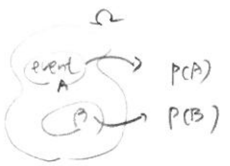
1. Probability models and axioms

- Elements of Probabilistic models
  - sample space Ω
  - probability law



- Probability Axioms
  - nonnegativity
  - additivity : $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$ * disjoint
  - normalization : $P(\Omega) = 1$
- Discrete Probability law
- Discrete uniform probability law

2. Conditioning and Bayes' rule

- $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, new universe B. $P(B) > 0$
- Multiplication rule

$$P(\cap_{i=1}^{n} A_i) = P(A_1)\, P(A_2|A_1) \cdots P(A_n | \cap_{i=1}^{n-1} A_i)$$

- Total Probability theorem

$$P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B)$$
$$= P(A_1) P(B|A_1) + \cdots + P(A_n) P(B|A_n)$$

* $A_1 \cup A_2 \cdots \cup A_n = \Omega$

- Inference and Bayes' rule

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i) P(B|A_i)}{\sum_j P(A_j) P(B|A_j)}$$

3. Independence

- Independence

$P(A \cap B) = P(A) P(B)$, $P(A|B) = P(A)$

$P(\cap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$ $\quad S = \{1, 2, \cdots, n\}$

* Pairwise $\neq$ independence

4. Counting

- Counting Principle : tree 그려나가기
- $n!$, $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$, $\dfrac{n!}{n_1! n_2! \cdots n_r!}$ — Partition

5. Discrete random variables, probability mass functions expectation

- Random variables
  - outcome $\xrightarrow{X}$ value , $w \in \Omega$, s.t $X(w) = x$

    w $\quad$ x

- Probability mass function
- Binomial pmf, geometric, Poisson pmf
- Expectation : center of gravity
- Variance : how far X from mean

6. Discrete random variable examples. Joint PMFs.

- Joint PMFs
  - $P_{X,Y}(x,y) = P(X=x, Y=y)$
  - $P_X(x) = \sum_y P_{X,Y}(x,y)$
  - $E[g(X,Y)] = \sum_x \sum_y g(x,y) P_{X,Y}(x,y)$
- Conditional PMFs and expectation

$P_{X|A}(x) = P(X=x | A)$

$E[X|A] = \sum_x x\, P_{X|A}(x)$

- Total expectation theorem

$E[X] = P(A_1) E[X|A_1] + \cdots + P(A_n) E[X|A_n]$

$(A_1 \cup A_2 \cdots \cup A_n = \Omega)$

disjoint event $A_1, \cdots, A_n$

7. Multiple discrete random variables : expectation, conditioning, independence

- Independent random variables

$P_{X|A}(x) = P_X(x)$

$P_{X,Y}(x,y) = P_X(x) P_Y(y)$ for all x, y

$P_{X,Y,Z}(x,y,z) = P_X(x) P_{Y|X}(y|x) P_{Z|X,Y}(z|x,y)$

이면 indep $= P_X(x) P_Y(y) P_Z(z)$ for all x, y, z

8. Continuous random variables

- Continuous r.v.'s and pdf's

$P(a \le X \le b) = \int_a^b f_X(x)\, dx$

$P(x \le X \le x + \delta) = \int_x^{x+\delta} f_X(s)\, ds \simeq f_X(x) \cdot \delta$

- mean and variances
- CDF $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)\, dt$

- mixed distribution
- Gaussian (normal) pdf

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad N(\mu, \sigma^2)$$

└ normality is preserved by linear transformation

$Y = aX + b$

↳ normal.

9. Multiple continuous random variables

- Joint PDF $f_{X,Y}(x,y)$: $P((X,Y) \in S) = \iint_S f_{X,Y}(x,y)\,dx\,dy$
  - Interpretation : volume

    $P(x \le X \le x+\delta,\ y \le Y \le y+\delta) \approx f_{X,Y}(x,y) \cdot \delta^2$
  - Expectations

    $$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$$
  - from joint to the marginal

    $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

- Conditioning
  - $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$, $f_Y(y) > 0$

    ($P(x \le X \le x+\delta \mid Y \approx y) \approx f_{X|Y}(x|y) \cdot \delta$

    For given y, conditional PDF is a normalized section of the joint PDF
  - If independent, $f_{X,Y} = f_X f_Y$

    $f_{X|Y}(x|y) = f_X(x)$

10. Continuous Bayes rule, derived distributions

- Bayes rule

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x) f_{Y|X}(y|x)}{\int_x f_X(x) f_{Y|X}(y|t)\,dx}$$

