디지털 통신 시스템의 성능 지표

확률과 랜덤변수

황승훈
동국대학교

Ⅰ. 랜덤신호와 통신시스템

지난 2016년 여름은 유난히도 무더웠던 것 같다. 그래서인지 기상 예보에 대해 일반인들의 관심도 많았고 경우에 따라서는 예측이 빗나감에 따른 불만의 소리도 있었다. 이와 같이 실제 생활에 있어서 기온이 몇 도일지 비가 올지 안 올지 미리 어떤 결과에 대해 이야기하는 것은 매우 불확실하다. 그런데, 실생활에서처럼 공학 문제에서도 마찬가지로 불확실한 사건(event)들이 일어나기 마련이다. 따라서 기온이 몇 도인지 정확한 값을 도출하기 어렵듯이 공학 문제의 정확한 값을 구해낼 수 있는 결정론적인(deterministic) 식을 유도해내기 어려운 경우가 많다. 우리는 이러한 경우를 무작위적(random)이라고 이야기한다.

전자전기공학부의 기초과목인 신호 및 시스템(signal and systems)에서 흔히 신호(signal)를 시간에 대한 함수로 정의를 하는데, 가장 대표적인 경우로 확정 신호와 랜덤 신호로 분류한다[1]. 〈그림 1〉과 같이 확정 신호는 sin(ωt)와 같이 시간 t_1이 정해지면 sin(ωt_1)이란 값이 확실하게 '결정'된다. 이에 반해 랜덤 신호는 그림1의 잡음(noise)와 같이 시간이 정해지더라도 그 때의 값을 확실하게 알 수 없는 경우이다. 시스템이란 입력 신호가 들어가서 출력 신호를 내보내주는 블랙박스와 같은 것으로 보통의 경우 입력 신호와 시스템을 정의하는 임펄스 응답을 알고 있는 경우 출력 신호를 구할 수 있게 된다.

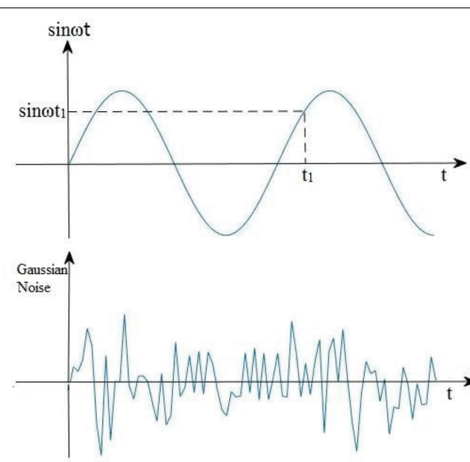

그림 1. 확정 신호와 랜덤신호

통신 시스템의 경우 〈그림 2〉와 같이 보통 전송부, 채널, 수신부와 같이 세 개로 나눌 수 있다[2]. 전송부에서는 통신의 대상인 정보(information)를 반송파(carrier)에 실어주는 변조(modulation) 과정이 핵심이며 수신부에서는 이와 반대로 반송파에 실려있는 정보를 다시 분리

해내는 복조(demodulation) 과정을 통해 정보를 전달하고자 하는 통신의 임무를 완성하게 된다. 채널은 전송부와 수신부 사이를 차지하고 있는 매체로서 유선과 무선 채널로 분류할 수 있다. 정도의 차이가 있겠지만 채널은 보통 잡음이나 간섭과 같이 정보의 전달이라고 하는 통신의 동작을 방해하는 요소들이 섞여 들어오는 지점이다.

