Basics of Probability

Logic-based Systems: Assumptions

- Assertions are **completely accurate**
- Observations are **complete** and **error-free**
- Conclusions are **equally viable**
- Inference rules are **truth-preserving**

Each of these may be unrealistic when modelling complex systems

Logic-Based Systems: Assumptions I

- Assertions are **completely accurate**
  - Initial intuition: inaccuracy can be eliminated by increasing specificity

if car won't start **then** battery flat

$\downarrow$

if car won't start **and** lights OK **and** starter motor OK **and** ...

  - Untenable for really complex systems (**Qualification** problem)
- Observations are **complete** and **error-free**

Logic-Based Systems: Assumptions II

  - Random and systematic errors are common
  - Some tests are impossible, too costly, or too dangerous
- Conclusions are **equally viable**
  - Alternative ligands may be indicated as binding well to a target protein
  - But, both may not be equally viable as new drugs (toxicity, ease of manufacture, . . .)

Logic-Based Systems: Assumptions III

- Inference rules are **truth-preserving**
  - No role for uncertainty
  - Even if inference rules are sound, any errors (however small) in the assertions or observations can compound with chaining

Solution: Explicit Methods for Handling Uncertainty

- **Joint probability distributions**
- 'Certainty factors' attached to rules
- Dempster-Shafer Theory
- Non-monotonic reasoning
- Possibility Theory
- **Bayesian Networks**

Simple logic-based systems: Terminology

- **Variables** (features): usually boolean, discrete or continuous
- **Atomic Proposition**: assignment of a value to a variable

$$Cavity = true \qquad Weather = sunny \qquad X = 5.1$$

- **Compound Proposition**: atomic propositions combined with standard logical connectives $\neg, \wedge, \vee$
- **State** (interpretation): assignment of values to all variables
  - Mutually exclusive (at most one is actually the case)
  - Collectively exhaustive (at least one must be the case)

To Add Probability

- Variables $\rightarrow$ **Random Variables** (intuitively, a variable that takes one of several values, depending on some observed outcome)
- Probabilities are applied to propositions

$$\mathrm{P}(\mathit{Weather} = \mathit{sunny}) = 0.70$$

- State: assignment of values to all the random variables (sometimes called an **atomic event**)

Probabilities I

– Toss a coin

$$P(H) = \frac{1}{2} \qquad P(T) =$$

$$P(H) = \frac{1}{4} \qquad P(T) =$$

– Toss a coin several times

$$P(H) = \frac{1}{2} \qquad P(HHH) =$$

$$P(H) = \frac{1}{2} \qquad P(identical\ result\ 4\ tosses) =$$

$$P(H) = \frac{1}{2} \qquad P(at\ least\ 3\ H\ on\ 4) =$$

Axioms of Probability

- $0 \leq \mathrm{P}(a) \leq 1$
- $\mathrm{P}(\mathit{true}) = 1$
- The sum of probabilities $= 1$. So, $P(A) + P(\neg A) = 1$

Prior Probability

Unconditional and **prior probability** of a proposition is the degree of belief accorded to it
in the absence of any other information

$\mathrm{P}(\mathit{Cavity} = \mathit{true})$ or simply $\mathrm{P}(\mathit{cavity})$
$\mathrm{P}((\mathit{Cavity} = \mathit{false}) \wedge (\mathit{Weather} = \mathit{sunny}))$ or simply
$\mathrm{P}(\neg\mathit{cavity}, (\mathit{Weather} = \mathit{sunny}))$

Probability Distribution

- For a discrete random variable a probability distribution tells us how the probability is distributed over the set of events for which the random variable takes a particular value:

$$\mathrm{P}(\mathit{Weather} = \mathit{sunny}) = 0.70$$
$$\mathrm{P}(\mathit{Weather} = \mathit{rain}) = 0.20$$
$$\mathrm{P}(\mathit{Weather} = \mathit{cloudy}) = 0.08$$
$$\mathrm{P}(\mathit{Weather} = \mathit{snow}) = 0.02$$

- For the continuous case, a probability density function (p.d.f.) often can be used. We will not be looking at this case.

