Artificial Intelligence

Quantifying Uncertainty & Probabilistic Reasoning

Dr. Bilgin Avenoğlu

Acting under Uncertainty

- Agents in the real world need to handle uncertainty, whether due to partial observability, nondeterminism, or adversaries.

$$\boxed{Toothache \Rightarrow Cavity\,.}$$

$$\boxed{Toothache \Rightarrow Cavity \lor GumProblem \lor Abscess \ldots}$$

$$\boxed{Cavity \Rightarrow Toothache\,.}$$

- The connection between toothaches and cavities is not a strict logical consequence in either direction.
- The agent's knowledge can at best provide only a degree of belief in the relevant sentences.
- Our main tool for dealing with degrees of belief is probability theory.

Basic Probability Notation

- In probability theory, the set of all possible worlds: sample space.
  - if we are about to roll two (distinguishable) dice, there are 36 possible worlds to consider: (1,1), (1,2), ..., (6,6).
  - The Greek letter Ω (uppercase omega) is used to refer to the sample space, and ω (lowercase omega) refers to elements of the space or particular possible worlds.

$$\boxed{0 \leq P(\omega) \leq 1 \text{ for every } \omega \text{ and } \sum_{\omega \in \Omega} P(\omega) = 1\,.}$$

- The probability of a proposition is defined to be the sum of the probabilities of the worlds in which it holds:

$$\boxed{\text{For any proposition } \phi,\ P(\phi) = \sum_{\omega \in \phi} P(\omega)\,.}$$

  - P(Total = 11) = P((5, 6)) + P((6, 5)) = 1/36 + 1/36 = 1/18

Basic Probability Notation

- Probabilities such as P(Total=11) and P(doubles) are called unconditional or prior probabilities,
  - they refer to degrees of belief in propositions in the absence of any other information.
- Probabilities such as "rolling doubles given that the first die is a 5" are called conditional or posterior probability.
  - P(doubles | Die_1 = 5), where the "|" is pronounced "given.
- Conditional probabilities are defined in terms of unconditional probabilities for any propositions *a* and *b*, we have

$$\boxed{P(a|b) = \frac{P(a \wedge b)}{P(b)},} \longrightarrow \boxed{P(a \wedge b) = P(a|b)P(b).}$$

product rule

Basic Probability Notation

- Variables are random variables : a function that maps from the domain of possible worlds Ω to some range
  - The range of Total for two dice is the set {2,...,12} and the range of Die_1 is {1,...,6}.
- Ranges can be sets of arbitrary tokens: Weather {sun, rain, cloud, snow}
- Variables can have infinite ranges - discrete or continuous : *NumberOfAtomsInUniverse* $\geq 10^{70}$
- We can use connectives of propositional logic : $\boxed{P(cavity \mid \neg toothache \wedge teen) = 0.1\,.}$
- We may want to talk about the probabilities of all the possible values of a random variable:

$$\boxed{\begin{aligned} &P(Weather = sun) = 0.6 \\ &P(Weather = rain) = 0.1 \\ &P(Weather = cloud) = 0.29 \\ &P(Weather = snow) = 0.01\,, \end{aligned}}$$

probability distribution

$$\boxed{\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle\,,}$$

Joint probability distribution

- We need notation for distributions on multiple variables.
  - P(Weather, Cavity) denotes the probabilities of all combinations of the values of Weather and Cavity.

$$\boxed{\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather \mid Cavity)\mathbf{P}(Cavity),}$$

$$\boxed{\begin{aligned}
&P(W=sun \wedge C=true) = P(W=sun|C=true)\,P(C=true)\\
&P(W=rain \wedge C=true) = P(W=rain|C=true)\,P(C=true)\\
&P(W=cloud \wedge C=true) = P(W=cloud|C=true)\,P(C=true)\\
&P(W=snow \wedge C=true) = P(W=snow|C=true)\,P(C=true)\\
&P(W=sun \wedge C=false) = P(W=sun|C=false)\,P(C=false)\\
&P(W=rain \wedge C=false) = P(W=rain|C=false)\,P(C=false)\\
&P(W=cloud \wedge C=false) = P(W=cloud|C=false)\,P(C=false)\\
&P(W=snow \wedge C=false) = P(W=snow|C=false)\,P(C=false)\,.
\end{aligned}}$$

