Motivation for probabilistic reasoning

- Modeling of uncertainty
  – Inherent randomness (e.g., radioactive decay)
  – Gross statistical dependencies of complex deterministic world (e.g., coin toss)

- Probability as guardian of commonsense reasoning

- Many empirical successes: robotics, language, speech, bioinformatics, etc.

Review

- Discrete random variable X
  Domain of possible values $\{x_1, x_2, \ldots, x_m\}$
  Ex: month M, $\{m_1 = \text{January}, m_2 = \text{February}, \ldots, m_{12} = \text{December}\}$

- Unconditional (prior) probability $P(X = x_i)$

- Basic axioms:

  (i) $P(X = x_i) \geq 0$

  (ii) $\sum_i P(X = x_i) = 1$

  (iii) $P(X = x_i \text{ or } X = x_j) = P(X = x_i) + P(X = x_j)$ if $x_i \neq x_j$
  Probabilities add for the union of mutually exclusive events.

- Conditional (or posterior) probabilities
  $P(X = x_i | Y = y_j)$ probability that $X = x_i$ given that $Y = y_j$
  In general, $P(X = x_i | Y = y_j) \neq P(X = x_i)$.

- Dependent random variables
  Ex: weather W, $\{w_1 = \text{sunny}, w_2 = \text{rainy}\}$

$$\begin{aligned} P(W = \text{sunny}) &= 0.9 \\ P(W = \text{sunny} | M = \text{August}) &= 0.97 \\ P(W = \text{sunny} | M = \text{January}) &= 0.83 \end{aligned}$$

- Independent random variables
  Ex: day of week D, $\{d_1 = \text{Sun}, d_2 = \text{Mon}, \ldots, d_7 = \text{Sat}\}$

$$\begin{aligned} P(W = \text{sunny}) &= 0.9 \\ P(W = \text{sunny} | d = \text{Sun}) &= 0.9 \\ P(W = \text{sunny} | d = \text{Mon}) &= 0.9 \\ &\vdots \\ P(W = \text{sunny} | d = \text{Sat}) &= 0.9 \end{aligned}$$

- Conditionally independent random variables
  Ex: Binary random variables
  R Did Robert ace the test?
  S Did Samantha ace the test?
  T Was the test very easy?

$$\begin{aligned} P(R=1) &< P(R=1|S=1) \\ P(R=1|T=1) &= P(R=1|T=1,S=1) \end{aligned}$$

  Here R and S are *not* independent random variables, but they are *conditionally* independent given T.

- Conditionally dependent random variables
  Ex: Binary random variables
  B Was there a burglary?
  E Was there an earthquake?
  A Did the alarm go off?

$$\begin{aligned} P(B=1) &= P(B=1|E=1) = P(B=1|E=0) \\ P(B=1|A=1) &> P(B=1|E=1,A=1) \end{aligned}$$

  Here B and E are independent random variables, but they are *conditionally* dependent given A.

- Same axioms hold for conditional probabilities:
  (i) $P(X=x_iY=y_j) \geq 0$
  (ii) $\sum_i P(X=x_i|Y=y_j) = 1$ *Note:* this sum is over i, not j!

- Joint probabilities
  $P(X=x_i, Y=y_j)$ probability that $X=x_i$ and $Y=y_j$

- Product rule

$$\text{For all } i,j: \quad \begin{aligned} P(X=x_i,Y=y_j) &= P(X=x_i|Y=y_j)P(Y=y_j) \\ P(X=x_i,Y=y_j) &= P(Y=y_j|X=x_i)P(X=x_i) \end{aligned}$$

- Marginalization

$$\begin{aligned} P(X=x_i) &= \sum_j P(X=x_i,Y=y_j) \\ P(X=x_i,Y=y_j) &= \sum_k P(X=x_i,Y=y_j,Z=z_k) \end{aligned}$$

- Assessing probabilities
  It is generally easier for experts to assess conditional probabilities (e.g., the chances of single outcomes, conditioned on one or more potentially informative events) than joint probabilities (e.g., the chances of multiple simultaneous outcomes in the absence of potentially relevant context).

- Shorthand notations
  (i) Implied universality
  $P(X,Y) = P(X|Y)P(Y) = P(Y|X)P(X)$
  Implies that equality holds for all assignments $X=x_i$ and $Y=y_j$
  (ii) Implied assignment
  $P(x,y,z) = P(X=x,Y=y,Z=z)$

- Generalized product rule
  $P(A, B, C, D, \ldots) = P(A)P(B|A)P(C|A, B)P(D|A, B, C)\ldots$

- Bayes rule
  Equating the two expressions of the product rule for $P(X, Y)$, we can write:

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

- Bayes rule (conditioned on an additional event Z)

$$P(X|Y, Z) \;=\; \frac{P(Y|X, Z)P(X|Z)}{P(Y|Z)}$$

  Compare to the equation above, and notice how the additional event Z appears in every term on the right hand side of the conditioning bar.

Alarm example

- Binary random variables
  B Was there a burglary?
  E Was there an earthquake?
  A Did the alarm go off?

- Joint distribution
  $P(B, E, A) = P(B)P(E)P(A|B, E)$

- Prior knowledge
  Burglaries are rare: $P(B = 1) = 0.001$
  Earthquakes are rare: $P(E = 1|B = 0) = P(E = 1|B = 1) = 0.002$
  Burglaries and earthquakes are independent events: $P(E|B) = P(E)$.
  The alarm is likelier to sound from a burglary than an earthquake:

B	E	$P(A = 1\|B, E)$
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

  Probabilities of complementary events are easy to compute:

$$\begin{aligned} P(B = 0) &= 1 - P(B = 1) \;=\; 0.999 \\ P(E = 0) &= 1 - P(E = 1) \;=\; 0.998 \\ P(A = 0|B, E) &= 1 - P(A = 1|B, E) \end{aligned}$$

- Inference
  Do the rules of probability capture commonsense patterns of reasoning?
  Let's compare $P(B = 1)$, $P(B = 1|A = 1)$, and $P(B = 1|A = 1, E = 1)$.
  We are given $P(B = 1) = 0.001$. How to compute the others?

