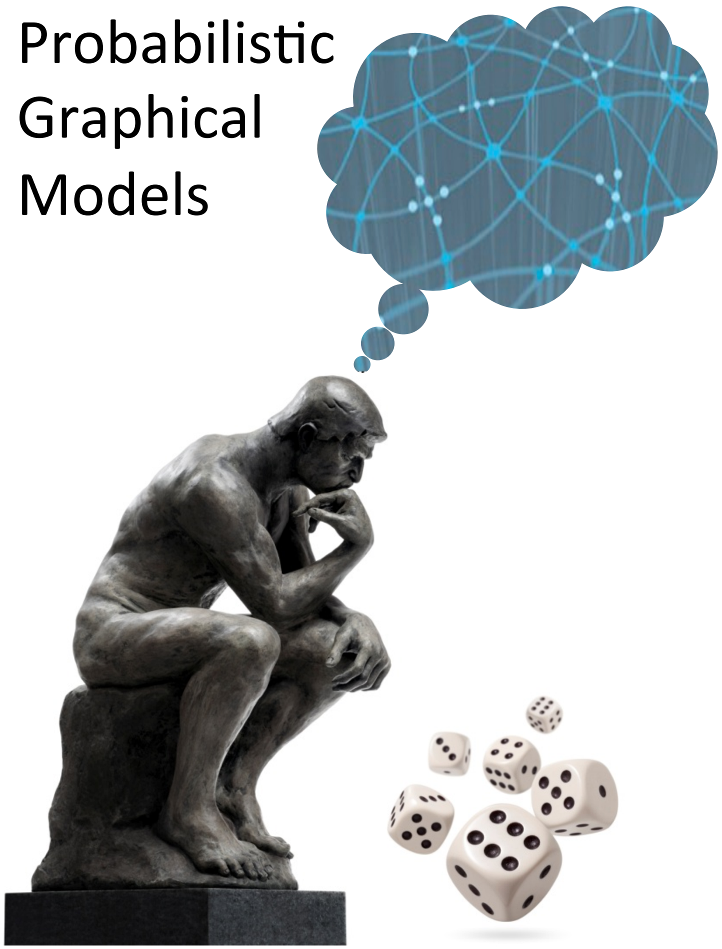
Probabilistic
Graphical
Models



Learning

BN Structure

BIC Score and Asymptotic Consistency

Penalizing Complexity

Bayesian information criterion

\# training instances

$$\text{score}_{BIC}(\mathcal{G} : \mathcal{D}) = \underbrace{\ell(\hat{\boldsymbol{\theta}}_{\mathcal{G}} : \mathcal{D})}_{\text{score}_L(\mathcal{G} : \mathcal{D})} - \frac{\log M}{2}\text{Dim}[\mathcal{G}]$$

\# independent params

- Tradeoff between fit to data and model complexity

(MDL criterion) ← minimum description length

Asymptotic Behavior

$$\ell(\hat{\boldsymbol{\theta}}_{\mathcal{G}} : \mathcal{D}) - \frac{\log M}{2}\mathrm{Dim}[\mathcal{G}]$$

=

$$\boxed{M\sum_{i=1}^{n}\mathbf{I}_{\hat{P}}(X_i;\mathbf{Pa}^{G}_{X_i})} - \underbrace{M\sum_{i}\boldsymbol{H}_{\hat{P}}(X_i)}_{\text{independent of } G} - \boxed{\frac{\log M}{2}\mathrm{Dim}[\mathcal{G}]}$$

- Mutual information grows linearly with M while complexity grows logarithmically with M
  - As M grows, more emphasis is given to fit to data

Consistency

$\hat{P} \to P^*$ G^*

$$\to M \sum_{i=1}^{n} \mathbf{I}_{\hat{P}}(X_i; \mathbf{Pa}_{X_i}^{G}) - M \sum_{i} \boldsymbol{H}_{\hat{P}}(X_i) - \frac{\log M}{2} \mathrm{Dim}[\mathcal{G}]$$

$\to \sim \mathbf{I}_{P^*}(X_i ; Pa_{X_i}^{G})$

- As M→∞, the true structure G* (or any I-equivalent structure) maximizes the score
  – Asymptotically, spurious edges will not contribute to likelihood and will be penalized
  – Required edges will be added due to linear growth of likelihood term compared to logarithmic growth of model complexity

Summary

- BIC score explicitly penalizes model complexity (# of independent parameters)
  - Its negation often called MDL
- BIC is asymptotically consistent:
  - If data generated by G^*, networks I-equivalent to G^* will have highest score as M grows to ∞