Joint Probability Distribution

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

where

toothache: I have a toothache
cavity: I have a cavity
catch: the dentist's probe catches in my tooth

Conditional (Posterior) Probability

$\mathrm{P}(a|b)$: the probability of a given that all we know is b

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

Example

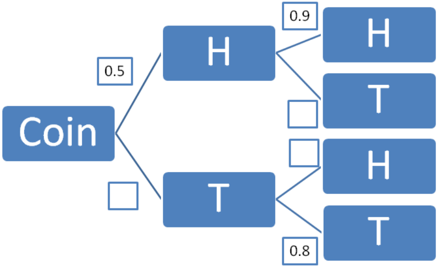
$$
\begin{aligned}
\mathrm{P}(\mathit{cavity}|\mathit{toothache}) &= \frac{\mathrm{P}(\mathit{cavity} \wedge \mathit{toothache})}{\mathrm{P}(\mathit{toothache})} \\
&= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} \\
&= 0.60
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{P}(\neg \mathit{cavity}|\mathit{toothache}) &= \frac{\mathrm{P}(\neg \mathit{cavity} \wedge \mathit{toothache})}{\mathrm{P}(\mathit{toothache})} \\
&= 0.40
\end{aligned}
$$

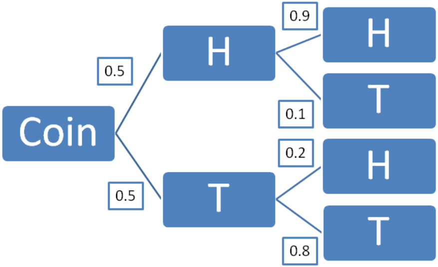
Probability Trees

In general, many (all?) probability questions are solvable by drawing out probability trees

Conditional Probabilities: Coins

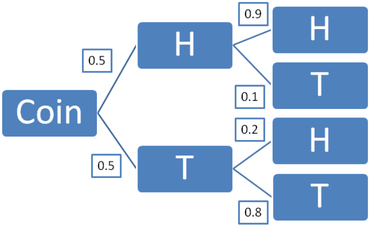


Conditional Probabilities: Coins



$$P(X_2 = H) = 0.9 \times 0.5 + 0.5 \times 0.2 = 0.55$$

Conditional Probabilities: Coins I



$$
\begin{aligned}
P(X_2 = H) &= P(X_2 = H|X_1 = H)P(X_1 = H) \\
&+ P(X_2 = H|X_1 = T)P(X_1 = T)
\end{aligned}
$$

Conditional Probabilities: Coins II

In general:

$$P(Y) = \sum_i P(Y|X = X_i)P(X = X_i)$$

Conditional Probabilities: Weather

$$
\begin{aligned}
P(D_1 = sunny) &= \text{Prob. Day1 is sunny} = 0.9 \\
P(D_2 = sunny | D_1 = sunny) &= 0.8. \text{ Then:} \\
P(D_2 = rainy | D_1 = sunny) &= \ldots
\end{aligned}
$$

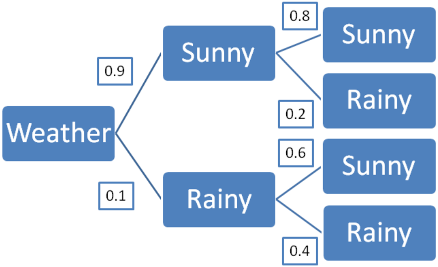
Dependent Events

$$P(D_1 = sunny) \ = \ \text{Prob. Day1 is sunny} \ = \ 0.9$$
$$P(D_2 = sunny|D_1 = sunny) \ = \ 0.8. \text{ Then:}$$
$$P(D_2 = rainy|D_1 = sunny) \ = \ 0.2$$

