담당과목

- Lecturer: 김태현 (taehyun@snu.ac.kr, 사무실: 301동 407호)

- 선형대수: 2주차 1/13(월) ~ 1/17(금) 오후 세션
  - 교재: Introduction to Linear Algebra, International edition Gilbert Strang (2019)

- 최적화: 3주차 1/20(월)~1/22(수) 오후, 1/29(수) 오후, 1/30(목) 오전 세션
  - 교재: Linear Algebra and Learning from Data, Gilbert Strang (2019)

Mean, Variance, and Probability

- Mean m: 평균값 또는 기대값
- Variance (분산) σ^2: m으로부터의 거리의 제곱의 평균
- Probability (확률): 결과 또는 측정값 x가 n개의 다른 값 $x_1, x_2, \ldots x_n$을 가질 수 있을 때, $p_1, p_2, \ldots, p_n$은 각각의 값이 나올 확률을 표기하며 $\sum_{i=1}^{n} p_i = 1$을 만족함

Mean, Variance, and Probability

- 2가지 다른 경우의 Mean
  - Example) 대학 1학년생들의 나이
  - Sample values: 18, 17, 18, 19, 17
  - Sample mean: $\frac{1}{5}(18+17+18+19+17) = 17.8$
  - 확률분포: 17 (20%), 18 (50%), 19 (30%)
  - Expectation value (기대값): $E[x] = 17 \cdot 0.2 + 18 \cdot 0.5 + 19 \cdot 0.3 = 18.1$

- 표본의 평균: $m = \mu = \frac{1}{N}(x_1 + x_2 + \cdots + x_N)$
- 기대값: $m = E[x] = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$
- Law of Large Numbers (LNN)
  - 표본의 크기 N이 커짐에 따라 표본의 평균값 μ 은 확률 분포의 기대값 $E[x]$ 에 수렴한다.
  - 이것은 동전의 뒷면이 앞면보다 많이 나왔다고 해서 다음 표본이 앞면이 나올 가능성이 높다는 것을 의미하지는 않음.

Mean, Variance, and Probability

- Variance (분산)
  - Variance σ^2: 이론적으로 기대되는 분산
  - Sample variance S^2: 실제 표본들이 표본의 평균으로부터 떨어진 거리의 제곱의 평균

$$S^2 = \frac{1}{N-1}[(x_1 - m)^2 + \cdots + (x_N - m)^2]$$

    - Example: sample= 18, 17, 18, 19, 17, mean $m = 17.8$

$$S^2 = \frac{1}{4}[(.2)^2 + (-.8)^2 + (.2)^2 + (1.2)^2 + (-.8)^2] = \frac{1}{4}(2.8) = 0.7$$

    - $\sum_{i=1}^{N}(x_i - m)^2 = \sum_{i=1}^{N}\left(x_i^2 - 2mx_i + m^2\right) = \sum_{i=1}^{N} x_i^2 - 2m\sum_{i=1}^{N} x_i + Nm^2$ $= \sum_{i=1}^{N} x_i^2 - 2m(Nm) + Nm^2 = \sum_{i=1}^{N} x_i^2 - Nm^2$
  - $\sigma^2 = E[(x - m)^2] = p_1(x_1 - m)^2 + \cdots + p_n(x_n - m)^2$
    - Example: 확률분포: 17 (20%), 18 (50%), 19 (30%), $m = E[x] = 18.1$
    - $\sigma^2 = 0.2(-1.1)^2 + 0.5(-0.1)^2 + 0.3(0.9)^2 = 0.2 \cdot 1.21 + 0.5 \cdot 0.01 + 0.3 \cdot 0.81 = 0.49$
    - $\sigma^2 = E[x^2] - (E[x])^2 = \sum_{i=1}^{n} p_i \cdot x_i^2 - (\sum_{i=1}^{n} p_i \cdot x_i)^2$

Continuous Probability Distribution

- 이산확률분포
  - Example: 나이 분포 17, 18, 19
- 연속확률분포
  - Example: 나이를 년단위 대신 일단위 또는 그 이하의 단위로 측정시, 확률변수 x는 17~20 사이의 거의 연속적인 값을 가질 수 있음
  - 확률밀도함수 (pdf, probability density function): $p(x) = \frac{dF}{dx}$
    - 확률밀도함수는 그 지점에서의 확률이 아니라 구간의 적분이 확률에 해당하므로 특정한 위치의 값은 1보다 큰 값을 가질 수 있음
  - 누적분포함수 (cumulative distribution function) $F(x)$
    - $a \leq x \leq b$일 확률: $\int_a^b p(x)dx = F(b) - F(a)$
- 균등분포 (uniform distribution)
  - Example: x가 17~20 사이의 실수를 가질 확률이 균등분포를 이룰 경우: $p(x) = \begin{cases} \frac{1}{3}, 17 \leq x \leq 20 \\ 0, \text{otherwise} \end{cases}$

