$P(\vec{X}) = P(X_1, X_2, \cdots, X_d) \quad \vec{X} \in \mathbb{R}^d$

$P_\theta(X) = P(X; \theta) = P(X|\theta)$

Sampling

$P_\theta(X)$ known $\Rightarrow$ Samples: $X^1, X^2, \cdots, X^m \in \mathbb{R}^d$

Learning

Data $X^1, X^2, \cdots, X^m$ known $\Rightarrow$ find $P_\theta(X)$

- find $P_\theta(X)$: { find θ: parameter learning; find structure + parameters (BN: directed edges; MRF: undirected edges) }

parameter learning (find θ)

find θ that maximizes the probability of occurrence of $\{X^1, X^2, \cdots, X^m\} = D$

- $\theta^* = \underset{\theta}{\operatorname{argmax}} \Pr(X^1, X^2, \cdots, X^n)$

$X^1, X^2, \cdots, X^m$ are i.i.d samples of $P(X)$

$\Pr(X^1, X^2, \cdots, X^n) = \Pr(X^1)\Pr(X^2)\cdots P(X^m)$

independent

$= P_\theta(X^1) P_\theta(X^2) \cdots P_\theta(X^m)$

all are samples from $P_\theta(X)$

$= \prod_{i=1}^{m} P_\theta(X^m) = L(\theta; D)$

parameters; data

$L(\theta; P) = L(\theta)$ likelihood function

$X^1, X^2, \ldots X^m$ $\quad \theta = Pr(X = H)$

H T T H T H

$$Pr(X = X^i) = \begin{cases} \theta & X^i = H \\ 1-\theta & X^i = T \end{cases}$$

$$\mathcal{L}(\theta) = \prod_{i=1}^{m} Pr(X = X^i)$$

$$= \theta^{\#(X^i = H)} (1-\theta)^{\#(X^i = T)}$$

$$= \theta^{n_H} (1-\theta)^{n_T} \qquad n_H + n_T = m$$

Assume $n_H, n_T > 0$

$$\underset{\theta}{\text{maximize}}\ \mathcal{L}(\theta) = \max_{\theta} \theta^{n_H} (1-\theta)^{n_T}$$

$$= \frac{d}{d\theta} \mathcal{L}(\theta) = n_H \theta^{n_H - 1} (1-\theta)^{n_T} - n_T \theta^{n_H} (1-\theta)^{n_T - 1}$$

$$= \theta^{n_H - 1} (1-\theta)^{n_T - 1} \left[n_H (1-\theta) - n_T \theta \right]$$

$\theta \in [0,1] \Rightarrow$ check for $\theta = 1, \emptyset$

$\theta = 0$

$\theta = 1$

$n_H (1-\theta) - n_T \theta = 0 \Rightarrow n_H = \theta (n_H + n_T)$

$$\Rightarrow \theta = \frac{n_H}{n_H + n_T} = \frac{n_H}{m}$$

$\theta = 0, \quad \theta = 1, \quad \theta = \frac{n_H}{m}$

$\mathcal{L}(\theta) = \theta^{n_H} (1-\theta)^{n_T} =$ $\quad \theta = 0 \Rightarrow 0$

$\theta = 1 \Rightarrow 0$

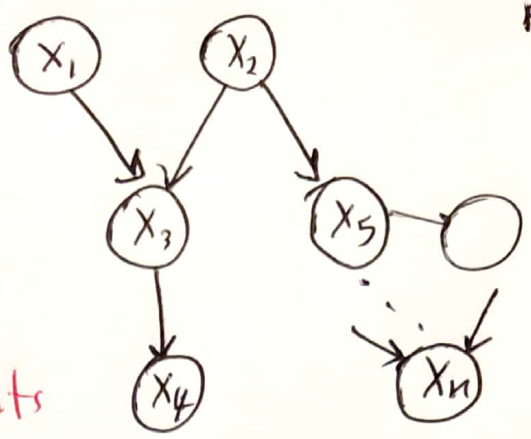
$\theta = \frac{n_H}{m} \Rightarrow \left(\frac{n_H}{m}\right)^{n_H} \left(1 - \frac{n_H}{m}\right)^{n_T} > 0$

if $n_H, n_T > 0$

$$\Rightarrow \boxed{\theta^* = \frac{n_H}{m}}$$

$$P_\theta(X) = P_\theta(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P_\theta(X_i \mid X_{P_i})$$

X_{P_i}: parents of X_i



Data: $X^1, X^2, \dots, X^m = (X_1^1, X_2^1, \dots, X_n^1), (X_1^2, X_2^2, \dots, X_n^2), \dots, (X_1^m, X_2^m, \dots, X_n^m)$

$$ll(\theta) = \sum_{k=1}^{m} \log P_\theta(X^k) = \sum_{k=1}^{m} \log \prod_{i=1}^{n} P_\theta(X_i^k \mid X_{P_i}^k)$$

$$= \sum_{k=1}^{m} \sum_{i=1}^{n} \log P_\theta(X_i^k \mid X_{P_i}^k)$$

1: Each CPD has its own parameters

$$\theta = (\theta_1, \theta_2, \dots, \theta_n)$$

$$P_\theta(X_i \mid X_{P_i}) = P_{\theta_i}(X_i \mid X_{P_i})$$

$$\Rightarrow ll(\theta) = \sum_{k=1}^{m} \sum_{i=1}^{n} \log P_\theta(X_i^k \mid X_{P_i}^k)$$

$$= \sum_{k=1}^{m} \sum_{i=1}^{n} \log P_{\theta_i}(X_i^k \mid X_{P_i}^k)$$

$$= \sum_{i=1}^{n} \underbrace{\sum_{k=1}^{m} \log P_{\theta_i}(X_i^k \mid X_{P_i}^k)}_{\text{local log-likelihood}}$$

$$\frac{\partial}{\partial \theta_j} ll(\theta) = \frac{\partial}{\partial \theta_j} ll(\theta_1, \theta_2, \dots, \theta_n)$$

