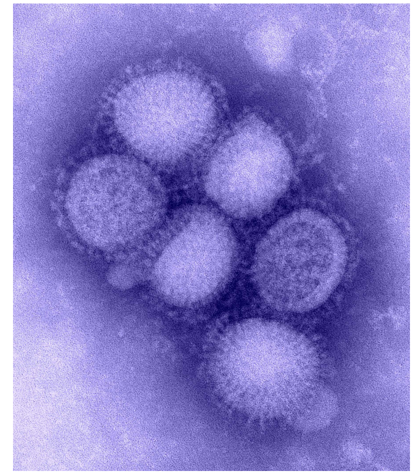
Reasoning with Uncertainty

Topics:

- Why is uncertainty important?
- How do we represent and reason with uncertain knowledge?
- Progress in research:
    - 1980s: rule-based representation of uncertainty (MYCIN, Prospector)
    - 1990s to present: graphical models, probabilistic expert systems (Munin, Promedas)
    - latest developments: integration of probability theory and logic

Why important: biomedical

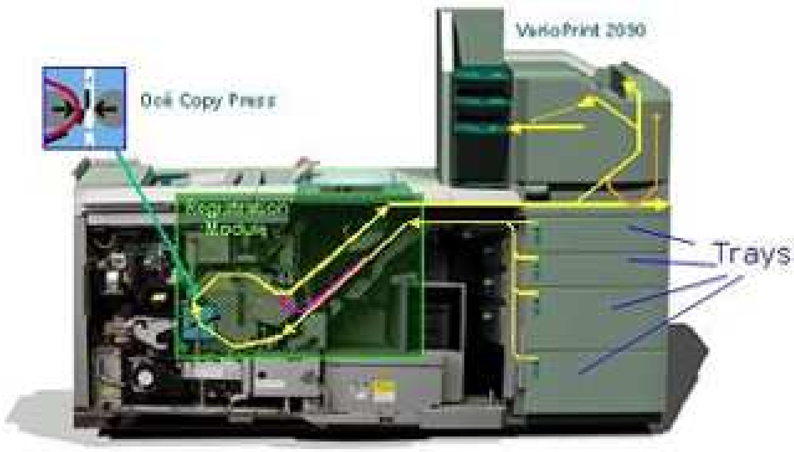
Have you got Mexican Flu?

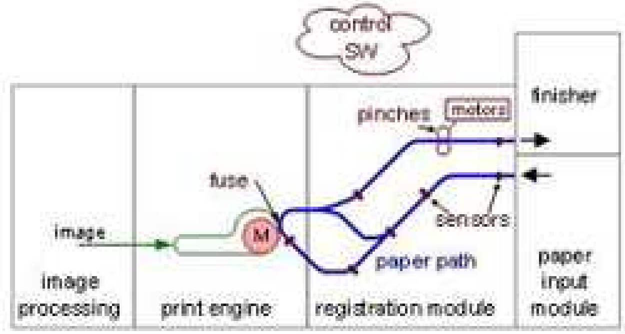


- M: mexican flu; C: chills; S: sore throat
- Probability of mexican flu given sore throat?

