CSCI 5822 Assignment 3

Evan Sidrow

February 22, 2018

Part 1

When flipped, a coin comes up heads with probability θ, which we treat as a random variable that we do not observe. We have a uniform prior for θ, and observe 7 different coin flips. Using this information:

$$P(\theta|D) = \frac{P(D|\theta) * P(\theta)}{Z}$$
$$P(\theta|D) = \frac{(N_H + N_T)!}{N_H! N_T!} \frac{\theta^{N_H}(1-\theta)^{N_T}}{Z}$$

Where N_H is the number of heads observed and N_T is the number of tails observed. But note that this is a Beta distribution, so $Z = 1$, and

$$P(\theta|D) \sim Beta(N_H + 1, N_T + 1)$$

Assume that we observed the sequence $[H, T, T, H, T, T, T]$. Then, our posterior after each trial is shown Figure 1:

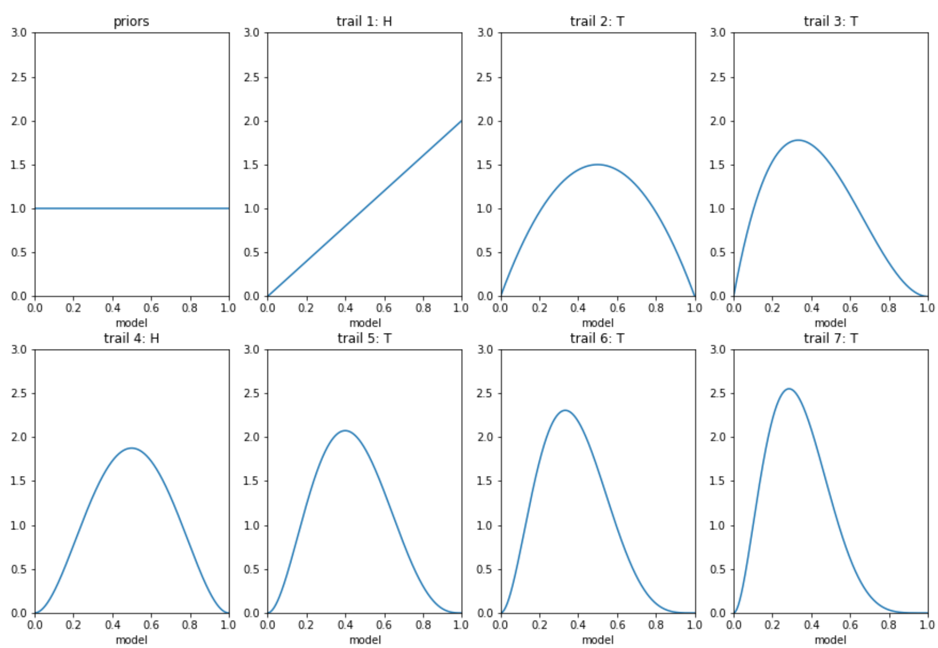

Figure 1: Posterior Distributions for θ after each Observation of a Coin Flip

Figure 1 was made using the python code below:

```
import matplotlib.pyplot as plt
from scipy import stats
import numpy as np

# