CS60045 Artificial Intelligence
Autumn 2023

Reasoning under Uncertainty

Random variables

We consider only finitely many random variables $V_1, V_2, \ldots, V_n$ for a given problem.

The number n of rv's we are dealing with may be large.

For our purpose, we assume that each rv is a proposition, that is, a statement about some constant object in the problem we like to solve.

The domain of each V_i is $\{T, F\}$.

We write $\Pr[V_i = T]$ as $\Pr[V_i]$, and $\Pr[V_i = F]$ as $\Pr[\neg V_i]$.

Our rv's are, in general, not like the outcomes of tosses of fair coins.

That is, $\Pr[V_i]$ and $\Pr[\neg V_i]$ may have arbitrary values in the real interval $[0, 1]$.

Of course, we must always have $\Pr[V_i] + \Pr[\neg V_i] = 1$.

Notational inconsistency: In the notation $\Pr[V]$, we may refer to V as a variable, and also as the constant $V = T$. The context would make it clear which interpretation we are talking about.

Joint distribution

For truth values $\theta_1, \theta_2, \ldots, \theta_n$, we write the joint probability as

$$\begin{aligned} & \Pr[V_1 = \theta_1 \wedge V_2 = \theta_2 \wedge \cdots \wedge V_n = \theta_n] \\ = \; & \Pr[V_1 = \theta_1, V_2 = \theta_2, \ldots, V_n = \theta_n] \\ = \; & \Pr[\theta_1, \theta_2, \ldots, \theta_n]. \end{aligned}$$

For example, $\Pr[A \wedge \neg B \wedge \neg C \wedge D] = \Pr[A, \neg B, \neg C, D] = \Pr[T, F, F, T]$. We will always have

$$\sum \Pr[V_1, V_2, \ldots, V_n] = 1.$$

Our rv's are not necessarily independent, that is, in general,

$$\Pr[V_1 = \theta_1, V_2 = \theta_2, \ldots, V_n = \theta_n] \neq \Pr[V_1 = \theta_1] \times \Pr[V_2 = \theta_2] \times \cdots \times \Pr[V_n = \theta_n].$$

The complete joint distribution on n variables will have 2^n entries. This may be prohibitively large. Even gathering data for so many possibilities is a practical infeasibility.

In order to make progress, we will oftentimes make some assumptions of independence that can be derived from or that may approximate real-life situations.

Marginal probability

For a subset $\mathcal{V}$ of the rv's, the probability that the variables in $\mathcal{V}$ assume given values can be obtained by fixing these truth values of the variables in $\mathcal{V}$ and by summing the joint probabilities for all possible values of the variables not in $\mathcal{V}$:

$$\Pr[\mathcal{V}] = \sum_{\mathcal{V}} \Pr[V_1, V_2, \ldots, V_n].$$

Example: For five variables A, B, C, D, E, the probability of $B = T$, $C = T$, and $E = F$ is

$$\Pr[B, C, \neg E] = \Pr[A, B, C, D, \neg E] + \Pr[A, B, C, \neg D, \neg E] + \Pr[\neg A, B, C, D, \neg E] + \Pr[\neg A, B, C, \neg D, \neg E].$$

Marginal distributions may again be infeasible to compute (for large n), because the sum may involve a huge number of terms.

Conditional probability

Let $\mathcal{V}$ and $\mathcal{W}$ be sets of rv's. The conditional probability that the variables in $\mathcal{V}$ assume specific values given that the variables in $\mathcal{W}$ assume specific values is

$$Pr[\mathcal{V} \mid \mathcal{W}] = \frac{\Pr[\mathcal{V}, \mathcal{W}]}{\Pr[\mathcal{W}]}.$$

The probabilities on the right side are marginal probabilities.

Example: For five variables A, B, C, D, E, the probability that $C = T$ given that $B = T$, $D = T$, and $E = F$ is

$$\begin{aligned} & \Pr[C \mid B, D, \neg E] \\ = \quad & \frac{\Pr[B, C, D, \neg E]}{\Pr[B, D, \neg E]} \\ = \quad & \frac{\Pr[A, B, C, D, \neg E] + \Pr[\neg A, B, C, D, \neg E]}{\Pr[A, B, C, D, \neg E] + \Pr[\neg A, B, C, D, \neg E] + \Pr[A, B, \neg C, D, \neg E] + \Pr[\neg A, B, \neg C, D, \neg E]}. \end{aligned}$$

Conditional probability: A cost saving

Continue with the example on the last slide.

$$\Pr[C \mid B, D, \neg E] = \frac{\Pr[B, C, D, \neg E]}{\Pr[B, D, \neg E]} = \frac{\Pr[A, B, C, D, \neg E] + \Pr[\neg A, B, C, D, \neg E]}{\Pr[B, D, \neg E]}.$$

There is no need to explicitly compute the denominator. Instead we may compute

$$\Pr[\neg C \mid B, D, \neg E] = \frac{\Pr[B, \neg C, D, \neg E]}{\Pr[B, D, \neg E]} = \frac{\Pr[A, B, \neg C, D, \neg E] + \Pr[\neg A, B, \neg C, D, \neg E]}{\Pr[B, D, \neg E]}.$$

Since

$$\Pr[C \mid B, D, \neg E] + \Pr[\neg C \mid B, D, \neg E] = 1,$$

the individual probabilities can be obtained by eliminating the common denominator.

