18.05 Problem Set 2, Spring 2014 Solutions

Problem 1. (10 pts.)

(a) Listing the gender of older child first, our sample space is $\{BB, BG, GB, GG\}$. The event "the older child is a girl" is $\{GB, GG\}$ and the event "both children are girls" is $\{GG\}$. Thus the probability that both children are girls given the older child is a girl is $\frac{1}{2}$.

(b) The event "at least one child is a boy" is $\{BB, BG, GB\}$ so the probability that both children are boys is $\frac{1}{3}$.

Problem 2. (10 pts.) The defense will try to make the case that this is likely to be a case of random mis-identification. So we look for the probability a random taxi the witness sees as blue is actually blue.

This is a question of 'inverting' conditional probability. We know

$$P(\text{the witness sees blue}|\text{the car is blue})$$

but we'd like to know

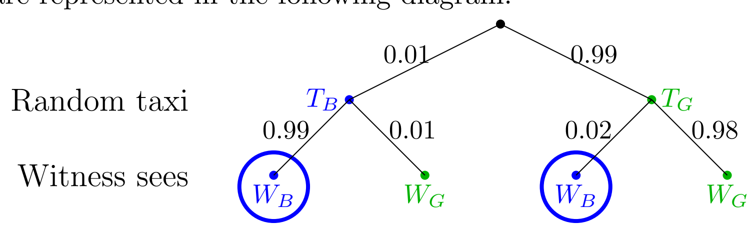
$$P(\text{the car is blue}|\text{the witness sees blue}).$$

Our first job is to translate this to symbols.

Let W_b = 'witness sees a blue taxi' and let W_g = 'witness sees a green car'. Further, let T_b = 'taxi is blue' and let T_g = 'taxi is green'. With this notation we want to find $P(T_b|W_b)$. We will compute this using Bayes' formula

$$P(T_b|W_b) = \frac{P(W_b|T_b) \cdot P(T_b)}{P(W_b)}.$$

All the pieces are represented in the following diagram.



We can determine each factor in the right side of Bayes' formula:
We are given $P(T_b) = .01$ (and $P(T_g) = .99$).
We are given, $P(W_b|T_b) = .99$ and $P(W_b|T_g) = .02$.
We compute $P(W_b)$ using the law of total probability:

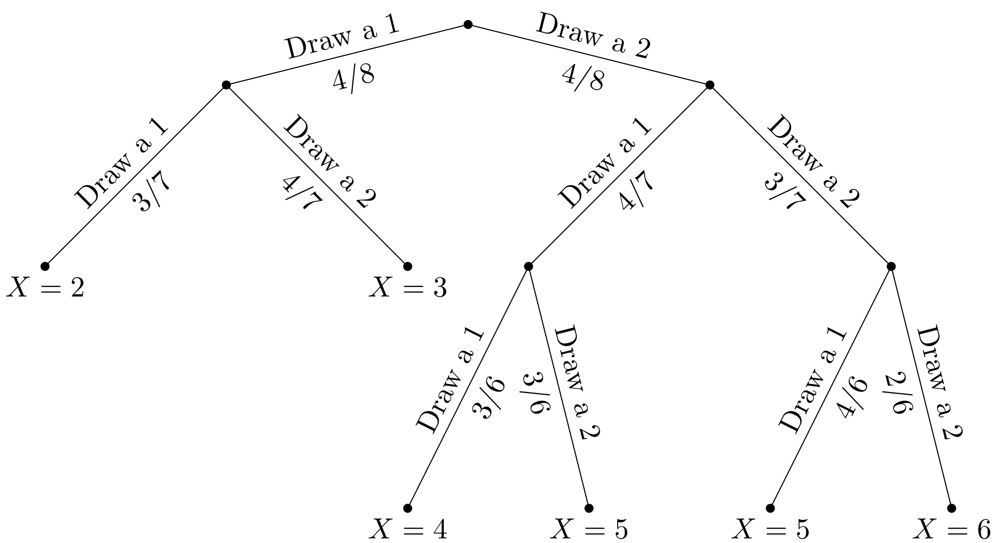
$$P(W_b) = P(W_b|T_b)P(T_b) + P(W_b|T_g)P(T_g) = .99 \times .01 + .02 \times .99 = .99 \times .03.$$

