# Recitation: Bayes Nets and Friends


## Lyle Ungar and Tony Liu

*Heavily adapted from slides by Mitch Marcus*


# Recitation Plan

- Naïve Bayes Exercise
- LDA Example
- Bayes Net Exercises
- HMM Example

# Recall: Naïve Bayes

# Recall: Naïve Bayes

- **Conditional independence assumption**
- **As MAP estimator (uses prior for smoothing)**
  - Contrast MLE – what's the problem?

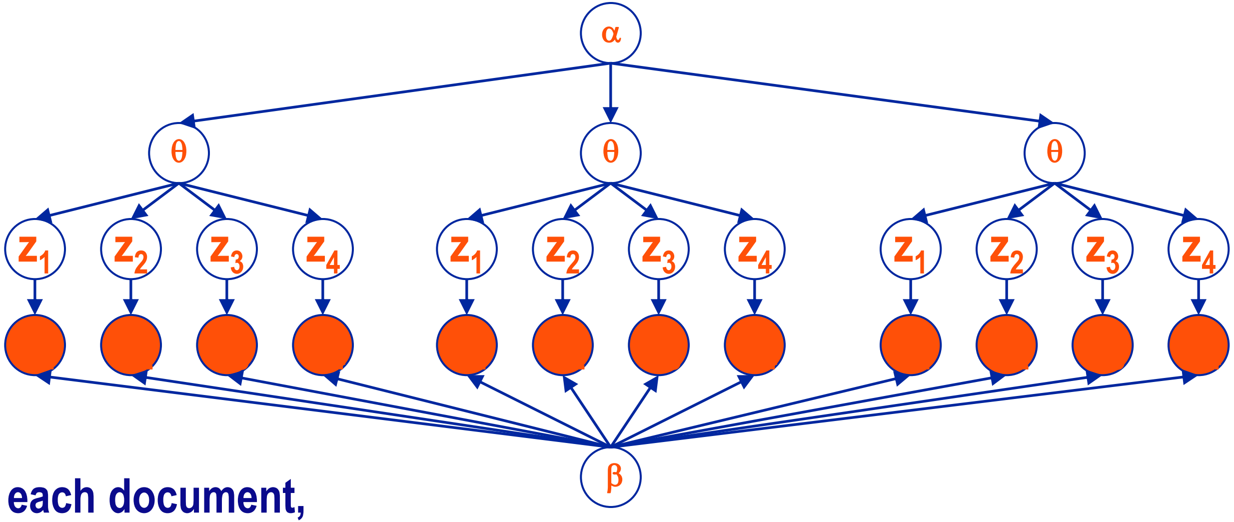
# Naïve Bayes Exercise

Consider the two class problem where class label $Y \in \{T, F\}$ and each training example X has 2 binary attributes $X1, X2 \in \{T, F\}$. How many parameters will you need to know/estimate if you are to classify an example using the Naive Bayes classifier?

