10-601 Machine Learning, Fall 2009: Homework 1 Solutions

Due: Wednesday, September 2^{nd}, 10:30 am

Instructions There are 5 questions on this assignment worth the total of 120 points. The last question involves some basic programming. Please hand in a hard copy at the beginning of the class; your code should be printed and attached to the write-up. Refer to the webpage for policies regarding collaboration, due dates, and extensions.

1 Monty's Haunted House [10 pts]

You are in a haunted house and you are stuck in front of three doors. A ghost appears and tells you "Your hope is behind one of these doors. There is only one door that opens to the outside and the two other doors have deadly monsters behind them. You must choose one door."

You choose the first door. The ghost tells you "Wait! I will give you some more information." The ghost opens the second door and shows you that there was a horrible monster behind it, then asks you "Would you like to change your mind and take the third door instead?"

Which strategy is better: to stick with the first door, or to change to the third door? For each of the following possibilities, determine probabilities that the exit is behind the first and the third door, given that the ghost opened the second door.

1. The ghost uses the same strategy as discussed in class. He always opens a door you have not picked with a monster behind it. If both of the unopened doors hide monsters, he picks each of them with equal probability. [2 pts]

★ **SOLUTION:** We define the following random variables:
O : the event that describes a door opening to outside. O take values from the set {1,2,3} and probability of each door being the exit door is $\mathbf{P}(O=1) = \mathbf{P}(O=2) = \mathbf{P}(O=3) = 1/3$.

G : the event that describes the ghost picking a door after you picked door 1, can take {2,3}. Note that the ghost is not going to open the first door since you picked it. The conditional probability of the ghost opening each door given which door opens to exit, $\mathbf{P}(G \mid O)$, is described in the conditional probability table Tab. 1.

O	G	$\mathbf{P}(G \mid O)$
1	2	1/2
1	3	1/2
2	2	0
2	3	1
3	2	1
3	3	0

Table 1: The conditional probability table $\mathbf{P}(G \mid O)$ for the first strategy.

To be able to decide whether we should switch or not, we need to calculate the conditional probabilities $\mathbf{P}(O \mid G = 2)$ for $O = 1$ and $O = 3$

$$\mathbf{P}(O = 1 \mid G = 2) = \frac{\mathbf{P}(G = 2 \mid O = 1)\, \mathbf{P}(O = 1)}{\mathbf{P}(G = 2 \mid O = 1)\, \mathbf{P}(O = 1) + \mathbf{P}(G = 2 \mid O = 3)\, \mathbf{P}(O = 3)}$$

From the table we look up $\mathbf{P}(G = 2 \mid O = 1)$:

$$\begin{aligned}\mathbf{P}(O = 1 \mid G = 2) &= \frac{1/2 \times 1/3}{1/2 \times 1/3 + 1 \times 1/3} \\ &= \frac{1/2 \times 1/3}{1/2} = 1/3\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbf{P}(O = 3 \mid G = 2) &= \frac{\mathbf{P}(G = 2 \mid O = 1)\, \mathbf{P}(O = 1)}{\mathbf{P}(G = 2 \mid O = 1)\, \mathbf{P}(O = 1) + \mathbf{P}(G = 2 \mid O = 3)\, \mathbf{P}(O = 3)} \\ &= \frac{1 \times 1/3}{1/2 \times 1/3 + 1 \times 1/3} \\ &= \frac{1 \times 1/3}{1/2} = 2/3\end{aligned}$$

Door 3 is more probable ($2/3 > 1/3$) to be the exit door, so it is better to switch.

2. The ghost has a slightly different strategy. If both of the unopened doors hide monsters, he always picks the second door. [4 pts]

★ **SOLUTION:** Based on this new strategy, the conditional probability has changed as in Tab. 2.

O	G	$\mathbf{P}(G \mid O)$
1	2	1
1	3	0
2	2	0
2	3	1
3	2	1
3	3	0

Table 2: The conditional probability table $\mathbf{P}(G \mid O)$ for the second strategy.

We calculate $\mathbf{P}(O \mid G = 2)$ with the second strategy:

$$\begin{aligned}\mathbf{P}(O = 1 \mid G = 2) &= \frac{1 \times 1/3}{1 \times 1/3 + 1 \times 1/3} = 1/2 \\ \mathbf{P}(O = 3 \mid G = 2) &= \frac{1 \times 1/3}{1 \times 1/3 + 1 \times 1/3} = 1/2\end{aligned}$$

Door 1 and door 3 has the exit equally likely, so you can switch or not switch, it would not matter.