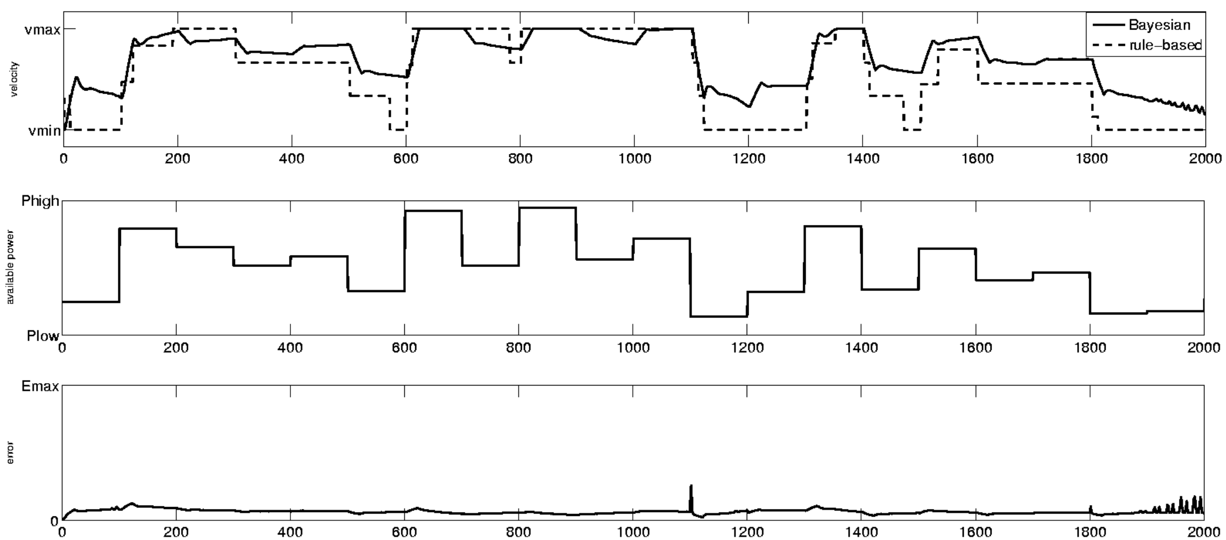
Why important: embedded systems

Control of behaviour of large production printer





Speed v given available power P and required energy:

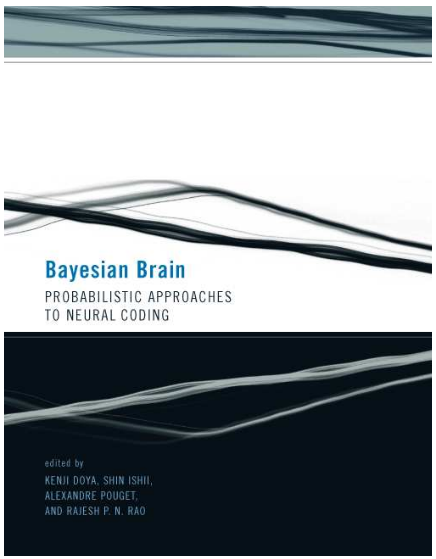


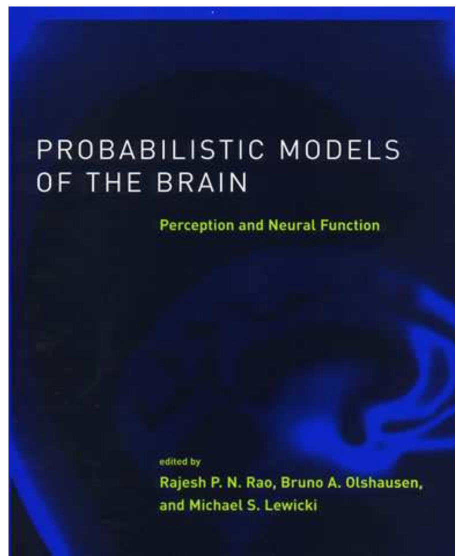
Why important: agents

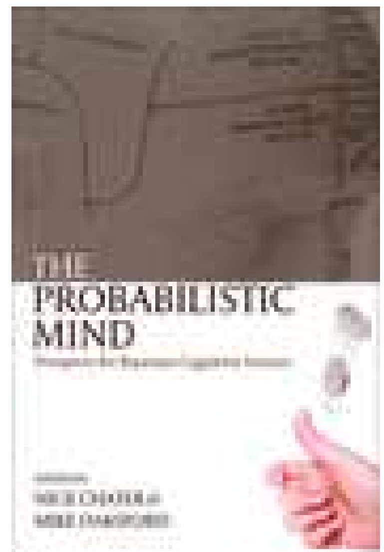
- Agents (robots) perceive an incomplete image of the world using sensors that are inherently unreliable
- Partially observable state

Why important: human reasoning

- Probabilistic theories of human brain function and cognition are dominant







Representation of uncertainty

- Representation of uncertainty is clearly important!
- How to do it? For example rule based:
  - e: evidence
  - h: hypothesis

$$e_1 \wedge \cdots \wedge e_n \to h_x$$

If e_1, $e_2, \ldots, e_n$ are true (observed), then conclusion h is true with certainty x

  - How to proceed when e_i, $i = 1, \ldots, n$ are uncertain?

