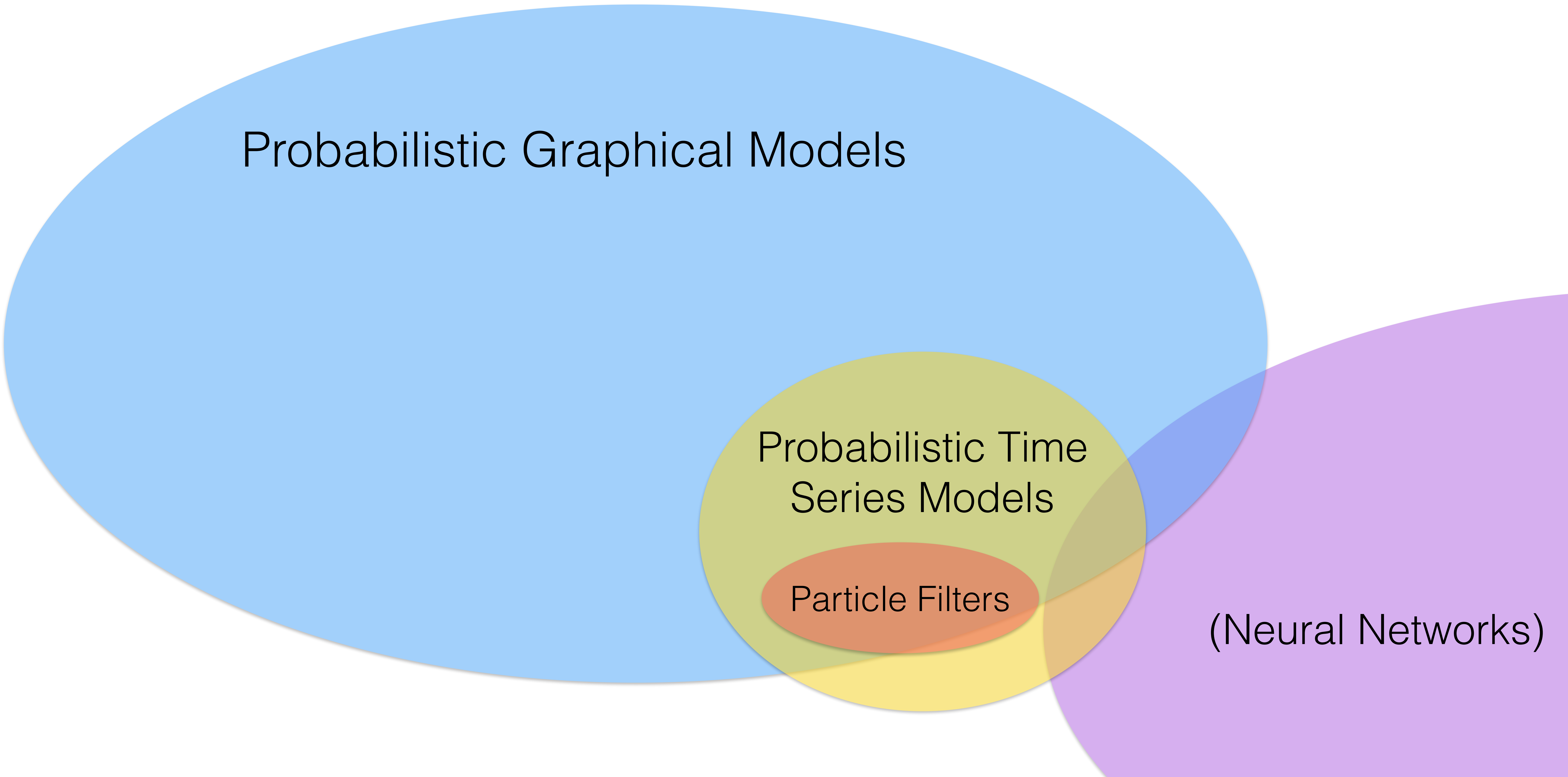
Probabilistic Graphical Models and Bayesian Networks

Artificial Intelligence
Bert Huang
Virginia Tech

Concept Map for Segment



Outline

- Probabilistic graphical models
- Bayesian networks
- Inference in Bayes nets

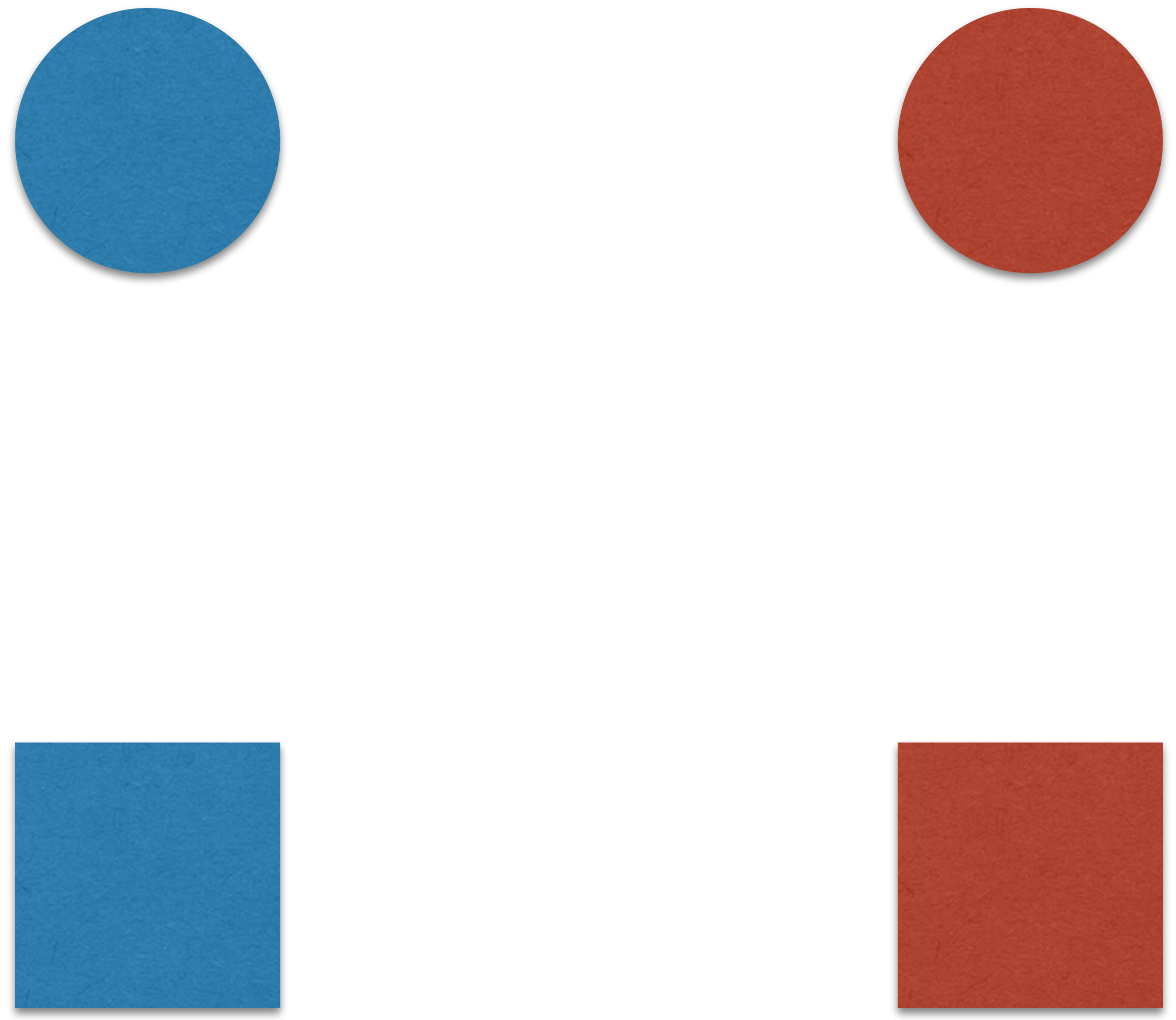
Probabilistic Graphical Models

- PGMs represent probability distributions
- They encode conditional independence structure with graphs
- They enable graph algorithms for inference and learning

Probability Identities

- Random variables in caps (**A**)
  - values in lowercase: **A = a** or just **a** for shorthand
- $\boxed{P(a \mid b)} = P(a, b) / P(b)$ conditional probability
- $\boxed{P(a, b)} = P(a \mid b)\, P(b)$ joint probability
- $P(b \mid a) = P(a \mid b)\, P(b) / P(a)$

