Uncertainty
(Chapter 13, Russell & Norvig)

Introduction to Artificial Intelligence
CS 150
Lecture 14
• Last Programming assignment will be handed out later this week.
• I am doing probability today because we want you to implement a Naïve Bayes spam filter.
• To understand a Naïve Bayes classifier, you need to understand Bayes.
• To understand Bayes, you have to understand probability….so neural nets will have to wait until next week…
This Lecture

- Probability
- Syntax
- Semantics
- Inference rules
- How to build a Naïve Bayes classifier

Reading

Chapter 13
The real world: Things go wrong

Consider a plan for changing a tire after getting a flat using operators of RemoveTire(x), PutOnTire(x), InflateTire(x)

- Incomplete information
  - Unknown preconditions, e.g., Is the spare actually intact?
  - Disjunctive effects, e.g., Inflating a tire with a pump may cause the tire to inflate or a slow hiss or the tire may burst or …

- Incorrect information
  - Current state incorrect, e.g., spare NOT intact
  - Missing/incorrect postconditions (effects) in operators

- Qualification problem:
  - can never finish listing all the required preconditions and possible conditional outcomes of actions
Possible Solutions

• Conditional planning
  – Plan to obtain information (observation actions)
  – Subplan for each contingency, e.g.,
    
    \[
    \text{[Check(Tire1),}
    
    \text{[IF Intact(Tire1)}
    
    \text{THEN [Inflate(Tire1)]}
    
    \text{ELSE [CallAAA] ]}
    \]

  – Expensive because it plans for many unlikely cases

• Monitoring/Replanning
  – Assume normal states, outcomes
  – Check progress during execution, replan if necessary
  – Unanticipated outcomes may lead to failure (e.g., no AAA card)

• In general, some monitoring is unavoidable
Dealing with Uncertainty Head-on: Probability

- Let action $A_t = \text{leave for airport } t \text{ minutes before flight from Lindbergh Field}$
- Will $A_t$ get me there on time?

Problems:
1. Partial observability (road state, other drivers' plans, etc.)
2. Noisy sensors (traffic reports)
3. Uncertainty in action outcomes (turn key, car doesn’t start, etc.)
4. Immense complexity of modeling and predicting traffic

Hence a purely logical approach either
1) risks falsehood: “$A_{90}$ will get me there on time,” or
2) leads to conclusions that are too weak for decision making:
   “$A_{90}$ will get me there on time if there's no accident on I-5 and it doesn't rain and my tires remain intact, etc., etc.”

($A_{1440}$ might reasonably be said to get me there on time but I'd have to stay overnight in the airport …)
Methods for handling uncertainty

Default or nonmonotonic logic:
  Assume my car does not have a flat tire
  Assume $A_{125}$ works unless contradicted by evidence
Issues: What assumptions are reasonable? How to handle contradiction?

Rules with fudge factors:
  $A_{90} \rightarrow 0.3$ get there on time
  Sprinkler $\rightarrow 0.99$ WetGrass
  WetGrass $\rightarrow 0.7$ Rain
Issues: Problems with rule combination, e.g., does Sprinkler cause Rain??

Probability
  Given the available evidence,
  $A_{90}$ will get me there on time with probability 0.95

(Fuzzy logic handles degree of truth NOT uncertainty
  e.g., WetGrass is true to degree 0.2)
Fuzzy Logic in Real World

Fuzzy Logic Rice Cooker
Made exclusively for Williams-Sonoma, this cooker operates on microchip technology and turns out perfectly cooked rice every time. Programmed settings include sushi rice, brown rice, regular rice, soft grains, slow cooking (for stews and soups), steaming (for vegetables and fish) and quick cooking. The nonstick bowl sits in a cupped heating pad, which ensures even cooking. A timer lets you preset finish times up to 24 hours in advance; the cooker keeps rice warm for 12 hours. Steaming tray, rice paddle, measuring cup and booklet with steaming chart and recipes included. 585W, 5-cup cap. uncooked (10-cup cooked); 13” x 10¼” x 9” high. #07-2195378 \textdollar199.00
Probability

Probabilistic assertions summarize effects of
- **Ignorance**: lack of relevant facts, initial conditions, etc.
- **Laziness**: failure to enumerate exceptions, qualifications, etc.

