chapter 6. probability and Distributions

- Construction of a probability space.

Probability space $(\Omega, \mathcal{A}, P)$

Ω : sample space = 실험을 했을 때 나올수 있는 모든 가능한 결과의 집합.

↳ Event : sample space의 부분집합 (특정한 outcome의 집합)

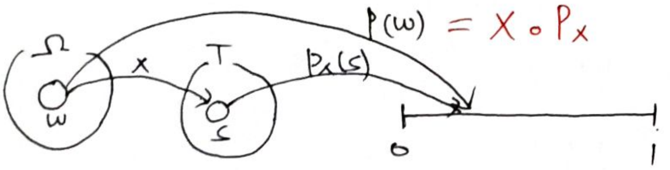
$\mathcal{A}$: Event space = 모든 가능한 event의 집합.

P : probability = 그 event가 일어날 신뢰도의 정도.

$$\begin{cases} 0 \le PA \le 1 \\ P(\Omega) = 1 \\ P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} P(A_n) \end{cases}$$

(A_k 끼리의 교집합이 없을 때 성립)

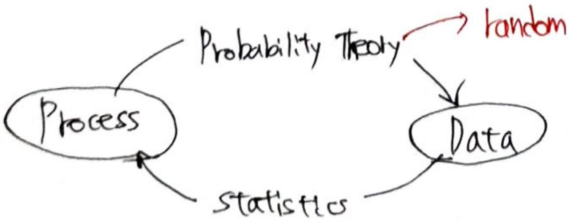
Random variable (변수)



P_x : random variable X의 distribution.

$$P_x(S) = P(X \in S) = P(X^{-1}(S))$$
$$= P(\{\omega \in \Omega : X(\omega) \in S\})$$

Probability and statistics.



Probability Theory → random variable로 모델링된 불확실성을 가지는 기본적인 process

statistics → Data에서 출발.

- Discrete and continous Probabilities.

Distribution of random variable
- Discrete random variable ⇒ Pmf $P(X = x)$
- Continous 〃 〃 ⇒
  - cdf $P(X \le x)$
  - pdf $f_x(x) = \dfrac{dF_x(x)}{dx}$

Discrete probabilities

- Univariate probability mass function $P(X=x)$
- Bivariate probability " " $P(x,y) = P(X=x, Y=y)$

Continous probabilities

- Pdf : $f : R^D \to R$ 조건 ① $f(x) \ge 0$ ② $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- Cdf : $F_X(x) = P(X_1 \le x_1, \cdots X_D \le x_D)$

○ Sum Rule, Product Rule, and Bayes' Theorem.

$p(x,y)$: X, Y의 random variable의 joint distribution.

$p(x), p(y)$: marginal distribution (각각의 distribution)

$p(y|x)$: conditional distribution of Y given X.

Sum Rule
- $p(x) = \sum_{y \in Y} p(x,y)$ y: discrete.
- $p(x) = \int_Y p(x,y)\,dy$ y: continous.

Product rule : $p(x,y) = p(y|x)\,p(x) = p(x|y)\,p(y)$

Bayes' therom : $\underbrace{p(x|y)}_{\text{posterior}} = \dfrac{\overbrace{p(y|x)}^{\text{likelihood}}\ \overbrace{p(x)}^{\text{prior}}}{\underbrace{p(y)}_{\text{evidence}}}$

: y가 주어져 있을 때 그것을 만들어낸 x가 어떤 메커니즘으로 만들어졌는지 결론을 도출하는 방법을 알려줌.

X: 관찰되지 않은 변수

Y: 관찰된 변수.

○ Summary statistics and Independence

Expected value : $E_X[g(x)] = \begin{cases} \int_X g(x)\,p(x)\,dx. \\ \sum_{x \in X} g(x)\,p(x) \end{cases}$

$X : [X_1, X_2 \cdots X_D]^T \to E_X[g(x)] = \begin{bmatrix} E_{X_1}[g(x_1)] \\ \vdots \\ E_{X_D}[g(x_D)] \end{bmatrix}$

linear operator → superposition property를 만족

Sommary Statics and Independence

Statistical Independence

X, Y : random variable.

