

STAT 325 Introduction to Probability Models Spring 2010

HW6 due Tues April 13

Topics: Bayes' Theorem

Show your method of solution for all of the following.

1. Suppose that three associates prepare legal briefs for a law firm. Lorenzo prepares 50% of the briefs and makes an error 10% of the time. Marci prepares 35% of the briefs and makes an error 20% of the time. Nancy prepares 15% of the briefs and makes an error 5% of the time.

a) Determine the percentage of legal briefs prepared by this law firm that contain an error. (Note that this is equivalent to the probability that a randomly selected legal brief contains an error.)

Let E denote the event that the legal brief contains an error, let L denote that Lorenzo prepared it, M that Marci prepared it, and N that Nancy prepared it. By the law of total probability, $\Pr(E) = \Pr(E|L)\Pr(L) + \Pr(E|M)\Pr(M) + \Pr(E|N)\Pr(N) = (.1)(.5) + (.2)(.35) + (.05)(.15) = .1275$. So, 12.75% of legal briefs prepared by this law firm contain an error.

b) Given that a (randomly selected) legal brief contains an error, what is the (updated) probability that Lorenzo prepared it?

$\Pr(L|E) = \Pr(E|L)\Pr(L) / \Pr(E)$, where $\Pr(E)$ was found in (a). This gives:

$$\Pr(L|E) = (.1)(.5) / .1275 = .05 / .1275 \approx .392$$

c) Given that a (randomly selected) legal brief contains an error, what is the (updated) probability that Marci prepared it?

$\Pr(M|E) = \Pr(E|M)\Pr(M) / \Pr(E)$, where $\Pr(E)$ was found in (a). This gives:

$$\Pr(M|E) = (.2)(.35) / .1275 = .07 / .1275 \approx .549$$

d) Given that a (randomly selected) legal brief contains an error, what is the (updated) probability that Nancy prepared it?

$\Pr(N|E) = \Pr(E|N)\Pr(N) / \Pr(E)$, where $\Pr(E)$ was found in (a). This gives:

$$\Pr(N|E) = (.05)(.15) / .1275 = .0075 / .1275 \approx .059$$

e) Now suppose that a randomly selected legal brief does *not* contain an error. Given this information, which associate is most likely to have prepared the brief? Which is least likely? Support your answers with relevant probability calculations.

$\Pr(E^c) = \Pr(E^c|L)\Pr(L) + \Pr(E^c|M)\Pr(M) + \Pr(E^c|N)\Pr(N) = (.9)(.5) + (.8)(.35) + (.95)(.15) = .8725$. Of course, we also could have found:

$$\Pr(E^c) = 1 - \Pr(E) = 1 - .1275 = .8725.$$

Then Bayes' Theorem establishes that:

$$\Pr(L|E^c) = \Pr(E^c|L)\Pr(L) / \Pr(E^c) = (.9)(.5) / .8725 = .45/.8725 \approx .516$$

$$\Pr(M|E^c) = \Pr(E^c|M)\Pr(M) / \Pr(E^c) = (.8)(.35) / .8725 = .28/.8725 \approx .321$$

$$\Pr(N|E^c) = \Pr(E^c|N)\Pr(N) / \Pr(E^c) = (.95)(.15) / .8725 = .1425/.8725 \approx .163$$

So, given that the brief does not contain an error, Lorenzo is most likely to have prepared it and Nancy is least likely.

2. Many spam filter algorithms are based on Bayes' Theorem. As a very simplified example, suppose that 15% of spam messages that you receive contain the word "greeting" and 1% of non-spam messages that you receive contain the word "greeting." Also suppose that 25% of the messages that you receive are spam. Consider selecting an email message at random, and define the following events: $G = \{\text{message contains the word "greeting"}\}$, $M = \{\text{message is spam}\}$.

a) Express the given values .15, .01, and .25 as probabilities in terms of these events.

$$\Pr(G|M) = .15 \quad \Pr(G|M^c) = .01 \quad \Pr(M) = .25$$

b) Determine the probability that a randomly selected e-mail message contains the word "greeting."

$$\text{By the law of total probability, } \Pr(G) = \Pr(G|M)\Pr(M) + \Pr(G|M^c)\Pr(M^c) = (.15)(.25) + (.01)(.75) = .045.$$

c) Given that a randomly selected e-mail message contains the word "greeting," what is the probability that the message is spam?

$$\text{Bayes' Thm gives that: } \Pr(M|G) = \Pr(G|M)\Pr(M) / [\Pr(G|M)\Pr(M) + \Pr(G|M^c)\Pr(M^c)] = (.15)(.25) / .045 = .0375/.045 = 5/6 \approx .833.$$

d) How many times more likely is the message to be spam given that it contains the word "greeting," as compared to the prior probability that the message is spam?

The ratio is $.833 / .25 = 10/3 \approx 3.333$, so the message is more than 3 times as likely to be spam if it is learned that the message contains the word "greeting."

3. Suppose that a box contains R red balls and 2 green balls. Your prior belief is that R is equally likely to be 0, 1, 2, 3, or 4. You select two balls from the box at random, and you find that both balls are green.

a) Determine the updated probabilities for the various possible values of R , in light of the evidence that both balls selected from the box were green. Identify the most likely and least likely values of R , in light of this evidence.