Subjective or Bayesian probability:
Probabilities relate propositions to one's own state of knowledge

  e.g., \( P(A_{90} \text{ succeeds} \mid \text{no reported accidents}) = 0.97 \)

These are **NOT** assertions about the world, but represent belief about the
whether the assertion is true.

Probabilities of propositions change with new evidence:

  e.g., \( P(A_{90} \mid \text{no reported accidents, 5 a.m.}) = 0.99 \)

  (Analogous to logical entailment status; I.e., does \( \text{KB} \models \alpha \) )
Making decisions under uncertainty

• Suppose I believe the following:
  \[ P(A_{30} \text{ gets me there on time } | \ldots ) = 0.05 \]
  \[ P(A_{60} \text{ gets me there on time } | \ldots ) = 0.70 \]
  \[ P(A_{100} \text{ gets me there on time } | \ldots ) = 0.95 \]
  \[ P(A_{1440} \text{ gets me there on time } | \ldots ) = 0.9999 \]

• Which action to choose?

• Depends on my preferences for missing flight vs. airport cuisine, etc.

• *Utility theory* is used to represent and infer preferences

• *Decision theory = utility theory + probability theory*
Unconditional Probability

• Let A be a proposition, \( P(A) \) denotes the unconditional probability that A is true.

• Example: if Male denotes the proposition that a particular person is male, then \( P(\text{Male}) = 0.5 \) mean that without any other information, the probability of that person being male is 0.5 (a 50% chance).

• Alternatively, if a population is sampled, then 50% of the people will be male.

• Of course, with additional information (e.g. that the person is a CS151 student), the “posterior probability” will likely be different.
Axioms of probability

For any propositions $A$, $B$

1. $0 \leq P(A) \leq 1$
2. $P(\text{True}) = 1$ and $P(\text{False}) = 0$
3. $P(A \lor B) = P(A) + P(B) - P(A \land B)$

de Finetti (1931): an agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome.
Syntax

Similar to PROPOSITIONAL LOGIC: possible worlds defined by assignment of values to random variables.

Note: To make things confusing variables have first letter upper-case, and symbol values are lower-case

Propositional or Boolean random variables
  e.g., Cavity (do I have a cavity?)
  Include propositional logic expressions
  e.g., ¬Burglary ∨ Earthquake

Multivalued random variables
  e.g., Weather is one of <sunny,rain,cloudy,snow>
  Values must be exhaustive and mutually exclusive

A proposition is constructed by assignment of a value:
  e.g., Weather = sunny; also Cavity = true for clarity
Prior or unconditional probabilities of propositions

e.g., \( P(\text{Cavity}) = P(\text{Cavity}=\text{TRUE}) = 0.1 \)
\( P(\text{Weather}=\text{sunny}) = 0.72 \)
correspond to belief prior to arrival of any (new) evidence

Probability distribution gives probabilities of all possible values of the random variable.

Weather is one of \(<\text{sunny, rain, cloudy, snow}>\)
\( P(\text{Weather}) = <0.72, 0.1, 0.08, 0.1> \)
(normalized, i.e., sums to 1)
Joint probability distribution

Joint probability distribution for a set of variables gives values for each possible assignment to all the variables.

\[ P(\text{Toothache, Cavity}) \] is a 2 by 2 matrix.

