

가우스(1855)의 동전 한 닢

이문호 (전북대학교)

우리 선조들은 가끔 점을 쳐서 내일이 길흉화복(吉凶禍福)을 예측했다. 오늘날에는 복권 한 장에 마음을 졸이며 대박을 기다리는 현대인들에 이르기까지, 우리 인류는 항상 확률적인 상황에 직면해 왔다. ‘확률’이라는 말이 개입되는 순간부터, 우리의 삶은 하나의 도박이 되는 것이다. 병원에서 질병 감염 여부를 검사할 때나 법적 증거로 DNA 유전자를 감식할 때, 거기에는 항상 ‘확률’적 요인이 숨어있다. 그 중심에 가우스(Carl Fredrich Gauss, 1777-1855, 독일)가 있다. EU 통합 전 독일의 10 Mark 화폐 주인공 가우스, 가우스는 독일의 자존심이다. 고대부터 인간은 무엇인가를 결정할 때 확률적 결정에 따른다. 본고에서는 가우스 확률분포의 기원을 추적 요약하였다.

I. 서론

랜덤 프로세스(Random Process)인 Stochastic 확률신호와 결정된 신호(Determinant Signal)은 분명히 서로 상반되는 개념이다. 그런데, 우리는 무슨 수로 Stochastic 확률신호를 찾을 것인가?

한마디로, 불확실한 신호처리를 확실하게 측정(measure)하여 측정횟수를 무한대로 증가시키면 확률적으로 가우스 분포를 갖는다.

우리가 겪는 일상사들은 확률을 좌우하는 변수들이 너무 많기 때문에 정확한 확률을 예견한다는 것은 애초부터 불가능한 일이다. 그러나 동전과 주사위를 던졌을 때 모든 눈금들이 동일한 확률로 나온다는 이 간단한 원리를 마음속 깊이 새겨둔다면, 잘못된 통계자료로 인한 오류는 상당부분 극복될 수 있다. 정상에서 벗어난 통계자료가 편향(Biased)된 방향

에 따라 ‘비관’과 ‘낙관’이 좌우된다. 세월이 흐를수록 우리의 주변 상황을 좌우하는 확률변수들이 날이 갈수록 많아진다. 그러므로 이들을 종합하여 어느 쪽으로도 편향되지 않은 결과를 도출해 내는 능력이 필요하다. 즉, 가우스 확률분포가 중요한 이유이다.

주식 시세를 예견할 때나 세계 경제 더블 딥(Double Dip)을 예측할 때, 복권이나 Project에 응모할 때, 시험을 보고 결과를 기다릴 때도 정답결과를 맞출 확률은 수많은 숨어있는 변수 때문에 그리 높지 않다.

본고에서는 Semi Deterministic 관점에서 가우스 분포를 유도하는데 2장에서는 가우스 생애를 살펴보고, 3장에서는 벨(Bell) 곡선의 기원을, 4장에서는 가우스 확률분포를 증명하고 5장에서는 결론을 맺는다.

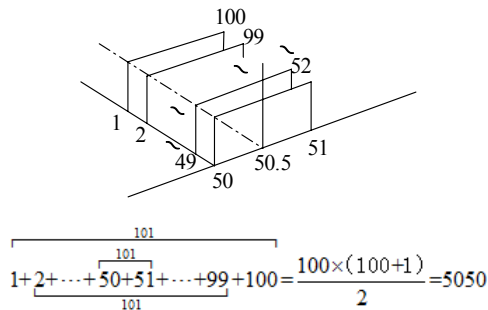
II. 가우스의 생애



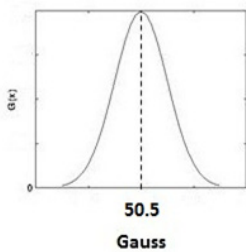
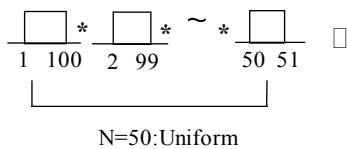
◀ 가우스는 천재적인 수학자였다. 초등학교 시절 1부터 100까지 합을 구하는 문제를 간단히 해결한 일화는 그의 천재성을 보여준 단편적인 예이다. 즉, $\frac{n(n+1)}{2} = \frac{100(101)}{2} = 5,050$.

