

CS 583: PROBABILISTIC GRAPHICAL MODELS

TOPIC: CONDITIONAL PROBABILITY DISTRIBUTIONS (CPDs)



Mustafa Bilgic



<http://www.cs.iit.edu/~mbilgic>



<https://twitter.com/bilgicm>

CPDs

- So far, we assumed tabular representation of Conditional Probability Distributions (CPDs); these are also called Conditional Probability Tables (CPTs)
- A CPD encodes $P(X \mid Pa(X))$ where $\sum P(X \mid Pa(X)) = 1$
- CPTs have significant disadvantages
 - Cannot handle variables with infinite domains
 - The number of independent parameters needed is $|Val(X)-1| * |Val(Pa(X))|$, which is exponential in the number of parents
 - Ignores the local structure

EXAMPLE – MEDICAL DIAGNOSIS

- A patient can have one of many diseases; each disease is represented as a binary variable (Present/NotPresent)
- A symptom, e.g., Fever, can have many causes; this requires many disease nodes to be the parents of a symptom
- If Fever is caused by 10 different diseases, then the CPT for Fever requires $2^{10} = 1,024$ independent parameters (assuming a patient can have more than one disease at the same time, and we represent each disease as present/notpresent; and fever is also binary)

KEY INSIGHT

- A CPD has to represent $P(X \mid Pa(X))$ where $\sum P(X \mid Pa(X)) = 1$
- CPD does not have to list all possible combinations
- CPD is a function that maps $(x, Pa(x))$ to a conditional distribution $P(x \mid Pa(x))$
- This representation guarantees that the joint is a well-defined probability distribution

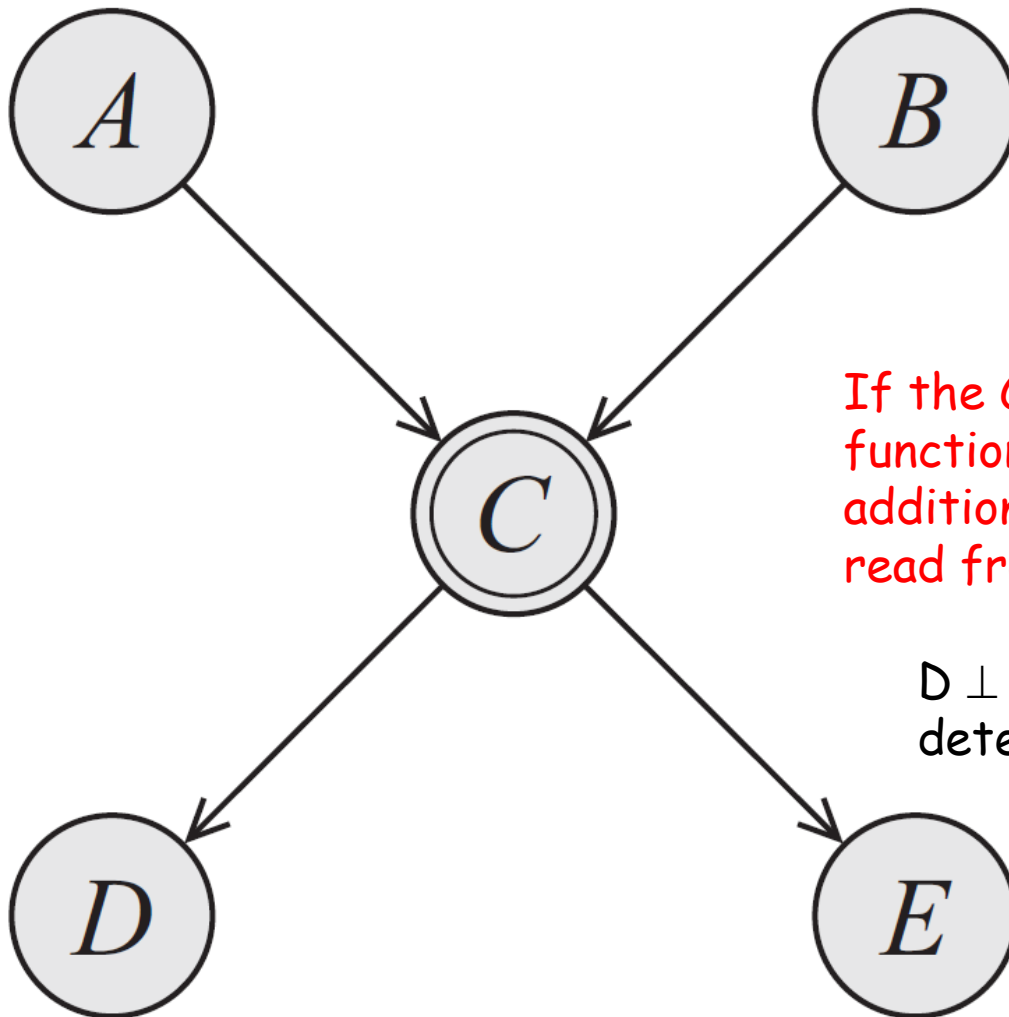
TYPES OF CPDS WE'LL SEE

- Deterministic CPDs
- Context-Specific CPDs
 - Tree CPDs
 - Rule CPDs
- Causal independence
 - The noisy-or model
 - Logistic CPD
- Continuous variables
 - Linear Gaussian CPDs

