1/23/14

GRAPHS + PROB REVIEW + BN

① Graph Concepts

→ G is an ordered pair or tuple of two sets

$G = (V, E)$

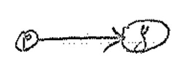
set of vertices/nodes

Set of edges/arcs

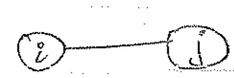
→ $V = \{\underbrace{A, B, C, \ldots}_{\text{names of vertices}}, Z\}$

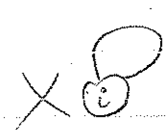
$V = \{v_1, v_2, \ldots, v_N\}$

→ Edge Set: $E = \{(i,j) \mid i \in V, j \in V\}$ Directed graphs



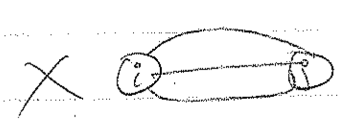
$E = \{\{i,j\} \mid i \in V, j \in V\}$ Undirected graphs





typical assumptions: 1) No self loops $i \neq j$ $\forall e = (i,j) \in E$

2) No repeated edge $e_1 \neq e_2$ $\forall e_1, e_2 \in E$



→ Weighted Graph $G = (V, E, W)$

where $W = \{w_1, w_2, \dots, w_{|E|}\}$

weight associated with edge e_1

$$i \xrightarrow{w_{ij}} j$$

$$i \overset{w_{ij}}{\text{———}} j$$

→ Neighbour → Undirected G: $N(i) = \{j \mid \{i,j\} \in E\}$

→ Parent/Child (Directed graphs)

$(i,j) \in E \Rightarrow$ i is a parent of j
j is a child of i

→ Walk: Sequence of vertices $v_1 v_2 \dots v_k$
s.t. $(v_i, v_j) \in E$
or $\{v_i, v_j\} \in E$

→ Path: Walk with no repeated nodes

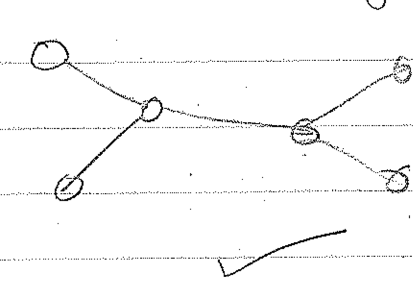
→ Cycle: Path with $v_1 = v_k$ (start = end)

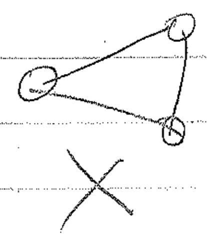
→ Connected component in G:

largest set of vertices $S \subseteq V$ s.t
$\forall\, i,j \in S$ $\exists$ $i-j$ path (undirected)
or $i \to j$ path (directed)

→ Connected Graph $\doteq$ #Components $= 1$

→ Tree: undirected G; no cycles

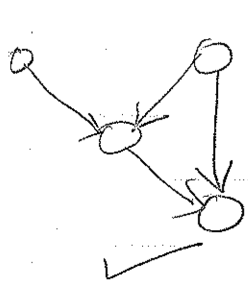


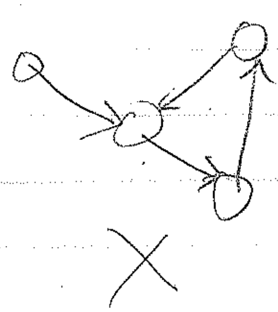


→ Spanning Tree of G is a graph $T = (V_T, E_T)$

s.t $V_T = V$
$E_T \subseteq E$
& there are no cycles in E_T

→ Directed Acyclic Graph (DAG)
directed G; no directed cycles





Ancestor: i is an ancestor of j if $\exists$ $i \to j$ path

Descendant: j is a descendant of i if $\exists$ $i \to j$ path

② Probability Refresher

[As an appendix in the end; Notes from F'13: 4984/5984]

(3) Bayes Nets

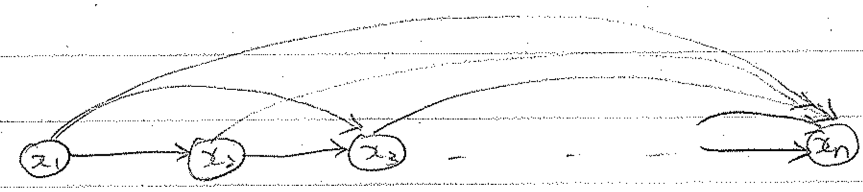
(3.1) Let's start with Chain Rule

$$P(x_1, \ldots, x_n) = P(x_1) P(x_2 \ldots x_n | x_1)$$
$$= P(x_1) P(x_2 | x_1) P(x_3 \ldots x_n | x_1, x_2)$$
$$\ldots$$
$$= P(x_1) \prod_{j=2}^{n} P(x_j | x_1 \ldots x_{j-1})$$