3. Finally, suppose that if both of the unopened doors hide monsters, the ghost always picks the third door.[4 pts]

O	G	$\mathbf{P}(G \mid O)$
1	2	0
1	3	1
2	2	0
2	3	1
3	2	1
3	3	0

Table 3: The conditional probability table of $\mathbf{P}(G \mid O)$ for the third strategy.

★ **SOLUTION:** The conditional probability based on the third strategy is given in Tab. 3. We recalculate $\mathbf{P}(O \mid G = 2)$,

$$\mathbf{P}(O = 1 \mid G = 2) = \frac{1 \times 1/3}{0 \times 1/3 + 1 \times 1/3} = 0$$
$$\mathbf{P}(O = 3 \mid G = 2) = \frac{1 \times 1/3}{0 \times 1/3 + 1 \times 1/3} = 1$$

The exit door is definitely door 3, so you should switch.

Common mistake: Forgetting to normalize the probabilities.

2 Medical Testing [15 pts]

There is a disease which affects 1 in 500 people. A $100.00 dollar blood test can help reveal whether a person has the disease. A positive outcome indicates that the person *may* have the disease. The test has perfect sensitivity (true positive rate), i.e., a person who has the disease tests positive 100% of the time. However, the test has 99% specificity (true negative rate), i.e., a healthy person tests positive 1% of the time.

1. A randomly selected individual is tested and the result is positive. What is the probability of the individual having the disease? [5 pts]

★ **SOLUTION:** We define the following random variables:

T_1: the outcome of the first test, takes values {+,-}. $T_1 = +$ indicates test 1 outcome is positive.

T_2^+: the outcome of the first test, takes values {+, -}.

D: the random variable describing whether the person has the disease or not, take values {yes, no}.

We are given the following information:

$$\mathbf{P}(D = \text{yes}) = 1/500$$
$$\mathbf{P}(T_1 = + \mid D = \text{yes}) = 1$$
$$\mathbf{P}(T_1 = + \mid D = \text{no}) = 0.01$$
$$\mathbf{P}(T_2 = + \mid D = \text{yes}) = 1$$
$$\mathbf{P}(T_2 = + \mid D = \text{no}) = 0$$

We want to calculate whether the person has the disease or not given s/he tested positive in the first test, that is $\mathbf{P}(D = \text{yes} \mid T_1 = +)$. First let's calculate the probability of a person testing positive in the first test:

$$\begin{aligned}\mathbf{P}(T_1 = +) &= \mathbf{P}(T_1 = + \mid D = \text{yes})\,\mathbf{P}(D = \text{yes}) + \mathbf{P}(T_1 = + \mid D = \text{no})\,\mathbf{P}(D = \text{no}) \\ &= 1 \times 1/500 + 0.01 \times 499/500 \approx 0.012\end{aligned}$$

Note that we will this probability in 2.2 as well. Now using the Bayes rule again $\mathbf{P}(D = \text{yes} \mid T_1 = +)$:

$$\begin{aligned}
\mathbf{P}(D = \text{yes} \mid T_1 = +) &= \frac{\mathbf{P}(T_1 = + \mid D = \text{yes})\,\mathbf{P}(D = \text{yes})}{\mathbf{P}(T_1 = +)} \\
&= \frac{\mathbf{P}(T_1 = + \mid D = \text{yes})\,\mathbf{P}(D = \text{yes})}{\mathbf{P}(T_1 = +)} \\
&= \frac{1 \times 1/500}{0.012} \\
&\approx 0.1669
\end{aligned}$$

2. There is a second more expensive test which costs \$10,000.00 dollars but is exact with 100% sensitivity and specificity. If we require all people who test positive with the less expensive test to be tested with the more expensive test, what is the expected cost to check whether an individual has the disease? [5 pts]

★ **SOLUTION:** Let c_1 and c_2 be the costs of first and second tests, respectively and C be the random variable which denotes the total cost that needs to be spent to check whether an individual have the disease or not.

$$C = \begin{cases} c_1 & \text{if the person takes the first test only} \\ c_1 + c_2 & \text{if the person takes the first test and the second test} \end{cases}$$

An individual will take the second test only if the first test resulted positive, therefore:

$$\begin{aligned}
\mathbf{P}(C = c_1) &= \mathbf{P}(T_1 = -) \\
\mathbf{P}(C = c_1 + c_2) &= \mathbf{P}(T_1 = +)
\end{aligned}$$