- 3 A
- 5 B
- 7 C
- 8 D

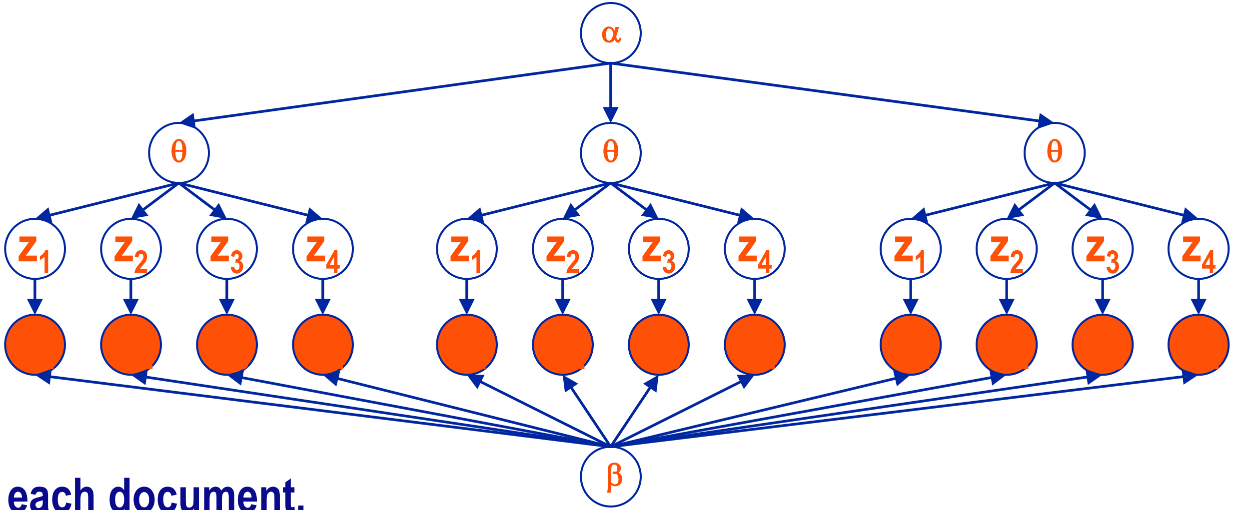
# Recall: The LDA Model

# Recall: The LDA Model



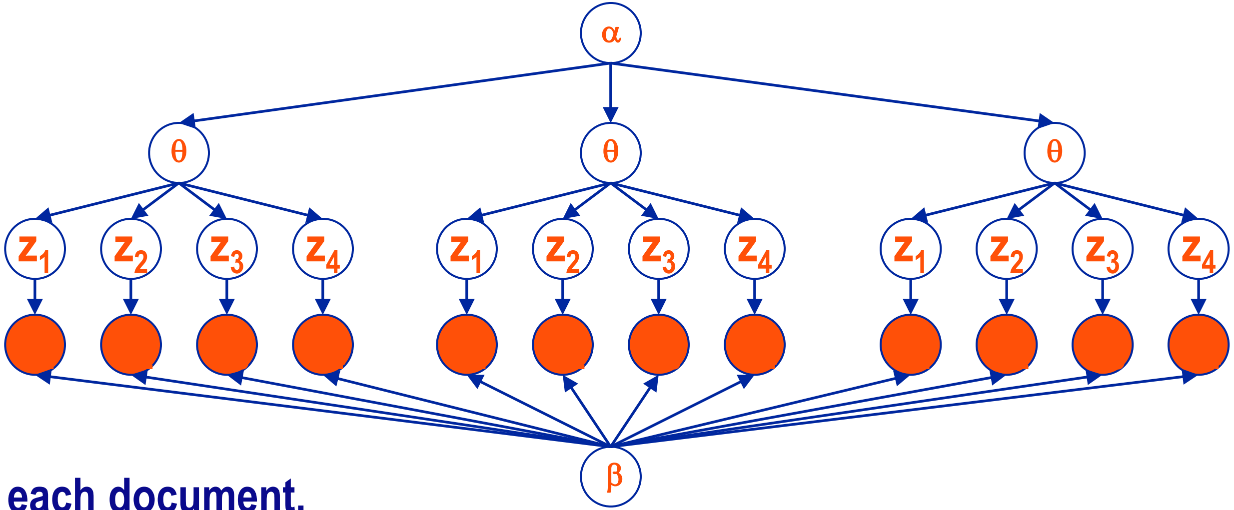

- For each document,

# Recall: The LDA Model



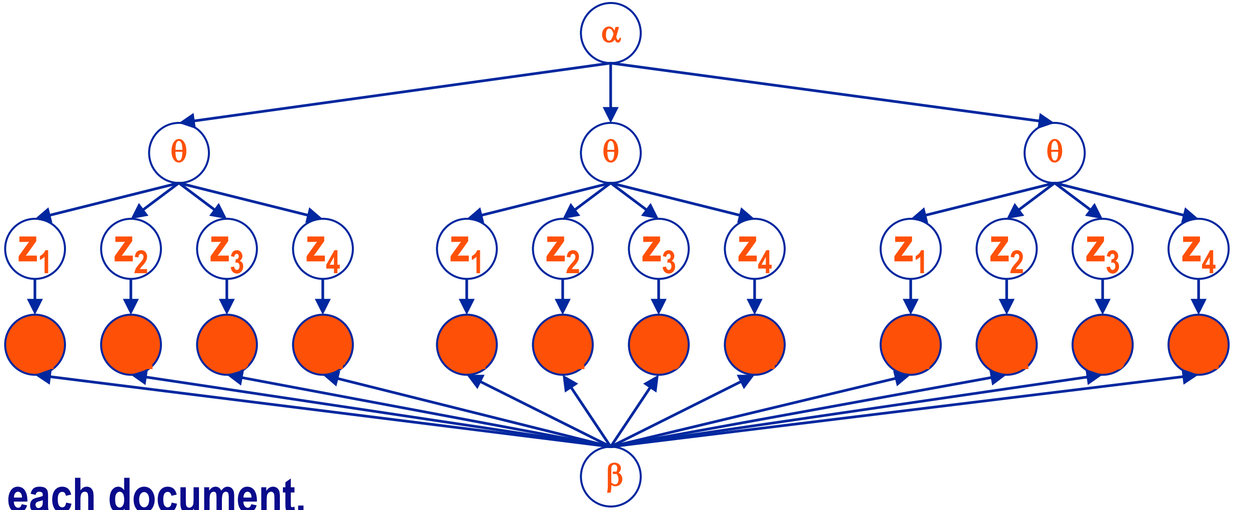

- **For each document,**