The chain rule

$\mathcal{W} = \{W_1, W_2, \ldots, W_k\}$. Then:

$$\begin{aligned}
\Pr[\mathcal{V}, \mathcal{W}] &= \Pr[\mathcal{V}, W_1, W_2, \ldots, W_k] \\
&= \Pr[\mathcal{V} \mid W_1, W_2, \ldots, W_k] \Pr[W_1, W_2, \ldots, W_k] \\
&= \Pr[\mathcal{V} \mid W_1, W_2, \ldots, W_k] \Pr[W_1 \mid W_2, \ldots, W_k] \Pr[W_2, \ldots, W_k] \\
&= \Pr[\mathcal{V} \mid W_1, W_2, \ldots, W_k] \Pr[W_1 \mid W_2, \ldots, W_k] \Pr[W_2 \mid W_3, \ldots, W_k] \Pr[W_3, \ldots, W_k] \\
&\cdots \Pr[\mathcal{V} \mid W_1, W_2, \ldots, W_k] \Pr[W_1 \mid W_2, \ldots, W_k] \Pr[W_2 \mid W_3, \ldots, W_k] \cdots \Pr[W_{k-1} \mid W_k] \Pr[W_k].
\end{aligned}$$

Example: For five variables A, B, C, D, E, we have

$$\Pr[\neg A, C, \neg D, E] = \Pr[\neg A \mid C, \neg D, E] \Pr[C \mid \neg D, E] \Pr[D \mid E] \Pr[E].$$

Bayes rule

Let $\mathcal{V}$ and $\mathcal{W}$ two sets of rv's. By the chain rule, we have

$$\Pr[\mathcal{V}, \mathcal{W}] = \Pr[\mathcal{V} \mid \mathcal{W}]\, P[\mathcal{W}] = \Pr[\mathcal{W} \mid \mathcal{V}]\, P[\mathcal{V}].$$

Therefore

$$\boxed{\Pr[\mathcal{V} \mid \mathcal{W}] = \frac{\Pr[\mathcal{W} \mid \mathcal{V}]\, P[\mathcal{V}]}{P[\mathcal{W}]}.}$$

Suppose that $\mathcal{V}$ are cause variables, and $\mathcal{W}$ are effect variables.

The probability $\Pr[\mathcal{W} \mid \mathcal{V}]$ is the probability of the effects given the causes [**prior probability**].

The probability $\Pr[\mathcal{V} \mid \mathcal{W}]$ is the probability of the causes given the effects [**posterior probability**].

The Bayes rule connects these two probabilities.

Example domain: Medical diagnosis

A doctor learns in a medical college (and later during experience buildup)

- What are the diseases? How probable these are in a population?
- What are the probable symptoms of each disease?

These are all prior information.

You do not go to a doctor for knowing these information.

You tell your symptoms to the doctor.

The doctor diagnoses what is/are the most probable causes of your medical condition.

Symptoms (like fever or abdominal pain) may be shared by many illnesses.

The combination you have gives the doctor a best guess for what happened to you. He may prescribe medicine or further clinical tests to confirm his diagnosis.

A doctor converts his/her prior knowledge to case-specific posterior knowledge.

Another example domain: Paleontology

A dinosaur fossil or a fossilized dinosaur footprint is discovered.

The team of paleontologists plan to determine several facts from the fossil.

- The age of the dinosaur.
- The type of the dinosaur.
- What the habits of the dinosaur were.
- How the animal lived.
- How the animal died.
- . . .

All the team knows is a set of prior probabilities associated with the animals.

The questions are all posterior in nature.

The team needs to make good guesses from the evidence they have (they may generate more evidence by modern tools like CAT scans).

Yet another example domain: Forensic investigation

A crime (like murder of bank heist) is committed.

Detectives rush to the crime scene to gather whatever tiny evidences are left by the perpetrator(s).

They have a vast prior knowledge of criminals, crime scenes, type of evidences, and so on.

Their job is to convert the collected evidences in view of their knowledge to potential suspect(s).

Once again, this is a case of prior-to-posterior conversion.

Probabilistic inference

The last three examples illustrate the process.

You have a set $\mathcal{V}$ of random variables.

Specific values for a subset $\mathcal{E}$ of these variables are available as evidence.

For a random variable V, we need to compute the probability of V (and $\neg V$) given the known values of the evidence variables $\mathcal{E}$.

$$\Pr[V \mid \mathcal{E}] = \frac{\Pr[V, \mathcal{E}]}{\Pr[\mathcal{E}]}$$

The problem is well-defined and well-understood.

The computations may be infeasible if there are many variables.

The tiny cost saving made by avoiding an explicit computation of the denominator cannot address this infeasibility.

Probabilistic inference: A toy example

Consider the following joint distribution of five rv's A, B, C, D, E. The table lists the $2^5 = 32$ probabilities against the values a, b, c, d, e of the rv's.

a	b	c	d	e	$\Pr[a, b, c, d, e]$
T	T	T	T	T	0.005
T	T	T	T	F	0.095
T	T	T	F	T	0.021
T	T	T	F	F	0.001
T	T	F	T	T	0.010
T	T	F	T	F	0.025
T	T	F	F	T	0.008
T	T	F	F	F	0.007
T	F	T	T	T	0.015
T	F	T	T	F	0.123
T	F	T	F	T	0.012
T	F	T	F	F	0.022
T	F	F	T	T	0.067
T	F	F	T	F	0.057
T	F	F	F	T	0.016
T	F	F	F	F	0.032

a	b	c	d	e	$\Pr[a, b, c, d, e]$
F	T	T	T	T	0.024
F	T	T	T	F	0.012
F	T	T	F	T	0.018
F	T	T	F	F	0.051
F	T	F	T	T	0.019
F	T	F	T	F	0.003
F	T	F	F	T	0.020
F	T	F	F	F	0.017
F	F	T	T	T	0.079
F	F	T	T	F	0.014
F	F	T	F	T	0.021
F	F	T	F	F	0.102
F	F	F	T	T	0.007
F	F	F	T	F	0.009
F	F	F	F	T	0.047
F	F	F	F	F	0.041