- Derived distribution

ex) $g(X,Y) = Y/X$, $X+Y$, $aX+b$

- Two-step procedure : $aX+b$, $Y = g(X)$ 등 $= X^3$
  - Get CDF of Y
  - Differentiate to get $f_Y(y) = \frac{dF_Y}{dy}(y)$, compute directly or graphically.
  - └ $Y = aX+b$ case

    ∗ $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$

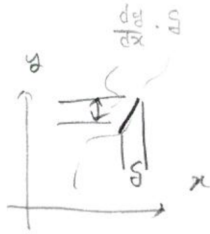
11. Derived distributions, convolution, covariance and correlation

- A general formula $Y = g(X)$

  g is strictly monotonic

$$\delta f_X(x) = \delta f_Y(y) \left| \frac{dg}{dx}(x) \right|$$

$$f_X(x) = f_Y(y) \left| \frac{dg}{dx}(x) \right|$$



- Case 1. $X + Y = Z$, $f_Z(z)$
  - └ convolution formula (or graphically)

    ┌ two step

    └ formula or flipping graph!

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-\cancel{x})\,dx$$

$$\left(p_Z(z) = \sum_x p_X(x) p_Y(w-x) \right)$$

- Case 2. $Z = Y/X$
  - └ finding CDF, diff !
    - └ graphically ...

- Two independent normal r.v.'s X, Y : sum $X+Y=Z$ normal!

- Covariance

$$\mathrm{cov}(X,Y) = E[(X-E(X))(Y-E[Y])]$$

$$= E[XY] - E[X]E[Y]$$

- $\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{var}(X_i) + \sum_{(i,j):i \ne j} \mathrm{cov}(X_i, X_j)$

- correlation coefficient

$$\rho = E\left[\frac{(X-E[X])}{\sigma_X} \cdot \frac{(Y-E[Y])}{\sigma_Y} \right] = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

- $-1 \le \rho \le 1$, $\rho \uparrow$ → strong relation

  linearly related

12. Iterated expectations, sum of random number of r.v.'s

- Conditional expectations

$E[X|Y=y] = \sum_x x\, p_{X|Y}(x|y)$ - r.v of Y.

$E[E[X|Y]] = E[X]$ ~ total expectation theorem

- var(X|Y) and its expectation

$\mathrm{var}(X|Y)$: r.v of Y $= \mathrm{Var}(X|Y=y)$

- $\mathrm{var}(X) = \underset{\text{avg var within sections}}{E[\mathrm{var}(X|Y)]} + \underset{\text{var between sections}}{\mathrm{var}(E[X|Y])}$

- Sum of random number of independent r.v.'s : X_i, N

$E[Y] = E[E[Y|N]] = E[N]E[X]$ $(Y = X_1 + \dots + X_N)$

$\mathrm{var}(Y) = E[N]\mathrm{var}(X) + (E[X])^2 \mathrm{var}(N)$

13. Bernoulli Process

- binomial with params p and n : k, fixed n

$$P_S(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (k=0,\dots,n)$$

$$E[S] = np, \quad var(S) = np(1-p)$$

- Geometric with parameter p : n, $(k=1)$

$$P_T(t) = (1-p)^{t-1} p \quad t = 1, 2, \dots$$

$$E[T] = \frac{1}{p}, \quad var(T) = \frac{(1-p)}{p^2}$$

- Independence properties of the Bernoulli Process
- Properties of the kth Arrival time
  - $Y_k = T_1 + \dots + T_k$, $P_{Y_k}(t) = \binom{t-1}{k-1} p^k (1-p)^{t-k}$, $t = k, k+1, \dots$

    Pscal pmf of order k $E[Y_k] = \frac{k}{p}$ $var(Y_k) = \frac{(1-p)}{p^2} \cdot k$
- Splitting and merging of Bernoulli Processes
- Poisson approximation to the Binomial

Poisson r.v Z with param λ (Poisson approximation)

$$\begin{cases} P_Z(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k = 0, 1, \dots) \\ P_S(k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \rightarrow P_Z(k) \end{cases}$$

converges as $n \to \infty$ and $p = \lambda/n$

14. Poisson Process

- Definition of the poisson process (arrival process with rate λ)

  (Time-homogeneity) : $P(k, \tau)$

  (Independence)

  (Small interval probabilities : $P(k, \tau)$
  - $P(0, \tau) = 1 - \lambda\delta$
  - $P(1, \tau) = \lambda\delta$
  - $P(k, \tau) = 0$
- r.v with Poisson process
  - $P_{N_\tau}(k) = P(k, \tau) = e^{-\lambda\tau} \frac{(\lambda\tau)^k}{k!}$, $k = 0, 1, \dots$

