Day 7

Outline

1. Two-way tables and diagnostic testing
2. Conditional probability
3. Tree Diagrams
4. Bayes' Rule
5. Solving conditional probability problems

Two-Way Table

Gender compared to handedness

	Handed		
	Left	Right	
Female	7	46	53
Male	5	63	68
	12	109	121

Figure 1: Table Example

Example

Helsinki Heart Study

2035 men in control group had 84 heart attacks 2046 men in special drug name group had 54 heart attacks

Referring to the table above, these are the correct mappings:

- Male: placebo
- Female: special drug
- Left: Heart attack
- Right: No heart attack

Diagnostic Testing

Formal definition: an examination to identify an individual's specific areas of weakness and strength in order determine a condition, disease or illness.

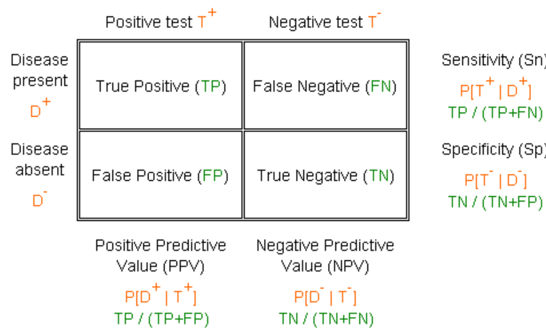


Figure 2: Diagnostic Testing Table

Machine Learning Terminology: we evaluate on a "training set" in which number if actual positive and actual negative is known in advance.

We compute:

- Sensitivity(recall): proportion of actual positive classified correctly $\frac{TP}{TP+FN}$.
- Specificity: proportion of actual negative classified correctly $\frac{TN}{TN+FP}$

These are both properties of our test/algorithm.

- Positive Predictive Value(PPV, precision): proportion of positive tests that were actually positive $\frac{TP}{TP+FP}$
- Negative Predictive Value(NPV): proportion of negative tests that are actually negative $\frac{TN}{TN+FN}$

Also depend on prevalence (base rate) $\frac{Actual\,positive}{Actual\,positive+Actual\,negative}$

Example

- 300 units
- 83% prevalence
- TP = 200
- FP = 10
- FN = 50
- TN = 40

Compute:

- Sensitivity: $\frac{200}{200+50} = 80\%$
- Specificity: $\frac{40}{40+10} = 80\%$
- PPV = $\frac{200}{200+10} = 95.20\%$
- NPV = $\frac{40}{40+50} = 44.4\%$

In Class Example

- 300 units
- 3% prevalence
- TP = 8
- FP = 58
- FN = 2
- TN = 232

Answers

- Sens: $\frac{8}{10} = 80\%$
- Spec: $\frac{232}{232+58} = 80\%$
- PPV: $\frac{8}{8+58} = 12.1\%$
- NPV: $\frac{232}{232+2} = 99.2\%$

Conditional Probability

The conditional probability of event "B" given event "A", denote denoted $P(B|A)$, is the probability of event "B", looking only at outcomes in A.

$P(B|A) = \frac{number\,of\,outcomes\,in\,A \cap B}{number\,of\,outcomes\,in\,A}$ when all outcomes are equally likely

More generally: $P(B|A) = \frac{P(A \cap B)}{P(A)} P(A) > 0$

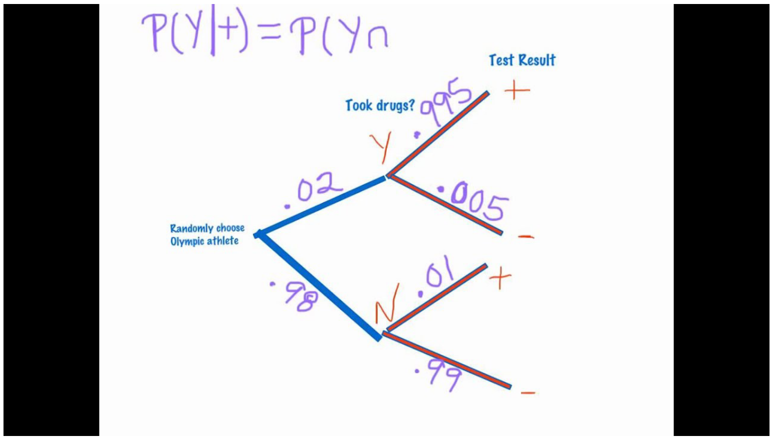


Figure 3: Conditional Probability Diagram

Independent events: $P(A \cap B) = P(A) \times P(B)$

Conditional probability: $P(A \cap B) = P(A) \times P(B|A)$

So: "A" and "B" are independent when $P(B) = P(B|A)$ dependent when $P(B) \neq P(B|A)$

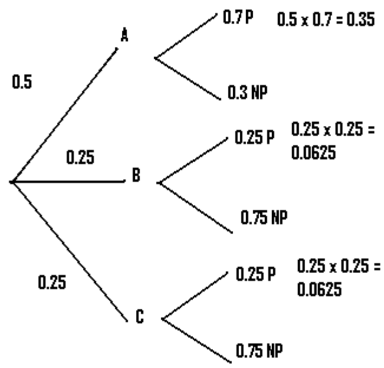
Example

1. What is the probability that a randomly selected person in the control group had a heart attack?
2. What is the probability that a randomly selected heart attack victim was in the control group?

Both of these are different questions

1. $P(heart\,attack|control) = \frac{84}{2035}$
2. $P(control|heart\,attack) = \frac{84}{140}$

Tree Diagram



- $\cdot = node$
- $\overline{\hspace{2em}} = branch$

Each node represents an event.

Each branch represents probability of getting to next node, given that we got to previous node.

Ending node is the **terminal node**.

Splits in the branches must add up to one. Please refer to node "A" when it splits between "NP" and "P".

Notes on Probabilities

1. Probability of leaving a node is 1. The sum of probabilities on all branches exiting a node = 1.
2. Probability of getting to a terminal node is a product of probabilities along the branch path to it.
3. Probability of event "B" is sum of probabilities at all terminal nodes including "B"

Bayes' Rule

"Switch" between $P(A|B)$ and $P(B|A)$

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

$P(A \cap B) = P(B|A) \times P(A)$

By fancy set theory logic:

- $P(B) = P(B \cap A) + P(B \cap A^c)$
- $P(B \cap A) = P(B|A) \times P(A)$
- $P(B \cap A^c) = P(B|A^c) \times P(A^c)$

$P(A|B) = \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A^c) \times P(A^c)}$

Use Bayes' Rule to find $P(A|B)$ given $P(B|A)$, $P(B|A^c)$, and $P(A)$

e.g) Find PPV given sensitivit, 1 - spec & prevalence