Inference Using Full Joint Distributions

Start with the joint distribution:

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬cavity	**.016**	**.064**	**.144**	**.576**

For any proposition φ, sum the atomic events where it is true:

$$P(\varphi) = \sum_{\omega:\omega\models\varphi} P(\omega)$$

Inference Using Full Joint Distributions

Start with the joint distribution:

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	**.108**	**.012**	**.072**	**.008**
$\neg cavity$	**.016**	**.064**	**.144**	**.576**

For any proposition φ, sum the atomic events where it is true:

$$P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$$

$$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Inference Using Full Joint Distributions

Start with the joint distribution:

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬cavity	**.016**	**.064**	**.144**	**.576**

For any proposition φ, sum the atomic events where it is true:

$$P(\varphi) = \Sigma_{\omega : \omega \models \varphi} P(\omega)$$

$$P(cavity \vee toothache) = 0.108+0.012+0.072+0.008+0.016+0.064 = 0.28$$

Inference Using Full Joint Distributions

Start with the joint distribution:

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	**.108**	**.012**	**.072**	**.008**
$\neg cavity$	**.016**	**.064**	**.144**	**.576**

Can also compute conditional probabilities:

$$\begin{aligned} P(cavity \mid toothache) &= \frac{P(cavity \wedge toothache)}{P(toothache)} \\ &= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 . \end{aligned}$$

$$\begin{aligned} P(\neg cavity \mid toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} \\ &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4 . \end{aligned}$$

Normalization

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬cavity	**.016**	**.064**	**.144**	**.576**

Denominator can be viewed as a normalization constant α

$$\begin{aligned} \mathbf{P}(Cavity \mid toothache) &= \alpha\,\mathbf{P}(Cavity, toothache) \\ &= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\ &= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle] \\ &= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle \end{aligned}$$

Normalize by dividing each one by 0.12 + 0.08, getting the true probabilities of 0.6 and 0.4.

General idea: compute distribution on query variable by fixing evidence variables and summing over hidden variables

We can calculate P(*Cavity*|*toothache*) even if we don't know the value of *P*(*toothache*)!

Inference Using Full Joint Distributions

Let $\mathbf{X}$ be all the variables. Typically, we want
the posterior joint distribution of the query variables $\mathbf{Y}$
given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$

Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y} \mid \mathbf{E} = \mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}) = \alpha \Sigma_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}, \mathbf{H} = \mathbf{h})$$

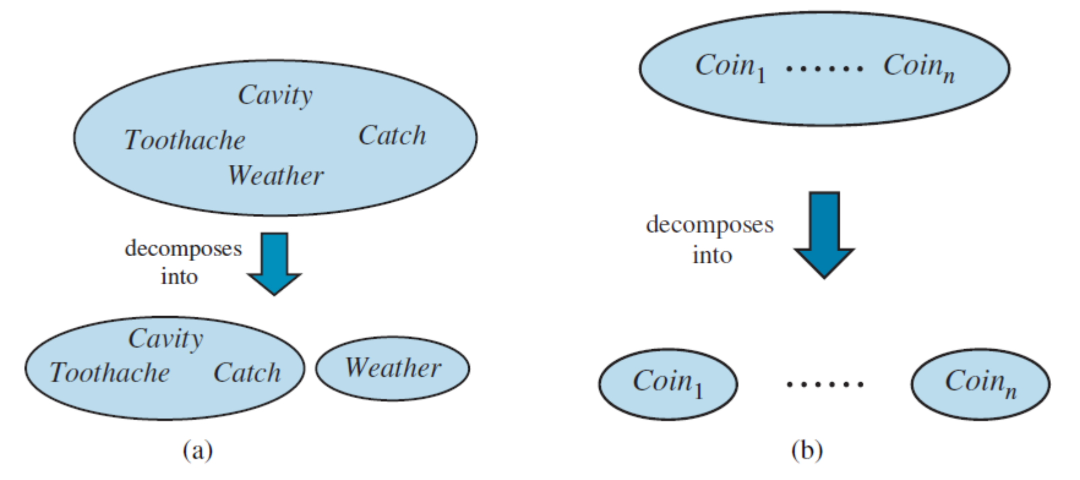
$$\begin{aligned} \mathbf{P}(Cavity \mid toothache) &= \alpha \mathbf{P}(Cavity, toothache) \\ &= \alpha [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \end{aligned}$$

The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables

Obvious problems:
1) Worst-case time complexity $O(d^n)$ where d is the largest arity
2) Space complexity $O(d^n)$ to store the joint distribution
3) How to find the numbers for $O(d^n)$ entries???

Independence

- Two examples of factoring a large joint distribution into smaller distributions, using absolute independence. (a) Weather and dental problems are independent. (b) Coin flips are independent.



product rule

$$P(a \wedge b) = P(a \mid b) P(b) .$$

- $P(a \mid b) = P(a)$ or $P(b \mid a) = P(b)$ or $P(a \wedge b) = P(a)P(b)$.
- one's dental problems influence the weather thus:
  - $P(toothache, catch, cavity, cloud) = P(cloud \mid toothache, catch, cavity)\, P(toothache, catch, cavity)$.
  - $P(cloud \mid toothache, catch, cavity) = P(cloud)$.
  - $P(toothache, catch, cavity, cloud) = P(cloud)P(toothache, catch, cavity)$

Bayes' Rule and Its Use

- Bayes' rule is derived from the product rule
- $P(a \wedge b) = P(a \mid b) P(b)$ and $P(a \wedge b) = P(b \mid a) P(a)$.
- Equating the two right-hand sides and dividing by $P(a)$, we get

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} .$$

- Often, we perceive as evidence the effect of some unknown cause and we would like to determine that cause. In that case, Bayes' rule becomes

$$P(cause|effect) = \frac{P(effect\ |cause)P(cause)}{P(effect)}$$

- The conditional probability $P(effect|cause)$ quantifies the relationship in the **causal** direction, whereas $P(cause|effect)$ describes the **diagnostic** direction.

Bayes' Rule and Its Use

- For example, a doctor knows that the disease meningitis causes a patient to have a stiff neck, say, 70% of the time. The doctor also knows some unconditional facts: the prior probability that any patient has meningitis is 1/50,000, and the prior probability that any patient has a stiff neck is 1%. Letting *s* be the proposition that the patient has a stiff neck and *m* be the proposition that the patient has meningitis, we have

$$P(s|m) = 0.7$$
$$P(m) = 1/50000$$
$$P(s) = 0.01$$

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.7 \times 1/50000}{0.01} = 0.0014$$

- That is, we expect only 0.14% of patients with a stiff neck to have meningitis. Notice that even though a stiff neck is quite strongly indicated by meningitis (with probability 0.7), the probability of meningitis in patients with stiff necks remains small. This is because the prior probability of stiff necks (from any cause) is much higher than the prior for meningitis.

Bayes Rule with Normalization

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

$$P(\neg B|A) = \frac{P(A|\neg B) \cdot P(\neg B)}{P(A)}$$

Given that A is true, B must either be true or false, which means that $P(B|A)$ + $P(\neg B|A) = 1$.

Hence, we can add the two equations above to give

$$1 = \frac{P(A|B) \cdot P(B)}{P(A)} + \frac{P(A|\neg B) \cdot P(B)}{P(A)}$$

$$\therefore P(A) = P(A|B) \cdot P(B) + P(A|\neg B) \cdot P(\neg B)$$

Now we can replace $P(A)$ in the equation for Bayes' theorem, to give

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\neg B) \cdot P(\neg B)}$$

This equation is often written as follows:

$$P(B|A) = \alpha \cdot P(A|B) \cdot P(B)$$

where α represents the normalizing constant:

$$\alpha = \frac{1}{P(A|B) \cdot P(B) + P(A|\neg B) \cdot P(\neg B)}$$

The general form of Bayes' rule with normalization is

$$\boxed{\mathbf{P}(Y \mid X) = \alpha \, \mathbf{P}(X \mid Y) \mathbf{P}(Y),}$$

Example

- A : "I have a high temperature" -> P(A) = 0.001
- B : "I have a cold," -> P(B) = 0.0001
- P(A|B) = 0.8
- The likelihood that a person will have a high temperature (A) if she does not have a cold (¬B) -> P(A|¬B) = 0.00099.

