W2

Conditional Probability

$P(X,Y) = P(X)P(Y)$ - independent

$\sum_x p(x_i|y) = 1$

$P(A_i|B)P(B) = P(B|A_i)P(A_i)$

conditional & marginal distribution

$P(X_a, X_b) = \sum P(X_a, X_b, X_c)$

$p(a_1, \cdots, a_n) = p(a_1|a_2 \cdots, a_n) \cdots p(a_{n-1}|a_n)p(a_n)$

$\underline{n!}$

indepent: $P(A|B) = P(A)$

Bayes Rule

Conditional Independent

$X \perp Y$: $P(X,Y) = P(X)P(Y)$

$X \nearrow X \perp Y | Z$

$\searrow X \perp Z | Y$

$X \perp Y$
$X \perp Z$ $\Leftrightarrow X \perp (Y, Z)$

$(X \perp Y|Z) \& (X \perp Z|Y) \Rightarrow X \perp (Y, Z)$

$X \perp Y|Z \; := P(X, Y|Z) = P(X|Z)P(Y|Z)$

10.4 Week 3.

1) Markovian (G)

$\forall\ V \in$ vetrices (G)

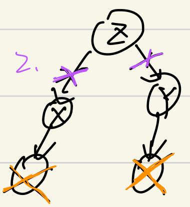
$V \perp$ non-descendants (V) / Parents (V)

2) <u>Algorithm</u>   G $(X \perp Y \mid Z)$

G1. Deletes all leaves that are not in X, Y or Z

2. remove all outgoing edges of Z $\longrightarrow$ G'

$X \perp Y \mid Z$ if X and Y are not joined in G'



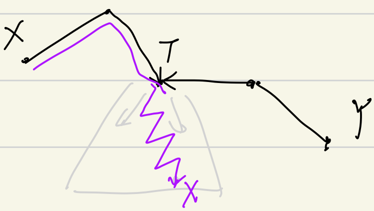
<u>proof</u>:

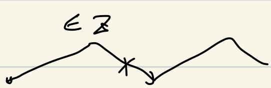
$\rightarrow T \rightarrow$ $T \notin Z$

$\leftarrow T \leftarrow$ $T \notin Z$

$\swarrow T \searrow$ $T \notin Z$

$\searrow T \swarrow$ if T or at least its decendents are observed





some edges r not delete

Exercise:

X, Y are two non-adjacent nodes of G

Z = set of all ancestors of X

$\cup$ " " " " Y

并 excluding X, Y

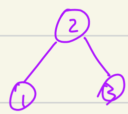
Prove: $X \perp Y \mid Z$

Fix v $\quad p_{x,v} = P_r\{$RW started at x visit v "eventually"$\}$

$$\forall x \quad p_{x,v} = \frac{1}{\deg(x)} \sum_{y \in N(x)} p_{y,v} \quad ; \quad p_{v,v} = 1.$$

$$p = (p_{x_1,v}, \cdots, p_{x_n,v}) \in \mathbb{R}^n$$

$$A \cdot p = p$$



$$A = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

i-th element of Ap

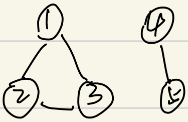
$$= \sum_{y \in N(x)} \frac{1}{\deg(x_i)} p_{y,v}$$

Target: find solutions to $Ap = p$

p should be an eigenvector for eigenvalues 1

"one candidate is $p = (1, \cdots, 1)$" linear independence

Claim if G is connective, p is unique



$$\begin{vmatrix} -\lambda & 1 & 1 & & \\ 1 & -\lambda & 1 & & 0 \\ 1 & 1 & -\lambda & & \\ & 0 & & -\lambda & 1 \\ & & & 1 & -\lambda \end{vmatrix}$$

$\to 2 \quad 1$

$\to 1$

$(\lambda-1)^2 \ldots$

find 2 linear independent eignvectors ?

1.1

$$P(X_{t+1} \mid X_t, X_{t-1}, \dots, X_0) = P(X_{t+1} \mid X_t)$$

$$T(x,y) = P(X_{t+1}=y \mid X_t=x)$$

distri π : $\pi(x)T(x,y) = \pi(y)T(y,x)$

$$\underbrace{P(X_t=x)P(X_{t+1}=y \mid X_t=x)}_{} = P(X_t=y)P(X_{t+1}=x \mid X_t=y)$$

$$P(X_{t+1}=y, X_t=x) = P(X_{t+1}=x, X_t=y)$$

$$P(X_t=x_0) = \pi(x_0)$$

$$P(X_t=x_0, \dots, X_{t+k}=x_k) = P(X_t=x_k, \dots, X_{t+k}=x_0) \text{ :}$$

$$
\begin{aligned}
P(X_t=x_0, \dots, X_{t+k}=x_k) &= \underbrace{P(X_t=x_0)}_{}\underbrace{P(X_{t+1}=x_1 \mid X_t=x_0)}_{} \cdots \underbrace{P(X_{t+k}=x_k \mid X_{t+k-1}=x_{k-1})}_{} \\
&= \underbrace{\pi(x_0) \cdot T(x_0,x_1)}_{} \cdot \cdots \cdot T(x_{k-1},x_k) \\
&= T(x_1,x_0)\underbrace{\pi(x_1)T(x_1,x_2)}_{T(x_2,x_1)\pi(x_2)} \cdots T(x_{k-1},x_k) \\
&\quad \cdots \overbrace{\qquad\qquad} \\
&= T(x_1,x_0)T(x_2,x_1) \cdots T(x_k,x_{k-1})\pi(x_k) \\
&= P(X_{t+k}=x_0 \mid X_{t+k-1}=x_1)P(X_{t+k-1}=x_1 \mid X_{t+k-2}=x_2) \cdots P(X_{t+1}=x_{k-1} \mid X_k=x_k)P(X_t=x_k) \\
&= P(X_t=x_k, \dots, X_{t+k}=x_0)
\end{aligned}
$$

2. $\underbrace{\frac{1}{|\mathcal{Y}|}}_{\pi(y)}T(x,y) = T(y,x)\underbrace{\frac{1}{|\mathcal{Y}|}}_{\pi(x)} \quad \forall x,y \in \mathcal{Y}$

$\pi(x)$ = Uniform distribution over $\mathcal{Y}$.