And the expected cost is:

$$\begin{aligned}
\mathbf{E}[C] &= c_1\mathbf{P}(C = c_1) + (c_1 + c_2)\mathbf{P}(C = c_1 + c_2) \\
&= c_1\mathbf{P}(T_1 = -) + (c_1 + c_2)\mathbf{P}(T_1 = +) \\
&= 100(1 - \mathbf{P}(T_1 = +)) + (c_1 + c_2)\mathbf{P}(T_1 = +) \\
&= c_1 - c_1\mathbf{P}(T_1 = +) + c_1\mathbf{P}(T_1 = +) + c_2\mathbf{P}(T_1 = +) \\
&= c_1 + c_2\mathbf{P}(T_1 = +) \\
&= 100 + 10000 \times 0.012 \\
&\approx \$220
\end{aligned}$$

3. A pharmaceutical company is attempting to decrease the cost of the second (perfect) test. How much would it have to make the second test cost, so that the first test is no longer needed? That is, at what cost is it cheaper simply to use the perfect test alone, instead of screening with the cheaper test as described in part 2? [5 pts]

★ **SOLUTION:** The expected cost of the new test should be at most equal to the expected cost of screening with the first test and then using the second test.

$$
\begin{aligned}
c_{new} &\leq E(C) \\
&\leq c_1 \mathbf{P}(C = c_1) + (c_1 + c_{new})\mathbf{P}(C = c_1 + c_2) \\
&\leq c_1(1 - \mathbf{P}(T_1 = +)) + (c_1 + c_{new})\mathbf{P}(T_1 = +) \\
&\leq c_1 - c_1\mathbf{P}(T_1 = +) + c_{new}\mathbf{P}(T_1 = +) + c_{new}\mathbf{P}(T_1 = +) \\
&\leq c_1 + c_{new}\mathbf{P}(T_1 = +) \\
&\leq 100 + c_{new} \times 0.012
\end{aligned}
$$

If we solve for c_new for the boundary case

$$
\begin{aligned}
c_{new} &= 100 + c_{new} \times 0.012 \\
c_{new} &\approx 101.21
\end{aligned}
$$

So if the second new test costs as cheap as \$101.21, there will be no more need the screening with the first test.

Warning: A common misunderstanding was comparing to the new test scenario to the previous scenario where test 2 costs 10,000. That leads to the answer of the expected cost calculated in section 2. We did not deduct points if you did it that way thinking some people might misunderstood the question. However, the correct solution is given above and the answer is \$101.21.

3 Products of Expectations [20 pts]

We showed in class that $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$ if X and Y are independent.

1. Prove that the converse is also true when X and Y are binary, i.e., if X and Y take values in $\{0, 1\}$ and $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$, then X and Y are independent. [10 pts]

★ **SOLUTION:**

$$
\mathbf{E}[X] = \mathbf{P}(X = 0) \times 0 + \mathbf{P}(X = 1) \times 1 = \mathbf{P}(X = 1) \tag{3.1}
$$

$$
\mathbf{E}[Y] = \mathbf{P}(Y = 1) \tag{3.2}
$$

$$
\mathbf{E}[XY] = \mathbf{P}(X = 1, Y = 1) \tag{3.3}
$$

We are given $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$. Therefore,

$$
\begin{aligned}
\mathbf{E}[XY] &= \mathbf{E}[X]\mathbf{E}[Y] \\
\mathbf{P}(X = 1, Y = 1) &= \mathbf{P}(X = 1)\mathbf{P}(Y = 1)
\end{aligned}
$$

In order to show they are independent we need to show $\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$ holds for all four combinations of X, Y values. For case $X = 1, Y = 1$, we already know $\mathbf{P}(X = 1, Y = 1) = \mathbf{P}(X = 1)\mathbf{P}(Y = 1)$

Case $X = 0, Y = 1$

$$
\begin{aligned}
\mathbf{P}(X = 0, Y = 1) &= \mathbf{P}(Y = 1) - \mathbf{P}(X = 1, Y = 1) \\
&= \mathbf{P}(Y = 1) - \mathbf{P}(X = 1)\mathbf{P}(Y = 1) \\
&= \mathbf{P}(Y = 1)(1 - \mathbf{P}(X = 1)) \\
&= \mathbf{P}(Y = 1)\mathbf{P}(X = 0)
\end{aligned}
$$