$$
\begin{aligned}
P(B|A) &= \frac{P(A|B)\cdot P(B)}{P(A|B)\cdot P(B) + P(A|\neg B)\cdot P(\neg B)} \\
&= \frac{0.8\cdot 0.0001}{0.8\cdot 0.001 + 0.00099\cdot 0.9999} \\
&= \frac{0.00008}{0.001069901} \\
&= 0.075
\end{aligned}
$$

$$
\begin{aligned}
P(\neg B|A) &= \frac{P(A|\neg B)\cdot P(\neg B)}{P(A|\neg B)\cdot P(\neg B) + P(A|B)\cdot P(B)} \\
&= \frac{0.00099\cdot 0.9999}{0.00099\cdot 0.9999 + 0.8\cdot 0.0001} \\
&= \frac{0.000989901}{0.001069901} \\
&= 0.925
\end{aligned}
$$

Naïve Bayes

- https://www.youtube.com/watch?v=O2L2Uv9pdDA

- The fundamental Naive Bayes assumption is that each feature makes an:
  - independent
  - equal
- contribution to the outcome.

Example

The Bayes Naive classifier selects the most likely classification V_{nb} given the attribute values $a_1, a_2, \ldots a_n$. This results in:

$$V_{nb} = \text{argmax}_{v_j \in V} P(v_j) \prod P(a_i|v_j) \tag{1}$$

We generally estimate $P(a_i|v_j)$ using m-estimates:

$$P(a_i|v_j) = \frac{n_c + mp}{n + m} \tag{2}$$

where:

n =	the number of training examples for which $v = v_j$
n_c =	number of examples for which $v = v_j$ and $a = a_i$
p =	a priori estimate for $P(a_i|v_j)$
m =	the equivalent sample size

Classify "Red Domestic SUV"

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

$P(\text{Red}|\text{Yes})$, $P(\text{SUV}|\text{Yes})$, $P(\text{Domestic}|\text{Yes})$,

$P(\text{Red}|\text{No})$, $P(\text{SUV}|\text{No})$, and $P(\text{Domestic}|\text{No})$

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

```
Yes:
    Red:
        n = 5
        n_c= 3
        p = .5
        m = 3
    SUV:
        n = 5
        n_c = 1
        p = .5
        m = 3
    Domestic:
        n = 5
        n_c = 2
        p = .5
        m = 3
No:
    Red:
        n = 5
        n_c = 2
        p = .5
        m = 3
    SUV:
        n = 5
        n_c = 3
        p = .5
        m = 3
    Domestic:
        n = 5
        n_c = 3
        p = .5
        m =3
```

$n =$	the number of training examples for which $v = v_j$
$n_c =$	number of examples for which $v = v_j$ and $a = a_i$
$p =$	a priori estimate for $P(a_i \vert v_j)$
$m =$	the equivalent sample size

Looking at $P(Red|Yes)$, we have 5 cases where v_j = Yes , and in 3 of those cases a_i = Red. So for $P(Red|Yes)$, n = 5 and n_c = 3. Note that all attribute are binary (two possible values). We are assuming no other information so, p = 1 / (number-of-attribute-values) = 0.5 for all of our attributes. Our m value is arbitrary, (We will use m = 3) but consistent for all attributes. Now we simply apply eqauation (3) using the precomputed values of n , n_c, p, and m.

