25-1

Machine Learning Programming

3주차

소프트웨어학부 최종환 교수님

한림대학교

학습 목표

- 확률과 랜덤 변수
- 기댓값과 분산
- 랜덤벡터
- 가우시안 분포
- 엔트로피
- 최대 우도 확률 추정

확률이란?

- 경험적 확률
  - 오늘 밤에 무릎이 아프니 내일 비 올 가능성은 70%이다.
- 통계적 확률
  - 5달 동안 비가 내린 날이 10, 15, 9, 3, 5일이었으므로, 이번 달에는 8.4일 비가 내릴 것이다.
- 고전적인 수학적 확률
  - 교수님의 생일을 맞출 확률은 1/365 이다.
- 공리에 바탕을 둔 확률

확률의 정의

감기 때문에 너무 아파서 뭐라는지 집중이 하나도 안됨...ㅜ
너무 아파 뒤지겠어 ㅆㅂ 거의 코로나 급이야 설마... 진짜 코로나는 아니겠지?

- 무작위 실험(random experiment)
  - 동일한 조건 아래에서 여러 번 반복할 수 있고,
  - 그 결과(outcome)들을 관찰할 수 있으며,
  - 관찰 가능한 모든 결과들을 알 수 있는 실험

- 표본 공간(sample space)
  - Random experiment 를 통해 관찰될 수 있는 모든 결과들을 원소(element)로 갖는 집합

- 사건(event)
  - Sample space 의 부분집합

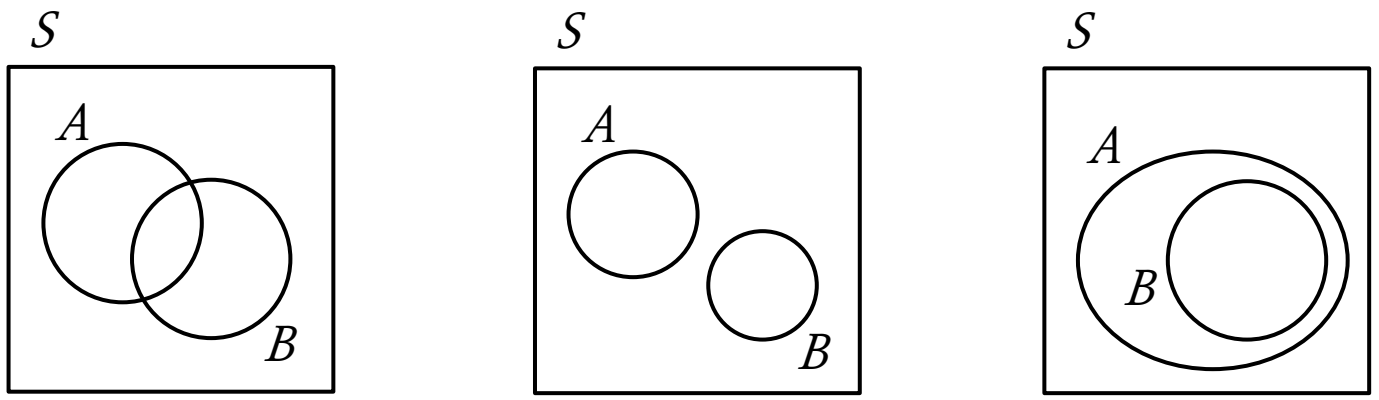
확률의 정의

- 이산 표본 공간(discrete sample space)
  - Finite 또는 countably infinite 표본공간
    셀 수 있는 무한
- 연속 표본 공간(continuous sample space)
  - Uncountably infinite 표본공간
    셀 수 없는 무한
- 예제
  - 동전 1회 던지기의 표본 공간은 이산적이다 (O/X)
  - 주사위 1회 굴리기의 표본 공간은 이산적이다 (O/X)
  - 동전 앞면 나올 때까지 던지기의 표본 공간은 연속적이다 (O/X)
  - 어떤 사람의 키를 측정하는 것의 표본 공간은 연속적이다 (O/X)

확률의 정의

- 상호 배타적(mutually exclusive)
  - 두 개의 event 가 서로 공통인 원소를 갖지 않는 경우

- 예제
  - 표본공간 S 에서 사건 A 와 B 가 상호 배타적인 것은?