DETERMINISTIC CPDS

- X is a deterministic function of its parents $\text{Pa}(X)$
- There is a deterministic function $f: \text{Val}(\text{Pa}(X)) \rightarrow \text{Val}(X)$ such that
 - $P(x \mid \text{Pa}(x)) = 1$ if $x = f(\text{Pa}(x))$
0 otherwise
- Examples
 - X is (OR, AND, XOR, ...) of its parents
 - X is average of its parents

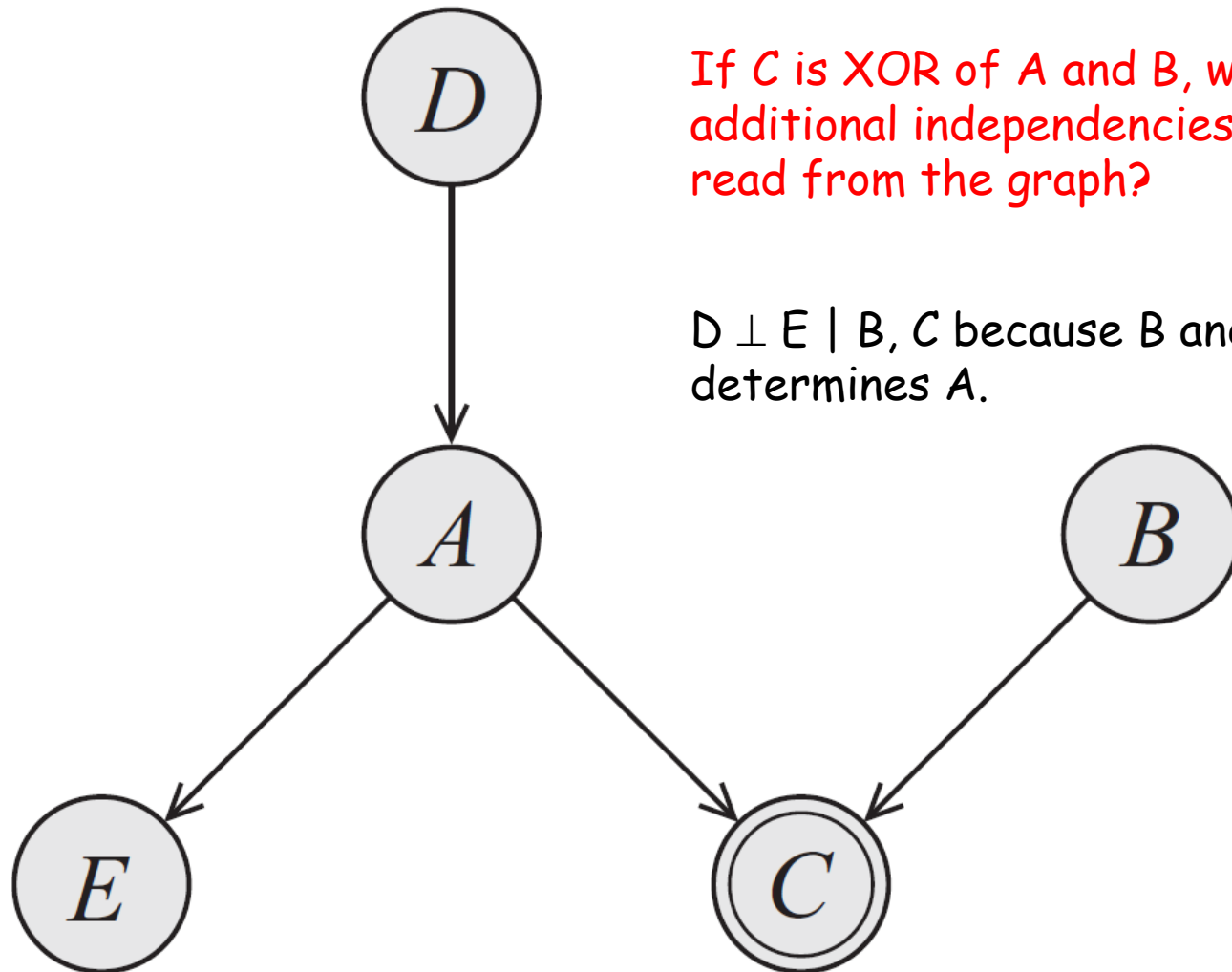
INDEPENDENCIES



If the CPD of C is a deterministic function of A and B , what additional independencies can we read from the graph?

$D \perp E \mid A, B$ because A and B determines C .

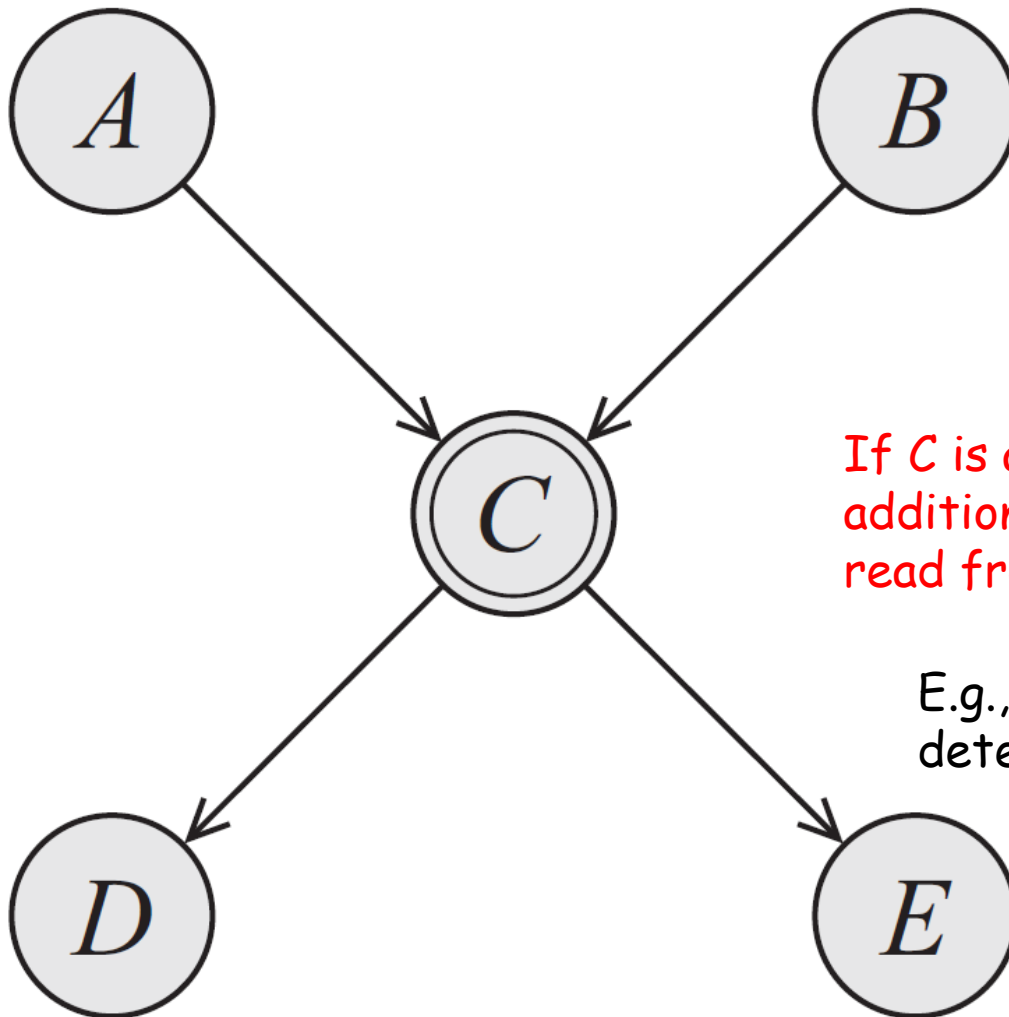
INDEPENDENCIES



If C is XOR of A and B , what additional independencies can we read from the graph?

$D \perp E \mid B, C$ because B and C determines A .

INDEPENDENCIES



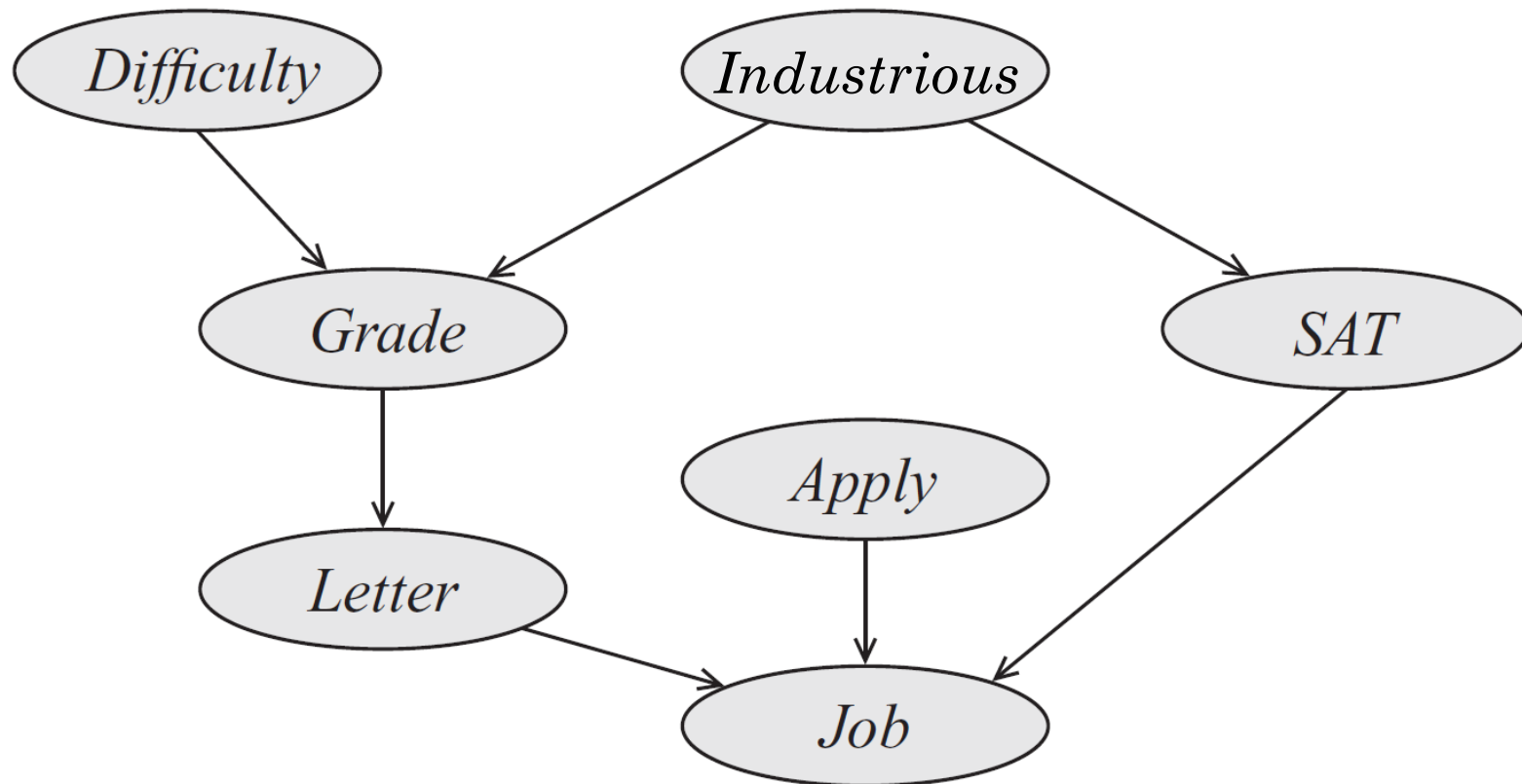
If C is an OR of A and B , what additional independencies can we read from the graph?

