Directed Graphical Models (part 1 - Discrete random variables)

Francisco Pereira

Filipe Rodrigues

DTU Management Engineering
Department of Management Engineering

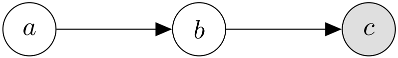
Outline

- Representation
- The concept of inference
- Conditional Probability Tables (CPTs)
- D-Separation

(Based on Michael Jordan, David Blei)

PGM representation

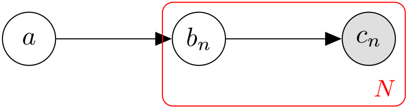
- An example graphical model



- **Nodes** represent random variables
  - shaded nodes correspond to observed variables
  - unshaded nodes denote unobserved variables (also known as hidden or latent variables)
- **Edges** express probabilistic relationships between the variables

PGM representation

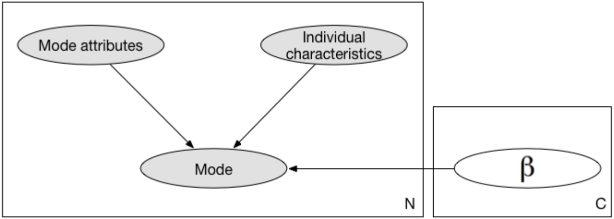
- An example graphical model



- **Nodes** represent random variables
  - shaded nodes correspond to observed variables
  - unshaded nodes denote unobserved variables (also known as hidden or latent variables)
- **Edges** express probabilistic relationships between the variables
- **Plates** indicate repetition

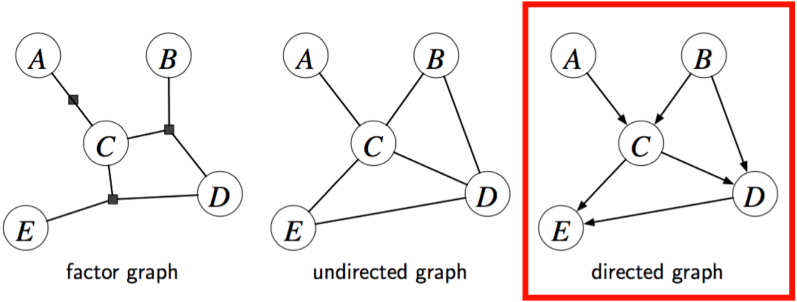
PGM representation

- A practical example



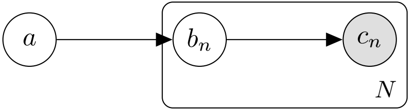
PGM representation

- There are other kinds of graphical models...



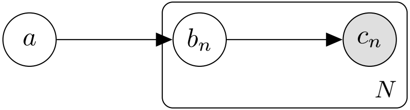
  - Each has different properties and expressiveness
- We will mainly consider **directed** graphical models in this and coming lectures!

PGM Representation



- PGMs represent a set of **conditional independence** relationships

PGM Representation

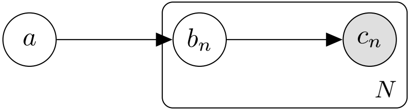


- PGMs represent a set of **conditional independence** relationships

$$c_n \perp\!\!\!\perp a \,|\, b_n \quad (c_n \text{ is conditional independent of } a \text{ given } b_n)$$

  - if we observed b_n, then observing a tell us nothing about c_n

PGM Representation



- PGMs represent a set of **conditional independence** relationships

$$c_n \perp\!\!\!\perp a \,|\, b_n \quad (c_n \text{ is conditional independent of } a \text{ given } b_n)$$

  - if we observed b_n, then observing a tell us nothing about c_n

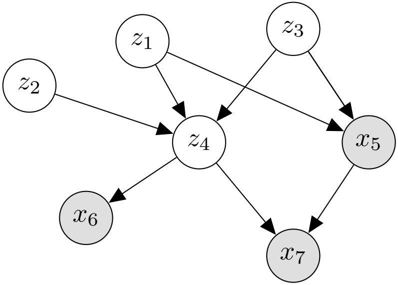
- A PGM specifies a **joint distribution** over variables and how it factorizes:

$$p(a, \mathbf{b}, \mathbf{c}) = p(a) \prod_{n=1}^{N} p(b_n|a)\, p(c_n|b_n)$$

where $\mathbf{b} = \{b_n\}_{n=1}^N$ and $\mathbf{c} = \{c_n\}_{n=1}^N$.