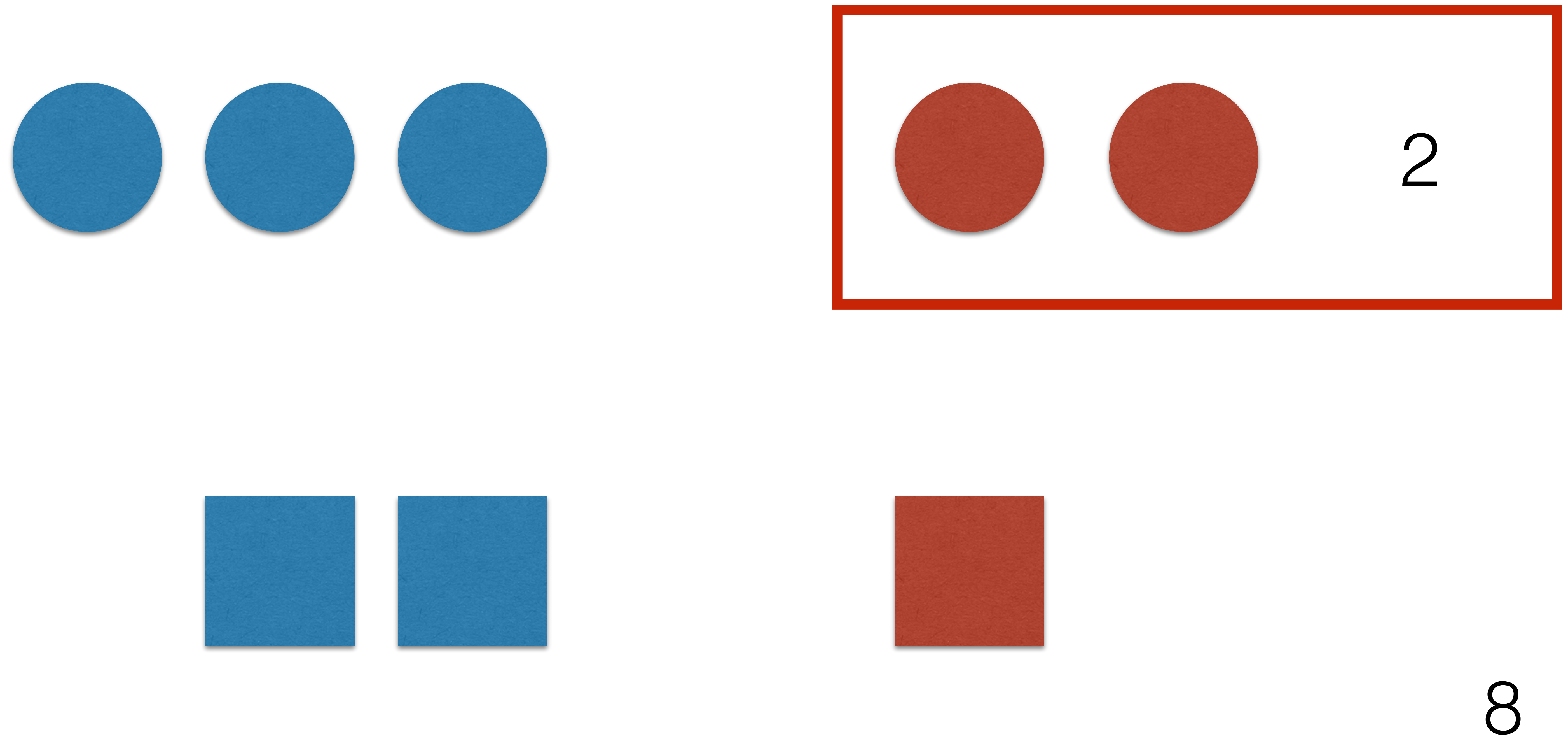
Probability via Counting



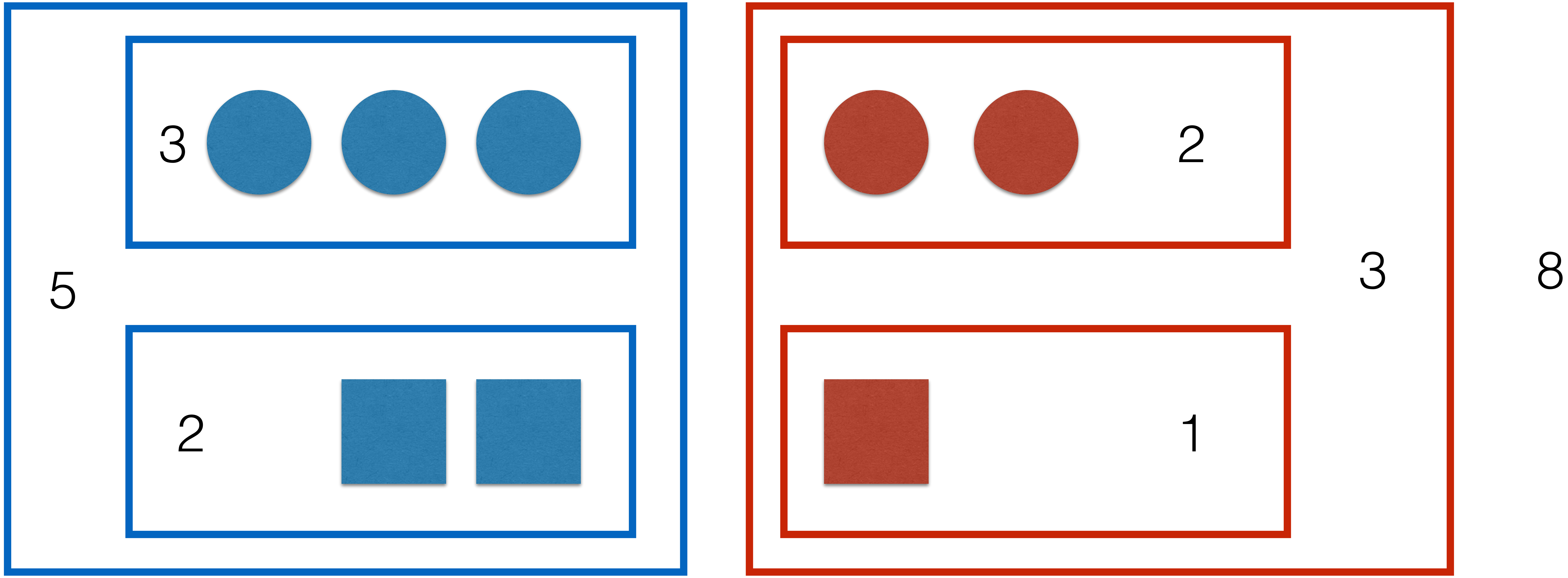
Probability via Counting

P(circle, red)

$=2/8 = 0.25$



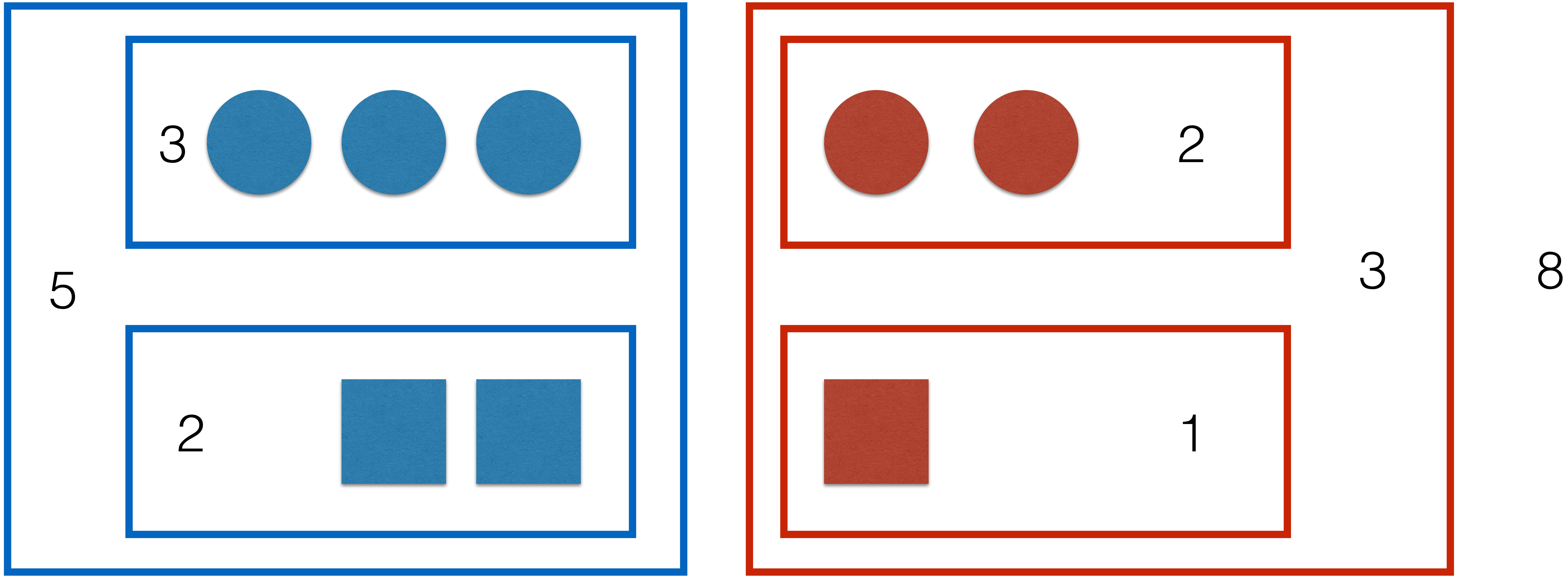
Probability via Counting



P(circle | red) = P(circle, red) / P(red)

