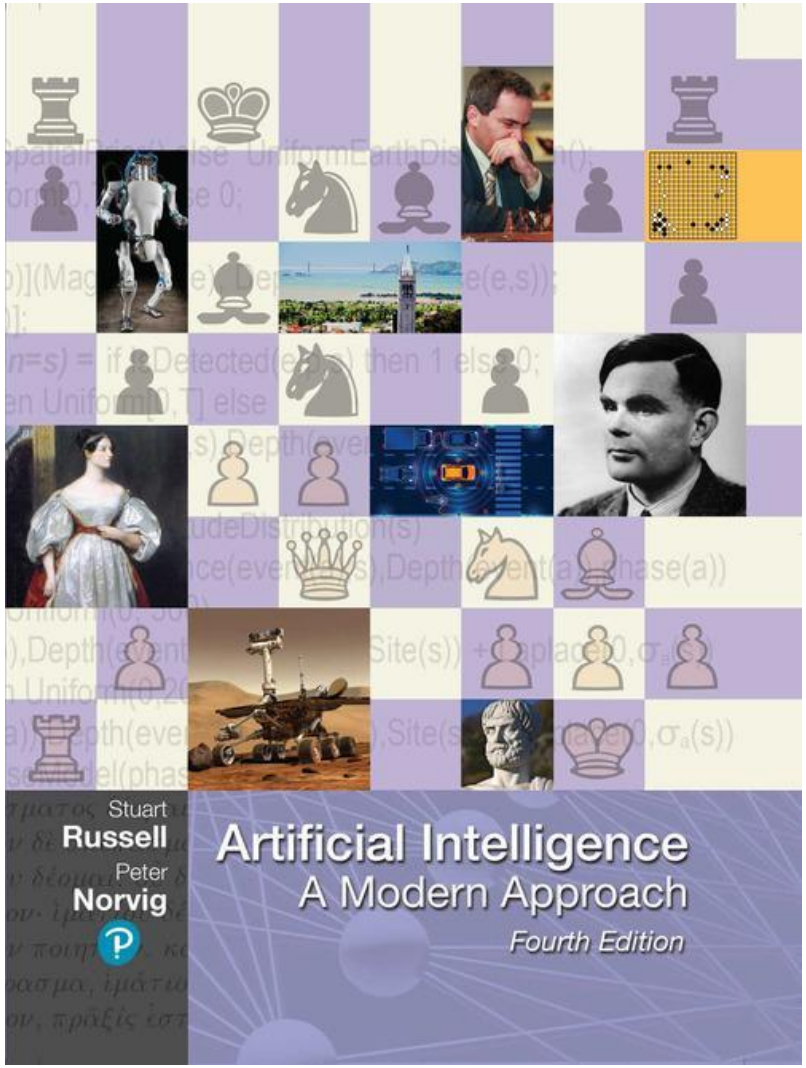
Artificial Intelligence: A Modern Approach

Fourth Edition



Chapter 12

Quantifying Uncertainty

Outline

- Acting Under Uncertainty
- Basic Probability Notation
- Inference Using Full Joint Distributions
- Independence
- Bayes' Rule and Its Use
- Naive Bayes Models
- The Wumpus World Revisited

Acting Under Uncertainty

- Real world problems contain **uncertainties due to:**
  - partial observability,
  - nondeterminism, or
  - adversaries.

- Example of dental diagnosis using propositional logic

$$Toothache \Rightarrow Cavity.$$

- However inaccurate, not all patients with toothaches have cavities

$$Toothache \Rightarrow Cavity \vee GumProblem \vee Abscess...$$

- In order to make the rule true, we have to add an almost unlimited list of possible problems.

- The only way to fix the rule is to make it logically exhaustive

Acting Under Uncertainty

- An agent strives to choose the right thing to do—the rational decision—depends on both the relative importance of various goals and the likelihood that, and degree to which, they will be achieved.

- Large domains such as medical diagnosis fail to three main reasons:
  - **Laziness:** It is too much work to list the complete set of antecedents or consequents needed to ensure an exceptionless rule
  - **Theoretical ignorance:** Medical science has no complete theory for the domain
  - **Practical ignorance:** Even if we know all the rules, we might be uncertain about a particular patient because not all the necessary tests have been or can be run.

- An agent only has a degree of belief in the relevant sentences.

Acting Under Uncertainty

- **Probability theory**
  - tool to deal with degrees of belief of relevant sentences.
  - summarizes the uncertainty that comes from our laziness and ignorance

- **Uncertainty and rational decisions**
  - An requires **preference** among **different possible outcomes** of various plans
  - **Utility Theory**: the quality of the outcome being useful
    - Every state has a degree of usefulness/utility
    - Higher utility is preferred
  - **Decision Theory:** Preferences (Utility Theory) combined with probabilities
    - *Decision theory* = *probability theory* + *utility theory*.
    - agent is **rational** if and only if it chooses the action that **yields the highest expected utility**, averaged over all the **possible outcomes** of the action.
    - principle of maximum expected utility (MEU).