  - Choose the topic distribution $\theta \sim \text{Dirichlet}(\alpha)$

# Recall: The LDA Model



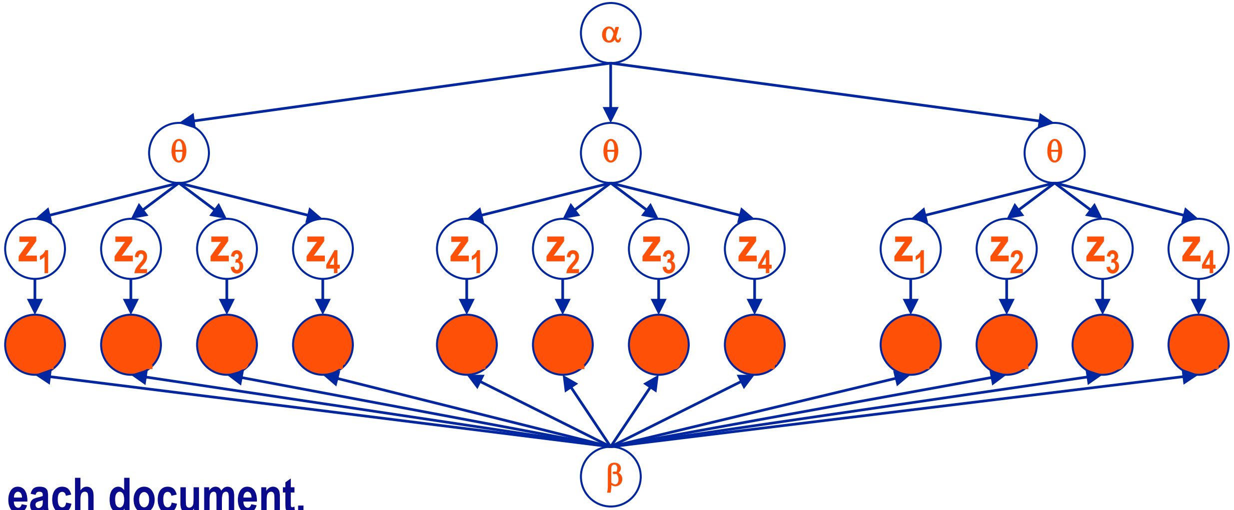

- **For each document,**
  - Choose the topic distribution $\theta \sim \text{Dirichlet}(\alpha)$
  - For each of the N words $w_n$:

# Recall: The LDA Model

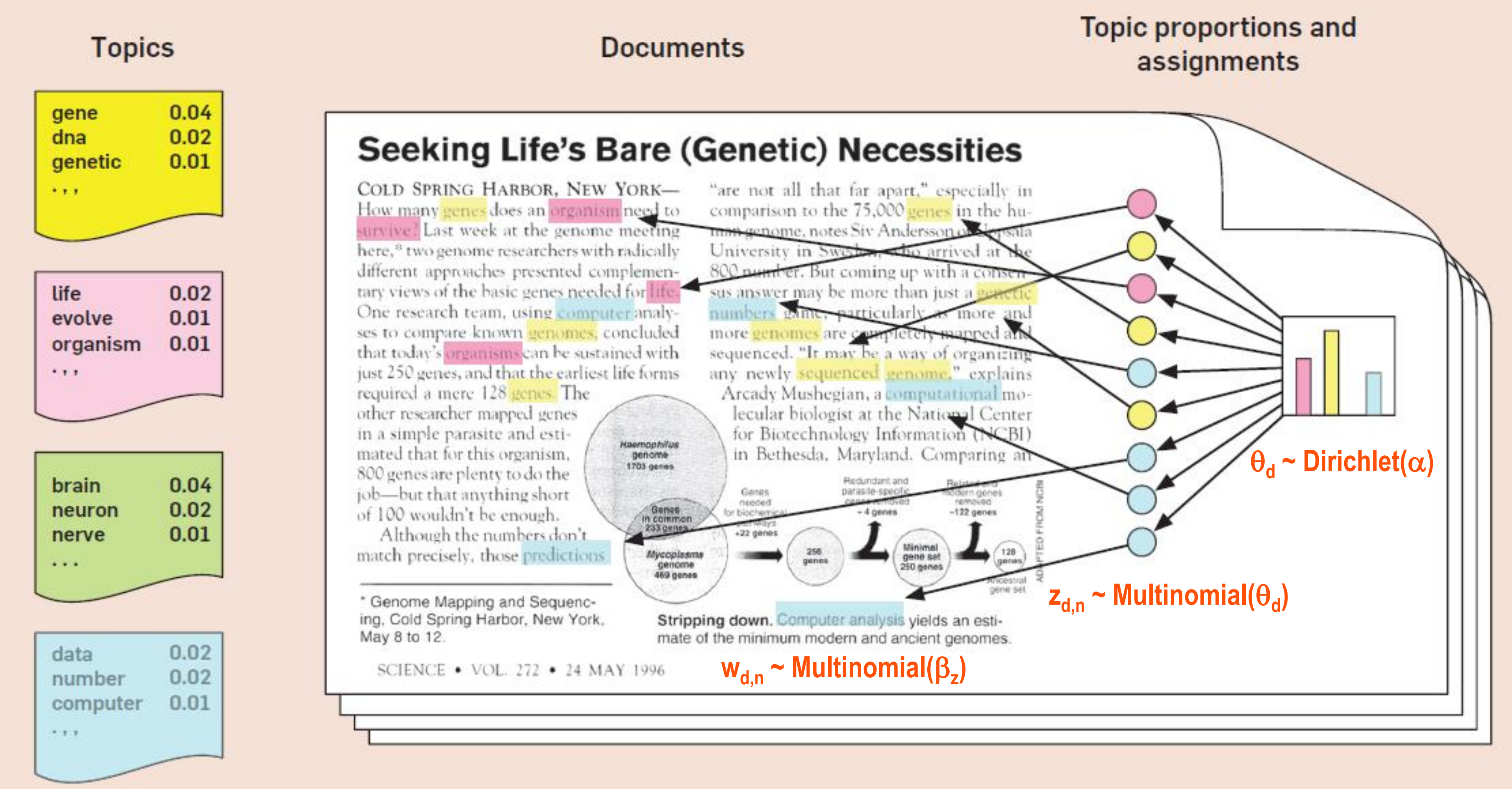



- **For each document,**
  - Choose the topic distribution $\theta \sim \text{Dirichlet}(\alpha)$
  - For each of the N words $w_n$:
    - Choose a topic $z \sim \text{Multinomial}(\theta)$

# Recall: The LDA Model




- **For each document,**
  - Choose the topic distribution $\theta \sim \text{Dirichlet}(\alpha)$
  - For each of the N words $w_n$:
    - Choose a topic $z \sim \text{Multinomial}(\theta)$
    - Then choose a word $w_n \sim \text{Multinomial}(\beta_z)$
      - Where each topic has a different parameter vector $\beta$ for the words




David Blei 2012: Probabilistic Topic Models (on course website)

# LDA Parameter Estimation

- Given a corpus of documents, find the parameters $\alpha$ and $\beta$ which maximize the likelihood of the observed data (words in documents), marginalizing over the hidden variables $\theta$, z

**$\theta$: topic distribution for the document,**
**z: topic for each word in the document**

- **E-step:**
  - Compute $p(\boldsymbol{\theta},\mathbf{z}|\mathbf{w},\alpha,\beta)$, the posterior of the hidden variables $(\theta,\mathbf{z})$ given each document **w**, and parameters $\alpha$ and $\beta$.
- **M-step**
  - Estimate parameters $\alpha$ and $\beta$ given the current hidden variable distribution estimates