Conditional Probabilities: Weather

$$
\begin{aligned}
&P(D_1 = sunny) \;=\; \text{Prob. Day1 is sunny} \;=\; 0.9\\
&P(D_2 = sunny | D_1 = sunny) \;=\; 0.8. \text{ Then:}\\
&P(D_2 = rainy | D_1 = sunny) \;=\; 0.2\\
&\text{Now, if:}\\
&P(D_2 = sunny | D_1 = rainy) \;=\; 0.6. \text{ Then:}\\
&P(D_2 = rainy | D_1 = rainy) \;=\; 0.4\\
&\text{Now, what is } P(D_2 = sunny)?
\end{aligned}
$$

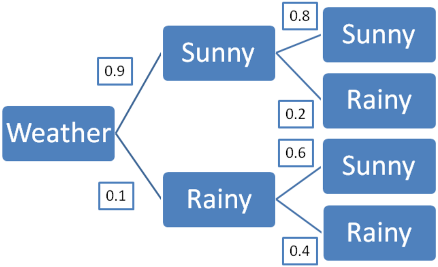
Conditional Probabilities: Weather I



$P(D_2 = sunny) \; = \; 0.9 \times 0.8 \; + \; 0.1 \times 0.6 \; = \; 0.78$

Now, what is $P(D_3 = sunny)$?

Conditional Probabilities: Weather I

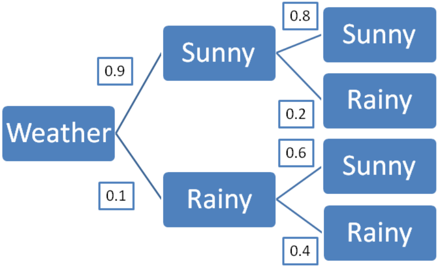


$$P(D_2 = sunny) \;=\; 0.9 \times 0.8 \;+\; 0.1 \times 0.6 \;=\; 0.78$$

Conditional Probabilities: Weather II

$$
\begin{aligned}
P(D_3 = sunny) &= P(D_3 = sunny | D_2 = sunny) P(D_2 = sunny) \\
&+ P(D_3 = sunny | D_2 = rainy) P(D_2 = rainy \\
&= 0.8 \times 0.78 + 0.6 \times 0.22 \\
&= 0.756
\end{aligned}
$$

Conditional Probabilities: Weather I

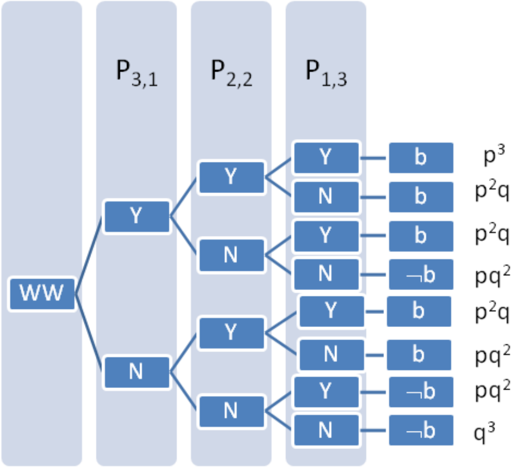


$$P(D_2 = sunny) \; = \; 0.9 \times 0.8 \; + \; 0.1 \times 0.6 \; = \; 0.78$$

Conditional Probabilities: Weather II

$$
\begin{aligned}
P(D_3 = sunny) &= P(D_3 = sunny | D_2 = sunny) P(D_2 = sunny) \\
&+ P(D_3 = sunny | D_2 = rainy) P(D_2 = rainy \\
&= 0.8 \times 0.78 + 0.6 \times 0.22 \\
&= 0.756
\end{aligned}
$$

Conditional Probabilities: Wumpus I



$$P(P_{1,3} = y|b) = \frac{(P_{1,3} = y, b)}{P(b)} = \frac{p^3 + 2p^2q}{p^3 + 3p^2q + pq^2}$$