이때, 벌써 가우스분포를 직감한 것은 아닐까?

간단한 증명을 보이면 1과 100의 합은 101이고 2와 99의 합은 101이며, 이것이 50개가 있다. 즉 직폭이 1, 대역폭이 1, 평균치가 0.5인 Uniform이 50개가 있는데, 이를 50개 중첩시킨 가우스 분포 꼴, <그림 1>과 같이 Gauss 분포 꼴이 된다.



Convolution



〈그림 1〉 1,2,3,4,5 … , 96,97,98,99,100 까지 합이 Uniform 분포와 중첩된 경우 Gauss 분포 꼴

독일의 수학자이자 과학자로 ‘수학의 왕자’라 일컬어진다. 벽돌공의 아들로 태어나 어려운 생활을 했다. 가우스의 천재적 재능을 일찍 안 브룬스바이크 공(公)의 도움으로 대학입학 전까지 학업에 전념할 수 있었다. 아카데미 시절 이항정리, 산술-기하 평균, 소수론과 보데의 법칙 등을 발견하였다. 이후 1795년 괴팅겐 대학에 입학한 후 더욱 왕성한 활동을 했다. 2차 형식에 관한 상호법칙과 최소자승법을 발견했고 복소수 평면을 도입했다. 대학을 졸업하던 해인 1798년, 열여덟의 나이로 2천년 이상 관심사였던 정17각형의 작도법을 발견해내기도 했다. 이 발견으로 수학적 전체가 흥분하기도 했다. 그 외 가우스기호, 가우스함수 등 수많은 발견을 했다. 무엇보다도 1799년 박사논문에서 발표한 ‘대수학의 기본정리’는 수세기 동안 수학자들이 이해하지 못했던 것이다.

1801년 출판된 ‘수론연구’(Disquisitiones Arithmeticae)는 정수론과 관련된 업적들을 총망라한다. 특히 방정식 $x^n - 1 = 0$ 의 해를 구하는 것과 원주의 n등분 작도 가능성과의 관계를 정수론적으로 설명하고 있다. 정17각형의 작도는 한 예이다. 수학사에서 18세기와 19세기의 구분은 가우스를 경계로 한다고 해도 과언이 아니다. 복소수를 수학적 대상으로 인식, 기호화해 엄격한 수체계를 확립했다.

그는 천문학에서도 두드러졌다. 1801년 소행성 세레스를

발견하고 그 궤도를 최소자승법으로 계산해 다음번 출현을 정확히 예측한 바 있다. 이후 괴팅겐 천문대장으로 재직한다. 가우스분포를 발견하는 등 현대 통계학의 기초를 다진 사람 중 한명이다. 1820년대에는 베버와의 공동연구를 통해 광학, 자기학, 기계학, 음향학 등 많은 물리학 분야에 업적을 남겼다. 가우스는 정수론-우리가 잘 아는 연립방정식의 가우스 소거법-, 해석학, 미분기하학, 측지학, 자기학, 천문학, 광학, 확률론 등 광범위한 분야에 걸쳐 업적을 남겼고 오랫동안 강한 영향력을 행사했다. 다방면에 걸친 학문적 관심을 가진 수학자는 가우스가 마지막이 아닌가 싶다. 가우스 이후 수학의 분야가 발전해 개개의 분야로 전문화하는 경향을 보였기 때문이다.

