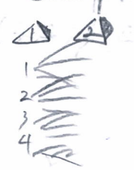
probabilistic Systems Analysis & Applied Probability

by John Tsitsiklis

Lecture 1 (Probability models, Axioms of prob)

PR = probability

- **Sample space "Ω"**
  - list of possible outcomes
    - ↳ mutually exclusive
    - collectively exhaustive
  - Art to be at the right granularity
  - Two rolls of tetrahedral dice
    - ↳ Sample space vs sequential description
    - (△:△) = (3, 4) ≠ (4, 3)
    - finite? (Dice)
    - infinite? (Dart)



- **PR axioms**
  - Event : a subset of the sample space
    - PR is assigned to events.
  - Axioms
    1. Nonnegativity : $P(A) \geq 0$
    2. Normalization : $P(\Omega) = 1$
    3. Additivity : If $A \cap B = \phi$,

$$P(A \cup B) = P(A) + P(B)$$

intersection of A & B = $A \cap B$.

union of A & B = $A \cup B$.

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

"complement" of A.

∴ $P(A^c) = 1 - P(A) \leq 1$.

A & A^c are "disjoint" (no elements in both A & A^c)

- $P(A \cup B \cup C) = P((A \cup B) \cup C)$ for disjoint sets $= P(A) + P(B) + P(C)$

- **Discrete uniform law**
  - Let all outcomes equally likely
  - $P(A) = \dfrac{\#\text{ of elements of } A}{\text{total }\#\text{ of sample points}}$
  - Computing PR ≡ counting
  - Defines fair coins, fair dice, well-shuffled decks

- **Continuous uniform law**
  - ⇒ "Probability = Area"

- Prob law : Exp. w/ countably infinite sample space

$P(n) = 2^{-n} \Rightarrow \left\{ \begin{array}{l} P(\Omega) = \frac{\frac{1}{2}}{1-\frac{1}{2}} = 1 \\ P(\text{even}) = \frac{1}{3} \end{array} \right\}$ (n = integer)

→ $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \ldots$

If All is disjoint

Lec 2 Conditional Probability & Bayes' rule

$P(A|B)$ = PR of A given that B occurred

B is our new universe.

Def. Assuming $P(B) \neq 0$, $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

→ $P(A \cap B) = P(B)\,P(A|B)$

↳ (freq of $A \cap B$) = (freq of B) × (fraction of A wrt B)

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

If $A \cap B = \phi$ → $P(A \cup B|C) = P(A|C) + P(B|C)$

ex) A : Airplane in the air
B : Radar Alarm.

$P(A) = 0.05$
- $P(B|A) = 0.99$
- $P(B^c|A) = 0.01$ — false negative problem

$P(A^c) = 0.95$
- $P(B|A^c) = 0.10$ → false Alarm
- $P(B^c|A^c) = 0.90$

$P(A \cap B) = P(A)\,P(B|A) = 0.05 \times 0.99 = 0.0495$

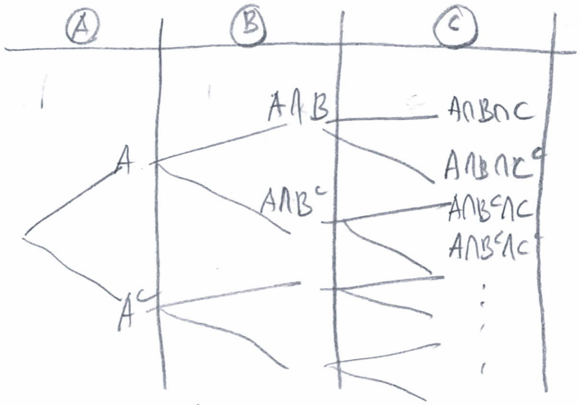
$P(B) = P(A \cap B) + \underbrace{P(A^c \cap B)}_{P(A^c)P(B|A^c)} = 0.05 \times 0.99 + 0.95 \times 0.1 = 0.1445$.

Q. Given your radar alarm → PR of existence of plane

⇒ $P(A|B) = P(A \cap B)/P(B) = \sim 34\%$.

- Multiplication rule (generalized)

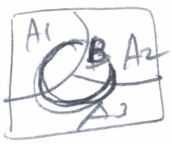
$P(A \cap B \cap C) = P(A) \times P(B|A) \times P(C|A \cap B)$



- Total PR theorem

$P(B) = P(A_1) P(B|A_1) + \cdots = \sum_i \boxed{P(A_i)} P(B|A_i)$

Add to 1



likelihood of different possible scenario (A_i)

average of PR of B in the different worlds (or scenarios)

- Bayes' rule

prior PR $P(A_i)$ → initial beliefs

we know $P(B|A_i)$ for each i

wish to compute $P(A_i|B)$

⇒ revise "beliefs", given that B occurred

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

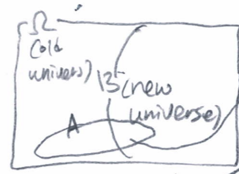
$$= \frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}$$

Scenario	cause-effect model	observation
A_i	$\xrightarrow{P(B\|A_i)}$	B
	inference	

Tells about $P(A_i|B)$

How one can learn from experience in systematic way?!