E.g., $D \perp E \mid A=T$ because $A=T$ determines C .

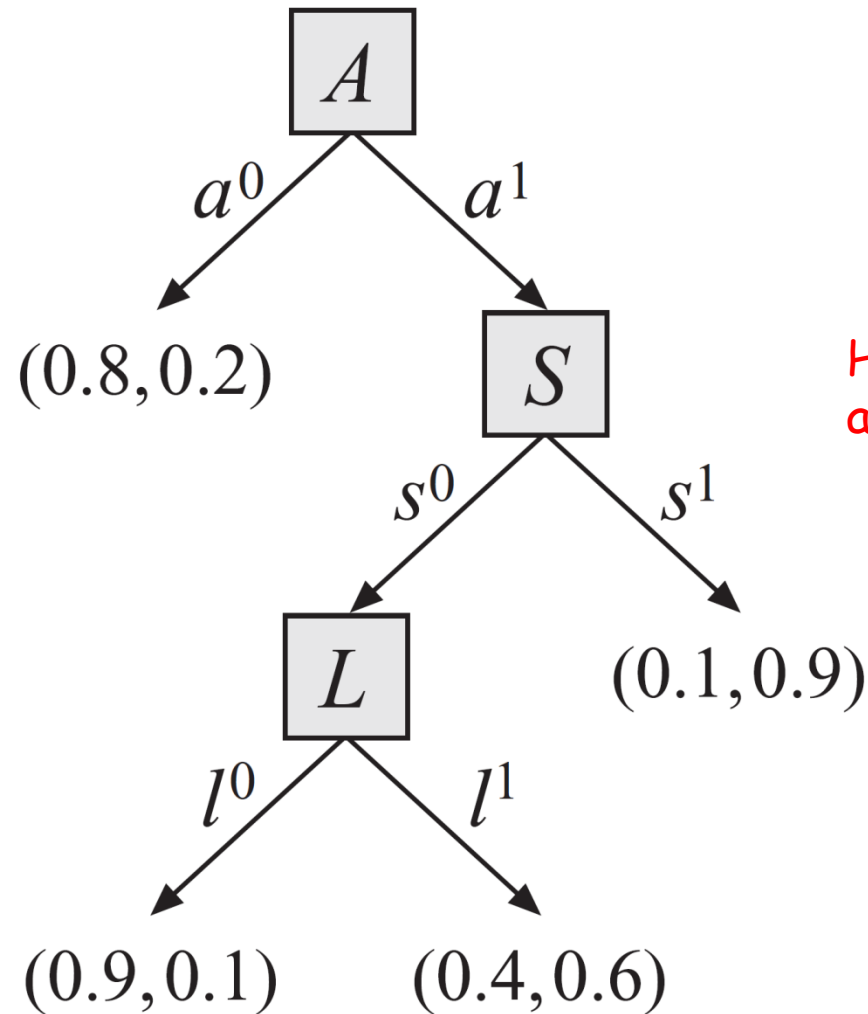
CONTEXT-SPECIFIC INDEPENDENCE

- **Definition:** Let X , Y , Z be pairwise disjoint sets of variables, C be a set of variables (that might overlap with $X \cup Y \cup Z$), and let $c \in \text{val}(C)$. We say X and Y are *contextually independent* given Z and the context c if
 - $P(X \mid Y, Z, c) = P(X \mid Z, c)$ whenever $P(Y, Z, c) > 0$.

