Introduction to Bayesian Inference

Will Penny

Short Course on Bayesian Inference, Virginia Tech, 24th January 2012
Bayes rule

Given marginal probabilities \( p(A) \), \( p(B) \), and the joint probability \( p(A, B) \), we can write the conditional probabilities

\[
\begin{align*}
    p(B|A) &= \frac{p(A, B)}{p(A)} \\
    p(A|B) &= \frac{p(A, B)}{p(B)}
\end{align*}
\]

This is known as the product rule. Eliminating \( p(A, B) \) gives Bayes rule

\[
p(B|A) = \frac{p(A|B)p(B)}{p(A)}
\]
The terms in Bayes rule

\[ p(B|A) = \frac{p(A|B)p(B)}{p(A)} \]

are referred to as the prior, \( p(B) \), the likelihood, \( p(A|B) \), and the posterior, \( p(B|A) \).

The probability \( p(A) \) is a normalisation term and can be found by marginalisation. For example,

\[
p(A = 1) = \sum_B p(A = 1, B) = p(A = 1, B = 0) + p(A = 1, B = 1) = p(A = 1|B = 0)p(B = 0) + p(A = 1|B = 1)p(B = 1)
\]

This is known as the sum rule.
Bayes rule

We can also write Bayes rule as

\[ p(B|A) = \frac{p(A|B)p(B)}{\sum_{B'} p(A|B')p(B')} \]

This makes use of the sum and product rules.

Bayes rule is the extension of Boolean logic to uncertain events.
Medical Decision Making


They commonly occur in arteries at the base of the brain.
Sensitivity and Specificity

Given patient 1’s symptoms, the prior probability of A (prior to MRA) is believed to be 90%.

For As bigger than 6mm MRA has a sensitivity and specificity of 95% and 92%.

What then is the probability of A given a negative test result, $T$?
Medical Decision Making
The clinician believes the probability of aneurysm prior to the MRA test to be

\[ p(A = 1) = 0.9 \]

MRA test sensitivity and specificity are

\[ p(T = 1|A = 1) = 0.95 \]
\[ p(T = 0|A = 0) = 0.92 \]

The false negative rate is therefore

\[ p(T = 0|A = 1) = 1 - p(T = 1|A = 1) = 0.08 \]

The probability of A given a negative test can be found from Bayes rule

\[ p(A = 1|T = 0) = \frac{p(T = 0|A = 1)p(A = 1)}{p(T = 0|A = 1)p(A = 1) + p(T = 0|A = 0)p(A = 0)} \]

This is the proportion of false negatives to false negatives plus true negatives.
Joint Probability

A prior of 0.9 means that of 1000 people that present to the clinician with the same symptoms he believes that 900 of them will have an aneurysm.

<table>
<thead>
<tr>
<th></th>
<th>$T = 0$</th>
<th>$T = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A = 0$</td>
<td>92</td>
<td>8</td>
</tr>
<tr>
<td>$A = 1$</td>
<td>45</td>
<td>855</td>
</tr>
<tr>
<td></td>
<td>137</td>
<td>863</td>
</tr>
</tbody>
</table>

The clinician’s belief that a patient has an aneurysm after a negative test is $45/137=0.3285$.

The inner table above is the joint probability $p(A, T)$ (if we divide by 1000).
### Medical Decision Making

#### Negative test result

<table>
<thead>
<tr>
<th>Prior (clinical) probability</th>
<th>0.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posterior probability</td>
<td>(\frac{(1 - \text{sensitivity}) \times \text{prior probability}}{(1 - \text{sensitivity}) \times \text{prior probability} + \text{specificity} \times (1 - \text{prior probability})})</td>
</tr>
<tr>
<td></td>
<td>(\frac{(1 - 0.95) \times 0.90}{(1 - 0.95) \times 0.90 + 0.92 \times (1 - 0.90)})</td>
</tr>
<tr>
<td>Posterior probability</td>
<td>0.3285</td>
</tr>
</tbody>
</table>

#### Positive test result

<table>
<thead>
<tr>
<th>Prior (clinical) probability</th>
<th>0.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posterior probability</td>
<td>(\frac{\text{sensitivity} \times \text{prior probability}}{(\text{sensitivity} \times \text{prior probability}) + (1 - \text{specificity}) \times (1 - \text{prior probability})})</td>
</tr>
<tr>
<td></td>
<td>(\frac{0.95 \times 0.90}{(0.95 \times 0.90) + (1 - 0.92) \times (1 - 0.90)})</td>
</tr>
<tr>
<td>Posterior probability</td>
<td>0.9907</td>
</tr>
</tbody>
</table>

**Fig 3** Probability of a posterior communicating artery aneurysm given a negative or positive result from magnetic resonance angiography and a prior clinical probability of 90%. Sensitivity and specificity of angiography are 95% and 92% respectively. Probabilities are expressed between 0.0 (0%) and 1.0 (100%).
Medical Decision Making

![Graph showing influence of prior clinical probability on the probability of a disease after a negative or positive test result. Test sensitivity and specificity are 95% and 92% respectively.]

A negative MRA cannot therefore be used to exclude a diagnosis of A in this case.
Odds Ratios

If \( p \) is the probability of an event then the odds \( R \) of that event are

\[
R = \frac{p}{1 - p}
\]

\( R \) is also referred to as an Odds Ratio.

Conversely,

\[
p = \frac{R}{R + 1}
\]
Bayes rule can be usefully expressed in the form of odds ratios. Considering first a positive test result, the posterior odds that the subject has an aneurysm are given by

\[
\frac{p(A = 1|T = 1)}{p(A = 0|T = 1)} = \frac{p(T = 1|A = 1) \cdot p(A = 1)}{p(T = 1|A = 0) \cdot p(A = 0)}
\]

where the prior odds are

\[
\frac{p(A = 1)}{p(A = 0)} = 9
\]

and the likelihood ratio is

\[
\frac{p(T = 1|A = 1)}{p(T = 1|A = 0)} = \frac{sens}{1 - spec} = 11.88
\]

The posterior odds is therefore \(11.88 \times 9 = 106.88\).
Odds Ratios

For a negative test result we have

\[
p(A = 1|T = 0) \quad \frac{p(T = 0|A = 1) \ p(A = 1)}{p(A = 0|T = 0)} = \frac{p(T = 0|A = 0) \ p(A = 0)}
\]

Here the likelihood ratio is \((1 - \text{sens})/\text{spec} = 0.054\), so the posterior odds are \(0.054 \times 9 = 0.49\).

The posterior probability of an aneurysm given positive and negative test results are given by \(p = R/(R + 1)\) which are \(0.9907\) and \(0.3285\). These are, of course, the same as before.
Multiple Causes and Observations

Multiple potential causes (eg. $x_1$, $x_2$) and observations ($x_3$, $x_4$ eg. headache, oculomotor palsy, double vision, drooping eye lids, blood in CSF)

![Diagram](image)
Generative Models
For a probabilistic generative model

The joint probability of all variables, $x$, can be written down as

$$p(x) = \prod_{i=1}^{5} p(x_i | pa[x_i])$$

where $pa[x_i]$ are the parents of $x_i$. If there are no cycles we have a Direct Acyclic Graph (DAG), also known as a Bayesian network (Jensen, 2000; Pearl, 1988).
Joint Probability

A DAG specifies the joint probability of all variables.

\[ p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2)p(x_3|x_1)p(x_4|x_1, x_2)p(x_5|x_4) \]

See Chapter 8 in Bishop (2006) for more examples. All other variables can be gotten from the joint probability via marginalisation.
Marginalisation

\[ p(x_1) = \int p(x_1, x_2) \, dx_2 \]

\[ p(x_1, x_2) \]

\[ x_2 \]

\[ x_1 \]

\[ p(x_1) \]

\[ x_1 \]
Marginalisation

\[ p(x_1, x_2) = \int \int \int p(x_1, x_2, x_3, x_4, x_5) \, dx_3 \, dx_4 \, dx_5 \]

\[ p(x_4) = \int \int \int \int p(x_1, x_2, x_3, x_4, x_5) \, dx_1 \, dx_2 \, dx_3 \, dx_5 \]

\[ 1 = \int \int \int \int \int p(x_1, x_2, x_3, x_4, x_5) \, dx_1 \, dx_2 \, dx_3 \, dx_4 \, dx_5 \]

\[ p(x_1) = \sum_{x_2} p(x_1, x_2) \]

\[ p(x_2 = 3, x_3 = 4) = \sum_{x_1} p(x_1, x_2 = 3, x_3 = 4) \]
Generative Models