Probabilistic inference: A toy example

Suppose that B and $\neg E$ are supplied as evidence. We want to find the probability of C.

$$\Pr[C \mid B, \neg E] = \frac{\Pr[B, C, \neg E]}{\Pr[B, \neg E]}.$$

$$\Pr[\neg C \mid B, \neg E] = \frac{\Pr[B, \neg C, \neg E]}{\Pr[B, \neg E]}.$$

$$\begin{aligned}
\Pr[B, C, \neg E] &= \Pr[A, B, C, D, \neg E] + \Pr[A, B, C, \neg D, \neg E] + \Pr[\neg A, B, C, D, \neg E] + \Pr[\neg A, B, C, \neg D, \neg E] \\
&= 0.095 + 0.001 + 0.012 + 0.051 \\
&= 0.159.
\end{aligned}$$

$$\begin{aligned}
\Pr[B, \neg C, \neg E] &= \Pr[A, B, \neg C, D, \neg E] + \Pr[A, B, \neg C, \neg D, \neg E] + \Pr[\neg A, B, \neg C, D, \neg E] + \Pr[\neg A, B, \neg C, \neg D, \neg E] \\
&= 0.025 + 0.007 + 0.003 + 0.017 \\
&= 0.052.
\end{aligned}$$

$$\begin{aligned}
\Pr[C \mid B, \neg E] &= \frac{0.159}{0.159 + 0.052} \\
&\approx 0.754.
\end{aligned}$$

Conditional independence

For n variables, the size of the joint distribution is 2^n.

This may be infeasibly large for large values of n.

If all the variables were independent, only n tables with two entries each would suffice.

The calculation would still involve an exponential number of steps.

Moreover, the assumption that all variables are independent is nowhere near a model of reality.

In view of this, belief networks are introduced.

These networks depend on the notion of conditional independence.

Conditional independence is natural in many real-life examples.

Even if the assumption of conditional independence is not fully accurate, the approximations it produces are often a reasonable model of reality.

Conditional independence

Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be sets of random variables.

$\mathcal{U}$ is said to be **conditionally independent** of $\mathcal{V}$, given (the values of the variables in) $\mathcal{W}$, if:

$$\Pr[\mathcal{U} \mid \mathcal{V}, \mathcal{W}] = \Pr[\mathcal{U} \mid \mathcal{W}]$$

We denote this as $\mathit{IND}(\mathcal{U}, \mathcal{V} \mid \mathcal{W})$.

This means that the variables $\mathcal{V}$ do not supply more information on the variables $\mathcal{U}$ than is already provided by $\mathcal{W}$.

Therefore, in the probability calculations of the form $\Pr[\mathcal{U} \mid \mathcal{V}, \mathcal{W}]$, it suffices to compute $\Pr[\mathcal{U} \mid \mathcal{W}]$, completely ignoring the variables $\mathcal{V}$.

Conditional independence

Let $\mathcal{U}$ consist of a single variable U, and $\mathcal{V}$ a single variable V. Then, the definition of conditional probability gives $\Pr[U \mid V, \mathcal{W}] = \Pr[U, V \mid \mathcal{W}] / \Pr[V \mid \mathcal{W}]$. By conditional independence, $\Pr[U \mid V, \mathcal{W}] = \Pr[U \mid \mathcal{W}]$. Therefore, we have:

$$\Pr[U, V \mid \mathcal{W}] = \Pr[U \mid \mathcal{W}] \ \Pr[V \mid \mathcal{W}]$$

For any number of variables $V_1, V_2, \ldots, V_k$ mutually conditionally independent given $\mathcal{W}$, we have:

$$\Pr[V_1, V_2, \ldots, V_k \mid \mathcal{W}] = \Pr[V_1 \mid \mathcal{W}] \ \Pr[V_2 \mid \mathcal{W}] \ \cdots \ \Pr[V_k \mid \mathcal{W}]$$

If $\mathcal{W}$ is the empty set, then $V_1, V_2, \ldots, V_k$ are unconditionally independent of one another, and we have:

$$\Pr[V_1, V_2, \ldots, V_k] = \Pr[V_1] \ \Pr[V_2] \ \cdots \ \Pr[V_k]$$

Bayes networks

Also called **Bayesian networks**, **belief networks**, and **causal networks**.

A Bayes network on n random variables $V_1, V_2, \ldots, V_n$ is a DAG (directed acyclic graph) such that:

- The graph contains n nodes, one for each variable.
- The edges in the graph are direct causal links. For every directed edge (U, V) in the graph, U is called a **parent** of V. The set of all parents of V is denoted by $\mathcal{P}(V)$.
- Let V be a node, and $\mathcal{W}$ a set of nodes neither of which is a descendant of V. Then, V is conditionally independent of $\mathcal{W}$ given $\mathcal{P}(V)$, that is, $IND(V, \mathcal{W} \mid \mathcal{P}(V))$, that is, $\Pr[V \mid \mathcal{W}, \mathcal{P}(V)] = \Pr[V \mid \mathcal{P}(V)]$.

