Lecture 2

Conditional Probability, Independence and Bayes Theorem

By Mr. Leonard Thuo

Overview

- Conditional probability
- Independence
- Law of total probability
- Bayes Theorem

Conditional probability

Consider screening randomly-selected individuals for HIV. For each person we can define two events: the test status and the HIV status.

$$H+, H- = \text{ HIV status}$$

$$T+, T- = \text{ test status}$$

In practice, we frequently will be interested in specific conditional probabilities:

$$P(H+|T+) = \text{ prob of being HIV+,}$$
$$\text{conditional on having a positive test}$$

$$P(T+|H+) = \text{ prob of having a positive test,}$$
$$\text{conditional on being HIV+}$$

Definition of conditional probability

For any two events A and B, where $P(B) > 0$, the conditional probability of event A, given that event B already has occurred, is given by

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

A direct result of this definition is the multiplication law, which can be used to find a joint probability of $A \cap B$:

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

Example HIV testing

Recall the HIV testing situation. The sample space is

$$\Omega = (H+, T+), (H+, T-), (H-, T+), (H-, T-)$$

Suppose the probabilities are as follows:

	H+	H-
T+	.10	.04
T-	.01	.85

Find the following: $P(H+), P(T+), P(H + \cup T+), P(H + \cap T+)$

Computing Conditional probabilities

1. What is the probability of being HIV$+$, given that the test result is positive? (This is called the predictive value of a positive test). $P(H+|T+) = \; = .71$
2. What is the probability of being HIV-, given that the test result is positive? Use the complement law. Ans: .29
3. What is the probability of testing positive, given that true HIV status is positive? (This is called sensitivity of a diagnostic test). $P(T+|H+) = \; = .91$

Example 1

The multiplication law can be used to calculate a joint probability $P(A \cap B)$ when $P(A|B)$ and $P(B)$ are known:

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

An urn contains 5 red balls and 6 white balls. You select two balls at random, without replacement. What is the probability of selecting 2 red balls?

$$Ans : \frac{5}{11} * \frac{4}{10} = .18$$

Example 2

In a particular community, the prevalence of HIV is .10. Among those with HIV, it is found that 40% also have hepatitis C (HCV). What is the probability that a randomly selected person is infected with both HIV and HCV?

$$\begin{aligned} Ans : P(HIV \cap HCV) &= P(HCV|HIV)P(HIV) \\ &= 0.4 * 0.1 = 0.04 \text{ or } 4\% \end{aligned}$$

Independent events

Two events A and B are said to be independent if

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

The intuition behind this can be seen as follows. We can say that A and B are independent if conditioning on B does not affect the probability of A occurring. That is,

$$P(A|B) = P(A)$$

Calculating Conditional Probabilities

Example 1

Suppose the population prevalence of TB is .01. Two randomly-selected individuals are tested for TB. What is the probability that: (a) both have TB? (b) neither has TB? (c) exactly one has TB? Ans: (a) .0001; (b) .9801; (c) .0198

Answer

Possible combinations (T+T+)(T+T-)(T-T+)(T-T-)

$$1.P(T+T+) = 0.01 * 0.01$$

$$2.P(T-T-) = 0.99 * 0.99$$

$$3.P((T+T-)or(T-T+) = 0.01 * 0.99 + 0.99 * 0.01$$

Example 2

A gene expression array has 10,000 cells that each measure RNA expression of individual genes. The false positive rate for any given cell i.e., the probability of indicating gene activity when it is in fact absent is 1 in 5000. The expression array is applied to a neutral medium, where it is known that none of the genes on the array will be active. What is the probability that the array will correctly indicate no activity for each of the 10,000 genes? Ans: .135

Answer

$P(T+|D-)$ in one cell=$1/5000$=0.0002 In 10,000 cell this probability is 0.0002^{10000}
Probability $P(T-|D-)$=1-0.0002 in each of the cell and for 10,000 probability is $(1-0.0002)^{10,000}$

Example 3

Suppose that the probability of contracting HIV in one act of sexual intercourse is 1 in 500. If a person has 100 sexual encounters with an HIV-infected individual, what is the probability of contracting HIV?

Law of total probability

Suppose the events $B_1, B_2, ..., B_n$ are disjoint and exhaustive. By exhaustive, we mean that $P(B_1 \cup B_2 \cup . ... \cup B_n) = 1$ Then for any event A,

$$\begin{aligned} P(A) = P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_n) \\ = \sum_{i=1}^{n} P(A \cap B_i) \end{aligned}$$

Because $P(A \cap B_i) = P(A|B_i)P(B_i)$, we also have

$$\begin{aligned} P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + ... + P(A|B_n)P(B_n) \\ = \sum_{i=1}^{n} P(A|B_i)P(B_i) \end{aligned}$$

Law of total probability in words

Says that the probability of an event A is a weighted average of conditional probabilities taken over a set of mutually exclusive and exhaustive events.
The conditional probabilities are $P(A|B_i)$ and they are weighted by $P(B_i)$.

