와 다른 점은 임펄스성 잡음의 발생 확률과는 무관하다는

것이다. 효율적인  $I$ 의 범위는  $\Gamma < 1$ 이다. 정의를 생각 해볼 때 1보다 크다면 임펄스성 잡음의 파워보다 가 우시안 잡음의 파워가 더 크다는 의미를 가지게 되므로 무의미한 범위이다.

그림1과 2는 각각 파라미터 조합에 따른 Complementary Cumulative Density Function (CCDF) 또는 Exceeding Probability (EP) 와 파라미터에 따른 임펄스성 잡음 샘플이다. PDF가 아니라 EP 를 사용하는 이유는 PDF보다는 EP가 보다 더 임펄스 성 잡음 파라미터의 특징을 찾기 쉽기 때문이다. EP 그래프에 따르면,  $A$ 값이 작아질수록 EP그래프의 평 평한 부분이 점점 더 내려가는 것을 관찰 할 수 있다. 또한  $\Gamma$ 값이 커질수록 그래프가 오른쪽으로 이동하는

것을 볼 수 있다. 또한 샘플을 관찰하면 계속되는 Gaussian 잡음 가운데에 갑작스러운 임펄스 형태의 잡음이 나옴을 알 수 있다. 임펄스 잡음의 크기는  $A$ 와  $\Gamma$ 로 조절 할 수 있으나 대부분  $A$ 의 영향력이 크다.

### III. 기존의 테스트 방법

관찰된 샘플들이 원하는 랜덤 변수를 따르는지에 관한 연구는 이미 되었다<sup>[10-12]</sup>. 본 절에서 이에 관한 이론을 언급하고자 한다. 또한 기존의 방법들을 언급 하면서 실제로 구현할 때 어떠한 것이 복잡도를 높이는 지에 대해 언급한다.

#### 3.1 Chi-Square 테스트 방법

기존의 Chi-Square 테스트 방법은 아래의 가정으로부터 시작된다.

- $H_0$ : 관찰된 샘플들이 원하는 분포를 따르는 경우
- $H_1$ : 관찰된 샘플들이 원하는 분포를 따르지 않는 경우

위의 가정은 모든 테스트 방법에 적용할 수 있다. Chi-Square 방법은 흔히 연속분포 보다 이산분포에 많이 쓰인다. 만약 원하는 분포가 연속분포일 경우, 연속 분포에 구간을 두어 이산 분포화시키는 방법을 쓸 수 있다. 예를 들어, 본 논문에서 언급되는 미들턴 class A 노이즈 모델도 연속 분포들 중 한 가지이므로 이것을 각 구간을 두어 이산화 시킨다. 이산 분포화된 모델은 각각 관찰 가능한 값마다 관찰된 값들을 히스토그램화 시킨다. 즉, 이론상의 값과 실제 관찰된 샘플들로 만든 히스토그램을 비교하는 것이다. 여기서 이론상의 값은 0보다 크고 1보다 작으므로 반드시 각 값마다 샘플 수를 곱한다. 이를 이용하여 아래의 식을 사용한다.

$$X^2 = \sum_{i=1}^r \frac{(O_i - E_i)^2}{E_i} \tag{2}$$

위의 식에서  $O_i$ 는 관찰된  $i$ 번째 빈도수,  $E_i$ 는 원하는 분포의  $i$ 번째 빈도수,  $r$ 은 관찰되는 값의 전체수를 의미한다. 이 테스트 방법이 Chi-Square 테스트라고 이름이 붙혀진 이유는 위의 식이 Chi-Square 분포를 따르기 때문이다. 위의 식의 제곱 안에 있는 것은 샘플 수가 많으면 많을수록 중심 극한 이론(Central Limit Theorem)을 따른다. 즉, 위의 식은 자유도(degree of freedom)가  $r-1$  인 Chi-square 분포를 따르게 된다. 또

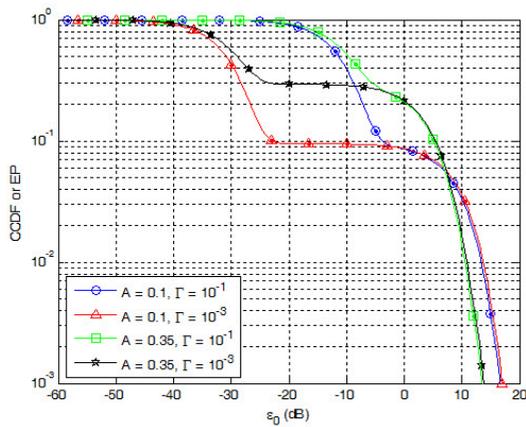


그림 1.  $\Gamma$ 와  $A$ 의 조합에 따른 CCDF 그래프  
Fig. 1. CCDF of impulsive noises with various  $\Gamma$  and  $A$  parameters

