

Mathematical Biostatistics BootCamp: Lecture 5, Conditional Probability

Brian Caffo

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University

August 16, 2012

Table of contents

- 1 Table of contents
- 2 Conditional probability
- 3 Conditional densities
- 4 Bayes' Rule
- 5 Diagnostic tests
- 6 DLRs

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

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

- Notice that if A and B are independent, then

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

Example

- Consider our die roll example
- $B = \{1, 3, 5\}$
- $A = \{1\}$

$$\begin{aligned}P(\text{one given that roll is odd}) &= P(A \mid B) \\&= \frac{P(A \cap B)}{P(B)} \\&= \frac{P(A)}{P(B)} \\&= \frac{1/6}{3/6} = \frac{1}{3}\end{aligned}$$

Conditional densities and mass functions

- Conditional densities or mass functions of one variable conditional on the value of another
- Let $f(x, y)$ be a bivariate density or mass function for random variables X and Y
- Let $f(x)$ and $f(y)$ be the associated marginal mass function or densities disregarding the other variables

$$f(y) = \int f(x, y) dx \quad \text{or} \quad f(y) = \sum_x f(x, y) dx.$$

- Then the **conditional** density or mass function *given that* $Y = y$ is given by

$$f(x | y) = f(x, y) / f(y)$$

- It is easy to see that, in the discrete case, the definition of conditional probability is exactly as in the definition for conditional events where $A =$ the event that $X = x$ and $B =$ the event that $Y = y$
- The continuous definition is a little harder to motivate, since the events $X = x$ and $Y = y$ each have probability 0
- However, a useful motivation can be performed by taking the appropriate limits as follows
- Define $A = \{X \leq x\}$ while $B = \{Y \in [y, y + \epsilon]\}$

Continued

$$\begin{aligned}P(X \leq x \mid Y \in [y, y + \epsilon]) &= P(A \mid B) = \frac{P(A \cap B)}{P(B)} \\&= \frac{P(X \leq x, Y \in [y, y + \epsilon])}{P(Y \in [y, y + \epsilon])} \\&= \frac{\int_y^{y+\epsilon} \int_{-\infty}^x f(x, y) dx dy}{\int_y^{y+\epsilon} f(y) dy} \\&= \frac{\epsilon \int_y^{y+\epsilon} \int_{-\infty}^x f(x, y) dx dy}{\epsilon \int_y^{y+\epsilon} f(y) dy}\end{aligned}$$

$$\begin{aligned}
 &= \frac{\frac{\int_{-\infty}^{y+\epsilon} \int_{-\infty}^x f(x,y) dx dy - \int_{-\infty}^y \int_{-\infty}^x f(x,y) dx dy}{\epsilon}}{\frac{\int_{-\infty}^{y+\epsilon} f(y) dy - \int_{-\infty}^y f(y) dy}{\epsilon}} \\
 &= \frac{\frac{g_1(y+\epsilon) - g_1(y)}{\epsilon}}{\frac{g_2(y+\epsilon) - g_2(y)}{\epsilon}}
 \end{aligned}$$

where

$$g_1(y) = \int_{-\infty}^y \int_{-\infty}^x f(x,y) dx dy \quad \text{and} \quad g_2(y) = \int_{-\infty}^y f(y) dy.$$

- Notice that the limit of the numerator and denominator tends to g'_1 and g'_2 as ϵ gets smaller and smaller
- Hence we have that the conditional distribution function is

$$P(X \leq x \mid Y = y) = \frac{\int_{-\infty}^x f(x, y) dx}{f(y)}.$$

- Now, taking the derivative with respect to x yields the conditional density

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$

Geometrically

- Geometrically, the conditional density is obtained by taking the relevant slice of the joint density and appropriately renormalizing it
- This idea extends to any other line, or even non-linear functions

Example

- Let $f(x, y) = ye^{-xy-y}$ for $0 \leq x$ and $0 \leq y$
- Then note

$$f(y) = \int_0^{\infty} f(x, y) dx = e^{-y} \int_0^{\infty} ye^{-xy} dx = e^{-y}$$

- Therefore

$$f(x | y) = f(x, y)/f(y) = \frac{ye^{-xy-y}}{e^{-y}} = ye^{-xy}$$

Example

- Let $f(x, y) = 1/\pi r^2$ for $x^2 + y^2 \leq r^2$
- X and Y are uniform on a circle with radius r
- What is the conditional density of X given that $Y = 0$?
- Probably easiest to think geometrically

$$f(x \mid y = 0) \propto 1 \quad \text{for} \quad -r \leq x \leq r$$

- Therefore

$$f(x \mid y = 0) = \frac{1}{2r} \quad \text{for} \quad -r \leq x \leq r$$

Bayes' rule

- Let $f(x | y)$ be the conditional density or mass function for X given that $Y = y$
- Let $f(y)$ be the marginal distribution for y
- Then if y is continuous

$$f(y | x) = \frac{f(x | y)f(y)}{\int f(x | t)f(t)dt}$$

- If y is discrete

$$f(y | x) = \frac{f(x | y)f(y)}{\sum_y f(x | t)f(t)}$$

- Bayes' rule relates the conditional density of $f(y | x)$ to the $f(x | y)$ and $f(y)$
- A special case of this kind relationship is for two sets A and B , which yields that

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

Proof:

- Let X be an indicator that event A has occurred
- Let Y be an indicator that event B has occurred
- Plug into the discrete version of Bayes' rule

Example: 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, $P(+ \mid D)$
- The **specificity** is the probability that the test is negative given that the subject does not have the disease, $P(- \mid D^c)$

More definitions

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

$$sensitivity / (1 - specificity)$$

- The **diagnostic likelihood ratio of a negative test**, labeled DLR_- , is $P(- | D)/P(- | D^c)$, 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 $P(D \mid +)$ given the sensitivity, $P(+ \mid D) = .997$, the specificity, $P(- \mid D^c) = .985$, and the prevalence $P(D) = .001$

Using Bayes' formula

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

- 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

$$P(D \mid +) = \frac{P(+ \mid D)P(D)}{P(+ \mid D)P(D) + P(+ \mid D^c)P(D^c)}$$

and

$$P(D^c \mid +) = \frac{P(+ \mid D^c)P(D^c)}{P(+ \mid D)P(D) + P(+ \mid D^c)P(D^c)}.$$

- Therefore

$$\frac{P(D \mid +)}{P(D^c \mid +)} = \frac{P(+ \mid D)}{P(+ \mid D^c)} \times \frac{P(D)}{P(D^c)}$$

ie

post-test odds of $D = DLR_+ \times$ 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