2/3 2/8 3/8

Probability via Counting



P(circle | red) P(red) = P(circle, red)

2/3 3/8 2/8

Probability Identities

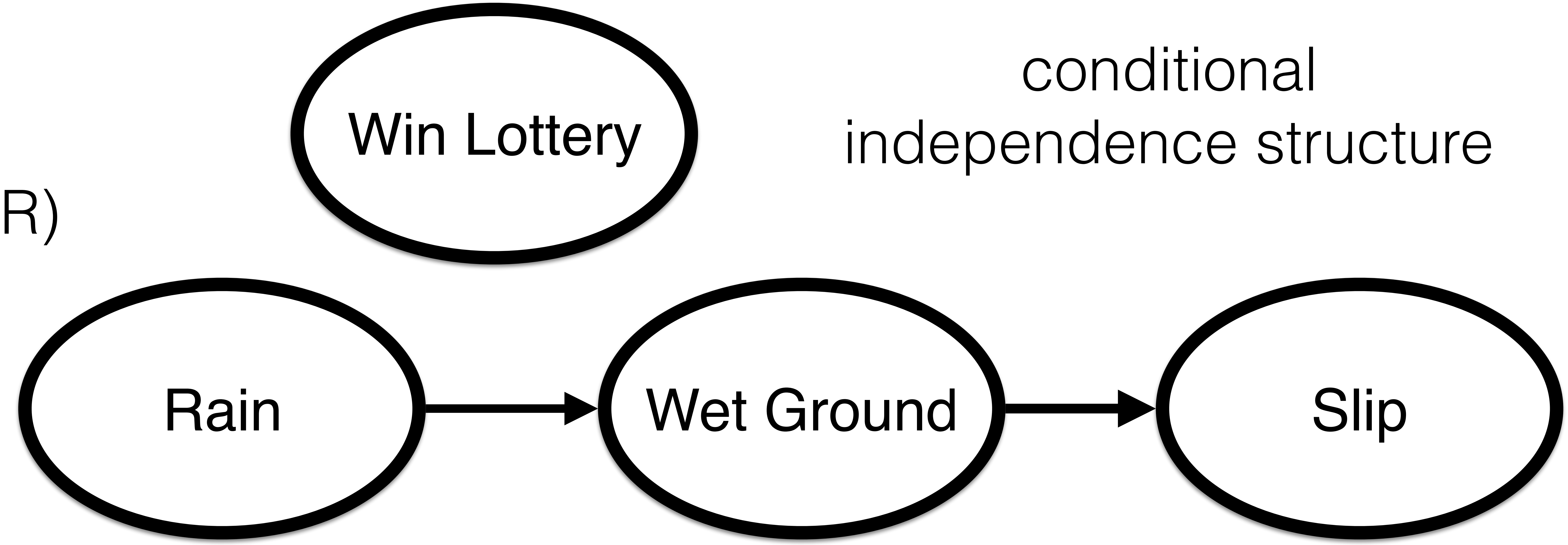
- Random variables in caps (**A**)
    - values in lowercase: **A = a** or just **a** for shorthand
- $P(a \mid b) = P(a, b) / P(b)$
- $P(a, b) = P(a \mid b) P(b)$
- $P(b \mid a) = P(a \mid b) P(b) / P(a)$

Bayesian Networks

$P(L, R, W)$

$= P(L)\,P(R)\,P(W \mid R)$

conditional independence structure



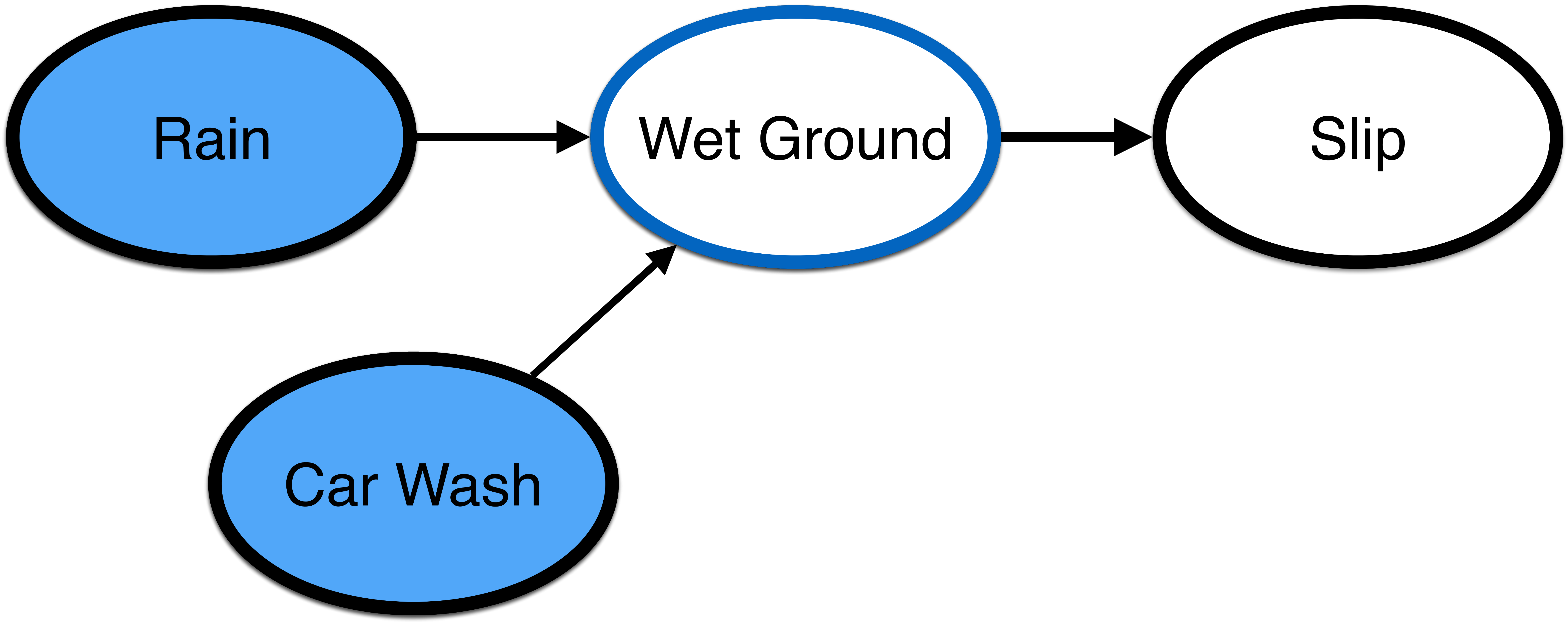
$P(L, R, W, S) = P(L)\,P(R)\,P(W \mid R)\,P(S \mid W)$