Lecture 3. Independence



$P(A|B)$; conditional PR.

mult — $P(A \cap B) = P(B) \cdot P(A|B)$

total PR — $P(B) = P(A)P(B|A) + P(A^c)P(B|A^c)$

Bayes — $P(A_i|B) = \dfrac{P(A_i)P(B|A_i)}{P(B)}$

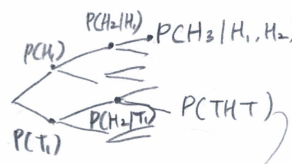
which one really occur? ≒ "inference"

different scenario

different observation

(ex) 3 tosses of a balanced coin;

$P(H) = p \quad P(T) = 1-p$



$P(THT) \stackrel{?}{=} P(T_1) P(H_2|T_1) P(T_3|T_1,H_2)$

$= (1-p) \times p \times (1-p)$

$P(1\text{ head}) = 3 \times \begin{pmatrix} P(THT) \\ P(TTH) \\ P(HTT) \end{pmatrix} = 3(1-p) \times p \times (1-p)$

$P(\text{first toss} = H \mid 1\text{ head}) = \dfrac{P(1\text{st toss }H \ \&\ 1\text{ head})}{P(1\text{ head})} = \dfrac{1}{3}$

- Independence of two events

when $P(B|A) = P(B)$

then $P(A \cap B) = P(A) \times P(B)$

independent ≠ disjoint

- Conditioning may affect independence

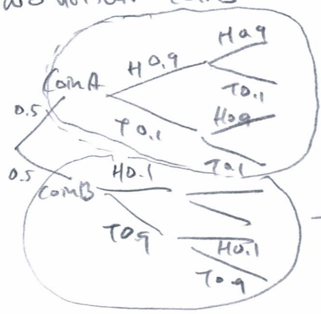
$P(A \cap B|C) = P(A|C)P(B|C)$

Def of independence in conditional universe.

→ in C, A & B are disjoint!

∴ A & B are not independent

- two unfair coins



$\Rightarrow$ Are the coin flips independent? Yes

$\Rightarrow$ Also independent flip-coins Yes

→ If we know which coin is C → Indep

→ If we don't know?

$P(\text{toss } 11 = H) = \frac{1}{2} \times 0.9 + \frac{1}{2} \times 0.1 = \frac{1}{2}$.
(A) (B)

$\neq P(\text{toss } 11 \mid \text{first 10 tosses are Heads})$

inference calculation!

≈ 90% ($\because$ it should be coin A!)

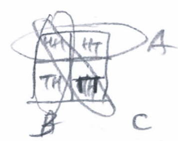
→ Observation changed our prior belief!!

- Independence vs pairwise independence

A : First toss = H

B : Second toss = H $\quad P(A) = P(B) = 1/2$

C : 1st 2nd gives same result



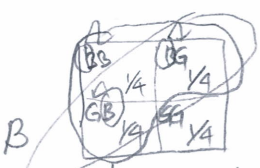
$P(C) = 1/2$ → pairwise indep.

$P(C \cap A) = 1/4$

$P(A \cap B \cap C) = 1/4$ → not indep.

$P(C \mid A \cap B) = 1$

- King's Sibling ; The King comes from a family of two children, what is the PR that his sibling is female?



At least one boy (A)

$\hookrightarrow 2/3 = P(B|A)$

$= \frac{P(A \cap B)}{P(A)} = \frac{2/4}{3/4} = \frac{2}{3}$.

Lecture 4. Counting

- Discrete uniform law

  - $P(A) = \frac{|A|}{|\Omega|}$ → counting

- Basic Counting Principle

  - r stages

    n_i choices at stage i

    → # of choices $= n_1 n_2 n_3 \cdots n_r$

  - permutation $n!$

  - Number of subset 2^n

- Combinations

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{Binomial coeffs.}$$

- Binomial Probabilities

  $\hookrightarrow \binom{n}{k} p^k (1-p)^{n-k}$

- partitions → $\frac{n!}{n_1! n_2! \cdots n_r!}$

Lecture 5. Discrete Random Variables (I)

- $\boxed{\text{Random Variables}}$: A function from Ω to real number

- PMF (Prob. Mass function)

  $P_X(x) = P(X = x) = P(\{\omega \in \Omega \text{ s.t. } X(\omega) = x\})$

  $\therefore P_X(x) \geq 0$ & $\sum_x P_X(x) = 1$.