Ⅲ. 벨(Bell) 곡선의 기원

1778년 피에르 시몽 드 라플라스는 측정오차의 확률분포를 보여주는 곡선 그래프를 최초로 제안하였다. 그 후 1808년에 ‘종 모양 곡선’이라 불리는 무작위오차의 정규분포 곡선이 로버트 에이드리언 Robert Adrian이라는 무명의 미국인에 의해 발견되었으며, 1년 뒤 1809년 독일의 수학자이자 물리학자, 그리고 천문학자였던 칼 프리드리히 가우스도 독립적으로 정규분포 곡선의 형태를 발견하였다. 에이드리언은 항해술과 천문학, 측지학 등의 분야에서 얻어진 측정오차로부터 곡선의 방정식을 유도하였고, 가우스는 천체의 운동을 측정할 때 나타나는 오차의 분포법칙을 설명하는 과정에서 에이드리언과 동일한 곡선방정식을 얻었다. 그러나 현대에 와서 정규분포 곡선은 ‘가우스 분포’라는 명칭으로 굳어지게 되었다.

1810년, 라플라스는 파리 학술원에서 그 유명한 ‘중앙수렴 정리’를 발표하였다. 그의 이론은 1812년에 완전히 정립되었으며, 이로부터 측정오차의 분포 곡선과 정규분포 곡선 사이의 상호 관계도 성공적으로 규명되었다.

종 모양 곡선이 측정오차의 확률분포 곡선으로 대두되기 이전에, 드무아브르(De Moivre)는 불연속확률을 계산하던 중 이와 동일한 모양의 곡선방정식을 유도하여 《기회의 원칙》이라는 학술지 1733호에 발표하였다. 당시 드무아브르는 동전 던지기 게임의 경우처럼 발생할 확률과 발생하지 않을 확률이 동일한 사건들에 대하여, 각각의 발생 횟수를 확률적으로 예견하는 연구에 몰두하고 있었다. 이는 동전을 여러 차례 던졌을 때 앞면이 나타나는 횟수를 확률분포도로부터 예견하는 것과 비슷한 작업이라고 할 수 있다. 지금부터 드무아브르의 논리를 따져가면서, 특히 실행 횟수가 매우 많아졌을 때 나타나는 현상들을 주의 깊게 살펴보기로 하자.



〈그림 2〉 EU 통합 전 독일 10마르크 지폐 인물 가우스

동전을 단 한번만 던진 경우, 앞면이 나오는 횟수는 1 또는 0이다. 이 때, 앞면이 한번 나올 확률과 0번 나올 확률은 완전히 동일하므로 확률분포 그래프 상에서 각각의 경우에 해당되는 막대그래프를 그린다면 두 경우 모두 동일한 길이(1/2, 또는 0.5)의 막대가 그려질 것이다. 〈표 1〉 동전을 두 번 던진 경우에는 앞면이 나올 수 있는 횟수가 0,1, 또는 2인데 앞면이 한번 나올 확률은 0번, 또는 두 번 나올 확률의 두 배이므로 막대의 길이가 균일하지 않다.

이 때 수식은

$$nC_r p^k q^{n-k} = nC_r \left(\frac{1}{2}\right)^r \left(\frac{1}{2}\right)^{n-r} = nC_r \left(\frac{1}{2}\right)^n \quad (1)$$

와 같다.

동전을 던지는 횟수가 증가함에 따라, 나타날 수 있는 결과의 종류도 다양해지기 때문에 그래프는 점점 옆으로 퍼진 모양을 하게 된다. 또한 결과의 종류가 많을수록 각각에 해당되는 확률값이 작아지므로 그래프의 키는 점차 작아진다. 시행 횟수를 늘려갈수록 이러한 현상은 더욱 두드러지게 나타나서, 동전을 스물다섯 번 던지는 경우 그래프에서 가장 긴 막대의 길이는 0.155 정도에 불과하다. (이 값은 앞면이 열두 번, 또는 열세 번 나올 확률에 해당한다.)

이 확률을 계산하는 방법은 드무아브르가 활동하던 시대에도 이미 알려져 있었지만, 손으로 직접 계산하기에는 엄청나게 번거로운 작업이었다. 드무아브르는 계산의 편이를 위해 정규분포 곡선으로 둘러싸인 면적을 계산하였는데, 이는 특정 범위 내의 결과가 나타날 확률을(비록 근삿값이긴 하지만) 매우 빠르게 계산할 수 있는 획기적인 방법이었다. 그가 연구하던 것은 측정오차의 확률분포가 아니라 불연속 확률분포였지만, 어쨌거나 그는 이로부터 정규분포 곡선의 공식을 최초로 유도해 낸 수학자가 되었다.