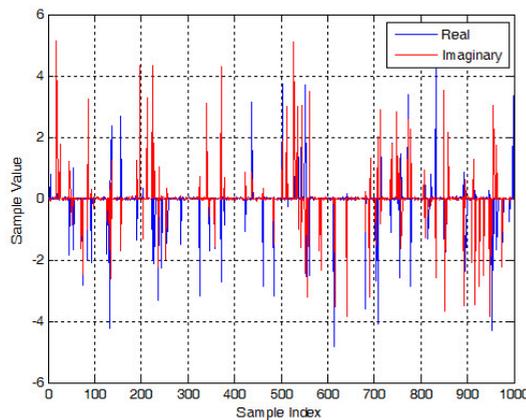


그림 2.  $\Gamma$ 와  $A$ 가 각각 0.001 과 0.1인 값을 지닐 때의 임펄스성 잡음 샘플  
Fig. 2. Impulsive noise samples with  $\Gamma = 0.001$  and  $A = 0.1$

한 값의 빈도수가 적을수록 요구되는 샘플 수도 역시 적어진다는 특징이 있다.

위의 식과 가정들을 판별하기 위해 유의수준 (Significance level,  $\alpha$ )을 사용하여 판별한다. 여기서 유의 할 점은  $r$ 에 따라 Chi-square 분포의 모양이 바뀐다는 점이다. 유의 수준과 비교를 위해 Chi-square 분포의 누적 밀도 함수(CDF)를 구하여야 한다. 하지만 특수한 경우를 제외하고 이는 식을 얻기 힘들다. 또한, 특정  $r$  값이 어느 경우에도 효율적이라는 결과도 없기 때문에 수치적인 접근 방법도 힘들다. 이에 따라 이론적으로 성립이 되지만 구현 관점에서 보면 큰 복잡도를 가진다는 단점을 관찰 할 수 있다.

### 3.2. 기존의 Kolmogorov-Smirnov 테스트 방법

Kolmogorov-Smirnov 테스트 방법은 누적 밀도 함수로부터 구분 방법을 얻으므로 아래와 같이 가정할 수 있다.

$$H_0 : F(x) = F_0(x)$$

$$H_1 : F(x) \neq F_0(x)$$

$F(x)$ 는 기준으로 삼는 분포의 누적 밀도 함수이고  $F_0(x)$ 는 샘플로 만들어진 분포의 누적 밀도 함수이다. Kolmogorov-Smirnov 테스트 방법은 원하는 분포의 누적 밀도 함수와 관측된 샘플의 누적 밀도 함수를 비교하여 누적 밀도 함수들 간의 거리의 차를 이용하여 얻는 방법이다. 수식적 알고리즘은 아래와 같다.

1.  $D = \max |F(x) - F_0(x)|$
2.  $\Pr(D) > 1 - \alpha$  이면  $H_0(x)$  결정
3.  $\Pr(D) \leq 1 - \alpha$  이면  $H_1(x)$  결정

여기서  $\Pr(D)$ 는 Kolmogorov-Smirnov 누적 밀도 함수로 식 (4)에서 표현된 분포를 갖는다.

$$\Pr(D) = \frac{\sqrt{2\pi}}{x} \sum_{k=1}^{\infty} e^{-\frac{(2k-1)^2 \pi^2}{8x^2}} \quad (4)$$

식 (3)에서 알고리즘의 D의 의미는 그림 3에 표현된다. 이론적인 누적 밀도 함수와 샘플을 이용한 누적 밀도 함수를 이용하여 보면 최대오차인 D가 작으면 작을수록 비교하고자 하는 원래의 분포와 동일하다는 것을 알 수 있다. 알고리즘을 설명하면, 먼저 원하는 함수와 샘플들을 이용하여 누적 밀도 함수들을 만들고 각 거리의 차를 구한다. 이 구한 차를  $\Pr(D)$ 에 넣