- How to get PMF?

  ① Collect all possible outcomes for which X is equal to x

  ② add their probability

  ③ repeat for all x.

- Binomial PMF $\Rightarrow P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$

  X = number of heads in 'n' independent coin tosses

- Expectation : $E(X) = \sum_x x P_X(x)$

- Variance : $Var(X) = E[(X - E[X])^2]$

$$= \sum_x (x - E[x])^2 P_X(x)$$

$$= E[X^2] - E[X]^2$$

- Std $\sigma_X = \sqrt{Var(X)}$

(Lect. 6 Discrete Random Var. II)

- Conditional PMF & Expectation

$$\begin{cases} P_{X|A}(x) = P(X = x \mid A) \\ E[X \mid A] = \sum_x x P_{X|A}(x) \end{cases}$$

- Geometric PMF

X : # of independent coin toss until first head

$$P_X(k) = (1-p)^{k-1} p$$

$$E[X] = \sum_{k=1}^{\infty} k P_X(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p$$

- Total Expectation Theorem (A_1, A_2, A_3 → disjoint partition)

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)$$

↓

$$P_X(x) = P(A_1)P_{X|A_1}(x) + P(A_2)P_{X|A_2}(x) + P(A_3)P_{X|A_3}(x)$$

↓

$$E[X] = P(A_1)E(X|A_1) + P(A_2)E(X|A_2) + P(A_3)E(X|A_3)$$

- Joint PMF

$$P_{X,Y}(x,y) = P(X = x \,\&\, Y = y)$$

- $\sum_x \sum_y P_{XY}(x,y) = 1$

- $P_X(x) = \sum_y P_{XY}(x,y)$

⇒ $P_{X|Y}(x|y) = P(X = x \mid Y = y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$

Conditional PMF

- $\sum_x P_{X|Y}(x|y) = 1$

Conditional PMF = $\frac{\text{Joint PMF}}{\text{PMF}}$

(Lect. 7 Multiple random variables)

- Independent Random Variables

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

X, Y, Z are indep if

$P_{XYZ}(xyz) = P_X(x) P_Y(y) P_Z(z)$

- Expectation

→ $E[g(X,Y)] = \sum_x \sum_y g(x,y) P_{X,Y}(x,y)$

if X, Y independent

↳ $E[XY] = E(X)E(Y)$

$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

- Variances

→ $Var(aX) = Var(X) \times a^2$

$Var(X+a) = Var(X)$

If X, Y indep ⇒ $Var(X+Y) = Var(X) + Var(Y)$

- Binomial Mean & Variance

$$\begin{cases} X = \text{\# of successes in } n \text{ indep. trial} \\ E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{k} = pn \end{cases}$$

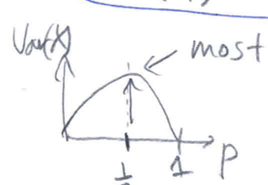
$X_i = \begin{cases} 1 & \text{if success in trial } i \\ 0 & \text{otherwise} \end{cases}$ $\quad (\therefore X = \sum_i X_i)$

$E[X_i] = 1 \cdot p + 0 \cdot (1-p) = p$

$E(X) = n \cdot p$

$Var[X_i] = p(1-p)^2 + (1-p)(0-p)^2 = p(1-p)$

$Var(X) = np(1-p)$



- Hat problem :

Lect. 8 Continuous Random Variables

- A continuous Random Variables (RV) is described by probability density fcn.

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

$\therefore \int_{-\infty}^{\infty} f_X(x)\,dx = 1$

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

$P(X \in B) = \int_B f_X(x)\,dx$ for nice "set" B

- $E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$

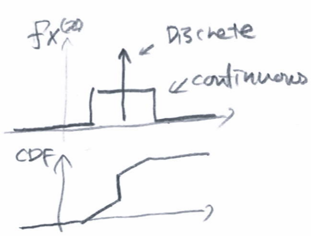
$E(g(x)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$

$Var(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - E(x))^2 f_X(x)\,dx$

- Cumulative Distribution function ⇒ Probability

$F_X(x) = P_X(X \le x) = \int_{-\infty}^{x} f_X(x)\,dx$

- Mixed Distribution



- Gaussian PDF.
  - Standard normal $N(0,1)$: $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
  - $E(x) = 0$ $V(x) = 1$
  - General model $N(\mu, \sigma^2)$

  ↳ $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

☆ The constellation of concepts

$P_X(x)$ PMF	$f_X(x)$ PDF
$F_X(x)$ CDF	
$E(X)$, $Var(X)$	

Expand to

Joint PR	$P_{X,Y}(x,y)$	$f_{X,Y}(x,y)$
Conditional PR.	$P_{X\|Y}(x,y)$	$f_{X\|Y}(x,y)$

Lect. 9 Multiple Cont. Rand. Var. & PDFs

- Joint PDF

$P((X,Y) \in S) = \iint_S f_{X,Y}(x,y)\,dx\,dy$

$E[g(X,Y)] = \iint g(x,y) f_{X,Y}(x,y)\,dx\,dy$

X, Y is independent if ...