$$P(Red|Yes) = \frac{3 + 3 * .5}{5 + 3} = .56 \qquad P(Red|No) = \frac{2 + 3 * .5}{5 + 3} = .43$$

$$P(SUV|Yes) = \frac{1 + 3 * .5}{5 + 3} = .31 \qquad P(SUV|No) = \frac{3 + 3 * .5}{5 + 3} = .56$$

$$P(Domestic|Yes) = \frac{2 + 3 * .5}{5 + 3} = .43 \qquad P(Domestic|No) = \frac{3 + 3 * .5}{5 + 3} = .56$$

We have $P(Yes) = .5$ and $P(No) = .5$, so we can apply equation (2). For $v = Yes$, we have

```
P(Yes) * P(Red | Yes) * P(SUV | Yes) * P(Domestic|Yes)

    =  .5 * .56 * .31 * .43   = .037
```

$$\boxed{V_{nb} = \text{argmax}_{v_j \in V} P(v_j) \prod P(a_i|v_j)}$$

and for $v = No$, we have

```
P(No) * P(Red | No) * P(SUV | No) * P (Domestic | No)

    = .5 * .43 * .56 * .56 = .069
```

Since $0.069 > 0.037$, our example gets classified as 'NO'

Bayesian Belief Networks

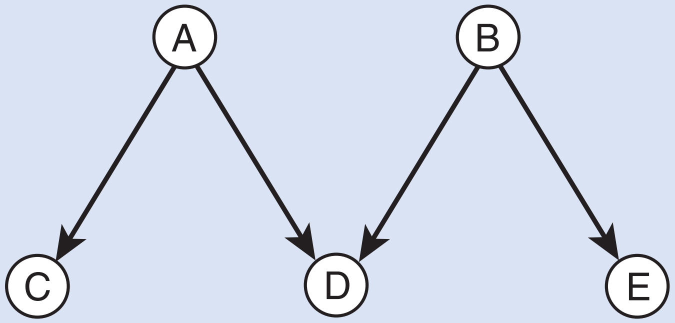
- If A and B are independent, then the probability that A and B will both occur can be calculated very simply:

$$\boxed{P(A \wedge B) = P(A).P(B)}$$

- We know that this equation does not hold if A depends on B :

$$\boxed{P(B|A) = \frac{P(B \wedge A)}{P(A)}}$$

- A Bayesian belief network is an acyclic directed graph,
  - the nodes in the graph represent evidence or hypotheses,
  - an arc that connects two nodes represents a dependence between those two nodes.



Bayesian Belief Networks

- The following conditional probabilities might be used in the network

$P(A) = 0.1$

$P(B) = 0.7$

$P(C|A) = 0.2$

$P(C|\neg A) = 0.4$

$P(D|A \wedge B) = 0.5$

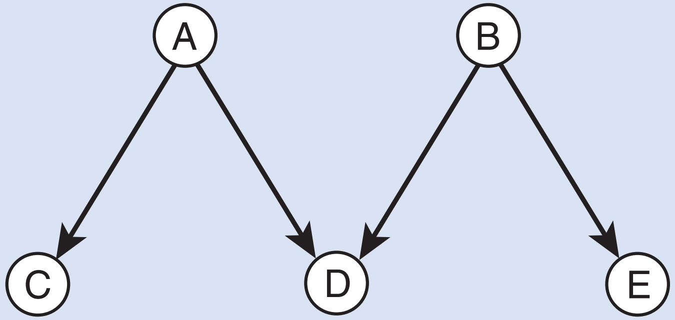
$P(D|A \wedge \neg B) = 0.4$

$P(D|\neg A \wedge B) = 0.2$

$P(D|\neg A \wedge \neg B) = 0.0001$

$P(E|B) = 0.2$

$P(E|\neg B) = 0.1$



P(A)
0.1

P(B)
0.7

A	P(C)
true	0.2
false	0.4

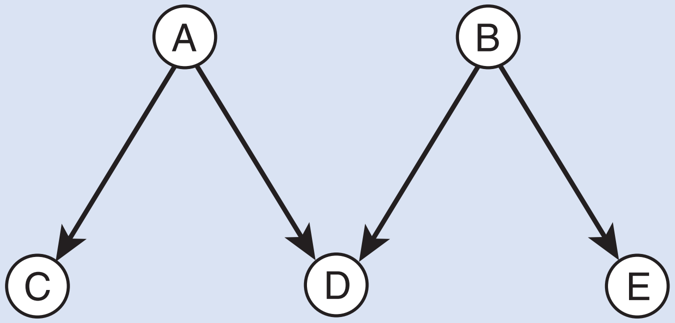
B	P(E)
true	0.2
false	0.1

A	B	P(D)
true	true	0.5
true	false	0.4
false	true	0.2
false	false	0.0001

Bayesian Belief Networks

A joint probability can be calculated from the Bayesian belief network using the definition of conditional probability:

$$P(B|A) = \frac{P(B \wedge A)}{P(A)}$$



Hence,

$$P(A,B,C,D,E) = P(E|A,B,C,D) \cdot P(A,B,C,D)$$

We can apply this rule recursively to obtain

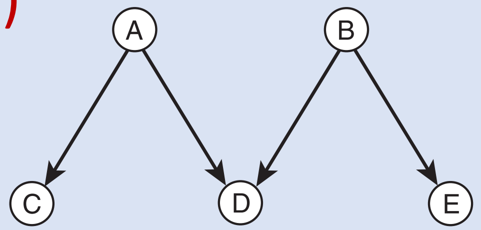
$$P(A,B,C,D,E) = P(E|A,B,C,D,) \cdot P(D|A,B,C,) \cdot P(C|A,B) \cdot P(B|A) \cdot P(A)$$

In fact, the nature of our belief network allows us to simplify this expression, and because we know that, for example, E is not dependent on A, C, or D, we can reduce $P(E|A,B,C,D)$ to $P(E|B)$.

$$P(A,B,C,D,E) = P(E|B) \cdot P(D|A,B) \cdot P(C|A) \cdot P(B) \cdot P(A)$$

Bayesian Belief Networks

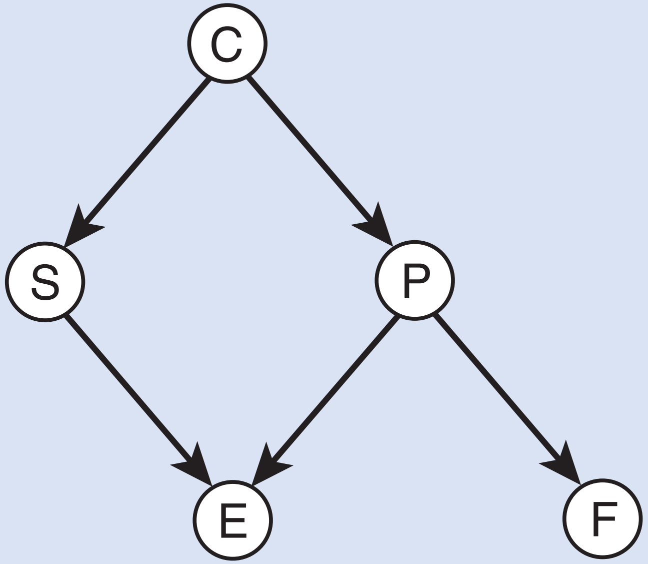
- If we used the same method blindly on the expression P(E,D,C,B,A)
  - $P(E,D,C,B,A) = P(A|E,D,C,B) * P(B|E,D,C) * P(C|E,D) * P(D|E) * P(E)$



- This is not correct because
  - E is dependent on B, and so we need to include P(E|B).
  - D is dependent on A and B, which is not reflected in this expression.
- The nodes must be ordered in the expression in such a way that if a node X is dependent on another node Y, then Y appears before X in the joint.
- We could have used any ordering in which A and B appear before C, D, and E;
  - B,A,E,D,C would have worked equally well, for example.

Example: Life at College

- If you go to college, this will affect the likelihood that you will study and the likelihood that you will party.
- Studying and partying affect your chances of exam success, and partying affects your chances of having fun.