Mean and Variance of $p(x)$

- 연속확률분포
  - Mean: $m = E[x] = \int xp(x)dx$
  - Variance: $\sigma^2 = E[(x-m)^2] = \int p(x)(x-m)^2 dx$

- 균등분포 (uniform distribution)
  - $0 \le x \le a$ 구간의 균등분포인 경우
  - pdf: $p(x) = \frac{1}{a}$
  - 누적분포: $F(x) = \frac{x}{a}$
  - Mean $m = \int_0^a \frac{1}{a} x dx = \frac{a}{2}$
  - Variance: $\sigma^2 = \int_0^a \frac{1}{a}\left(x - \frac{a}{2}\right)^2 dx = \frac{a^2}{12}$

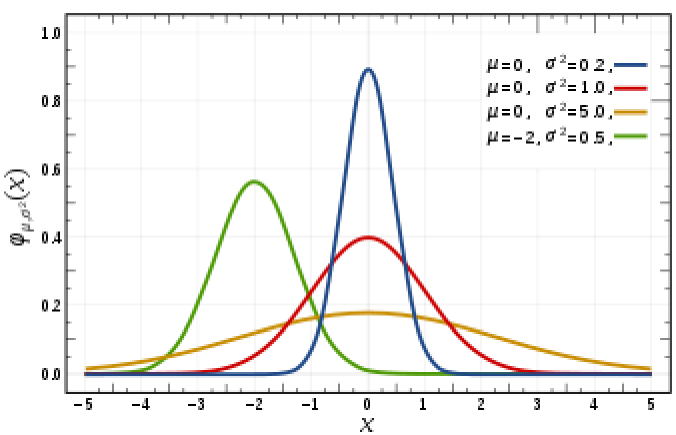
정규분포 (Normal Distribution)

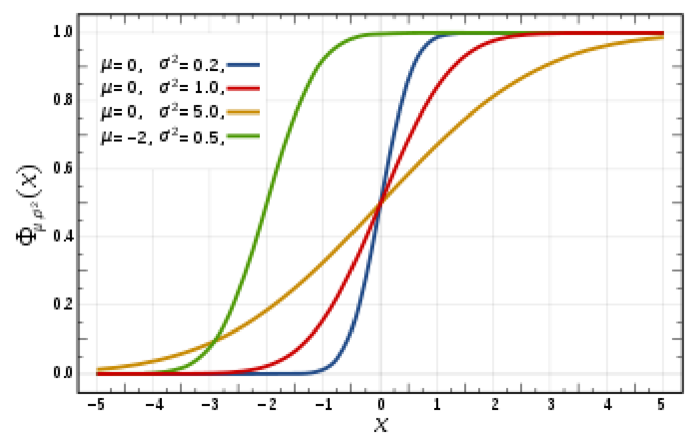
- 정규분포 (normal distribution)
  - Gaussian distribution이라고도 불림
  - 동전 던지기와 같이 개별적인 실험은 정규분포와 상관없는 확률분포를 가지더라도, N개 샘플의 평균값은 정규분포에 가까워지게 됨.
  - Central Limit Theorem: 샘플의 개수 N이 무한대로 갈수록 N개 샘플의 평균값은 정규분포로 수렴하게 됨

- 표준정규분포 (standard normal distribution)
  - 평균이 0, 분산이 1인 정규분포: $\mathbf{N}(0,1)$
  - $p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

정규분포 (Normal Distribution)

- 일반적인 정규분포
  - 평균 m, 분산 σ^2 ➔ $\mathbf{N}(m, \sigma)$
  - $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}$
  - 샘플이 $\pm\sigma$ 이내에 들어올 확률: 약 68.3%
  - 샘플이 $\pm 2\sigma$ 이내에 들어올 확률: 약 95.4%
  - 샘플이 $\pm 3\sigma$ 이내에 들어올 확률: 약 99.7%





From https://en.wikipedia.org/wiki/Normal_distribution

기대값과 분산의 성질

- x_1과 x_2가 독립적인 2개의 확률 변수를 나타낸다고 가정
  - $E[cx_1 + b] = cE[x_1] + b$
  - $Var[cx_1 + b] = c^2 Var[x_1]$
  - $E[x_1 + x_2] = E[x_1] + E[x_2]$
  - $Var[x_1 + x_2] = Var[x_1] + Var[x_2]$
- $x_1, \dots, x_n$가 평균 m, 분산 σ^2를 가지는 모집단의 샘플을 나타내는 경우, $\bar{x}$ 는 이들의 평균값 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$을 나타낸다고 가정
  - $E[\bar{x}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{n \cdot m}{n} = m$
  - $Var[\bar{x}] = Var\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var[x_i] = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$