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

- Conditioning

$P(x \le X \le x+\delta) \approx f_X(x)\,\delta$

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

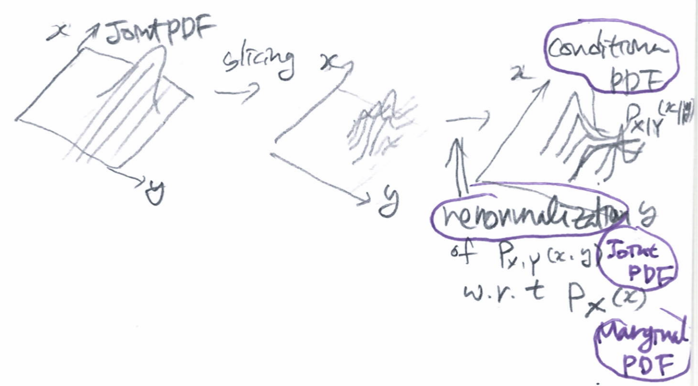
Definition

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$ if $f_Y(y) > 0$

Conditional PDF, Joint PDF, Marginal PDF.

⇒ For given y, conditional PDF is a (normalized) "section" of the joint PDF.

If independent → $f_{X|Y}(x,y) = f_X(x)$



10. Continuous Bayes Rule ; Derived Distribution

- The Bayes variations

$$\underset{\text{unknown RV}}{X} \xrightarrow[P_X(x)]{} \boxed{\overset{\text{model}}{P_{Y|X}(y|x)}} \xrightarrow{} Y$$

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

$$\boxed{P_{X|Y}(x|y)\ ?}$$

(unknown random variable) (observable)

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

$$= P_Y(y)P_{X|Y}(x|y)$$

$$\Rightarrow \boxed{P_{X|Y}(x|y) = \frac{P_X(x)P_{Y|X}(y|x)}{P_Y(y)}}\ \boxed{\text{Inference}}$$

$$\Rightarrow P_{Y}(y) = \sum_x P_X(x) P_{Y|X}(y|x)$$

Continuous

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

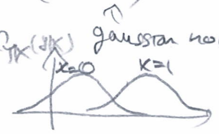
$$\& \ f_Y(y) = \int_x f_X(x)f_{Y|X}(y|x)\,dx.$$

example
- X: some signal "prior" $f(x)$
- Y: Noisy version of X
- $f_{Y|X}(y|x)$: model of noise.

Discrete X, Continuous Y

- X : discrete signal (0, 1)
- Y : noisy version of X ($X+W$)
- $f_{Y|X}(y|x)$: cont. noise model

gaussian noise



Cont X, Discrete Y

- X : a cont. signal, "prior" $f_X(x)$ (intensity of light beam)
- Y : discrete r.v. affected by X (photon count)
- $P_{Y|X}(y|x)$: model of discrete r.v

- Derived Distribution

$$f_{X,Y}(x,y) \xrightarrow[G(X,Y)=\frac{Y}{X}]{} f_G(g)\ ?$$

① Discrete case → $Y = g(X)$ then,

$$P_Y(y) = P(g(X)=y)$$
$$= \sum_{x:g(x)=y} P_X(x)$$

② Continuous case

$$f_Y(y) = \frac{dF_Y}{dy}(y) = \frac{d}{dy}(P(Y\le y))$$

← CDF

11. Derived Dist, Convolution, Covariance & Correlation

- A general formula

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

- Distribution of $X+Y=W$

$$P_W(w) = P(X+Y=w)$$
$$= \sum_x P(X=x)P(Y=w-x)$$
$$= \sum_x P_X(x)P_Y(w-x)$$

- Continuous case ?

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w-x)\,dx$$

- Two - independent normal random variables

$$\begin{cases} X \sim N(\mu_x, \sigma_x^2) \\ Y \sim N(\mu_y, \sigma_y^2) \end{cases}$$

then, $f_{X,Y}(x,y) = f_X(x) \times f_Y(y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}\right)$

- Sum of independent normal r.v.

$$f_W(w) = \int f_X(x) f_Y(w-x)\,dx = ce^{-\gamma w^2}$$

(for $\mu_x = \mu_y = 0$ case)

- Covariance !!

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

Independent ? $\mathrm{cov}(X,Y) = 0$

$\Rightarrow$

$\not\Leftarrow$