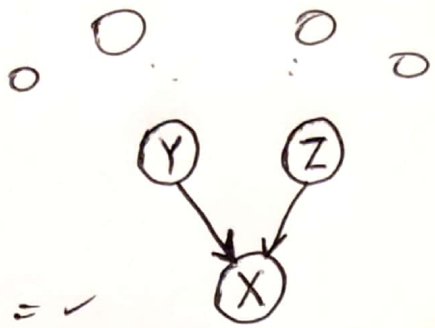
$$= \sum_{k=1}^{m} \frac{\partial}{\partial \theta_j} \log P_{\theta_j}(X_i^k \mid X_{P_i}^k)$$

$$= \sum_{k=1}^{m} \frac{\frac{\partial}{\partial \theta_j} P_{\theta_j}(X_i^k \mid X_{P_i}^k)}{P_{\theta_j}(X_i^k \mid X_{P_i}^k)}$$

gradient w.r.t. θ_j

$\Rightarrow$ each θ_j can be found independently of other θ_i's

no share parameters among CPDs
+
table representation



$X \in \{1, 2, \ldots, c\}$

$Y, Z \in \{0, 1\}$

$\gamma_{00}^{i} = \checkmark$

$\gamma_{01}^{i} = Pr(X = i \mid Y = 0, Z = 1)$

$\gamma_{10}^{i} = Pr(X = i \mid Y = 1, Z = 0)$

$\gamma_{11}^{i} = Pr(X = i \mid Y = 1, Z = 1)$

$\sum_{i} \gamma_{10}^{i} = 1$

local likelihood $\sum_{k=1}^{m} \log P_{\gamma}(X^k \mid Y^k, Z^k)$

$$\sum_{k=1}^{m} \log \gamma_{Y^k, Z^k}^{X^k}$$

$$\sum_{y=0}^{1} \sum_{z=0}^{1} \sum_{x=1}^{c} (\log \gamma_{yz}^{x}) \cdot \#(X^k = x, Y^k = y, Z^k = z)$$

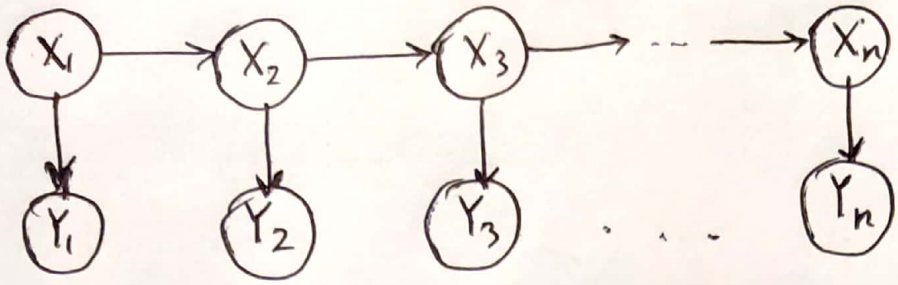
sufficient statistics

$$\boxed{\sum_{x} \gamma_{yz}^{x} = 1}$$

$$\gamma_{yz}^{x} = \frac{\#(X^k = x, Y^k = y, Z^k = z)}{\#(Y^k = y, Z^k = z)}$$

$$= \frac{\sum_{k=1}^{m} 1(X^k = x, Y^k = y, Z^k = z)}{\sum_{k=1}^{m} 1(Y^k = y, Z^k = z)}$$

Shared Parameters



$$P_\theta(X_1 - X_n, Y_1 - Y_n) = P_\alpha(X_1) \prod_{i=2}^{n} P_\beta(X_i \mid X_{i-1}) \prod_{i=1}^{n} P_\gamma(Y_i \mid X_i)$$

$\theta = (\alpha, \beta, \gamma)$

$$ll(\theta) = \sum_{k=1}^{m} \log P_\alpha(X_1^k) + \sum_{k=1}^{m} \sum_{i=2}^{n} \log P_\beta(X_i^k \mid X_{i-1}^k) + \sum_{k=1}^{m} \sum_{i=1}^{n} \log P_\gamma(Y_i^k \mid X_i^k)$$

$$\frac{\partial ll(\theta)}{\partial \beta} = \sum_{k=1}^{m} \sum_{i=2}^{n} \frac{\partial}{\partial \beta} \log P_\beta(X_i^k \mid X_{i-1}^k)$$

table representation $X_i \in \{0,1\}$

$\beta_0 = \beta_{00}, \beta_{10}, \beta_{11}, \beta_{01}$

$\beta_{01} = P(X_i = 0 \mid X_{i-1} = 1)$ independent of i

ML solution

$$\beta_{01}^* = \frac{\#(X_i^k = 0, X_{i-1}^k = 1)}{\#(X_{i-1}^k = 1)}$$

$$= \frac{\sum_{i=2}^{n} \sum_{k=1}^{m} 1(X_i^k = 0, X_{i-1}^k = 1)}{\sum_{i=2}^{n} \sum_{k=1}^{m} 1(X_{i-1}^k = 1)}$$