C = that you will go to college

S = that you will study

P = that you will party

E = that you will be successful in your exams

F = that you will have fun

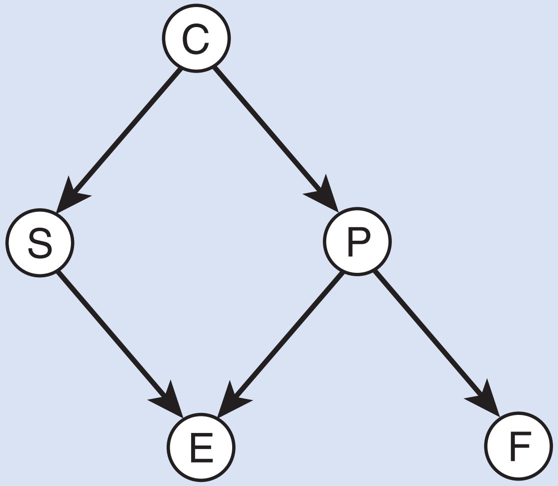
Example: Life at College

- There is a dependence between F and C, but because it is not a direct dependence, no information needs to be stored about it.
- Obtain values such as P(¬C) by using the fact that

$$\boxed{P(\neg C) = 1 - P(C) = 1 - 0.2 = 0.8.}$$

- if P is true, then P(F) = 0.9

$$\boxed{P(F|P) = 0.9.}$$



P(C)
0.2

P	P(F)
true	0.9
false	0.7

C	P(S)
true	0.8
false	0.2

C	P(P)
true	0.6
false	0.5

S	P	P(E)
true	true	0.6
true	false	0.9
false	true	0.1
false	false	0.2

Example: Life at College

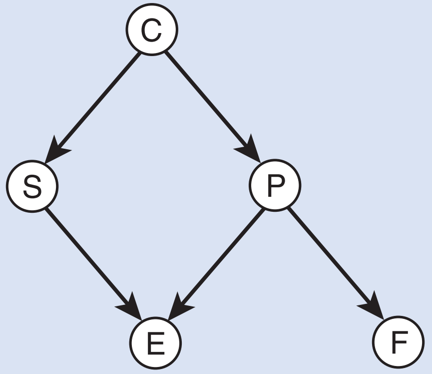
- The probability that you will go to college and that you will study and be successful in your exams, but will not party or have fun.

$$\boxed{P(C = \text{true}, S = \text{true}, P = \text{false}, E = \text{true}, F = \text{false})}$$

$$\boxed{P(C, S, \neg P, E, \neg F)}$$

- Where E is the evidence on which each x_i is dependent - in other words, in the Bayesian belief network, E consists of the nodes that are parents of x_i.

$$\boxed{P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | E)}$$



Example: Life at College

$$P(x_1,\ldots,x_n)=\prod_{i=1}^{n}P(x_i|E)$$

$$P(C, S, \neg P, E, \neg F)$$

$$\begin{aligned}P(C,S,\neg P,E,\neg F) &= P(C)\cdot P(S|C)\cdot P(\neg P|C)\cdot P(E|S\wedge\neg P)\cdot P(\neg F|\neg P)\\ &= 0.2\cdot 0.8\cdot 0.4\cdot 0.9\cdot 0.3\\ &= 0.01728\end{aligned}$$

The probability that you will go to college and that you will study and be successful in your exams, but will not party or have fun.

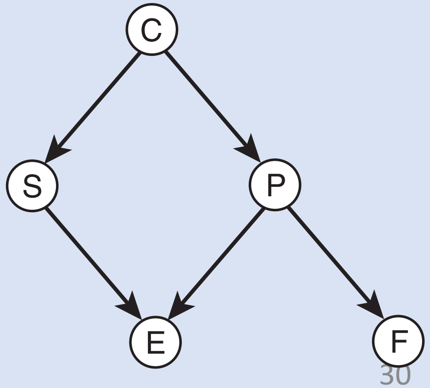
P(C)
0.2

P	P(F)
true	0.9
false	0.7

C	P(S)
true	0.8
false	0.2

C	P(P)
true	0.6
false	0.5

S	P	P(E)
true	true	0.6
true	false	0.9
false	true	0.1
false	false	0.2



Example: Life at College

- The probability that you will have success in your exams if you have fun and study at college, but don't party.