<table>
<thead>
<tr>
<th></th>
<th>Toothache=true</th>
<th>Toothache = false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity=true</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Cavity=false</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

**NOTE:** Elements in table sum to 1 → 3 independent numbers.
The importance of independence

- $P(Weather, Cavity)$ is a 4 by 2 matrix of values:

<table>
<thead>
<tr>
<th>Weather =</th>
<th>sunny</th>
<th>rain</th>
<th>cloudy</th>
<th>snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity=True</td>
<td>.072</td>
<td>.01</td>
<td>.008</td>
<td>.01</td>
</tr>
<tr>
<td>Cavity=False</td>
<td>.648</td>
<td>.09</td>
<td>.072</td>
<td>.09</td>
</tr>
</tbody>
</table>

- But these are independent (usually):
  - Recall that if $X$ and $Y$ are independent then:
  - $P(X=x, Y=y) = P(X=x) * P(Y=y)$

$Weather$ is one of $<sunny, rain, cloudy, snow>$

$P(Weather) = <0.72, 0.1, 0.08, 0.1> \text{ and } P(Cavity) = .1$
The importance of independence

If we compute the *marginals* (sums of rows or columns), we see:

<table>
<thead>
<tr>
<th>Weather =</th>
<th>sunny</th>
<th>rain</th>
<th>cloudy</th>
<th>snow</th>
<th>marginals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity=True</td>
<td>.072</td>
<td>.01</td>
<td>.008</td>
<td>.01</td>
<td>.1</td>
</tr>
<tr>
<td>Cavity=False</td>
<td>.648</td>
<td>.09</td>
<td>.072</td>
<td>.09</td>
<td>.9</td>
</tr>
<tr>
<td>marginals</td>
<td>.72</td>
<td>.1</td>
<td>.08</td>
<td>.1</td>
<td></td>
</tr>
</tbody>
</table>

Each entry is just the product of the marginals (*because* they’re independent)!

So, instead of using 7 numbers, we can represent these by just recording 4 numbers, 3 from Weather and 1 from Cavity:

Weather is one of <sunny,rain,cloudy,snow>

\[
P(\text{Weather}) = <0.72, 0.1, 0.08, (1- \text{Sum of the others})>
\]

\[
P(\text{Cavity}) = <.1, (1-.1)>
\]

This can be much more efficient (more later…)
Conditional Probabilities

- **Conditional** or posterior probabilities
  
  e.g., $P(\text{Cavity} \mid \text{Toothache}) = 0.8$
  
  What is the probability of having a cavity given that the patient has a toothache? (as another example: $P(\text{Male} \mid \text{CS151Student}) = ??$)

- If we know more, e.g., dental probe catches, then we have
  
  $P(\text{Cavity} \mid \text{Toothache}, \text{Catch}) = 0.9$ (say)

- If we know even more, e.g., Cavity is also given, then we have
  
  $P(\text{Cavity} \mid \text{Toothache}, \text{Cavity}) = 1$
  
  Note: the less specific belief remains valid after more evidence arrives, but is not always useful.

- New evidence may be irrelevant, allowing simplification, e.g.,
  
  $P(\text{Cavity} \mid \text{Toothache}, \text{PadresWin}) = P(\text{Cavity} \mid \text{Toothache}) = 0.8$
  
  (again, this is because Cavities and Baseball are independent)
Conditional Probabilities

- **Conditional** or posterior probabilities
  e.g., $P(\text{Cavity} \mid \text{Toothache}) = 0.8$
  What is the probability of having a cavity given that the patient has a toothache?
- **Definition of conditional probability:**
  $$P(A \mid B) = \frac{P(A, B)}{P(B)} \quad \text{if} \quad P(B) \neq 0$$

I.e., $P(A \mid B)$ means, our universe has shrunk to the one in which $B$ is true.

$P(A \mid B)$ is the ratio of the **red** area to the **yellow** area.
Conditional Probabilities

- **Conditional** or posterior probabilities
  - e.g., \( P(\text{Cavity} \mid \text{Toothache}) = 0.8 \)
  - What is the probability of having a cavity given that the patient has a toothache?