확률의 정의

- 멱집합(power set)
  - 표본 공간 S 의 모든 부분집합을 원소로 갖는 집합
  - $\mathcal{P}(S) \coloneqq \{U | U \subseteq S\}$

- 확률(probability)
  - 아래의 3가지 공리(axiom)에 따라 표본 공간 S 의 모든 사건들을 실수(real number)에 대응시키는 함수(function). 즉, $\mathrm{P}: \mathcal{P}(S) \to \mathbb{R}$
    ① $\mathrm{P}\{\mathrm{A}\} \geq 0$
    ② $\mathrm{P}\{S\} = 1$
    ③ $\mathrm{P}\{\mathrm{A} \cup B\} = \mathrm{P}\{\mathrm{A}\} + \mathrm{P}\{\mathrm{B}\}$ 이 세개가 공리임.

확률의 정의

- 확률의 특징
  - $P\{\emptyset\} = 0$
  - $P\{A\} = 1 - P\{A^C\}$

확률 변수

- 확률 변수(random variable)
  - 표본 공간을 구성하는 각 원소를 하나의 실수에 대응시키는 함수
  - 확률 실험의 결과가 숫자가 아니더라도 실험 결과를 수치화 할 수 있음

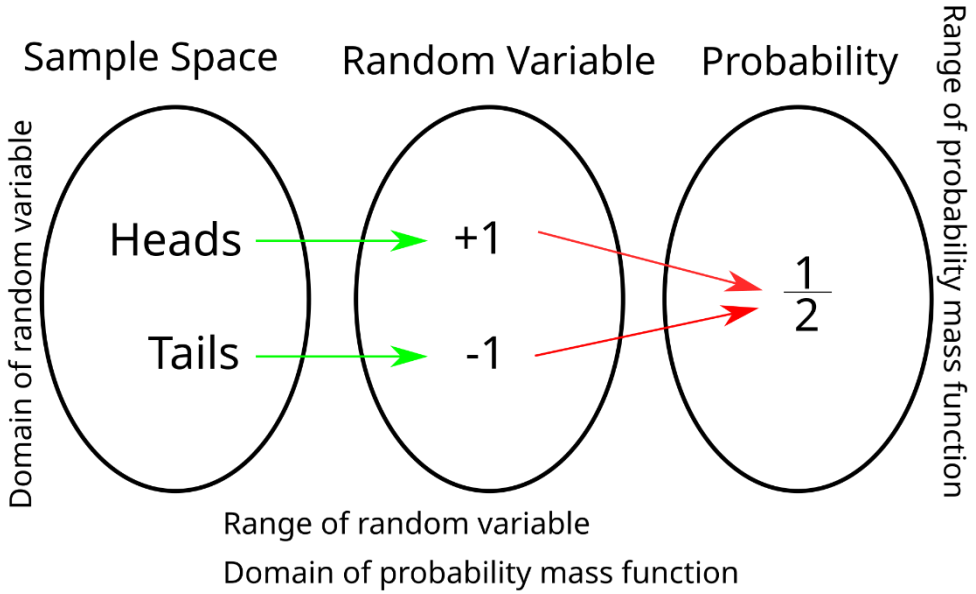


그림 출처: https://en.wikipedia.org/wiki/Random_variable

누적 분포 함수

- 누적 분포 함수(cumulative distribution function, CDF)
  - 랜덤 변수 X 가 특정값 x 보다 작은 값을 가질 확률
  - $F_X(x) \coloneqq P\{X \leq x\}$

- 누적 분포 함수의 특징
  - 단조 증가 함수(monotonically non-decreasing function)
    - ✓ $a > b \Rightarrow F_X(a) \geq F_X(b)$
  - $\lim_{x \to \infty} F_X(x) = 1$
  - $\lim_{x \to -\infty} F_X(x) = 0$

확률 밀도 함수

- 확률 밀도 함수(probability density function, PDF)
  - 다음의 조건을 만족하는 적분 가능한 함수
  - $\int_{-\infty}^{x} p_X(x)dx = F_X(x)$

- CDF 가 미분가능한 함수인 경우
  - $p_X(x) = \frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P\{x < X \leq x+\Delta x\}}{\Delta x}$