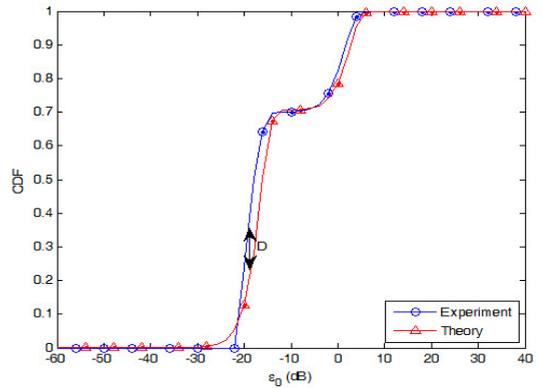


그림 3. 이론적인 누적 밀도 함수와 샘플을 이용한 누적 밀도 함수의 비교  
Fig. 3. Comparison of theoretical and empirical PDFs of impulsive noise

어서 이것이  $1-\alpha$ 을 기준으로 가정들을 기각할지 채택할지 결정한다. 이 테스트 방법이 Chi-Square 방법보다 좋은 점은 연속 분포를 이산 분포화 시킬 필요가 없다는 점이다. 두 번째는 Chi-Square 테스트에서 이산 분포화하면 생기는 문제점이 얼마나 관측 값의 수를 정하는 것인데 그 문제도 없어진다. 하지만  $\Pr(D)$ 을 사용한 판별 방식은 정밀하게 분포를 구분하지 못할 수 있다는 단점을 지닌다. 또한 식 (4)가 무한대의 항으로 표현되기 때문에, 컴퓨터로  $\Pr(D)$ 를 구현하려면 높은 복잡도가 요구된다.

### 3.3. 제안하는 Kolmogorov-Smirnov 테스트 방법

그림 4는 미들턴 class A 잡음과 Complex Gaussian 잡음의 진폭을 dB 스케일에 따라 보여준다. 가우시안 분포를 임펄스성 잡음의 분포와 비교할 때

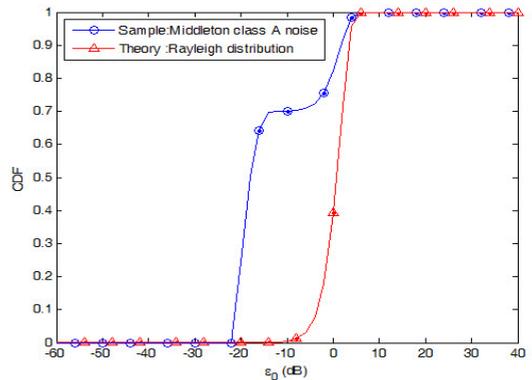


그림 4.  $\Gamma$ 와 A가 각각 0.0005 과 0.35인 값을 지닐 때의 Middleton class A 잡음과 콤플렉스 가우시안 잡음의 진폭의 누적 밀도 함수  
Fig. 4. Middleton class A noise CDF with  $\Gamma = 0.0005$  and  $A = 0.35$  and Rayleigh CDF

누적 밀도 함수에서 큰 오차를 보여준다. 그러므로 들어온 샘플들을 가우시안 누적 밀도 함수와 비교하여 잡음을 분류하는 방법을 제안한다. 특히 기존의 Kolmogorov-Smirnov 테스트 방법보다 낮은 오류율 및 복잡도를 위한 방법을 제안한다. 단, 본 논문의 목적인 임펄스성 잡음의 분류를 위해 다음과 같은 가정을 한다.

- $H_0$  : Impulsive noise
- $H_1$  : Complex Gaussian noise

만약  $H_0$ 가 맞다면 임펄스성 잡음으로 분류되고 그렇지 않다면 기존의 가우시안 잡음으로 분류가 된다. 위와 같은 가정을 한 이유는 가정의 목적이 임펄스성 잡음의 분류이고, 또한 임펄스성 잡음의 파라미터에 의해 분포가 정해지지 않기 때문이다. 그러므로 위와 같은 가정을 사용한다. 기본적인 알고리즘은 3.2절의 수식적 알고리즘과 크게 다르지 않다. 다만,  $\Pr(D)$  부분만 다르게 제안되었다. 기존의 Kolmogorov-Smirnov 방식에선 Kolmogorov-Smirnov 누적 밀도 함수가 사용되어 가정을 검증하지만 이 절에서는 다른 방법으로 보다 더 복잡도가 낮으며 더 성능이 좋은 알고리즘을 만들려고 한다.

