

# Mat 354

## Homework 4

Kenny Roffo

Due September 28, 2015

1. Imagine a bucket is filled with marbles labeled 1 through 10. For each label  $i$ , there are  $i$  marbles with that label. Say we pull 5 marbles out with replacement. Let  $A$  be the event that the maximum is 8 and let  $B$  be the event that the minimum is 3.

a) How many marbles are in the bucket?

$$\begin{aligned} N &= 10 \times 10 + 9 \times 9 + 8 \times 8 + \dots + 1 \times 1 \\ &= 385 \end{aligned}$$

- b) Determine the probability that a single draw from the bucket yields a result no greater than 8.

$$\begin{aligned} P &= \frac{8 \times 8 + 7 \times 7 + \dots + 1 \times 1}{10 \times 10 + 9 \times 9 + \dots + 1 \times 1} \\ &= \frac{204}{385} \\ &= 0.5299 \end{aligned}$$

A general formula for finding the probability the maximum will be at most  $n$  is

$$P(\max \leq n) = \frac{n \times n + (n-1) \times (n-1) + \dots + 1 \times 1}{385}$$

c) Determine  $P(A)$ :

The probability that the max is 8 is 1 minus that probability the max is not 8.

We see

$$\begin{aligned}
 P(\max = 8) &= 1 - \overline{P(A)} \\
 &= 1 - P(\min \geq 9 \cup \max \leq 7) \\
 &= 1 - P(\overline{\max \leq 8} \cup \max \leq 7) \\
 &= 1 - (P(\overline{\max \leq 8}) + P(\max \leq 7)) && \text{Since these events are independent} \\
 &= 1 - (1 - P(\max \leq 8) + P(\max \leq 7)) \\
 &= P(\max \leq 8) - P(\max \leq 7) && \text{Using our general formula} \\
 &= 0.5299^5 - 0.3636^5 && \text{Raised to the 5th since these numbers are for inc} \\
 &= 0.0354
 \end{aligned}$$

d) Determine  $P(B)$ :

$$\begin{aligned}
 P(\max = 3) &= P(\overline{\max \leq 2} \cap \max \leq 3) \\
 &= P(\overline{\max \leq 2}) \times P(\max \leq 3) \\
 &= (1 - P(\max \leq 2)) \times P(\max \leq 3) \\
 &= (1 - 0.0130^5) \times (0.03636^5) \\
 &= 6.358 \times 10^{-8}
 \end{aligned}$$

e) Determine the probability that the maximum is 8 given the minimum is 3:

Given the minimum is 3, we know there are no 2s or 1s, thus we can calculate this the same way as we did part b, but instead of dividing by 385, we will divide by the number of marbles which are neither 2 nor 1, and also we will exclude 1 and 2 from the numerator too:

$$\begin{aligned}
 P &= \frac{8 \times 8 + 7 \times 7 + \dots + 3 \times 3}{10 \times 10 + 9 \times 9 + \dots + 3 \times 3} \\
 &= \frac{199}{380} \\
 &= 0.5234
 \end{aligned}$$

- f) Determine the probability that the sum of the integers is 49 (a mean of 9.8):

The only way this can happen is if there are four 10s and one 9. Since the order in which they appear does not matter, there are 5 ways this can occur. Thus

$$\begin{aligned} P &= \frac{5}{385} \\ &= 0.0130 \end{aligned}$$

2. Consider the following survey:

**Survey**

- 1) Was your mother born in an even-numbered year?
- 2) If you answer YES to Question 1: Was your best friend born on a weekend (Saturday or Sunday)?  
If you answer NO to Question 1: Have you ever cheated on a test?

Consider the two following events:

$E$  = the answer to Question 1 is YES       $Y$  = the answer to Question 2 is YES

$E$  and  $\bar{E}$  form a partition of the sample space.

- a) Use common sense to determine  $P(E)$ :

$$P(E) = \frac{1}{2}$$

- b) Use common sense to determine  $P(Y|E)$ :

$$P(Y|E) = \frac{2}{7}$$

- c) Convince yourself that  $P(Y|\bar{E}) = p$  ( $p$  has no value yet, but is soon to be estimated.)  
Done.

- d) Use the Law of Total Probability to obtain an expression for  $P(Y)$  in terms of  $P(Y|E)$ ,  $P(Y|\bar{E})$ ,  $P(E)$  and  $P(\bar{E})$ :

The Law of Total Probability tells us that

$$P(Y) = P(E)P(Y|E) + P(\bar{E})P(Y|\bar{E})$$

- e) You conduct the survey. 20% of people answered YES, so  $P(Y) \approx 0.20$ . Use this to estimate  $p = P(Y|\overline{E})$ :

$$\begin{aligned} P(Y|\overline{E}) &\approx \frac{P(Y) - P(E)P(Y|E)}{P(\overline{E})} \\ &= \frac{0.20 - (\frac{1}{2})(\frac{2}{7})}{1 - \frac{1}{2}} \\ &= 0.1143 \end{aligned}$$

3. A fisherman gets lost. The probability of him being in one of three regions, is  $P(R_1) = 0.05$  for the first region,  $P(R_2) = 0.70$  for the second region and  $P(R_3) = 0.25$  for the third region. Allowing  $F_i$  to be the event that the fisherman is found in the  $i^{th}$  region, the search effectiveness of each area is given by  $P(F_1|R_1) = 0.50$ ,  $P(F_2|R_2) = 0.80$  and  $P(F_3|R_3) = 0.30$ .

- a) Determine the probability the fisherman is found.

$$\begin{aligned} P(F_1 \cup F_2 \cup F_3) &= P(R_1)P(F_1 \cup F_2 \cup F_3|R_1) + P(R_2)P(F_1 \cup F_2 \cup F_3|R_2) + P(R_3)P(F_1 \cup F_2 \cup F_3|R_3) \\ &= P(R_1)P(F_1|R_1) + P(R_2)P(F_2|R_2) + P(R_3)P(F_3|R_3) \\ &= (0.05)(0.50) + (0.70)(0.80) + (0.25)(0.30) \\ &= 0.66 \end{aligned}$$

- b) Given the fisherman is undetected, what is the probability he is in each region?

$$\begin{aligned} P(R_i|\overline{F_1 \cup F_2 \cup F_3}) &= \frac{P(R_i \cap \overline{F_1 \cup F_2 \cup F_3})}{P(\overline{F_1 \cup F_2 \cup F_3})} \\ &= \frac{P(R_i)P(\overline{F_1 \cup F_2 \cup F_3}|R_i)}{1 - P(F_1 \cup F_2 \cup F_3)} \\ &= \frac{P(R_i)(1 - P(F_1 \cup F_2 \cup F_3|R_i))}{1 - P(F_1 \cup F_2 \cup F_3)} \\ &= \frac{P(R_i)(1 - P(F_i|R_i))}{1 - P(F_1 \cup F_2 \cup F_3)} \end{aligned}$$

Applying this to each specific  $R_i$  we get:

$$P(R_1|\overline{F_1 \cup F_2 \cup F_3}) = 0.0735$$

$$P(R_2|\overline{F_1 \cup F_2 \cup F_3}) = 0.4118$$

$$P(R_3|\overline{F_1 \cup F_2 \cup F_3}) = 0.5147$$

- c) Why is it unnecessary to calculate the probability that the fisherman was found in each region given that the fisherman was found?

If the fisherman has been found, then you will know where, or else you will not have found him. But you did find him, so you must know his location.

4. Consider a healthy woman who has a hemophiliac brother. Neither she nor her husband has hemophilia, but she may be a carrier of the hemophilia gene. It is a long established fact that a healthy woman with a hemophiliac brother has a 50-50 chance of carrying the hemophilia gene. Genetics tells us that, if the mother is a carrier, the probability of her having a healthy boy (health is defined here with respect to hemophilia only) is  $1/2$ . Of course if she is not a carrier, the son is always healthy.

- a) New information appears when this woman gives birth to her first son. Suppose the first son is a healthy child. Determine the probability that she is a carrier.

Let  $H$  be the event that the son is healthy, and let  $C$  be the event that the mother is a carrier. Then we have

$$P(H|C) = 0.5$$

$$P(H|\overline{C}) = 1$$

$$P(C) = P(\overline{C}) = 0.5$$

Note that  $C$  and  $\overline{C}$  partition the sample space, and so

$$P(H) = P(C)P(H|C) + P(\overline{C})P(H|\overline{C})$$

Now we have

$$\begin{aligned}
 P(C|H) &= \frac{P(C \cap H)}{P(H)} \\
 &= \frac{P(H|C)P(C)}{P(C)P(H|C) + P(\overline{C})P(H|\overline{C})} \\
 &= \frac{(0.5)(0.5)}{(0.5)(0.5) + (0.5)(1)} \\
 &= \frac{0.25}{0.75} \\
 &= 0.1875
 \end{aligned}$$

- b) Suppose she has a second, healthy boy. Now what is the probability that she is a carrier.