IV. Gauss의 분포 증명

디지털 통신에서 가장 중요한 가우스 분포는 어떻게 증명할 것인가^[1-5]?



〈표 1〉 동전 한 닢을 던질 때의 시행횟수

1 회 시행	2 회 시행	4 회 시행	25 회 시행
$1C_0 \left(\frac{1}{2}\right)^1 = 0.5$ $1C_1 \left(\frac{1}{2}\right)^1 = 0.5$	$2C_0 \left(\frac{1}{2}\right)^2 = 0.25$ $2C_1 \left(\frac{1}{2}\right)^2 = 0.5$ $2C_2 \left(\frac{1}{2}\right)^2 = 0.25$	$4C_0 \left(\frac{1}{2}\right)^4 = 0.06$ $4C_1 \left(\frac{1}{2}\right)^4 = 0.25$ $4C_2 \left(\frac{1}{2}\right)^4 = 0.375$ $4C_3 \left(\frac{1}{2}\right)^4 = 0.25$	$25C_{10} \left(\frac{1}{2}\right)^{25} = 0.1033$ $25C_{15} \left(\frac{1}{2}\right)^{25} = 0.1033$ $25C_{20} \left(\frac{1}{2}\right)^{25} = 0.0064$



N개의 요소로 된 한 계에 대해서 각 요소들이 흥미 대상의 기준을 만족하는 확률 p와 그것을 만족하지 않는 확률 q를 가졌다면 n의 요소들이 그 기준을 만족하고 나머지 (N-n)이 그렇지 않을 확률은

$$P_N(n) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(n-\bar{n})^2}{2\sigma^2}} \quad n-\bar{n} \ll \sigma^2 \text{ (일 때)} \quad (2)$$

(여기서 $\bar{n} = pN, \sigma = \sqrt{Npq}$)

로 나타낼 수 있다. 한편, 지수확률분포와 감마함수로 가우스 분포 증명은 [8]에서 다루었다.

위의 사실을 확인하기 위하여 Gauss 분포, (1)식을 유도한다.

(2)식을 유도해 나가기 위해선 “Taylor 급수 전개” 라고 하는 수학적 방법을 사용하여야 한다. 즉, 어떤 점 $x = x_0$ 에서 어떤 함수 $f(x)$ 와 모든 미분들의 값을 알고 있다면 이 공식을 통해 다른 점 $x = x_1$ 에서의 함수값을 계산할 수 있다.

▶ (참고) 여기서 Taylor 급수 전개를 사용한다.

연속적이고 일가함수인 $f(x)$ 는 다음과 같은 모양의 x 의 다항식으로 표현할 수 있다.

$$f(x) = C_0 + C_1(x-a) + C_2(x-a)^2 + C_3(x-a)^3 + \dots + C_n(x-a)^n + \dots \quad (3)$$

(여기서 a는 임의의 상수, C_0, C_1, C_2, \dots 도 상수)

x 에 관한 식 (3)의 부분을 취하면

$$f'(x) = C_1 + 2C_2(x-a) + 3C_3(x-a)^2 + \dots + nC_n(x-a)^{n-1} + \dots$$

$$f''(x) = 2C_2 + 6C_3(x-a) + \dots + n(n-1)C_n(x-a)^{n-2} + \dots$$

$$f^{(n)}(x) = n!C_n + \frac{(n+1)!}{1!}C_{n+1}(x-a) + \dots$$

이 된다.

점 $x = a$ 에서 이들의 각각을 계산하면

$$f(a) = C_0$$

$$f'(a) = C_1$$

$$f''(a) = 2!C_2$$

$$f^{(n)}(a) = n!C_n$$

임을 알 수 있다.