$$D = \max |F(x) - F_0(x)| \rightarrow 0 \tag{5}$$

식 (5)는 샘플수가 많으면 많을수록 이론적으로 샘플에 의한 최대 오차거리가 0에 가까워짐을 보여준다. 또한 같은 분포를 지닌 샘플을 다수 얻기 때문에 중심

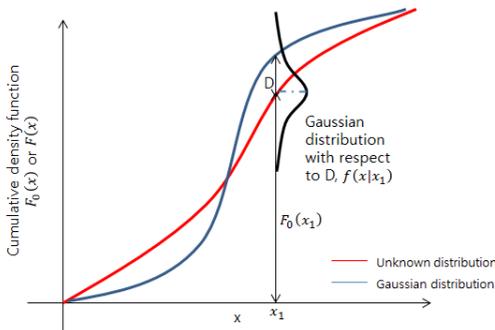


그림 5. 누적 밀도 함수가  $F(x)$ 인 가우시안 분포 (파란색)와  $F_0(x)$ 인 샘플에 의한 분포 (빨간색)의 최대오차에 대한 가우시안 분포 (검은색)  
 Fig. 5. Gaussian distribution of maximum difference between cumulative density function  $F(x)$  and distribution of samples  $F_0(x)$ .

극한 정리를 이용할 수 있다. 그림 5는 최대오차  $D$ 가 일어난  $x_1$ 에서의  $D$ 에 따른 가우시안 분포인  $f(x|x_1)$ 을 보여준다.  $f(x|x_1)$ 의 평균은  $F_0(x_1)$ 이고  $F_0(x_1)$ 은 최대오차 지점의 샘플들의 누적 밀도 확률이다. 1개의 샘플의 경우,  $x_1$ 보다 작을 확률은  $F_0(x_1)$ 이고  $x_1$ 보다 클 확률은  $1 - F_0(x_1)$ 이다. 이것은 Bernulli 분포의 성공확률과 실패확률로 모델링이 가능하므로 한 개의 샘플로 이루어진  $f(x|x_1)$ 의 분산은  $F_0(x_1)(1 - F_0(x_1))$ 이다. 그리고  $x_1$  지점에서 샘플이 전체  $n$ 개에서  $k$ 개가 사용된 경우에 분산은 아래의 식으로 유도할 수 있다.

$$\begin{aligned} \text{var}(f(x|x_1)) &= \text{var}\left(\frac{1}{k} \sum_{i=1}^k f(x|x_{1i})\right) = \frac{1}{k^2} \sum_{i=1}^k \text{var}(f(x|x_{1i})) \\ &= \frac{1}{k^2} k F_0(x_1)(1 - F_0(x_1)) = \frac{1}{k} F_0(x_1)(1 - F_0(x_1)) \end{aligned} \tag{6}$$

식 (6)에서  $k$ 개의 샘플에 의한 분포는  $f(x|x_1)$ 이다.  $f(x|x_{1i})$ 는 1번째 샘플의  $x_1$ 에서의 확률밀도함수이다. 언급했듯이  $f(x|x_{1i})$ 의 평균은  $F_0(x_1)$ 이고 분산은  $F_0(x_1)(1 - F_0(x_1))$ 이다. 그러므로 식 (6)에 따라  $f(x|x_1)$ 의 분산은  $F_0(x_1)(1 - F_0(x_1))/k$ 이 된다. 1개의 샘플로 이루어진  $f(x|x_{1i})$ 과 비교하면 분산이 샘플 수의 역수에 비례한다는 것을 알 수 있다. 이것은 샘플 수가 많아질수록 더 좁은 형태의  $f(x|x_1)$ 를 보여준다. 그리고  $f(x|x_1)$ 는 많은 수의 샘플로 이루어지므로 가우시안분포로 모델링이 가능하다. 이를 기반으로 검정을 위한 확률은 아래와 같이 유도할 수 있다.

$$\begin{aligned} \text{Hypothesis decision} &= \Pr(|x - F_0(x_1)| > D|x_1) \\ &= \int_{F_0(x_1)+D}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-F_0(x_1))^2}{2\sigma^2}} dx \\ &\quad + \int_{-\infty}^{F_0(x_1)-D} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-F_0(x_1))^2}{2\sigma^2}} dx \\ &= 2Q\left(\frac{D}{\sigma^2}\right) \end{aligned} \tag{7}$$