- **Definition of conditional probability:**
  \[
P(A \mid B) = \frac{P(A, B)}{P(B)} \quad \text{if} \quad P(B) \neq 0
  \]

<table>
<thead>
<tr>
<th></th>
<th>Toothache=true</th>
<th>Toothache = false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity=true</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Cavity=false</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

\[
P(\text{Cavity} \mid \text{Toothache}) = \frac{0.04}{(0.04+.01)} = 0.8
\]
Definition of conditional probability:

\[ P(A | B) = \frac{P(A, B)}{P(B)} \quad \text{if} \quad P(B) \neq 0 \]

Product rule gives an alternative formulation:

\[ P(A, B) = P(A | B)P(B) = P(B|A)P(A) \]

Example:

\[
\begin{align*}
P(\text{Male}) &= 0.5 \\
P(\text{CS151Student}) &= 0.0037 \\
P(\text{Male} | \text{CS151Student}) &= 0.9
\end{align*}
\]

What is the probability of being a male CS151 student?

\[
\begin{align*}
P(\text{Male, CS151Student}) &= P(\text{Male} | \text{CS151Student})P(\text{CS151Student}) \\
&= 0.0037 \times 0.9 = 0.0033
\end{align*}
\]
Conditional probability cont.

A general version holds for whole distributions, e.g.,
\[
P(\text{Weather}, \text{Cavity}) = P(\text{Weather} | \text{Cavity}) \cdot P(\text{Cavity})
\]
(View as a 4 X 2 set of equations, not matrix multiplication - The book calls this a pointwise product)

The **Chain rule** is derived by successive application of product rule:
\[
P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1}) \cdot P(X_n | X_1, \ldots, X_{n-1})
\]
\[
= P(X_1, \ldots, X_{n-2}) \cdot P(X_{n-1} | X_1, \ldots, X_{n-2}) \cdot P(X_n | X_1, \ldots, X_{n-1})
\]
\[
= \ldots
\]
\[
= \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})
\]
Bayes Rule

From product rule $P(A \land B) = P(A|B)P(B) = P(B|A)P(A)$, we can obtain Bayes' rule

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

Why is this useful???

For assessing diagnostic probability from causal probability:

- e.g. Cavities as the cause of toothaches
- Sleeping late as the cause for missing class
- Meningitis as the cause of a stiff neck

$$P(Cause | Effect) = \frac{P(Effect | Cause)P(Cause)}{P(Effect)}$$
Bayes Rule: Example

\[ P(Cause \mid Effect) = \frac{P(Effect \mid Cause)P(Cause)}{P(Effect)} \]

Let \( M \) be meningitis, \( S \) be stiff neck

\[
P(M) = 0.0001
\]

\[
P(S) = 0.1
\]

\[
P(S \mid M) = 0.8
\]

\[
P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008
\]

Note: posterior probability of meningitis still very small!
Bayes Rule: Another Example

Let \( A \) be AIDS, \( T \) be positive test result for AIDS

\[ P(T|A) = 0.99 \text{ “true positive” rate of test (aka a “hit”) } \]

\[ P(T|\neg A) = 0.005 \text{ “false positive” rate of test (aka a “false alarm”) } \]

\[ P(\neg T|A) = 1 - 0.99 = 0.01 \text{ “false negative” or “miss” } \]

\[ P(\neg T|\neg A) = 0.995 \text{ “correct rejection” } \]

Seems like a pretty good test, right?
Bayes Rule: Another Example

Let $A$ be AIDS, $T$ be positive test result for AIDS

$P(T|A) = 0.99$ “true positive” rate of test (aka a “hit”)

$P(T|\neg A) = 0.005$ “false positive” rate of test (aka a “false alarm”)

$P(\neg T | A) = 1 - 0.99 = 0.01$ “false negative” or “miss”

$P(\neg T | \neg A) = 0.995$ “correct rejection”

Now, suppose you are from a low-risk group, so $P(A) = 0.001$, and you get a positive test result. Should you worry?

Let’s compute it using Bayes’ rule: $P(A|T) = \frac{P(T|A)P(A)}{P(T)}$

Hmmm. Wait a minute: How do we know $P(T)$?

Recall there are two possibilities: Either you have AIDS or you don’t.

Therefore: $P(A|T) + P(\neg A|T) = 1$

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

So, $P(A|T) + P(\neg A|T) = \frac{P(T|A)P(A)}{P(T)} + \frac{P(T|\neg A)P(\neg A)}{P(T)} = 1$
Bayes Rule: Another Example

\[ P(A|T) = P(T|A)*P(A)/P(T) \]

Hmmm. Wait a minute: How do we know \( P(T) \)?
Recall there are two possibilities: Either you have AIDS or you don’t.
Therefore: \[ \textbf{P}(A|T) + P(\neg A|T) = 1. \]
Now, by Bayes’ rule:
\[ P(\neg A|T) = P(T|\neg A)*P(\neg A)/P(T) \]
So,
\[ P(A|T) + P(\neg A|T) = P(T|A)*P(A)/P(T) + P(T|\neg A)*P(\neg A)/P(T) = 1 \]
\[ \Rightarrow \ P(T|A)*P(A) + P(T|\neg A)*P(\neg A) = P(T) \]
\[ \Rightarrow \textbf{The denominator is always the sum of the numerators} \textbf{ if they are exhaustive of the possibilities - a normalizing constant to make the probabilities sum to 1!} \]
\[ \Rightarrow \textbf{We can compute the numerators and then normalize afterwards!} \]
Bayes Rule: Back to the Example

Let $A$ be AIDS, $T$ be positive test result for AIDS

$P(T|A)=0.99$ “true positive” rate of test (aka a “hit”)

$P(T|\neg A)= 0.005$ “false positive” rate of test (aka a “false alarm”)

$P(\neg T | A) = 1-0.99=.01$ “false negative” or “miss”

$P(\neg T|\neg A) = 0.995$ “correct rejection”

Now, suppose you are from a low-risk group, so $P(A) = .001$, and you get a positive test result. Should you worry?

$P(A|T) \propto P(T|A)*P(A) = 0.99*.001 = .00099$

$P(\neg A|T) \propto P(T|\neg A)*P(\neg A) = 0.005*.999 = 0.004995$

$P(A|T) = .00099/(.00099+.004995) = 0.165$

*Maybe* you should worry, or maybe you should get re-tested…. 
Let $A$ be AIDS, $T$ be positive test result for AIDS

$P(T|A) = 0.99$ “true positive” rate of test (aka a “hit”)

$P(T|\neg A) = 0.005$ “false positive” rate of test (aka a “false alarm”)

$P(\neg T | A) = 1 - 0.99 = 0.01$ “false negative” or “miss”

$P(\neg T | \neg A) = 0.995$ “correct rejection”

Now, suppose you are from a high-risk group, so $P(A) = 0.01$, and you get a positive test result. Should you worry?

$P(A|T) \propto P(T|A)P(A) = 0.99 * 0.01 = 0.0099$

$P(\neg A|T) \propto P(T|\neg A)P(\neg A) = 0.005 * 0.99 = 0.00495$

$P(A|T) = 0.0099 / (0.0099 + 0.00495) = 0.667$

$\Rightarrow$ You should worry!
Bayes Rule: Another Example

Let $A$ be AIDS, $T$ be positive test result for AIDS
$P(T|A)=0.99$ “true positive” rate of test (aka a “hit”)
$P(T|\neg A)=0.005$ “false positive” rate of test (aka a “false alarm”)
$P(\neg T|A) = 1-0.99=.01$ “false negative” or “miss”
$P(\neg T|\neg A) = 0.995$ “correct rejection”

Now, suppose you are from a low-risk group, so $P(A) = .001$, and you get a positive test result. Should you worry?

Let’s compute it using Bayes’ rule: $P(A|T) = P(T|A)*P(A)/P(T)$

Hmmm. Wait a minute: How do we know $P(T)$?

Recall there are two possibilities: Either you have AIDS or you don’t.

