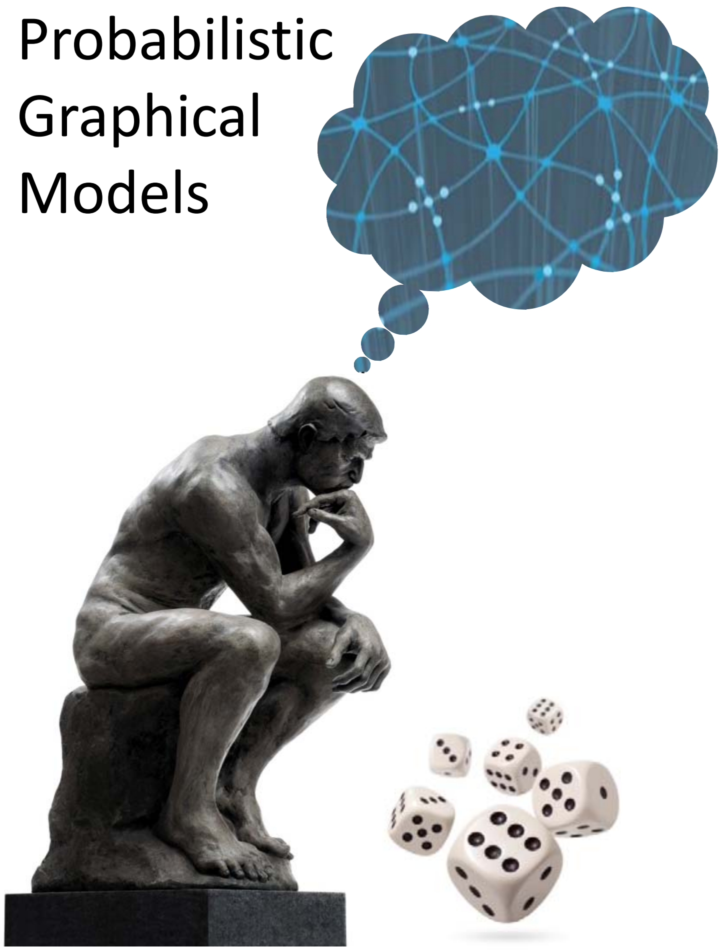
Probabilistic
Graphical
Models

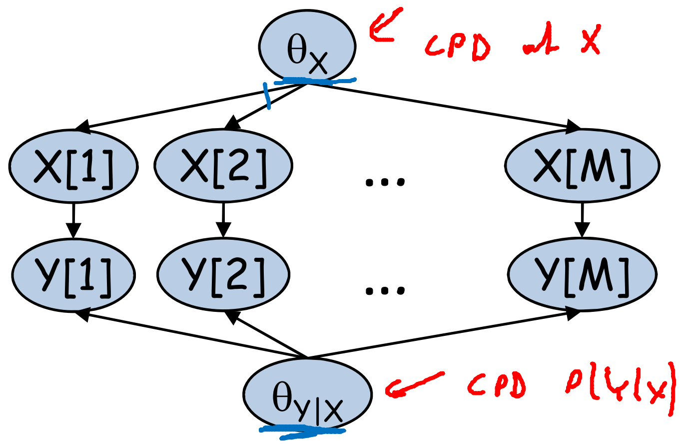


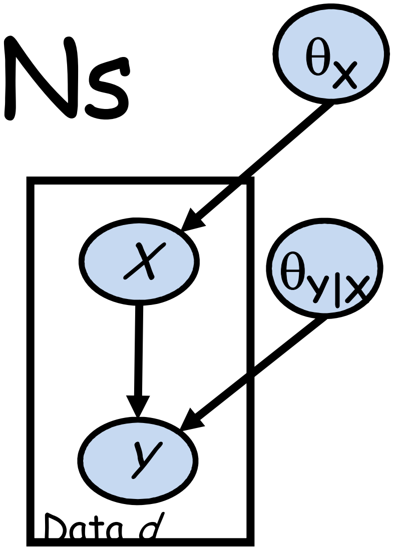
Learning

Parameter Estimation

Bayesian Estimation for BNs

Bayesian Estimation in BNs

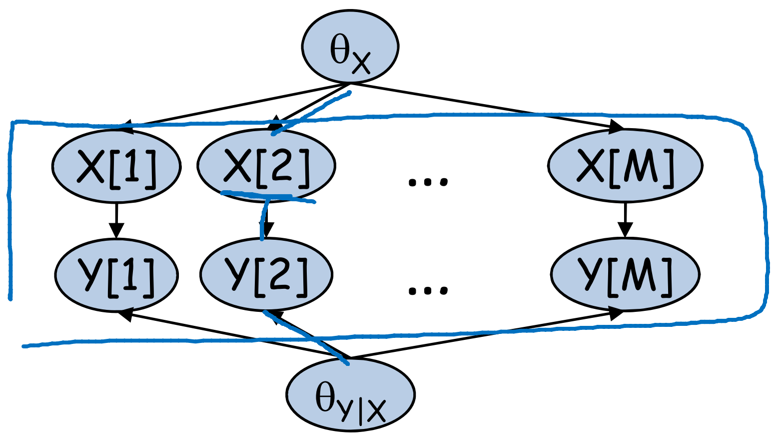


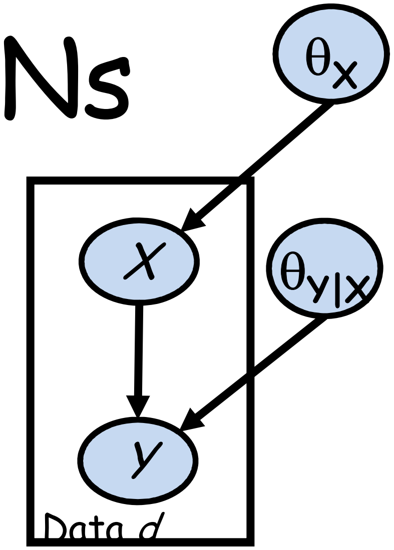


- Instances are independent given the parameters
  - (X[m'],Y[m']) are d-separated from (X[m],Y[m]) given θ