Joint distribution in a Bayes network

Theorem: Let $V_1, V_2, \ldots, V_n$ be all the variables, that is, vertices (not their values) in a Bayes network. Then:

$$\Pr[V_1, V_2, \ldots, V_n] = \prod_{i=1}^{n} \Pr[V_i \mid \mathcal{P}(V_i)]$$

Proof Since the Bayes network is a DAG, we have a topological ordering of the vertices in the network. Let $V_1, V_2, \ldots, V_n$ be the *reverse* of such an ordering. By the chain rule, we have

$$\Pr[V_1, V_2, \ldots, V_n] = \Pr[V_1 \mid V_2, V_3, \ldots, V_n] \; \Pr[V_2 \mid V_3, V_4, \ldots, V_n] \; \cdots \; \Pr[V_{n-1} \mid V_n] \; \Pr[V_n].$$

By the chosen ordering, neither of $V_{i+1}, V_{i+2}, \ldots, V_n$ is a descendant of V_i. By conditional independence in a Bayes network, we then have $\Pr[V_i \mid V_{i+1}, V_{i+2}, \ldots, V_n] = \Pr[V_i \mid \mathcal{P}(V_i)]$, so

$$\Pr[V_1, V_2, \ldots, V_n] = \Pr[V_1 \mid \mathcal{P}(V_1)] \; \Pr[V_2 \mid \mathcal{P}(V_2)] \; \cdots \; \Pr[V_{n-1} \mid \mathcal{P}(V_{n-1})] \; \Pr[V_n \mid \mathcal{P}(V_n)].$$

Finally, note that V_n has no parents, so $\Pr[V_n \mid \mathcal{P}(V_n)] = \Pr[V_n]$. ◀

Conditional Probability Tables (CPTs)

The formula for the joint distribution indicates that it suffices to store the probabilities $\Pr[V_i \mid \mathcal{P}(V_i)]$ only, for each i.

If $k > 0$ is the number of parents of V_i, then the table against V_i contains 2^k rows.

Each row gives $\Pr[V_i \mid \mathcal{P}(V_i)]$ for one truth assignment of $\mathcal{P}(V_i)$.

We can calculate $\Pr[\neg V_i \mid \mathcal{P}(V_i)] = 1 - \Pr[V_i \mid \mathcal{P}(V_i)]$, so we do not need a separate table of 2^k rows for storing these.

If V_i has no parents (that is, $k = 0$), then only the unconditional probability $\Pr[V_i]$ is stored, and we calculate $\Pr[\neg V_i] = 1 - \Pr[V_i]$.

Example

Professor Foojit lives in an apartment building with two elderly neighbors David and Emil on the same floor. Foojit is worried about protecting his latest research findings, so he has set up a burglar alarm in his apartment. If he is not home, somebody needs to react to the alarm. Both David and Emil agree to inform police independently in case they hear an alarm. However, the alarm is sometimes triggered by cockroaches too. David can hear the alarm if he is at home and not in the wash room, whereas Emil is a bit short of hearing, so she can miss the sound if her music is playing. Finally, both David and Emil may mistakenly take the ring tone of the security guard or of the janitor as the alarm sound.

Consider the following random variables:

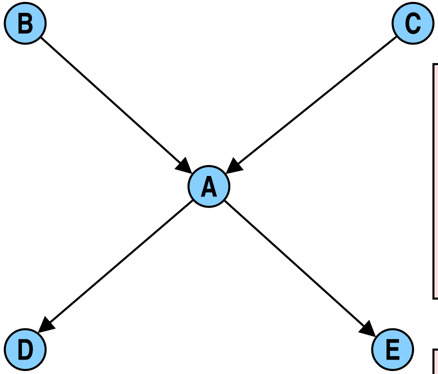
- A: The alarm rings.
- B: There is a burglar in Foojit's apartment.
- C: Cockroaches are active in Foojit's apartment.
- D: David hears and interprets the alarm correctly and calls the police.
- E: Emil hears and interprets the alarm correctly and calls the police.

The direct causal relations are: $B \to A$, $C \to A$, $A \to D$, and $A \to E$.

Example: The Bayes network with CPTs

Pr[B = T] = 0.005

Pr[C = T] = 0.020



B	C	Pr[A = T | B,C]
T	T	0.900
T	F	0.800
F	T	0.250
F	F	0.000

A	Pr[D = T | A]
T	0.750
F	0.200

A	Pr[E = T | A]
T	0.600
F	0.100

Example: Probability calculations

- Pr[There is a burglary. There is no cockroach activity. The alarm rings. Neither David nor Emil calls the police.]

$$\begin{aligned}\Pr[A, B, \neg C, \neg D, \neg E] &= \Pr[A \mid B, \neg C] \Pr[B] \Pr[\neg C] \Pr[\neg D \mid A] \Pr[\neg E \mid A] \\ &= 0.800 \times 0.005 \times (1 - 0.020) \times (1 - 0.750) \times (1 - 0.600) \\ &= 0.000392\end{aligned}$$

- Pr[There is a burglary. There is no cockroach activity. The alarm rings. Only David calls the police.]

$$\begin{aligned}\Pr[A, B, \neg C, D, \neg E] &= \Pr[A \mid B, \neg C] \Pr[B] \Pr[\neg C] \Pr[D \mid A] \Pr[\neg E \mid A] \\ &= 0.800 \times 0.005 \times (1 - 0.020) \times 0.750 \times (1 - 0.600) \\ &= 0.001176\end{aligned}$$

- Pr[There is a burglary. There is no cockroach activity. The alarm rings. The police is called.]