식 (7)의  $\sigma^2 = F_0(x_1)(1 - F_0(x_1))/k$ 이다. 식 (7)은 최대오차가 발견된  $x_1$ 에서 오차  $D$ 보다 클 확률을 보여준다. 식 (7)에서 사용된  $F_0(x_1)$ ,  $D$ 와  $k$ 은 각각

샘플들에 의한  $x_1$ 에서의 누적 밀도 확률,  $F_0(x)$ 와  $F(x)$  사이의 최대 오차 그리고  $x_1$  지점에서 사용한 샘플 수이다. 즉 Kolmogorov-Smirnov를 사용하지 않고도 필요한 파라미터들을 모두 구할 수 있다. 아래의 알고리즘은 Kolmogorov-Smirnov의 누적 밀도 함수를 사용하지 않고도 잡음을 분류하는 알고리즘이다.

$$1. D = \max|F(x) - F_0(x)| \rightarrow N\left(0, \frac{F_0(x_1) \times (1 - F_0(x_1))}{n}\right)$$

$$2. 1 - \alpha < 1 - 2Q\left(\frac{D}{\sqrt{F_0(x_1) \times (1 - F_0(x_1)) / n}}\right) \text{ 이면 } H_0 \text{ 결정}$$

$$3. 1 - \alpha \geq 1 - 2Q\left(\frac{D}{\sqrt{F_0(x_1) \times (1 - F_0(x_1)) / n}}\right) \text{ 이면 } H_1 \text{ 결정} \quad (8)$$

식 (8)을 관찰하면, 알고리즘 1은 최대오차거리를 찾기 위해 가우시안 분포의 누적 밀도 함수  $F(x)$ 와 샘플로 만든 누적 밀도 함수  $F_0(x)$ 를 비교하여 최대 오차  $D$ 를 찾는다. 찾은 최대오차는 평균이 0이고 분산이  $F_0(x_1)(1 - F_0(x_1))/k$ 인 가우시안분포로 근사화가 가능하다. 알고리즘 1에서 찾은 최대오차  $D$ 를 사용하여 알고리즘 2 및 3에서 식 (7)에서 구한 확률로 유의수준과 비교한다. 식 (3)과 달리 식 (8)에 설명된 알고리즘은 성공확률로 각 가정이 성립하는지 판단한다. 계산결과가  $1 - \alpha$ 보다 작으면 가우시안잡음이라 결정하고 크면 임펄스성 잡음이라 결정한다.

#### IV. 실험 결과

본 논문에서 제안된 Kolmogorov-Smirnov 테스트 방법을 사용하여 임펄스성 잡음 분류를 실험하였다. 가정을 위해  $F_0(x)$ 는 콤플렉스 가우시안 잡음을 사용하며, 사용될 미들턴 class A 잡음의 파라미터 조합 ( $A, \Gamma$ )는  $A$ 가 0.1과 0.8일 때  $\Gamma$ 의 범위는  $0.0001 \leq \Gamma \leq 0.5$  이다. 여러 종류의 임펄스성 잡음을 테스트하기 위해 임펄스성이 약한 큰  $A$ 와  $\Gamma$ 부터 임펄스성이 매우 강한 작은  $A$ 와  $\Gamma$ 를 사용하였다. 예를 들어, 파라미터 조합이 (0.8, 0.5)이면 임펄스성이 매우 약하고 (0.1, 1) 이나 (0.1, 0.0005) 등은 임펄스성이 강하다. 샘플 수가 10개여도 기존의 Kolmogorov-Smirnov 검정은 가우시안 잡음을 분류하기 때문에 이 실험에서 샘플 수는 10개이다. 유의수준은 0.1이고 100000번 시행한다. 검정 오류율은 아래와 같다.

$$ER = \frac{\text{The error trials}}{\text{The total trials}}$$

ER에서 The total trials는 총 시행횟수를 의미하며, The error trials는 모든 시행 중에서 잘못 판단한 시행 횟수를 의미한다.