- Parameters for individual variables are independent a priori $P(\theta) = \prod_i P(\theta_{X_i|Pa(X_i)})$

Daphne Koller

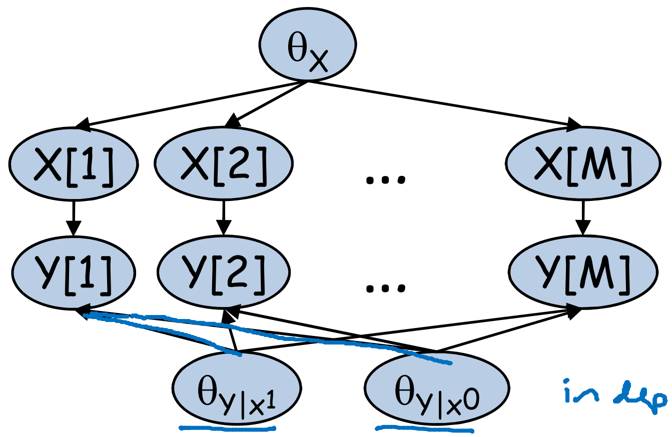
Bayesian Estimation in BNs

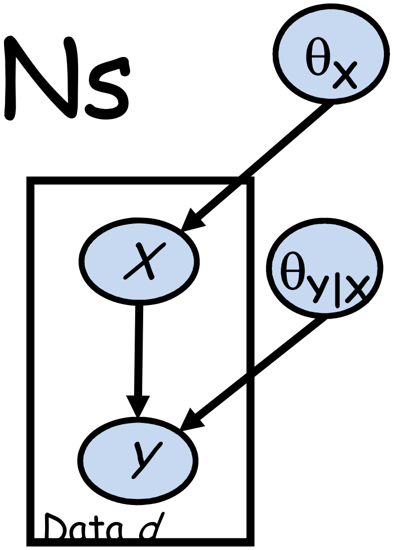




- Posteriors of θ are independent given complete data
  - Complete data d-separates parameters for different CPDs
  - $P(\theta_X, \theta_{Y|X} \mid D) = P(\theta_X \mid D) P(\theta_{Y|X} \mid D)$
  - As in MLE, we can solve each estimation problem separately