Case $X = 1, Y = 0$

$$\begin{aligned}
\mathbf{P}(X=1, Y=0) &= \mathbf{P}(X=1) - \mathbf{P}(X=1, Y=1) \\
&= \mathbf{P}(X=1) - \mathbf{P}(X=1)\,\mathbf{P}(Y=1) \\
&= \mathbf{P}(X=1)\,(1 - \mathbf{P}(Y=1)) \\
&= \mathbf{P}(X=1)\,\mathbf{P}(Y=0)
\end{aligned}$$

Case $X = 0, Y = 0$

$$\begin{aligned}
\mathbf{P}(X=0, Y=0) &= \mathbf{P}(Y=0) - \mathbf{P}(X=1, Y=0) \\
&= \mathbf{P}(Y=0) - \mathbf{P}(X=1)\,\mathbf{P}(Y=0) \\
&= \mathbf{P}(Y=0)\,(1 - \mathbf{P}(X=1)) \\
&= \mathbf{P}(Y=0)\,\mathbf{P}(X=0)
\end{aligned}$$

Since for all cases, we were able to show independence, we conclude: if $\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y]$ holds and X, Y both take binary values $\{0, 1\}$ then X and Y are independent from each other.

2. Does the converse hold if X is required to take values in $\{0, 1\}$, but Y can be arbitrary? Prove or give a counterexample. [10 pts]

★ **SOLUTION:** Note it does not hold. Following is a counter example:
Let Y be a random variable which takes values $\{-1, 0, 1\}$. The joint probability table $\mathbf{P}(X, Y)$ is given in Tab. 4.

X	Y	$\mathbf{P}(X,Y)$
0	-1	0.1
0	0	0.4
0	1	0.1
1	-1	0
1	0	0.4
1	1	0

Table 4: The joint probability distribution for $\mathbf{P}(X, Y)$.

$\mathbf{E}[XY] = \sum_{xy} xy\mathbf{P}(X=x, Y=y) = 1 \times 1 \times \mathbf{P}(X=1, Y=1) + 1 \times -1 \times \mathbf{P}(X=1, Y=-1) = 1 \times 1 \times 0 + 1 \times -1 \times 0 = 0$ (notice all other pairs have X=0 or Y=0)

$$\begin{aligned}
\mathbf{P}(Y=-1) &= 0.1 \\
\mathbf{P}(Y=0) &= 0.8 \\
\mathbf{P}(Y=1) &= 0.1
\end{aligned}$$

$\mathbf{E}[Y] = \sum_y y\mathbf{P}(Y=y) = -1 \times \mathbf{P}(Y=-1) + 1 \times \mathbf{P}(Y=1) + 0 \times \mathbf{P}(Y=0) = -0.1 + 0.1 = 0$

Therefore $\mathbf{E}[XY] = \mathbf{E}[Y] = 0$.
Notice $\mathbf{P}(X=0) = 0.6$, $\mathbf{P}(Y=-1) = 0.1$, however $\mathbf{P}(X=0, Y=-1) = 0.1$
However $\mathbf{P}(X=0, Y=-1) \neq \mathbf{P}(X=0)\,\mathbf{P}(Y=-1)$ since

4 Balls and Bins [15 pts]

Suppose we have n bins and m balls. We throw balls into bins independently at random, so that each ball is equally likely to fall into any of the bins.

1. What is the probability of the first ball falling into the first bin? [2 pts]

   ★ **SOLUTION:** It is equally likely for a ball to fall in any of the bins, so the probability that first ball falling into the first bin is $\frac{1}{n}$.

2. What is the expected number of balls in the first bin? [3 pts]

   Hint 1: Define an indicator random variable representing whether the i-th ball fell into the first bin:

$$X_i = \begin{cases} 1 & \text{if } i\text{-th ball fell into the first bin} \\ 0 & \text{otherwise.} \end{cases}$$

   Hint 2: Use linearity of expectation.

   ★ **SOLUTION:** Define X_i as described in Hint 1. Let Y be the total number of balls that fall into first bin: $Y = X_1 + \ldots + X_m$. The expected number of balls:

$$\mathbf{E}[Y] = \sum_{i=1}^{m} \mathbf{E}[X_i] = \sum_{i=1}^{m} \mathbf{P}(X_i) \times 1 = m \times \frac{1}{n} = \frac{m}{n}$$

3. What is the probability that the first bin is empty? [5 pts]

   ★ **SOLUTION:** Let Y and X_i be the same as defined in 4.2. For the first bin to be empty none of the balls should fall into the first bin, $Y = 0$.
   $\mathbf{P}(Y = 0) = \mathbf{P}(X_1 = 0, \ldots, X_m = 0) = \prod_{i=1}^{m} \mathbf{P}(X_i = 0) = (1 - \frac{1}{n})^m = (\frac{n-1}{n})^m$

4. What is the expected number of empty bins? [5 pts]

   Hint 3: Define an indicator for the event "bin j is empty" and use linearity of expectations.

   ★ **SOLUTION:** For each n bins we define an indicator random variable for the event "bin j is empty".

$$Y_j = \begin{cases} 1 & \text{if } j\text{-th bin is empty} \\ 0 & \text{otherwise.} \end{cases}$$

   Let Z be the random variables denoting the number of empty bins. $Z = Y_1 + \ldots + Y_m$. Then the expected number of empty bins:

$$\mathbf{E}[Z] = \mathbf{E}\left[\sum_{j=1}^{n} Y_j\right] = \sum_{j=1}^{m} \mathbf{E}[Y_j] = \sum_{j=1}^{n} \mathbf{P}(Y_j) \times 1 = \sum_{j=1}^{n} (\frac{n-1}{n})^m = n(\frac{n-1}{n})^m$$

5 Tacky Programming [60 points]

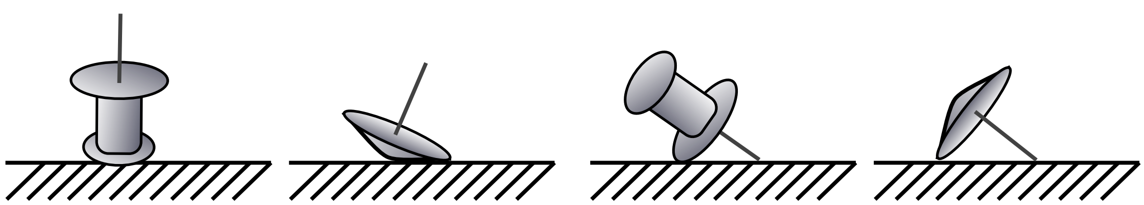
In this problem you will have an opportunity to become familiar with your numerical computing environment. We highly recommend that you use Matlab and will only provide help with Matlab code. You may obtain Matlab from `http://www.cmu.edu/computing/software/all/matlab/index.html` or by using many of the facilitized systems on campus. For this problem you will need to download the data from `http://www.cs.cmu.edu/~ggordon/10601/hws/hw1/hw1data.zip`. For all plots, please include titles and axis labels (see Matlab commands `xlabel`, `ylabel`, and `title`). Please include all code with your homework submission.

★ **SOLUTION:** I posted matlab solution code for all parts at `http://www.cs.cmu.edu/~ggordon/10601/hws/hw1/code_soln.m`.

Problem Setup: A fun game, among graduate students, is tossing thumbtacks[1] to see if they land facing up or facing down as seen in Fig. 1. We can treat the orientation of a thumbtack as random variable Y which can either be *up* or *down*. To make the game more interesting, a friend of mine commissioned a multinational team of unbiased experimental physicists to build the perfect set of 100 uniquely tuned thumbtacks. For each thumbtack $X \in \{1, \ldots, 100\}$ the probability that the thumbtack will land facing up is given by:

$$\mathbf{P}(Y = \text{up} \mid X = x) = \frac{x}{100} \tag{5.1}$$

Hence, thumbtack 1 lands facing up with probability 0.01 and thumbtack 100 lands facing up with probability 1. After cleaning out my office, I filled three jars, $J \in \{1, 2, 3\}$, with varying amounts of each type of thumbtack. I then thoroughly and meticulously shook each jar. You can find the counts of each type of thumbtack for all three jars in `jars.csv`. The first column is the thumbtack id X, and the remaining 3 columns contain the counts for each jar.



(a) Up Thumbtacks (b) Down Thumbtacks

Figure 1: Thumbtacks landing up and down

5.1 Looking at the Data [10 pts]

To familiarize yourself with the problem use the `bar` function in Matlab to plot:

1. the *probability* $Y =$ up for each type of thumbtack

   ★ **SOLUTION:** See Fig. 2(a) for plot.

2. the *probability* of drawing each type of thumbtack from jar 1

   ★ **SOLUTION:** See Fig. 2(b) for plot.

3. the *probability* of drawing each type of thumbtack from jar 2

[1]Careful! Thumbtacks are believed to be sharp.