From PGMs to joint distributions

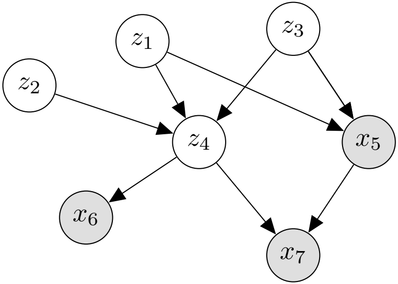
- Another example



- Corresponding factorization of the joint distribution:

From PGMs to joint distributions

- Another example

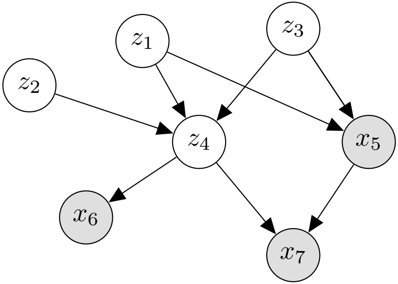


- Corresponding factorization of the joint distribution:

$$\begin{aligned} p(z_1, z_2, z_3, z_4, x_5, x_6, x_7) = {} & p(z_1)\, p(z_2)\, p(z_3)\, p(z_4|z_1, z_2, z_3) \\ & \times p(x_5|z_1, z_3)\, p(x_6|z_4)\, p(x_7|z_4, x_5) \end{aligned}$$

Inference

- **Model + Data $\rightarrow$ Inference**
- Answer various types of questions about the data by computing the posterior distribution of the latent variables given the observed ones



- Example: $p(z_2|x_5, x_6, x_7) = ?$

Inference

- **Exact** inference
  - Set of latent variables $\mathbf{z} = \{z_m\}_{m=1}^{M}$
  - Observed variables $\mathbf{x} = \{x_n\}_{n=1}^{N}$
  - Using Bayes' theorem, the **posterior distribution** of $\mathbf{z}$ can be computed as

$$\overbrace{p(\mathbf{z}|\mathbf{x})}^{\text{posterior}} = \frac{\overbrace{p(\mathbf{x},\mathbf{z})}^{\text{joint}}}{p(\mathbf{x})} = \frac{\overbrace{p(\mathbf{x}|\mathbf{z})}^{\text{likelihood}}\ \overbrace{p(\mathbf{z})}^{\text{prior}}}{\underbrace{p(\mathbf{x})}_{\text{evidence}}}$$

Inference

- **Exact** inference
  - Set of latent variables $\mathbf{z} = \{z_m\}_{m=1}^{M}$
  - Observed variables $\mathbf{x} = \{x_n\}_{n=1}^{N}$
  - Using Bayes' theorem, the **posterior distribution** of $\mathbf{z}$ can be computed as

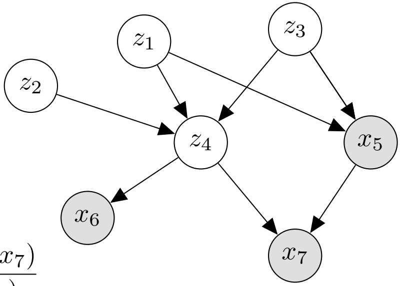
$$\overbrace{p(\mathbf{z}|\mathbf{x})}^{\text{posterior}} = \frac{\overbrace{p(\mathbf{x},\mathbf{z})}^{\text{joint}}}{p(\mathbf{x})} = \frac{\overbrace{p(\mathbf{x}|\mathbf{z})}^{\text{likelihood}}\ \overbrace{p(\mathbf{z})}^{\text{prior}}}{\underbrace{p(\mathbf{x})}_{\text{evidence}}}$$

  - The **model evidence**, or marginal likelihood, can be computed by making use of the sum rule of probability to give

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{x}|\mathbf{z})\, p(\mathbf{z})$$

Inference

- Returning to the previous example...



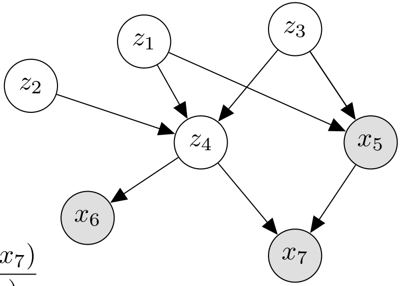
- Assuming discrete variables:

$$p(z_2|x_5, x_6, x_7) = \frac{p(z_2, x_5, x_6, x_7)}{p(x_5, x_6, x_7)}$$
$$\propto \sum_{z_1} \sum_{z_3} \sum_{z_4} p(z_1, z_2, z_3, z_4, x_5, x_6, x_7)$$

- Can be computed exactly (using the sum rule of probability)!

Inference

- Returning to the previous example...