Bayesian Estimation in BNs





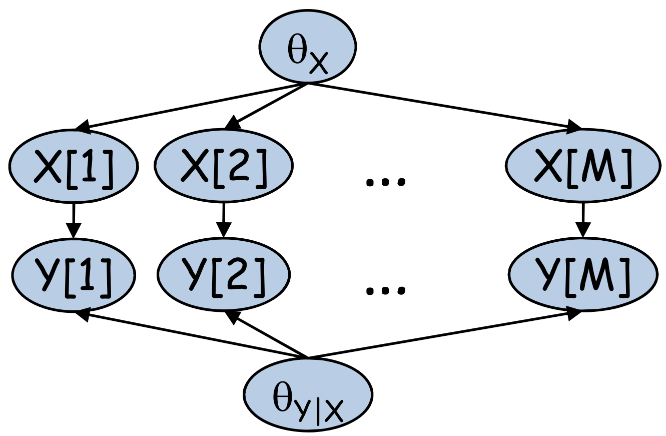
$P(\theta_X \mid D) \cdot P(\theta_{Y|x} \mid D) \cdot P(\theta_{Y|x} \mid D)$

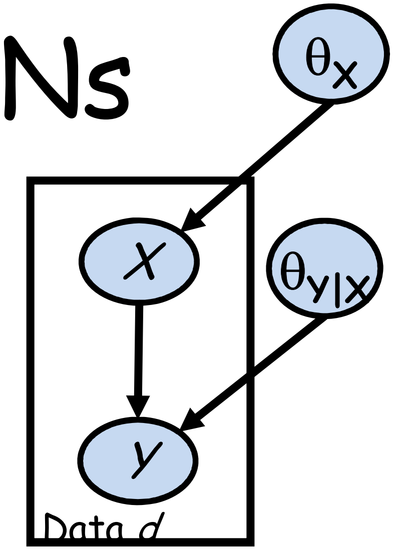
independent a priori

- Posteriors of θ are independent given complete data
  - Also holds for parameters within families
  - Note context specific independence between $\theta_{Y|x^1}$ and $\theta_{Y|x^0}$ when given both X's and Y's

Daphne Koller

Bayesian Estimation in BNs



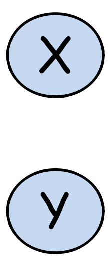


- Posteriors of θ can be computed independently
  - assignment to x's parents u
  - For multinomial $\theta_{X|\mathbf{u}}$ if prior is Dirichlet$(\alpha_{x^1|\mathbf{u}},\ldots,\alpha_{x^k|\mathbf{u}})$
  - posterior is Dirichlet$(\alpha_{x^1|\mathbf{u}}+M[x^1,\mathbf{u}],\ldots,\alpha_{x^k|\mathbf{u}}+M[x^k,\mathbf{u}])$

Daphne Koller

Assessing Priors for BNs

- We need hyperparameter $\alpha_{x|\mathbf{u}}$ for each node X, value x, and parent assignment **u**
  - Prior network with parameters Θ_0
  - Equivalent sample size parameter α
  - $\alpha_{x|\mathbf{u}} := \alpha \cdot P(x,\mathbf{u}|\Theta_0)$