$\cancel{P(S \mid W, R)}$

Bayesian Networks

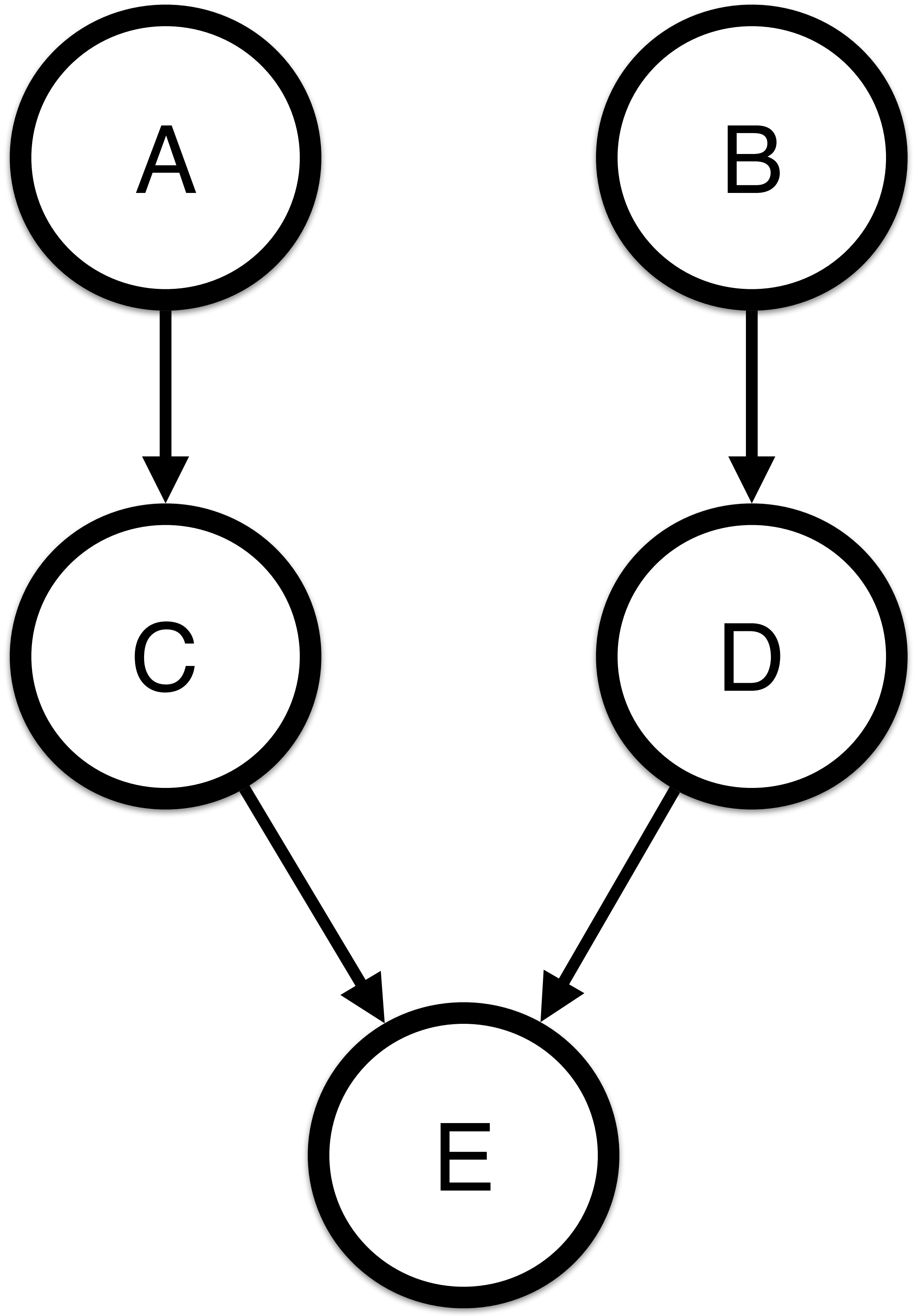
$P(R, W, S, C) = P(R)\, P(C)\, \boxed{P(W \mid C, R)}\, P(S \mid W)$