$$\begin{aligned}&\Pr[A, B, \neg C, D, \neg E] + \Pr[A, B, \neg C, \neg D, E] + \Pr[A, B, \neg C, D, E] \\ &= \Pr[A \mid B, \neg C] \Pr[B] \Pr[\neg C] \Big(\Pr[D \mid A] \Pr[\neg E \mid A] + \Pr[\neg D \mid A] \Pr[E \mid A] + \Pr[D \mid A] \Pr[E \mid A] \Big) \\ &= \Pr[A \mid B, \neg C] \Pr[B] \Pr[\neg C] \Big(1 - \Pr[\neg D \mid A] \Pr[\neg E \mid A] \Big) \\ &= 0.800 \times 0.005 \times (1 - 0.020) \times \Big(1 - (1 - 0.750) \times (1 - 0.600) \Big) \\ &= 0.003528\end{aligned}$$

Example: Marginal distributions

- $\Pr[A] = \Pr[A \mid B, C]\Pr[B, C] + \Pr[A \mid B, \neg C]\Pr[B, \neg C] + \Pr[A \mid \neg B, C]\Pr[\neg B, C] + \Pr[A \mid \neg B, \neg C]\Pr[\neg B, \neg C] = 0.900 \times 0.005 \times 0.020 + 0.800 \times 0.005 \times (1 - 0.020) + 0.250 \times (1 - 0.005) \times 0.020 + 0.000 \times (1 - 0.005) \times (1 - 0.020) = 0.008985.$
- $\Pr[\neg A] = 1 - \Pr[A] = 1 - 0.008985 = 0.991015.$

- $\Pr[B] = 0.005$
- $\Pr[\neg B] = 1 - 0.005 = 0.995$

- $\Pr[C] = 0.020$
- $\Pr[\neg C] = 1 - 0.020 = 0.980.$

- $\Pr[D] = \Pr[D \mid A]\Pr[A] + \Pr[D \mid \neg A]\Pr[\neg A] = 0.750 \times 0.008985 + 0.200 \times 0.991015 = 0.20494175.$
- $\Pr[\neg D] = 1 - \Pr[D] = 1 - 0.20494175 = 0.79505825.$

- $\Pr[E] = \Pr[E \mid A]\Pr[A] + \Pr[E \mid \neg A]\Pr[\neg A] = 0.600 \times 0.008985 + 0.100 \times 0.991015 = 0.1044925.$
- $\Pr[\neg E] = 1 - \Pr[E] = 1 - 0.1044925 = 0.8955075.$

Example: Marginal distributions

- $\Pr[A, B] = \Pr[A, B, C] + \Pr[A, B, \neg C] = \Pr[A \mid B, C] \Pr[B, C] + \Pr[A \mid B, \neg C] \Pr[B, \neg C] = \Pr[A \mid B, C] \Pr[B] \Pr[C] + \Pr[A \mid B, \neg C] \Pr[B] \Pr[\neg C] = 0.900 \times 0.005 \times 0.020 + 0.800 \times 0.005 \times (1 - 0.020) = 0.00401$
- $\Pr[A, \neg B] = \Pr[A \mid \neg B, C] \Pr[\neg B] \Pr[C] + \Pr[A \mid \neg B, \neg C] \Pr[\neg B] \Pr[\neg C] = 0.004975$
- $\Pr[\neg A, B] = \Pr[B] - \Pr[A, B] = 0.00099$
- $\Pr[\neg A, \neg B] = \Pr[\neg B] - \Pr[A, \neg B] = 0.990025$

- $\Pr[A, C] = \Pr[A, B, C] + \Pr[A, \neg B, C] = \Pr[A \mid B, C] \Pr[B] \Pr[C] + \Pr[A \mid \neg B, C] \Pr[\neg B] \Pr[C] = 0.005065$
- $\Pr[A, \neg C] = \Pr[A \mid B, \neg C] \Pr[B] \Pr[\neg C] + \Pr[A \mid \neg B, \neg C] \Pr[\neg B] \Pr[\neg C] = 0.00392$
- $\Pr[\neg A, C] = \Pr[C] - \Pr[A, C] = 0.014935$
- $\Pr[\neg A, \neg C] = \Pr[\neg C] - \Pr[A, \neg C] = 0.97608$

Example: Marginal distributions

- $\Pr[A, D] = \Pr[D \mid A] \Pr[A] = 0.750 \times 0.008985 = 0.00673875$
- $\Pr[A, \neg D] = \Pr[A] - \Pr[A, D] = 0.008985 - 0.00673875 = 0.00224625$
- $\Pr[\neg A, D] = \Pr[D \mid \neg A] \Pr[\neg A] = 0.200 \times 0.991015 = 0.198203$
- $\Pr[\neg A, \neg D] = \Pr[\neg A] - \Pr[\neg A, D] = 0.991015 - 0.198203 = 792812$

- $\Pr[A, E] = \Pr[E \mid A] \Pr[A] = 0.600 \times 0.008985 = 0.005391$
- $\Pr[A, \neg E] = \Pr[A] - \Pr[A, E] = 0.008985 - 0.005391 = 0.003594$
- $\Pr[\neg A, E] = \Pr[E \mid \neg A] \Pr[\neg A] = 0.100 \times 0.991015 = 0.0991015$
- $\Pr[\neg A, \neg E] = \Pr[\neg A] - \Pr[\neg A, E] = 0.991015 - 0.0991015 = 0.8919135$