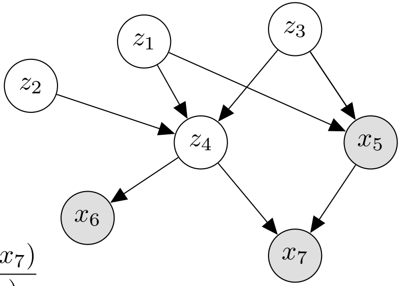
- Assuming discrete variables:

$$p(z_2|x_5, x_6, x_7) = \frac{p(z_2, x_5, x_6, x_7)}{p(x_5, x_6, x_7)}$$
$$\propto \sum_{z_1} \sum_{z_3} \sum_{z_4} p(z_1, z_2, z_3, z_4, x_5, x_6, x_7)$$

- Can be computed exactly (using the sum rule of probability)! *well, not always*

Inference

- Returning to the previous example...

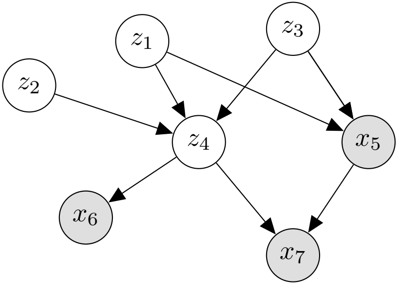


- Assuming discrete variables:

$$p(z_2|x_5, x_6, x_7) = \frac{p(z_2, x_5, x_6, x_7)}{p(x_5, x_6, x_7)}$$
$$\propto \sum_{z_1} \sum_{z_3} \sum_{z_4} p(z_1, z_2, z_3, z_4, x_5, x_6, x_7)$$

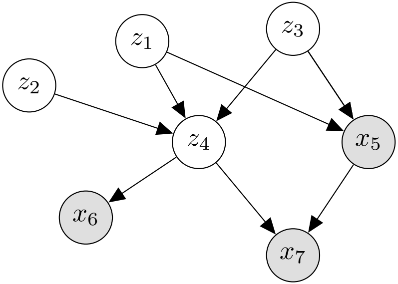
- Can be computed exactly (using the sum rule of probability)! *well, not always*
- Notice that, in this case, we don't need to compute $p(x_5, x_6, x_7)$! We can just renormalize the numerator in the end

Inference



- What if x_6 is missing?

Inference



- What if x_6 is missing?
  - No problem! Just marginalize over its values

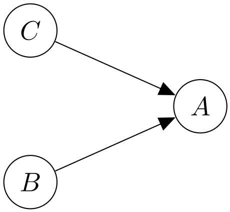
$$p(z_2|x_5, x_7) = \frac{p(z_2, x_5, x_7)}{p(x_5, x_7)}$$
$$\propto \sum_{z_1} \sum_{z_3} \sum_{z_4} \sum_{x_6} p(z_1, z_2, z_3, z_4, x_5, x_6, x_7)$$

- PGMs provide a consistent way of handling **missing data**

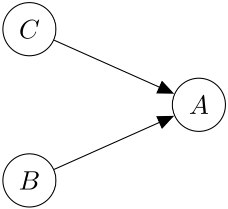
Conditional Probability Tables (CPTs)

- For now, our PGMs have only discrete random variables
- Each node has associated a **Conditional Probability Table**
  - It maps all possible values of its incoming set of arcs...
  - ...to all possible values of the node itself

- For example



Conditional Probability Tables (CPTs)



- This relationship could be defined by:

	$A=0$	$A=1$
$B=0, C=0$	0.7	0.3
$B=0, C=1$	0.3	0.7
$B=1, C=0$	0.5	0.5
$B=1, C=1$	0.1	0.9

Table: **P(A|B,C)**

- It factorizes as

$$P(A,B,C) = P(A|B,C)P(B)P(C)$$

$B=0$	$B=1$
0.4	0.6

Table:
P(B)

$C=0$	$C=1$
0.7	0.3

Table:
P(C)

Conditional Probability Tables (CPTs)

- Imagine we observe $B = 1$. Let's calculate
  - $P(A|B = 1)$

Conditional Probability Tables (CPTs)

- Imagine we observe $B = 1$. Let's calculate
  - $P(A|B = 1)$

- Remember that, for C, we have

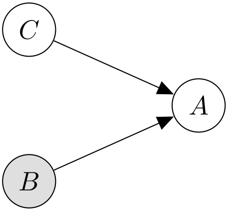
$C = 0$	$C = 1$
0.7	0.3

Table: **P(C)**

Conditional Probability Tables (CPTs)