Acting Under Uncertainty

- Function of a decision-theoretic agent that selects rational actions.

function DT-AGENT(*percept*) **returns** an *action*
persistent: *belief state*, probabilistic beliefs about the current state of the world
    action, the agent's action
update *belief state* based on *action* and *percept*
calculate outcome probabilities for actions,
    given action descriptions and current *belief state*
select *action* with highest expected utility
    given probabilities of outcomes and utility information
return *action*

Basic Probability Notation

- For our agent to represent and use probabilistic information, we need a formal language.

- **Sample space**: the set of all possible worlds
  - The possible worlds are mutually exclusive and exhaustive

- A fully specified probability model associates a numerical probability $P(\omega)$ with each possible world.
- The basic axioms of probability theory say that every possible world has a probability between 0 and 1 and that the total probability of the set of possible worlds is 1:

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega \text{ and } \omega \in \Omega$$

- **Unconditional or prior probability**: degrees of belief in propositions in the absence of any other information

Basic Probability Notation

- **Conditional or posterior probability:** given evidence that has happened, degree of belief of new event
  - Make use of unconditional probabilities

- Probability of a given b:

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

- Can also written as:

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

- Example of rolling fair dice, rolling doubles when the first dice is 5

$$P(doubles|Die_1 = 5) = \frac{P(doubles \wedge Die_1 = 5)}{P(Die_1 = 5)} \ .$$

Basic Probability Notation

- **Factored representation:** possible world is represented by a set of variable/value pairs.
  - Variables in probability theory are called random variables, and their names begin with an uppercase letter. ($Total$ and Die_1)

- Sometimes we will want to talk about the probabilities of all the possible values of a random variable. We could write:

$$P(Weather = sun) = 0.6$$
$$P(Weather = rain) = 0.1$$
$$P(Weather = cloud) = 0.29$$
$$P(Weather = snow) = 0.01\ ,$$

- Abbreviation of this will be:

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

- **P** statement defines a **probability distribution** for the random variable *Weather*

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:

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

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:

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

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

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:

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

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

Inference Using Full Joint Distributions

Start with the joint distribution:

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

Can also compute conditional probabilities:

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

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

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

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

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:

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

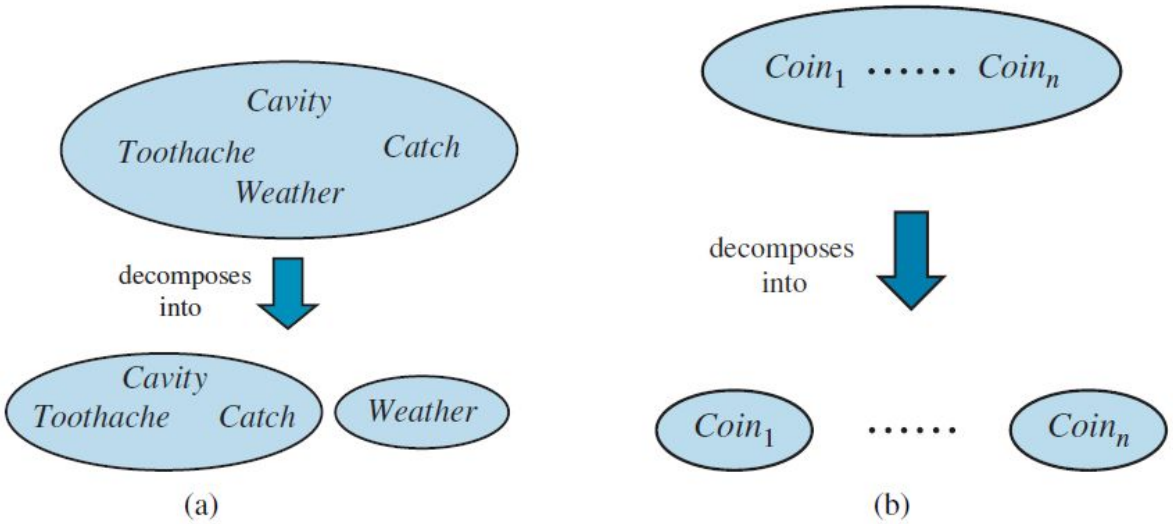
The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables

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.