Application of law of total probability

Consider MRI for breast CA, where M = 0,1, 2 is true degree of malignancy (0=none, 1=benign, 2=malignant). Suppose prevalence of each underlying status, for the population being screened, is 0.80 (none), 0.15 (benign), 0.05 (malignant). Upon radiology scan, a biopsy is either ordered (B) or not ordered (B^c), depending on the judgment of the radiologist. Suppose a particular radiologist orders biopsies with the following probabilities:

$$P(B|M = 0) = .10$$

$$P(B|M = 1) = .40$$

$$P(B|M = 2) = .90$$

What is the probability that a patient undergoing breast MRI will be referred for a biopsy? Ans = .19

Diagnostic test

Consider the following events:

- D = "Disease is present"
- D^c = "Disease is absent"
- T^+ = "Positive test result (test detects disease)"
- T^- = "Negative test result (test does not detect disease)"

In diagnostic-testing situations, the following "performance parameters" of the diagnostic procedure under consideration will be available:

- $P(T^+|D)$ = "Sensitivity (true positive rate) of the test"
- $P(T^+|D^c)$ = "Probability of a false positive test result"
- $P(T^-|D)$ = "Probability of a false negative test result"
- $P(T^-|D^c)$ = "Specificity (or true-negative rate) of the test"

Cont:Diagnostic test

To derive estimates of the predictive values of a positive test (PVP) and predictive value of a negative test PVN , where $PVN = P(D^c|T^-)$ we will need an estimate of the overall probability of disease in the general population. This is called the prevalence of the disease P(D).
Goal: Find $P(D|T^+)$ the predictive value of a positive test result (or PVP), that is, find the probability that a subject has the disease given a positive test.

Example

In a large study 1820 individuals with or without tuberculosis (ascertained via an independent test) had an X-ray performed on them in order to ascertain the predictive ability of this examination. The situation is presented in the table below.

Xray results	TB yes	TB No	Total
Positive	22	51	73
Negative	18	1739	1747
Total	30	1790	1820

Consider that the prevalence of the disease in the general population is $P(D) = 0{:}000093$ (i.e., 9.3 cases in 100,000).

Cont:example

Sensitivity of Xray test $= P(T^+|D) = \frac{22}{30} = 0.7333.$

Specificity of Xray test $P(T^-|D^c) = \frac{1739}{1790} = 0.9715.$

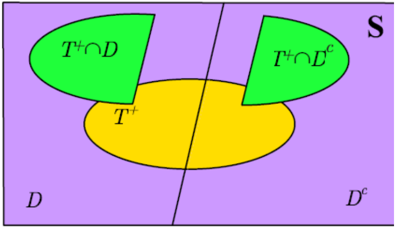
False Negative Rate $P(T^-|D) = \frac{8}{30} = 0.2667$

False Positive Rate $P(T^+|D^-) = \frac{51}{1790} = 0.0285$

Now let's calculate $P(D|T^+)$. By the definition of conditional probability,

$$P(D|T^+) = \frac{P(D \cap T^+)}{P(T^+)}$$

Since we do not know $P(T^+)$ let us consult the figure below.



Cont:

From the Figure it is seen that $T^+ = (T^+ \cap D) \cup (T^+ \cap D^c)$ so that

$$P(T^+) = P[(T^+ \cap D) \cup (T^+ \cap D^c)]$$
$$= P(T^+ \cap D) + P(T^+ \cap D^c)$$
$$= P(T^+|D)P(D) + P(T^+|D^c)P(D^c)$$

since $D \cap T^+$ and $T^+ \cap D^c$ are mutually exclusive events (using the additive rule). Then substituting above we have

$$P(D|T^+) = \frac{P(T^+|D)P(D)}{P(T^+|D)P(D) + P(T^+|D^c)P(D^c)}$$
$$= \frac{sensitivity * prevalence}{sensitivity * prevalence + falsepositive * (1 - prevalence)}$$
$$= 0.00239$$

For every 100,000 positive x-rays, only 239 signal true cases of tuberculosis. This is called the "false positive paradox".

Bayes theorem

Let $B_1, ..., B_n$ be a set of disjoint and exhaustive events. Then for some event A, the conditional probability $P(B_j|A)$ is

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

Example

In the previous example, among those referred for biopsy, what is the probability of having a no tumor? Ans: .42. What is the probability of having a benign tumor? Ans: .32. What is the probability of having a malignant tumor? Ans: .24.