Bayesian Reasoning

6.873/HST951
Peter Szolovits

Simplest Example

- Relationship between a diagnostic conclusion and a diagnostic test

<table>
<thead>
<tr>
<th></th>
<th>Disease Present</th>
<th>Disease Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Positive</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Test Negative</td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

FP+TN          
True Negative
False Positive
Disease Absent

FN+TN          
False Negative
True Positive
Disease Present

Definitions

<table>
<thead>
<tr>
<th></th>
<th>Disease Present</th>
<th>Disease Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Positive</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Test Negative</td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

TP+FN          
True Positive
False Negative
Disease Present

FP+TN          
False Positive
True Negative
Disease Absent

Sensitivity (true positive rate): \( \frac{TP}{TP+FN} \)
False negative rate: \( 1-\text{Sensitivity} = \frac{FN}{TP+FN} \)
Specificity (true negative rate): \( \frac{TN}{FP+TN} \)
False positive rate: \( 1-\text{Specificity} = \frac{FP}{FP+TN} \)
Positive Predictive Value: \( \frac{TP}{TP+FP} \)
Negative Predictive Value: \( \frac{TN}{FN+TN} \)

Test Thresholds

Test Thresholds Change Trade-off between Sensitivity and Specificity

Wonderful Test
**Receiver Operator Characteristic (ROC) Curve**

**What makes a better test?**

![ROC Curve Diagram]

**How certain are we after a test?**

**Rationality**

- Behavior is a continued sequence of choices, interspersed by the world’s responses
- Best action is to make the choice with the greatest expected value
- … decision analysis

**Example: Acute Renal Failure**

- Choice of a handful (8) of therapies (antibiotics, steroids, surgery, etc.)
- Choice of a handful (3) of invasive tests (biopsies, IVP, etc.)
- Choice of 27 diagnostic “questions” (patient characteristics, history, lab values, etc.)
- Underlying cause is one of 14 diseases
  - We assume one and only one disease

**Entropy of a distribution**

\[ H(P_1, \ldots, P_n) = \sum_{j=1}^{n} P_j \log_2 P_j \]

For example:
- \( H(0.5, 0.5) = 1.0 \)
- \( H(0.1, 0.9) = 0.47 \)
- \( H(0.01, 0.99) = 0.08 \)
- \( H(0.001, 0.999) = 0.01 \)

\( H(0.33, 0.33, 0.33) = 1.58 (?) \)
- \( H(0.005, 0.455, 0.5) = 1.04 \)
- \( H(0.005, 0.995, 0.0) = 0.045 \)

(!) should use \( \log_2 \)
Interacting with ARF in 1973

Question 1: What is the patient's age?
1. 0-10
2. 11-20
3. 21-30
4. 31-50
5. 51-70
6. Over 70

Reply: 5

The current distribution is:
Disease       Probability
FARF         0.58
IBSTR        0.22
ATN          0.09

Question 2: What is the patient's sex?
1. Male
2. Pregnant Female
3. Non-pregnant Female

Reply: 1

Local Sensitivity Analysis

Case-specific Likelihood Ratios

DeDombal, et al. Experience 1970’s & 80’s

- “Idiot Bayes” for appendicitis
  1. Based on expert estimates -- lousy
  2. Statistics -- better than docs
  3. Different hospital -- lousy again
  4. Retrained on local statistics -- good
Probabilistic Models

• What to represent?
  – Disease
  – Finding (signs, symptoms, labs, radiology, …)
  – Syndromes
  – History, predisposing conditions
  – Treatments
    • modify disease, cause new symptoms, …
  – (Outcomes, preferences, …)

State Space

• Set of random variables
  • Possible values of each
  • Assignment of probability to every possible combination of values of all variables
  • \( p(v_1=a_1, v_2=a_2, v_3=a_3, \ldots) \)

Questions of Interest

• Given a set of values of certain variables, what is the probability that certain other variables have certain other values?
• E.g., \( p(v_1=a_1, v_7=a_7|v_2=a_2, v_4=a_4) \)
  \( =\frac{p(v_1=a_1, v_7=a_7, v_2=a_2, v_4=a_4)}{p(v_2=a_2, v_4=a_4)} \)
• We don’t care about all other variables
  – marginalize; i.e., sum over them all

Computational Cost

• For \( n \) binary variables, we need probability assignments to \( 2^n \) states.
• In programs such as DXPLAIN, \( n \) is on the order of thousands.
• Need to be very careful and clever
  – simple models
  – approximate solution techniques

Independence

• Two random variables are independent iff \( p(A&B)=p(A)p(B) \)
• Usually, however, variables may depend on others, but we are still interested whether they have a conditional dependence
• Two random variables are conditionally independent if for a conditioning variable \( D \), \( p(A&B|D)=p(A|D)p(B|D) \)