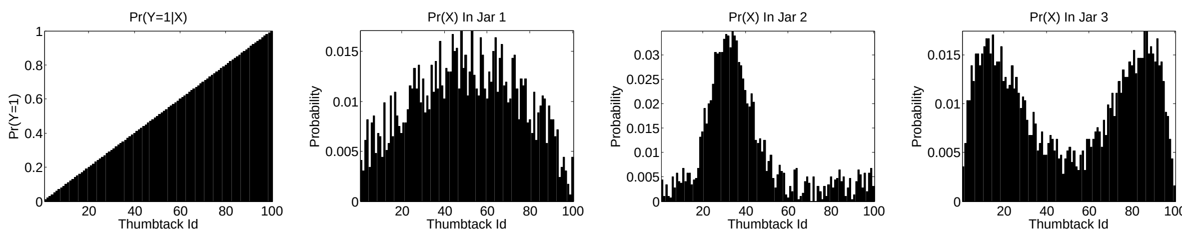
★ **SOLUTION:** See Fig. 2(c) for plot.

4. the *probability* of drawing each type of thumbtack from jar 3

★ **SOLUTION:** See Fig. 2(d) for plot.

To save paper, please consider plotting several plots on the same page (see `help subplot` for how to do this in Matlab).

★ **SOLUTION:** This question was intended to help you become familiar with Matlab. You needed to load in the data file and normalize the counts. The resulting plots are all in Fig. 2.



(a) 5.1.1 (b) 5.1.2 (c) 5.1.3 (d) 5.1.4

Figure 2: Plots for question 5.1

5.2 The Likelihood of Multiple Observations [5 pts]

Using the `bar` function in Matlab, plot the probability of obtaining the sequence (u, u, d, u, d) of random tosses for each type of thumbtack.

★ **SOLUTION:** To solve this we must first define a likelihood function. There are two reasonable answers to this question. If we assume that I am interested in the order of the sequence then we obtain:

$$\mathbf{P}\left(Y = (u, u, d, u, d) \,|\, X = i\right) = \left(\frac{i}{100}\right)^3 \left(1 - \frac{i}{100}\right)^2,$$

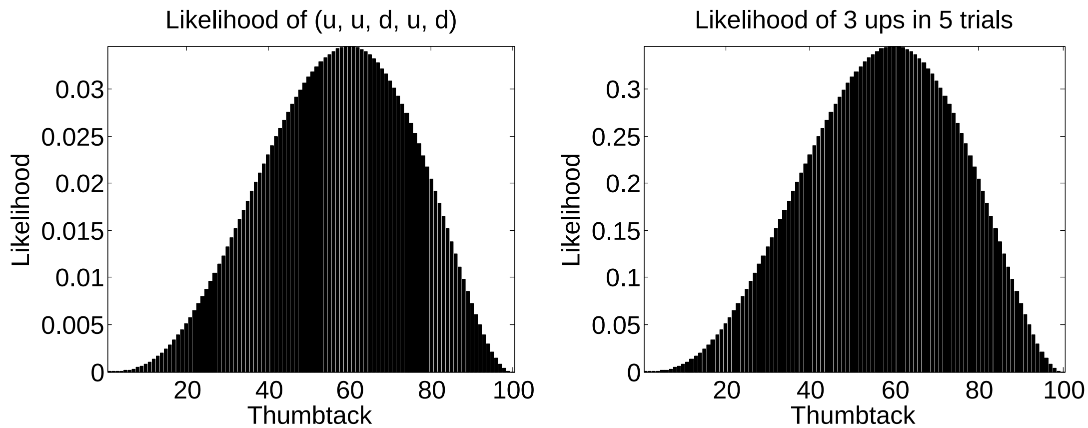
which is plotted in Fig. 3(a). In practice we probably care more about the probability of the event in which 3 out of 5 flips landed facing up. This leads to the similar likelihood function (also known as the Binomial Likelihood):

$$\mathbf{P}\left(\text{3 up in 5 trials} \,|\, X = i\right) = \binom{5}{3} \left(\frac{i}{100}\right)^3 \left(1 - \frac{i}{100}\right)^2,$$

which is plotted in Fig. 3(b). In both cases, the shape of the likelihood function remains the same.

5.3 Writing Bayes Rule [5 pts]

Suppose I give you $\mathbf{P}(Y = \text{up} \,|\, X)$ and $\mathbf{P}(X)$. Write down equations I would use to compute $\mathbf{P}(X \,|\, Y = \text{down})$ and $\mathbf{P}(X \,|\, Y = \text{up})$.