$P(X \mid \text{Parents}(X))$



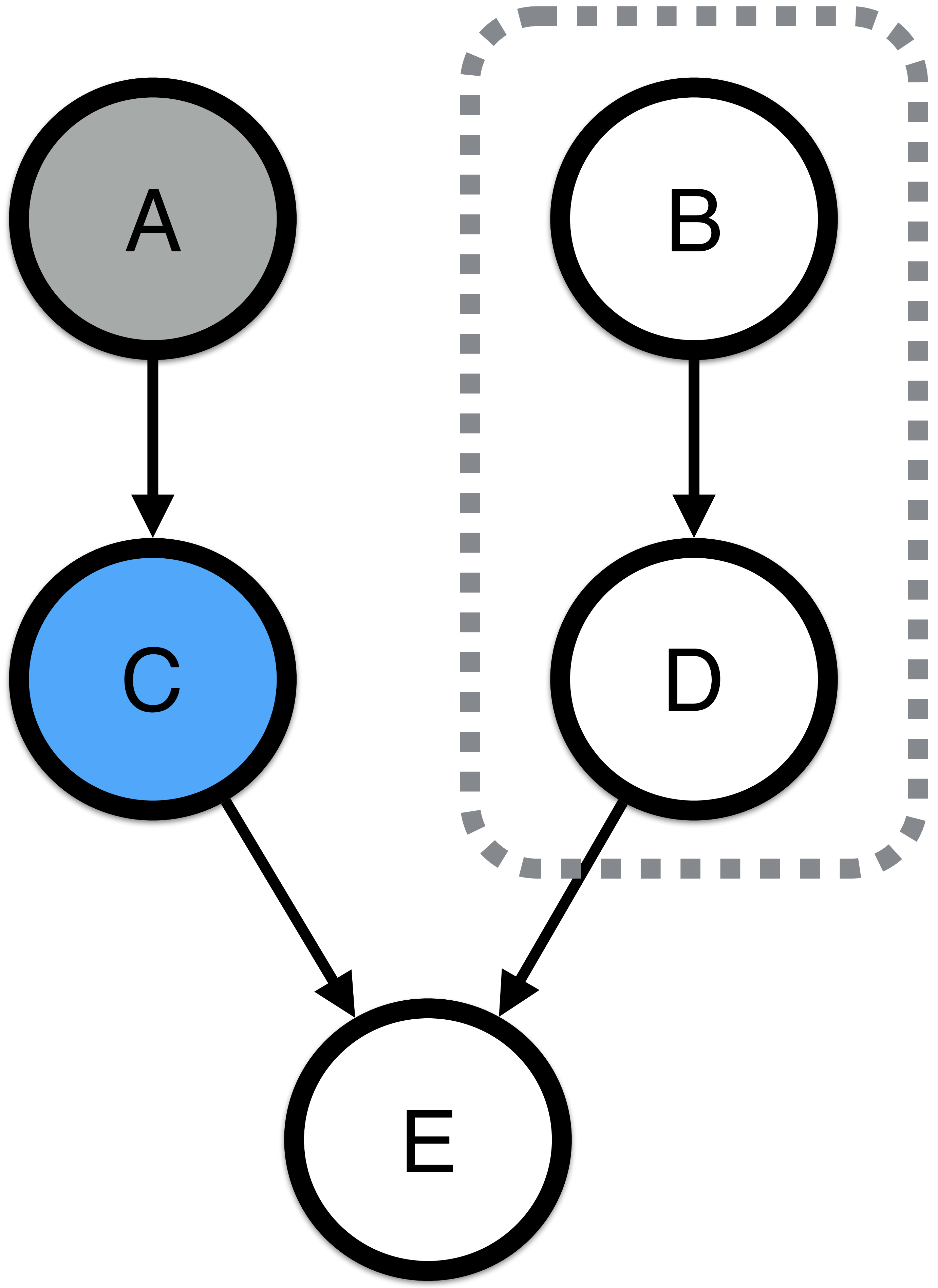
Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**



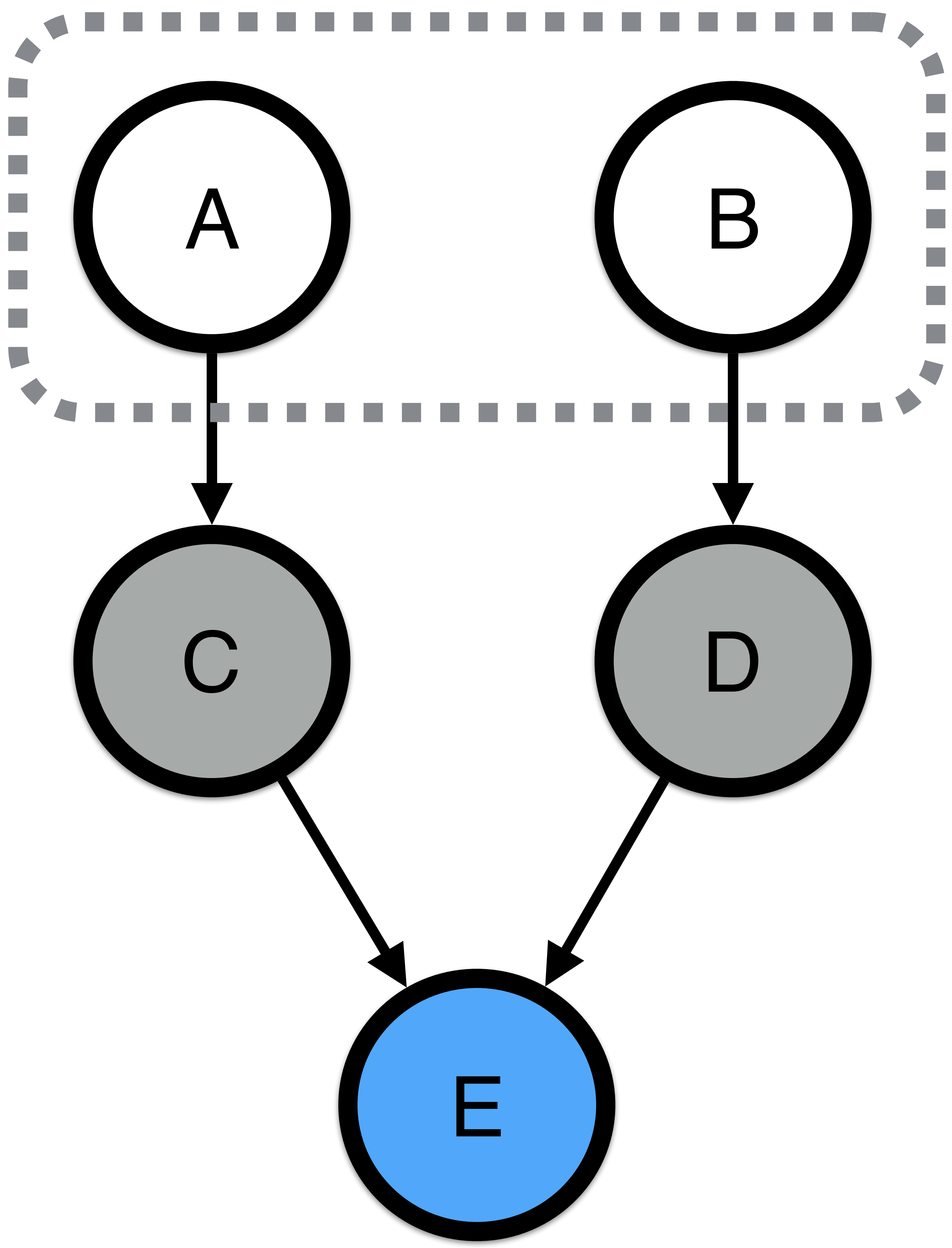
Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**



