Statistics 410, Fall 2011

Solutions to homework 3, due 10/12/11

3.15 We are assuming that $P(E) = P(E|F)$. The second number is $P(E \cap F)/P(F)$. Solving

$$P(E) = \frac{P(E \cap F)}{P(F)}$$

gives $P(E \cap F) = P(E) \cdot P(F)$. Now we calculate

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E) \cdot P(F)}{P(E)} = P(F).$$

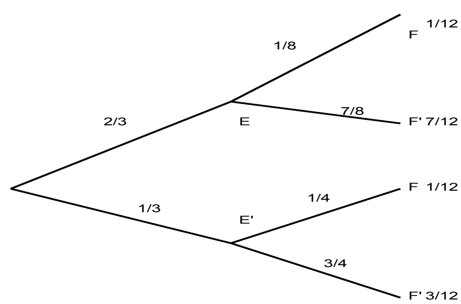
3.16 We assume E and F are independent. Then $P(E|F) = P(E)$, and as above, $P(E \cap F) = P(E) \cdot P(F)$. We also note that

$$P(E) = P(E|F)P(F) + P(E|F')P(F') = P(E)P(F) + P(E|F')(1 - P(F)).$$

If we subtract $P(E)P(F)$ from both sides, we get

$$P(E)(1 - P(F)) = P(E|F')(1 - P(F)).$$

Dividing by $1 - P(F)$, we get $P(E) = P(E|F')$. *Note:* this only makes sense when $P(F') \neq 0$.



3.18

3.20 Following the reasoning of example 3.33, we get

$$\frac{364 \cdot 363 \cdot \cdots \cdot 350}{365^{15}} = .716$$

is the probability that no two people share a birthday. So the probability that at least two share a birthday is $1 - .716 = .284$.

3.21 If you do the calculation with 23 people, you discover the probability that no two people have the same birthday is .493, which is less than .5. On the other hand, if one does the calculation for 22 people, then the probability that no two people have the same birthday is .524, which is greater than .5.

So there need to be at least 23 people at a meeting for the probability that 2 share a birthday to be greater than .5.

3.23 We use the formula from page 108.

$$PV^+ = \frac{P(\text{disease and positive})}{P(positive)} = \frac{\frac{1}{63}.8}{\frac{1}{63}.8 + \frac{62}{63}.1} = \frac{.013}{.013 + .098} = \frac{.013}{.111} = .112.$$

This says that the risk of breast cancer before the exam is $1/63 = .016$, and after a positive exam it is .112 - roughly 7 times as high.

The probability of having breast cancer with a negative test is:

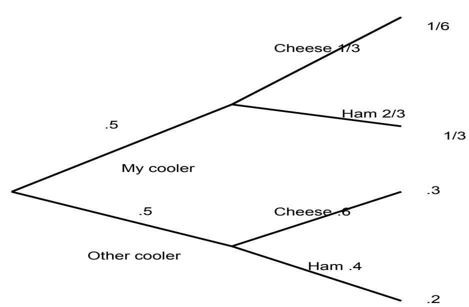
$$P(\text{cancer}|\text{negative}) = \frac{P(\text{negative and cancer})}{P(negative)} = \frac{\frac{1}{63}.2}{\frac{1}{63}.2 + \frac{62}{63}.9} = \frac{.003}{.003 + .886} = \frac{.003}{.889} \approx .003.$$

This is roughly 3 tenths of a percent, roughly 1/5 of the risk without a test.

3.24 It will be helpful to have a tree diagram. The probability will be

$$\frac{1/3}{1/3 + .2} = .414$$

by using the terminal probaiblities in the tree diagram.



3.28 We wish to find the probability that the cab was blue, given a witness identifying it as blue. Let E be the event that the cab was blue, and F be the event that it was identified as blue by the witness.

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E) \cdot .8}{P(E) \cdot .8 + P(E') \cdot .2} = \frac{.15 \cdot .8}{.15 \cdot .8 + .85 \cdot .2} = \frac{.12}{.29} = .414.$$

So the probability that the cab was actually blue was near 42%.