(a) (u, u, d, u, d) (b) 3 out of 5 flips land up

Figure 3: Plots for question 5.2

★ **SOLUTION:** This is a straight forward application of Bayes Rule:

$$\mathbf{P}(X \mid Y) = \frac{\mathbf{P}(Y \mid X)\,\mathbf{P}(X)}{\mathbf{P}(Y)} = \frac{\mathbf{P}(Y \mid X)\,\mathbf{P}(X)}{\sum_{i=1}^{100} \mathbf{P}(Y, X = i)} = \frac{\mathbf{P}(Y \mid X)\,\mathbf{P}(X)}{\sum_{i=1}^{100} \mathbf{P}(Y \mid X = i)\,\mathbf{P}(X = i)} \tag{5.2}$$

Thus we can easily compute both cases as follows:

$$\mathbf{P}(X \mid Y = \text{down}) = \frac{(1 - \mathbf{P}(Y = \text{up} \mid X))\,\mathbf{P}(X)}{\sum_{i=1}^{100} (1 - \mathbf{P}(Y = \text{up} \mid X = i))\,\mathbf{P}(X = i)}$$

$$\mathbf{P}(X \mid Y = \text{up}) = \frac{\mathbf{P}(Y = \text{up} \mid X)\,\mathbf{P}(X)}{\sum_{i=1}^{100} \mathbf{P}(Y = \text{up} \mid X = i)\,\mathbf{P}(X = i)}$$

5.4 Posterior Dependence on Number of Observations [20 pts]

Suppose I randomly select a thumbtack from the first jar $J = 1$. For each of the following scenarios, plot the posterior distribution over thumbtack types given the outcomes of the tosses in that scenario

1. I toss the thumbtack only once and it lands facing down.

   ★ **SOLUTION:** See Fig. 4(a).

2. I toss the thumbtack 3 times and each time it lands facing down.

   ★ **SOLUTION:** See Fig. 4(b).

3. I toss the thumbtack 5 times and 3 times it lands facing up and 2 times it lands facing down.

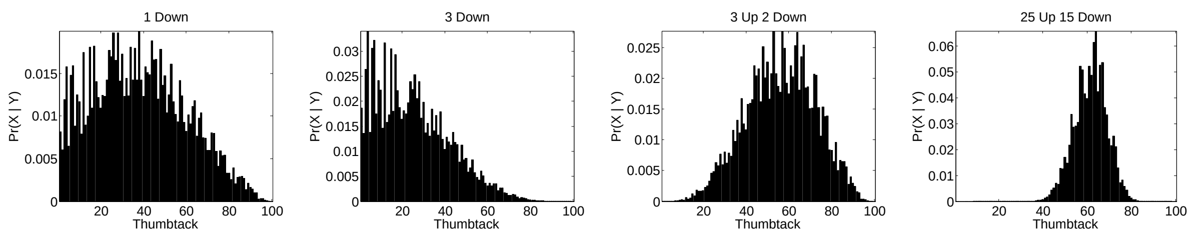
★ **SOLUTION:** See Fig. 4(c).

4. I toss the thumbtack 40 times and 25 times it lands facing up and 15 times it lands facing down.

★ **SOLUTION:** See Fig. 4(d).

Based on the last (largest) set of tosses, which thumbtack type X do you think I selected? How does increasing the number of observations affect the posterior distribution? How does the change in posterior distribution relate to your confidence in the prediction of which thumbtack type was used?

★ **SOLUTION:** The purpose behind this problem was to practice using the Bayesian method of incorporating observations into a prior to predict a posterior. Here you should have used ideas from questions 5.2 and 5.3 to compute the posterior.



(a) 5.4.1 (b) 5.4.2 (c) 5.4.3 (d) 5.4.4

Figure 4: Plots for question 5.4

With very few observations we see that the posterior is closely related to the prior. As we increase the number of observations we quickly "forget" the prior and our distribution becomes increasingly peaked around the thumbtacks that are most likely with respect to the observations. Effectively, by increasing the number of observations we become more confident in our estimate. In 5.5.4 we find that thumbtack number 64 is the most likely.

5.5 Marginalizing The Unobserved [20 pts]

Suppose I randomly select a jar with probability $\mathbf{P}(J=1)=0.25$, $\mathbf{P}(J=2)=0.25$, and $\mathbf{P}(J=3)=0.50$. Then from the chosen jar, I randomly draw a thumbtack. I then toss the thumbtack 10 times and it lands facing up 8 times and facing down 2 times.