그림 6과 7은 파라미터 조합에 따른 실험 결과를 보여 준다. 그림 6, 7에서 기존의 Kolmogorov-Smirnov 검정 방법은 임펄스성 성격이 강한 작은  $A$ 와  $\Gamma$  값에서 비교적 적은 오류율을 보여준다. 하지만 제안한 Kolmogorov-Smirnov 검정 방법은 더 적은 오류율을 보여주며 큰  $\Gamma$ 에서도 적은 오류율을 보여준다. 또한, 그림 7의 모든  $\Gamma$ 인 경우는 대부분 약한 임펄스성 성격을 지니지만 환경에 상관없이 기존의 Kolmogorov-Smirnov 검정 방법에 비해 낮은 오류율

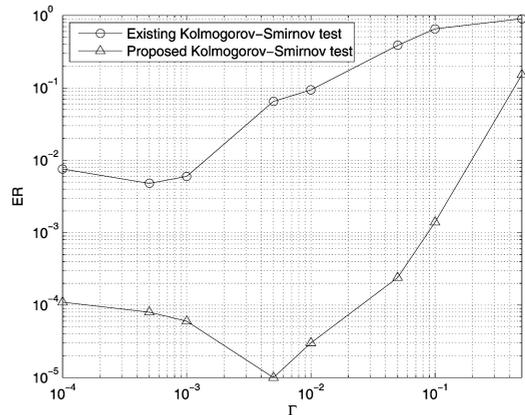


그림 6.  $A$ 가 0.1일 때 다양한  $\Gamma$ 에 따른 실험결과  
Fig. 6. Simulation results of various  $\Gamma$  with  $A = 1$

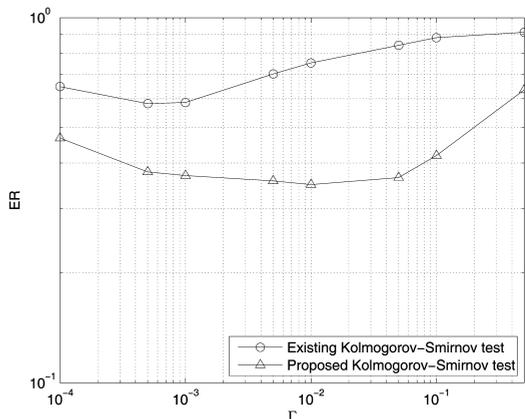


그림 7.  $A$ 가 0.8일 때 다양한  $\Gamma$ 에 따른 실험결과  
Fig. 7. Simulation results of various  $\Gamma$  with  $A = 0.8$

을 보여준다. 즉, 그림 6과 7은 제안한 알고리즘이 임펄스성이 강할 때는 오류없이 분류가 가능함을 보여주며 임펄스성이 약한 환경에서도 기존의 알고리즘에 비해 더 적은 오류를 보여준다.

Kolmogorov-Smirnov의 누적 밀도 함수는 식 (4)처럼 무한개의 지수함수로 표현된다. 그리고 식 (8)에서  $Q(x) = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)$ 이므로  $\operatorname{erfc}(x)$ 를 지수함수로 표현한다.  $\operatorname{erfc}(x)$ 는 아래와 같이 지수함수의 합으로 상한화시킬 수 있다<sup>13)</sup>.

$$\operatorname{erfc}(x) \leq \sum_{i=1}^N \frac{2(\theta_i - \theta_{i-1})}{\pi} e^{-\frac{x^2}{\sin(\theta_i)^2}} \quad (9)$$

식 (9)에서  $\theta_i = \frac{i\pi}{2N}$ 이고  $N$ 은 지수함수의 갯수이다.  $N$ 이 3이면  $\operatorname{erfc}(x)$ 는 아래와 같이 상한화 된다.

$$\operatorname{erfc}(x) \leq \frac{1}{3}e^{-4x^2} + \frac{1}{3}e^{-\frac{4}{3}x^2} + \frac{1}{3}e^{-x^2} \quad (10)$$

복잡도 비교를 위해 식 (4)의  $k=1,2,3$  항들과 식 (10)을 비교하여 mean square error(MSE)를 이용하여 복잡도를 비교한다. MSE는 아래처럼 표현된다.