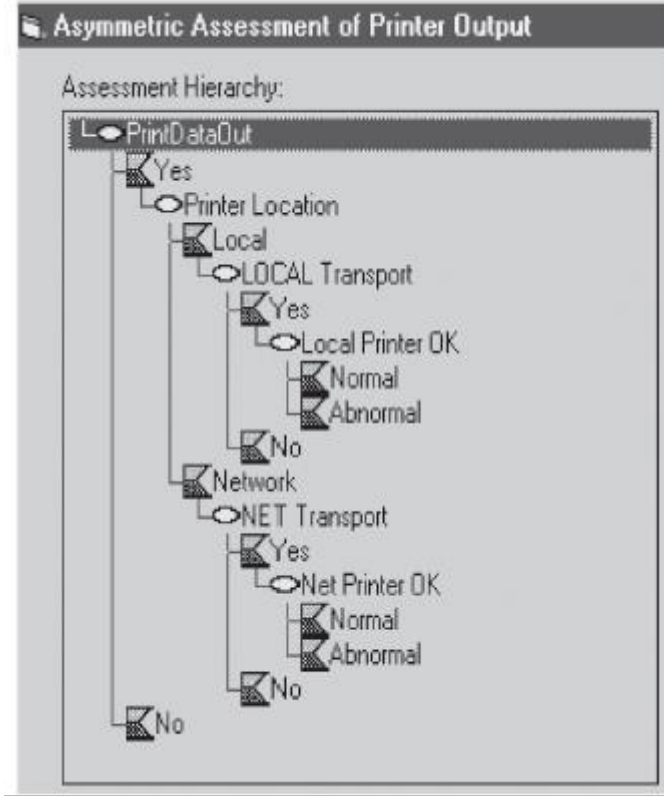
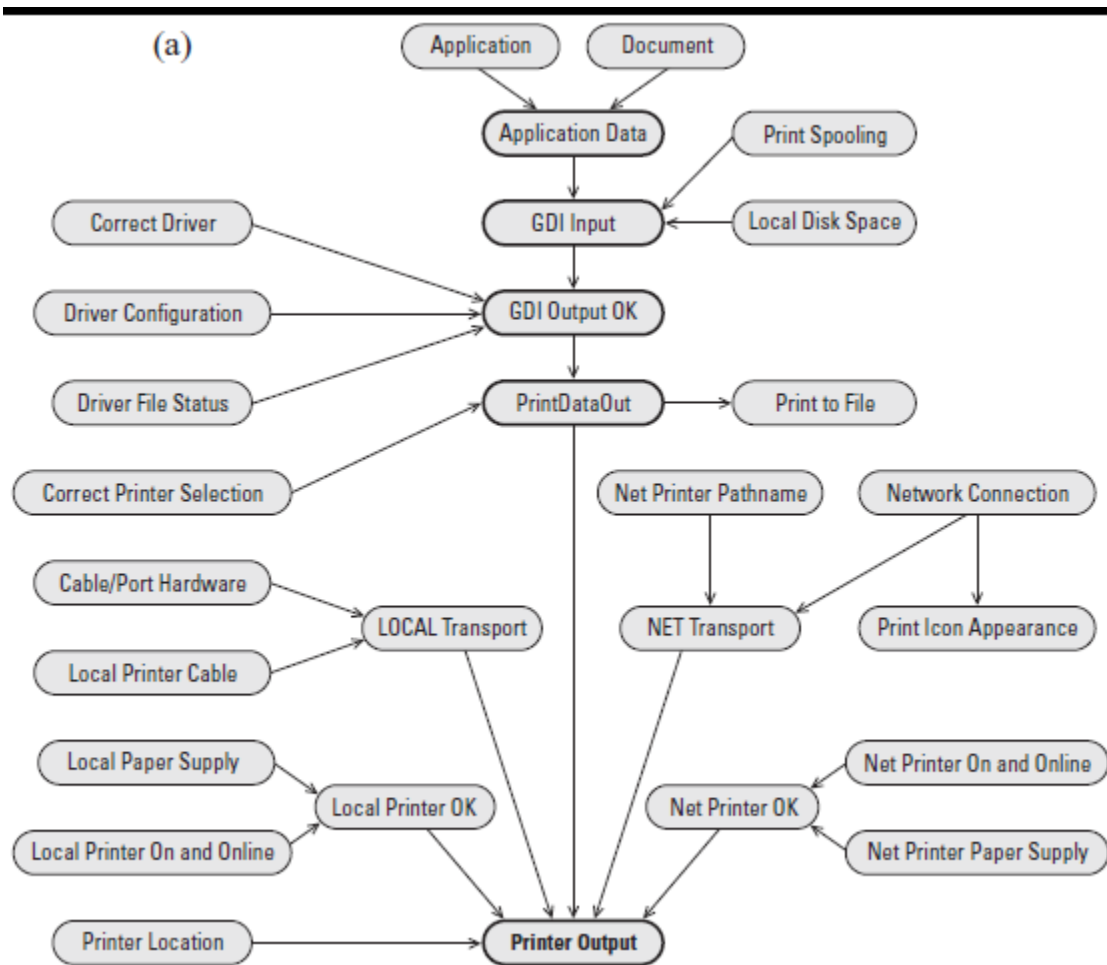
EXAMPLE: JOB APPLICATION



TREE-CPDS



How many parameters are needed?



