CSC384 Summer 2018

Week 9 - Reasoning Under Uncertainty

Ilir Dema

University of Toronto

July 16, 2018

Overview

1. Normalization
2. Independence
3. Bayes Nets

Normalization

- Normalizing a vector of non-negative numbers (with at least one positive component) is described as the process of dividing all components of the vector by the sum of all components.
- Example:
  `normalize([2,2.5,1.5,4])=[0.2,0.25,0.15,0.4]`
- After normalizing the vector of numbers sums to 1
  - Exactly what is needed for these numbers to specify a probability distribution.

PROPERTIES OF NORMALIZATION

- $\texttt{normalize}([x_1,\ldots,x_n]) = [\frac{x_1}{\alpha},\ldots,\frac{x_n}{\alpha}]$ for $\alpha = \sum_{i=1}^{n} x_i$.
- $[x_1,\ldots,x_n] = \alpha\texttt{normalize}([x_1,\ldots,x_n])$
- For any $\beta > 0$, $\texttt{normalize}([x_1,\ldots,x_n]) = \texttt{normalize}([\beta x_1,\ldots,\beta x_n])$.
- `normalize` map is idempotent, i.e. multiple applciations of it no longer will change the (already) normalized vector.

Variable Independence

- With feature vectors we often want to state collection of independencies or conditional independencies
- $V_1 = 1$ is independent of $V_2 = 1$

  $V_1 = 1$ is independent of $V_2 = 2$

  $V_1 = 2$ is independent of $V_2 = 1$

  $V_1 = 2$ is independent of $V_2 = 2$

  ...
- (Different features are independent irrespective of the specific values they take).
- So we often use statements of **variable independence**

Variable Independence

- $P(V_1|V_2) = P(V_1)$ (V_1 and V_2 are independent)
- $P(V_1|V_2, V_3) = P(V_1, V_3)$ (V_1 is conditionally independent of V_2 given V_3)
- It means that the independence holds no matter what value the variable takes.
  - $\forall d_1 \in Dom[V_1], \forall d_2 \in Dom[V_2],\ P(V_1 = d_1|V_2 = d_2) = P(V_1 = d_1)$
  - $\forall d_1 \in Dom[V_1], \forall d_2 \in Dom[V_2], \forall d_3 \in Dom[V_3],\ P(V_1 = d_1|V_2 = d_2, V_3 = d_3) = P(V_1 = d_1|V_3 = d_3)$

Probabilities over variables

- $P(V_1, V_2)$ for variable V_1 and V_2 refers to a set of probabilities, one probability for each pair of values value of V_1 and V_2.
    - It specifies $P(V_1 = d_1, V_2 = d_2)$ for all $d_1 \in Dom[V_1]$ and $d_2 \in Dom[V_2]$
    - E.g., if $Dom[V_1] = Dom[V_2] = \{1,2,3\}$ then $P(V_1, V_2)$ will be a vector of 9 numbers $[P(V_1 = 1, V_2 = 1), P(V_1 = 1, V_2 = 2), P(V_1 = 1, V_2 = 3), P(V_1 = 2, V_2 = 1), P(V_1 = 2, V_2 = 2), P(V_1 = 2, V_2 = 3), P(V_1 = 3, V_2 = 1), P(V_1 = 3, V_2 = 2), P(V_1 = 3, V_2 = 3))]$
    - This vector of probabilities specifies the joint distribution of V_1 and V_2

Conditional Probabilities over Variables

- $P(V_1 | V_2, V_3)$ specifies a **collection** of distributions over V_1, one for each $d_2 \in Dom[V_2]$ and $d_3 \in Dom[V_3]$.
- E.g., if $Dom[V_1] = Dom[V_2] = Dom[V_3] = \{1, 2, 3\}$ then $P(V_1 | V_2, V_3)$ will specify 27 values:

$$\begin{array}{l}
P(V_1 = 1 | V_2 = 1, V_3 = 1), P(V_1 = 2 | V_2 = 1, V_3 = 1), P(V_1 = 3 | V_2 = 1 V_3 = 1) \\
P(V_1 = 1 | V_2 = 1, V_3 = 2), P(V_1 = 2 | V_2 = 1, V_3 = 2), P(V_1 = 3 | V_2 = 1 V_3 = 2) \\
P(V_1 = 1 | V_2 = 1, V_3 = 3), P(V_1 = 2 | V_2 = 1, V_3 = 3), P(V_1 = 3 | V_2 = 1 V_3 = 3) \\
P(V_1 = 1 | V_2 = 2, V_3 = 1), P(V_1 = 2 | V_2 = 2, V_3 = 1), P(V_1 = 3 | V_2 = 2 V_3 = 1) \\
P(V_1 = 1 | V_2 = 2, V_3 = 2), P(V_1 = 2 | V_2 = 2, V_3 = 2), P(V_1 = 3 | V_2 = 2 V_3 = 2) \\
P(V_1 = 1 | V_2 = 2, V_3 = 3), P(V_1 = 2 | V_2 = 2, V_3 = 3), P(V_1 = 3 | V_2 = 2 V_3 = 3) \\
P(V_1 = 1 | V_2 = 3, V_3 = 1), P(V_1 = 2 | V_2 = 3, V_3 = 1), P(V_1 = 3 | V_2 = 3 V_3 = 1) \\
P(V_1 = 1 | V_2 = 3, V_3 = 2), P(V_1 = 2 | V_2 = 3, V_3 = 2), P(V_1 = 3 | V_2 = 3 V_3 = 2) \\
P(V_1 = 1 | V_2 = 3, V_3 = 3), P(V_1 = 2 | V_2 = 3, V_3 = 3), P(V_1 = 3 | V_2 = 3 V_3 = 3)
\end{array}$$

- The values in each row form a different probability distribution.

Conditional Probabilities over Variables

- Useful to think of $P(V_i)$ as a function. Give it a value for V_i it returns a number (a probability). These numbers form a probability distribution. The numbers can be stored in a table.
- Similarly $P(V_1|V_2, V_3)$ is also a function. Give it three values, one for V_1, V_2 and V_3, it will return a number (a conditional probability). Note that for each fixed value of V_2 and V_3 this function specifies a probability distribution over the values of V_1.

  $P(V_1|V_2 = 1, V_3 = 1)$ - a vector of probabilities, one for each assignment to V_1
  $P(V_1|V_2 = 1, V_3 = 2)$ - another distribution over V_1

Conditional Probabilities over Variables

- Recall set notation used in last lecture. Say we have some variables, X taking values in $Dom[X]$ and Y taking values in $Dom[Y]$.
- Then universe $U = Dom[X] \times Dom[Y]$.
- The fact that X and Y are independent, means the events $X = x$ and $Y = y$ are independent. Then the independency theorem implies

  $$P(x, y) = P(x)P(y)$$

  Equivalently, $P(x|y) = P(x)$.
- A generalization using induction, gives **the chain rule**: $P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

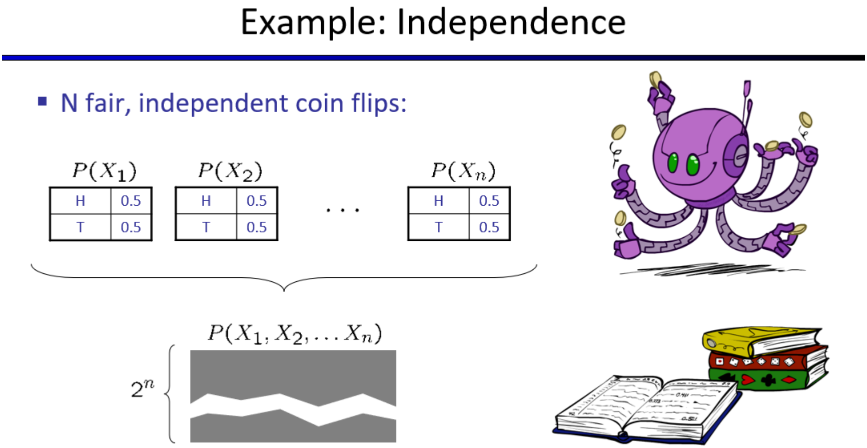
$P_2(T,W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

Credits: Pieter Abbeel



Example: Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

$P(X_2)$

H	0.5
T	0.5

$\ldots$

$P(X_n)$

H	0.5
T	0.5

$P(X_1, X_2, \ldots X_n)$

2^n

Credits: Pieter Abbeel

Conditional Independence and the Chain Rule

- E.g., Traffic, Umbrella, Rain

  `Dom[Traffic] = {Heavy, Normal}`

  `Dom[Umbrella] = {Used, Not used}`

  `Dom[Rain] = {Yes, No}`

- Trivial decomposition

  $\texttt{P(Traffic, Rain, Umbrella)} =$
  $\texttt{P(Rain)P(Traffic}|\texttt{Rain)P(Umbrella}|\texttt{Rain, Traffic)}$

Unconditional Independence

- Unconditional Independence is quite rare in most situations
    - `P(Rain=Yes|Traffic=Heavy) = P(Rain=Yes)`
      No, heavy traffic is evidence for rain.
    - `P(Umbrella=Used|Traffic=Heavy)`
      ` = P(Umbrella=Used)`
      No, heavy traffic is evidence for rain which would influence Umbrella usage
    - `P(Umbrella=Used|Rain=Yes)`
      Definitely not, Rain is main reason for using the Umbrella
- Conditional Independence quite common.
    - `P(Traffic=Heavy, Umbrella=Used| Rain=Yes) =`
      `P(Traffic=Heavy|Rain=Yes)P(Umbrella=Used|Rain=Yes)`
      Yes, once we know the status of Raining, heavy Traffic and Umbrella usage are independent of each other.