$$\text{MSE} = (f(x) - \hat{f}(x))^2 \quad (11)$$

식 (11)에서  $f(x)$ 는 원래 함수이고  $\hat{f}(x)$ 는 대략

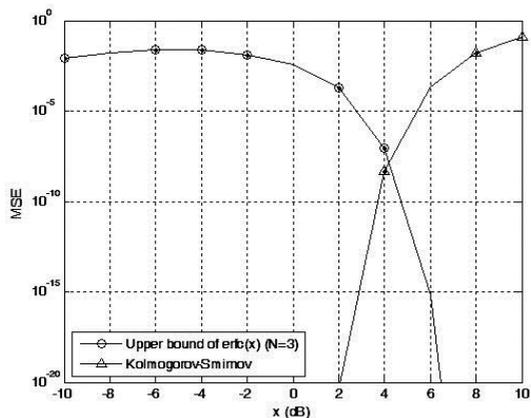


그림 8. 상한화한  $\operatorname{erfc}(x)$ 와 Kolmogorov-Smirnov의 거리에 따른 MSE비교  
Fig. 8. MSE comparison between upper bound of  $\operatorname{erfc}(x)$  and distance of Kolmogorovo-Smirnov

화 또는 상한화된 함수이다. Kolmogorov-Smirnov의 경우에 무한개까지 더한 함수가  $f(x)$ 이고  $k=1,2,3$  항들의 합이  $\hat{f}(x)$ 이다.  $\operatorname{erfc}(x)$ 의 경우에  $\operatorname{erfc}(x)$ 가  $f(x)$ 이고 식 (10)이  $\hat{f}(x)$ 이다.

그림 8은  $\operatorname{erfc}(x)$ 와 Kolmogorov-Smirnov의 대략화시킨 함수들의 MSE를 그렸다. 모든  $x$ 에 대해  $\operatorname{erfc}(x)$ 는 낮은 MSE를 보여주는 반면에 Kolmogorov-Smirnov는  $x$ 가 -4 dB 이상일 때 큰 오차를 보여준다. 이 사실을 뒷받침해주기 위해 그림9와 10을 보여준다. 그림 9는  $\operatorname{erfc}(x)$ 와 상한화된 함수들을 그렸으며  $\operatorname{erfc}(x)$  및 상한함수가 모두 감소함수이고, 두 함수들의 차이도 적음을 보여준다. 반면에 그림 10은  $x$ 가 증

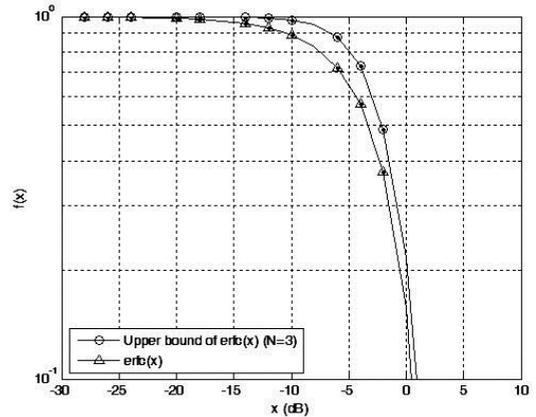


그림 9.  $\operatorname{erfc}(x)$ 의 상한함수와  $\operatorname{erfc}(x)$ 의 그림  
Fig. 9. plotting the  $\operatorname{erfc}(x)$  and upper bound of  $\operatorname{erfc}(x)$

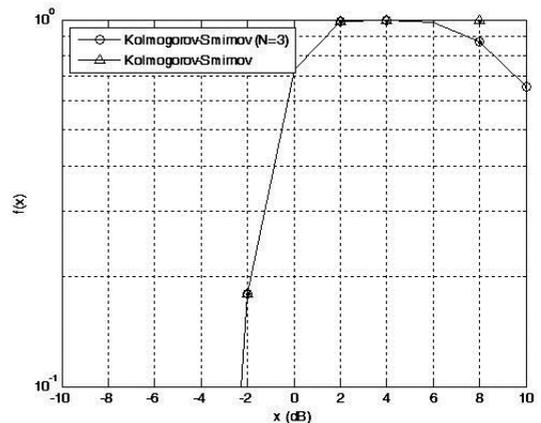


그림 10. 처음 3개의 항들만 고려한 Kolmogorov-Smirnov (파란색) 누적 밀도 함수와 무한개의 항을 고려한 Kolmogorov-Smirnov 누적 밀도 함수  
Fig. 10. The first three terms of Kolmogorov-Smirnov cumulative density function (blue) and conventional Kolmogorov-Smirnov cumulative density function (red)

가할 때 증가함수이지만 -4 dB 이상이 될 때 다시 감소하는 현상을 보여준다.  $x$ 가 10 dB 일때는 0.35의 큰 오차를 보이므로 Kolmogorov-Smirnov 누적 밀도 함수는 많은 항들이 필요하다는 것을 알 수 있다.