- Imagine we observe $B = 1$. Let's calculate
  - $P(A|B = 1)$

- Graphical Model



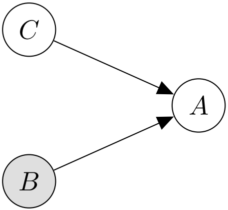
- Remember that, for C, we have

$C = 0$	$C = 1$
0.7	0.3

Table: **P(C)**

Conditional Probability Tables (CPTs)

- Imagine we observe $B = 1$. Let's calculate
  - $P(A|B = 1)$
- Graphical Model



$$P(A|B = 1) =?$$

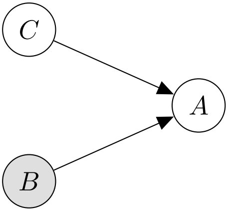
- Remember that, for C, we have

$C = 0$	$C = 1$
0.7	0.3

Table: **P(C)**

Conditional Probability Tables (CPTs)

- Imagine we observe $B = 1$. Let's calculate
  - $P(A|B = 1)$
- Graphical Model



$$P(A|B = 1) = ?$$

$$P(A|B = 1) = \frac{\sum_C P(A, B = 1, C)}{P(B = 1)}$$

- Remember that, for C, we have

$C = 0$	$C = 1$
0.7	0.3

Table: **P(C)**

Conditional Probability Tables (CPTs)

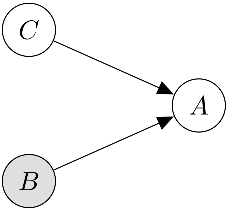
- Imagine we observe $B = 1$. Let's calculate
  - $P(A|B = 1)$

- Remember that, for C, we have

$C = 0$	$C = 1$
0.7	0.3

Table: **P(C)**

- Graphical Model



$$P(A|B = 1) = ?$$

$$P(A|B = 1) = \frac{\sum_C P(A, B = 1, C)}{P(B = 1)} = \frac{\sum_C P(A|B = 1, C)\cancel{P(B = 1)}P(C)}{\cancel{P(B = 1)}}$$

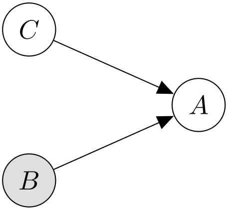
Conditional Probability Tables (CPTs)

- Imagine we observe $B = 1$. Let's calculate
  - $P(A|B = 1)$
- Remember that, for C, we have

$C = 0$	$C = 1$
0.7	0.3

Table: **P(C)**

- Graphical Model



$$P(A|B = 1) =?$$

$$P(A|B = 1) = \frac{\sum_C P(A, B = 1, C)}{P(B = 1)} = \frac{\sum_C P(A|B = 1, C)\cancel{P(B = 1)}P(C)}{\cancel{P(B = 1)}}$$

$$= \sum_C P(A|B = 1, C)P(C) =$$

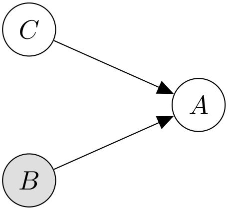
Conditional Probability Tables (CPTs)

- Imagine we observe $B = 1$. Let's calculate
  - $P(A|B = 1)$
- Remember that, for C, we have

$C = 0$	$C = 1$
0.7	0.3

Table: **P(C)**

- Graphical Model



$$P(A|B = 1) = ?$$

$$P(A|B=1) = \frac{\sum_C P(A, B=1, C)}{P(B=1)} = \frac{\sum_C P(A|B=1, C)\cancel{P(B=1)}P(C)}{\cancel{P(B=1)}}$$

$$= \sum_C P(A|B=1, C)P(C) = P(A|B=1, C=1)*0.3 + P(A|B=1, C=0)*0.7$$

Conditional Probability Tables (CPTs)

$$P(A|B=1) = P(A|B=1, C=1) * 0.3 + P(A|B=1, C=0) * 0.7$$

- Considering $P(A|B,C)$:

	$A=0$	$A=1$
$B=0, C=0$	0.7	0.3
$B=0, C=1$	0.3	0.7
$B=1, C=0$	0.5	0.5
$B=1, C=1$	0.1	0.9

- We have:

$P(A=1|B=1) = 0.9 * 0.3 + 0.5 * 0.7 = 0.62$ and $P(A=0|B=1) = 0.1 * 0.3 + 0.5 * 0.7 = 0.38$

- Thus $P(A|B=1)$ will be:

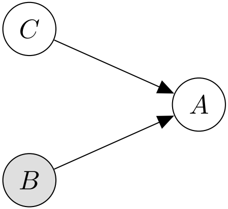
$A=0$	$A=1$
0.38	0.62

Playtime!