$p(x, y) = p(x)p(y) \xrightarrow{\text{만족한다면}} p(y|x) = p(y)$

$\cancel{p}\ \mathrm{COV}_{X,Y}[x, y] = 0.$

covariance ⇒ linear dependent만 measure.

여기서의 independent → linear, non-linear 상관없이 independent.

$V_{X,Y}[x+y] = V_X[x] + V_Y[y] + \underbrace{\mathrm{COV}_{XX}(x, y)}_{0} + \underbrace{\mathrm{COV}_{xy}(y, x)}_{0}$

Transpose.

Conditional Independence.

X, Y: random variable / Z (given)

$p(x, y \mid z) = p(x|z)\,p(y|z) \quad = \quad X \perp\!\!\!\perp Y \mid Z$

$p(x, y | z) \overset{\text{product rule}}{=} p(x|y, z)\,p(y|z) \Rightarrow p(x|y, z) = p(x|z)$

※ z가 주어졌을때 y가 주어지든 아니든 x의 distribution 변화 X

⇒ y가 x에 영향 끼치지 X.

Inner product of random variable.

$\langle X, Y \rangle = \mathrm{COV}[x, y] \Rightarrow$ bilinear, symmetric (교환법칙) → inner product 조건들에 만족.

random variable의 length : $\|X\| = \sigma[x]$ length ↑ ⇒ $\sigma[x]$ ↑ ⇒ uncertain ↑.

random variables의 angle : $\cos\theta = \dfrac{\mathrm{COV}[x, y]}{\sqrt{V[x]V[y]}}$: correlation

Gaussian Distribution.

Univariate Gaussian Distribution → mean μ, variance σ^2 있으면 형태가 다 정해진다

$p(x|\mu, \sigma^2)$

Multivariate Gaussian Distribution → Mean Vector μ, covariance Σ " " "

$p(x|\mu, \Sigma)$

Marginal distribution of Gaussian distribution
Conditional " "
→ Gaussian.

x. y의 joint distribution.

$p(x, y) = \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right)$

$\rightarrow p(x) = \mathcal{N}(x \mid \mu_x, \Sigma_{xx})$

$p(x|y) = \mathcal{N}(\mu_{x|y}, \Sigma_{x|y})$

Mean.

$$E_X[x] = \begin{bmatrix} E_{X_1}[x_1] \\ \vdots \\ E_{X_D}[x_D] \end{bmatrix} \qquad E_{X_D}[x_d] = \begin{cases} \int_X x_d \, p(x_d) \, dx_d \\ \sum_{x_i \in X} x_i \, p(x_d = x_i) \end{cases}$$

Covariance.

$$\begin{cases} \text{Univariate} \Rightarrow \mathrm{Cov}_{X,Y}[x,y] = E(xy) - E(x)E(y) \\ \text{Multivariate} \Rightarrow \mathrm{Cov}(x,y) = E(xy^T) - E(x)E(y)^T \end{cases}$$

Variance

$$V_X(x) = E_X(xx^T) - E_X(x)E_X(x^T) \Rightarrow \text{Covariance matrix}$$

↳ symmetric, positive definite.

Correlation.

$$\mathrm{corr}(x,y) = \frac{\mathrm{Cov}(x,y)}{\sqrt{V(x)V(y)}} \in [-1, 1]$$

Empirical mean and covariance.

⇒ N개의 data를 $X_1 \dots X_N$ 까지 random variable 생성. → distribution을 구함.

