Reasoning Under Uncertainty: Independence

Alan Mackworth

UBC CS 322 – Uncertainty 3

March 15, 2013

Textbook §6.2

Announcements

- Assignment 4 posted today. Due Wed. April 3.
- Exercise 10 posted today. Marginal and Conditional Independence, AIspace Belief and Decision App.

Lecture Overview

- Recap
  - Conditioning & Inference by Enumeration
  - Bayes Rule & The Chain Rule
- Independence
  - Marginal Independence
  - Conditional Independence

Recap: Conditioning

- Conditioning: revise beliefs based on new observations

- We need to integrate two sources of knowledge
  - Prior probability distribution P(X): all background knowledge
  - New evidence e

- Combine the two to form a posterior probability distribution
  - The conditional probability P(h|e)

Recap: Example for conditioning

- You have a prior for the joint distribution of weather and temperature, and the marginal distribution of temperature

Possible world	*Weather*	*Temperature*	*$\mu(w)$*
w_1	sunny	hot	0.10
w_2	sunny	mild	0.20
w_3	sunny	cold	0.10
~~w_4~~	~~cloudy~~	~~hot~~	~~0.05~~
~~w_5~~	~~cloudy~~	~~mild~~	~~0.35~~
~~w_6~~	~~cloudy~~	~~cold~~	~~0.20~~

T	*P(T|W=sunny)*
hot	?
mild	?
cold	?

- Now, you look outside and see that it's sunny
  – You are certain that you're in world w_1, w_2, or w_3

Recap: Example for conditioning

- You have a prior for the joint distribution of weather and temperature, and the marginal distribution of temperature

Possible world	*Weather*	*Temperature*	*$\mu(w)$*
w_1	sunny	hot	0.10
w_2	sunny	mild	0.20
w_3	sunny	cold	0.10
~~w_4~~	~~cloudy~~	~~hot~~	~~0.05~~
~~w_5~~	~~cloudy~~	~~mild~~	~~0.35~~
~~w_6~~	~~cloudy~~	~~cold~~	~~0.20~~

T	*P(T|W=sunny)*
hot	0.10/0.40=0.25
mild	0.20/0.40=0.50
cold	0.10/0.40=0.25

- Now, you look outside and see that it's sunny
  - You are certain that you're in world w_1, w_2, or w_3
  - To get the conditional probability, you simply renormalize to sum to 1
  - 0.10+0.20+0.10=0.40

Recap: Conditional probability

Definition (conditional probability)
The conditional probability of formula h given evidence e is

$$P(h|e) = \frac{P(h \wedge e)}{P(e)}$$

E.g. $P(T = hot|W = sunny) = \frac{P(T=hot \wedge W=sunny)}{P(W=sunny)}$

Possible world	*Weather*	*Temperature*	*$\mu(w)$*
w_1	sunny	hot	0.10
w_2	sunny	mild	0.20
w_3	sunny	cold	0.10
~~w_4~~	~~cloudy~~	~~hot~~	~~0.05~~
~~w_5~~	~~cloudy~~	~~mild~~	~~0.35~~
~~w_6~~	~~cloudy~~	~~cold~~	~~0.20~~

T	*P(T|W=sunny)*
hot	0.10/0.40=0.25
mild	0.20/0.40=0.50
cold	0.10/0.40=0.25

Recap: Inference by Enumeration

- Great, we can compute arbitrary probabilities now!
- Given
  - Prior joint probability distribution (JPD) on set of variables X
  - specific values e for the evidence variables E (subset of X)
- We want to compute
  - posterior joint distribution of query variables Y (a subset of X) given evidence e
- Step 1: Condition to get distribution P(X|e)
- Step 2: Marginalize to get distribution P(Y|e)
- Generally applicable, but memory-heavy and slow

Recap: Bayes rule and Chain Rule

Theorem (Bayes theorem, or Bayes rule)