- Solve $P(C|B = 1)$
- Estimated time: 20 min

Conditional Probability Tables (CPTs)

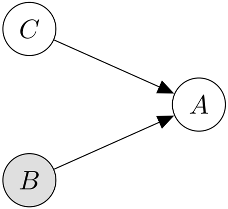
- Indeed, B and C are independent. Just look at the factorization...



$$P(A, B, C) = P(A|B, C)P(B)P(C)$$

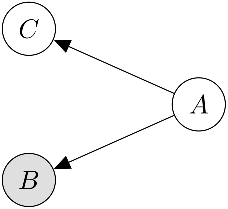
Conditional Probability Tables (CPTs)

- Indeed, B and C are independent. Just look at the factorization...



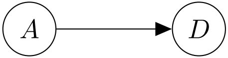
$$P(A, B, C) = P(A|B, C)P(B)P(C)$$

- What if we had instead...



$$P(A, B, C) = P(B|A)P(C|A)P(A)$$

Conditional Probability Tables (CPTs)

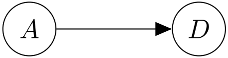


- Another relationship, another CPT:

	$D = 0$	$D = 1$
$A = 0$	0.6	0.4
$A = 1$	0.2	0.8

Table: **P(D|A)**

Conditional Probability Tables (CPTs)



- Another relationship, another CPT:

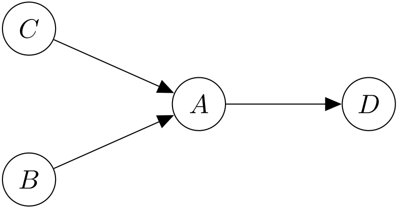
	$D=0$	$D=1$
$A=0$	0.6	0.4
$A=1$	0.2	0.8

Table: **P(D|A)**

- If A is observed, then we can get D directly
- We can conclude that $D \perp\!\!\!\perp B, C|A$
- And also $P(A, B, C, D) = P(A|B, C)P(D|A)P(B)P(C)$

Conditional Probability Tables (CPTs)

- Full model:



- Of course, if we do **not** observe A, then D will depend on the values of B and C.

- Another relationship, another CPT:

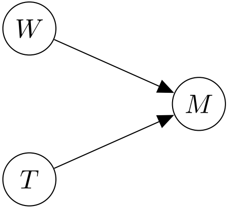
	$D = 0$	$D = 1$
$A = 0$	0.6	0.4
$A = 10$	0.2	0.8

Table: **P(D|A)**

Some magic rules

	We want	We have	We do
1.	$P(A,B)$	$P(A,B,C)$	$P(A,B) = \sum_C P(A,B,C)$
2.	$P(A\|B,C)$	$P(A,B,C)$	$P(A\|B,C) = \frac{P(A,B,C)}{\sum_A P(A,B,C)}$
3.	$P(A\|B)$	$P(A,B,C)$	$P(A\|B) = \frac{\sum_C P(A,B,C)}{\sum_C \sum_A P(A,B,C)}$
4.	$P(A\|B)$	$P(B\|A), P(A)$	$P(A\|B) = \frac{P(B\|A)P(A)}{\sum_A P(B\|A)P(A)}$
5.	$P(A\|B)$	$P(A\|B,C), P(C)$	$P(A\|B) = \sum_C P(A\|B,C)P(C)$

Mode choice - a possible story



	$W=r$	$W=s$
	0.7	0.3

Table: P(W)

	$T=y$	$T=n$
	0.3	0.7

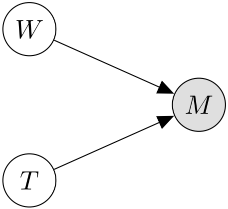
Table: P(T)

- Every day, John needs to make a simple mode choice, M: should he come to work by car or bike?
  - It depends on the schedule constraints, T (e.g. a meeting in a far away place may imply the need of a car)
  - It depends on the weather, W (Sunny or Rainy).