- PDF 를 이용한 확률 표현
  - $P\{a < X \leq b\} = \int_a^b p_X(x)dx$

확률 밀도 함수

- 확률 밀도 함수의 특징
  - $p_X(x) \geq 0$
  - $\int_{-\infty}^{\infty} p_X(x)dx = 1$

결합 확률 함수

- 결합 누적 분포 함수(joint cumulative distribution function)
  - 두 확률 변수 X 와 Y 의 결합 사건(joint event)의 확률
  - $F_{XY}(x,y) \coloneqq P\{(X \le x) \cap (Y \le y)\} = P\{X \le x, Y \le y\}$

- 결합 누적 분포 함수의 특징
  - $0 \le F_{XY}(x,y) \le 1$
  - $F_{XY}(-\infty,\infty) = 1$
  - $F_{XY}(-\infty,y) = F_{XY}(x,-\infty) = 0$
  - $F_{XY}(x,\infty) = F_X(x)$
  - $F_{XY}(\infty,y) = F_Y(y)$

결합 확률 함수

- 결합 확률 밀도 함수(joint probability density function)
  - 결합 누적 분포 함수로부터 다음과 같이 정의
  - $F_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} p_{XY}(u,v)du\,dv$

- 결합 누적 분포 함수가 미분가능한 경우
  - $p_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y} = \lim_{\Delta x,\Delta y \to 0} \frac{P\{x<X\leq x+\Delta x, y<Y\leq y+\Delta y\}}{\Delta x \Delta y}$

- 결합 확률 밀도 함수의 특징
  - $p_{XY}(x,y) \geq 0$
  - $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{XY}(u,v)du\,dv = 1$

주변 분포

- 주변 확률 밀도 함수(marginal probability density function)
  - 확률 변수 X 와 Y 의 결합 확률 밀도 함수로부터 계산되는 확률 밀도 함수
  - $p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x,y)dy$
  - $p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x,y)dx$

- 주변 누적 분포 함수(marginal cumulative distribution function)
  - $F_X(x) = \lim_{y\to\infty} F_{XY}(x,y)$
  - $F_Y(y) = \lim_{x\to\infty} F_{XY}(x,y)$

조건부 확률 함수

- 조건부 확률(conditional probability)
  - 사건 B 가 관찰되었을 때, 사건 A 가 관찰될 확률
  - $P\{A|B\} \coloneqq \frac{P\{A,B\}}{P\{B\}}$
- 조건부 확률을 이용한 결합 확률 표현
  - $P\{A,B\} = P\{A|B\}P\{B\}$
- 조건부 확률의 확장
  - 사건 B 와 C 가 관찰되었을 때, 사건 A 가 관찰될 확률
  - $P\{A|B,C\} = \frac{P\{A,B,C\}}{P\{B,C\}} = \frac{P\{A,B|C\}P\{C\}}{P\{B|C\}P\{C\}} = \frac{P\{A,B|C\}}{P\{B|C\}}$

확률의 연쇄 법칙

- 확률의 연쇄 법칙(chain rule)

$$P\{A_1, A_2, \ldots, A_N\} = P\{A_N\} \prod_{i=1}^{N-1} P\{A_i | A_{i+1:N}\}$$

독립

- 사건의 독립(independent)
  - 두 사건 A 와 B 에 대하여 다음이 성립하는 경우
  - $P\{A,B\} = P\{A\}P\{B\}$
- 독립과 결합 확률 함수
  - $F_{XY}(x,y) = F_X(x)F_Y(y)$
  - $p_{XY}(x,y) = p_X(x)p_Y(y)$
- 독립과 조건부 확률 함수
  - $P\{A|B\} = P\{A\}$
  - $P\{B|A\} = P\{B\}$

베이즈 정리

- 사건 B_i (i=1,⋯,n) 이 상호 배타적이고, 그 합집합이 표본 공간과 동일한 경우를 가정하자. 즉,
  - $i \neq j \Rightarrow P\{B_i, B_j\} = 0$
  - $S = \cup_{i=1}^{n} B_i$

- 이 때, 임의의 사건 A 의 확률은 다음과 같이 표현하고, 이러한 표현을 전확률(total probability)라고 한다.

$$P\{A\} = \sum_{i=1}^{n} P\{A, B_i\} = \sum_{i=1}^{n} P\{A|B_i\}P\{B_i\}$$

베이즈 정리

- 사건 A 를 조건으로 하는 임의의 사건 B_i 의 조건부 확률

상수

$$P\{B_i|A\} = \frac{P\{A|B_i\}P\{B_i\}}{\sum_{i=1}^{n} P\{A|B_i\}P\{B_i\}}$$

상수

P(M|D): 베이즈 정리 식(?)