$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)}$$

E.g., $P(fire|\text{alarm}) = \frac{P(alarm|fire) \times P(fire)}{P(alarm)}$

Theorem (Chain Rule)

$$P(f_n \wedge \cdots \wedge f_1) = \prod_{i=1}^{n} P(fi|f_{i-1} \wedge \cdots \wedge f_1)$$

E.g. $P(A,B,C,D) = P(A) \times P(B|A) \times P(C|A,B) \times P(D|A,B,C)$

Lecture Overview

- Recap
  - Conditioning & Inference by Enumeration
  - Bayes Rule & The Chain Rule
- Independence
  - Marginal Independence
  - Conditional Independence

Marginal Independence: example

- Some variables are independent:
  - Knowing the value of one does not tell you anything about the other
  - Example: variables W (weather) and R (result of a die throw)
    - Let's compare P(W) vs. P(W | R = 6)
- What is P(W=cloudy) ?

0.066 0.1 0.4 0.6

Weather W	*Result R*	*P(W,R)*
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal Independence: example

- Some variables are independent:
  - Knowing the value of one does not tell you anything about the other
  - Example: variables W (weather) and R (result of a die throw)
    - Let's compare P(W) vs. P(W | R = 6)
- What is P(W=cloudy) ?
  - P(W=cloudy) = $\sum_{r \in dom(R)}$ P(W=cloudy, R = r)

    = 0.1+0.1+0.1+0.1+0.1+0.1 = 0.6
- What is P(W=cloudy|R=6) ?

0.066/0.166 0.1/0.166

0.066+0.1 0.1/0.6

Weather W	*Result R*	*P(W,R)*
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal Independence: example

- Some variables are independent:
  - Knowing the value of one does not tell you anything about the other
  - Example: variables W (weather) and R (result of a die throw)
    - Let's compare P(W) vs. P(W | R = 6)
- What is P(W=cloudy) ?
  - P(W=cloudy) = $\Sigma_{r \in dom(R)}$ P(W=cloudy, R = r)
    = 0.1+0.1+0.1+0.1+0.1+0.1 = 0.6
- What is P(W=cloudy|R=6) ?
  - P(W=cloudy|R=6) = $\frac{P(W=cloudy \wedge R=6)}{P(R=6)}$
  - P(W=cloudy ∧ R=6) = 0.1 (from table)
  - P(R=6) = 0.166 (marginal, 0.1+0.066)
  - Thus, P(W=cloudy|R=6) = 0.1/0.166 = 0.6

Weather W	*Result R*	*P(W,R)*
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal Independence: example

- Some variables are independent:
  - Knowing the value of one does not tell you anything about the other
  - Example: variables W (weather) and R (result of a die throw)
    - Let's compare P(W) vs. P(W | R = 6)
- What is P(W=cloudy) ?
  - P(W=cloudy) = $\Sigma_{r \in dom(R)}$ P(W=cloudy, R = r)

    = 0.1+0.1+0.1+0.1+0.1+0.1 = 0.6
- What is P(W=cloudy|R=6) ?
  - P(W=cloudy|R=6) = $\frac{P(W=cloudy \wedge R=6)}{P(R=6)}$
  - P(W=cloudy ∧ R=6) = 0.1 (from table)
  - P(R=6) = 0.166 (marginal, 0.1+0.066)
  - Thus, P(W=cloudy|R=6) = 0.1/0.166 = 0.6

Weather W	*Result R*	*P(W,R)*
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Marginal Independence: example

- Some variables are independent:
  - Knowing the value of one does not tell you anything about the other
  - Example: variables W (weather) and R (result of a die throw)
    - Let's compare P(W) vs. P(W | R = 6)
    - The two distributions are identical
    - Knowing the result of the die does not change our belief in the weather