Example: Marginal distributions

- $\Pr[B, D] = \Pr[A, B, D] + \Pr[\neg A, B, D] = \Pr[D \mid A, B] \Pr[A, B] + \Pr[D \mid \neg A, B] \Pr[\neg A, B] = \Pr[D \mid A] \Pr[A, B] + \Pr[D \mid \neg A] \Pr[\neg A, B] = 0.750 \times 0.00401 + 0.200 \times 0.00099 = 0.0032055$
- $\Pr[B, \neg D] = \Pr[B] - \Pr[B, D] = 0.005 - 0.0032055 = 0.0017945$
- $\Pr[\neg B, D] = \Pr[A, \neg B, D] + \Pr[\neg A, \neg B, D] = \Pr[D \mid A, \neg B] \Pr[A, \neg B] + \Pr[D \mid \neg A, \neg B] \Pr[\neg A, \neg B] = \Pr[D \mid A] \Pr[A, \neg B] + \Pr[D \mid \neg A] \Pr[\neg A, \neg B] = 0.750 \times 0.004975 + 0.200 \times 0.990025 = 0.20173625$
- $\Pr[\neg B, \neg D] = \Pr[\neg B] - \Pr[\neg B, D] = 0.995 - 0.20173625 = 0.79326375$

- $\Pr[B, E] = \Pr[A, B, E] + \Pr[\neg A, B, E] = \Pr[E \mid A, B] \Pr[A, B] + \Pr[E \mid \neg A, B] \Pr[\neg A, B] = \Pr[E \mid A] \Pr[A, B] + \Pr[E \mid \neg A] \Pr[\neg A, B] = 0.600 \times 0.00401 + 0.100 \times 0.00099 = 0.002505$
- $\Pr[B, \neg E] = \Pr[B] - \Pr[B, E] = 0.005 - 0.002505 = 0.002495$
- $\Pr[\neg B, E] = \Pr[A, \neg B, E] + \Pr[\neg A, \neg B, E] = \Pr[E \mid A, \neg B] \Pr[A, \neg B] + \Pr[E \mid \neg A, \neg B] \Pr[\neg A, \neg B] = \Pr[E \mid A] \Pr[A, \neg B] + \Pr[E \mid \neg A] \Pr[\neg A, \neg B] = 0.600 \times 0.004975 + 0.100 \times 0.990025 = 0.1019875$
- $\Pr[\neg B, \neg E] = \Pr[\neg B] - \Pr[\neg B, E] = 0.995 - 0.1019875 = 0.8930125$

Example: Marginal distributions

- $\Pr[A, D, E] = \Pr[D \mid A, E] \Pr[E \mid A] \Pr[A] = \Pr[D \mid A] \Pr[E \mid A] \Pr[A] = 0.750 \times 0.600 \times 0.008985 = 0.00404325$
- $\Pr[A, D, \neg E] = \Pr[D \mid A] \Pr[\neg E \mid A] \Pr[A] = 0.750 \times (1 - 0.600) \times 0.008985 = 0.0026955$
- $\Pr[A, \neg D, E] = \Pr[\neg D \mid A] \Pr[E \mid A] \Pr[A] = (1 - 0.750) \times 0.600 \times 0.008985 = 0.00134775$
- $\Pr[A, \neg D, \neg E] = \Pr[\neg D \mid A] \Pr[\neg E \mid A] \Pr[A] = (1 - 0.750) \times (1 - 0.600) \times 0.008985 = 0.0008985$
- $\Pr[\neg A, D, E] = \Pr[D \mid \neg A] \Pr[E \mid \neg A] \Pr[\neg A] = 0.200 \times 0.100 \times 0.991015 = 0.0198203$
- $\Pr[\neg A, D, \neg E] = \Pr[D \mid \neg A] \Pr[\neg E \mid \neg A] \Pr[\neg A] = 0.200 \times (1 - 0.100) \times 0.991015 = 0.1783827$
- $\Pr[\neg A, \neg D, E] = \Pr[\neg D \mid \neg A] \Pr[E \mid \neg A] \Pr[\neg A] = (1 - 0.200) \times 0.100 \times 0.991015 = 0.0792812$
- $\Pr[\neg A, \neg D, \neg E] = \Pr[\neg D \mid \neg A] \Pr[\neg E \mid \neg A] \Pr[\neg A] = (1 - 0.200) \times (1 - 0.100) \times 0.991015 = 0.7135308$

- $\Pr[D, E] = \Pr[A, D, E] + \Pr[\neg A, D, E] = 0.00404325 + 0.0198203 = 0.02386355$
- $\Pr[D, \neg E] = \Pr[A, D, \neg E] + \Pr[\neg A, D, \neg E] = 0.0026955 + 0.1783827 = 0.1810782$
- $\Pr[\neg D, E] = \Pr[A, \neg D, E] + \Pr[\neg A, \neg D, E] = 0.00134775 + 0.0792812 = 0.08062895$
- $\Pr[\neg D, \neg E] = \Pr[A, \neg D, \neg E] + \Pr[\neg A, \neg D, \neg E] = 0.0008985 + 0.7135308 = 0.7144293$

Note: D and E are **not** independent.