샘플링

- 샘플링(sampling)
  - 확률 분포를 이용하여 결과를 수집(=데이터를 생성)하는 과정
  - 샘플링을 통해 얻어진 데이터를 샘플(sample)이라고도 부름

- 샘플 표기법
  - $x \sim p_X(x)$

- Independent and Identical Distributed, iid
  - 여러 개의 샘플이 동일한 확률 분포에서 독립적으로 추출된 경우

기대값

- 기댓값(expectation)
  - 확률 변수 X의 기댓값은 다음과 같이 정의된다.

P(M|D): 베이즈 정리 식(?)ㅅ

$$\mathbb{E}[X] \coloneqq \int_{-\infty}^{\infty} x p_X(x) dx$$

  - 확률 변수 X에 대하여 새로운 확률 변수 Y=g(X) 가 정의되는 경우, 확률 변수 Y 의 기댓값은 다음과 같다.

$$\mathbb{E}[Y] \coloneqq \int_{-\infty}^{\infty} g(x) p_X(x) dx$$

기댓값

- 기댓값의 특징
  - 확률 변수가 상수(constant)인 경우, $\mathbb{E}[X] = \mathbb{E}[c] = c$
  - $\mathbb{E}[ag(X) + bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)]$

공분산

- 분산(variance)

$$Var(X) \coloneqq \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

- 표준편차(standard deviation)

$$\sigma_X \coloneqq \sqrt{Var(X)}$$

- 공분산(covariance)

$$Cov(X, Y) \coloneqq \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

상관계수

- 상관계수(correlation coefficient)
  - 두 확률 변수의 선형 (반)비례 정도를 나타내는 수치

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

- 상관계수의 특징
  - 두 확률 변수가 독립이면 $\rho_{XY} = 0$
  - $\rho_{XY} = 0$ 이 두 확률 변수의 독립을 보장하지는 않음

상관계수

- (예제) $\rho_{XY} = 0$ 이 두 확률 변수의 독립을 보장하지는 않음
- 두 확률 변수 X, Y가 다음과 같이 정의되었다고 하자.
  - $p_X(x) = \frac{1}{3}$ for all $x \in \{-1,0,1\}$
  - $Y \coloneqq X^2$
- 그러면, 두 확률 변수의 상관계수는 0임
  - $\mathbb{E}[X] = 0$
  - $\mathbb{E}[Y] = \frac{2}{3}$
  - $\mathbb{E}[XY] = \mathbb{E}[X^3] = 0$
  - $Cov(X,Y) = 0$
- 하지만, 두 확률 변수는 독립은 아님
  - $p_{XY}(-1,1) = \frac{1}{3}$
  - $p_X(-1)p_Y(1) = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$

조건부 기댓값

- 조건부 기댓값(conditional expectation)

$$\mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x p_{X|Y}(x|y) dx$$

- 조건부 분산(conditional variance)

$$Var(X|Y=y) = \mathbb{E}[(X - \mathbb{E}[X|Y=y])^2|Y=y]$$
$$= \mathbb{E}[X^2|Y=y] - (\mathbb{E}[X|Y=y])^2$$

랜덤벡터

- 확률 변수를 원소로 갖는 벡터(vector)

- 랜덤벡터의 CDF

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = P\{X_1 \le x_1,\dots,X_n \le x_n\}$$

- 랜덤벡터의 PDF

$$F_X(x) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} p_{X_1,\dots,X_n}(u_1,\dots,u_n)du_n\dots du_2\,du_1$$

랜덤벡터

- 랜덤벡터의 주변 확률 밀도 함수
- 예를 들어, 랜덤벡터 $X = [X_1, X_2, X_3]^T$ 인 경우:

$$p_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{\infty} p_X(x) dx_3$$

$$p_{X_1}(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_X(x) dx_3 \, dx_2$$

