Lab 4: Probability

Goals

- Get practice solving probability problems

Question #1 A bag has 3 black marbles and 4 white marbles. You reach into the bag and draw a marble, and then draw another marble without replacing the first one.

a. What is the chance that the first draw is a white marble?
For each draw the sample space is {B, W}.
f is the first draw, and s is the second.
We assume that the probability of drawing each of the seven marbles is equal, and four are white so $\Pr(f = \text{white}) = \frac{4}{7}$

b. If the first draw was a white marble, what is the chance that the second draw is a black marble?
s represents the second draw. After a white marble is removed the bag will have 3 black and 3 white marbles. $\Pr(s = \text{black} \mid f = \text{white}) = \frac{3}{6}$

c. What is the chance that both marbles that you have removed are the same color as each other?

For the entire experiment the sample space is made up of four outcomes:
$f = \text{B}$ and $s = \text{B}$
$f = \text{B}$ and $s = \text{W}$
$f = \text{W}$ and $s = \text{B}$
$f = \text{W}$ and $s = \text{W}$

We could say that $S = \{BB, BW, WB, WW\}$.
Note that (in addition to the statements we calculated for parts a and b) we have:

$$\Pr(f = \text{black}) = \frac{3}{7}$$

Outcomes BB and WW are the outcomes that correspond to A. We need to calculate:

$$\Pr(A) = \Pr(BB \text{ or } WW)$$

Note that these events are mutually exclusive.

$$\Pr(A) = \Pr(f = \text{black and } s = \text{black}) + \Pr(f = \text{white and } s = \text{white})$$

$\Pr(f = \text{black and } s = \text{black})$ and $\Pr(f = \text{white and } s = \text{white})$ are joint probabilities – the probability of two simple events both occurring.

We use the general multiplication rule:

$$\begin{aligned}
\Pr(f = \text{black and } s = \text{black}) &= \Pr(f = \text{black}) \Pr(s = \text{black} \mid f = \text{black}) \\
\Pr(f = \text{white and } s = \text{white}) &= \Pr(f = \text{white}) \Pr(s = \text{white} \mid f = \text{white})
\end{aligned}$$

Note that when we draw without replacement, we are altering the population of marbles in the bag, so the probabilities for the second draw will not be identical to the probability statements for the first draw.

We can calculate $\Pr(s = \text{black} \mid f = \text{black})$, because we know that if $f = \text{black}$ then the bag will contain 2 black and 4 white marbles after removing the first marble.

$$\begin{aligned}
\Pr(s = \text{white} \mid f = \text{white}) &= \frac{3}{6} \\
\Pr(s = \text{black} \mid f = \text{black}) &= \frac{2}{6}
\end{aligned}$$

$$\begin{aligned}
\Pr(f = \text{black and } s = \text{black}) &= \Pr(f = \text{black}) \Pr(s = \text{black} \mid f = \text{black}) \\
&= \left(\frac{3}{7}\right)\left(\frac{2}{6}\right) \\
&= \frac{1}{7} \\
\Pr(f = \text{white and } s = \text{white}) &= \Pr(f = \text{white}) \Pr(s = \text{white} \mid f = \text{white}) \\
&= \left(\frac{4}{7}\right)\left(\frac{3}{6}\right) \\
&= \frac{2}{7}
\end{aligned}$$

$$\begin{aligned}
\Pr(A) &= \Pr(BB) + \Pr(WW) \\
&= \frac{1}{7} + \frac{2}{7} \\
&= \frac{3}{7}
\end{aligned}$$

Question #2 In order for an annual plant to flower, it must germinate and then survive for several weeks. A group of botanists estimate that 70% of seeds will germinate in a particular grassland. One quarter of the plants that germinate will survive until they flower.

What is the chance that a randomly selected seed from this grassland will flower?

Sample space = $\{F, \text{not } F\}$ for "flower" and "not flower" respectively.

Event of interest = F.

Probability to calculate = $\Pr(F) = \Pr(G \text{ and } S)$ where G indicates germination, and S is the event that a plant survives after germination until flowering.