	$M=b$	$M=c$
$T=y, W=r$	0.1	0.9
$T=y, W=s$	0.2	0.8
$T=n, W=r$	0.3	0.7
$T=n, W=s$	0.8	0.2

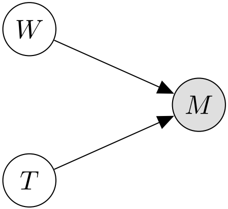
Table: P(M|T, W)

Mode choice - a possible story



- We observe the mode $M = c$
- What is the probability that it is raining?
  - $P(W = r|M = c)$
- Notice that we have $P(M, W, T) = P(M|W, T)P(W)P(T)$

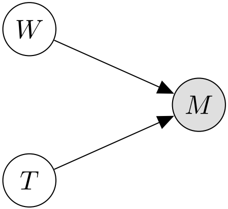
Mode choice - a possible story



- We observe the mode $M = c$
- What is the probability that it is raining?
  - $P(W = r|M = c)$
- Notice that we have $P(M, W, T) = P(M|W, T)P(W)P(T)$

$$P(W = r|M = c) \overset{1,2}{=} \frac{\sum_{t=\{y,n\}} P(M = c, W = r, T = t)}{P(M = c)} =$$

Mode choice - a possible story

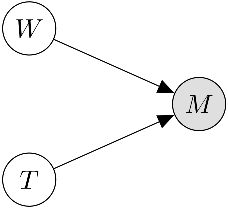


- We observe the mode $M = c$
- What is the probability that it is raining?
  - $P(W = r|M = c)$
- Notice that we have $P(M, W, T) = P(M|W, T)P(W)P(T)$

$$P(W = r|M = c) \overset{1,2}{=} \frac{\sum_{t=\{y,n\}} P(M = c, W = r, T = t)}{P(M = c)} = \frac{\sum_{t=\{y,n\}} P(M = c|W = r, T = t)P(W = r)P(T = t)}{P(M = c)} =$$

$$\overset{3}{=} \frac{\sum_{t=\{y,n\}} P(M = c|W = r, T = t)P(W = r)P(T = t)}{\sum_{w=\{s,r\}} \sum_{t=\{y,n\}} P(M = c, W = w, T = t)} =$$

Mode choice - a possible story

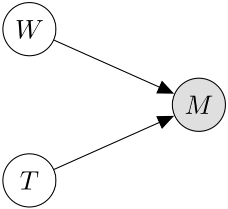


- We observe the mode $M = c$
- What is the probability that it is raining?
  - $P(W = r|M = c)$
- Notice that we have $P(M, W, T) = P(M|W, T)P(W)P(T)$

$$P(W = r|M = c) \stackrel{1,2}{=} \frac{\sum_{t=\{y,n\}} P(M = c, W = r, T = t)}{P(M = c)} = \frac{\sum_{t=\{y,n\}} P(M = c|W = r, T = t)P(W = r)P(T = t)}{P(M = c)} =$$

$$\stackrel{3}{=} \frac{\sum_{t=\{y,n\}} P(M = c|W = r, T = t)P(W = r)P(T = t)}{\sum_{w=\{s,r\}} \sum_{t=\{y,n\}} P(M = c, W = w, T = t)} = \frac{\sum_{t=\{y,n\}} P(M = c|W = r, T = t)P(W = r)P(T = t)}{\sum_{W} \sum_{T} P(M = c|W = w, T = t)P(W = w)P(T = t)} =$$

Mode choice - a possible story



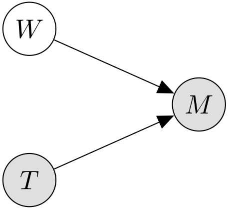
- We observe the mode $M = c$
- What is the probability that it is raining?
  - $P(W = r|M = c)$
- Notice that we have $P(M, W, T) = P(M|W, T)P(W)P(T)$

$$P(W = r|M = c) \stackrel{1,2}{=} \frac{\sum_{t=\{y,n\}} P(M = c, W = r, T = t)}{P(M = c)} = \frac{\sum_{t=\{y,n\}} P(M = c|W = r, T = t)P(W = r)P(T = t)}{P(M = c)} =$$

$$\stackrel{3}{=} \frac{\sum_{t=\{y,n\}} P(M = c|W = r, T = t)P(W = r)P(T = t)}{\sum_{w=\{s,r\}} \sum_{t=\{y,n\}} P(M = c, W = w, T = t)} = \frac{\sum_{t=\{y,n\}} P(M = c|W = r, T = t)P(W = r)P(T = t)}{\sum_W \sum_T P(M = c|W = w, T = t)P(W = w)P(T = t)} =$$

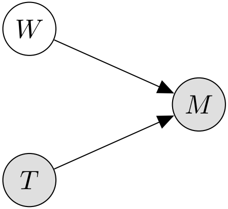
$$= \frac{(0.9 * 0.3 + 0.7 * 0.7) * 0.7}{(0.9 * 0.3 + 0.7 * 0.7) * 0.7 + (0.8 * 0.3 + 0.2 * 0.7) * 0.3} = \frac{0.532}{0.646} = 0.824$$