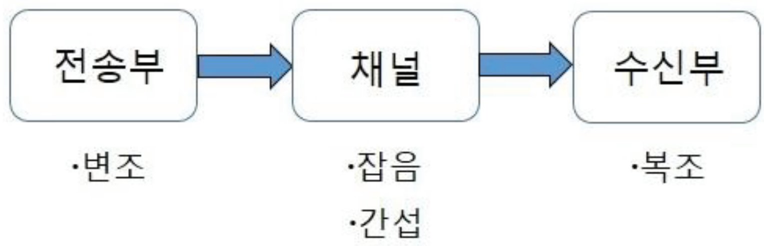

그림 2. 통신 시스템의 블록 다이어그램

다시 말하면 전송부와 수신부에서 잡음원이 전혀 없다고 가정하는 경우, 수신된 정보에 발생하는 오류의 원인은 채널이 될 것이며 정보 신호에 랜덤신호인 잡음이나 간섭이 더해지거나 곱해지게 된다. 따라서 정보 신호가 디지털인 디지털 통신 시스템의 경우 수신 정보 신호는 보내진 −1이나 1의 값이 아닌 1이나 −1의 값으로 잘못 결정되게 된다. 일반적으로 디지털 통신 시스템의 가장 대표적인 성능 지표는 전체 전송부에서 보내진 비트수 대비 수신부에서 결정된 비트들 중에 오류가 난 비트수인 비트 오류 확률(bit error probability)로 나타내어 지는데, 본 고에서는 이와 같이 성능 지표로 사용되는 확률과 이를 위해 제안된 랜덤 변수에 대한 개념을 소개하고, 대표적인 누적 분포 함수와 정규분포를 예로 들어 설명하고 마무리하고자 한다.

II. 확률과 랜덤변수

먼저 확률에 대한 이야기를 꺼내기 전에 확률 이론이 바탕으로 두고 있는 집합론(set theory)에 대한 이야기를 하고자 한다. 집합론은 이미 고등학교 교과과정에서 학습한 내용이므로 본 고에 도움이 되는 몇 가지 집합론의 개념을 확인하고자 한다[3].

- 집합(set): 어떤 조건에 따라 결정되는 요소의 모임
- 원소(element): 위 집합을 이루는 각각의 요소
- 부분집합(subset): 두 집합 A, B에서 $x \in A$인 임의의 원소 x에 대하여 반드시 $x \in B$일 때, A를 B의 부분집합이라 하고, $A \subset B$ 또는 $B \supset A$로 나타내며, 이것을 A는 B에 포함된다, 또는 B는 A를 포함한다고 한다.
- 전체집합(Universal set): 최초의 집합 U를 고정하고, 이것을 근거로 하여 그 모든 부분집합만을 고찰할 때 이 U를 전체집합이라고 한다.

〈그림 3〉은 집합을 그림으로 표현하는 방법인 벤 다이어그램인데 그 안에 전체집합, 부분집합, 원소 등이 표현되어 있다. 일반적으로 전체집합은 사각형으로 표현하고 집합은 대문자, 원소는 소문자로 표시한다.