Let's try to draw this as a graph:

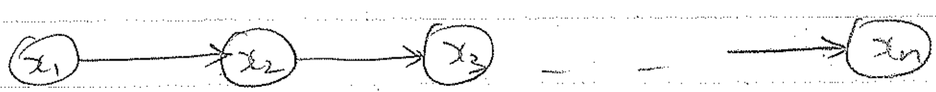
→ 1 node per variable

→ an edge $x_i \to x_j$ if $j > i$ ⇒ x_i appears to right of condition sign



$$\# \text{edges} = 1 + 2 + \ldots + n-1 = \frac{n(n-1)}{2} = {}^nC_2 \quad \text{complete graph}$$

How about a sparse graph?



Look familiar? Markov Chain!

(3.2) Chain Rule vs Markov Chain

STRUCTURE

Chain Rule:

$$\begin{tikzcd} x_1 \arrow[r] \arrow[rrr, bend left] & x_2 \arrow[r, dash, dashed] \arrow[rr, bend left] & \cdots \arrow[r] & x_n \end{tikzcd}$$

Markov Chain:

$$x_1 \to x_2 \to \cdots \to x_n$$

FACTORIZATION

Chain Rule:

$$P(x_1 \ldots x_n) = P(x_1) \prod_{j=2}^{n} P(x_j \mid x_1 \ldots, x_{j-1})$$

Markov Chain:

$$P(x_1 \ldots x_n) = P(x_1) \prod_{j=2}^{n} P(x_j \mid x_{j-1})$$

ASSUMPTION

Chain Rule: None

Markov Chain:

$$P(x_j \mid x_1 \ldots x_{j-1}) = P(x_j \mid x_{j-1}) \quad \forall j$$

$\Rightarrow X_{j-2} \perp X_j \mid X_{j-1}$

Future is indep of past given present

9/5/13

1 Probability Concepts

→ Sample Space (The space of events)

$\Omega = \{H, T\}$ (for a coin)

$= \{\text{spam}, \text{no-spam}\}$ (for email)

$= \{\text{car}, \text{boat}, \text{person}, \ldots\}$ for an image

→ Random Variable (Mapping from Sample Space to numbers)

Discrete

$X : \Omega \to \{0, 1, \ldots, k\}$

$: \Omega \to \mathbb{R}$ (continuous)

→ Notation: X, Y random variables (capital) (e.g Y could be label we give to an email)

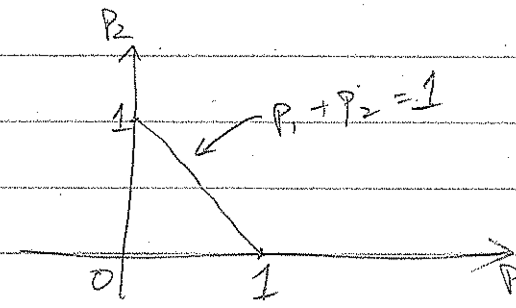
x, y: their states (e.g 0 or 1)

→ Probability Mass $\sum_{x=0}^{k} P(X = x) = 1$

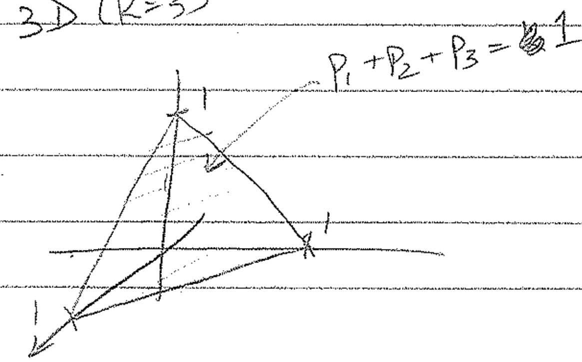
Sometimes it is useful to think of prob as a vector

$$\vec{P} = \begin{bmatrix} P_0 \\ \vdots \\ P_k \end{bmatrix} \qquad \left. \begin{array}{l} \vec{P} \in \mathbb{R}^k \\ \vec{P} \geq 0 \\ \sum p_i = 1 \end{array} \right\} \longleftarrow \text{SIMPLEX}$$

2D (K=2)



3D (K=3)



For Continuous R.Vs Prob. Density Function

$$\int_x p(x)\,dx = 1$$

$p(x) \geq 0$

$p(x)$ can be ≥ 1

Same symbol p; Discrete or Continuous, clear from context.