Conditional Probabilities: Wumpus II

That is:

$$P(P_{1,3} = y|b) = \frac{p^2 + 2pq}{p^2 + 3pq + q^2}$$

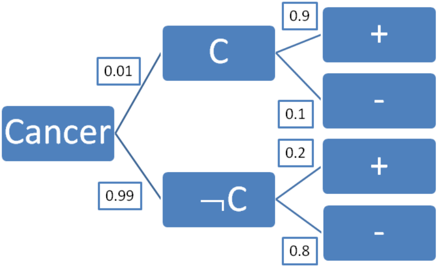
If $p = 0.2$, then $q = 0.8$ and:

$$P(P_{1,3} = y|b) = \frac{0.04 + 2(0.2)(0.8)}{(0.04) + 3(0.2)(0.8) + (0.8)(0.8)} = \frac{0.36}{0.04 + 0.48 + 0.64}$$

Conditional Probabilities: Cancer

$P(C) = 0.01$. Then $P(\neg C) = 0.99$
Suppose there is a test for this, and $P(+|C) = 0.9$. Again, $P(-|C) = 0.1$.
Also, let $P(+|\neg C) = 0.2$. Then $P(-|\neg C) = 0.8$
Now, what can we say about $P(C|+)$

Conditional Probabilities: Cancer I



By definition of conditional probability:
$P(C, +) = P(C|+) \times P(+)$.

Conditional Probabilities: Cancer II

So, we want:
$P(C|+) = \frac{P(C,+)}{P(+)}$
Now $P(C,+) = 0.01 \times 0.9$, and $P(+) = (0.01 \times 0.9 + 0.99 \times 0.2)$.
So $P(C|+) = 0.009/0.207 = 0.043$.
We have just done the calculations for a formula known as Bayes' Rule
Since $P(C,+) = P(+,C)$, we have:
$P(C,+) = P(+,C) = P(+|C) \times P(C)$. So, we have:
$P(C|+) \times P(+) = P(+|C) \times P(C)$. Or:
$P(C|+) = \frac{P(+|C) \times P(C)}{P(+)}$
This is Bayes' formula or Bayes' Rule.

Product Rule

Equivalent to the previous equation is the following, known as the **product rule**

$$
\begin{aligned}
\mathrm{P}(a \wedge b) &= \mathrm{P}(a|b)\mathrm{P}(b) \\
\mathrm{P}(b \wedge a) &= \mathrm{P}(b|a)\mathrm{P}(a)
\end{aligned}
$$

Note on Boldface Notation

- X denotes a random variable but $\mathbf{X}$ denotes a set of random variables
- $\mathrm{P}(X = x)$ denotes the probability that random variable X takes the value x, but $\mathbf{P}(X)$ denotes a probability distribution over X
  - $\mathbf{P}(Weather) = \langle 0.70, 0.20, 0.08, 0.02 \rangle$
  - $\mathbf{P}(X, Y) = \mathbf{P}(X|Y)\mathbf{P}(Y)$ stands for a set of equations (not matrix multiplication)

$$
\begin{aligned}
&P(X = x_1 \wedge Y = y_1) = P(X = x_1|Y = y_1)P(Y = y_1)\\
&P(X = x_1 \wedge Y = y_2) = P(X = x_1|Y = y_2)P(Y = y_2)\\
&\ldots
\end{aligned}
$$

Marginalisation and Conditioning

- Marginalisation or **summing out** for any sets of variables $\mathbf{Y}, \mathbf{Z}$

$$\mathbf{P}(\mathbf{Y}) = \sum_{Z \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, Z)$$

  This follows from the Law of Total Probability.

- Conditioning (variant of marginalisation, using the product rule)

$$\mathbf{P}(\mathbf{Y}) = \sum_{Z \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}|Z)P(Z)$$

Some Important Relations

Total Probability:

$$P(Y) = \sum_x P(Y, X = x) = \sum_x P(Y|X = x)P(X = x)$$

Conditional form of this:

$$P(Y|e) = \sum_x P(Y, X = x|e) = \sum_x P(Y|X = x, e)P(X = x|e)$$

Also:

$$P(\neg A|B) = 1 - P(A|B)$$

But:

$$P(A|\neg B) \neq 1 - P(A|B)$$

Normalisation

$$\begin{aligned} \mathrm{P}(\textit{cavity}|\textit{toothache}) &= \frac{\mathrm{P}(\textit{cavity} \wedge \textit{toothache})}{\mathrm{P}(\textit{toothache})} \\ \mathrm{P}(\neg\textit{cavity}|\textit{toothache}) &= \frac{\mathrm{P}(\neg\textit{cavity} \wedge \textit{toothache})}{\mathrm{P}(\textit{toothache})} \end{aligned}$$