Therefore: $P(A|T) + P(\neg A|T) = 1$.

$P(\neg A|T) = P(T|\neg A)*P(\neg A)/P(T)$

So, $P(A|T) + P(\neg A|T) = P(T|A)*P(A)/P(T) + P(T|\neg A)*P(\neg A)/P(T) = 1$
A generalized Bayes Rule

- More general version conditionalized on some background evidence $E$

$$P(A \mid B, E) = \frac{P(B \mid A, E)P(A \mid E)}{P(B \mid E)}$$
Full joint distributions

A **complete probability model** specifies every entry in the joint distribution for all the variables $X = X_1, \ldots, X_n$
i.e., a probability for each possible world $X_1 = x_1, \ldots, X_n = x_n$

E.g., suppose **Toothache** and **Cavity** are the random variables:

<table>
<thead>
<tr>
<th></th>
<th>Toothache=true</th>
<th>Toothache = false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity=true</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Cavity=false</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Possible worlds are mutually exclusive $\Rightarrow P(w_1 \land w_2) = 0$
Possible worlds are exhaustive $\Rightarrow w_1 \lor \ldots \lor w_n$ is True
hence $\sum_i P(w_i) = 1$
Using the full joint distribution

<table>
<thead>
<tr>
<th></th>
<th>Toothache=true</th>
<th>Toothache = false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity=true</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Cavity=false</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

What is the unconditional probability of having a Cavity?

\[
P(\text{Cavity}) = P(\text{Cavity} \land \text{Toothache}) + P(\text{Cavity} \land \neg \text{Toothache})
\]
\[
= 0.04 + 0.06 = 0.1
\]

What is the probability of having either a cavity or a Toothache?

\[
P(\text{Cavity} \lor \text{Toothache})
\]
\[
= P(\text{Cavity,Toothache}) + P(\text{Cavity,} \neg \text{Toothache}) + P(\neg \text{Cavity,Toothache})
\]
\[
= 0.04 + 0.06 + 0.01 = 0.11
\]
Using the full joint distribution

<table>
<thead>
<tr>
<th></th>
<th>Toothache=true</th>
<th>Toothache = false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity=true</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Cavity=false</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

What is the probability of having a cavity given that you already have a toothache?

\[
P(\text{Cavity} \mid \text{Toothache}) = \frac{P(\text{Cavity} \land \text{Toothache})}{P(\text{Toothache})} = \frac{0.04}{0.04 + 0.01} = 0.8
\]
Normalization

Suppose we wish to compute a posterior distribution over random variable $A$ given $B=b$, and suppose $A$ has possible values $a_1 \ldots a_m$

We can apply Bayes' rule for each value of $A$:

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

$$\ldots$$

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

Adding these up, and noting that $\sum_i P(A=a_i | B=b) = 1$:

$$P(B=b) = \sum_i P(B=b | A=a_i)P(A=a_i)$$

This is the normalization factor denoted $\alpha = 1/P(B=b)$:

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

Typically compute an unnormalized distribution, normalize at end

e.g., suppose $P(B=b | A)P(A) = <0.4,0.2,0.2>$

then $P(A|B=b) = \alpha <0.4,0.2,0.2>$

$$= <0.4,0.2,0.2> / (0.4+0.2+0.2) = <0.5,0.25,0.25>$$
Marginalization

Given a conditional distribution $P(X \mid Z)$, we can create the unconditional distribution $P(X)$ by marginalization:

$$P(X) = \sum_z P(X \mid Z=z) \cdot P(Z=z) = \sum_z P(X, Z=z)$$

In general, given a joint distribution over a set of variables, the distribution over any subset (called a marginal distribution for historical reasons) can be calculated by summing out the other variables.
Conditioning

Introducing a variable as an extra condition:

\[ P(X|Y) = \sum_z P(X | Y, Z=z) P(Z=z | Y) \]

Why is this useful??