$\Rightarrow$ uncertainty propagation/inference/reasoning

Theory

- We need a basic ‘theory’, e.g.
  - Certainty-factor model (Mycin)
  - Subjective Bayesian method (Prospector) – not discussed
  - Dempster-Shafer theory – not discussed
  - Probability theory
- This theory should tell us how to draw inferences with uncertainty statements

Rule-based uncertain knowledge

Early, simple approach – certainty-factor calculus:

- *fever* $\wedge$ *myalgia* $\longrightarrow$ *flu*$_{\mathrm{CF}=0.8}$

- Example how its works:
  - $\mathrm{CF}(\textit{fever}, e) = 0.6$;
    $\mathrm{CF}(\textit{myalgia}, e) = 1$
    (e is evidence; background knowledge)
  - Combination functions:
    $$\begin{aligned}\mathrm{CF}(\textit{flu}, e) \\ &= 0.8 \cdot \max\{0, \min\{\mathrm{CF}(\textit{fever}, e), \mathrm{CF}(\textit{myalgia}, e)\}\} \\ &= 0.8 \cdot \max\{0, \min\{0.6, 1\}\} = 0.48\end{aligned}$$

Certainty factor calculus

- Developed by E.H. Shortliffe and B.G. Buchanan for rule-based expert systems
- Applied in MYCIN, the expert system for the diagnosis of infectious disease
- Probability theory was seen as unsatisfactory:
  - Not enough data to obtain sufficient statistics
  - Medical knowledge must be explicitly represented
  - Line of reasoning should be explained by the system

Inference rules

- Define combination functions $f_\wedge$, $f_\vee$, f_{prop}, f_{co}, where:
  - $f_\wedge$: combines uncertainty w.r.t. conjunctions of uncertain evidence
  - $f_\vee$: combines uncertainty w.r.t. disjunctions of uncertain evidence
  - f_{co}: combines uncertainty for two co-concluding rules:

$$e_1 \longrightarrow h_x \qquad \textit{contact\_chicken} \longrightarrow \textit{flu}_{0.01}$$

$$e_2 \longrightarrow h_y \qquad \textit{train\_contact\_humans} \longrightarrow \textit{flu}_{0.1}$$

  - f_{prop}: propagation of uncertain evidence e to a hypothesis h

Certainty factor calculus

- Weak relationship to probability theory
- Certainty factors (CFs): subjective estimates of uncertainty with $\mathrm{CF}(x, e) \in [-1, 1]$ ($\mathrm{CF}(x, e) = -1$ false, $\mathrm{CF}(x, e) = 0$ unknown, and $\mathrm{CF}(x, e) = 1$ true)
- CF-calculus offers fill-in for combination functions: $f_\wedge$, $f_\vee$, f_{co}, f_{prop}

Combination functions

- $f_\wedge$
  - rule: $e_1 \wedge e_2 \rightarrow h_{\mathsf{CF}(h,e)}$ with
  - uncertain evidence $\mathsf{CF}(e_1, e')$ and $\mathsf{CF}(e_2, e')$

  then:

$$\mathsf{CF}(e_1 \wedge e_2, e') = \min\{\mathsf{CF}(e_1, e'), \mathsf{CF}(e_2, e')\}$$

- $f_\vee$
  - rule: $e_1 \vee_2 \rightarrow h_{\mathsf{CF}(h,e)}$ with
  - uncertain evidence $\mathsf{CF}(e_1, e')$ and $\mathsf{CF}(e_2, e')$

  then:

$$\mathsf{CF}(e_1 \vee e_2, e') = \max\{\mathsf{CF}(e_1, e'), \mathsf{CF}(e_2, e')\}$$

Combination functions

- f_{prop}
  - rule $e \rightarrow h_{\text{CF}(h,e)}$
  - uncertain evidence w.r.t. e, i.e. $\text{CF}(e, e')$ (e' includes all evidence so far)

then:

$$\text{CF}(h, e') = \text{CF}(h, e) \cdot \max\{0, \text{CF}(e, e')\}$$

Combination functions

- f_{co}:
  - two rules:

    $e_1 \rightarrow h_{\text{CF}(h,e_1)}$
    $e_2 \rightarrow h_{\text{CF}(h,e_2)}$
  - uncertain evidence $\text{CF}(e_1, e')$ and $\text{CF}(e_2, e')$
  - Let $\text{CF}(h, e'_1) = x$ via rule 1 and $\text{CF}(h, e'_2) = y$ via rule 2 (using f_{prop})
  - Then:

$$\text{CF}(h, e') = \begin{cases} x + y(1 - x) & \text{if } x, y \geq 0 \\ x + y(1 + x) & \text{if } x, y < 0 \\ \frac{x+y}{1-\min\{|x|,|y|\}} & \text{otherwise} \end{cases}$$

Example

$$\mathcal{R} = \{R_1 : \textit{flu} \longrightarrow \textit{fever}_{\mathrm{CF}(\textit{fever},\textit{flu})=0.8},$$
$$R_2 : \textit{common-cold} \longrightarrow \textit{fever}_{\mathrm{CF}(\textit{fever},\textit{common-cold})=0.3}\}$$

- Evidence: $\mathrm{CF}(\textit{flu}, e') = 0.6$ and $\mathrm{CF}(\textit{common-cold}, e') = 1$
- What is the certainty factor for *fever*?

Solution

Application of f_{prop}

Evidence: $\mathrm{CF}(\textit{flu}, e') = 0.6$ and $\mathrm{CF}(\textit{common-cold}, e') = 1$

For rule R_1:

$$\begin{aligned} \text{CF}(\textit{fever}, e_1') &= \text{CF}(\textit{fever}, \textit{flu}) \cdot \max\{0, \text{CF}(\textit{flu}, e')\} \\ &= 0.8 \cdot 0.6 = 0.48 \end{aligned}$$

for rule R_2 this yields $\text{CF}(\textit{fever}, e_2') = 0.3$

Application of f_{co}:

$$\begin{aligned} \text{CF}(\textit{fever}, e') &= \text{CF}(\textit{fever}, e_1') + \text{CF}(\textit{fever}, e_2')(1 - \text{CF}(\textit{fever}, e_1')) \\ &= 0.48 + 0.3(1 - 0.48) = 0.636 \end{aligned}$$

However . . .

$$\textit{fever} \wedge \textit{myalgia} \longrightarrow \textit{flu}_{\mathrm{CF}=0.8}$$

- How likely is the occurrence of fever or myalgia given that the patient has flu?
- How likely is the occurrence of fever or myalgia in the absence of flu?
- How likely is the presence of flu when just fever is present?
- How likely is the presence of no flu when just fever is present?

Problems with the CF model

- CF model requires rules to be encoded in the direction in which they are used.
- CF reasoning becomes unsound if strong assumptions fail to hold (consequence of combination functions)
- Assumption of modularity: A rule *if* e *then* h conforms to the following:
  - Detachment: given e we can conclude h no matter how we established e
  - Locality: given e we can conclude h no matter what else we know to be true
- Holds for logic but not for probability theory!
- Illogical results are obtained such as the dependence of a diagnosis on the order in which findings are entered

The inevitability of probability theory

Probability theory is nothing but common sense reduced to calculation.
Laplace, 1819

- Basic postulates for any measure of belief (Cox, 1946; Jaynes, 2003):
  1. Representation of degrees of plausibility by real numbers
  2. Qualitative correspondence with common sense
  3. Consistency
- Axioms of probability theory follow as a logical consequence from these postulates
- If you do not reason according to Probability Theory, you can be made to act irrationally (de Finetti)

Probability space

- A probability space represents our uncertainty regarding an *experiment* (DB query) and consists of:
  - A *sample space* Ω consisting of a set of outcomes
  - A *probability measure* P which is a real function of the subsets of Ω
- A set of outcomes $A \subseteq \Omega$ is called an event
- $P(A)$ represents how likely it is that an experiment's outcome will be a member of A.