Data given:

$$\begin{aligned}
\Pr(G) &= 0.7 \\
\Pr(S \mid G) &= 0.25 \\
\Pr(G \text{ and } S) &= \Pr(G)\Pr(S \mid G) \\
&= 0.7 \times 0.25 \\
&= \frac{7}{40} \\
&= .175
\end{aligned}$$

Question #3 (This is question 19 from the text): DNA is made up of A, C, G, and T. Different chromosomal regions have different frequencies of these nucleotides. Assume that one region has 20% A, 30%C, 30%G, and 20%T. In a second region, the four nucleotides are equally frequent.

a. If you choose a nucleotide randomly from each region, what is the probability that both nucleotides will be the same base?

$S = \{AA, AC, AG, AT, CA, CC, CG, CT, GA, GC, GG, GT, TA, TC, TG, TT\}$

These are joint events. $AC \rightarrow [f = A \text{ and } s = C]$.

Event B (A is a nucleotide, so we won't call our event A). $B = \{AA, CC, GG, TT\}$

$$\Pr(f = A) = \Pr(f = T) = 0.2$$

$$\Pr(f = C) = \Pr(f = G) = 0.3$$

$$\Pr(s = A) = \Pr(s = T) = \Pr(s = C) = \Pr(s = G) = 0.25$$

The events that correspond to the values of f and s are independent events.

$$\begin{aligned}
\Pr(B) &= \Pr(AA \text{ or } CC \text{ or } GG \text{ or } TT) \\
&= \Pr(AA) + \Pr(CC) + \Pr(GG) + \Pr(TT) \\
&= \Pr(f = A)\Pr(s = A) + \Pr(f = C)\Pr(s = C)\ldots \\
&\quad + \Pr(f = G)\Pr(s = G) + \Pr(f = T)\Pr(s = T) \\
&= (.2)(.25) + (.3)(.25) + (.3)(.25) + (.2)(.25) \\
&= .25
\end{aligned}$$

b. Assume that nucleotides occur independently within each region and you random sample a three-nucleotide sequence from each of the 2 regions. What is the chance that these two triples are identical?

Event C is the Event B occurs in three out of three independent trials.

$$\Pr(C) = \Pr(B)\Pr(B)\Pr(B) = \frac{1}{64}$$

Question #4 (Assignment problem # 25 on page 125) A seed randomly blows around a variable habitat: If it lands on high-quality soil it has a 0.8 chance of survival. If it lands on medium-quality soil it has a 0.3 chance of survival. A low-quality soil gives it only a 0.1 chance of survival. The three soil types (high, medium, and low) are present in proportions of 30:20:50, respectively. The probability of landing in a soil type is simply the proportion of the environment that is that habitat type.

a. Draw a probability tree to determine the probabilities of survival in all circumstances.
s =survival
d =death (this is "not survival")
h = high
m = mid
l = low

$$S = \{s \text{ and } h, d \text{ and } h, s \text{ and } m, d \text{ and } m, s \text{ and } l, d \text{ and } l\}$$

$$\begin{aligned}
\Pr(h \mid \text{lands}) &= .3 \\
\Pr(m \mid \text{lands}) &= .2 \\
\Pr(l \mid \text{lands}) &= .5 \\
\Pr(s \mid h) &= .8 \\
\Pr(s \mid m) &= .3 \\
\Pr(s \mid l) &= .1
\end{aligned}$$

(note the first 3 statements add up to 1, the second 3 don't)

First level is habitat types. Second is survival.

If you start to build the tree and find the you don't know the probabilities, try to reverse the order of events.