Weather W	*Result R*	*P(W,R)*
sunny	1	0.066
sunny	2	0.066
sunny	3	0.066
sunny	4	0.066
sunny	5	0.066
sunny	6	0.066
cloudy	1	0.1
cloudy	2	0.1
cloudy	3	0.1
cloudy	4	0.1
cloudy	5	0.1
cloudy	6	0.1

Weather W	*P(W)*
sunny	0.4
cloudy	0.6

Weather W	*P(W|R=6)*
sunny	0.066/0.166=0.4
cloudy	0.1/0.166=0.6

Marginal Independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$
\begin{aligned}
& P(X = xi|Y = y_j) \\
= & P(X = xi|Y = y_k) \\
= & P(X = xi)
\end{aligned}
$$

- Intuitively: if X and Y are marginally independent, then
  - learning that Y=y does not change your belief in X
  - and this is true for all values y that Y could take

- For example, weather is marginally independent of the result of a dice throw

Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$
\begin{aligned}
& P(X = xi | Y = y_j) \\
= & P(X = xi | Y = y_k) \\
= & P(X = xi)
\end{aligned}
$$

- Results C_1 and C_2 of two tosses of a fair coin
- Are C_1 and C_2 marginally independent?

yes no

C_1	C_2	$P(C_1, C_2)$
heads	heads	0.25
heads	tails	0.25
tails	heads	0.25
tails	tails	0.25

Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$
\begin{aligned}
& P(X = xi | Y = y_j) \\
= & P(X = xi | Y = y_k) \\
= & P(X = xi)
\end{aligned}
$$

- Results C_1 and C_2 of two tosses of a fair coin
- Are C_1 and C_2 marginally independent?
  - Yes. All probabilities in the definition above are 0.5.

C_1	C_2	$P(C_1, C_2)$
heads	heads	0.25
heads	tails	0.25
tails	heads	0.25
tails	tails	0.25

Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$
\begin{aligned}
&P(X = xi | Y = y_j) \\
&= P(X = xi | Y = y_k) \\
&= P(X = xi)
\end{aligned}
$$

- Are Weather and Temperature marginally independent?

yes no

Weather W	*Temperature T*	*P(W,T)*
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$P(X = xi \mid Y = y_j) \\ = P(X = xi \mid Y = y_k) \\ = P(X = xi)$$

- Are Weather and Temperature marginally independent?
  - No. We saw before that knowing the Weather changes our belief about the Temperature
  - E.g. P(hot) = 0.10+0.05=0.15
    P(hot|cloudy) = 0.05/0.6 ≅ 0.083

Weather W	*Temperature T*	*P(W,T)*
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$P(X = xi|Y = y_j) \\ = P(X = xi|Y = y_k) \\ = P(X = xi)$$

- Intuitively (without numbers):
  - Boolean random variable "Canucks win the Stanley Cup this season"
  - Numerical random variable "Canucks' revenue last season" ?
  - Are the two marginally independent?

yes no

Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$P(X = xi|Y = y_j)$$
$$= P(X = xi|Y = y_k)$$
$$= P(X = xi)$$

- Intuitively (without numbers):
  - Boolean random variable "Canucks win the Stanley Cup this season"
  - Numerical random variable "Canucks' revenue last season" ?
  - Are the two marginally independent?
    - No! Without revenue they cannot afford to keep their best players

Exploiting marginal independence

- Recall the product rule: $P(f_2 \wedge f_1) = P(f_2|f_1) \times P(f_1)$
- Thus, $P(X = x \wedge Y = y) = P(X = x|Y = y) \times P(Y = y)$

- If X and Y are marginally independent, then
$$P(X = x) = P(X = x|Y = y)$$

- We thus have $P(X = x \wedge Y = y) = P(X = x) \times P(Y = y)$
  - In distribution form: $P(X, Y) = P(X) \times P(Y)$

- In general, if X_1, …, X_n are marginally independent, then we can represent their JPD as a product of marginal distributions

Exploiting marginal independence

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions

$$P(X_1, \ldots, Xn) = \prod_{i=1}^{n} P(Xi)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, Xn)$ have?