그러므로 다항식 표현 (3)에서의 모든 계수들은

$$C_n = \frac{1}{n!} f^{(n)}(a)$$

로 주어진다.

이 결과를 다항식 (3)에 대입하면

$$\begin{aligned} f(x) &= f(a) + f'(a)(x-a) + \frac{1}{2!} f''(a)(x-a)^2 + \dots \\ &\quad + \frac{1}{n!} f^{(n)}(a)(x-a)^n + \dots \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(a)(x-a)^n \end{aligned}$$

을 얻는다.

이것을 점 $x = a$ 에 관한 함수 $f(x)$ 의 “Taylor 급수 전개” 이다.

따라서

$$\begin{aligned} f(x_1) &= f(x_0) + \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m f}{\partial x^m} \Big|_{x=x_0} (x_1-x_0)^m \\ &= f(x_0) + \frac{\partial f}{\partial x} \Big|_{x=x_0} (x_1-x_0) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \Big|_{x=x_0} (x_1-x_0)^2 + \dots \quad (4) \end{aligned}$$

이때 Taylor 급수전개를 사용하여 $P(n)$ 에 대한 한 공식을 유도할 때 그 결과의 정확성을 최대로 하기 위해서는 전개식의 처음 몇 개항의 정확성을 증가시켜야한다. 즉, 아래에 열거한 기준을 만족하도록 할 것이다.

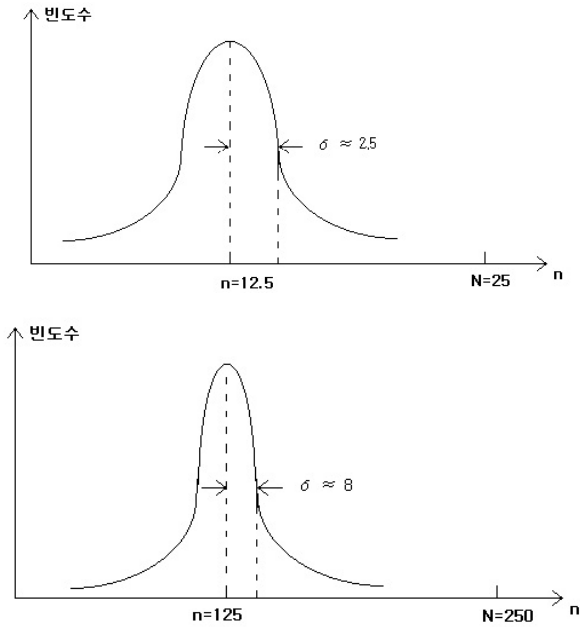
- (1) 가능한 한 매끄러운 함수를 택하여야한다.
- (2) 기준점 n_0 를 가능한 한 우리가 흥미를 갖는 n최대의 값에 가깝도록 택하여야한다.
- (3) 첫 번째 기준을 만족시키기 위해서 $P(n)$ 의 로그를 전개한다. 왜냐하면 함수의 로그가 함수자체보다 천천히 변하기 때문이다.

여기에서 $n_{\text{최대}} = \bar{n}$ 이라 가정하자.

왜냐하면 계(system)가 더욱 커질수록 분포는 더욱 뾰족하게 되고 $n_{\text{최대}}$ 와 \bar{n} 사이의 차이는 더욱 적어지게 되기 때문이다.

그러므로 큰계(N의 큰 값)에 대해서는 위의 근사가 정당화되어진다.

Taylor 급수 공식에 의하여 $P(n)$ 의 로그를 점 $n = \bar{n}$ 주위로 전개하여 보면



〈그림 3〉 가우스 분포와 variance 변화관계

$$\begin{aligned} \ln P(n) &= \ln P(\bar{n}) + \frac{\partial}{\partial n} \ln P(n) \Big|_{n=\bar{n}} (n - \bar{n}) \\ &+ \frac{1}{2} \frac{\partial^2}{\partial n^2} \ln P(n) \Big|_{n=n_0} (n - \bar{n})^2 + \dots \quad (5) \end{aligned}$$

(5)식에서 1차미분은 극대값에서 0이다.