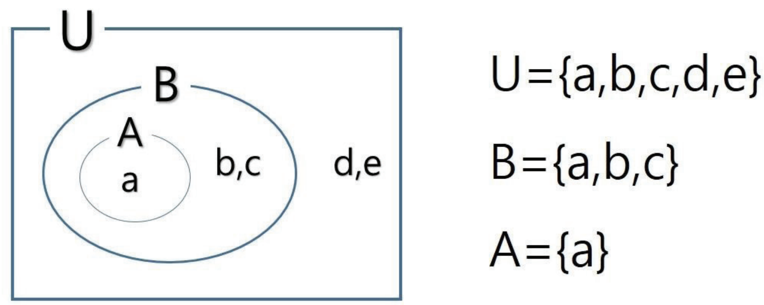

그림 3. 벤 다이어그램

이제 다시 확률로 돌아오도록 하자. 확률은 음이 아닌 실수 값으로서 세가지 주요 요소들이 먼저 정의가 되어야 한다. 첫째 실험(experiment), 둘째 결과(outcome), 셋째 사건(event)이다 [4]. 이 세가지는 서로간에 유기적인 연관을 가지며 또한 집합론과도 대응이 된다. 예를 들어 주사위 던지기, 동전 던지기, 플레잉 카드(트럼프) 중에서 한 장 뽑기와 같이 실험을 생각해볼 수 있다. 이와 같은 실험을 실제로 직접 수행해보는 것을 실행(trial)이라고 하는데 항상 실행에는 결과(outcome)가 얻어지게 된다. 위 예로든 실험에서는 주사위 여섯 면 중에서 한 면의 주사위 값, 동전의 앞이나 뒷면, 카드 52장중에 뽑힌 한 장이 결과로 얻어지게 된다. 그런데 각 실험에 대해 여러 번의 실행을 통해 얻어질 수 있는 모든 결과들을 생각해 볼 수 있으며 이를 표본공간(sample space)라고 부른다. 이 표본공간은 전체집합에 해당하며, 표본공간을 이루는 결과들을 원소라고 생각할 수 있으므로 그 중의 일부는 전체집합의 부분집합에 해당되며 사건이라고 부른다. 이제 사건에 대해서 비로소 확률을 할당할 수 있게 되는 것이다.

주사위 짝수 값이 나오는 사건을 예로 들어보자. 주사위를 던지는 실험에서 나오는 모든 경우의 수로 이루어진 전체집합 (표본공간)을 {1,2,3,4,5,6}이라고 원소 나열법으로 표현할 수 있고 부분집합인 짝수 값이 나오는 사건은 {2,4,6}으로 나타낼 수 있다. 이제 비로소 사건 {2,4,6}에 확률을 할당할 수 있으므로 P{2,4,6}으로 쓸 수 있게 되는 것이다. 물론 {2,4,6}=E라고 표현하면 P(E)라고도 나타낼 수 있다.

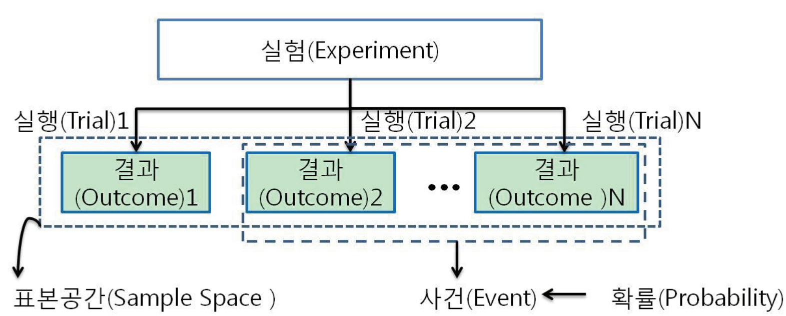

그림 4. 확률 모델

〈그림4〉에서 집합론과 관련 지어 실험, 표본공간, 사건을 정의하고 결론적으로 사건에 확률을 할당할 수 있다는 것을 보였다. 그런데 예로 들은 주사위의 경우와 다르게 52장의 원소를 가지는 플레잉 카드 중에서 한 장 뽑기와 같은 실험의 경우 사건으로 나타내는 것이 다소 복잡하게 된다. 더 복잡한 경우의 실험과 이에 따른 사건들도 많이 등장하게 된다. 따라서 사건을 어떻게 하면 간단하게 표현할 수 있을지에 대한 고민을 하게 되고 이를 위해 고안된 도구를 랜덤 변수

(random variable)라고 한다.

이제 랜덤 변수에 대해 소개하고자 한다. 위에서 이야기한 것처럼 랜덤변수는 어떤 물리적인 의미를 가진 개념이라기보다는 복잡함을 줄여주기 위해 즉 편리하기 해 고안해낸 도구이다. 그런데, 확률을 할당하기 위해 집합론을 살펴본 것처럼 랜덤 변수를 정의하기 위해서는 함수의 개념을 확인해야 한다[3].