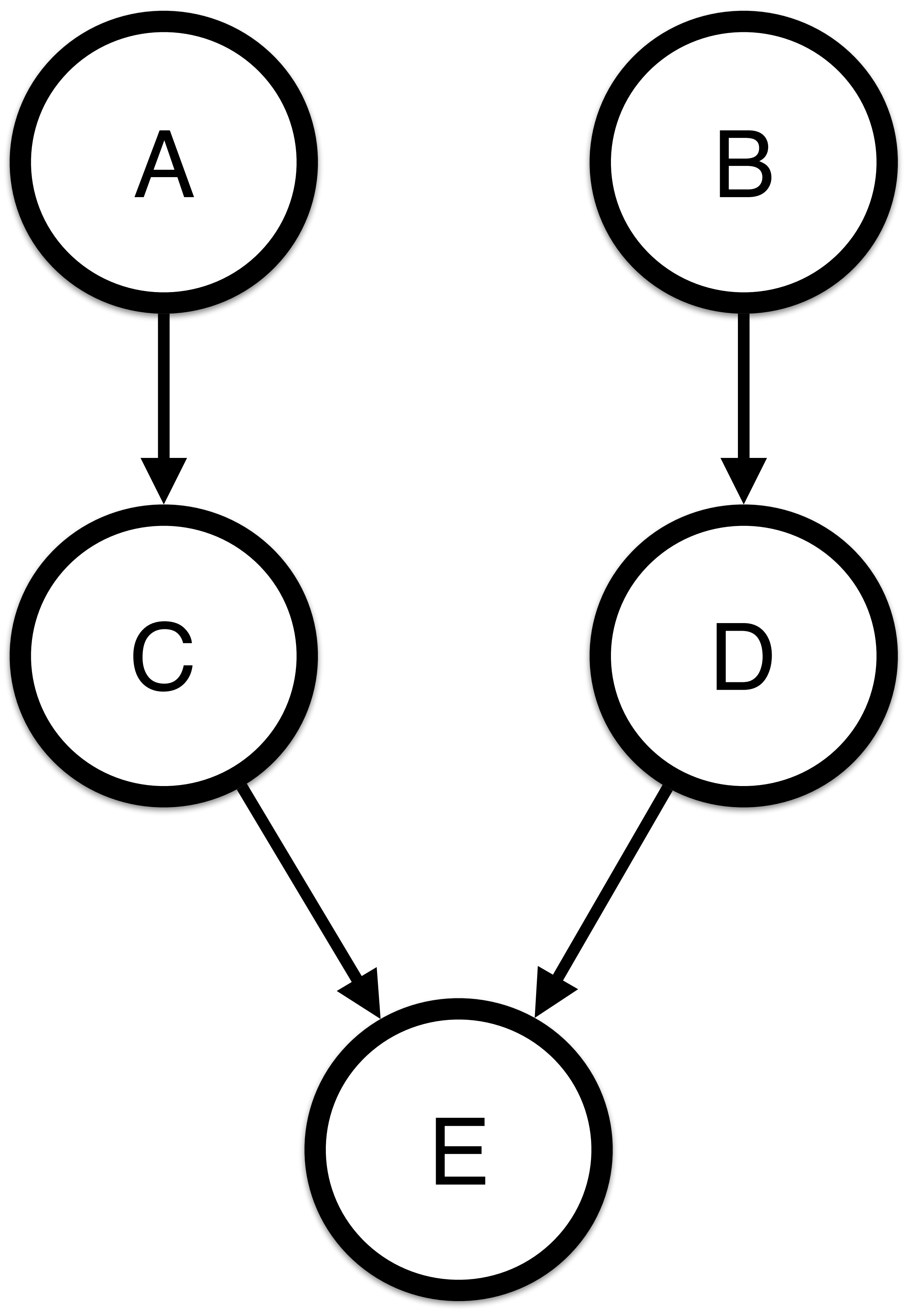
Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**



