BIOS 4120: Introduction to Biostatistics
Breheny

Lab #6 (solutions)

Note: We will not be using R for this lab.

Today's lab is about probability. You have learned some very useful properties about probabilities in class, now it's time to provide some examples of when these properties come into play. Then we'll talk a little bit about conditional probabilities, and when they come in handy. Last we will do an example or two of Bayes' rule, learning how to calculate sensitivity and specificity along the way.

1 Probability Rules

- **Addition Rule:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Complement Rule:** $P(A^C) = 1 - P(A)$
- **Multiplication Rule:** $P(A \cap B) = P(A)P(B|A)$ or $P(A \cap B) = P(B)P(A|B)$
  - *If A,B are independent:* $P(A \cap B) = P(A)P(B)$
- **Law of Total Probability:** $P(A) = P(A \cap B) + P(A \cap B^C)$
- **Bayes' Rule:** $P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^C)P(B|A^C)}$
  - *More on this later*

2 Probability Question Examples

For each of the following examples, state which principles/rules of probability your using and find a numerical answer.

a) Conditional Probability Card Game – Guess my card

Demonstrate how to use conditional probabilities to solve the problem.

b) In a class of 101 students, 30 are freshmen and 41 are sophomores. Find the probability that a student picked from this group at random is either a freshman or a sophomore.

This is the addition rule for mutually exclusive events.

Note that P(freshman) = 30/101 and P(sophomore) = 41/101. Thus

P(freshman or sophomore) = 30/101 + 41/101 = 71/101

This makes sense since 71 of the 101 students are freshmen or sophomores.

c) Suppose two 6-sided dice are rolled. What is the probability that I roll two sixes? What assumption(s) are you making about the dice?

This is the multiplicative rule in the independent events case. P(2 sixes) = P(six)*P(six) = 1/36.

You are assuming independence and fair dice.

d) A different class, still with 101 students, has 40 juniors. Also, the class has 50 females, 22 of whom are juniors. Find the probability that a student picked from this group at random is either a junior or a female.

This is the addition rule with non-mutually exclusive events.

Note that P(junior) = 40/101 and P(female) = 50/101, and P(junior and female) = 22/101. Thus

P(junior or female) = 40/101 + 50/101 – 22/101 = 68/101

This makes sense since 68 of the 101 students are juniors or female.

Not sure why? When we add 40 juniors to 50 females and get a total of 90, we have overcounted. The 22 female juniors were counted twice; 90 minus 22 makes 68 students who are juniors or female.

e) Blind man's bluff: I'm going to show you my card, then shuffle the cards again (with my card included) and deal one to you. You have to determine if your card will be higher or lower than mine. Assume aces are high. Knowing my card, (before I hand you a card), what is the probability that your card is greater than or equal to my card? Would you bet that your card is greater than or equal to mine?

f) (I will pass out cards now). Check your card. Write down whether or not you won this round of blind man's bluff (i.e. whether or not your card beat my card).

The proportion of correct answers in the class should be close to the probability they found in e). **See table to see probabilities for whether you will beat my card.** Note: since there is no replacement of cards in the deck, the probabilities below are a bit off, but they hold under replacement = TRUE as n goes to infinity (n being the number of students participating)

```
      mycard        sph
 [1,]      2 1.00000000
 [2,]      3 0.92307692
 [3,]      4 0.84615385
 [4,]      5 0.76923077
 [5,]      6 0.69230769
 [6,]      7 0.61538462
 [7,]      8 0.53846154
 [8,]      9 0.46153846
 [9,]     10 0.38461538
[10,]     11 0.30769231
[11,]     12 0.23076923
[12,]     13 0.15384615
[13,]     14 0.07692308
```