- $P(a \mid b) = P(a)$ 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 and conditional independence

$$\mathrm{P}(Cavity|toothache \wedge catch)$$
$$= a\,\mathrm{P}(toothache \wedge catch|Cavity)\mathrm{P}(Cavity)$$
$$= a\,\mathrm{P}(toothache|Cavity)\mathrm{P}(catch|Cavity)\mathrm{P}(Cavity)$$

This is an example of a naive Bayes model:

$$\mathrm{P}(Cause, Effect_1, \ldots, Effect_n) = \mathrm{P}(Cause)\prod_i \mathrm{P}(Effect_i|Cause)$$

Total number of parameters is linear in n

Naïve Bayes Models

- The full joint distribution can be written as

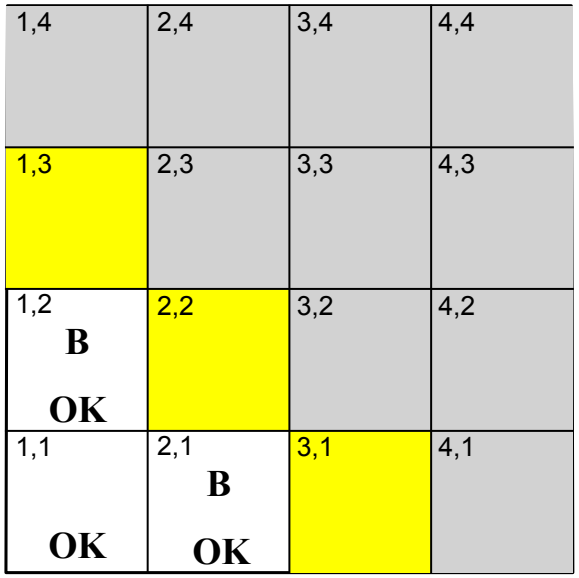
$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i \mid Cause)$$

- Such a probability distribution is called a naive Bayes model—"naive" because it is often used (as a simplifying assumption) in cases where the "effect" variables are not strictly independent given the cause variable.

- Call the observed effects **E**=**e**, while the remaining effect variables **Y** are unobserved

$$\begin{aligned}
\mathbf{P}(Cause \mid \mathbf{e}) &= \alpha \sum_{\mathbf{y}} \mathbf{P}(Cause)\mathbf{P}(\mathbf{y} \mid Cause)\left(\prod_j \mathbf{P}(e_j \mid Cause)\right) \\
&= \alpha \mathbf{P}(Cause)\left(\prod_j \mathbf{P}(e_j \mid Cause)\right)\sum_{\mathbf{y}} \mathbf{P}(\mathbf{y} \mid Cause) \\
&= \alpha \mathbf{P}(Cause)\prod_j \mathbf{P}(e_j \mid Cause)
\end{aligned}$$

The Wumpus World Revisited



$P_{ij} = true$ iff $[i, j]$ contains a pit

$B_{ij} = true$ iff $[i, j]$ is breezy

Include only $B_{1,1}$, $B_{1,2}$, $B_{2,1}$ in the probability model

Specifying the probability model

The full joint distribution is $\mathrm{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Apply product rule: $\mathrm{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})\mathrm{P}(P_{1,1}, \ldots, P_{4,4})$

(Do it this way to get $P(Effect|Cause)$.)

First term: 1 if pits are adjacent to breezes, 0 otherwise

Second term: pits are placed randomly, probability 0.2 per square:

$$\mathrm{P}(P_{1,1}, \ldots, P_{4,4}) = \prod_{i,j=1,1}^{4,4} \mathrm{P}(P_{i,j}) = 0.2^n \times 0.8^{16-n}$$

for n pits.

Observations and query

We know the following facts:

$$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$$

$$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$$

Query is $\mathrm{P}(P_{1,3} | known, b)$

Define $Unknown$ = P_{ij}s other than $P_{1,3}$ and $Known$

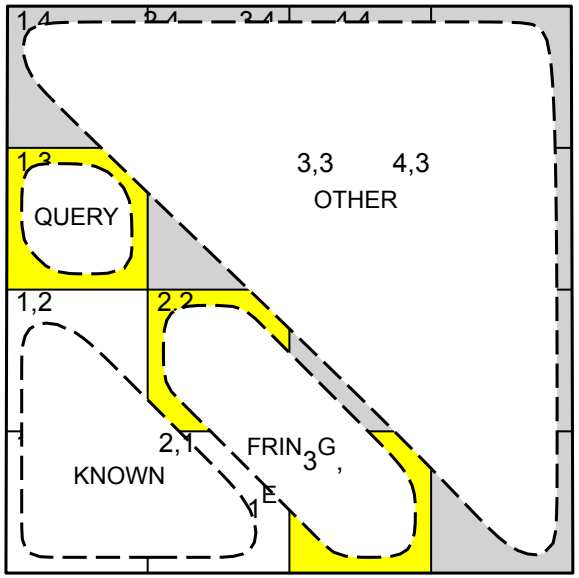
For inference by enumeration, we have

$$\mathrm{P}(P_{1,3} | known, b) = a\Sigma_{unknown}\mathrm{P}(P_{1,3}, unknown, known, b)$$

Grows exponentially with number of squares!