If \( x_5 \) is observed and we want to know \( x_3 \) then

\[
p(x_3 | x_5) = \frac{p(x_3, x_5)}{p(x_5)}
\]

Necessary probabilities obtained via marginalisation. This can be implemented efficiently using local computations via 'belief propagation'.
Did I Leave The Sprinkler On?

A single observation with multiple potential causes (not mutually exclusive). Both rain, $r$, and the sprinkler, $s$, can cause my lawn to be wet, $w$.

$$p(w, r, s) = p(r)p(s)p(w| r, s)$$
Did I Leave The Sprinkler On?

The probability that the sprinkler was on given I’ve seen the lawn is wet is given by Bayes rule

\[
p(s = 1 | w = 1) = \frac{p(w = 1 | s = 1)p(s = 1)}{p(w = 1)} = \frac{p(w = 1, s = 1)}{p(w = 1, s = 1) + p(w = 1, s = 0)}
\]

where the joint probabilities are obtained from marginalisation

\[
p(w = 1, s = 1) = \sum_{r=0}^{1} p(w = 1, r, s = 1)
\]

\[
p(w = 1, s = 0) = \sum_{r=0}^{1} p(w = 1, r, s = 0)
\]

and from the generative model we have

\[
p(w, r, s) = p(r)p(s)p(w | r, s)
\]
Look next door

Rain $r$ will make my lawn wet $w_1$ and nextdoors $w_2$ whereas the sprinkler $s$ only affects mine.

\[ p(w_1, w_2, r, s) = p(r)p(s)p(w_1 | r, s)p(w_2 | r) \]
After looking next door

Use Bayes rule again

\[
p(s = 1|w_1 = 1, w_2 = 1) = \frac{p(w_1 = 1, w_2 = 1, s = 1)}{p(w_1 = 1, w_2 = 1, s = 1) + p(w_1 = 1, w_2 = 1, s = 0)}
\]

with joint probabilities from marginalisation

\[
p(w_1 = 1, w_2 = 1, s = 1) = \sum_{r=0}^{1} p(w_1 = 1, w_2 = 1, r, s = 1)
\]

\[
p(w_1 = 1, w_2 = 1, s = 0) = \sum_{r=0}^{1} p(w_1 = 1, w_2 = 1, r, s = 0)
\]
Numerical Example

Bayesian models force us to be explicit about exactly what it is we believe.

\[ p(r = 1) = 0.01 \]
\[ p(s = 1) = 0.02 \]
\[ p(w = 1 | r = 0, s = 0) = 0.001 \]
\[ p(w = 1 | r = 0, s = 1) = 0.97 \]
\[ p(w = 1 | r = 1, s = 0) = 0.90 \]
\[ p(w = 1 | r = 1, s = 1) = 0.99 \]

These numbers give

\[ p(s = 1 | w = 1) = 0.67 \]
\[ p(r = 1 | w = 1) = 0.31 \]
Explaining Away

Numbers same as before. In addition

\[ p(w_2 = 1 \mid r = 1) = 0.90 \]

Now we have

\[ p(s = 1 \mid w_1 = 1, w_2 = 1) = 0.21 \]
\[ p(r = 1 \mid w_1 = 1, w_2 = 1) = 0.80 \]

The fact that my grass is wet has been explained away by the rain (and the observation of my neighbours wet lawn).
The CHILD network

Proabilistic graphical model for newborn babies with congenital heart disease.

Decision making aid piloted at Great Ormond Street hospital (Spiegelhalter et al. 1993).
References