3 Screening/Diagnostic Test Examples

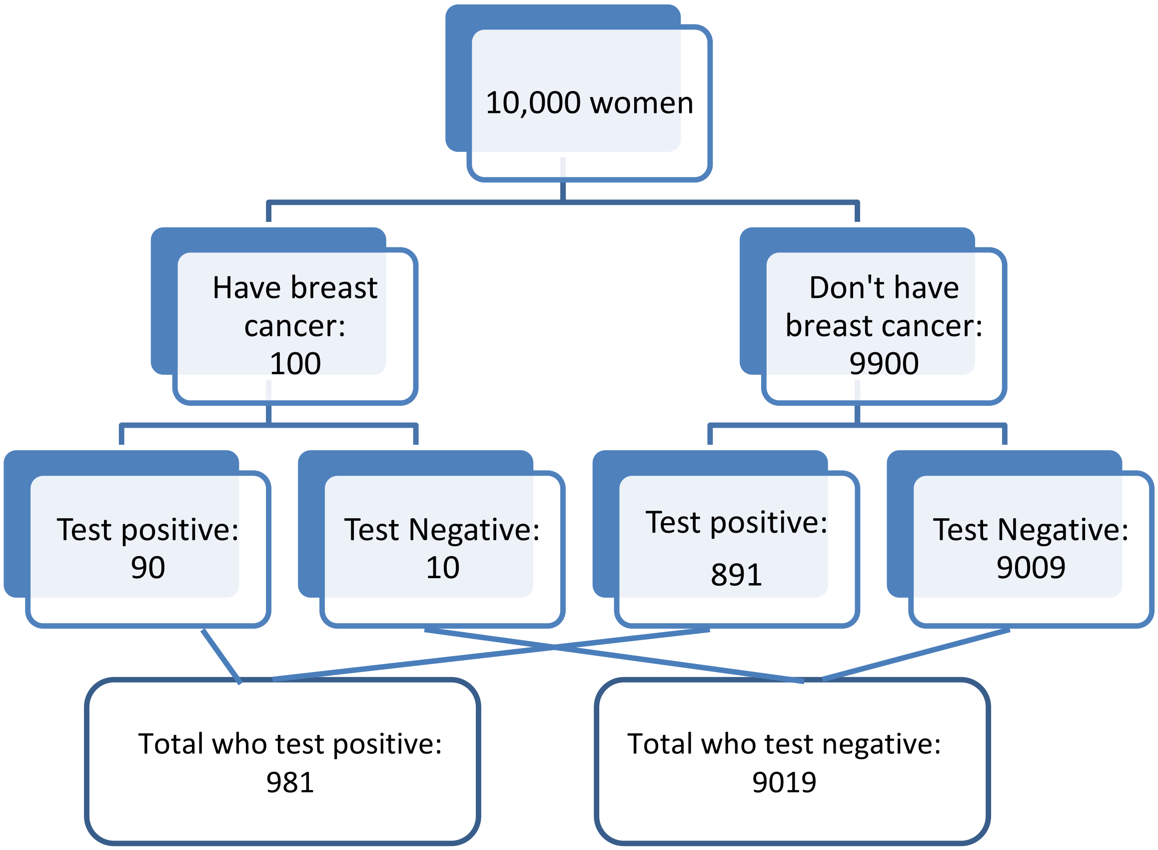
Example 1:
Scenario (from Gigerenzer 2007): The prevalence of breast cancer in women is 1%. A mammography has sensitivity of 90% (90% of women with breast cancer will test positive) and a false positive rate of 9%.

Of the four choices below, which best describes what you would tell a patient with a positive mammogram?

a) It is not certain you have breast cancer, yet the probability is about 81%.
b) Out of 10 women who test positive as you did, about 9 have breast cancer.
c) Out of 10 women who test positive as you did, only about 1 has breast cancer.
d) The chance you have breast cancer is about 1%.

In Gigerenzer, it was found that most practicing physicians answered either A or B. Only 21% answered C, which is the correct answer.
(http://www.bbc.com/news/magazine-28166019)

- **Probability tree**



Now let's look at the probability a person has the disease given that they test positive.

$$P(cancer|+) = \frac{90}{90+891} = \frac{90}{981} = 0.092$$

- **Sensitivity, specificity, prevalence**
  - **Prevalence:** P(cancer) = 0.01
  - **Sensitivity:** P(+ | cancer) = 0.9
  - **False Positive rate:** P(+ | no cancer) = 0.09
  - **Specificity**: P(- | no cancer) = 1-FPR = 1-0.09 = 0.91

Example 2:

The following data are taken from a study investigating the use of a technique called radionuclide ventriculography as a diagnostic test for detecting coronary artery disease. Assume the participants were randomly selected from the population.

Test	**Disease**		**Total**
	Present (D)	**Absent (D^c)**	
+	302	80	382
-	179	372	551
Total	481	452	933

- **Find P(+ | D)**
  - Otherwise known as the *sensitivity*

$$P(+|\, D) = \frac{302}{481} = 0.6279$$

- **Find P(D)**
  - Otherwise known as *prevalence*

$$P(D) = \frac{481}{933} = 0.5155$$

- **Find P(+ | D^c)**
  - Otherwise known as the *false positive rate*

$$P(+|\, D^c) = \frac{80}{452} = 0.1770$$

- **Find P(- | D^c)**
  - Otherwise known as the *specificity*

$$P(-|\, D^c) = \frac{372}{452} = 0.8230 = 1 - FPR$$

- **Find P(D | +)** (from the table and with Baye's Theorem)
  - Otherwise known *predictive value positive*

$$P(D|+) = \frac{P(D)P(+|D)}{P(D)P(+|D) + P(D^C)P(+|D^C)} = \frac{.5155 * .6279}{.5155 * .6279 + (1 - .5155) * (1 - .8230)} = .791$$