RULE CPDS

- Definition: A rule-based CPD $P(X | \text{Pa}_X)$ is a set of rules \mathcal{R} such that
 - For each rule $r \in \mathcal{R}$, $\text{Scope}[r] \subseteq \{X\} \cup \text{Pa}_X$
 - For each assignment to (x, \mathbf{u}) , we have precisely one rule $\langle \mathbf{c}; p \rangle \in \mathcal{R}$ such that \mathbf{c} is compatible with (x, \mathbf{u}) . In this case, $P(X=x | \text{Pa}_X=\mathbf{u}) = p$.
 - The resulting CPD $P(X | \text{Pa}_X)$ is a legal CPD,
 - $\sum_x P(x | \mathbf{u}) = 1$
- More general than tree CPDs
 - Every tree CPD can be represented as a rule-based CPD but the converse is not true

EXAMPLE: RULE CPDS

$$\rho_1: \langle a^1, b^1, x^0; 0.1 \rangle$$

$$\rho_3: \langle a^0, c^1, x^0; 0.2 \rangle$$

$$\rho_5: \langle b^0, c^0, x^0; 0.3 \rangle$$

$$\rho_7: \langle a^1, b^0, c^1, x^0; 0.4 \rangle$$

$$\rho_9: \langle a^0, b^1, c^0; 0.5 \rangle$$

$$\rho_2: \langle a^1, b^1, x^1; 0.9 \rangle$$

$$\rho_4: \langle a^0, c^1, x^1; 0.8 \rangle$$

$$\rho_6: \langle b^0, c^0, x^1; 0.7 \rangle$$

$$\rho_8: \langle a^1, b^0, c^1, x^1; 0.6 \rangle$$

X	$a^0 b^0 c^0$	$a^0 b^0 c^1$	$a^0 b^1 c^0$	$a^0 b^1 c^1$	$a^1 b^0 c^0$	$a^1 b^0 c^1$	$a^1 b^1 c^0$	$a^1 b^1 c^1$
x^0	0.3	0.2	0.5	0.2	0.3	0.4	0.1	0.1
x^1	0.7	0.8	0.5	0.8	0.7	0.6	0.9	0.9

INDEPENDENCE OF CAUSAL MODELS

- Variable of interest Y depends on a number of causes X_1, \dots, X_k
- Even though the interaction between X_i and Y can be arbitrary, it is often reasonable to assume that the combined influence of X_i on Y is a simple combination of the individual influences of X_i on Y in isolation.
 - Noisy-or model
 - Logistic CPD

NOISY-OR MODEL

- A professor writes a good letter if
 - The student asked good questions in class
 - The student wrote a good final paper
- However, the professor
 - Might forget that student asked good questions, with 0.2 probability
 - Might not be able to read student's handwriting, with 0.1 probability
- What is the probability that the professor will write a good letter if the student
 - Did not ask good questions, and did not write a good final paper?
 - Asked good questions but did not write a good final paper?
 - Did not ask good questions but wrote a good final paper?
 - Asked good questions, and wrote a good final paper?

NOISY-OR MODEL

Q, F	l^0	l^1
q^0, f^0	1	0
q^0, f^1	0.1	0.9
q^1, f^0	0.2	0.8
q^1, f^1	?	?

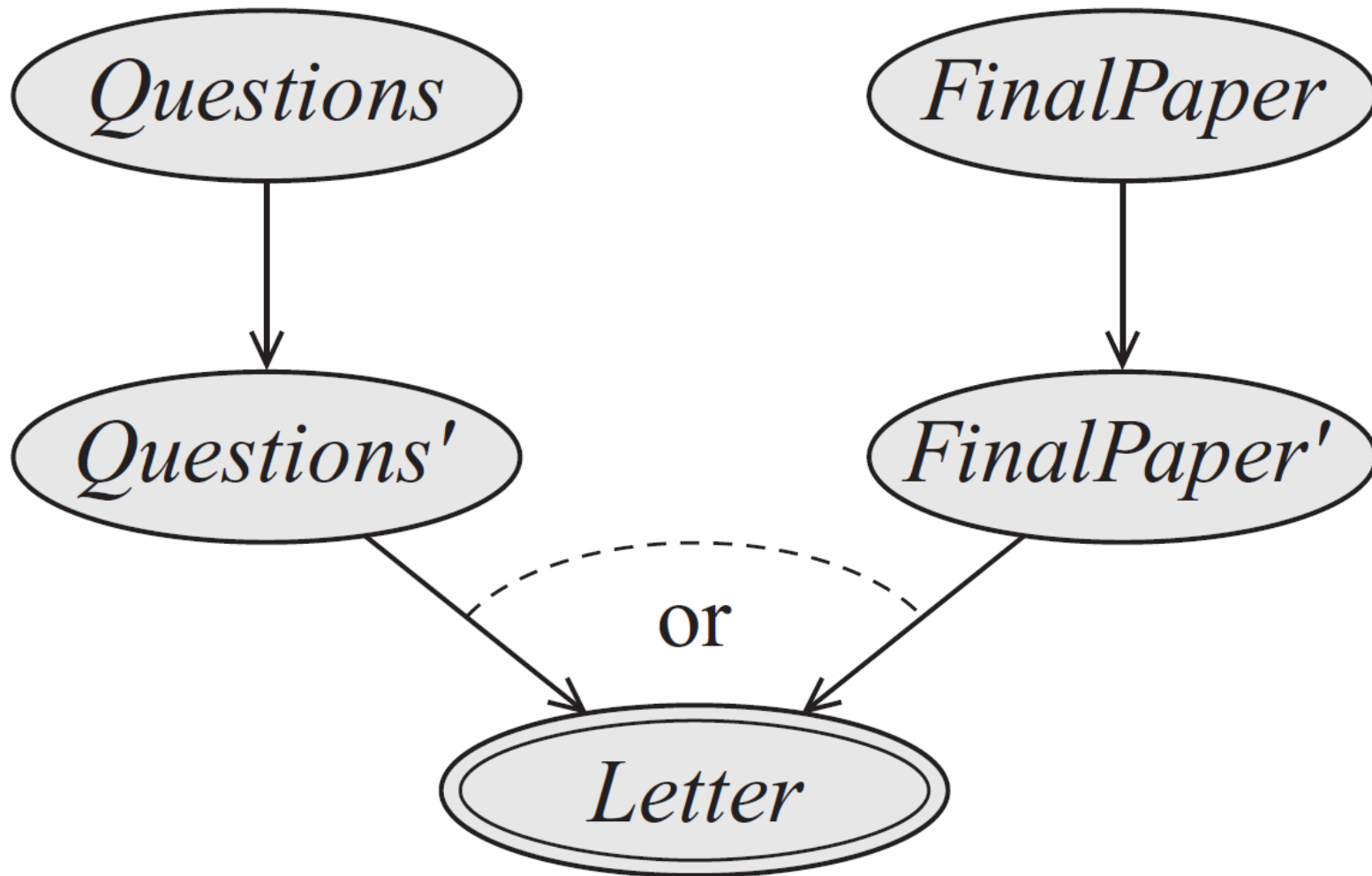
For the professor to write a bad letter when student asked good questions and a good final paper, the professor

- Forgets the student's participation AND
- Cannot read the student's handwriting

$$P(l^0 | q^1, f^1) = 0.1 * 0.2 = 0.02$$

$$P(l^1 | q^1, f^1) = 1 - 0.02 = 0.98$$

NOISY-OR MODEL



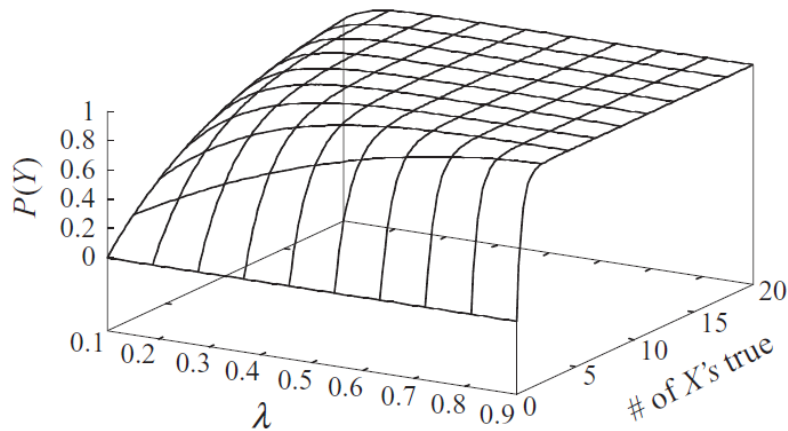
NOISY-OR MODEL

- Let Y be a binary random variable with k binary parents X_1, \dots, X_k . The CPD $P(Y | X_1, \dots, X_k)$ is a noisy-or if there are $k+1$ parameters $\lambda_0, \lambda_1, \dots, \lambda_k$ such that

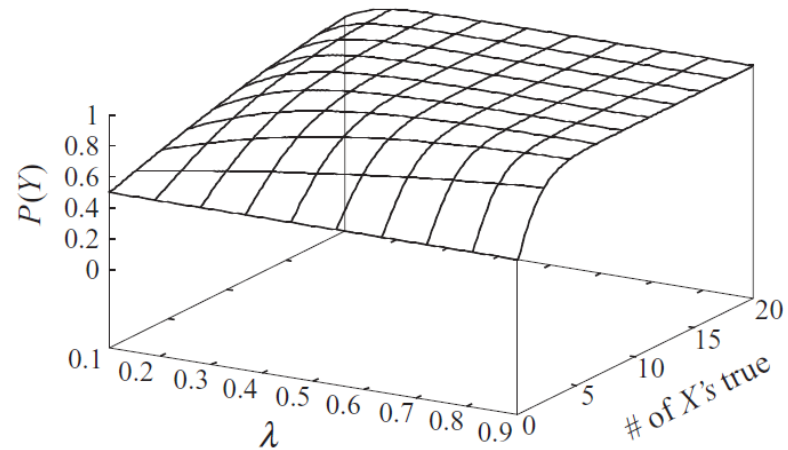
$$P(y^0 | X_1, \dots, X_k) = (1 - \lambda_0) \prod_{i: X_i = x_i^1} (1 - \lambda_i)$$

$$P(y^1 | X_1, \dots, X_k) = 1 - (1 - \lambda_0) \prod_{i: X_i = x_i^1} (1 - \lambda_i)$$