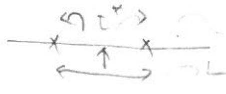
    $E[N_\tau] = \lambda\tau$, $var(N_\tau) = \lambda\tau$
  - $f_T(t) = \lambda e^{-\lambda t}$ $(t \ge 0)$ $E[T] = \frac{1}{\lambda}$, $var(T) = \frac{1}{\lambda^2}$
- Independence Properties of the poisson process
- Properties of the kth arrival time

  $Y_k = T_1 + \dots + T_k$

  $E[Y_k] = k \cdot \frac{1}{\lambda}$, $var(Y_k) = k \cdot \frac{1}{\lambda^2}$

  $$f_{Y_k}(y) = \frac{(\lambda y)^{k-1} \cdot e^{-\lambda y}}{(k-1)!} \cdot \lambda = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}$$

  Erlang pdf of order k
- Splitting and merging of poisson process
- The random incidence paradox



양쪽 끝 poisson process (or Erlang of order two)

16. Markov Process

- Specification
  - states $S = \{1, \dots, m\}$ (Set of states S)
  - transition probability P_{ij}, $(P_{ij} > 0)$ (from state i to j)
  - Independent from the past

$$P(X_{n+1} = j \mid X_n = i, X_{n-1}, \dots, X_0) = P(X_{n+1} = j \mid X_n = i) = P_{ij}$$

- n-step transition probability (recursion)

$$r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1) P_{kj}, \quad \text{for } n > 1, \text{ all } i, j$$

- Decomposition : recurrent (classes) and transient (states)

- Periodicity : states can be grouped in $d > 1$, S_k lead to S_{k+1}

  (All transitions from one group lead to the next group)

- Steady-state convergence theorem

  ㄴ recurrent, aperiodic $\rightarrow \pi_j$

$$\begin{cases} \pi_j = \lim_{n \to \infty} r_{ij}(n), \quad \text{for all } i \\ \pi_j = \sum_{k=1}^{m} \pi_k P_{kj} \quad (j = 1, \dots, m) \\ 1 = \sum_{k=1}^{m} \pi_k \end{cases}$$

- Steady-State Probabilities as Expected State frequencies.

  π_j : (Long run) frequency of being in j

  $\pi_k P_{kj}$: frequency of transitions $k \rightarrow j$

  $\sum_k \pi_k P_{kj}$: frequency of transitions into j

- Birth-death Processes

$$\pi_{i+1} = \pi_i \frac{p}{q} = \pi_i \rho$$

$$\pi_i = \pi_0 \rho^i, \quad (i = 0, \dots, m)$$

$$\pi_0 = 1 - \rho, \quad E[X_n] = \frac{\rho}{1-\rho} \quad \text{(steady-state)}$$

$$\begin{tikzcd} i \arrow[r, bend left, "p_i"] & i+1 \arrow[l, bend left, "q_{i+1}"] \end{tikzcd}$$

- Absorption probability Equations (to state s)

$$a_i = \sum_{j=1}^{m} P_{ij} a_j, \quad a_s = 1$$

  ㄴ for all other i

- Expected time to Absorption (to state s)

$$\mu_i = 1 + \sum_{j=1}^{m} P_{ij} \mu_j$$

- Mean first passage and recurrence times
  - passage i to s : $t_i = 1 + \sum_j P_{ij} t_j$ $(i \ne s)$

    $t_s = 0$
  - recurrence s : $t_s^* = 1 + \sum_j P_{sj} t_j$

19. Limit theorems

- Markov Inequality : Nonnegative values X,

$$P(X \ge a) \le \frac{E(X)}{a}, \text{ for all } a>0.$$

- Chebyshev Inequality

$$P(|X-\mu| \ge c) \le \frac{\sigma^2}{c^2} \quad \text{for all } c>0$$

- Weak law of large numbers

$X_1, X_2, \cdots$ iid r.v with mean μ.

$$P(|M_n - \mu| \ge \epsilon) = P\left(\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| \ge \epsilon\right) \to 0$$

as $n \to \infty$

- Convergence in probability

- Convergence of a deterministic sequence

└ a_n converges to a, $\lim_{n\to\infty} a_n = a$

if for every $\epsilon>0$, there exists n_0 s.t

$$|a_n - a| \le \epsilon, \text{ for all } n \ge n_0$$

- Convergence in probability

$$\lim_{n\to\infty} P(|Y_n - a| \ge \epsilon) = 0 \quad , \text{ for every } \epsilon>0$$