- The denominator in both cases is the same
- It can be viewed as a normalisation constant α

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

General Inference Procedure

Let X be a random variable. We want to know its probabilities, given some values $\mathbf{e}$ for observable evidence variables $\mathbf{E}$. Let $\mathbf{Y}$ be the remaining unobserved variables. The query $\mathbf{P}(X|\mathbf{e})$ can be answered by:

$$\mathbf{P}(X|\mathbf{e}) = \alpha\mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Toothache Again

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Inference: Example

$$
\begin{aligned}
\mathbf{P}(\mathit{Cavity}|\mathit{toothache}) &= \alpha\mathbf{P}(\mathit{Cavity}, \mathit{toothache}) \\
&= \alpha\sum\mathbf{P}(\mathit{Cavity}, \mathit{toothache}, \mathit{Catch}) \\
&= \alpha[\mathbf{P}(\mathit{Cavity}, \mathit{toothache}, \mathit{catch}) + \\
&\qquad \mathbf{P}(\mathit{Cavity}, \mathit{toothache}, \neg\mathit{catch})] \\
&= \alpha[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle] \\
&= \alpha\langle 0.12, 0.08\rangle \\
&= \langle 0.60, 0.40\rangle
\end{aligned}
$$

Independence

- Propositions a and b are independent if and only if

$$\mathrm{P}(a \wedge b) = \mathrm{P}(a)\mathrm{P}(b)$$

- Or, by the product rule

$$\mathrm{P}(a|b) = \mathrm{P}(a) \quad \text{and} \quad \mathrm{P}(b|a) = \mathrm{P}(b)$$

- Between variables, this is

$$\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y) \quad \text{and}$$
$$\mathbf{P}(X|Y) = \mathbf{P}(X) \quad \text{and} \quad \mathbf{P}(Y|X) = \mathbf{P}(Y)$$

Illustration of Independence

Suppose we add a fourth variable: *Weather* with 4 values. We know (product rule) that:

$$\begin{aligned} & \mathrm{P}(\textit{toothache}, \textit{catch}, \textit{cavity}, \textit{Weather} = \textit{cloudy}) \\ = & \mathrm{P}(\textit{Weather} = \textit{cloudy} | \textit{toothache}, \textit{catch}, \textit{cavity}) \\ = & \mathrm{P}(\textit{Weather} = \textit{cloudy}) \mathrm{P}(\textit{toothache}, \textit{catch}, \textit{cavity}) \end{aligned}$$

We can do the same for every value of the 4 variables. That is:

$$\begin{aligned} & \mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity}, \textit{Weather}) \\ = & \mathbf{P}(\textit{Weather}) \mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity}) \end{aligned}$$

Bayes' Rule I

$P(C|+) = \frac{P(+|C) \times P(C)}{P(+)}$
In general:
$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$
Here is some terminology:
$P(A)$ is called the p̌rior probability of A
$P(A|B)$ is called the p̌osterior probability of A given B
$P(B|A)$ is called the ľikelihood of B given A
$P(B)$ is called the m̌arginal likelihood of B
$P(B) = \sum_a P(B|A = a)P(A = a)$.
(This is the rule of total probability.)
Conditional form of Bayes' rule:
$P(A|B, e) = \frac{P(B|A, e) \times P(A|e)}{P(B|e)}$ (Check!)