NOISY-OR MODEL



(a)



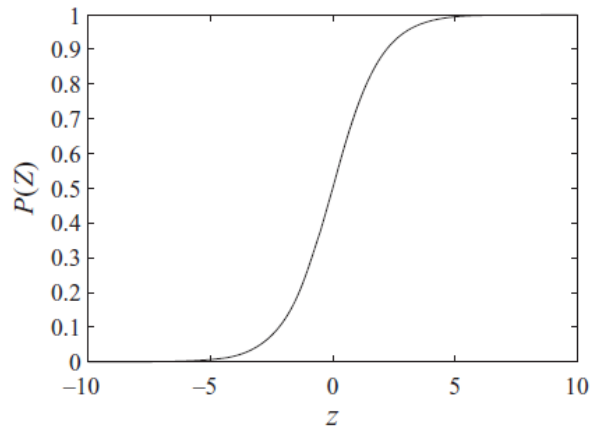
(b)

LOGISTIC CPD

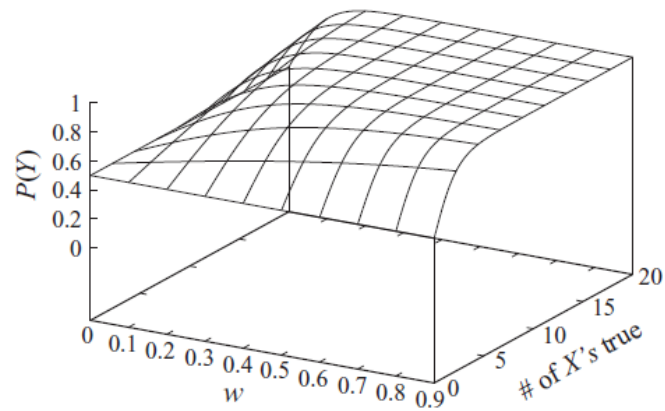
- Let Y be a binary variable with k parents: X_1, \dots, X_k that take on numerical values. The CPD $P(Y | X_1, \dots, X_k)$ is a *logistic CPD* if there are $k+1$ weights w_0, w_1, \dots, w_k such that

$$\begin{aligned} P(y^1 | X_1, \dots, X_k) &= \text{sigmoid} \left(w_0 + \sum_{i=1}^k w_i X_i \right) \\ &= \frac{e^{w_0 + \sum_{i=1}^k w_i X_i}}{1 + e^{w_0 + \sum_{i=1}^k w_i X_i}} \end{aligned}$$

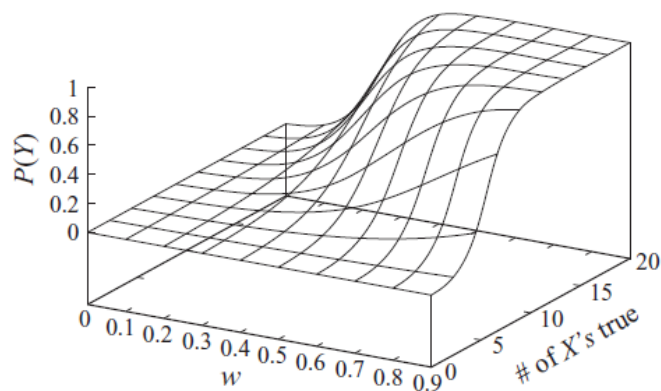
LOGISTIC CPD



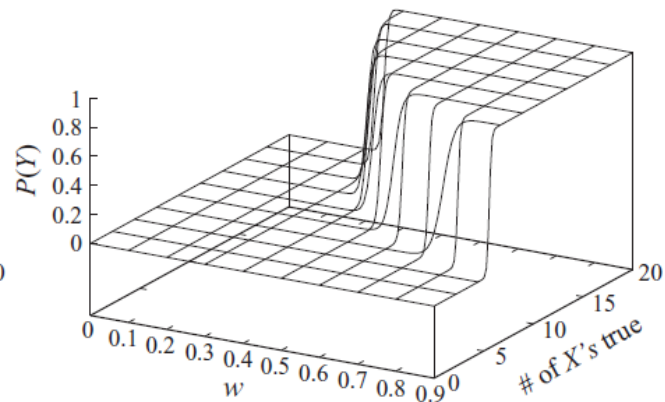
(a)



(b)



(c)



(d)

LINEAR GAUSSIAN CPD

- Let Y be a continuous variable with continuous parents X_1, \dots, X_k . We say that Y has a linear Gaussian model if there are parameters β_0, \dots, β_k and σ^2 such that

$$p(Y \mid x_1, \dots, x_k) = \mathcal{N}(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k; \sigma^2)$$