랜덤벡터

- 랜덤벡터의 조건부 누적 분포 함수

$$P\{X \leq x | Y = y\} = \int_{-\infty}^{x} p_{X|Y}(u|y)du$$

- 랜덤벡터의 조건부 확률 밀도 함수

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_{X_1,\ldots,X_n,Y_1,\ldots,Y_m}(x_1,\ldots,x_n,y_1,\ldots,y_m)}{p_{Y_1,\ldots,Y_m}(y_1,\ldots,y_m)}$$

랜덤벡터

- 확률 연쇄 법칙을 이용한 랜덤벡터의 PDF 전개

$$p_{X_1,\dots,X_n}(x_1,\dots,x_n) = p_{X_1,\dots,X_n}(x_1,\dots,x_{n-1}|x_n)p_{X_n}(x_n)$$

$$= p_{X_n}(x_n)\prod_{i=1}^{n-1} p_{X_1,\dots,X_n}(x_1,\dots,x_i|x_{i+1})$$

- 랜덤벡터의 확률 변수들이 독립인 경우

$$p_{X_1,\dots,X_n}(x_1,\dots,x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$$

랜덤벡터

- 랜덤벡터의 기댓값 벡터

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix} = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])^T$$

- 랜덤벡터의 공분산 행렬

$$Cov(X) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

$$\sigma_{ij} = \sigma_{ji} = Cov(X_i, X_j) = \mathbb{E}\big[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\big]$$

표본 평균

- 어떤 랜덤벡터에 대해 N개의 iid 샘플이 주어진다면, 확률 밀도 함수는 다음과 같이 근사화(approximation)될 수 있음

$$p_X(x) \approx \frac{1}{N}\sum_{i=1}^{N}\delta(x-x_i), \qquad \delta(x-a)=\begin{cases}\infty, & x=a\\ 0, & x\neq a\end{cases}$$

- 랜덤벡터의 기댓값은 표본평균으로 근사화될 수 있음

$$\mathbb{E}[X]=\int_{-\infty}^{\infty}xp_X(x)dx \approx \int_{-\infty}^{\infty}x\frac{1}{N}\sum_{i=1}^{N}\delta(x-x_i)\,dx$$

$$=\frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}x\delta(x-x_i)dx=\frac{1}{N}\sum_{i=1}^{N}x_i$$

가우시안 분포

- 기댓값과 분산만으로 PDF가 정의되는 확률 분포

단변량

$$p_X(x) = N(x|\mu, \sigma^2) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- 랜덤 벡터에 대한 가우시안 PDF

$$p_X(x) = N(x|\mu, \Sigma)$$

공분산의 역행렬

$$\equiv (2\pi)^{-\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

- 위 식에서 $\mu \in \mathbb{R}^k, \Sigma \in \mathbb{R}^{k\times k}$, and Σ is positive−definite

여러가지 분포

- Binomial distribution

$$f(x|n,p) = \frac{n!}{x!\,(n-x)!}p^x(1-p)^{n-x}$$

- Student's t−distribution

$$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

- F−distribution

$$f(x|d_1,d_2) = \frac{1}{B\left(\frac{d_1}{2},\frac{d_2}{2}\right)}\left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1}\left(1+\frac{d_1}{d_2}x\right)^{-\frac{d_1+d_2}{2}}$$

마르코프 시퀀스

- 마르코프(Markov) 시퀀스는 마르코프 성질(Markov property)을 만족하는 랜덤 시퀀스를 의미함

- 마르코프 성질

$$p(X_{n+1}|X_n, X_{n-1}, \dots, X_1) = p(X_{n+1}|X_n)$$

- 즉, 마르코프(Markov) 시퀀스는 현재의 확률 정보가 주어진 조건 하에서 미래와 과거가 조건부 독립인 랜덤 시퀀스임

엔트로피

- 정보량은 확률 실험에서 특정 사건의 발생 여부를 예측하기 어려운 정도를 나타내는 정량화 기법임

$$h(x) = -\log p(x)$$

- 엔트로피(entropy)는 정보량의 기댓값임

$$\mathcal{H}(p) = \mathbb{E}_{p(x)}[-\log p(x)] = -\int p(x) \log p(x)\, dx$$

엔트로피

- 가우시안 분포에 대한 엔트로피

$$\log p(x) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\mathcal{H}(p) = -\int p(x)\log p(x)\,dx$$

$$= \int p(x)\frac{1}{2}\log(2\pi\sigma^2)\,dx + \int p(x)\frac{(x-\mu)^2}{2\sigma^2}dx$$