즉,

$$\frac{\partial}{\partial n} \ln P(n) \Big|_{n=\bar{n}} = 0$$

2차미분은 미분의 정의와 이항공식 $P(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n}$ 에 의해 주어진 $P(n)$ 의 값을 이용하여 계산할 수 있다. 즉, $\ln P(n)$ 의 1차미분의 표현은

$$\begin{aligned} \frac{\partial \ln P(n)}{\partial n} &\approx \frac{\Delta \ln P(n)}{\Delta n} = \frac{\ln P(n+1) - \ln P(n)}{(n+1) - n} \\ &= \ln P(n+1) - \ln P(n) \\ &= \ln \frac{N!}{(n+1)!(N-n-1)!} P^{n+1} q^{N-n-1} \\ &\quad - \ln \frac{N!}{n!(N-n)!} p^n q^{N-n} \\ &= \ln \frac{n!(N-n)!}{(n+1)!(N-n-1)!} \frac{P^{n+1} q^{N-n-1}}{P^n q^{N-n}} \\ &= \ln \frac{N-n}{n+1} \frac{p}{q} \quad (6) \end{aligned}$$

이 된다.

1차미분에 대한 이 표현은 다음과 같이 2차미분을 계산하

는데 사용된다.

$$\begin{aligned} \frac{\partial^2 \ln P(n)}{\partial n^2} &\approx \frac{\Delta \frac{\partial \ln P(n)}{\partial n}}{\Delta n} = \frac{\frac{\partial \ln P(n+1)}{\partial n} - \frac{\partial \ln P(n)}{\partial n}}{1} \\ &= \ln \frac{N-n-1}{n+2} \frac{p}{q} - \ln \frac{N-n}{n+1} \frac{p}{q} \\ &= \ln \frac{N-n-1}{N-n} \frac{n+1}{n+2} \\ &= \ln \left(1 - \frac{1}{N-n}\right) + \ln \left(1 - \frac{1}{n+2}\right) \quad (7) \end{aligned}$$

작은 x 에 대한 전개식 $\ln(1-x)$ 를 사용하고 n 에 비해 이를 무시하면, 이것은

$$\begin{aligned} \frac{\partial^2 \ln P(n)}{\partial n^2} \Big|_{n=\bar{n}} &\approx -\frac{1}{N-n} - \frac{1}{n+2} \Big|_{n=\bar{n}} \\ &\approx -\frac{1}{N-\bar{n}} - \frac{1}{\bar{n}} \Big|_{n=\bar{n}} \\ &= -\frac{1}{N-\bar{n}} - \frac{1}{\bar{n}} \\ &= -\left(\frac{1}{N-NP} + \frac{1}{NP}\right) \\ &= -\frac{1}{N} \left(\frac{1}{1-p} + \frac{1}{p}\right) \\ &= -\frac{1}{N} \left(\frac{1}{q} + \frac{1}{p}\right) \\ &= -\frac{1}{Npq} (p+q) \\ &= -\frac{1}{Npq} \quad (8) \end{aligned}$$

따라서,

$$\frac{\partial^2}{\partial n^2} \ln P(n) \Big|_{n=\bar{n}} = -\frac{1}{Npq} = -\frac{1}{\sigma^2} \quad (9)$$

똑같은 방법을 사용하여 3차항, 4차항 등을 구할 수 있다. 하지만,

$$|n - \bar{n}| \ll Npq \text{ 또는 } |n - \bar{n}| \ll \sigma^2$$

에 의해서 3차 및 그 보다 높은 차수의 항들은 전개식에서 처음 3개의 항들에 비해 무시할 수 있음을 알 수 있다.

따라서 (4)식은

$$\ln P(n) = \ln P(\bar{n}) - \frac{1}{2\sigma^2} (n - \bar{n})^2$$

또는

$$P(n) = P(\bar{n})e^{-\frac{(n-\bar{n})^2}{2\sigma^2}}$$

으로 쓸 수 있다. 여기에서 n의 모든 가능한 값의 확률의 합은 1과 같아야 한다.