Putting all this in Bayes' formula we get

$$P(T_b|W_b) = \frac{.99 \times .01}{.99 \times .03} = \boxed{\frac{1}{3}}$$

Ladies and gentlemen of the jury. The prosecutor tells you that the witness is nearly flawless in his ability to distinguish whether a taxi is green or blue. He claims that this implies that beyond a reasonable doubt the taxi involved in the hit and run was blue. However probability theory shows without any doubt that the probability a random taxi seen by the witness as blue is actually blue is only 1/3. This is considerably more than a reasonable doubt. In fact it is more probable than not that the taxi involved in the accident was green. If the probability doesn't fit you must acquit!

Problem 3. (10 pts.) The following tree shows all the possible values for X (note, below each edge in the tree is the conditional probability that edge).



At this point we have enough information to compute expectation.

$$E(X) = 2\left(\frac{3}{14}\right) + 3\left(\frac{2}{7}\right) + 4\left(\frac{1}{7}\right) + 5\left(\frac{1}{7}\right) + 5\left(\frac{1}{7}\right) + 6\left(\frac{1}{14}\right) = \frac{26}{7} \approx 3.7143$$

We can also give the probability distribution of X

X	2	3	4	5	6
p(X)	3/14	2/7	1/7	2/7	1/14

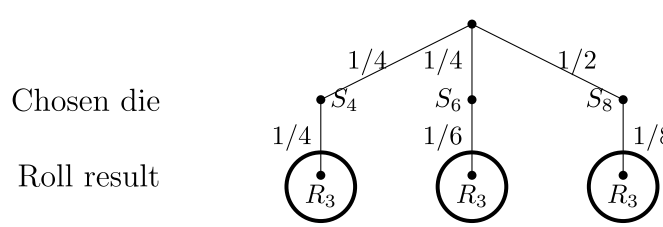
Problem 4. (10 pts.) **(a)**

S	4	6	8
p(X)	1/4	1/4	1/2

(b) Bayes' rule says

$$P(S = k | R = 3) = \frac{P(R = 3 | S = k)P(S = k)}{P(R = 3)}.$$

We summarize what we know in a tree. In the tree the notation S_4 means the 4-sided die ($S = 4$), likewise R_3 means a 3 was rolled ($R = 3$). Because we only care about the case $R = 3$ the tree does not include other possible rolls.



We have the following probabilities (you should identify them in the tree):

$$P(R=3|S=4)=1/4, \quad P(R=3|S=6)=1/6, \quad P(R=3|S=8)=1/8.$$

The law of total probability gives (again, see how the tree tells us this):

$$\begin{aligned} P(R=3) &= P(R=3|S=4)P(S=4)+P(R=3|S=6)P(S=6)+P(R=3|S=8)P(S=8) \\ &= \frac{1}{4}\cdot\frac{1}{4}+\frac{1}{6}\cdot\frac{1}{4}+\frac{1}{8}\cdot\frac{1}{2} \\ &= \frac{1}{6}. \end{aligned}$$

Hence

$$\begin{aligned} P(S=4|R=3) &= \frac{1/16}{1/6} = \boxed{\frac{3}{8}} \\ P(S=6|R=3) &= \frac{1/24}{1/6} = \boxed{\frac{1}{4}} \\ P(S=8|R=3) &= \frac{1/16}{1/6} = \boxed{\frac{3}{8}}. \end{aligned}$$

(c) In a similar vein, we have

$$P(R=6|S=4)=0, \quad P(R=6|S=6)=1/6, \quad P(R=6|S=8)=1/8.$$

and

$$P(R=6) = 0\cdot\frac{1}{4}+\frac{1}{6}\cdot\frac{1}{4}+\frac{1}{8}\cdot\frac{1}{2}=\frac{5}{48}.$$

So,

$$P(S=4|R=6)=0, \quad P(S=6|R=6)=\frac{1/24}{5/48}=\frac{2}{5}, \quad P(S=8|R=6)=\frac{1/16}{5/48}=\frac{3}{5}.$$

The eight-sided die is more likely. Note, the denominator is the same in each probability, i.e. the total probability $P(R=6)$, so all we had to check was the numerator.