Explaining away



- What if we **also** observe that the schedule is tight, $T = y$
- Should the probability that it is raining change?...
  - P(W=r|M=c, T=y)

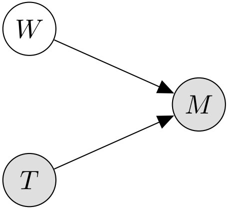
Explaining away



- What if we **also** observe that the schedule is tight, $T = y$
- Should the probability that it is raining change?...
  - P(W=r|M=c, T=y)

$$P(W=r|M=c,T=y) = \frac{P(M=c|W=r,T=y)P(W=r)\cancel{P(T=y)}}{\sum_{w=\{s,r\}} P(M=c|W=w,T=y)P(W=w)\cancel{P(T=y)}} =$$

Explaining away

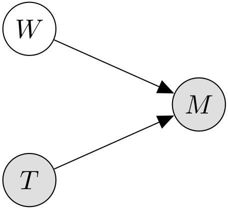


- What if we **also** observe that the schedule is tight, $T = y$
- Should the probability that it is raining change?...
  - P(W=r|M=c, T=y)

$$P(W=r|M=c,T=y) = \frac{P(M=c|W=r,T=y)P(W=r)\cancel{P(T=y)}}{\sum_{w=\{s,r\}} P(M=c|W=w,T=y)P(W=w)\cancel{P(T=y)}} =$$

$$=$$

Explaining away

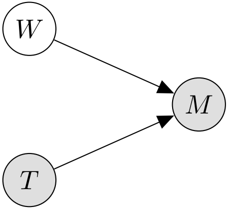


- What if we **also** observe that the schedule is tight, $T = y$
- Should the probability that it is raining change?...
  - P(W=r|M=c, T=y)

$$P(W=r|M=c,T=y) = \frac{P(M=c|W=r,T=y)P(W=r)\cancel{P(T=y)}}{\sum_{w=\{s,r\}} P(M=c|W=w,T=y)P(W=w)\cancel{P(T=y)}} =$$

$$= \frac{P(M=c|W=r,T=y)P(W=r)}{\sum_{w=\{s,r\}} P(M=c|W=w,T=y)P(W=w)} = \frac{0.9*0.7}{(0.9*0.7)+(0.8*0.3)} = \frac{0.63}{0.87} = 0.72$$

Explaining away



- What if we **also** observe that the schedule is tight, $T = y$
- Should the probability that it is raining change?...
  - P(W=r|M=c, T=y)

$$P(W=r|M=c,T=y) = \frac{P(M=c|W=r,T=y)P(W=r)\cancel{P(T=y)}}{\sum_{w=\{s,r\}} P(M=c|W=w,T=y)P(W=w)\cancel{P(T=y)}} =$$

$$= \frac{P(M=c|W=r,T=y)P(W=r)}{\sum_{w=\{s,r\}} P(M=c|W=w,T=y)P(W=w)} = \frac{0.9*0.7}{(0.9*0.7)+(0.8*0.3)} = \frac{0.63}{0.87} = 0.72$$

- What happened is that knowing that the choice of car was *explained away* by the fact that the schedule is tight.
- As if you believe less that John picks the car due to the rain.

Independence properties

- Just by analysing representation, we simplify the calculations!
  - Observed data vs Latent variables (color of node)
  - Arrow directions
  - Conditional independence rules (D-Separation)
- The Bayes net assumption says:

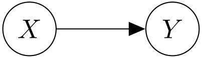
"Each variable is conditionally independent of its non-descendants, given its parents."

D-Separation

- When does **X** influence **Y**?

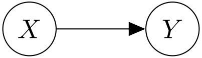
D-Separation

- When does **X** influence **Y**?
- Direct connection:

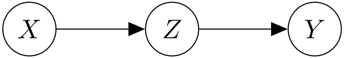


D-Separation

- When does **X** influence **Y**?
- Direct connection:



- Indirect connection:



(**Causal effect**)

D-Separation

- When does **X** influence **Y**?
- Direct connection:

$X \rightarrow Y$

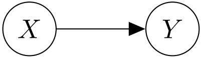
- Indirect connection:

$X \rightarrow Z \rightarrow Y$ (**Causal effect**)

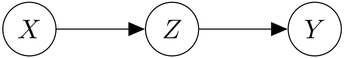
$X \leftarrow Z \leftarrow Y$ (**Evidential effect**)

D-Separation

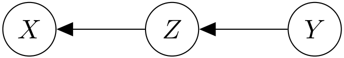
- When does **X** influence **Y**?
- Direct connection:



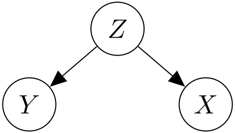
- Indirect connection:



(**Causal effect**)



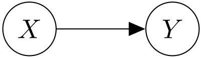
(**Evidential effect**)



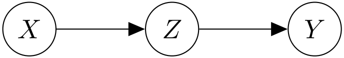
(**Common cause**)

D-Separation

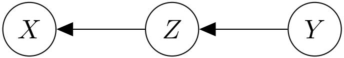
- When does **X** influence **Y**?
- Direct connection:



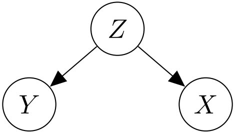
- Indirect connection:



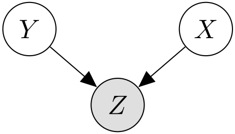
(Causal effect)



(Evidential effect)



(Common cause)



(Common effect)

D-Separation

- When influence can flow from X to Y via Z, we say that the trail X, Y, Z is *active* (otherwise, it is *blocked*).

D-Separation

- When influence can flow from X to Y via Z, we say that the trail X, Y, Z is *active* (otherwise, it is *blocked*).

Causal trail:	$X \rightarrow Z \rightarrow Y$	Active iff Z is not observed
Evidential trail:	$X \leftarrow Z \leftarrow Y$	Active iff Z is not observed
Common cause:	$Y \leftarrow Z \rightarrow X$	Active iff Z is not observed
Common effect:	$Y \rightarrow Z \leftarrow X$	Active iff Z **or one of its descendants are observed**

D-separation: a simple(r) algorithm

For any expression “Is X independent of Y given Z" (formally, $X \perp\!\!\!\perp Y|Z$)[1]

1. Draw the *ancestral graph*
   - It is the part of the original graph that has only the variable sets X, Y and Z, and all their ancestors among them
2. *Moralize* the graph by *marrying* the parents
   - For each pair of variables with a common child, draw an undirected edge (line) between them (If a variable has more than two parents, draw lines between every pair of parents.)
3. *Disorient* the graph by replacing all edges for undirected ones
4. Delete the variables Z (and any other observed variables not explicitly included in Z), and their edges.

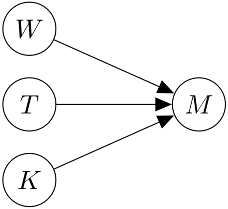
[1]Note that X, Y and Z can themselves be sets of variables!

D-separation: a simple(r) algorithm (cont.)

Analysis of the result

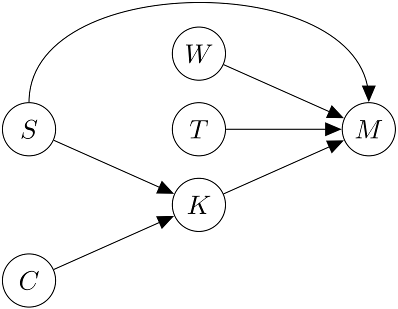
- If X and Y are **disconnected**, then they are conditionally independent given Z!
  - Being disconnected means that there is no possible path between X and Y in the resulting graph
- Otherwise, they are not proven to be independent

Mode choice - a possible story



- Every day, John needs to make a simple mode choice, M: should he come to work by car or bike?
    - It depends on the schedule constraints, T (e.g. a meeting in a far away place may imply the need of a car)
    - It depends on the weather, W (Sunny or Rainy).
    - It also depends on whether he needs to pickup and drop off his kids, K.

Mode choice - a possible story



- ...we can dig further in this problem
  - If there is no school on the calendar, C, of that day, he probably won't need to bring his kids at all.
  - His wife (spouse, S), may bring the kids
  - His wife may need to take the car (in which case, he has to take the kids by bike)

Playtime!

- Please prove:

$$C \not\perp\!\!\!\perp M | K$$

$$C \perp\!\!\!\perp M | \{K, S\}$$

$$\{W, T\} \perp\!\!\!\perp K$$

- Estimated time: 20 min

References

- Koller, D., and Friedman, N. (2009). Probabilistic graphical models: principles and techniques. MIT press.
- "CS 536: Introduction to Graphical Models". Weng-Keen Wong, EECS Oregon State University. http://classes.engr.oregonstate.edu/eecs/winter2015/cs536/slides/bayesnets3.2pp.pdf
- "d-separation: How to determine which variables are independent in a Bayes net". Jessica Noss. EECS MIT. http://web.mit.edu/jmn/www/6.034/d-separation.pdf