즉,

$$\begin{aligned} \sum_n P(n) &= 1 \\ &= \sum_n P(n) \Delta n (\because \Delta n = 1) \\ &\approx \int_{-\infty}^{\infty} P(n) dn \\ &= P(\bar{n}) \int_{-\infty}^{\infty} e^{-\frac{(n-\bar{n})^2}{2\sigma^2}} dn \\ &= P(\bar{n}) (\sqrt{2\pi} \sigma) \end{aligned}$$

그러므로

$$P(\bar{n}) = \frac{1}{\sqrt{2\pi} \sigma}$$

이 되고, 결국

$$P(n) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(n-\bar{n})^2}{2\sigma^2}} \quad (10)$$

이 되어 (2)식, 즉, 가우스 분포를 유도할 수가 있다. ■
한편, 가우스 분포와 엔트로피(Entropy) 관계를 부록에서 다루었다. [6-9]에서는 가우스와 같은 수준의 대학자들인 푸슈킨, 마코브, 체비쇼프, 푸리에, 아인슈타인, 구글의 래리 페이지 등의 이론을 간단히 비교분석하였다.

V. 결론

19세기 들어, 생명과학과 물리학에서 발견된 자연 현상들 중 상당수가 무작위오차의 정규분포법칙을 따른다는 사실이 알려졌다. 별의 위치¹⁾를 측정할 때 나타나는 오차가 무작위

1) 별의 위치 : 서울 밤하늘의 별은 잘 보이지 않는다. 전주 밤하늘의 별은 그런대로 보이는데, 제주도 한라산 밤하늘의 별은 반짝거린다. 히말라야 산자락 마을의 밤하늘별은 왕별. 손만 닿으면 곧 따낼 듯하다.

로 나타난다면, 극미의 세계에서 분자의 위치를 측정할 때 수반되는 오차값도 역시 무작위로 발생할 것이다. 별기에 과학자 람베르트 케틀러(Lambert Quetelet)는 사람의 키와 몸무게, 근력 등에서 나타나는 개인차가 평균값을 중심으로 정규분포 곡선을 보인다는 사실을 입증하였는데, 그 이후로 생물학자들은 케틀러의 주장을 정설로 받아들이고 있다.

물론, 뉴턴의 결정론적 세계관을 아직 고수하고 있는 학자들도 있다. 고대 그리스의 합리적 철학에 그 뿌리를 두고 있는 기계적 결정론은 정확성과 질서, 법칙, 그리고 필연성의 논리로 이 우주를 서술하고 있으며, 거기에는 ‘우연’이 끼어들 자리가 없다. 그러나 19세기에 확고한 기반을 다져놓은 통계이론, 특히 벨의 곡선은 뉴턴의 물리학과 더불어 범우주적 법칙으로 받아들여지고 있다. 무선통신 신호처리에서는 가우스 잡음을 기준으로 레일레이, 라이시안, 나카가미 등 확장된 확률분포가 실용화에 약방에 감초 격으로 응용되고 있다. 가우스의 직관에 의해서 동전(coin)한 낚을 던졌을 때를 확률적으로 해석한 이론, 1과 100의 합은 101이고 2와 99의 합은 101이며, 이것이 50개가 있다. 즉 진폭이 1, 대역폭이 1, 평균치가 0.5인 Uniform이 50개가 있는데, 이를 50개 중첩시킨 가우스 분포 꼴이 된다. 그 얼마나 아름다우면서도 간단한 결과인가?

감사의 말씀

본 연구는 한국연구재단(NRF)의 ‘세계 수준의 연구 중심 대학 (WCU) 육성 사업 (R32-2011-000-20014-0)’과 ‘기초 연구 사업 (FR 2010-0020942)’의 지원을 받아 수행되었음.

부 록

$$\binom{n}{k} = \frac{1}{\sqrt{2\pi npq}} \exp \frac{(k-pn)^2}{2pqn} \quad (11)$$

Lemma 1 ([10]). Let λ_n be a integer and $H(x) = -x \log_2 x - (1-x) \log_2 (1-x)$ be the entropy function and $0 < \lambda < 1$.