Example

- Suppose our experiment is to examine whether someone has a cold and its related symptom fever.
- The outcomes are defined by

$$\begin{aligned}\Omega &= \{(\text{cold}, \text{fever}), (\text{no cold}, \text{fever}), \\ &\quad (\text{cold}, \text{no fever}), (\text{no cold}, \text{no fever})\}\end{aligned}$$

- and we may define probabilities

$$\begin{aligned}P(\{(\text{cold}, \text{fever}), (\text{cold}, \text{no fever})\}) &= 0.001 \\ P(\{(\text{no cold}, \text{fever}), (\text{cold}, \text{fever})\}) &= 0.01 \\ &\vdots\end{aligned}$$

- A probability measure P can be completely described by assigning a probability to each event $\omega \in \Omega$

Axioms of probability theory

- P should obey three axioms:
  1. $P(A) \geq 0$ for all events A
  2. $P(\Omega) = 1$
  3. $P(A \cup B) = P(A) + P(B)$ for disjoint events A and B
- Some consequences:
  - $P(A) = 1 - P(\Omega \setminus A)$
  - $P(\emptyset) = 0$
  - If $A \subseteq B$ then $P(A) \leq P(B)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$
- Given these axioms and a completely defined probability measure any quantity of interest can be computed!

Random variables

- A random variable X is a function from the sample space Ω to the domain $\text{dom}(X)$ of X.
- $p_X\colon \text{dom}(X) \to \mathbb{R}$ is the density of X if for all $x \in \text{dom}(X)$:

$$p_X(x) = P(\{\omega\colon X(\omega) = x\})$$

- Often written as $P(X = x)$, $P(x)$, $p(x)$, $\ldots$.
- Random variables allow us to abstract away from events: $P(\text{no flu}) = P(\{\omega\colon Flu(\omega) = \text{no flu}\})$
- Random variables can either be discrete ($P(\text{Flu = no flu})$) or continuous ($P(\text{Temperature} \leq 38^\circ)$)
- In this course we only work with discrete random variables

Joint density

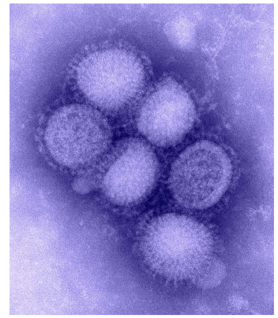
- The joint density for two random variables X and Y is given by

$$p_{XY}(x, y) = P(\{\omega \colon X(\omega) = x, Y(\omega) = y\})$$

- Often written as $P(X = x, Y = y)$, $P(x, y)$, $p(x, y)$, $\ldots$
- Generalizes to multiple random variables
- From now on we work with random variables and (joint) densities instead of events

Example

Have you got Mexican Flu?

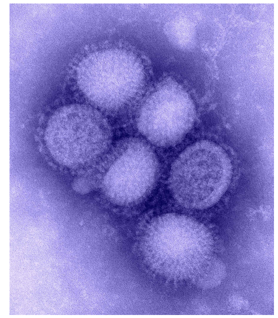


$$
\begin{aligned}
P(m, c, s) &= 0.009215 \\
P(m, \bar{c}, s) &= 0.000485 \\
P(m, c, \bar{s}) &= 0.000285 \\
P(m, \bar{c}, \bar{s}) &= 1.5 \cdot 10^{-5} \\
P(\bar{m}, c, s) &= 9.9 \cdot 10^{-6} \\
P(\bar{m}, \bar{c}, s) &= 0.0098901 \\
P(\bar{m}, c, \bar{s}) &= 0.0009801 \\
P(\bar{m}, \bar{c}, \bar{s}) &= 0.97912
\end{aligned}
$$

- M: mexican flu; C: chills; S: sore throat

Example

Have you got Mexican Flu?



$$
\begin{aligned}
P(m,c,s) &= 0.009215\\
P(m,\bar{c},s) &= 0.000485\\
P(m,c,\bar{s}) &= 0.000285\\
P(m,\bar{c},\bar{s}) &= 1.5\cdot 10^{-5}\\
P(\bar{m},c,s) &= 9.9\cdot 10^{-6}\\
P(\bar{m},\bar{c},s) &= 0.0098901\\
P(\bar{m},c,\bar{s}) &= 0.0009801\\
P(\bar{m},\bar{c},\bar{s}) &= 0.97912
\end{aligned}
$$

- M: mexican flu; C: chills; S: sore throat
- Probability of mexican flu and sore throat?

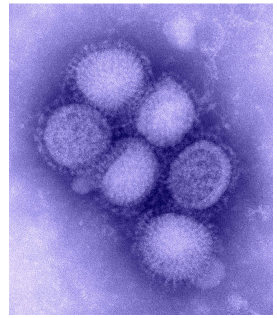
Marginalization

- Joint probability distribution $P(X) = P(X_1, X_2, \ldots, X_n)$
- U and V are mutually exclusive and collectively exhaustive subsets of X.
  - Marginalization:

$$P(u) = \sum_{v \in \text{dom}(v)} P(u, v)$$

Example

Have you got Mexican Flu?

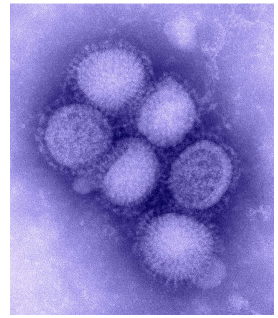


$$
\begin{aligned}
P(m,c,s) &= 0.009215\\
P(m,\bar{c},s) &= 0.000485\\
P(m,c,\bar{s}) &= 0.000285\\
P(m,\bar{c},\bar{s}) &= 1.5\cdot 10^{-5}\\
P(\bar{m},c,s) &= 9.9\cdot 10^{-6}\\
P(\bar{m},\bar{c},s) &= 0.0098901\\
P(\bar{m},c,\bar{s}) &= 0.0009801\\
P(\bar{m},\bar{c},\bar{s}) &= 0.97912
\end{aligned}
$$

- M: mexican flu; C: chills; S: sore throat
- Probability of mexican flu and sore throat?

Example

Have you got Mexican Flu?



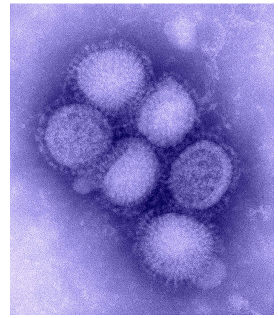
$$
\begin{aligned}
P(m,c,s) &= 0.009215\\
P(m,\bar{c},s) &= 0.000485\\
P(m,c,\bar{s}) &= 0.000285\\
P(m,\bar{c},\bar{s}) &= 1.5\cdot 10^{-5}\\
P(\bar{m},c,s) &= 9.9\cdot 10^{-6}\\
P(\bar{m},\bar{c},s) &= 0.0098901\\
P(\bar{m},c,\bar{s}) &= 0.0009801\\
P(\bar{m},\bar{c},\bar{s}) &= 0.97912
\end{aligned}
$$

- M: mexican flu; C: chills; S: sore throat
- Probability of mexican flu and sore throat?

$$
\begin{aligned}
P(m,s) &= P(m,c,s) + P(m,\bar{c},s)\\
&= 0.009215 + 0.000485\\
&= 0.0097
\end{aligned}
$$

Example

Have you got Mexican Flu?



$$
\begin{aligned}
P(m, c, s) &= 0.009215 \\
P(m, \bar{c}, s) &= 0.000485 \\
P(m, c, \bar{s}) &= 0.000285 \\
P(m, \bar{c}, \bar{s}) &= 1.5 \cdot 10^{-5} \\
P(\bar{m}, c, s) &= 9.9 \cdot 10^{-6} \\
P(\bar{m}, \bar{c}, s) &= 0.0098901 \\
P(\bar{m}, c, \bar{s}) &= 0.0009801 \\
P(\bar{m}, \bar{c}, \bar{s}) &= 0.97912
\end{aligned}
$$

- M: mexican flu; C: chills; S: sore throat
- Probability of mexican flu given sore throat?