(d) The only way to get $R=7$ is if we picked an octahedral die.

Problem 5. (10 pts.) Label the seats 1 to n going clockwise around the table. Let X_i be the Bernoulli random variable with value 1 if the person in seat i is shorter than his or her neighbors. Then $X=\sum_{i=1}^n X_i$ represents the total number of people who are shorter than both of their neighbors, and

$$E(X)=E(\sum_{i=1}^n X_i)=\sum_{i=1}^n E(X_i)$$

by linearity of expected value. Recall that this property of expected values holds even when the X_i are dependent, as is the case here!

Among 3 random people the probability that the middle one is the shortest is 1/3. Therefore $X_i \sim \text{Bernoulli}(1/3)$, which implies $E(X_i) = 1/3$. Therefore the expected number of people shorter than both their neighbors is

$$E(X) = \sum_{i=1}^{n} E(X_i) = \frac{n}{3}.$$

Alternate (and more complicated) solution
Suppose we number the n people based on their height, from shortest to tallest (so person 1 is the shortest and person n is the tallest). For person k, there are $n-k$ people who are taller. In order for person k's neighbors to be taller than person k, we can pick two neighbors from these $n-k$ in $\binom{n-k}{2}$ ways. Moreover, there are $\binom{n-1}{2}$ total ways to pick the neighbor of person k. Thus, the probability that person k is seated between two people taller than him/her is

$$p_k = \frac{\binom{n-k}{2}}{\binom{n-1}{2}}.$$

This formula is valid for $k = 1, 2, \ldots, n-2$ and for $k = n-1$ and $k = n$, we let $p_k = 0$ (since there is no way for the tallest person or the second tallest person to be sitting next to two people taller than themselves).

We can define n Bernoulli random variables, $X_1, \ldots, X_n$, as follows:

$$X_k = \begin{cases} 1 & \text{if person } k\text{'s neighbors are taller than person k} \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $E(X_k) = p_k$ for $k = 1, 2, \ldots, n$. The total number of people who are seated next to people taller than them is $X = X_1 + \cdots + X_{n-2}$, (again we ignore X_{n-1} and X_n since they have to be 0). So, we get, by linearity of expectation

$$E(X) = \sum_{k=1}^{n-2} \frac{\binom{n-k}{2}}{\binom{n-1}{2}}.$$

One can show (after a bit of algebra) that this sum is equal to $n/3$.

Had we ordered the people from tallest to shortest, then we would have

$$E(X) = \sum_{k=3}^{n} \frac{\binom{k-1}{2}}{\binom{n-1}{2}}$$

Problem 6. (10 pts.) **(a)** Any sequence of 50 0's and 1's is valid. However, most people do not put in any long runs that parts (c) and (d) will show happen frequently.

(c) Here is my code with comments

```
nflips = 50
ntrials = 10000
total = 0 # We'll keep a running total of all the trials' longest runs
```

```
for (j in 1:ntrials)
{
  # One trial consists of 50 flips
  trial = rbinom(nflips, 1, .5) # binomial(1,.5) = bernoulli(.5)
  # rle() finds the lengths of all the runs in trials.  We add the max to total
  total = total + max(rle(trial)$lengths)
}
# The average maximum run is the total/ntrials
aveMax = total/ntrials
print(aveMax)
```

My run of this code produced aveMax = 5.9645

(d) Instead of keeping a total we keep a count of the number of trials with a run of 8 or more

```
nflips = 50
ntrials = 10000
# We'll keep a running count of all the trials with a run of 8 or more
count = 0
for (j in 1:ntrials)
{
  trial = rbinom(nflips, 1, .5) # binomial(1,.5) = bernoulli(.5)
  count = count + (max(rle(trial)$lengths) >= 8)
}
# The probability of a run of 8 or more is count/ntrials
prob8 = count/ntrials
print(prob8)
```

My run of this code produced prob8 = 0.1618

18.05 Introduction to Probability and Statistics
Spring 2014