- Q ~> Unormalized target distribution

$R(x \mid x')$ ~> proposal distribution from which we know how to sample

Algo (MH). Initialize at x_0

Until converge

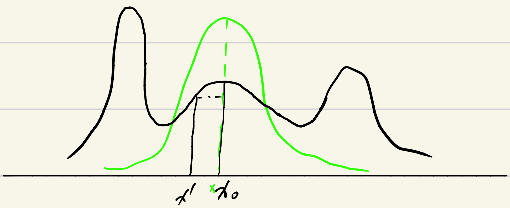
1 $\quad x = R(x \mid x_t)$

2 $\quad \alpha = \min\left(1, \dfrac{Q(x)\, P(x_t \mid x)}{Q(x_t)\, R(x \mid x_t)}\right)$

3. with this prob., accept $x_{t+1} = x$

otherwise, $x_{t+1} = x'$

$R(x \mid x') = \mathcal{N}(x; \mu = x', \sigma = 1) = \frac{1}{Z} \exp\left(-\frac{1}{2}(x - x')^2\right)$



(i) → see claim (not specific for MH)

(ii) $T(x, x')$?

$x_{t+1} \neq x_t$

$T(x_t, x_{t+1}) = P(X_{t+1} = x_{t+1} \mid X_t = x_t)$.

$=$ probability to sample x_{t+1} from $R(\cdot \mid x_t)$ $\times$ prob. of accepting x_{t+1}

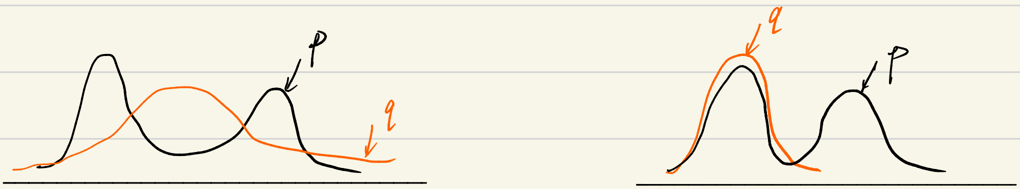
$= \dfrac{Q(x_t)}{Q(x_t)} R(x_{t+1} \mid x_t)\, \alpha$ $\quad \alpha = \min\left(\dfrac{Q(x_{t+1}) R(x_t \mid x_{t+1})}{Q(x_t) R(x_{t+1} \mid x_t)}, 1\right)$

$= \dfrac{1}{Q(x_t)} \min(\quad , \quad)$

by symmetric, $T(x_{t+1}, x_t) = \dfrac{1}{Q(x_{t+1})} \min(\quad , \quad)$

$Q(x_{t+1})\, T(x_{t+1}, x_t) = \dfrac{1}{Q(x_{t+1})} (\quad , \quad) = Q(x_t)\, T(x_t, x_{t+1})$

3.



$$KL(q\|p) = \int q(x) \ln\left(\frac{q(x)}{p(x)}\right) dx = \int q(x) \ln q(x)\, dx - \int q(x) \ln p(x)\, dx$$

$KL(p\|q)$

$f: \mathbb{R}^m \to \mathbb{R}^n$

$x = (x_1, \dots, x_m) \to (f_1(x), \dots, f_n(x))$

$$Df = \begin{array}{c} \partial f_1 \quad\quad\quad \partial f_n \\ \dfrac{}{\partial x_m} \begin{bmatrix} & \dfrac{\partial f_j}{\partial x_j} & \end{bmatrix}_{m \times n} \end{array}$$

$f: \mathbb{R}^m \to \mathbb{R}$

$$Df = \nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_m} \end{bmatrix}$$

$D(f + \alpha g) = Df + \alpha Dg$

1) $f: \mathbb{R}^n \to \mathbb{R}$

$f(x) = \|x\|_2^2$

$Df = \nabla f = \begin{bmatrix} 2x_1 \\ \vdots \\ 2x_m \end{bmatrix} = 2x$

(2) $f: \mathbb{R}^m \to \mathbb{R}^n$

$f(x) = Ax + b$

$b \in \mathbb{R}^n, \quad A \in \mathbb{R}^{n \times m}$

$Df = A^T$

(3) $f(x): \mathbb{R}^m \to \mathbb{R}$

$f(x) = x^T A x \quad A \in \mathbb{R}^{m \times m}$

$Df = (A + A^T)\, x$

4) $f: \mathbb{R}^2 \to \mathbb{R}$

$f(x_1, x_2) = x_1 \cdot x_2$

$Df = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$

—

$f: \mathbb{R}^m \to \mathbb{R}^n \qquad g \circ f: \mathbb{R}^m \to \mathbb{R}^k$

$g: \mathbb{R}^n \to \mathbb{R}^k$

$D(g \circ f)(x) = \boxed{(Df)(x)}\ (Dg)(f$