Creating Probability Distributions

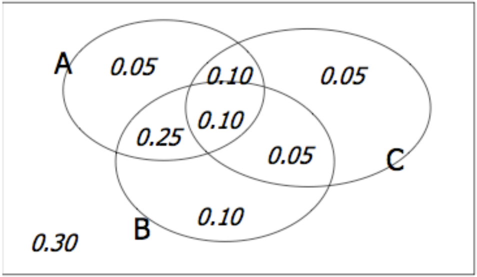
- A joint distribution records the probabilities that variables will hold particular values.
- They can be populated using expert knowledge, by using the axioms of probability, or by actual data.
- The sum of all the probabilities MUST be 1 in order to satisfy the axioms of probability.
- We can use normalization to convert raw counts of data into a legal probability distribution (i.e. into a distribution that sums to 1).

Definition

If X and Y are discrete random variables, the function given by $f(x,y) = P(X = x, Y = y)$ for each pair of values (x,y) is called the joint probability distribution of X and Y .

Creating Probability Distributions

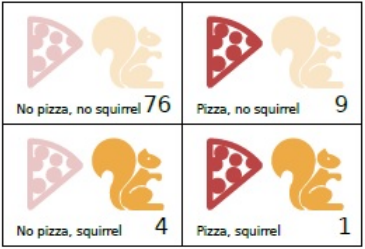
A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10



DATA ANALYSIS EXAMPLE

Now and then the young programmer Jacques transforms into a squirrel. To figure out why, he keeps a diary:

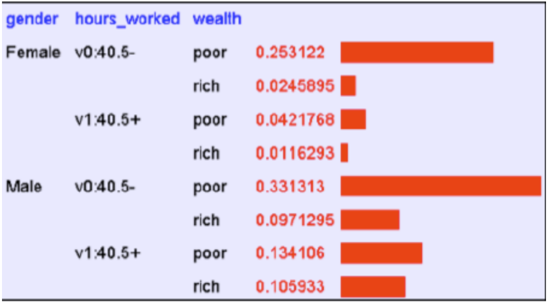
```
"False","pizza","ice cream","computer","work"
```



No pizza, no squirrel 76	Pizza, no squirrel 9
No pizza, squirrel 4	Pizza, squirrel 1

Joint distribution

- Once you have the joint distribution you can ask for the probability of any logical expression involving your attribute.
  $P(E) = \sum_{\text{rows matching } E} P(\text{row})$

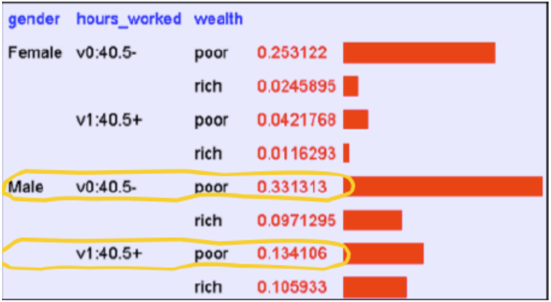


gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

Using Joint distribution

- Once you have the joint distribution you can ask for the probability of any logical expression involving your attribute.

  $P(\text{Wealth=poor, Gender=male}) = 0.331313 + 0.134106 = 0.465419$

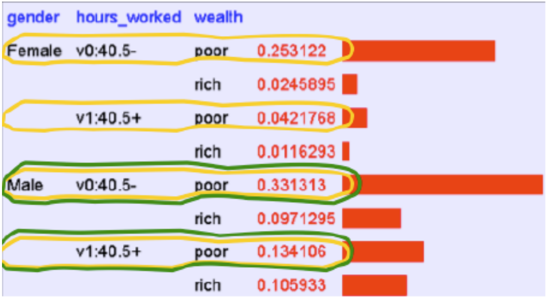


gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

INFERENCE USING JOINT DISTRIBUTION

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{\sum_{\text{rows matching } E_1 \text{ and } E_2} P(\text{row})}{\sum_{\text{rows matching } E_2} P(\text{row})}$$

$$P(Male|Poor) = 0.4654/0.7604 = 0.612$$



gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

Conditional Independence is Symmetric

- Assume $P(X|Y \cap Z) = P(X|Y)$.
- Then:

  $P(Z|X \cap Y)$

  $= P(X \cap Y|Z)P(Z)/P(X \cap Y)$ (Bayes rule)

  $= P(X|Y \cap Z)P(Y|Z)P(Z)/P(X \cap Y)P(Y)$ (Chain rule)

  $= P(X|Y)P(Y|Z)P(Z)/P(X|Y)P(Y)$ (by assumption)

  $= P(Y|Z)P(Z)/P(Y) = P(Z|Y)$ (Bayes rule)

CONDITIONAL PROBABILITIES OVER VARIABLES

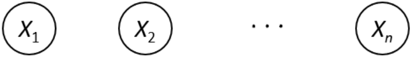
- Two problems with using full joint distribution tables as our probabilistic models:
    - Unless there are only a few variables, the joint is WAY too big to represent explicitly
    - Hard to learn (estimate) anything empirically about more than a few variables at a time
- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
    - More properly called graphical models
    - We describe how variables locally interact
    - Local interactions chain together to give global, indirect interactions

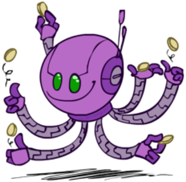
Graphical Model Notation

- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  - Similar to CSP constraints
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence

Example: Coin Flips

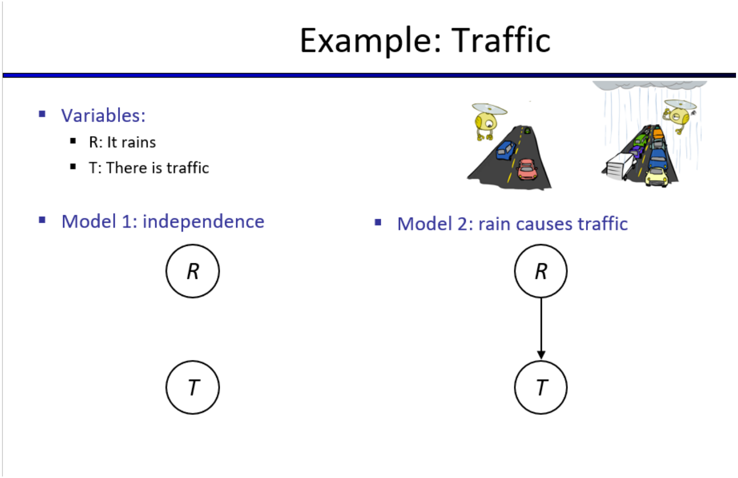
- N independent coin flips





- No interactions between variables: absolute independence

Credits: Pieter Abbeel



Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic

- Model 1: independence

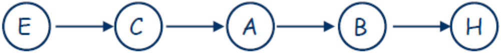
- Model 2: rain causes traffic

Credits: Pieter Abbeel

Exploiting Conditional Independence

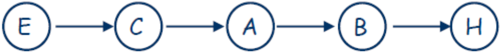
- Consider a story:

  If Craig woke up too early E, Craig probably needs coffee C; if C, Craig needs coffee, he's likely angry A. If A, there is an increased chance of an aneurysm (burst blood vessel) B. If B, Craig is quite likely to be hospitalized H.



E - Craig woke too early
A - Craig is angry
H - Craig hospitalized
C - Craig needs coffee
B - Craig burst a blood vessel

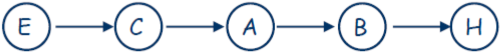
Conditional Independence In Our Story



E - Craig woke too early
C - Craig needs coffee
A - Craig is angry
B - Craig burst a blood vessel
H - Craig hospitalized

- If you learned any of E, C, A, or B, your assessment of $P(H)$ would change.
  - E.g., if any of these are seen to be true, you would increase $P(H)$ and decrease $P(\neg H)$.
  - So H is not **independent** of E, or C, or A, or B.
- But if you knew value of B (true or false), learning the value of E, C, or A, would not influence $P(H)$. Influence these factors have on H is mediated by their influence on B.
  - Craig doesn't get sent to the hospital because he's angry, he gets sent because he's had an aneurysm.
  - So H is **independent** of E, and C, and A, **given** B.