2^n $2n$ $2+n$ n^2

Exploiting marginal independence

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions

$$P(X_1, \ldots, Xn) = \prod_{i=1}^{n} P(Xi)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, Xn)$ have?
  - One entry for each possible world: 2^n

- How many entries would all the marginal distributions have combined?

2^n | $2n$ | $2+n$ | n^2

Exploiting marginal independence

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions

$$P(X_1, \ldots, Xn) = \prod_{i=1}^{n} P(Xi)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, Xn)$ have?
  - One entry for each possible world: 2^n

- How many entries would all the marginal distributions have combined?
  - Each of the n tables only has two entries $P(X_1 = \text{true})$ and $P(X_1 = \text{false})$
  - So, in total: 2n. Exponentially fewer than the JPD!

Lecture Overview

- Recap
  - Conditioning & Inference by Enumeration
  - Bayes Rule & The Chain Rule
- Independence
  - Marginal Independence
  - Conditional Independence

Follow-up Example

- Intuitively (without numbers):
  - Boolean random variable "Canucks win the Stanley Cup this season"
  - Numerical random variable "Canucks' revenue last season" ?
  - Are the two marginally independent?
    - No! Without revenue they cannot afford to keep their best players
  - But they are conditionally independent given the Canucks line-up
    - Once we know who is playing then learning their revenue last year won't change our belief in their chances

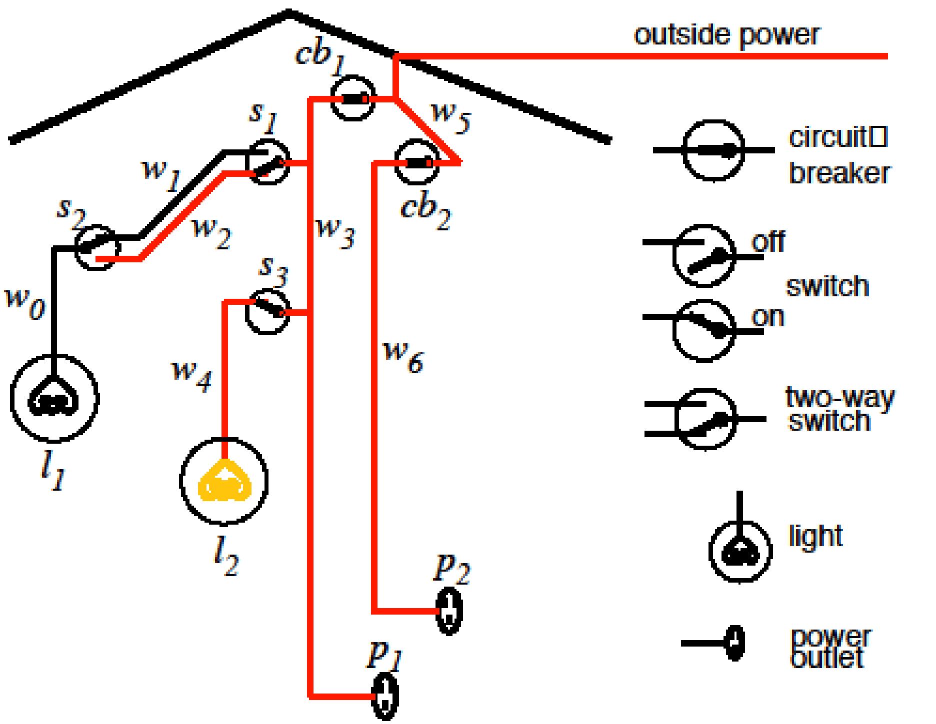
Conditional Independence

Definition (Conditional independence)
Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$ the following equation holds:

$$
\begin{aligned}
&P(X = xi|Y = y_j, Z = zm) \\
&= P(X = xi|Y = y_k, Z = zm) \\
&= P(X = xi|Z = zm)
\end{aligned}
$$