- 함수(function): 변수 x와 y 사이에 x의 값이 정해지면 따라서 y값이 정해진다는 관계가 있을 때, y는 x의 함수라고 한다.
- 정의역(domain of definition): 주어진 함수에 대해 그 함수가 정의되는 모든 수의 집합
- 공변역(codomain): 집합 X의 임의의 원소 x에 대응하는 집합 Y의 원소 y를 나타내는 함수 y=f(x)에서 집합 Y를 부르는 말이다.
- 치역(range): 함수가 취하는 값 전체의 집합
- 함수의 조건: 정의역의 한 원소에 공역의 원소가 오직 하나 대응되는 관계를 함수라고 한다.

랜덤 변수란 간단히 함수라고 말할 수 있는데, 정의구역을 표본공간으로 하고 공변역을 실수로 하는 함수라고 정의할 수 있다. 함수란 y=f(x)=2x와 같이 x와 y간의 관계를 설정하는 식이 있는데 랜덤 변수도 마찬가지로 해당 식에 의해서 정의구역인 표본공간의 원소인 각각의 결과들이 실수 수직선상의 어떤 한 점과 매핑을 하게 되는 것이다.

예를 들어 주사위를 던지는 실험의 경우 〈그림5〉와 같이 표본공간은 S={1,2,3,4,5,6}이므로 랜덤변수 Z, 즉 Z=Z(s)=2s에 의해 Z={2,4,6,8,10,12}로 매핑이 되게 된다. 하지만 이 경우에는 숫자가 숫자로 매핑이 되므로 랜덤 변수의 편리함이나 필요성을 확인하기가 쉽지 않다. 또 다른 예를 들면 52장의 플레잉 카드 중 한 장을 뽑는 실험에서는 표본공간이 52개의 원소를 가지며 각각의 원소를 지칭하기 위해서는 다이아몬드, 스페이드, 하트, 클로버 모양과 1부터 10까지의 숫자와 Jack, Queen, King까지의 조합으로 표현해야 한다. 하지만 이를 랜덤변수를 사용한다면 어떠한 룰에 따라서든 결국 실수 수직선위에 나타낼 수 있게 되고 사건들도 숫자들의 조합으로 나타낼 수 있다. 이와 같이 사건의 표현이 모두 수직선 위에서 나타날 수 있다는 점이 사건에 할당되는 확률을 다루는데 있어서 큰 편리함과 간소함을 가져다 주기 때문에 랜덤 변수를 활용하게 되는 것이다.

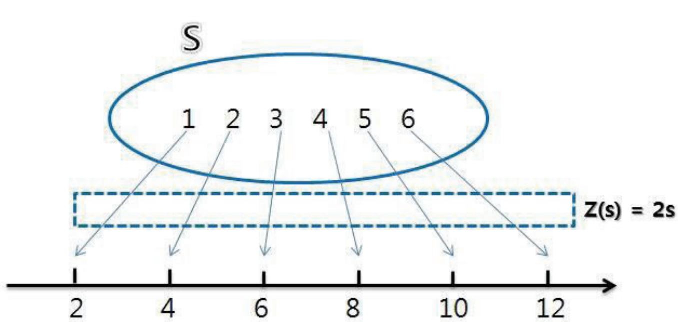

그림 5. 랜덤 변수

III. 누적분포함수와 정규분포

전장에서 랜덤변수를 정의하였으므로 이제 본격적으로 누적분포함수와 확률밀도함수에 대해 소개하고자 한다. 〈그림 5〉에서와 같이 랜덤변수를 통하여 어떠한 실험의 사건이라도 숫자를 원

소로 가지는 사건으로 표현이 될 수 있다. 따라서 랜덤변수는 실수 수직선상에 매핑이 되어있으므로 수의 범위로 사건을 손쉽게 표현할 수 있고 역으로 생각하면 이 수의 범위는 랜덤변수의 역함수를 통해 원래 사건으로 매핑이 되어 있음을 그림을 통해 쉽게 이해할 수 있다. 예를 들어 실수인 랜덤변수 Z가 어떤 실수 z보다 작거나 같은 범위라고 한다면 즉 $\{Z \le z\}$은 사건을 의미한다. 예를 들어 $\{Z \le 7\}$이라고 하면 Y의 부분집합인 사건 $\{2,4,6\}$을 의미하고 이는 X의 부분집합인 사건 $\{1,2,3\}$을 지칭하는 것이다. 이제 $\{Z \le z\}$에 확률을 할당하여 $P\{Z \le z\}$로 표현할 수 있는데 이를 랜덤변수 Z의 누적분포함수 (cumulative distribution function: CDF)라고 하고 $F_Z(z)$라고 표시한다. 그리고 누적분포함수를 z에 대해 미분한 것을 랜덤변수 Z의 확률밀도함수(probability density function: PDF)라고 하고 $f_Z(z) = dF_Z(z)/dz$로 표시한다[5].