Now we will replace  $P(C)$  with the result of part a, since we know the woman had a healthy boy already, thus we have

$$\begin{aligned}
 P(C) &= 0.1875 \\
 P(\overline{C}) &= 0.8125 \\
 P(H|C) &= 0.5 \\
 P(H|\overline{C}) &= 1
 \end{aligned}$$

Now we apply these new numbers to the same steps from part a:

$$\begin{aligned}
 \frac{P(H|C)P(C)}{P(C)P(H|C) + P(\overline{C})P(H|\overline{C})} &= \frac{(0.5)(0.1875)}{(0.5)(0.1875) + (0.8125)(1)} \\
 &= \frac{0.09375}{0.90625} \\
 &= 0.1034
 \end{aligned}$$

- c) Continue, for more healthy boys. What happens?

The probability that the mother is a carrier approaches zero asymptotically.

- d) How do things change if she has a hemophiliac son?

She is a carrier. Mathematically speaking,  $P(C) = 1$

5. (This will be a gross oversimplification, but the idea works.) Imagine that Facebook develops an algorithm for assisting with the identification of terrorists. (Mathematically this is not so different from the problem of figuring out whether a Netflix user is likely to enjoy *A Dolphin Tale*. It's more or less the same thing Target does when they cross-reference your lotion and vitamin purchases to estimate how likely it is that you're pregnant. Which Target has done. Which got one young lady found out by her parents.) Facebook has a lot of information about most people, and from that profile they create a predictive model for terrorism.

Rhetorical question: What would you do if you found out a person living near you were on the Facebook suspected terrorist list?

The probability a terrorist is (correctly) placed on this list by Facebook is very small – say 0.001 (a tenth of a percent). Terrorists don't generally behave in obviously different ways (especially with respect to Facebook habits), and identifying them is not easy. The Facebook algorithm isn't bad: The probability a non-terrorist is placed on the list is half of the probability for a terrorist.

In the population, 1 in 20,000 people are terrorist supporters: The probability of randomly choosing such a person when selecting randomly from the entire population of adults in the U.S. is  $1/20000 = 0.00005$ .

- a) Given a person identified by Facebook as a terrorist supporter, what is the probability this person (your neighbor?) really is a terrorist supporter?

Let  $T$  be the event that my neighbor is a terrorist, and let  $F$  be the event that Facebook adds my neighbor to the terrorist list. Then we have

$$\begin{aligned}P(T) &= 0.00005 \\P(\bar{T}) &= 0.99995 \\P(F|T) &= 0.001 \\P(F|\bar{T}) &= 0.0005\end{aligned}$$

Also, note that  $T$  and  $\bar{T}$  partition the sample space, and thus

$$P(F) = P(T)P(F|T) + P(\bar{T})P(F|\bar{T})$$

Now we find  $P(T|F)$ :

$$\begin{aligned} P(T|F) &= \frac{P(T \cap F)}{P(F)} \\ &= \frac{P(T)P(F|T)}{P(T)P(F|T) + P(\bar{T})P(F|\bar{T})} \\ &= \frac{(0.00005)(0.001)}{(0.00005)(0.001) + (0.99995)(0.0005)} \\ &= 9.9995 \times 10^{-5} \\ &\approx 0.00001 \end{aligned}$$

- b) Assuming there are 200 million adults in the U.S.: How many false and true positives are there in the population?

According to the given information, about 0.00005 of people are terrorists, and 0.001 of them are added to Facebook's list. Thus there are

$$200,000,000 \times 0.00005 \times 0.001 = 10$$

true positives. Also, about 0.99995 of people are not terrorists, and 0.0005 of them are added to Facebook's list. Thus there are

$$200,000,000 \times 0.99995 \times 0.0005 = 99995$$

false positives.

It sounds to me like Facebook is really wasting a lot of time and effort in this hypothetical situation.