The prior probabilities are: $\Pr(R = 0) = 1/5$, $\Pr(R = 1) = 1/5$, $\Pr(R = 2) = 1/5$, $\Pr(R = 3) = 1/5$, and $\Pr(R = 4) = 1/5$.

Let G_2 denote the evidence that two green balls were selected. Then:

$$\Pr(G_2 | R = 0) = 1 \quad \Pr(G_2 | R = 1) = \frac{\binom{2}{2}\binom{1}{0}}{\binom{3}{2}} = 1/3 \quad \Pr(G_2 | R = 2) = \frac{\binom{2}{2}\binom{2}{0}}{\binom{4}{2}} = 1/6$$

$$\Pr(G_2 | R = 3) = \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} = 1/10 \quad \Pr(G_2 | R = 4) = \frac{\binom{2}{2}\binom{4}{0}}{\binom{6}{2}} = 1/15$$

Now we can use Bayes' Thm to calculate the updated probabilities to be:

$$\Pr(R = 0 | G_2) =$$

$$\frac{\Pr(G_2 | R = 0)\Pr(R = 0)}{\Pr(G_2 | R = 0)\Pr(R = 0) + \Pr(G_2 | R = 1)\Pr(R = 1) + \Pr(G_2 | R = 2)\Pr(R = 2) + \Pr(G_2 | R = 3)\Pr(R = 3) + \Pr(G_2 | R = 4)\Pr(R = 4)}$$

$$= \frac{(1)(1/5)}{(1)(1/5) + (1/3)(1/5) + (1/6)(1/5) + (1/10)(1/5) + (1/15)(1/5)} = \frac{1/5}{1/3} = 3/5 = .6$$

$$\text{Similarly, } \Pr(R = 1 | G_2) = (1/3)(1/5) / (1/3) = 1/5 = .2$$

$$\Pr(R = 2 | G_2) = (1/6)(1/5) / (1/3) = 1/10 = .1$$

$$\Pr(R = 3 | G_2) = (1/10)(1/5) / (1/3) = 3/50 = .06$$

$$\Pr(R = 4 | G_2) = (1/15)(1/5) / (1/3) = 3/75 = .04$$

Given that 2 green balls are selected, the most likely possibility is that there are 0 red balls (prob .6) and the least likely possibility is that there are 4 red balls (prob .04).

b) Repeat a), in the event that your selection of two balls produces one red and one green ball.

Let E denote the evidence that 1 red ball and 1 green ball were selected. Then we can similarly calculate:

$$\Pr(R = 0 | E) =$$

$$\frac{\Pr(E | R = 0)\Pr(R = 0)}{\Pr(E | R = 0)\Pr(R = 0) + \Pr(E | R = 1)\Pr(R = 1) + \Pr(E | R = 2)\Pr(R = 2) + \Pr(E | R = 3)\Pr(R = 3) + \Pr(E | R = 4)\Pr(R = 4)}$$

$$= \frac{(0)(1/5)}{(0)(1/5) + (2/3)(1/5) + (4/6)(1/5) + (6/10)(1/5) + (8/15)(1/5)} = \frac{0}{37/75} = 0$$

$$\text{Similarly, } \Pr(R = 1 | E) = (2/3)(1/5) / (37/75) = 10/37 \approx .270$$

$$\Pr(R = 2 | E) = (4/6)(1/5) / (37/75) = 10/37 \approx .270$$

$$\Pr(R = 3 | E) = (6/10)(1/5) / (37/75) = 9/37 \approx .243$$

$$\Pr(R = 4 | E) = (8/15)(1/5) / (37/75) = 8/37 \approx .216$$

Given that 1 red ball and 1 green ball are selected, the most likely possibility is that there are 1 or 2 reds balls (prob .270 each) and the least likely possibility is that there are 4 red balls (prob .216).

c) Repeat a), in the event that both balls selected turn out to be red.

Similarly we can calculate:

$$\Pr(R = 0 | R_2) =$$

$$\frac{\Pr(R_2 | R = 0)\Pr(R = 0)}{\Pr(R_2 | R = 0)\Pr(R = 0) + \Pr(R_2 | R = 1)\Pr(R = 1) + \Pr(R_2 | R = 2)\Pr(R = 2) + \Pr(R_2 | R = 3)\Pr(R = 3) + \Pr(R_2 | R = 4)\Pr(R = 4)}$$

$$= \frac{(0)(1/5)}{(0)(1/5) + (0)(1/5) + (1/6)(1/5) + (3/10)(1/5) + (6/15)(1/5)} = \frac{0}{13/75} = 0$$

$$\text{Similarly, } \Pr(R = 1 | R_2) = (0)(1/5) / (13/75) = 0$$

$$\Pr(R = 2 | R_2) = (1/6)(1/5) / (13/75) = 5/26 \approx .192$$

$$\Pr(R = 3 | R_2) = (3/10)(1/5) / (13/75) = 9/26 \approx .346$$

$$\Pr(R = 4 | R_2) = (6/15)(1/5) / (13/75) = 6/13 \approx .462$$

Given that 2 red balls are selected, the most likely possibility is that there are 4 red balls (prob .462) and the least likely possibility is that there are 0 or 1 red balls (prob 0).