The probability statements along the branches (after the first level) are conditional probabilities!

b. What is the probability of survival of a seed (assuming that a seed lands)?

$$\begin{aligned}
\Pr(s \mid \text{lands}) &= \Pr(s \text{ and } h) + \Pr(s \text{ and } m) + \Pr(s \text{ and } l) \\
&= \Pr(h)\Pr(s \mid h) + \Pr(m)\Pr(s \mid m) + \Pr(l)\Pr(s \mid l) \\
&= (.3)(.8) + (.2)(.3) + (.5)(.1) \\
&= .35
\end{aligned}$$

c. Assume that a seed has a 0.2 change of dying before it lands. What is the overall probability of survival?

$$\begin{aligned}\Pr(s) &= \Pr(\text{lands})\Pr(s \mid \text{lands})\\ &= (1-.2)(.35)\\ &= (.8)(.35)\\ &= .28\end{aligned}$$

Question #5: Territorial males of a species of fish are more obvious to females and to predators. In a particular stream, 60% of these territorial males will die before the breeding season. Another set of "cryptic" males are smaller and do not defend territories. Only one quarter of these cryptic males will die before the breeding season. Most females prefer to mate with territorial males. If a territorial male is able to survive until breeding season, there is a 70% chance that he will fertilize a clutch. Only half of the cryptic males who survive until the breeding season will fertilize a clutch of eggs. You can assume that a male cannot fertilize more than one clutch and that all males die after the breeding season ends.

(a) If you find a (live) territorial male at the beginning of the **breeding** season (after all of the seasonal mortality events have occurred), what is the probability that he will fertilize a clutch of eggs?

(b) Imagine that you are able to track a cryptic male for his entire lifetime (starting at the beginning of the year **before** any mortality events occur). What is the probability that he would fertilize a clutch of eggs?

(c) CIRCLE ONE: TRUE or FALSE: The events "the male dies before breeding season" and "the male fertilizes a clutch of eggs" are mutually exclusive.

(d) CIRCLE ONE: TRUE or FALSE: The events "the male survives until breeding season" and "the male fertilizes a clutch of eggs" are independent.

Question #6

A disease is found in 8% of the population. There is a diagnostic test for the disease, but the test has a false positive rate of 10% (in other words 10% of healthy people will have positive result for the test), and a false negative rate of 5% (someone with the disease will not have a positive result for the test 5% of the time).

a. Which of the following statements is a correct summary of the data given in the problem (more than one can be correct!):

(a) $\Pr(\text{disease} \mid \text{pos. test}) = 0.08$ WRONG

(b) $\Pr(\text{disease}) = 0.08$ Correct

(c) $\Pr(\text{pos. test} \mid \text{disease}) = 0.1$ WRONG

(d) $\Pr(\text{pos. test} \mid \text{not disease}) = 0.1$ Correct

(e) $\Pr(\text{neg. test} \mid \text{disease}) = 0.1$ WRONG

(f) $\Pr(\text{neg. test} \mid \text{disease}) = 0.05$ Correct

b. We would like to the probability that someone with a positive test result actually has the the disease.

(a) Write down in notational form, the probability that you are being asked to calculate.
$\Pr(\text{disease} \mid \text{pos. test})$

(b) Calculate the probability.

Bayes rule:

$$\Pr(\text{disease} \mid \text{pos. test}) = \frac{\Pr(\text{disease}) \Pr(\text{pos. test} \mid \text{disease})}{\Pr(\text{pos. test})}$$

From the Law of Total Probability (and the fact the "disease" and "not disease" are mutually exclusive and exhaustive):

$$\Pr(\text{pos. test}) = \Pr(\text{disease}) \Pr(\text{pos. test} \mid \text{disease}) + \Pr(\text{not disease}) \Pr(\text{pos. test} \mid \text{not disease})$$

$$\begin{aligned}
\Pr(\text{not disease}) &= 1 - \Pr(\text{disease}) \\
&= .92 \\
\Pr(\text{pos. test} \mid \text{disease}) &= 1 - \Pr(\text{not pos. test} \mid \text{disease}) \\
&= 1 - 0.05 \\
&= .95
\end{aligned}$$

$$\begin{aligned}
\Pr(\text{pos. test}) &= (0.08) \times (.95) + (.92) \times (.1) \\
&= 0.076 + .092 \\
&= 0.168
\end{aligned}$$

So back to Bayes rule:

$$\Pr(\text{disease} \mid \text{pos. test}) = \frac{0.076}{0.168}$$