Θ_0 uniform

$\theta_X \sim \text{Dirichlet}(\frac{\alpha}{2}, \frac{\alpha}{2})$

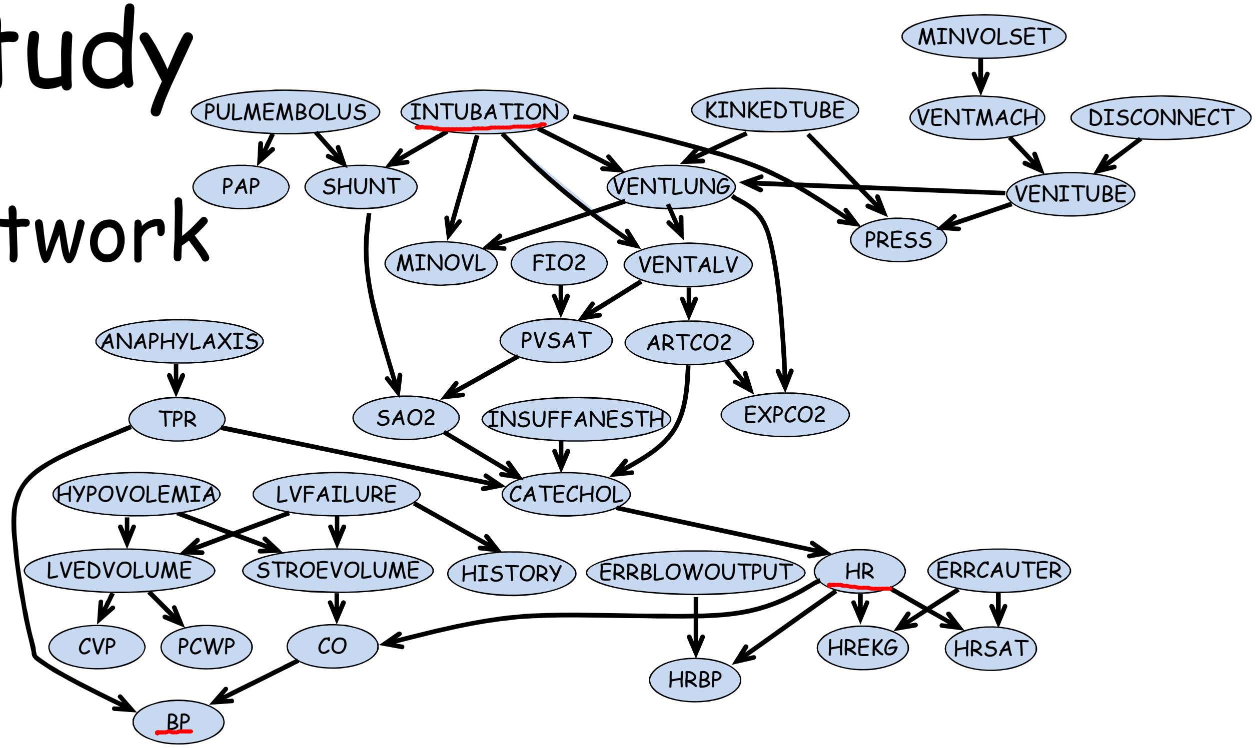
$\theta_{Y|x^0} \sim \text{Dirichlet}(\alpha/4, \alpha/4)$