- Bayes rule

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

In the numerator:

$$\begin{aligned}
P(A=1|B=1) &= \sum_{e\in\{0,1\}} P(A=1, E=e|B=1) \quad \text{(marginalization)} \\
&= \sum_{e\in\{0,1\}} P(A=1|E=e, B=1)P(E=e|B=1) \quad \text{(product rule)} \\
&= \sum_{e\in\{0,1\}} P(A=1|E=e, B=1)P(E=e) \quad \text{(independence)} \\
&= P(A=1|E=0, B=1)P(E=0) + P(A=1|E=1, B=1)P(E=1) \\
&= (0.94)(1-0.002) + (0.95)(0.002) \\
&= 0.94002
\end{aligned}$$

In the denominator:

$$\begin{aligned}
P(A=1) &= \sum_{e,b\in\{0,1\}} P(A=1, E=e, B=b) \quad \text{(marginalization)} \\
&= \sum_{e,b} P(A=1|E=e, B=b)P(E=e|B=b)P(B=b) \quad \text{(product rule)} \\
&= \sum_{e,b} P(A=1|E=e, B=b)P(E=e)P(B=b) \quad \text{(independence)} \\
&= P(A=1|E=0, B=0)P(E=0)P(B=0) + \text{ (three other terms)} \\
&= 0.00252
\end{aligned}$$

Substituting into Bayes rule:

$$P(B=1|A=1) = \frac{P(A=1|B=1)P(B=1)}{P(A=1)} = \frac{(0.94002)(0.001)}{(0.00252)} = 0.37$$

So far this agrees with commonsense: $P(B=1) = 0.001$ and $P(B=1|A=1) = 0.37$.
And indeed, we are much more likely to think that a burglary occurred given that the alarm sounded.

- Bayes rule (again)

$$P(B=1|A=1, E=1) = \frac{P(A=1|B=1, E=1)P(B=1|E=1)}{P(A=1|E=1)}$$

Compare this to the previous invocation of Bayes rule; in particular, notice how we have simply added the event $E=1$ to the right hand side of every conditioning bar in the equation.

In the denominator:

$$P(A=1|E=1) = \frac{P(A=1, E=1)}{P(E=1)}$$

The only unknown probability that we need to compute is $P(A=1, E=1)$. The calculation is very similar to earlier ones:

$$\begin{aligned}
P(A=1,E=1) &= \sum_b P(A=1,E=1,B=b) \quad \text{(marginalization)} \\
&= \sum_b P(A=1|E=1,B=b)P(E=1|B=b)P(B=b) \quad \text{(product rule)} \\
&= \sum_b P(A=1|E=1,B=b)P(E=1)P(B=b) \quad \text{(independence)} \\
&= P(A=1|E=1,B=0)P(E=1)P(B=0)\ + \\
&\quad\ P(A=1|E=1,B=1)P(E=1)P(B=1) \\
&= (0.29)(0.002)(1-0.001)+(0.95)(0.002)(0.001) \\
&= 0.00058
\end{aligned}$$

Substituting into Bayes rule again:

$$\begin{aligned}
P(B=1|A=1,E=1) &= \frac{P(A=1|B=1,E=1)P(B=1|E=1)}{P(A=1|E=1)} \\
&= \frac{P(A=1|B=1,E=1)P(B=1)}{P(A=1|E=1)} \quad \text{(independence)} \\
&= \frac{P(A=1|B=1,E=1)P(B=1)}{\frac{P(A=1,E=1)}{P(E=1}} \quad \text{(product rule)} \\
&= \frac{(0.95)(0.001)}{\left(\frac{0.00058}{0.002}\right)} \\
&= 0.0033
\end{aligned}$$

Again this agrees with commonsense patterns of reasoning. In particular, here is what we have computed and how it accords with common sense:

$$\begin{aligned}
P(B=1) &= 0.001 \quad \text{“Burglaries are rare. No need to worry.”} \\
P(B=1|A=1) &= 0.37 \quad \text{“Oh no! The alarm may have been triggered by a burglar.”} \\
P(B=1|A=1,E=1) &= 0.0033 \quad \text{“Nah, probably the alarm was triggered by the earthquake.”}
\end{aligned}$$

This is an example of *non-monotonic reasoning*: after learning that the alarm has sounded, we are *more* afraid that a burglary has occurred, but after learning (in addition) that an earthquake has occurred, we are *less* afraid.

This particular pattern of reasoning is known as *explaining away*. Note how the earthquake explains away the alarm, thus diminishing our fear of a burglary.

These probabilities also reveal an example of *conditional dependence*. Burglaries and earthquakes are independent events, but this independence does not hold when we condition on the sounding of the alarm. In particular:

$$\begin{aligned}
P(B) &= P(B|E) \quad \text{(independence)} \\
P(B|A) &\neq P(B|A,E) \quad \text{(conditional dependence)}
\end{aligned}$$

You should start to see patterns in these calculations as well. Notice how Bayes rule is invoked to infer causes from effects. Also notice the frequent consecutive steps involving marginalization, the product rule, and the appeal to independence. This suggests that many of these computations might be efficiently automated. (To be continued ...)