**You don't need to know the details;**
**Only what is hidden and what is observed;**
**And that EM works here.**

# LDA Exercise

**True or False? In LDA, the words in each document are assumed to be drawn from a Dirichlet distribution. These distributions can vary across documents.**

True **A**

False **B**

# Recall: Bayes Nets

# Recall: Bayes Nets

- **Local Markov Assumption**
  - given its parents, each node is conditionally independent of everything except its descendants
- **Active Trails**
- **D Separation**

# Active Trails

A trail $\{X_1, X_2, \cdots, X_k\}$ in the graph (no cycles) is an **active trail** if for each consecutive triplet in the trail:

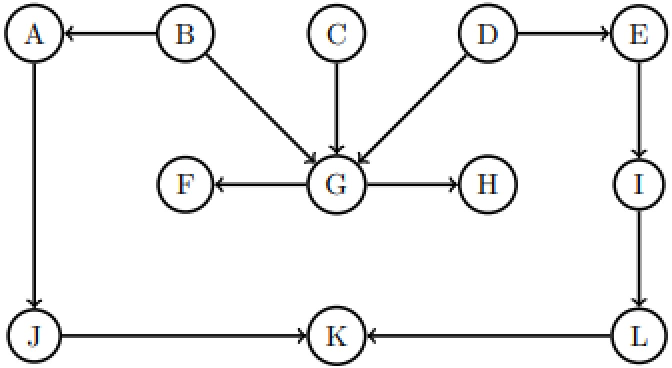
- $X_{i-1} \rightarrow X_i \rightarrow X_{i+1}$, and $X_i$ is not observed
  $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$, and $X_i$ is not observed
  $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$, and $X_i$ is not observed
  $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, and $X_i$ is observed or one of its descendants is observed

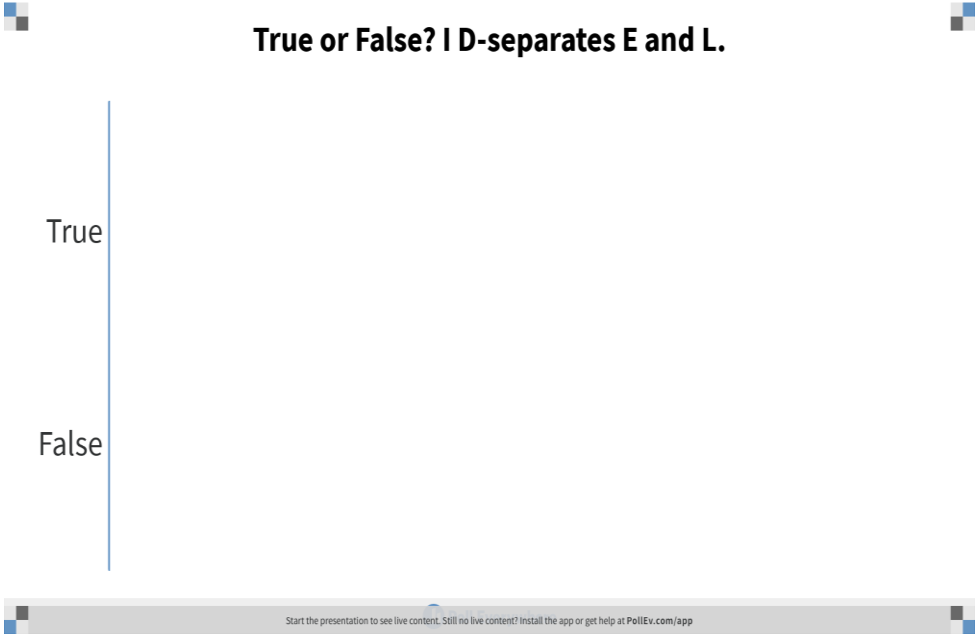
**Variables connected by active trails are not conditionally independent**

# D-separation

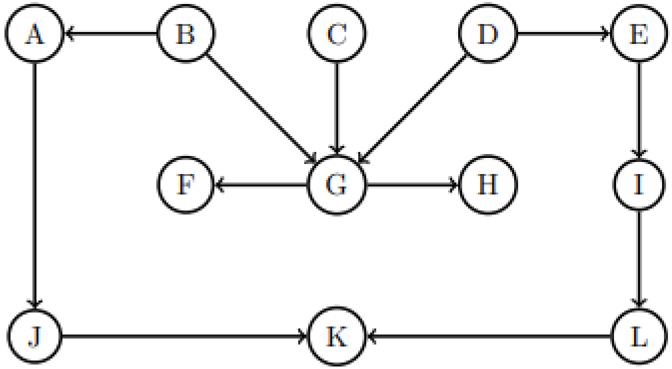
- Variables $X_i$ and $X_j$ are independent if there is no *active trail* between $X_i$ and $X_j$.
  - given a set of observed variables $O \subset \{X_1, \cdots, X_m\}$
  - O sometimes called a Markov Blanket