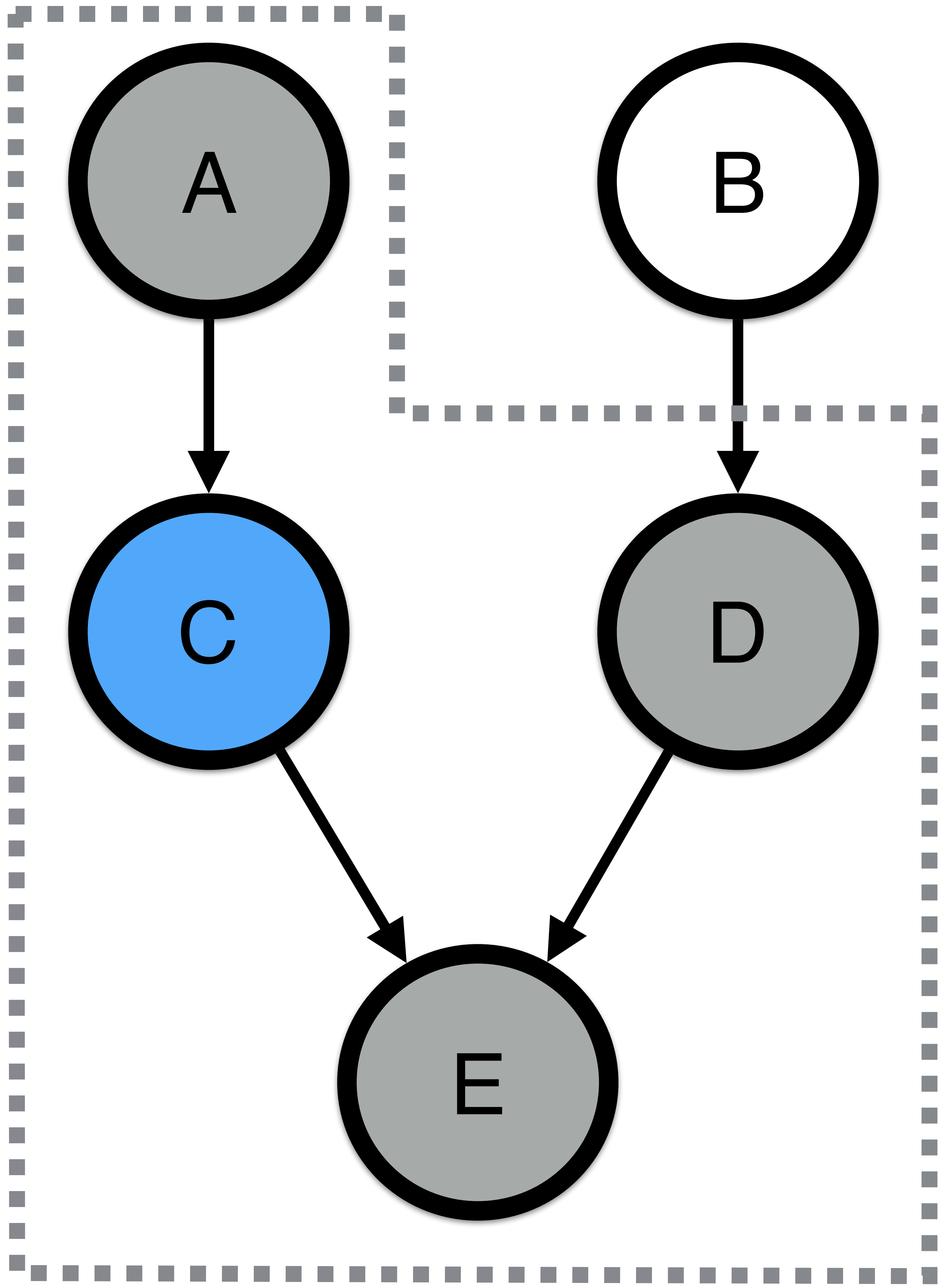
Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**
- Each variable is conditionally independent of any other variable given its **Markov blanket**
  - Parents, children, and children's parents



Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**
- Each variable is conditionally independent of any other variable given its **Markov blanket**
  - Parents, children, and children's parents



Inference

- Given a Bayesian Network describing P(X, Y, Z), what is P(Y)
  - First approach: **enumeration**

P(R, W, S, C) = P(R) P(C) P(W | C, R) P(S | W)

$$P(r|s) = \sum_w \sum_c P(r, w, s, c)/P(s)$$

$$P(r|s) \propto \sum_w \sum_c P(r)P(c)P(w|c, r)P(s|w)$$

$$P(r|s) \propto P(r) \sum_w P(s|w) \sum_c P(c)P(w|c, r)$$

$$O(2^n)$$

Second Approach: Variable Elimination

$$P(r|s) \propto \sum_w \sum_c P(r) P(c) P(w|c,r) P(s|w)$$

$$f_C(w) = \sum_c P(c) P(w|c,r)$$

$$P(r|s) \propto \sum_w P(r) P(s|w) f_c(w)$$

$$P(W, X, Y, Z) = P(W)P(X|W)P(Y|X)P(Z|Y)$$

$P(Y)$?

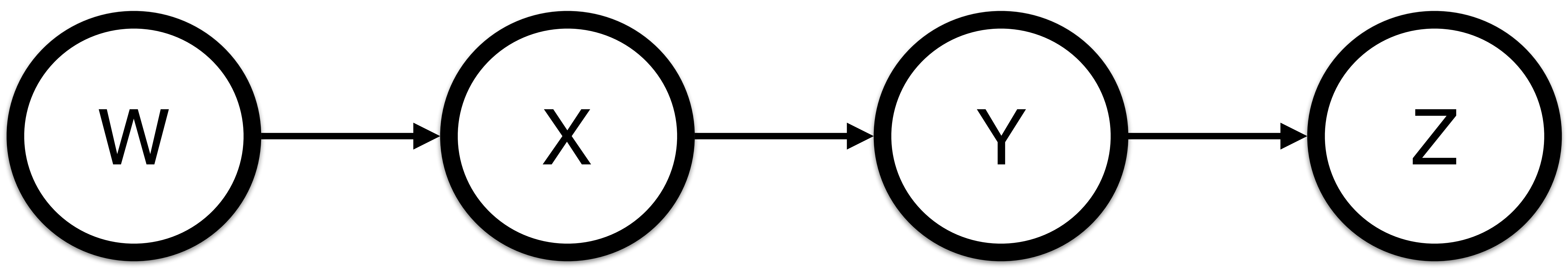
$$P(Y) = \sum_w \sum_x \sum_z P(w)P(x|w)P(Y|x)P(z|Y)$$

$$f_w(x) = \sum_w P(w)P(x|w)$$

$$P(Y) = \sum_x \sum_z f_w(x)P(Y|x)P(z|Y)$$

$$f_x(Y) = \sum_x f_w(x)P(Y|x)$$

$$P(Y) = \sum_z f_x(Y)P(z|Y)$$



$$P(Y) = \sum_{w}\sum_{x}\sum_{z} P(w)P(x|w)P(Y|x)P(z|Y)$$

$$f_w(x) = \sum_{w} P(w)P(x|w)$$

$$P(Y) = \sum_{x}\sum_{z} f_w(x)P(Y|x)P(z|Y)$$

$$f_x(Y) = \sum_{x} f_w(x)P(Y|x)$$

$$P(Y) = \sum_{z} f_x(Y)P(z|Y)$$

Variable Elimination

- Every variable that is not an ancestor of a query variable or evidence variable is irrelevant to the query
- Iterate:
  - choose variable to eliminate
  - sum terms relevant to variable, generate new factor
  - until no more variables to eliminate
- Exact inference is #P-Hard
  - in tree-structured BNs, linear time (in number of table entries)

Learning in Bayes Nets

- Super easy!
- Estimate each conditional probability by counting