Then

$$\frac{1}{\sqrt{8n\lambda(1-\lambda)}} 2^{nH_2(\lambda)} \leq \binom{n}{\lambda n} \leq \frac{1}{\sqrt{2\pi n\lambda(1-\lambda)}} 2^{nH_2(\lambda)} \quad (12)$$

Proof. The Stirling formula:

$$\sqrt{2\pi n} \frac{n+\frac{1}{2}e^{-n}+\frac{1}{12n}-\frac{1}{360n^3}}{n!} < n! < \sqrt{2\pi n} \frac{n+\frac{1}{2}e^{-n}+\frac{1}{12n}}{n!} \quad (13)$$

therefore,

$$\binom{n}{\lambda n} = \frac{n!}{\lambda n!(n-\lambda n)!} = \frac{1}{\sqrt{2\pi n\lambda(1-\lambda)}} 2^{nH_2(\lambda)} \quad (14)$$

Corollary 1. In condition of lemma 1

$$\binom{n}{[\lambda n]} = \frac{1}{\sqrt{2\pi n\lambda(1-\lambda)}} 2^{nH_2(\lambda)(1+\varepsilon_n)}, n \rightarrow \infty, \varepsilon_n \rightarrow 0. \quad (15)$$

Obviously,

$$p^k q^{n-k} = 2^{k \log_2 p + (n-k) \log_2(1-p)} \quad (16)$$

If $\frac{k}{n} = \lambda = p$, then from (14)

$$p^k q^{n-k} \binom{n}{k} = \frac{1}{\sqrt{2\pi n\lambda(1-\lambda)}} 2^{n(H_2(\lambda) + \frac{k}{n} \log_2 p + \frac{n-k}{n} \log_2(1-p))} \quad (17)$$

If $\frac{k}{n} = p + \theta_n$, $\theta_n \rightarrow 0$, $n \rightarrow \infty$, then

$$H_2(\lambda) + \frac{k}{n} \log_2 p + \frac{n-k}{n} \log_2(1-p) = \frac{\theta_n^2}{2npq} \quad (18)$$

What to prove it is necessary decomposed the function $\log x$ to Taylor series.

참고문헌

[1] A.Papoulis, Probability, Random Variables and Stochastics, Papoulis Pilois, 20인.
 [2] 이문호, 실용디지털통신, 영일 2003.
 [3] http://en.wikipedia.org/Carl_Fredrich_Gauss
 [4] J.G.Proakis, Digital Communications, Mc Grawill, 2001.
 [5] B.Sklar, Digital Communications, PHPTR, 2001.

[6] 이문호, "푸슈킨-체비쇼프-마코브-콜모고로프-펠레만, 러시아 상트페테르부르크 : 2011 IEEE 국제정보이론 심포지엄", 한국통신학회지 정보와 통신, Vol.28, No.9, 2011.
 [7] 이문호, "푸리에(1822) 그 후", 한국통신학회지 정보와 통신, Vol.28, No.10, 2011.
 [8] 이문호, "샤논(1948)을 넘어", 한국통신학회지 정보와 통신, Vol.28, No.11, 2011.
 [9] 이문호, "구글 : PageRank", 한국정보과학회 정보과학회지, Vol.29, No.11, 2011. 10
 [10] MacWilliams F.W. and Sloane N.W.A., The theory of error-correcting codes, North-Holland, Amsterdam, 1977.



이 문 호

1984년 2월 전남대 전기공학과 (공학박사), 통신기술사.
 1985년~1986년 미국 University of Minnesota (Post Doc.).
 1990년 일본 동경대학 정보통신 (공학박사).
 1970년~1980년 남양 MBC 송신소장.
 1980년 10월~2010년 2월 전북대학교 전자정보공학부 교수.
 2010년 2월~현재 전북대 WCU-II 연구책임 교수.
 <관심분야> 정보통신의 원형, 뿌리찾는 연구