3.29 (a) The blank on the lower left is $(1-x)(1-\text{specificity})$. The blank on the upper right is $x(1-\text{sensitivity})$.

(b)

$$P(\text{test positive}) = x \cdot \text{sensitivity} + (1-x)(1-\text{sensitivity}).$$

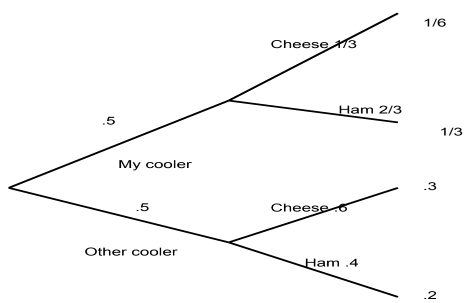
(c)

$$PV^+ = \frac{x \cdot \text{sensitivity}}{x \cdot \text{sensitivity} + (1-x)(1-\text{specificity})}.$$

(d) We assume the test is not perfect, in both ways. That is, the specificity is less than 1 and the sensitivity is also less than 1. One important pattern is that as x increases, so does PV^+. More important, as x decreases, PV^+ decreases too, and as x approaches 0, so does PV^+. So we get less positive predictive value for rarer diseases, and (all other things being useful) more positive predictive value for more common diseases.

In particular, depending on x, PV^+ can be any value between 0 and 1, both above and below x.

In all cases, the function is increasing with x, and sometimes it is concave up, sometimes concave down. In particular, if the sensitivity plus the specificity is greater than 1, the graph is concave down, and if less than 1, the graph is concave up.



3.52

b. $.2 + .4 = .6$.

c. $.4$.

d. $.2/.6 = 1/3$.

3.53 (a) Using Bayes' method,

$$P(\text{fair}|\text{three heads}) = \frac{P(\text{fair and three heads})}{P(\text{three heads})} = \frac{.8 \cdot \frac{1}{8}}{.8 \cdot \frac{1}{8} + .2 \cdot 1} = \frac{.1}{.3} = 1/3.$$

(b) Same calculation and method but now we replace 1/8 with 1/64 which is the chance of 6 heads in a row with a fair coin.

$$P(\text{fair}|\text{six heads}) = \frac{P(\text{fair and six heads})}{P(\text{six heads})} = \frac{.8 \cdot \frac{1}{64}}{.8 \cdot \frac{1}{64} + .2 \cdot 1} = \frac{.0125}{.2125} = .059.$$

3.75 This problem is somewhat badly posed. Without knowing the prevalance of prostate cancer in the population (rather than the number given, which is the prevalence of men who will eventually develop the disease) we can't answer the specificity, PV^+ or PV^- questions. We will pretend that .3 is the prevalence in the population to answer those questions.

(a) The prevalence of men who will develop prostate cancer at some point is .3.

(b) The sensitivity is .8.

(c) The specificity is 1/3.

(d) Since 2/3 of the positive tests don't have prostate cancer, 1/3 do. So that 1/3 of the positive test is $.3 \cdot .8 = .24$ of the population. Hence the other 2/3 of the positive tests represent .48 of the population. This number is (1 - prevalence) (1 - specificity) so

$$.48 = (1 - .3)(1 - S)$$

where S is the specificity. Solving for S, $1 - S = .48/.7$ so $S = 22/70$.