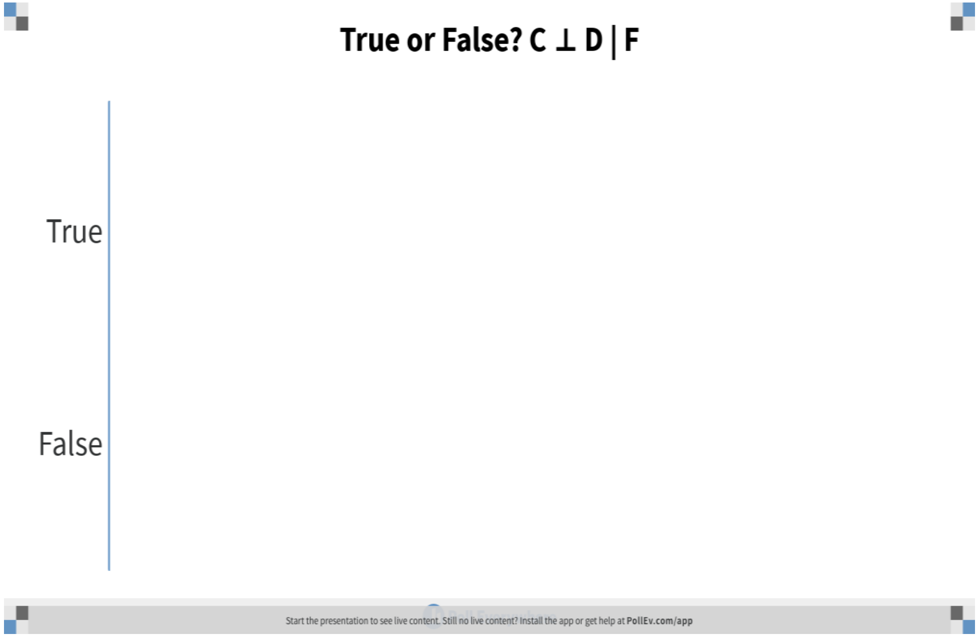
# Bayes Net Exercises



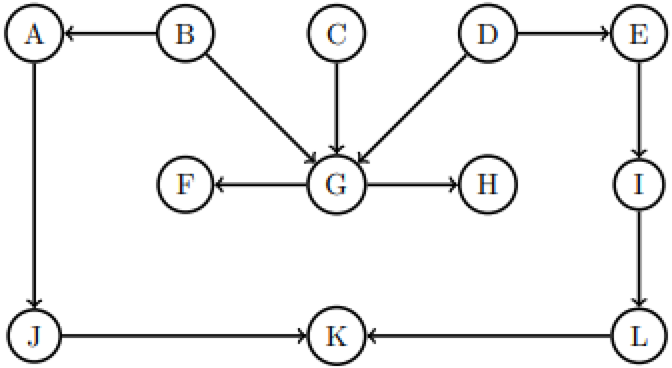



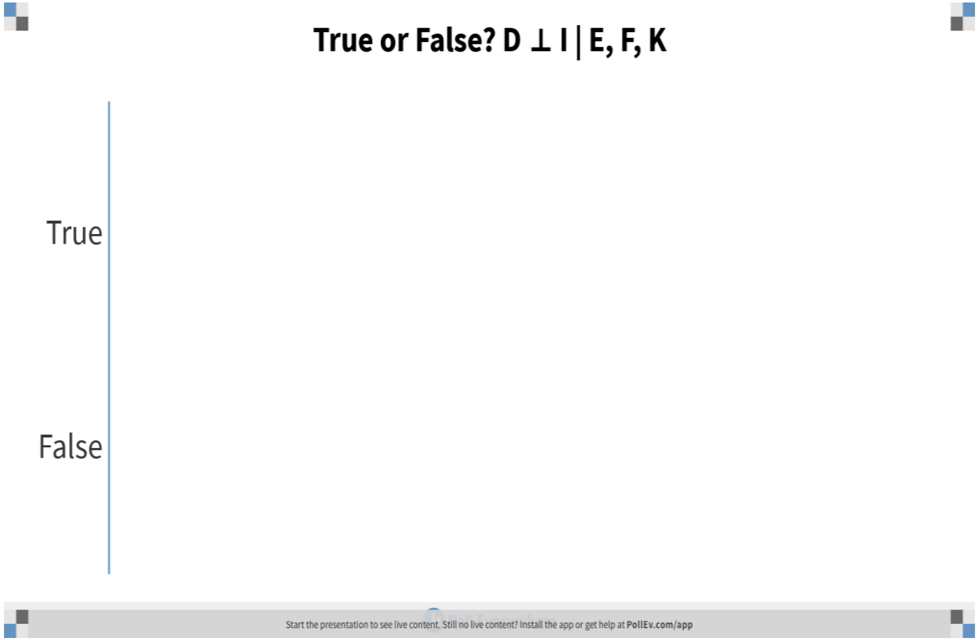
# Bayes Net Exercises



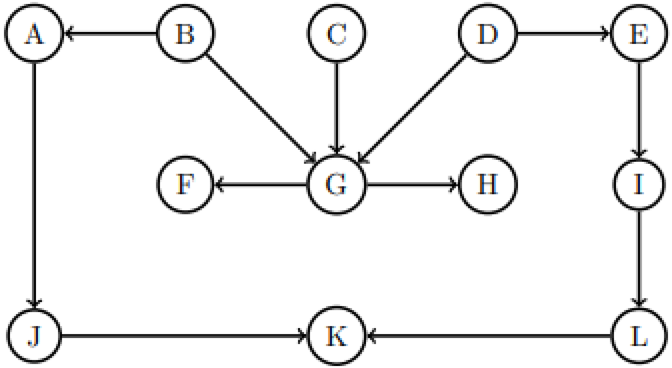



# Bayes Net Exercises



# Bayes Net Exercises




**What is the minimum number of parameters needed to represent the full joint probability P(A, B, C, D, E, F, G, H, I, J, K, L) in the above network if all the variables are binary?**

4095

20

23

24

29

# Recall: Hidden Markov Models

# Recall: Hidden Markov Models

- Markov assumption
- Parameterization

# Parameters of an HMM

- ***States***: A set of states $S = s_1, \ldots, s_k$
- ***Markov transition probabilities***: $A = a_{1,1}, a_{1,2}, \ldots, a_{k,k}$ Each $a_{i,j} = p(s_j \mid s_i)$ represents the probability of transitioning from state $s_i$ to $s_j$.
- ***Emission probabilities***: A set $B$ of functions of the form $b_i(o_t) = p(o|s_i)$ giving the probability of observation $o_t$ being emitted by $s_i$