Conditioning

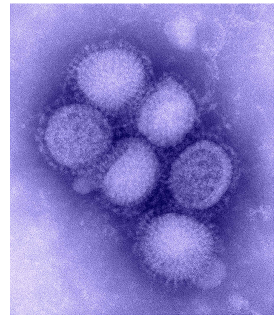
- Conditioning specifies how to revise beliefs based on new information.
- The conditional probability of a A given B is

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

  where $P(B) = \sum_a P(a, B)$.
- Information B rules out possible worlds incompatible with B and induces a new measure over possible worlds in which B holds
- Often, B is available evidence and A is a hypothesis of interest (e.g., disease given symptoms)

Example

Have you got Mexican Flu?

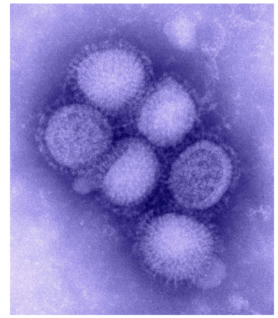


$$
\begin{aligned}
P(m, c, s) &= 0.009215 \\
P(m, \bar{c}, s) &= 0.000485 \\
P(m, c, \bar{s}) &= 0.000285 \\
P(m, \bar{c}, \bar{s}) &= 1.5 \cdot 10^{-5} \\
P(\bar{m}, c, s) &= 9.9 \cdot 10^{-6} \\
P(\bar{m}, \bar{c}, s) &= 0.0098901 \\
P(\bar{m}, c, \bar{s}) &= 0.0009801 \\
P(\bar{m}, \bar{c}, \bar{s}) &= 0.97912
\end{aligned}
$$

- M: mexican flu; C: chills; S: sore throat
- Probability of mexican flu given sore throat?

Example

Have you got Mexican Flu?



- M: mexican flu; C: chills; S: sore throat
- Probability of mexican flu given sore throat?

$$
\begin{aligned}
P(m,c,s) &= 0.009215\\
P(m,\bar{c},s) &= 0.000485\\
P(m,c,\bar{s}) &= 0.000285\\
P(m,\bar{c},\bar{s}) &= 1.5\cdot 10^{-5}\\
P(\bar{m},c,s) &= 9.9\cdot 10^{-6}\\
P(\bar{m},\bar{c},s) &= 0.0098901\\
P(\bar{m},c,\bar{s}) &= 0.0009801\\
P(\bar{m},\bar{c},\bar{s}) &= 0.97912
\end{aligned}
$$

$$
\begin{aligned}
P(m \mid s) &= P(m,s)/P(s)\\
&= 0.0097/0.0196\\
&= 0.495
\end{aligned}
$$

Product rule

- Conditional probability: $P(A|B) = \frac{P(A,B)}{P(B)}$
- Therefore: $P(A, B) = P(A|B)P(B)$

Chain rule

- Extension of the product rule:

$$
\begin{aligned}
&P(X_1, X_2, \ldots, X_n) \\
&= P(X_n \mid X_1, X_2, \ldots, X_{n-1}) \times P(X_1, X_2, \ldots, X_{n-1}) \\
&= P(X_n \mid X_1, X_2, \ldots, X_{n-1}) \times \\
&\qquad P(X_{n-1} \mid X_1, X_2, \ldots, X_{n-2}) \times P(X_1, X_2, \ldots, X_{n-2}) \\
&= P(X_n \mid X_1, X_2, \ldots, X_{n-1}) \times P(X_{n-1} \mid X_1, X_2, \ldots, X_{n-2}) \times \\
&\qquad \cdots \times P(X_3 \mid X_1, X_2) \times P(X_2 \mid X_1) \times P(X_1) \\
&= \prod_{i=1}^{n} P(X_i \mid X_1, \cdots, X_{i-1})
\end{aligned}
$$

Bayes' rule

- The chain rule and commutativity of conjunction ($P(A, B)$ is equivalent to $P(B, A)$) gives us:

$$P(A, B) = P(A \mid B) \times P(B) = P(B \mid A) \times P(A).$$

- If $P(B) \neq 0$, you can divide the right hand sides by $P(B)$:

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

- This is Bayes' rule.