- Intuitively: if X and Y are conditionally independent given Z, then
  - learning that Y=y does not change your belief in X when we already know Z=z
  - and this is true for all values y that Y could take and all values z that Z could take

Example for Conditional Independence

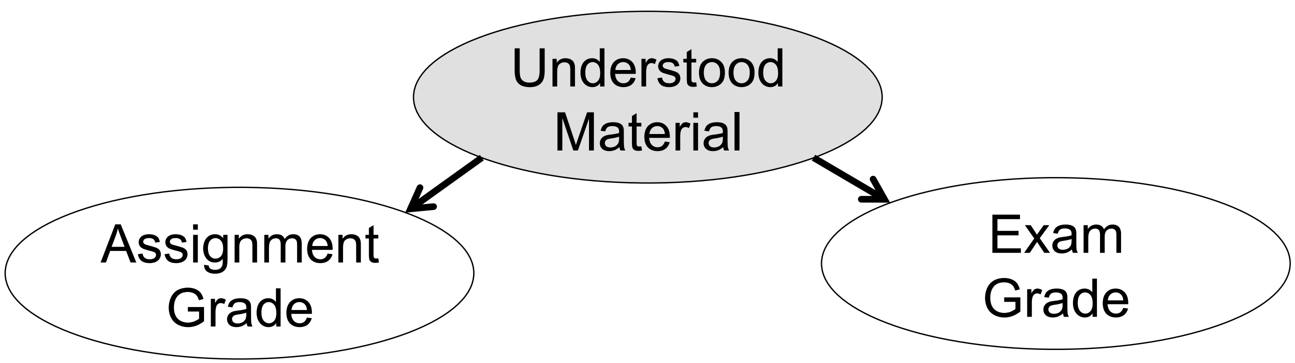


- Whether light l_1 is lit is conditionally independent from the position of switch s_2 given whether there is power in wire w_0
- Once we know Power(w_0) learning values for any other variable will not change our beliefs about Lit(l_1)
  - I.e., Lit(l_1) is independent of any other variable given Power(w_0)

Example: conditionally but not marginally independent

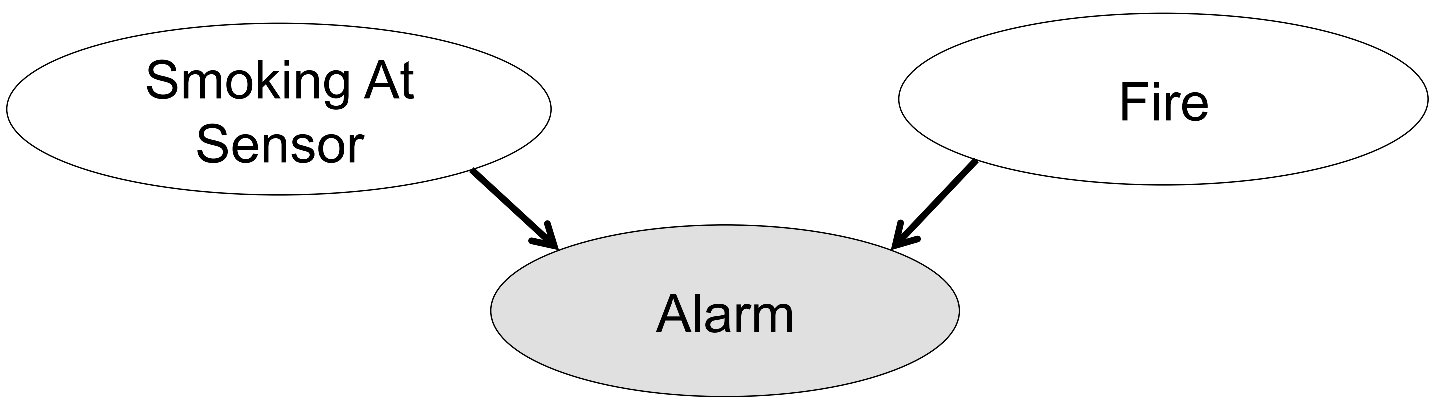
- ExamGrade and AssignmentGrade are not marginally independent
  - Students who do well on one typically do well on the other