다양한 문제들 속에서 마주치게 되는 랜덤 변수의 종류 중에 가장 대표적인 것은 가우시안 또는 노말 랜덤변수이다. 통신 시스템의 성능을 고려할 때 기본이 되는 가장 간단한 잡음이 AWGN(Additive White Gaussian Noise)인데 가우시안 랜덤변수 X는 다음과 같은 PDF를 가진다.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{(x-\mu)^2}{2\sigma^2}] \qquad (1)$$

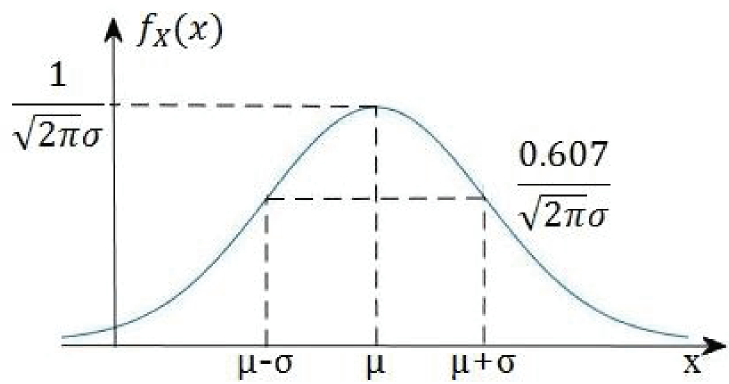

그림 6. 가우시안 PDF

〈그림 6〉에서 파라미터 μ는 랜덤변수의 평균(mean 또는 average value), 파라미터 σ^2는 분산(vaiance)이라고 불린다. 파라미터 σ는 표준편차(standard deviation)으로 분포의 퍼짐의 정도를 나타내는 지표가 된다. 가우시안 CDF는 식 (1)의 가우시안 PDF를 적분하여 구할 수 있는데 이 적분식을 손쉽게 계산해낼 수는 없다. 이를 간편히 하기 위해서 $\mu=0$, $\sigma^2=1$인 경우로 대표가 되는 가우시안 CDF를 정규 분포라 부르고 정규분포표에 그 적분값을 미리 계산하여둔다. 다음은 정규분포의 식이다. 보통 누적분포함수와 마찬가지로 F로 표시하지만 랜덤변수를 나타내는 첨자 X가 없음에 유의하도록 한다.

$$\mathrm{F(x)} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp[-\frac{z^2}{2}]\,dz \qquad (2)$$

그리고 정규분포와 임의의 μ, σ^2 값을 가지는 CDF간의 관계를 통해 우리는 손쉽게 적분식의 계산값을 정규분포표에서 찾아낼 수 있는 것이다.

$$F_X(x) = P\{X \le x\} = F\left(\frac{x-\mu}{\sigma}\right) \qquad (3)$$

전자, 전기, 통신공학에서 자주 사용되는 Q 함수라는 것이 있는데 누적분포와 다음과 같은 관계를 가진다.

$$Q(x) = 1 - F(x) \tag{4}$$

가우시안 잡음이 존재할 때 이진 시그날링을 사용하는 디지털 통신 시스템의 오류확률은 Q함수를 이용하여 다음과 같이 표현된다.