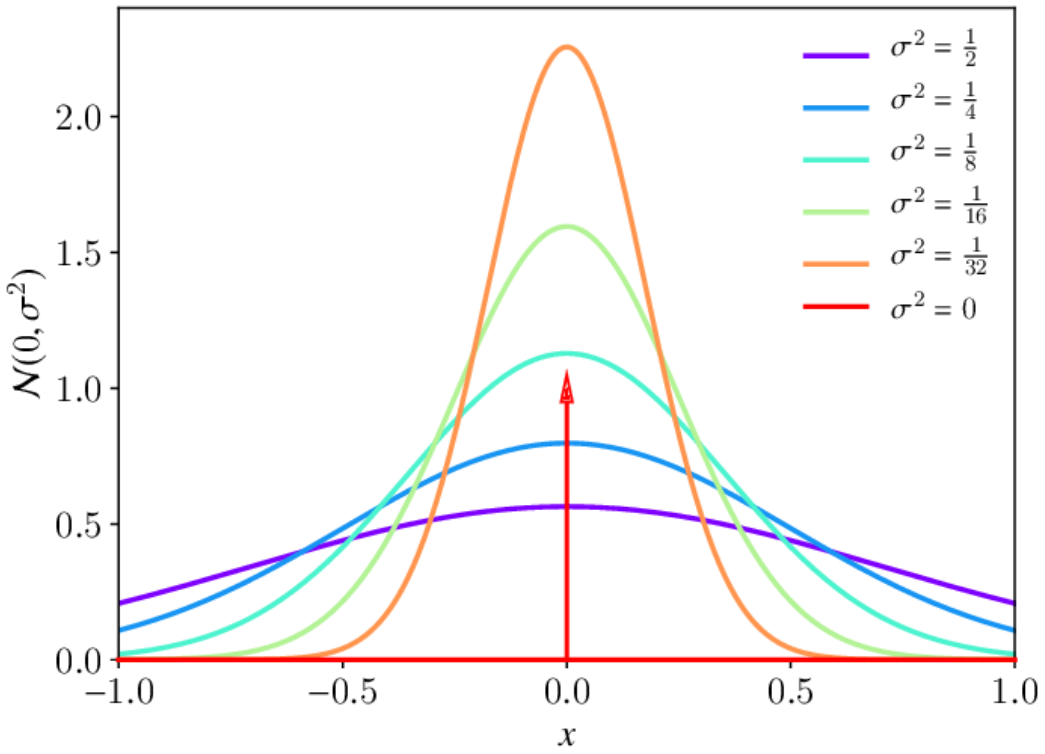
$$= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\int p(x)(x-\mu)^2 dx$$

$$= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\mathbb{E}[(x-\mu)^2]$$

$$= \frac{1}{2}(\log(2\pi\sigma^2) + 1)$$

엔트로피

$$\mathcal{H}(p) = \frac{1}{2}(\log(2\pi\sigma^2) + 1)$$



엔트로피

- 균등한 주사위

$$\mathcal{H}(p) = \sum_{i=1}^{6} \frac{1}{6} \log 6 = 0.778$$

- 짝의 주사위(짝수가 홀수보다 등장확률 2배)

$$\mathcal{H}(p) = \sum_{i=1}^{3} \frac{1}{9} \log 9 + \sum_{i=1}^{3} \frac{2}{9} \log \frac{9}{2} = 0.318 + 0.145 = 0.463$$