Intuition: it is often easier to assess each specific circumstance, e.g.,

\[
P(\text{RunOver} | \text{Cross})
= P(\text{RunOver} | \text{Cross, Light=green})P(\text{Light=green} | \text{Cross})
+ P(\text{RunOver} | \text{Cross, Light=yellow})P(\text{Light=yellow} | \text{Cross})
+ P(\text{RunOver} | \text{Cross, Light=red})P(\text{Light=red} | \text{Cross})
\]
Absolute Independence

• Two random variables $A$ and $B$ are (absolutely) **independent** iff
  \[ P(A, B) = P(A)P(B) \]

• Using product rule for $A$ & $B$ independent, we can show:
  \[ P(A, B) = P(A | B)P(B) = P(A)P(B) \]
  Therefore \[ P(A | B) = P(A) \]

• If $n$ Boolean variables are independent, the full joint is:
  \[ P(X_1, \ldots, X_n) = \prod_i P(X_i) \]
  Full joint is generally specified by $2^n - 1$ numbers, but when independent only $n$ numbers are needed.

• Absolute independence is a very strong requirement, seldom met!!
Conditional Independence

- Some evidence may be irrelevant, allowing simplification, e.g.,

\[ P(\text{Cavity} \mid \text{Toothache, PadresWin}) = P(\text{Cavity} \mid \text{Toothache}) = 0.8 \]

- This property is known as **Conditional Independence** and can be expressed as:

\[ P(X \mid Y, Z) = P(X \mid Z) \]

which says that \( X \) and \( Y \) independent given \( Z \).

- If I have a cavity, the probability that the dentist’s probe catches in it doesn't depend on whether I have a toothache:

1. \( P(\text{Catch} \mid \text{Toothache, Cavity}) = P(\text{Catch} \mid \text{Cavity}) \)
   i.e., \( \text{Catch} \) is conditionally independent of \( \text{Toothache} \) given \( \text{Cavity} \)
   This doesn’t say anything about \( P(\text{Catch} \mid \text{Toothache}) \)
   The same independence holds if I haven't got a cavity:

2. \( P(\text{Catch} \mid \text{Toothache, ~Cavity}) = P(\text{Catch} \mid \sim \text{Cavity}) \)
Conditional independence contd.

Equivalent statements to

\[ P(\text{Catch} \mid \text{Toothache, Cavity}) = P(\text{Catch} \mid \text{Cavity}) \]  

\( (\ast) \)

1.a \[ P(\text{Toothache} \mid \text{Catch, Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \]

\[
P(\text{Toothache} \mid \text{Catch, Cavity}) \\
= P(\text{Catch} \mid \text{Toothache, Cavity}) P(\text{Toothache} \mid \text{Cavity}) / P(\text{Catch} \mid \text{Cavity}) \\
= P(\text{Catch} \mid \text{Cavity}) P(\text{Toothache} \mid \text{Cavity}) / P(\text{Catch} \mid \text{Cavity}) \quad \text{(from \(\ast\))} \\
= P(\text{Toothache} \mid \text{Cavity})
\]

1.b \[ P(\text{Toothache, Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) \]

\[
P(\text{Toothache, Catch} \mid \text{Cavity}) \\
= P(\text{Toothache} \mid \text{Catch, Cavity}) P(\text{Catch} \mid \text{Cavity}) \quad \text{(product rule)} \\
= P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) \quad \text{(from 1a)}
\]
Using Conditional Independence

Full joint distribution can now be written as

\[
P(\text{Toothache}, \text{Catch}, \text{Cavity})
\]

\[
= (\text{Toothache}, \text{Catch} \mid \text{Cavity}) \cdot P(\text{Cavity})
\]

\[
= P(\text{Toothache} \mid \text{Cavity}) \cdot P(\text{Catch} \mid \text{Cavity}) \cdot P(\text{Cavity})
\]  

(from 1.b)

Specified by: \(2 + 2 + 1 = 5\) independent numbers

Compared to \(7\) for general joint

or \(3\) for unconditionally independent.