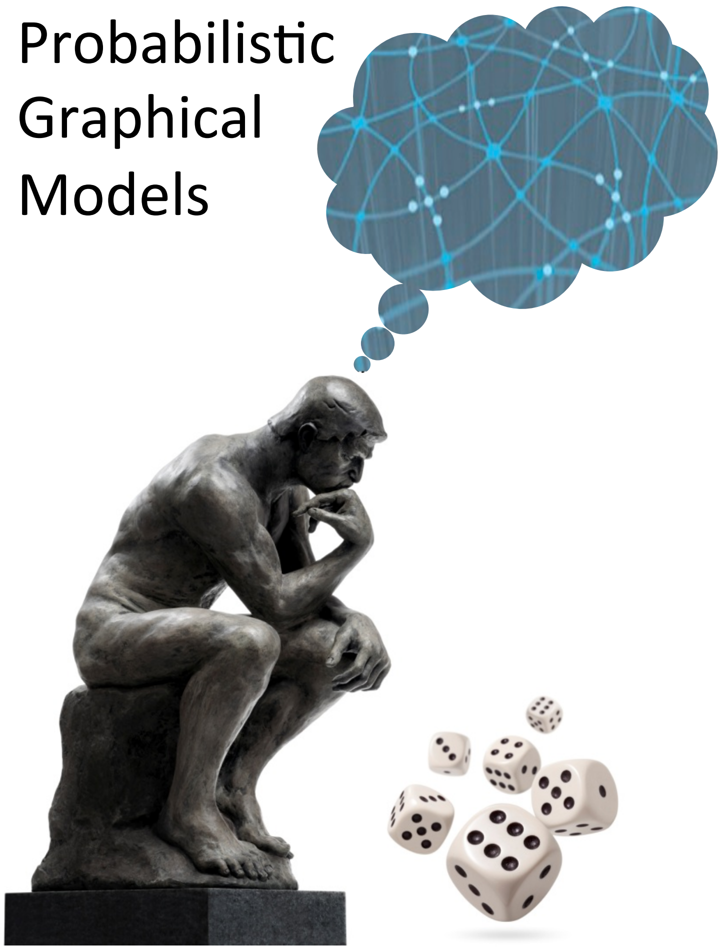
Probabilistic
Graphical
Models



Inference

Sampling Methods

MCMC for PGMs: The Gibbs Chain

Gibbs Chain

- Target distribution $P_\Phi(X_1,\ldots,X_n)$
- Markov chain state space: complete assignments $\mathbf{x}$ to $\mathbf{X} = \{X_1,\ldots,X_n\}$
- Transition model given starting state $\mathbf{x}$:
  – For $i=1,\ldots,n$
    - Sample $x_i \sim P_\Phi(X_i \mid \mathbf{x}_{-i})$

      assignmnt to all $X_1 \ldots X_n$ except X_i
  – Set $\mathbf{x}' = \mathbf{x}$

X_1	X_2	X_3
0	0	0
1	0	0
1	0	0
1	0	1

$p(X_1 \mid X_2=0, X_3=0)$
$p(X_2 \mid X_1=1, X_3=0)$
$p(X_3 \mid X_1=1, X_2=0)$

Example

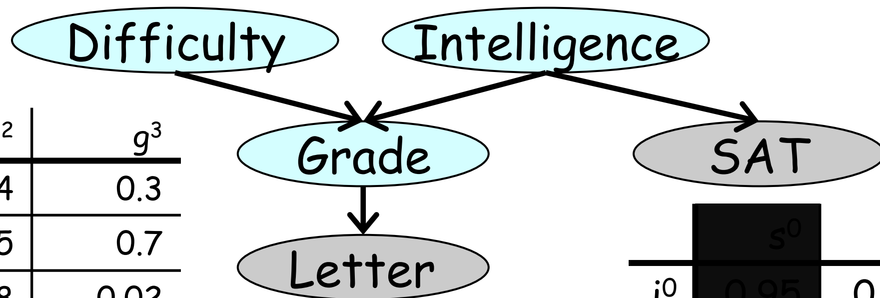
$P(D \mid i^0, g^{*}, l^0, s^1)$

$P_{\Phi}(D, I, G \mid s^1, l^0)$

$P(I \mid d^1, g^{*}, l^0, s^1)$

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

$P(G \mid d^1, i^1, l^0, s^1)$

$d^0 \; i^0 \; g^{*}$

$d^1 \; i^0 \; g^{*}$

$d^1 \; i^1 \; g^{*}$

$\rightarrow d^1 \; i^1 \; g^3$

	l^0	l^1
g^1	0.1	[redacted]
g^2	0.4	[redacted]
g^3	0.99	[redacted]