- 넘버원 주사위(오직 1만)

$$\mathcal{H}(p) = \sum_{i=1}^{5} 0 \log 0 + 1 \log 1 = 0 + 0 = 0$$

최대 우도 확률 추정

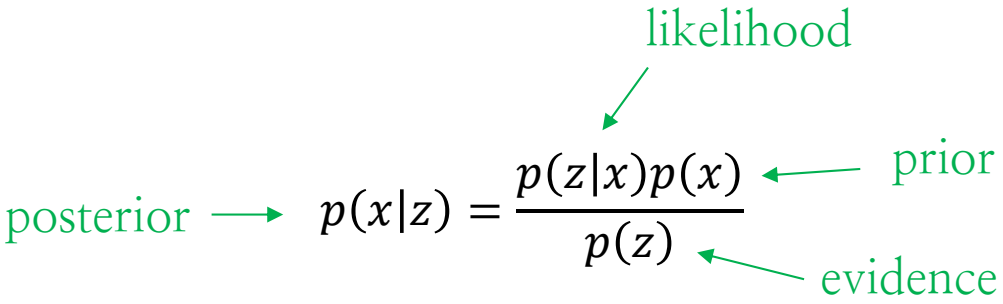
- 병원에서는 진찰을 받은 중증환자가 암 환자일 확률을 구하고자 함. 이 문제를 이진 분류 문제(1: 암환자, 0: 암환자 아님)으로 정의할 수 있음. 통계적으로 해결하기 위해 구하고자 하는 대상, 즉 암 환자일 확률을 θ 라고 하자.

- θ 의 값을 추정하기 위해, 병원에 방문한 100명의 중증환자 중 40명이 암 환자로 진단되었다고 가정하자.

- 상기 문제에서의 확률 추정은 3가지 방법으로 풀이 가능함
  - 최대 우도 확률 추정(maximum likelihood estimation, MLE)
  - 최대 사후 확률 추정(maximum a posteriori estimation, MAP)
  - 베이지안 추정(Bayesian estimation)

최대 우도 확률 추정

- 베이즈 정리(Bayes' theorem)



- 베이즈 정리를 이용한 문제 정의

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \quad \Rightarrow \quad p(\theta|D) \propto p(D|\theta)p(\theta)$$

최대 우도 확률 추정

- MLE 풀이

$$\theta^{MLE} = \underset{\theta}{\operatorname{argmax}}\, p(D|\theta)$$

- 이진 분류 문제이므로 간단한 베르누이 확률 분포(Bernoulli's distribution)를 이용하여 예측 모델을 설계함

$$p(Y = 1|\theta) = \theta$$

$$p(Y = 0|\theta) = 1 - \theta$$

- 그러면 $p(D|\theta)$는 이항 확률 분포로 표현됨

$$p(D|\theta) = \frac{n!}{k!\,(n-k)!}\theta^k(1-\theta)^{n-k}$$

$$= \frac{100!}{40!\,(60)!}\theta^{40}(1-\theta)^{60}$$

최대 우도 확률 추정

$$\theta^{MLE} = \underset{\theta}{\operatorname{argmax}}\, p(D|\theta)$$

$$= \underset{\theta}{\operatorname{argmax}} \left(\frac{100!}{40!\,(60)!} \theta^{40}(1-\theta)^{60} \right)$$

$$= \underset{\theta}{\operatorname{argmax}} (\theta^{40}(1-\theta)^{60})$$

$$= \underset{\theta}{\operatorname{argmax}} \log(\theta^{40}(1-\theta)^{60})$$

$$= \underset{\theta}{\operatorname{argmax}} (40 \log\theta + 60 \log(1-\theta))$$

$$\frac{d}{d\theta}(40 \log\theta + 60 \log(1-\theta)) = \frac{40}{\theta} - \frac{60}{1-\theta}$$

$$= \frac{100\theta - 40}{\theta(1-\theta)} \Rightarrow \theta^{MLE} = 0.4$$

최대 우도 확률 추정

- MAP 풀이

$$\theta^{MLE} = \underset{\theta}{\operatorname{argmax}}\, p(D|\theta)p(\theta)$$

- $p(D|\theta)$를 이항 확률 분포이고, $p(\theta)$를 beta distribution 로 모델링하자

$$p(D|\theta) = \frac{n!}{k!\,(n-k)!}\theta^k(1-\theta)^{n-k}$$

$$p(\theta) = \text{Beta}(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- 그러면, 사후 확률을 다음과 같이 표현할 수 있음

$$p(D|\theta)p(\theta) \propto \theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}$$