Independence was crucial to ARF

• Diseases were dependent; mutually exhaustive and exclusive.
• Questions were conditionally independent, given disease.
ARF model convenient

- Odds: $O(D) = P(D)/P(\neg D) = P(D)/(1-P(D))$
- Likelihood ratio: $L(S|D) = P(S|D)/P(S|\neg D)$
- Bayes: $O(D|S) = O(D)L(S|D)$
- Multiple evidence:
  - $O(D|S_1&S_2&\ldots) = O(D)L(S_1|D)L(S_2|D)\ldots$
- Log transform:
  - $W(D|S_1&S_2&\ldots) = W(D)+W(S_1|D)+W(S_2|D)+\ldots$

Side comment on likelihood ratio

- $L(S|D) = p(s|d)/p(s|\neg d)$ is constant only if $\neg d$ is a "fixed" entity
- If, as in ARF, we have $d_1, d_2, d_3, \ldots$, then
  $p(s|\neg d_j) = \sum_{i \neq j} p(d_i)p(s|d_i)$
- As probabilities vary over the $d_j$, $p(s|\neg d_j)$ will vary!

What if we made no assumptions in ARF?

- Any combination of diseases: $2^{14}=16K$
- Distinct probability for any combination of answers to any questions:
  $p(q_1=a_12&q_2=a_24&\ldots)$
  $3^27 = 7.7\times 10^{12}$
- $p(q_1=a_12&q_2=a_24&\ldots|d_1&d_2&\neg d_3&\ldots)$
  $2^14 \times 3^27 = 1.25\times 10^{17}$, just for ARF
- Simplification is essential!

Conditional Independence is not Independence

- $P(b\&c|a) = P(a\&b\&c)/P(a)$
- But
  $P(b\&c) = p(a)p(b\&c|a) + p(\neg a)p(b\&c|\neg a)$

Conditional independence is not independence

- Information may still "flow" from one observation to another, even if they are conditionally independent given a disease, unless the disease is known with certainty
- $p(D) = .2$
- $p(A|D) = .8$, $p(A|\neg D) = .1$
- $p(B|D) = .6$, $p(B|\neg D) = .1$
- \textit{a priori}, $p(A) = .16 + .08 = .24$, $p(B) = .12 + .04 = .16$
- $P(D|A) = .67$
- $P(B|A) = .40 + .03 = .43$, not .16!
- But, if $p(D) = 0$ or 1, no effect.

Conditional independence is not independence

- Two variables may be made \textit{conditionally dependent} when we learn about a common descendant
- $p(A) = .2$, $p(B) = .1$
- $p(C|A&B) = .8$, $p(C|A&\neg B) = .4$
- $p(C|\neg A&B) = .6$, $p(C|\neg A&\neg B) = .1$
- $p(C) = .2^*1^*.8+.2^*.9^*.4+.8^*.1^*.6+.8^*.9*1 = .208$
- If we observe C, $p(A&B) = .02^*.8/.208 = .077$, but $p(A) = .42$, $p(B) = .31$. $p(a|c) \neq p(a|c)p(b|c)$
We don’t want to model high-arity dependence

\[ P(C | A_1 \& \neg A_2 \& \neg A_3 \& A_4 \& \ldots) \]

• too many probabilities
• Can we simplify?
  – Noisy or
  – noisy and
  – noisy max/min
  – ?

Noisy or

• \( p(C | A \& B) = p(“C happened because of A”) \)
  * \( p(“C happened because of B”) \)
  * \( p(“C happened anyway”) \)
• \( (1-p(C | A \& B)) = (1-p_c(C | A)) \times (1-p_c(C | B)) \times (1-L) \)
• \( p_c(C | A) \leq p(C | A) \)

Simple Models
(Singly-Connected)

How to propagate likelihood information?
Like likelihood ratios.
(Pearl poly-tree algorithm)

Causality?

• Noisy-or (and, …)
• Bayes arrows are/are not causal
  – reversing an arrow adds new dependencies

General Case Bayes Nets
(Multiply-connected DAGs)

Cooper’s MCBN1

Factoring to Simplify Computation

\[
p(E = t, B = t) = \frac{\sum_{D, C} p(A, B = t, C, D, E = t) p(D | B, C) p(E | C)}{\sum_{A} p(A) p(B = t | A) p(C | A) p(D | B = t, C) p(E = t | C)}
\]

In this simple example, 12 instead of 32 multiplications
How to build Bayes Nets?

- David Heckerman
  Pathfinder/Intellipath, around 1990