Bayes' rule

- Why is Bayes' rule interesting?
- Often you have causal knowledge:

$$P(\text{symptom} \mid \text{disease}), P(\text{disease})$$
$$P(\text{alarm} \mid \text{fire}), P(\text{fire})$$
$$P(\text{image} \mid \text{a tree is in front of a car}), P(\text{a tree is in front of a car})$$

and want to do evidential reasoning:

$$P(\text{disease} \mid \text{symptom})$$
$$P(\text{fire} \mid \text{alarm})$$
$$P(\text{a tree is in front of a car} \mid \text{image})$$

- Reasoning 'against the direction of the arrows' is not possible using e.g. certainty factors.

Bayes' rule in practice

- A drug test is 99% sensitive (the test returns a positive result for a user 99% of the time)
- A drug test is 99% specific (the test returns a negative result for a non-user 99% of the time)
- Suppose that 0.5% of people are users of the drug
- If an individual tests positive, what is the probability they are a user?

Bayes' rule in practice

- d = drug user, p = positive test, $P(p \mid d) = 0.99$
- $P(\neg p \mid \neg d) = 0.99$, $p(d) = 0.005$.

$$
\begin{aligned}
P(d \mid p) &= \frac{P(p \mid d)P(d)}{P(p)} \\
&= \frac{P(p \mid d)P(d)}{P(p \mid d)P(d) + P(p \mid \neg d)p(\neg d)} \\
&= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.01 \cdot 0.995} \\
&= 33.2\%
\end{aligned}
$$

Independence

- Random variable X is independent of random variable Y if for all x and y

$$P(x \mid y) = P(x)$$

- This is written as $X \perp\!\!\!\perp Y$
- Examples:
  - Flu $\perp\!\!\!\perp$ Haircolor since $P(\text{Flu} \mid \text{Haircolor}) = P(\text{Flu})$.
  - Myalgia $\not\!\perp\!\!\!\perp$ Fever since $P(\text{Myalgia} \mid \text{Fever}) \neq P(\text{Myalgia})$.

Independence

- Independence is very powerful because it allows us to reason about aspects of a system in isolation.
- However, it does not often occur in complex systems. For example, try and think of two medical symptoms that are independent.
- A generalization of independence is conditional independence, where two aspects of a system become independent once we observe a third aspect.
- Conditional independence does often arise and can lead to significant representational and computational savings.

Conditional independence

- Random variable X is conditionally independent of random variable Y given random variable Z if

$$P(x \mid y, z) = P(x \mid z)$$

whenever $P(y, z) > 0$. That is, knowledge of Y doesn't affect your belief in the value of X, given a value of Z.

- This is written as $X \perp\!\!\!\perp Y \mid Z$
- Example:
  - Symptoms are conditionally independent given the disease:

$$\text{Myalgia} \perp\!\!\!\perp \text{Fever} \mid \text{Flu}$$

    since $P(\text{Myalgia} \mid \text{Fever}, \text{Flu}) = P(\text{Myalgia} \mid \text{Flu})$
  - Requires specification of two instead of four probabilities

Conditional independence

- An intuitive test of conditional independence (Paskin):

  Imagine that you know the value of Z and you are trying to guess the value of X. In your pocket is an envelope containing the value of Y . Would opening the envelope help you guess X? If not, then $X \perp\!\!\!\perp Y \mid Z$.

Late adoption

- Belief that the assignment of probabilities to events requires information that is not normally available (McCarthy and Hayes, 1969)
- No way to explicitly represent (medical) knowledge
- No way to have the system explain it's line of reasoning
- Failure of early systems based on probability theory (unrealistic independence assumptions versus intractability)

Breakthroughs for probability theory

- Bayesian interpretation of probability theory:
  - Probability is an agent's measure of belief in some proposition:
  - Probability of an event is determined not only by observed data but also by prior knowledge
  - Other agents may have different prior knowledge, as they may have had different experiences regarding a particular proposition.
- Explicit ways to represent probabilistic knowledge (Bayesian networks, influence diagrams)
- Algorithmic breakthroughs in probabilistic reasoning (Pearl, 1988)