Bayes' Rule Calculations I

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \times P(A)}{P(B)} \\ P(\neg A|B) &= \end{aligned}$$

Bayes' Rule Calculations I

$$
\begin{aligned}
P(A|B) &= \frac{P(B|A) \times P(A)}{P(B)} \\
P(\neg A|B) &= \frac{P(B|\neg A) \times P(\neg A)}{P(B)}
\end{aligned}
$$

So, $P(A|B) = \alpha P(B|A)P(A)$ and $P(\neg A|B) = \alpha P(B|\neg A)P(\neg A)$.
Since $P(A|B) + P(\neg A|B) = 1$, we get α as:

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

Although we knew this anyway, it still gives us a slightly different way of calculating with Bayes' rule, by postponing the calculation of the normalising factor α

Bayes' Rule with Multivalued Variables

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)}$$

Or, with normalisation:

$$\mathbf{P}(Y|X) = \alpha\mathbf{P}(X|Y)\mathbf{P}(Y)$$

Why Use Bayes' Rule?

- Causal knowledge such as $\mathrm{P}(\textit{effect}|\textit{cause})$ is often more readily available, and reliable, than diagnostic knowledge such as $\mathrm{P}(\textit{cause}|\textit{effect})$

  $\mathrm{P}(\textit{stiff neck}|\textit{meningitis})$ versus $\mathrm{P}(\textit{meningitis}|\textit{stiff neck})$

- Bayes' rule lets us use causal knowledge to make diagnostic inferences

  Use $\mathrm{P}(\textit{stiff neck}|\textit{meningitis})$ to derive $\mathrm{P}(\textit{meningitis}|\textit{stiff neck})$

Bayes' Rule with Several Variables

$$\mathbf{P}(Y_1, \cdots, Y_m | X_1, \cdots, X_n) = \\ \alpha \mathbf{P}(X_1, \cdots, X_n | Y_1, \cdots, Y_m) \mathbf{P}(Y_1, \cdots, Y_m)$$

- We need to know the conditional probabilities for all possible combinations of the X_i, Y_js
  - With just boolean variables, this is $2^n 2^m$
  - Just like using the full joint distribution
- Can we use the idea of independence to reduce these requirements?

Conditional Independence

- If X and Y are conditionally independent given Z then
  $\mathbf{P}(X, Y|Z) = \mathbf{P}(X|Z)\mathbf{P}(Y|Z) \quad \text{and}$
  $\mathbf{P}(X|Y, Z) = \mathbf{P}(X|Z) \quad \text{and} \quad \mathbf{P}(Y|X, Z)\mathbf{P}(Y|Z)$
- $X_1, \cdots, X_n$ are conditionally independent, given $Y_1, \cdots, Y_m$ if and only if

$$\mathbf{P}(X_1, \cdots, X_n|Y_1, \cdots, Y_m) = \\ \alpha\mathbf{P}(X_1|Y_1, \cdots, Y_m)\cdots\mathbf{P}(X_n|Y_1, \cdots, Y_m)$$

- For boolean variables, this reduces the representation to $2n2^m$ combinations

Benefits of Conditional Independence

- Conditional independence allows probabilistic systems to scale up (tabular representations of full joint distributions become very large)

  $$\begin{aligned}&\mathbf{P}(\mathit{Toothache}, \mathit{Catch}, \mathit{Cavity})\\ &\quad = \mathbf{P}(\mathit{Toothache}, \mathit{Catch} | \mathit{Cavity})\mathbf{P}(\mathit{Cavity})\\ &\quad = \mathbf{P}(\mathit{Toothache} | \mathit{Cavity})\mathbf{P}(\mathit{Catch} | \mathit{Cavity})\mathbf{P}(\mathit{Cavity})\end{aligned}$$

- Conditional independence information is much more commonly available than absolute independence. For example, *Toothache* and *Catch* are not independent *per se*. But they are, <u>given</u> the knowledge about the presence or absence of a cavity.

Summary

- Basic probability statements include **prior** and **conditional** probabilities over simple and compound propositions.
- The **full joint distribution** specifies the probability of each complete assignment of values to random variables.
- **Absolute independence** between subsets of random variables may allow the full joint to be factored into smaller joint distributions.
- Bayes' rule performs computations using conditional distributions.
- **Conditional independence** may allow the full joint into smaller conditional distributions. It may allow more efficient computations using Bayes' rule.