Example: Marginal distributions

- $\Pr[A, B, D, E] = \Pr[E \mid A, B, D] \Pr[D \mid A, B] \Pr[A, B] = \Pr[E \mid A] \Pr[D \mid A] \Pr[A, B] = 0.6 \times 0.75 \times 0.00401 = 0.0018045$
- $\Pr[\neg A, B, D, E] = \Pr[E \mid \neg A] \Pr[D \mid \neg A] \Pr[\neg A, B] = 0.1 \times 0.2 \times 0.00099 = 0.0000198$
- $\Pr[B, D, E] = \Pr[A, B, D, E] + \Pr[\neg A, B, D, E] = 0.0018045 + 0.0000198 = 0.0018243$

- $\Pr[A, \neg B, D, E] = \Pr[E \mid A] \Pr[D \mid A] \Pr[A, \neg B] = 0.6 \times 0.75 \times 0.004975 = 0.00223875$
- $\Pr[\neg A, \neg B, D, E] = \Pr[E \mid \neg A] \Pr[D \mid \neg A] \Pr[\neg A, \neg B] = 0.1 \times 0.2 \times 0.990025 = 0.0198005$
- $\Pr[\neg B, D, E] = \Pr[A, \neg B, D, E] + \Pr[\neg A, \neg B, D, E] = 0.00223875 + 0.0198005 = 0.02203925$

- $\Pr[D, E] = \Pr[B, D, E] + \Pr[\neg B, D, E] = 0.0018243 + 0.02203925 = 0.02386355$

Example: Conditional probability [Probabilistic inference]

- Probability that Emil calls given that there is a burglary: $\Pr[E \mid B] = \dfrac{\Pr[B, E]}{\Pr[B]} = \dfrac{0.002505}{0.005} = 0.501$
- Probability that there is a burglary given that Emil calls: $\Pr[B \mid E] = \dfrac{\Pr[B, E]}{\Pr[E]} = \dfrac{0.002505}{0.1044925} \approx 0.023973$
- Probability that both David and Emil call given that there is a burglary:
  $\Pr[D, E \mid B] = \dfrac{\Pr[B, D, E]}{\Pr[B]} = \dfrac{0.0018243}{0.005} = 0.36486$
- Probability that there is a burglary given that both David and Emil call:
  $\Pr[B \mid D, E] = \dfrac{\Pr[B, D, E]}{\Pr[D, E]} = \dfrac{0.0018243}{0.02386355} \approx 0.076447$
- Probability that there is a burglary given that the alarm rings and both David and Emil call:
  $\Pr[B \mid A, D, E] = \dfrac{\Pr[A, B, D, E]}{\Pr[A, D, E]} = \dfrac{0.0018045}{0.00404325} \approx 0.4462994$
- Probability that the alarm rings given that neither David nor Emil calls:
  $\Pr[A \mid \neg D, \neg E] = \dfrac{\Pr[A, \neg D, \neg E]}{\Pr[\neg D, \neg E]} = \dfrac{0.0008985}{0.7144293} \approx 0.00125765$

Conditional independencies in a Bayes network

From the definition of Bayes nets, it follows that:

Each variable in the net is independent of its non-descendants given its parents.

This is called **local Markov property**.

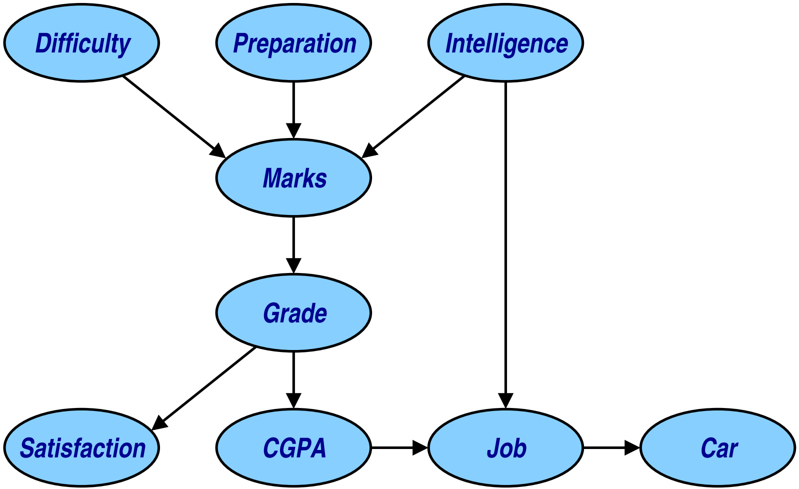
Examples in the Alarm network

- D is independent of E given A.
- D is not unconditionally independent of E.
- D is independent of B given A.
- B and C are unconditionally independent.

That is not all. A Bayes net offers more conditional independencies.

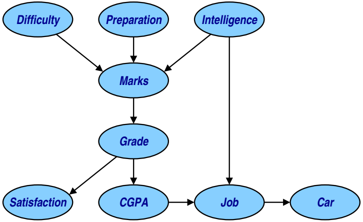
The alarm network is too small to explain all the cases.

Example: Your AI course may have an impact on what car you will own



You can Booleanize all the variables like ***Difficulty*** can be tough/easy, ***Grade*** can be high/low, and ***Car*** can be expensive/cheap.

Another conditional independency in a Bayes net



The **Markov blanket** of a variable X consists of

all the parents of X,
all the children of X, and
all the parents (except X) of these children.

Markov blanket property

Given the Markov blanket of a variable in a Bayes net, the variable is independent of all other variables in the net.

Examples

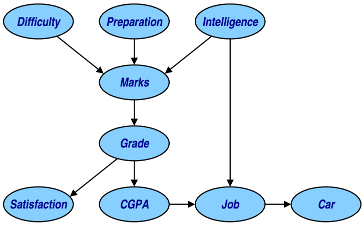
- The Markov blanket of ***Marks*** is $\{$***Difficulty***, ***Preparation***, ***Intelligence***, ***Grade***$\}$. So ***Marks*** is independent of ***CGPA*** given ***Difficulty***, ***Preparation***, ***Intelligence***, and ***Grade***.
- The Markov blanket of ***CGPA*** is $\{$***Grade***, ***Job***, ***Intelligence***$\}$. So ***CGPA*** is independent of ***Marks*** given ***Grade***, ***Job***, and ***Intelligence***.