- ***Initial state distribution***: the probability that $s_i$ is a start state

# Markov Model Example

$S_1 = [0, 1]$

| | Tomorrow's Weather | | |
|---|---|---|---|
| Today's Weather | | Sunny | Rainy |
| | Sunny | 0.8 | 0.2 |
| | Rainy | 0.6 | 0.4 |

**Markov Transition Matrix A**

**Steady state at [0.75, 0.25]**

# Hidden Markov Model Example

$S_1 = [0.5, 0.5]$

We observe:
(umbrella, no umbrella)

We can ask questions like:
- What is the joint probability of the states (rain, sun) and our observations?

| | Tomorrow's Weather | |
|---|---|---|
| Today's Weather | Sunny | Rainy |
| Sunny | 0.8 | 0.2 |
| Rainy | 0.6 | 0.4 |

Markov Transition Matrix A

| | Weather | |
|---|---|---|
| | Sunny | Rainy |
| Umbrella | 0.1 | 0.8 |
| No Umbrella | 0.9 | 0.2 |

Emission Probabilities B

# HMM Exercise

$$P(O_{t+1} = o_{t+1}, ..., O_T = o_T | O_1 = o_1, ..., O_t = o_t, S_t = k)$$
$$= P(O_{t+1} = o_{t+1}, ..., O_T = o_T | S_t = k)$$

**True or False? The above statement about hidden Markov models holds for all 1 ≤ t ≤ T and k.**

True **A**

False **B**