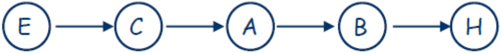
Conditional Independence In Our Story



E - Craig woke too early
C - Craig needs coffee
A - Craig is angry
B - Craig burst a blood vessel
H - Craig hospitalized

- Similarly
  - B is **independent** of E and C, **given** A.
  - A is **independent** of E, **given** C.
- This means that
  - $P(H|B, \{A, C, E\}) = P(H|B)$
    - i.e., for any subset of $\{A, C, E\}$ this relation holds.
  - $P(B|A, \{C, E\}) = P(B|A)$
  - $P(C|E)$ and $P(E)$ don't "simplify".

Conditional Independence In Our Story



E - Craig woke too early A - Craig is angry H - Craig hospitalized
C - Craig needs coffee B - Craig burst a blood vessel

- By the chain rule, for any instantiation of $H, \ldots, E$:

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

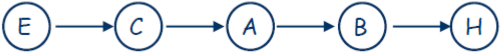
- By our independence assumptions:

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

- We can specify the full joint by specifying five local conditional distributions:

$$P(H|B); P(B|A); P(A|C); P(C|E); P(E)$$

Conditional Independence In Our Story

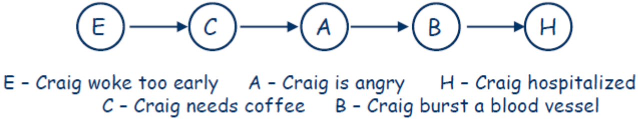


E - Craig woke too early A - Craig is angry H - Craig hospitalized
C - Craig needs coffee B - Craig burst a blood vessel

$P(E) = 0.7$	$P(C\|E) = 0.9$	$P(A\|C) = 0.3$	$P(B\|A) = 0.2$	$P(H\|B) = 0.9$
$P(\neg E) = 0.3$	$P(\neg C\|E) = 0.1$	$P(\neg A\|C) = 0.7$	$P(\neg B\|A) = 0.8$	$P(\neg H\|B) = 0.1$
	$P(C\|\neg E) = 0.5$	$P(A\|\neg C) = 1.0$	$P(B\|\neg A) = 0.1$	$P(H\|\neg B) = 0.1$
	$P(\neg C\|\neg E) = 0.5$	$P(\neg A\|\neg C) = 0.0$	$P(\neg B\|\neg A) = 0.9$	$P(\neg H\|\neg B) = 0.9$

- Specifying the joint distribution over E,C,A,B,H requires only 9 parameters (half the numbers are not needed since, e.g., $P(\neg A|C) + P(A|C) = 1$), instead of 32 for the explicit representation
  - linear in number of vars instead of exponential!
  - linear generally if dependence has a chain structure

Inference is easy



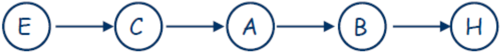
- Want to know $P(a)$? Use summation rule:

$$P(a) = \sum_{c_i \in Dom[C]} P(a|c_i)P(c_i)$$

$$= \sum_{c_i \in Dom[C]} P(a|c_i) \sum_{e_i \in Dom[E]} P(c_i|e_i)P(e_i)$$

These are all terms specified in our local distributions!

Inference is easy



E - Craig woke too early   A - Craig is angry   H - Craig hospitalized
C - Craig needs coffee   B - Craig burst a blood vessel

$P(E) = 0.7$	$P(C\|E) = 0.9$	$P(A\|C) = 0.3$	$P(B\|A) = 0.2$	$P(H\|B) = 0.9$
$P(\neg E) = 0.3$	$P(\neg C\|E) = 0.1$	$P(\neg A\|C) = 0.7$	$P(\neg B\|A) = 0.8$	$P(\neg H\|B) = 0.1$
	$P(C\|\neg E) = 0.5$	$P(A\|\neg C) = 1.0$	$P(B\|\neg A) = 0.1$	$P(H\|\neg B) = 0.1$
	$P(\neg C\|\neg E) = 0.5$	$P(\neg A\|\neg C) = 0.0$	$P(\neg B\|\neg A) = 0.9$	$P(\neg H\|\neg B) = 0.9$

- Computing probabilities in more concrete terms:

$$P(C) = P(C|E)P(E) + P(C|\neg E)P(\neg E)$$
$$= 0.9 \times 0.7 + 0.5 \times 0.3 = 0.78$$

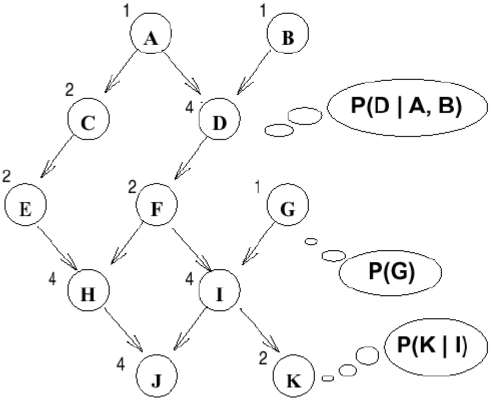
Bayesian Networks

- The structure above is a Bayesian network. A BN is a graphical representation of the direct dependencies over a set of variables, together with a set of conditional probability tables quantifying the strength of those influences.
- Bayes nets generalize the above ideas in very interesting ways, leading to effective means of representation and inference under uncertainty.