(e)

$$PV^+ = \frac{.3 \cdot .8}{.3 \cdot .8 + .7 \cdot (1 - 22/70)} = \frac{.24}{.24 + .48} = 1/3.$$

(f)

$$PV^- = \frac{.7 \cdot 22/70}{.7 \cdot 22/70 + .3 \cdot .2} = \frac{.22}{.22 + .06} = \frac{22}{28}.$$

4.3

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{27}{64} & 0 \leq x < 1 \\ \frac{54}{64} & 1 \leq x < 2 \\ \frac{63}{64} & 2 \leq x < 3 \\ 1 & 3 \leq x. \end{cases}$$

4.4

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{27} & 0 \leq x < 1 \\ \frac{7}{27} & 1 \leq x < 2 \\ \frac{19}{27} & 2 \leq x < 3 \\ 1 & 3 \leq x. \end{cases}$$

4.5

$$f(x) = \begin{cases} .75 & x = 0 \\ .25 & x = 1. \end{cases}$$

4.6

$$f(x) = \begin{cases} .25 & x = 1 \\ .5 & x = 3 \\ .25 & x = 5. \end{cases}$$

4.9

$$E(X) = \frac{27}{64}0 + \frac{27}{64}1 + \frac{9}{64}2 + \frac{1}{64}3 = \frac{48}{64} = .75.$$

$$\sigma_X = \sqrt{\frac{27}{64}(0-.75)^2 + \frac{27}{64}(1-.75)^2 + \frac{9}{64}(2-.75)^2 + \frac{1}{64}(3-.75)^2}$$
$$= \sqrt{\frac{3^5}{4^5} + \frac{3^3}{4^5} + \frac{9\cdot 25}{4^5} + \frac{3^7}{4^5}} = \sqrt{\frac{576}{4^5}} = \sqrt{9/16} = .75.$$

Note that it is easier to compute this by using Theorem 4.2.

$$E(X^2) = \frac{27}{64}0 + \frac{27}{64}1 + \frac{9}{64}4 + \frac{1}{64}9 = \frac{72}{64} = 1.125$$

Then the variance is $1.125 - .75^2 = .5625$, and the standard deviation is the square root of this: .75.

4.10

$$E(X) = \frac{1}{27}0 + \frac{6}{27}1 + \frac{12}{27}2 + \frac{8}{27}3 = \frac{54}{27} = 2.$$

To calculate σ,

$$E(X^2) = \frac{1}{27}0 + \frac{6}{27}1 + \frac{12}{27}4 + \frac{8}{27}9 = \frac{126}{27} = 14/3.$$

So the variance is $14/3 - 4 = 2/3$ and σ is $\sqrt{2/3} = .816$.

4.12 First we write down what f_X is. f_Y will be the same.

$$f(x) = \begin{cases} \frac{1}{16} & x = 0 \\ \frac{1}{4} & x = 1 \\ \frac{3}{8} & x = 2 \\ \frac{1}{4} & x = 3 \\ \frac{1}{16} & x = 4. \end{cases}$$

A quick calculation, or symmetry, tells us that the means of X and Y are 2. $X+Y$ is always 4, so has mean 4.

To calculate σ_X,

$$\sigma = \sqrt{\frac{1}{16}2^2 + \frac{1}{4}1^2 + \frac{3}{8}0^2 + \frac{1}{4}1^2 + \frac{1}{16}2^2} = \sqrt{\frac{16}{16}} = 1.$$

The σ for Y is the same, and the σ for $X+Y$ is 0.

4.14 First we write down what f_X is. $f_Y(n) = f_X(4-n)$.

$$f(x) = \begin{cases} .48^4 = .053 & x = 0 \\ 4 \cdot .52 \cdot .48^3 = .230 & x = 1 \\ 6 \cdot .52^2 \cdot .48^2 = .374 & x = 2 \\ 4 \cdot .52^3 \cdot .48 = .270 & x = 3 \\ .52^4 = .073 & x = 4 \end{cases}$$

$$E(X) = .053 \cdot 0 + .23 \cdot 1 + .374 \cdot 2 + .27 \cdot 3 + .073 \cdot 4 = 2.08.$$

It follows that $E(Y) = 1.92$.

$$\sigma_X = \sqrt{.053 \cdot 2.08^2 + .23 \cdot 1.08^2 + .374 \cdot .08^2 + .27 \cdot .92^2 + .073 \cdot 1.92^2} = .999$$

By symmetry, σ_Y will be the same.

Of course the mean of $X + Y$ is 4, and the standard deviation is 0.