## V. 결 론

본 논문에서 많은 전자기기들에 의해 영향을 받아 언급되는 임펄스성 잡음의 분류법에 대해 언급하였다. 임펄스성 잡음의 영향을 줄이기 위해 대표적으로 비선형 블락을 사용하며, 이것 때문에 임펄스성 잡음을 분류하는 시도가 요구된다. 특히, 여러 가지 분류법 중 대표적인 Kolmogorov-Smirnov 방법에서 높은 복잡도와 검증 오류 확률이라는 단점을 보완하기 위해 새로운 방법을 본 논문에서 제안했다. 그 결과, 가우시안 잡음뿐만이 아니라 임펄스성 잡음도 명확하게 분류할 수 있음을 실험결과를 통해 알 수 있었다. 특히, 이론적 접근이 어려운 Kolmogorov-Smirnov 테스트의 접근은 실용적인 측면에서 낮은 복잡도를 지닌 방법으로 접근한 것은 큰 의의를 지닐 수 있다.

## References

[1] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, no. 3, pp. 106-127, Aug. 1977.

[2] A. D. Spaulding and D. Middleton, "Optimum reception in an impulsive interference environment-part I: Coherent detection," *IEEE Trans. Commun.*, vol. COM-25, no. 9, pp. 910-923, Sept. 1977.

[3] 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. Inf. Theory*, vol. 45, no. 4, pp. 1129-1149, May 1999.

[4] K. Gulati, B. L. Evans, J. G. Andrews, and K. R. Tinsley, "Statistics of co-channel interference in a field of poisson and poisson-poisson clustered interferers," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6027-6222, Dec. 2010.

[5] T. S. Saleh, I. Marshland, and M. El-Tanany, "Suboptimal detectors for alpha-stable noise:

Simplifying design and improving performance," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2982-2989, Oct. 2012.

[6] S. M. Zabin and H. V. Poor, "Parameter estimation for middleton class a interference processes," *IEEE Trans. Commun.*, vol. 37, no. 10, pp. 1042-1051, Oct. 1989.

[7] S. M. Zabin and H. V. Poor, "Efficient estimation of class a noise parameters via the EM algorithm," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 60-72, Jan. 1991.

[8] J. Friedman, H. Messer, and J. Cardoso, "Robust parameter estimation of a deterministic signal in impulsive noise," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 935-942, Apr. 2000.

[9] K. S. Vastola, "Threshold detection in narrow-band non-gaussian noise," *IEEE Trans. Commun.*, vol. COM-32, no. 2, pp. 134-138, Feb. 1984.

[10] D. J. Gibbons, *Nonparametric Methods for Quantitative Analysis*, 3<sup>rd</sup> Ed., Amor Sciences, 1996.

[11] W. J. Conover, *Practical Nonparametric Statistics*, 3<sup>rd</sup> Ed., Wiley, 1999.

[12] J. H. Drew, A. G. Glen, and L. M. Leemis, "Computing the cumulative distribution function of the kolmogorov smirnov statistic," *Computational Statics & Data Analysis*, vol. 34, no. 1, pp. 1-15, Jul. 2000.

[13] M. Chiani and D. Dardari, "Improved exponential bounds and approximation for the Q-function with application to average error probability," in *Proc. IEEE GLOBECOM*, pp. 1399-1402, Taipei, Taiwan, Nov. 2002.

오 형 국 (Hyungkook Oh)



2012년 2월 : 한양대학교 전자  
통신공학과 졸업  
2012년 3월~현재 : 한양대학교  
전자통신공학과 박사과정  
<관심분야> 임펄스성 잡음에서  
의 통신시스템 설계, 이동통  
신 시스템

남 해 운 (Haewoon Nam)



1997년 2월 : 한양대학교 공학  
학사  
1999년 2월 : 서울대학교 공학  
석사  
1999년~2002년 : 삼성전자 통신  
연구소 선임연구원  
2006년 12월 : University of  
Texas at Austin 전기컴퓨터공학과 박사  
2006년~2010년 : Motorola Inc., Senior Staff  
Engineer  
2010년~2011년 : Apple Inc., Principal Engineer  
2011년~현재 : 한양대학교 전자통신공학과 교수  
<관심분야> 이동통신, 통신시스템설계, 임베디드 시  
스템