This is slightly non-local, but the most general result tells us more.

Trails and colliders

An undirected path in a Bayes net is called a **trail**. *Undirected* means you may go in the reverse directions of one or more arrows on the edges. Repeated nodes are not allowed on a trail.

An internal node X on a trail is called a **collider** if both the arrows on the edges incident upon X point to X.



Examples

Consider the trail:

Preparation $\rightarrow$ ***Marks*** $\leftarrow$ *Intelligence* $\rightarrow$ ***Job*** $\leftarrow$ *CGPA* $\leftarrow$ *Grade* $\rightarrow$ *Satisfaction*

Marks and *Job* are the colliders on this trail. *Intelligence* (also *Grade*) is not a collider because the two edges incident on it are directed away from it. *CGPA* is not a collider too because one edge incident on it goes in that node, whereas the other moves away from the node.

Neither *Marks* nor *Job* is a collider on the trail:

CGPA $\leftarrow$ *Grade* $\leftarrow$ *Marks* $\leftarrow$ *Intelligence* $\rightarrow$ *Job* $\rightarrow$ *Car*

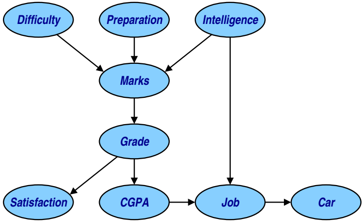
Unconditional D-connection and D-separation

D stands for directional (or dependency-driven).

Two variables are called **D-connected** if there exists a collider-free trail between them.

Two variables are are called **D-separated** if they are not D-connected (that is, if every trail connecting the two variables contains one or more colliders).

Examples



Although the trail between ***Preparation*** and ***Satisfaction*** on the last slide contains colliders (***Marks*** and ***Job***), these two variables are D-connected by the collider-free trail ***Preparation*** $\rightarrow$ ***Marks*** $\rightarrow$ ***Grade*** $\rightarrow$ ***Satisfaction***.

There are only two trails between ***Preparation*** and ***Intelligence***:

Preparation $\rightarrow$ ***Marks*** $\leftarrow$ ***Intelligence***

Preparation $\rightarrow$ ***Marks*** $\rightarrow$ ***Grade*** $\rightarrow$ ***CGPA*** $\rightarrow$ ***Job*** $\leftarrow$ ***Intelligence***

Both the trails contain colliders, so ***Preparation*** and ***Intelligence*** are D-separated.

Handling colliders using evidences

An evidence is a set E of variables whose values are observed (that is, known or given).

Let X and Y be two variables not in E.

Let V be a collider on an X-Y trail. V **blocks** the trail unconditionally.

If V or any descendant of V is in E, then that collider is cleared.

However, E creates a block on the trail if E contains a non-collider variable on the trail.

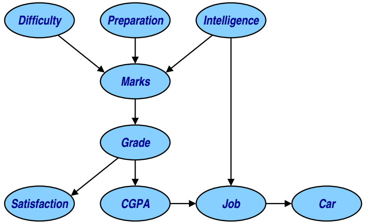
The X-Y trail is called **unblocked** given E if

(i) all colliders (if any) on the trail are cleared by E, and
(ii) E does not create a block at a non-collider node on the trail.

X and Y are called **D-connected given E** if there exists an unblocked X-Y trail given E.

X and Y are called **D-separated given E** if they are not D-connected given E, that is, if all X-Y trails are blocked by uncleared colliders and/or by non-collider variables in E.

Examples of conditional D-connection and D-separation



- The collider *Marks* on the trail

  Preparation → ***Marks*** ← *Intelligence*

  can be cleared by any one (or more) of the evidence variables *Marks*, *Grade*, *Satisfaction*, *CGPA*, *Job*, and *Car*. The collider *Job* on the trail

  Preparation → *Marks* → *Grade* → *CGPA* → ***Job*** ← *Intelligence*

  can be cleared by either *Job* or *Car* (or both) as evidence variable(s).
  It follows that *Preparation* and *Intelligence* are D-connected given *Car* as evidence.

- *Difficulty* and *CGPA* are unconditionally D-connected by the trail *Difficulty* → *Marks* → *Grade* → *CGPA*.

  Difficulty and *CGPA* are D-separated given *Marks* as evidence. First, this creates a new block on the trail:

  Difficulty → ***Marks*** → *Grade* → *CGPA*

  Second, it clears the collider *Marks* but not the collider *Job* on the trail:

  Difficulty → ***Marks*** ← *Intelligence* → ***Job*** ← *CGPA*

  There are no other trails between *Difficulty* and *CGPA*.

  Difficulty and *CGPA* are D-connected given *Marks* and *Car* as evidence. The additional evidence *Car* clears the collider *Job*, unblocking the second trail.

Global Markov property

In short: D-separation $\Rightarrow$ independent.

Theorem: *If two variables X and Y are D-separated given the evidence set E, then X and Y are independent given E.* ◀

Example: ***Difficulty*** and ***CGPA*** are independent given ***Marks*** as evidence.

Notes

- If E is empty in the theorem, then X and Y are unconditionally independent.
- This theorem can be generalized to mutually disjoint subsets X, Y, E of variables.

Exercise: Prove the local Markov property from the global Markov property.

Exercise: Prove the Markov blanket property from the global Markov property.