$$\mathrm{P(error)} = \mathrm{Q}\left(\sqrt{z}\right) \tag{5}$$

여기서 z는 신호 대 잡음비(signal-to-noise ratio) 또는 비트당에너지 대 잡음전력밀도비를 포함하는 값으로 변조 방식에 따라 달라질 수 있다. 자세한 오류 확률 식과 유도는 통신이론 책을 참조하기 바라며 z값이 커질 때 $\mathrm{F}\left(\sqrt{z}\right)$값이 커지므로 식 (4)에서 $\mathrm{Q}\left(\sqrt{z}\right)$값이 작아지므로 P(error) 값이 작아지게 된다. 따라서 신호의 전력이나 비트의 에너지 값이 커질수록 오류가 일어날 확률이 작아진다는 사실을 확인할 수 있다.

IV. 확률의 역사

몇 개의 수식이 나오면서 약간은 복잡해진 머리를 식히기 위해 본 장에서는 확률에 관련된 역사를 참고문헌 [6]에서 발췌하여 소개하고자 한다.

역사적으로 확률은 학자이자 도박꾼이었던 카르다노와 관련하여 시작되었다. 카르다노는 확률게임에서 확률을 상대방보다 더 잘 이해함으로써 승리할 가능성을 높일 수 있다는 주제로 1633년 '확률게임에 관한 책'을 출판하였다. 이 책에는 확률 이론을 최초로 체계적으로 다룬 학술적인 내용이 포함되어있다. 확률이론은 블레즈 파스칼의 주목을 받으면서 비상하기 시작했다. 확률게임이 중도에 끝난 경우 판돈을 어떻게 나누어야 하는지에 관한 문제에 대한 새로운 해법을 제시했다. 이 해법의 핵심이 오늘날 기대값(expectation)이라고 부르는 개념이다. 확률이론은 1713년 야코프 베르누이가 '추측의 기술'을 발표하면서 완전히 독립적인 영역으로 모습을 갖추었다. 1730년경 아브라함 드 무아브르는 동전을 반복하여 던지는 것과 관련된 확률에 대한 근사식을 도출하였는데 이것은 정규분포로 이어져서 종형 곡선이라고 불렀다. 이와 같은 종형곡선은 수학이론 분야 뿐만 아니라 사회과학의 경험적 데이터들에서도 모습을 드러내게 되었다.

IV. 결 론

본 고에서는 랜덤신호, 확률, 랜덤변수를 소개하면서 기본적인 확률과 랜덤변수 이론을 기술하였다. 아주 자세한 사항은 지면상 생략된 면이 있지만 처음 이 글을 쓰기 시작하면서 목적은 뼈대를 세우고 전체 건물의 모습을 독자에게 전달하고자 했으므로 그 처음의 뜻은 이룬 것 같다.

랜덤변수는 시간 파라미터를 포함하여 확장할 수 있는데 이를 랜덤 프로세스라고 부르며 시간에 대한 상관성(correlation), 시간에 대한 정상(stationary), 주파수 영역에 대한 고려 등을 추가적으로 학습할 수 있을 것이다. 아무쪼록 이 글이 디지털 통신 시스템의 성능 지표인 확률에 대한 보다 넓고 깊은 이해의 기초가 되었으면 하는 바램이다.

감사의 글

이 글은 사랑하는 아들과 제 강의를 듣는 수강학생들을 염두에 두며 작성됐고 이 글을 읽고 통신을 포함한 공학 전반에 대한 청소년들의 흥미와 저변이 확대되길 기원해본다.

참 고 문 헌

[1] Simon Haykin, Signals and systems, Wiley.

[2] Rodger E. Ziemer, Principles of Communications, seventh edition, Wiley.

[3] 두산백과

[4] Peyton Z. Peebles and Bertram Shi, Probability, random variables and random signal principles, Mc Graw Hill.

[5] Charles W. Therrien and Murali Tummala, Probability and random processes for electrical and computer engineers, second edition, CRC press.

[6] 이언 스튜어트, 세계를 바꾼 17가지 방정식, 사이언스북스