Bayesian Networks - Formal Definition

- A BN over variables $\{X_1, X_2, \ldots, X_n\}$ consists of:
  - a DAG (directed acyclic graph) whose nodes are the given variables
  - a set of CPTs (conditional probability tables) $P(X_i|Par(X_i))$ for each X_i
- Key notions:
  - parents of a node: $Par(X_i)$
  - children of a node
  - descendants of a node
  - ancestors of a node
  - family: set of nodes consisting of X_i and its parents
    CPTs are defined over families in the BN

Example (Binary valued Variables)



- A couple of the CPTS are "shown"

Credits: F. Bacchus

Semantics of Bayes Nets

- A Bayes net specifies that the joint distribution over all of the variables in the net can be written as the following product decomposition.

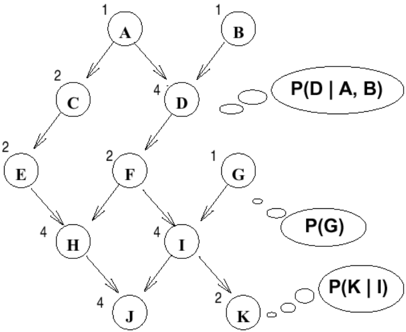
  $P(X_1, X_2, \ldots, X_n) =$
  $P(X_n | Par(X_n))P(X_{n-1} | Par(X_{n-1})) \ldots P(X_1 | Par(X_1))$
- Like other equations over variables this decomposition holds for any set of values $d_1, d_2, \ldots, d_n$ for the variables $X_1, X_2, \ldots, X_n$.

Example

- E.g., have X_1, X_2, X_3, $Dom[X_i] = \{a, b, c\}$, given $P(X_1, X_2, X_3) = P(X_3|X_2)P(X_2)P(X_1)$,

  can write
- $P(X_1 = a, X_2 = a, X_3 = a) = P(X_3 = a|X_2 = a)P(X_2 = a)P(X_1 = a)$
- $P(X_1 = a, X_2 = a, X_3 = b) = P(X_3 = b|X_2 = a)P(X_2 = a)P(X_1 = a)$
- ...

EXAMPLE (BINARY VALUED VARIABLES)



$$P(a, b, c, d, e, f, g, h, i, j, k)$$
$$= P(a)P(b)P(c|a)P(d|a, b)P(e|c)P(f|d)$$
$$P(g)P(h|e, f)P(i|f, g)P(j|h, i)P(k|i)$$

Explicit joint requires $2^{11} - 1 = 2047$ parameters

BN requires only 27 parameters (the number of entries for each CPT is listed)

Constructing a Bayes Net

- Note that this means we can compute the probability of any setting of the variables using only the information contained in the CPTs of the network.
- It is always possible to construct a Bayes net to represent any distribution over the variables $X_1, X_2, \ldots, X_n$, using any ordering of the variables.
- Take any ordering of the variables. From the chain rule we obtain.

$$P(X_1, \ldots, X_n) = P(X_n | X_1, \ldots, X_{n-1}) P(X_{n-1} | X_1, \ldots, X_{n-2}) \ldots P(X_1)$$

- Now for each X_i go through its conditioning set $X_1, \ldots, X_{i-1}$, and remove all variables X_j such that X_i is conditionally independent of X_j given the remaining variables.
- The final product will specify a Bayes net.

Constructing a Bayes Net

- The end result will be a product decomposition/Bayes net $P(X_n|Par(X_n))P(X_{n-1}|Par(X_{n-1}))\ldots Pr(X_1)$
- Now we specify the numeric values associated with each term $P(X_i|Par(X_i))$ in a CPT.
- Typically we represent the CPT as a table mapping each setting of $\{X_i, Par(X_i)\}$ to the probability of X_i taking that particular value given that the variables in $Par(X_i)$ have their specified values.
- If each variable has d different values,
  - We will need a table of size $d^{|\{X_i, Par(X_i)\}|}$.
  - That is, exponential in the size of the parent set.
- Note that the original chain rule

  $$P(X_1,\ldots,X_n) = P(X_n|X_1,\ldots,X_{n-1})P(X_{n-1}|X_1,\ldots,X_{n-2})\ldots P(X_1)$$

  requires as much space to represent as representing the probability of each individual atomic event.