Using conditional independence

Basic insight: observations are conditionally independent of other hidden squares given neighbouring hidden squares



Define $Unknown = Fringe \cup Other$

$\mathrm{P}(b|P_{1,3}, Known, Unknown) = \mathrm{P}(b|P_{1,3}, Known, Fringe)$

Manipulate query into a form where we can use this!

Using conditional independence contd.

$$\mathrm{P}(P_{1,3}|known, b) = a \sum_{unknown} \mathrm{P}(P_{1,3}, unknown, known, b)$$

$$= a \sum_{unknown} \mathrm{P}(b|P_{1,3}, known, unknown)\mathrm{P}(P_{1,3}, known, unknown)$$

$$= a \sum_{f\ ringe\ other} \mathrm{P}(b|known, P_{1,3}, fringe, other)\mathrm{P}(P_{1,3}, known, fringe, other)$$

$$\equiv a \sum_{fringe\ other} \mathrm{P}(b|known, P_{1,3}, fringe)\mathrm{P}(P_{1,3}, known, fringe, other)$$

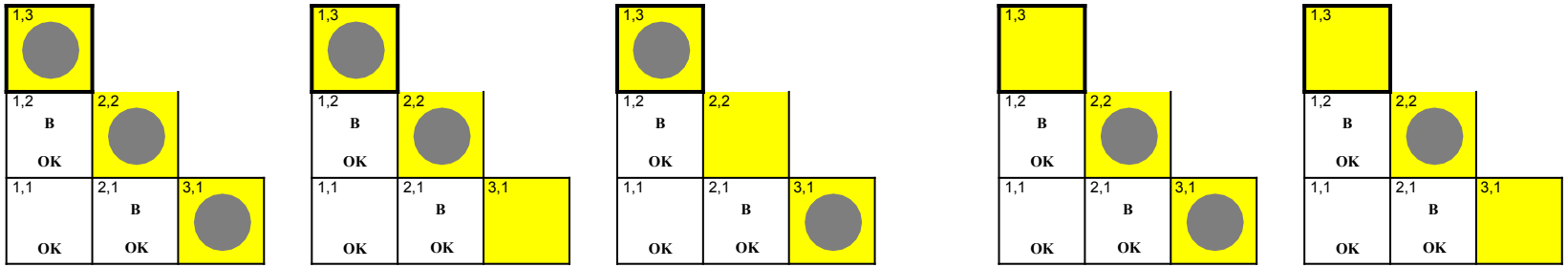
$$= a \sum_{f\ ringe\ other} \mathrm{P}(b|known, P_{1,3}, fringe)\mathrm{P}(P_{1,3})P(known)P(fringe)P(other)$$

$$= a \sum_{f\ ringe} \sum_{other} \mathrm{P}(b|known, P_{1,3}, fringe)\mathrm{P}(P_{1,3})P(known)P(fringe)P(other)$$

$$= a\, P(known)\mathrm{P}(P_{1,3}) \sum_{f\ ringe} \mathrm{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other)$$

$$= a\, \mathrm{P}(P_{1,3}) \sum_{f\ ringe} \mathrm{P}(b|known, P_{1,3}, fringe)P(fringe)$$

Using conditional independence contd.



$$\mathbf{P}(P_{1,3}|known, b) = \alpha\,(0.2(0.04 + 0.16 + 0.16), 0.8(0.04 + 0.16))$$

$$\approx (0.31, 0.69)$$

$$\mathbf{P}(P_{2,2}|known, b) \approx (0.86, 0.14)$$

Summary

- **Probabilities** express the agent's inability to reach a definite decision regarding the truth of a sentence.
- **Decision theory** combines the agent's beliefs and desires, defining the best action as the one that maximizes expected utility.
- Basic probability statements include **prior or unconditional probabilities** and **posterior or conditional probabilities** over simple and complex propositions.
- The axioms of probability constrain the probabilities of logically related propositions.
- The **full joint probability distribution** specifies the probability of each complete assignment of values to random variables
- **Absolute independence** between subsets of random variables allows the full joint distribution to be factored into smaller joint distributions, greatly reducing its complexity.
- **Bayes' rule** allows unknown probabilities to be computed from known conditional probabilities, usually in the causal direction.
- **Conditional independence** brought about by direct causal relationships in the domain allows the full joint distribution to be factored into smaller, conditional distributions.