최대 우도 확률 추정

- 베타 분포의 기댓값

$$\mathbb{E}_{\theta \sim Beta(\alpha,\beta)}[\theta] = \frac{\alpha}{\alpha + \beta}$$

- 기존 연구 결과에서(혹은 경험적으로) 암 환자일 확률이 20%임을 알고 있다면, 즉 $\alpha = 5, \beta = 20$ 라고 한다면

$$p(D|\theta)p(\theta) \propto \theta^{44}(1-\theta)^{79}$$

- 따라서,

$$\frac{d}{d\theta}(44\log\theta + 79\log(1-\theta)) = \frac{123\theta - 44}{\theta(1-\theta)}$$

$$\Rightarrow \theta^{MAP} = \frac{44}{123} \approx 0.357$$

최대 우도 확률 추정

- 베이지안 추론 풀이

$$\theta^* = \mathbb{E}[\theta|D]$$

- $p(D|\theta)$가 이항 확률 분포이고, $p(\theta)$가 beta distribution 이면, $p(D|\theta)p(\theta)$는 beta distribution가 됨이 통계학에서 알려져 있음. 이러한 특징으로 인해 베타 분포를 이항 분포의 켤레(conjugate)라고 함

$$p(D|\theta) = \frac{n!}{k!\,(n-k)!}\theta^k(1-\theta)^{n-k}$$

$$p(\theta) = \text{Beta}(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$p(\theta|D) = \text{Beta}(\theta|k+\alpha, n-k+\beta)$$

최대 우도 확률 추정

$$p(D|\theta)p(\theta) = \frac{n!}{k!\,(n-k)!}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}$$

$$= \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}$$

$$p(D) = \int p(D|\theta)p(\theta)d\theta$$

$$= \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int \theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}d\theta$$

$$= \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(k+\alpha)\Gamma(n-k+\beta)}{\Gamma(n+\alpha+\beta)}$$

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(k+\alpha)\Gamma(n-k+\beta)}\theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}$$

$$= \mathrm{Beta}(\theta|k+\alpha, n-k+\beta)$$

최대 우도 확률 추정

- 베타 분포의 기댓값

$$\mathbb{E}_{\theta \sim Beta(k+\alpha, n-k+\beta)}[\theta] = \frac{k+\alpha}{n+\alpha+\beta}$$

- 기존 연구 결과에서(혹은 경험적으로) 암 환자일 확률이 20%임을 알고 있다면, 즉 $\alpha = 5, \beta = 20$ 라고 한다면

$$\theta^* = \frac{40+5}{100+5+20} = \frac{45}{125} = 0.36$$

최대 우도 확률 추정

- 병원에서는 진찰을 받은 중증환자가 암 환자일 확률을 구하고자 함. 이 문제를 이진 분류 문제(1: 암환자, 0: 암환자 아님)으로 정의할 수 있음. 통계적으로 해결하기 위해 구하고자 하는 대상, 즉 암 환자일 확률을 θ 라고 하자.

- θ 의 값을 추정하기 위해, 병원에 방문한 100명의 중증환자 중 40명이 암 환자로 진단되었다고 가정하자.

- 상기 문제에서의 확률 추정을 3가지 방법으로 풀이한 결과
  - $\theta^{MLE} = 0.4$
  - $\theta^{MAP} \approx 0.357$
  - $\theta^{*} = 0.36$

연습 문제

- 병원에서는 진찰을 받은 중증환자가 암 환자일 확률을 구하고자 함. 이 문제를 이진 분류 문제(1: 암환자, 0: 암환자 아님)으로 정의할 수 있음. 통계적으로 해결하기 위해 구하고자 하는 대상, 즉 암 환자일 확률을 θ 라고 하자.

- θ 의 값을 추정하기 위해, 병원에 방문한 100만명의 중증환자 중 40만명이 암 환자로 진단되었다고 가정하자.

- 3가지 방법으로 θ 의 값을 추정하시오.
  - 최대 우도 확률 추정(maximum likelihood estimation, MLE)
  - 최대 사후 확률 추정(maximum a posteriori estimation, MAP)
  - 베이지안 추정(Bayesian estimation)

Q & A

Jonghwanc@hallym.ac.kr

한림대학교