Causal Intuition

- The BN can be constructed using an arbitrary ordering of the variables.
- However, some orderings will yield BN's with very large parent sets. This requires exponential space, and (as we will see later) exponential time to perform inference.
- Empirically, and conceptually, a good way to construct a BN is to use an ordering based on causality. This often yields a more natural and compact BN.

Causal Intuition

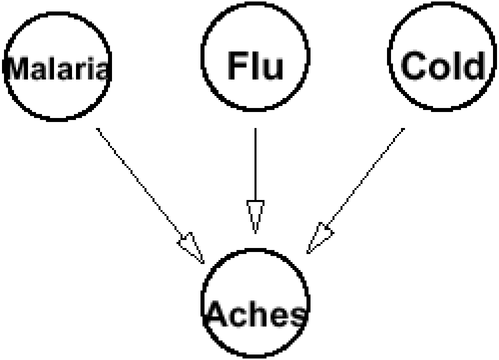
- Malaria, the flu and a cold all “cause” aches. So use the ordering that causes come before effects:

  Malaria, Flu, Cold, Aches
  $P(M, F, C, A) = P(A|M, F, C)P(C|M, F)P(F|M)P(M)$
- Each of these disease affects the probability of aches, so the first conditional probability cannot simplify.
- It is reasonable to assume that these diseases are independent of each other: having or not having one does not change the probability of having the others.
- So $P(C|M, F) = Pr(C)P(F|M) = P(F)$
- This gives us the simplified decomposition of the joint probablity

$$P(M, F, C, A) = P(A|M, F, C)P(C)P(F)P(M)$$

Causal Intuition

- This yields a fairly simple Bayes net.
- Only need one big CPT, involving the family of “Aches”.

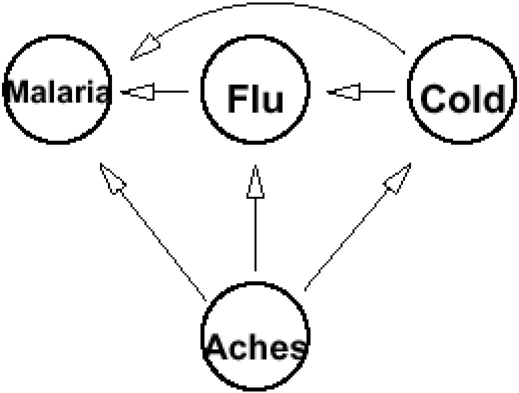


Causal Intuition

- Suppose we build the BN for distribution P using the opposite (non clausal) ordering
- i.e., we use ordering Aches, Cold, Flu, Malaria $P(A, C, F, M) = P(M|A, C, F)P(F|A, C)P(C|A)Pr(A)$
- We can't reduce $P(M|A, C, F)$. Probability of Malaria is clearly affected by knowing aches. What about knowing aches and Cold, or aches and Cold and Flu?
- Probability of Malaria is affected by both of these additional pieces of knowledge
- Knowing Cold and of Flu lowers the probability of Aches indicating Malaria since they "explain away" Aches!
- Similarly, we can't reduce $P(F|A, C)$ - Cold explains away Aches
- $P(C|A) \neq Pr(C)$ - clearly probability of Cold goes up with Aches

Causal Intuition

- Therefore, we obtain a much more complex Bayes net. In fact, we obtain no savings over explicitly representing the full joint distribution (i.e., representing the probability of every atomic event).

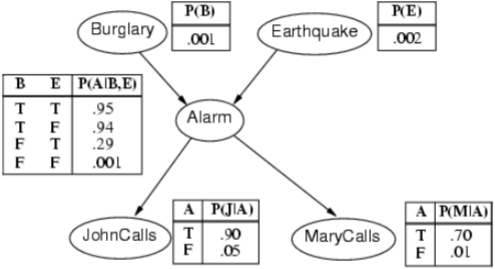


BURGLARY EXAMPLE

- You are at work, neighbor John calls to say your alarm is ringing, but neighbor Mary doesn't call. Sometimes alarm is set off by minor earthquakes. Is there a burglar?
- Variables:
  `Burglary, Earthquake, Alarm, JohnCalls, MaryCall`
- Domains: Each domain is `{True,False}`.
- Joint distribution: 2^5 possible combinations.
- But the alarm does not cause an earthquake, nor does Mary or John calling cause the alarm.
- Network topology reflects "causal" knowledge:
  - Burglar can cause the alarm to go off
  - An earthquake can cause the alarm to go off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