Sequence Y_n converges to a in probability

- Central Limit Theorem

$X_1, \cdots, X_n \to$ iid r.v common μ, σ

$$Z_n = \frac{X_1+\cdots+X_n - n\mu}{\sigma\sqrt{n}}$$

$$\lim_{n\to\infty} P(Z_n \le z) = \phi z \quad , \text{ for every } z$$

- De Moivre-laplace approximation to the <u>Binomial</u>

$$P(k \le S_n \le l) \approx \phi\left(\frac{l+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \phi\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

21. Bayesian Inference

- Terms ; Bayesian statistics (parameter estimation, hypothesis testing

- MAP rule, LMS estimation, Linear LMS estimation

- Summary : Prior f_Θ of unknown r.v Θ

model $p_{X|\Theta}$ (or $f_{X|\Theta}$) of the observation vector X.

After observing X, form posterior distribution of Θ. (using Bayes' rule)

estimator (output of Bayesian inference)

- MAP rule : $\text{Argmax}_\theta \, f_{\Theta|X}(\theta|x) = \hat{\theta}$

LMS : $\hat{\theta} = E[\Theta \mid X = x]$: conditional expectation

$$\left(\hat{\theta} = \text{argmax}_\theta \frac{f_{X|\Theta}(x|\theta) \cdot f_\Theta(\theta)}{\underbrace{f_X(x)}_{\text{const}}}\right)$$

- LMS estimation

- without information

$$E[(\Theta - E[\Theta])^2] \le E[(\Theta - \hat{\theta})^2] \quad \text{for all } \hat{\theta}$$

- Given any value x,

$$E[(\Theta - E[\Theta|X=x])^2 \mid X=x] \le E[(\Theta - \hat{\theta})^2 \mid X=x]$$

$$E[(\Theta - E[\Theta|X])^2] \le E[(\Theta - g(X))^2]$$

- Properties of the estimation error

- estimator $\hat{\Theta} = E[\Theta|X]$, error $\tilde{\Theta} = \hat{\Theta} - \Theta$

- $\text{cov}(\hat{\Theta}, \tilde{\Theta}) = 0$, $\text{var}(\Theta) = \text{var}(\hat{\Theta}) + \text{var}(\tilde{\Theta})$

- Linear LMS

estimator form $\hat{\Theta} = aX + b$, Minimize $E[(\Theta - aX - b)^2]$

$$\hat{\Theta}_L = E[\Theta] + \frac{\text{cov}(X,\Theta)}{\text{var}(X)}(X - E[X])$$

$$\rho = \frac{\text{cov}(\Theta, X)}{\sigma_\Theta \sigma_X} , \quad \text{var}(\hat{\Theta} - \Theta) = (1-\rho^2)\sigma_\Theta^2$$

$\dashrightarrow \rho \uparrow$ uncertainty $\downarrow$

(*) cleanest linear LMS example (- set derivatives to zero to find $a_1, \cdots, b$)

23. Classical Inference

- Terms : Classical statistics (parameter estimation, hypothesis testing, significance testing

- ML estimation (mean estimation, variance)

- Desirable properties of estimator

- unbiased, consistent, small mean squared error

$E[\hat{\Theta}_n] = \theta$ $\quad$ $\hat{\Theta}_n \to \theta$ $\quad$ ↳ $E[(\hat{\Theta} - \theta)^2] = \text{var}(\hat{\Theta}) + (\text{bias})^2$

<u>tradeoff</u>

- ML estimation

: Pick θ that makes data most likely

$$\hat{\theta}_{ML} = \text{argmax}_\theta \, p_X(x;\theta)$$

likelihood function of X parameterized by θ

- Estimate mean, variance

↓

sample mean

$$M_n = \frac{1}{n}(X_1+\cdots+X_n) = \hat{\Theta}$$

└ Bernoulli → upperbound $\frac{1}{2}$ (σ), ad hoc $\sqrt{\hat{\theta}(1-\hat{\theta})}$

→ generic

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\Theta}_n)^2$$

$\frac{1}{n-1}$ $\quad$ $\hat{\Theta}_n = M_n$

- Confidence Interval

: single answer may not be informative... Estimating sample mean, compute CIs using CLT $P(\hat{\Theta} - \frac{z\sigma}{\sqrt{n}} \le \theta \le \hat{\Theta}_n + \frac{z\sigma}{\sqrt{n}}) \approx 1-\alpha$