1. Plot the posterior distribution over thumbtack types given the outcome of the tosses.

★ **SOLUTION:** See Fig. 5(a) for the plot. To compute the posterior over the thumbtack we needed to marginalize out the jar J random variable. We can write the joint distribution of Y, J, and X as:

$$\mathbf{P}(Y,X,J)=\mathbf{P}(Y\,|\,X)\,\mathbf{P}(X\,|\,J)\,\mathbf{P}(J)$$

By the definition of a conditional we have:

$$\mathbf{P}(X\,|\,Y)=\frac{\mathbf{P}(Y\,|\,X)\sum_{j=1}^{3}\mathbf{P}(X\,|\,J=j)\,\mathbf{P}(J=j)}{\sum_{x=1}^{100}\mathbf{P}(Y\,|\,X=x)\sum_{j=1}^{3}\mathbf{P}(X=x\,|\,J=j)\,\mathbf{P}(J=j)}$$

Alternatively, we could simply construct a new prior on X as:

$$\mathbf{P}(X) = \sum_{j=1}^{3} \mathbf{P}(X, J = j) = \sum_{j=1}^{3} \mathbf{P}(X|J = j)\,\mathbf{P}(J = j)$$

and then applying Bayes rule as before except using the new prior:

$$\mathbf{P}(X \mid Y) = \frac{\mathbf{P}(Y \mid X)\,\mathbf{P}(X)}{\sum_{x=1}^{100} \mathbf{P}(Y \mid X = x)\,\mathbf{P}(X = x)}$$

2. Plot the posterior distribution over the jars given the outcome of the tosses.

★ **SOLUTION:** See Fig. 5(b) for the plot. Working directly from the definition of the joint and constructing the conditional we obtain:

$$\mathbf{P}(J \mid Y) = \frac{\sum_{x=1}^{100} \mathbf{P}(Y \mid X = x)\,\mathbf{P}(X = x \mid J)\,\mathbf{P}(J)}{\sum_{j=1}^{3} \sum_{x=1}^{100} \mathbf{P}(Y \mid X = x)\,\mathbf{P}(X = x \mid J = j)\,\mathbf{P}(J = j)}$$

Alternatively, we could construct a new likelihood by marginalizing out the dependence on the thumbtack:

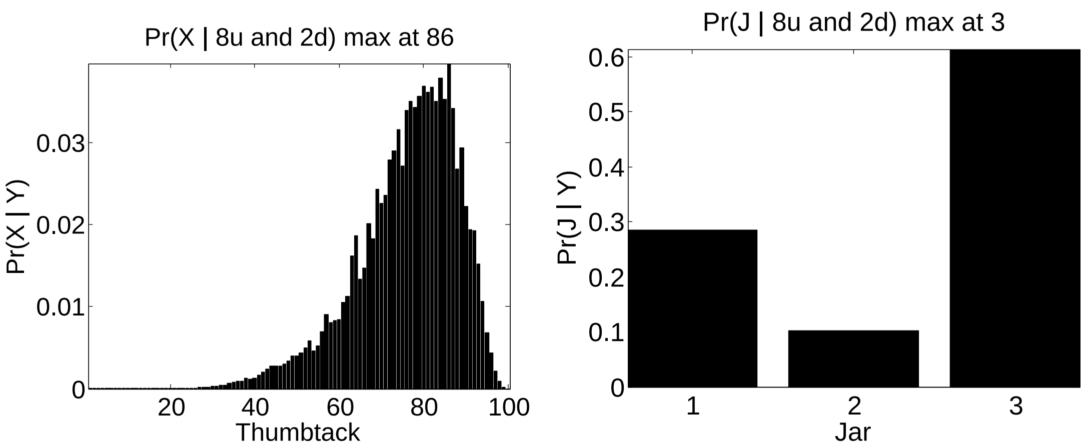
$$\mathbf{P}(Y \mid J) = \sum_{x=1}^{100} \mathbf{P}(Y \mid X = x)\,\mathbf{P}(X = x \mid J)$$

and then applying bayes rule:

$$\mathbf{P}(J \mid Y) = \frac{\mathbf{P}(Y \mid J)\,\mathbf{P}(J)}{\sum_{j=1}^{3} \mathbf{P}(Y \mid J = j)\,\mathbf{P}(J = j)}$$

3. Which thumbtack type attains the largest posterior? Which jar attains the largest posterior?

★ **SOLUTION:** The thumbtack type 86 obtains the largest posterior and jar 3 obtains the largest posterior.



(a) 5.5.1

(b) 5.5.2

Figure 5: Plots for question 5.5