	s^0	s^1
i^0	[redacted]	0.05
i^1	[redacted]	0.8

Computational Cost

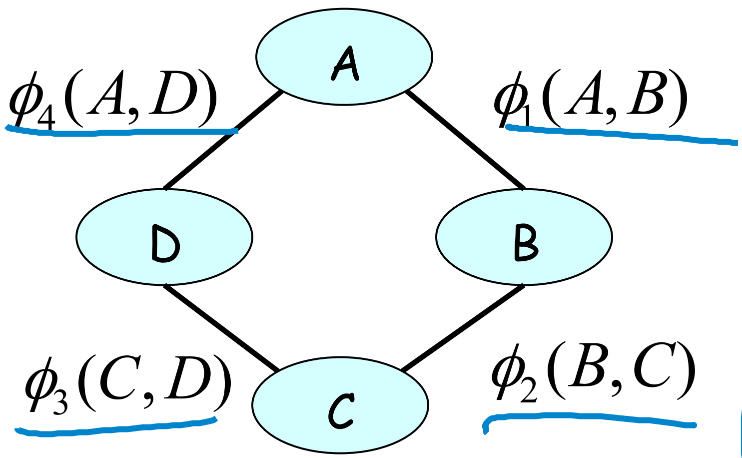
- For i=1,…,n
  - Sample $x_i \sim P_\Phi(X_i \mid \mathbf{x}_{-i})$

$$\underline{P_\Phi(X_i \mid \boldsymbol{x}_{-i})} = \frac{P_\Phi(\underline{X_i}, \underline{\boldsymbol{x}_{-i}})}{P_\Phi(\underline{\boldsymbol{x}_{-i}})} = \frac{\cancel{\frac{1}{Z}}\boxed{\tilde{P}_\Phi(X_i, \boldsymbol{x}_{-i})}}{\cancel{\frac{1}{Z}}\tilde{P}_\Phi(\boldsymbol{x}_{-i})}$$

complete assignment

product of factors

Another Example



$$P_\Phi(A \mid b,c,d) = \frac{\tilde{P}_\Phi(a,b,c,d)}{\underline{\sum_{A'} \tilde{P}_\Phi(A',b,c,d)}}$$

$$\frac{\phi_1(A,b)\cancel{\phi_2(b,c)}\cancel{\phi_3(c,d)}\phi_4(A,d)}{\underbrace{\sum_{A'} \phi_1(A',b)\cancel{\phi_2(b,c)}\cancel{\phi_3(c,d)}\phi_4(A',d)}_{\text{normalizing constant}}}$$

factors that involve A

$\propto \phi_1(A,b)\ \phi_4(A,d)$

Daphne Koller

Computational Cost Revisited

- For i=1,…,n
  - Sample $x_i \sim P_\Phi(X_i \mid \boldsymbol{x}_{-i})$

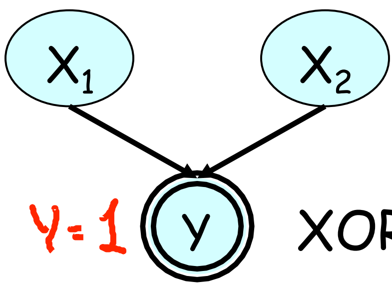
$$P_\Phi(\underline{X_i} \mid \underline{\boldsymbol{x}_{-i}}) = \frac{P_\Phi(X_i, \boldsymbol{x}_{-i})}{P_\Phi(\boldsymbol{x}_{-i})} = \frac{\tilde{P}_\Phi(X_i, \boldsymbol{x}_{-i})}{\tilde{P}_\Phi(\boldsymbol{x}_{-i})}$$

only X_i and its neighbors

$$\boxed{\propto \prod_{j : X_i \in Scope[C_j]} \phi_j(X_i, \boldsymbol{x}_{j,-i})}$$

factors that involve X_i

Gibbs Chain and Regularity



XOR

X_1	X_2	Y	Prob
~~0~~	~~0~~	~~0~~	~~0.25~~
0	1	1	0.25
1	0	1	0.25
~~1~~	~~1~~	~~0~~	~~0.25~~

- If all factors are positive, Gibbs chain is regular
- However, mixing can still be very slow

Summary

- Converts the hard problem of inference to a sequence of "easy" sampling steps
- Pros:
  - Probably the simplest Markov chain for PGMs
  - Computationally efficient to sample
- Cons:
  - Often slow to mix, esp. when probabilities are peaked
  - Only applies if we can sample from product of factors