$\rightarrow$ Expectation of $f(X)$ ← some function of R.V. X

$$E_p[f(X)] = \sum_{x=0}^{k} f(x)p(x) \quad \text{(Discrete)}$$

← [Weighted Average value of $f(X)$]

$$= \int_x f(x)p(x)\,dx$$

Can think of $E[f(X)]$ as inner-product (or linear operation) for discrete RVs.

$$E[f(X)] = [f(0) \; \cdots \; f(k)] \cdot \begin{bmatrix} p(0) \\ \vdots \\ p(k) \end{bmatrix} = \sum_{x=0}^{k} f(x)p(x)$$

$\rightarrow$ Mean Value of X under p: set $f(X) = X$

$$\mu = E[X] = \sum_{x=0}^{k} x\,p(x)$$

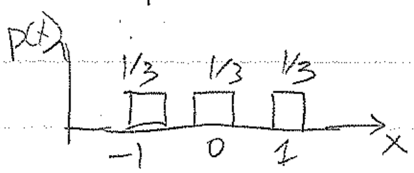
$= 3.5$ (for fair dice)

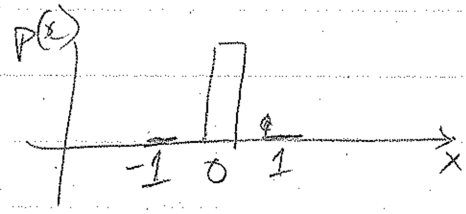
→ Variance: Estimate of "spread" around μ

$$f(x) = (x-\mu)^2$$

$$\text{Var}(X) = E[(X-\mu)^2]$$

Example:





$\mu = 0$

$$\text{Var}(X) = E[X^2]$$

$$= \frac{(-1)^2}{3} + \frac{0}{3} + \frac{1^2}{3}$$

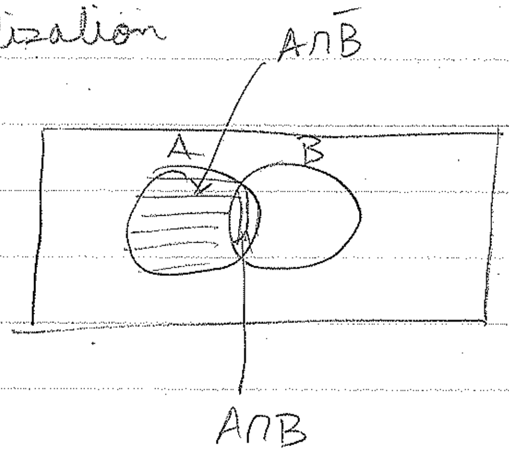
$$= \boxed{\frac{2}{3}}$$

$\mu = 0$

$$\text{Var}(X) = 0 + 0 + 0 = \boxed{0}$$

No spread.

→ Marginalization



$$A = (A \cap B) \cup (A \cap \bar{B})$$

$$\Rightarrow p(A) = p(A \cap B) + p(A \cap \bar{B}) - \underbrace{p(\{A \cap B\} \cap \{A \cap \bar{B}\})}_{0}$$

$$\Rightarrow p(A) = p(A \cap B) + p(A \cap \bar{B})$$

→ Conditional Prob.

$$P(Y=y \mid X=x) = \frac{P(Y=y, X=x)}{P(X=x)}$$

Chain Rule $P(X=x, Y=y) = P(Y=y \mid X=x) P(X=x)$

[Recursive Application]

$$P(X_1=x_1, X_2=x_2 \ldots, X_d=x_d, Y=y) = P(X_2=x_2, \ldots, X_d=x_d, Y=y \mid X_1=x_1) \cdot P(X_1=x_1)$$

$$= P(Y=y \mid X_1=x_1 \ldots, X_d=x_d) \cdot P(X_d=x_d \mid X_1=x_1 \ldots, X_{d-1}=x_{d-1}) \ldots P(X_2=x_2 \mid X_1=x_1) P(X_1=x_1)$$

→ Independence:

$$P(Y=y, X=x) = P(Y=y) \cdot P(X=x) \quad \boxed{\forall y, x}$$

Very Imp!

→ $\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_X)]$

where $\mu_X = E[X]$ $\mu_Y = E[Y]$

$$\text{Corr-Coeff}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$$

Bayes Rule

$$\underbrace{P(Y=y \mid X=x)}_{\text{Posterior}} = \frac{\overbrace{P(X=x \mid Y=y)}^{\text{Likelihood}}\ \overbrace{P(Y=y)}^{\text{Prior}}}{\underbrace{P(X=x)}_{\text{Evidence}}}$$