



Conditional Probability

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

Conditional probability, motivation

- The probability of getting a one when rolling a (standard) die is usually assumed to be one sixth
- Suppose you were given the extra information that the die roll was an odd number (hence 1, 3 or 5)
- *conditional on this new information*, the probability of a one is now one third

Conditional probability, definition

- Let B be an event so that $P(B) > 0$
- Then the conditional probability of an event A given that B has occurred is $P(A | B) = \frac{P(A \cap B)}{P(B)}$
- Notice that if A and B are independent, then $P(A | B) = \frac{P(A) P(B)}{P(B)} = P(A)$

Example

- Consider our die roll example
- $B = \{1, 3, 5\}$
- $A = \{1\}$
$$P(\text{one given that roll is odd}) = P(A \cap B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}$$

Bayes' rule

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

Diagnostic tests

- Let + and - be the events that the result of a diagnostic test is positive or negative respectively
- Let D and D^c be the event that the subject of the test has or does not have the disease respectively
- The **sensitivity** is the probability that the test is positive given that the subject actually has the disease, $P(+ | D)$
- The **specificity** is the probability that the test is negative given that the subject does not have the disease, $P(- | D^c)$

More definitions

- The **positive predictive value** is the probability that the subject has the disease given that the test is positive, $P(D \sim | \sim +)$
- The **negative predictive value** is the probability that the subject does not have the disease given that the test is negative, $P(D^c \sim | \sim -)$
- The **prevalence of the disease** is the marginal probability of disease, $P(D)$

More definitions

- The **diagnostic likelihood ratio of a positive test**, labeled DLR_{+} , is $P(+ | D) / P(+ | \neg D)$, which is the sensitivity / (1 - specificity)
- The **diagnostic likelihood ratio of a negative test**, labeled DLR_{-} , is $P(- | D) / P(- | \neg D)$, which is the (1 - sensitivity) / specificity

Example

- A study comparing the efficacy of HIV tests, reports on an experiment which concluded that HIV antibody tests have a sensitivity of 99.7% and a specificity of 98.5%
- Suppose that a subject, from a population with a .1% prevalence of HIV, receives a positive test result. What is the probability that this subject has HIV?
- Mathematically, we want $P(D \sim +)$ given the sensitivity, $P(+ \sim D) = .997$, the specificity, $P(- \sim \sim D^c) = .985$, and the prevalence $P(D) = .001$

Using Bayes' formula

$$\begin{aligned} P(D | +) &= \frac{P(+ | D)P(D)}{P(+ | D)P(D) + P(+ | \neg D)P(\neg D)} \\ &= \frac{P(+ | D)P(D)}{P(+ | D)P(D) + \{1 - P(+ | \neg D)\}\{1 - P(D)\}} \\ &= \frac{.997 \times .001}{.997 \times .001 + .015 \times .999} \\ &= .062 \end{aligned}$$

- In this population a positive test result only suggests a 6% probability that the subject has the disease
- (The positive predictive value is 6% for this test)

More on this example

- The low positive predictive value is due to low prevalence of disease and the somewhat modest specificity
- Suppose it was known that the subject was an intravenous drug user and routinely had intercourse with an HIV infected partner
- Notice that the evidence implied by a positive test result does not change because of the prevalence of disease in the subject's population, only our interpretation of that evidence changes

Likelihood ratios

- Using Bayes rule, we have $P(D \sim | \sim +) = \frac{P(+ \sim | \sim D)P(D)}{P(+ \sim | \sim D)P(D) + P(+ \sim | \sim D^c)P(D^c)}$
and $P(D^c \sim | \sim +) = \frac{P(+ \sim | \sim D^c)P(D^c)}{P(+ \sim | \sim D)P(D) + P(+ \sim | \sim D^c)P(D^c)}$.

Likelihood ratios

- Therefore $\frac{P(D \sim | \sim +)}{P(D^c \sim | \sim +)} = \frac{P(+ \sim | \sim D)}{P(+ \sim | \sim D^c)} \times \frac{P(D)}{P(D^c)}$
ie $\text{post-test odds of } D = \text{DLR}_+ \times \text{pre-test odds of } D$
- Similarly, DLR_- relates the decrease in the odds of the disease after a negative test result to the odds of disease prior to the test.

HIV example revisited

- Suppose a subject has a positive HIV test
- $DLR_+ = .997 / (1 - .985) \approx 66$
- The result of the positive test is that the odds of disease is now 66 times the pretest odds
- Or, equivalently, the hypothesis of disease is 66 times more supported by the data than the hypothesis of no disease

HIV example revisited

- Suppose that a subject has a negative test result
- $DLR_- = (1 - .997) / .985 \approx .003$
- Therefore, the post-test odds of disease is now .3\% of the pretest odds given the negative test.
- Or, the hypothesis of disease is supported .003 times that of the hypothesis of absence of disease given the negative test result