- But conditional on UnderstoodMaterial, they are independent
  - Variable UnderstoodMaterial is a common cause of variables ExamGrade and AssignmentGrade
  - UnderstoodMaterial shields any information we could get from AssignmentGrade



Example: marginally but not conditionally independent

- Two variables can be marginally but not conditionally independent
  - "Smoking At Sensor" S: resident smokes cigarette next to fire sensor
  - "Fire" F: there is a fire somewhere in the building
  - "Alarm" A: the fire alarm rings
  - S and F are marginally independent
    - Learning S=true or S=false does not change your belief in F
  - But they are not conditionally independent given alarm
    - If the alarm rings and you learn S=true your belief in F decreases



Conditional vs. Marginal Independence

- Two variables can be
  - Both marginally and conditionally independent
    - CanucksWinStanleyCup and Lit(l_1)
    - CanucksWinStanleyCup and Lit(l_1) given Power(w_0)
  - Neither marginally nor conditionally independent
    - Temperature and Cloudiness
    - Temperature and Cloudiness given Wind
  - Conditionally but not marginally independent
    - ExamGrade and AssignmentGrade
    - ExamGrade and AssignmentGrade given UnderstoodMaterial
  - Marginally but not conditionally independent
    - SmokingAtSensor and Fire
    - SmokingAtSensor and Fire given Alarm

Exploiting Conditional Independence

- Example 1: Boolean variables A,B,C
  - C is conditionally independent of A given B
  - We can then rewrite P(C | A,B) as P(C|B)

Definition (Conditional independence)
Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in$ dom(X), $y_j \in$ dom(Y), $y_k \in$ dom(Y) and $z_m \in$ dom(Z) the following equation holds:

$$
\begin{aligned}
& P(X = xi|Y = y_j, Z = zm) \\
= & P(X = xi|Y = y_k, Z = zm) \\
= & P(X = xi|Z = zm)
\end{aligned}
$$

Exploiting Conditional Independence

- Example 2: Boolean variables A,B,C,D
  - D is conditionally independent of A given C
  - D is conditionally independent of B given C
  - We can then rewrite P(D | A,B,C) as P(D|B,C)
  - And can further rewrite P(D|B,C) as P(D|C)

Definition (Conditional independence)
Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$ the following equation holds:

$$
\begin{aligned}
&P(X = xi | Y = y_j, Z = zm) \\
&= P(X = xi | Y = y_k, Z = zm) \\
&= P(X = xi | Z = zm)
\end{aligned}
$$

Exploiting Conditional Independence

- Recall the chain rule

Theorem (Chain Rule)

$$P(f_n \wedge \cdots \wedge f_1) = \prod_{i=1}^{n} P(fi | f_{i-1} \wedge \cdots \wedge f_1)$$

E.g. $P(A,B,C,D) = P(A) \times P(B|A) \times P(C|A,B) \times P(D|A,B,C)$

- If, e.g., D is conditionally independent of A and B given C, we can rewrite this as
  $P(A,B,C,D) = P(A) \times P(B|A) \times P(C|A,B) \times P(D|C)$

- Under independence, we gain compactness
  - The chain rule lets us represent the JPD as a product of conditional distributions
  - Conditional independence allows us to write them compactly

Learning Goals For Today's Class

- Define and use marginal independence
- Define and use conditional independence

- Assignment 4 available on Connect
  - Due in 2.5 weeks
  - Do the questions *early*
    - Right after the material for the question has been covered in class
    - This will help you stay on track