(하나의 관찰대상에서 data를 얻었으므로 하나의 distribution 따름)

⇒ 이것을 통해 구할것 $\begin{cases} \text{Empirial mean} \\ \quad \text{" covariance.} \end{cases}$

Product of Gaussian densities $\Rightarrow$ Gaussian distribution.

$$N(x|a, A)\, N(x|b, B) = c\, N(x|c, C)$$

Gaussian 2개의 pdf (분포)를 곱하면 또다른 Gaussian 생성.

Sum and weighted sum of random variables.

X, Y : independent Gaussian random variable.

$$\Rightarrow p(x+y) = N(\mu_x + \mu_y, \Sigma_x + \Sigma_y)$$

Mixture of Gaussian densities.

$$p(x) = \alpha \underbrace{p_1(x)}_{\text{Gau}} + (1-\alpha) \underbrace{p_2(x)}_{\text{Gau}} \Rightarrow \cancel{p(x)}$$

$E(x) = \alpha \mu_1 + (1-\alpha)\mu_2$

$V(x) = [\alpha \sigma_1 \text{ ~~~~ }]$

p(x)가 Gaussian distribution이 아닐수도 있지만, 서로 다른 Gaussian density들의 weighted sum이 될수있다.

Linear / Affine transformation.

$X \sim N(\mu, \Sigma)$ $\quad y = Ax \quad \Rightarrow$ $E(y) = A\mu$

$V(y) = A\Sigma A^T$

Conjugacy and Exponential family.

Bernoulli distribution : $p(x|u) = u^x (1-u)^{1-x}$

Binomial " : $p(m|N,u) = \binom{N}{m} u^m (1-u)^{N-m}$.

Beta " : $Beta(\alpha, \beta)$: $\alpha \uparrow$: 1쪽에서 일어날 확률 ↑

$\beta \uparrow$: 0 " " ↑

Conjugacy : $\underbrace{p(x|y)}_{\text{posterior}} = \dfrac{\overbrace{p(y|x)}^{\text{liklihood}}\ \overbrace{p(x)}^{\text{prior}}}{\underbrace{p(y)}_{\text{evidence}}}$ ⇒ prior는 liklihood function에 대해 conjugate 이다.

⇒ posterior distribution을 계산하기 위해 prior distribution의 파라미터만 변형하면 된다. = posterior 와 prior이 같은 type / form 이다.

Sufficient statistics : 모집단에 대한 추론을 하는데 필요한 모든 정보.

$$p(x|\theta) = h(x)\, g_\theta(\phi(x))$$

↳ $\phi(x)$가 주어지면 θ에 관해 통과시키면 θ에 대한 정보를 추출해냄

Distribution을 생각할때 abstraction의 3가지 level

① 파라미터와 분포를 모두 아는것

② 분포를 알지만 파라미터를 모르는것 ⇒ $N(\mu, \sigma^2)$에서 maximum liklihood를 통해 μ, σ^2 추정.

③ 분포를 모를 경우 ⇒ exponential family 사용.

Exponential family : $\theta \in R^D$에 parameterized된 확률분포의 family.

$$p(x|\theta) = h(x)\underbrace{\left[\exp\left(\underbrace{\langle \theta, \phi(x)\rangle}_{\theta^T\phi(x)} - A(\theta)\right)\right]}_{g_\theta(\phi(x))} \quad \propto \exp(\theta^T \phi(x))$$

property : conjugate 관계이고 둘다 exponential family이면 두식을 곱해도 exp로 표현가능
⇒ exponential family.

Change of Variables / Inverse Transform.

- Univariate case with monotonically increasing function u.

$$F_Y(y) = P(Y \le y) = P(u(x) \le y) = P(X \le u^{-1}(y)) = \int_a^{u^{-1}(y)} f(x)\,dx.$$

$$f(y) = \frac{d}{dy} F_y(y) = \frac{d}{dy}\int_a^{u^{-1}(y)} f(x)\,dx. = f_x(u^{-1}(y)) \cdot \left(\frac{d}{dy} u^{-1}(y)\right)$$

Leibniz rule