$$P(E|F \wedge \neg P \wedge S \wedge C)$$

- The assumption behind the Bayesian belief network is that because there is no direct connection between E and C, E is independent of C, given S and P.

$$P(E|C \wedge S \wedge P) \rightarrow P(E|S \wedge P) = 0.6$$

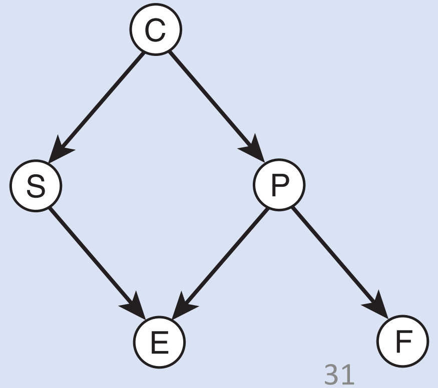
P(C)
0.2

P	P(F)
true	0.9
false	0.7

C	P(S)
true	0.8
false	0.2

C	P(P)
true	0.6
false	0.5

S	P	P(E)
true	true	0.6
true	false	0.9
false	true	0.1
false	false	0.2



Example: Life at College

$$\boxed{P(E|F \wedge \neg P \wedge S \wedge C)}$$

- can be simplified by dropping F and C to give

$$\boxed{P(E|S \wedge \neg P) = 0.9}$$

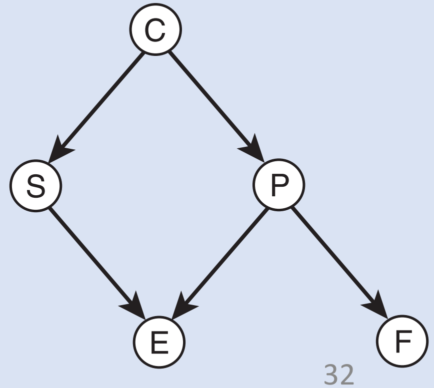
P(C)
0.2

P	P(F)
true	0.9
false	0.7

C	P(S)
true	0.8
false	0.2

C	P(P)
true	0.6
false	0.5

S	P	P(E)
true	true	0.6
true	false	0.9
false	true	0.1
false	false	0.2



Example: Life at College

- We know that you had fun and studied hard while at college and we know that you succeeded in your exams, but we want to know whether you partied or not.
- We know C, S, E, and F, but we do not know P.

$$
\begin{aligned}
P(C \wedge S \wedge P \wedge E \wedge F) &= P(C) \cdot P(S|C) \cdot P(P|C) \cdot P(E|S \wedge P) \cdot P(F|P) \\
&= 0.2 \cdot 0.8 \cdot 0.6 \cdot 0.6 \cdot 0.9 \\
&= 0.05184
\end{aligned}
$$

$$
\begin{aligned}
P(C \wedge S \wedge \neg P \wedge E \wedge F) &= P(C) \cdot P(S|C) \cdot P(\neg P|C) \cdot P(E|S \wedge \neg P) \cdot P(F|\neg P) \\
&= 0.2 \cdot 0.8 \cdot 0.4 \cdot 0.9 \cdot 0.7 \\
&= 0.04032
\end{aligned}
$$

- It is slightly more likely that you did party while at college than that you did not.

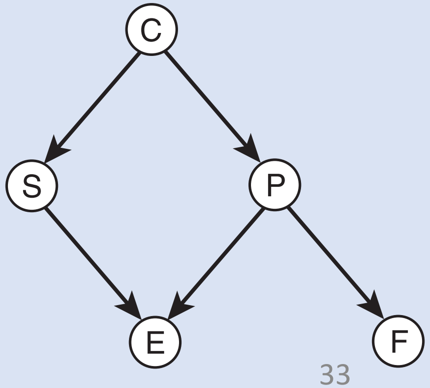
P(C)
0.2

P	P(F)
true	0.9
false	0.7

C	P(S)
true	0.8
false	0.2

C	P(P)
true	0.6
false	0.5

S	P	P(E)
true	true	0.6
true	false	0.9
false	true	0.1
false	false	0.2



The End!