N Coin Flips and $N \to \infty$

- Example
  - x는 1 또는 -1을 $p_1 = p_{-1} = \frac{1}{2}$의 확률로 가짐
  - $m = \frac{1}{2}(+1) + \frac{1}{2}(-1) = 0$, $\sigma^2 = \frac{1}{2}(1)^2 + \frac{1}{2}(-1)^2 = 1$
  - $A_N = (x_1 + x_2 + \cdots + x_N)/N$ 의 분포?
  - $m_N = E[A_N] = \frac{1}{N}(NE[x]) = 0$
  - $\sigma_N^2 = Var[A_N] = \frac{1}{N^2}(Var[x_1] + \cdots Var[x_N]) = \frac{Var[x]}{N} = \frac{1}{N}$
- 이항분포 (binomial distribution)
  - 시행횟수 N이 커질수록 정규분포로 수렴

확률 분포의 종류

- Binomial (이항분포): 동전을 n번 던질 경우
- Poisson: rare event
- Exponential: forgetting the past
- Gaussian (normal): 많은 시행에 대한 평균
- Log-normal: Logarithm has normal distribution
- Chi-squared: distance squared in n dimensions
- Multivariable Gaussian: vector에 대한 확률

Binomial Distribution

- 개별 시행에 대한 결과값이 1 또는 0
- 성공확률 $p_{1,1} = p$, 실패확률 $p_{0,1} = 1 - p = q$
- n번 시행 중 k번 성공할 확률: $p_{k,n} = \binom{n}{k} p^k (1-p)^{n-k}$
  - $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ n개의 샘플 중 순서 상관없이 k개만 뽑을 때 경우의 수
- 1번 시행시 평균: $\mu = E[x] = p$
- n번 시행시 평균: $\mu_n = E[x_1 + \cdots + x_n] = n\mu = np$
- 1번 시행시 분산: $\sigma^2 = E[(x-\mu)^2] = (1-p)(0-\mu)^2 + p(1-\mu)^2 = p(1-p)$
- n번 시행시 분산: $\sigma_n^2 = Var[x_1 + \cdots + x_n] = nVar[x] = np(1-p)$

포아송 분포

- 포아송 (Poisson) 분포
  - 많은 시행 중에 매우 드물게 발생하는 경우
  - 개별 성공확률이 매우 낮은 경우 ($p \to 0$), n번 반복하여 ($n \to \infty$)평균 λ번 성공하는 이항분포 관점에서 해석 가능 ➔ $\lambda = np$
  - n번 시행 중 k번 성공할 확률 $p_{k,n} = \binom{n}{k} p^k (1-p)^{n-k}$
    - n번 실패, 0번 성공할 확률: $p_{0,n} = (1-p)^n = \left(1 - \frac{\lambda}{n}\right)^n \xrightarrow{n\to\infty} e^{-\lambda}$
    - n번 실패, 1번 성공할 확률: $p_{1,n} = np(1-p)^{n-1} = \frac{\lambda}{1-p}\left(1 - \frac{\lambda}{n}\right)^n \xrightarrow{n\to\infty} \lambda e^{-\lambda}$
    - n번 실패, 2번 성공할 확률: $p_{2,n} = \frac{n(n-1)}{2!} p^2 (1-p)^{n-2} = \frac{(\lambda^2 - \lambda p)}{2(1-p)^2}\left(1 - \frac{\lambda}{n}\right)^n \xrightarrow{n\to\infty} \frac{1}{2}\lambda^2 e^{-\lambda}$
  - 포아송 확률은 $p_{k,n} = \frac{\lambda^k}{k!} e^{-\lambda}$에 가까워지게 됨.
  - 포아송 확률의 합은 당연히 1
  - 평균은 이항분포를 따르므로 $np = \lambda$
  - 분산도 이항분포를 따라서 $np(1-p) \approx \lambda$

Markov's Inequality

- Markov 부등식
  - 확률변수 x가 취할 수 있는 값이 0보다 크거나 같은 경우에만 해당함 $(x \geq 0)$
  - 평균값이 $E[x]$일 때, x가 a보다 클 확률 $\Pr(x \geq a)$은 $E[x]/a$ 보다 작거나 같음
  - $\Pr(x \geq a) \leq \frac{E[x]}{a}$

Chebyshev's Inequality

- Chebyshev 부등식
  - 확률변수 x의 평균값 $E[x] = m$ 와 분산σ^2이 알려져 있을 때, $|x - m|$가 a보다 클 확률 $\Pr(|x - m| \geq a)$ 은 σ^2/a^2 보다 작거나 같음
  - $\Pr(|x - m| \geq a) \leq \sigma$^2/$a^2$
  - 증명
    - 확률변수 x 와 같은 확률분포를 따르는 새로운 확률변수 $y = (x - m)^2$를 정의한 후, Markov 부등식을 적용