Burglary example

- Burglar can cause the alarm to go off
- An earthquake can cause the alarm to go off
- The alarm can cause Mary to call
- The alarm can cause John to call



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

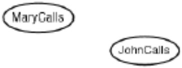
A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Number of parameters: $1 + 1 + 4 + 2 + 2 = 10$

Example of Constructing Bayes Network

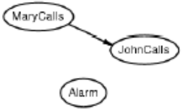
- Suppose we choose the ordering *M, J, A, B, E*
-



P*(J | M)* = ***P****(J)?*

Example continue …

- Suppose we choose the ordering *M, J, A, B, E*
-



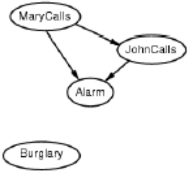
P*(J | M)* = ***P****(J)?*

No

P*(A | J, M)* = ***P****(A | J)*? ***P****(A | J, M)* = ***P****(A)*?

Example continue …

- Suppose we choose the ordering *M, J, A, B, E*
-



P(*J* | *M*) = ***P***(*J*)?

No

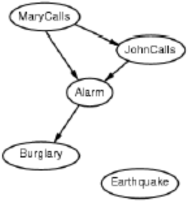
P(*A* | *J*, *M*) = ***P***(*A* | *J*)? ***P***(*A* | *J*, *M*) = ***P***(*A*)? **No**

P(*B* | *A*, *J*, *M*) = ***P***(*B* | *A*)?

P(*B* | *A*, *J*, *M*) = ***P***(*B*)?

Example continue ...

- Suppose we choose the ordering M, J, A, B, E
-



P(*J* | *M*) = ***P***(*J*)?

No

P(*A* | *J*, *M*) = ***P***(*A* | *J*)? ***P***(*A* | *J*, *M*) = ***P***(*A*)? **No**

P(*B* | *A*, *J*, *M*) = ***P***(*B* | *A*)? **Yes**

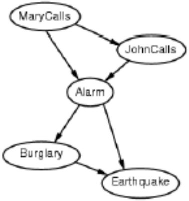
P(*B* | *A*, *J*, *M*) = ***P***(*B*)? **No**

P(*E* | *B*, *A* ,*J*, *M*) = ***P***(*E* | *A*)?

P(*E* | *B*, *A*, *J*, *M*) = ***P***(*E* | *A*, *B*)?

EXAMPLE CONTINUE ...

- Suppose we choose the ordering M, J, A, B, E
-



***P**(J | M) = **P**(J)?*

No

***P**(A | J, M) = **P**(A | J)? **P**(A | J, M) = **P**(A)?* **No**

***P**(B | A, J, M) = **P**(B | A)?* **Yes**

***P**(B | A, J, M) = **P**(B)?* **No**

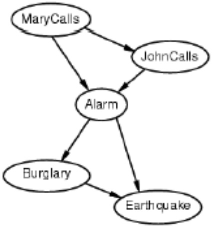
***P**(E | B, A ,J, M) = **P**(E | A)?* **No**

***P**(E | B, A, J, M) = **P**(E | A, B)?* **Yes**

Example continue ...

Example continue…

- Deciding conditional independence **is hard** in non-causal directions!
- (Causal models and conditional independence seem hardwired for humans!)
- Network is **less compact**: 1 + 2 + 4 + 2 + 4 = 13 numbers needed



INFERENCE IN BAYES NETS

- Given a Bayes net

  $P(X_1, X_2, \ldots, X_n)$

  $= P(X_n|Par(X_n))P(X_{n-1}|Par(X_{n-1}))\ldots P(X_1|Par(X_1))$
- and some evidence $E = \{$specific known values for some of the variables$\}$ we want to compute the new probability distribution

  $P(X_k|E)$
- That is, we want to figure out $P(X_k = d|E)$ for all $d \in Dom[X_k]$

Inference in Bayes Nets

- E.g., computing probability of different diseases given symptoms, computing probability of hail storms given different metrological evidence, etc.
- In such cases getting a good estimate of the probability of the unknown event allows us to respond more effectively (gamble rationally)

Inference in Bayes Nets

In the alarm example:

```
P(Burglary,Earthquake, Alarm, JohnCalls, MaryCalls)=
P(Earthquake) P(Burglary)
P(Alarm|Earthquake,Burglary)
P(JohnCalls|Alarm)| P(MaryCalls|Alarm)
```

And, e.g., we want to compute things like

```
P(Burglary=True| MaryCalls=false, JohnCalls=true)
```

So - next week - variable elimination.