$\theta_{Y|x^1} \sim \text{Dirichlet}(\alpha/4, \alpha/4)$

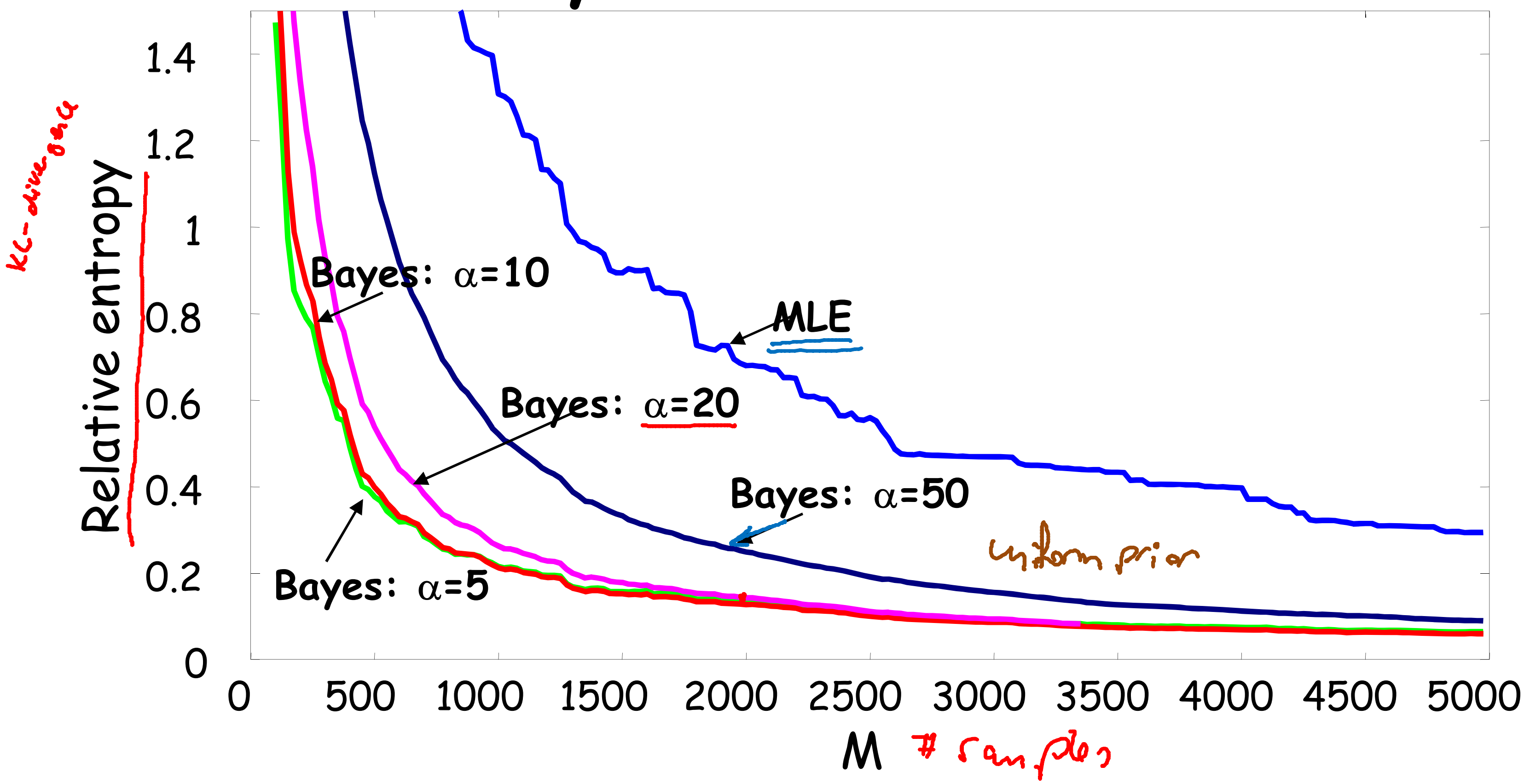
Daphne Koller

Case Study

- ICU-Alarm network
  - 37 variables
  - 504 params
- Experiment
  - Sample instances from network
  - Relearn parameters



Case Study: ICU Alarm Network



Daphne Koller

Summary

- In Bayesian networks, if parameters are independent a priori, then also independent in the posterior
- For multinomial BNs, estimation uses sufficient statistics M[x,**u**]

$$\underbrace{\hat{\theta}_{x|\boldsymbol{u}} = \frac{M[x,\boldsymbol{u}]}{M[\boldsymbol{u}]}}$$

MLE

$$P(x \mid \boldsymbol{u}, D) = \frac{\alpha_{x,\boldsymbol{u}} + M[x,\boldsymbol{u}]}{\underline{\alpha_{\boldsymbol{u}} + M[\boldsymbol{u}]}}$$

Bayesian (Dirichlet)

- Bayesian methods require choice of prior
  – can be elicited as prior network and equivalent sample size α