Lecture 10: Probability continued…

COSC 480 Data Science, Spring 2017
Michael Hay

Logistics

- Quiz 1
- Quiz 2
- Lab 3

Probability concepts

- Sample space and events
- Conditional probability P(A | B)
- Chain rule: P(A, B) = P(A | B) P(B)
- Independence: if A and B independent, P(A, B) = P(A) P(B)
- Random variables: map events to numbers
- Expected value: "average" value of r.v. Linearity of expected value.
- Total law of probability: $P(A) = P(A, B=b_1) + \ldots P(A, B=b_k)$

Another probability puzzle

- The probability that a woman 40 to 50 years old has breast cancer is 0.8%

- If a woman has breast cancer, the probability is 90% that she will have a positive mammogram

- If a woman does not have breast cancer, the probability is 7% that she will still have a positive mammogram

- Imagine a woman who has a positive mammogram. *What is the probability that she actually has breast cancer?*

http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

What does the doc say?

- Reaction from a department chief at a German university teaching hospital with more than 30 years of experience:

> *"[He] was visibly nervous while trying to figure out what he would tell the woman. After mulling the numbers over, he finally estimated the woman's probability of having breast cancer, given that she has a positive mammogram, to be* ***90 percent****. Nervously, he added, 'Oh, what nonsense. I can't do this. You should test my daughter; she is studying medicine.'"*

http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

Bayes' rule

Bayes' rule This simple but powerful rule follows from the definition of conditional probability and an application of the chain rule.

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

Let $\neg A$ denote the event "A did not occur," we can also write it as

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

A particular useful application is when event A is some hidden cause and event B is some observable symptom. Given the test result is positive (B), what is the probability the patient has the disease (A)?

Another probability puzzle

- The probability that a woman 40 to 50 years old has breast cancer is 0.8%

  P(cancer) = 0.008

- If a woman has breast cancer, the probability is 90% that she will have a positive mammogram

  P(pos|cancer) = 0.9

- If a woman does not have breast cancer, the probability is 7% that she will still have a positive mammogram

  P(pos|¬cancer) = 0.07

- Imagine a woman who has a positive mammogram. *What is the probability that she actually has breast cancer?*

  P(cancer|pos) = ???

http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

As for American doctors

- … 95 out of 100 estimated the woman's probability of having breast cancer to be somewhere around ***75 percent***.

http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University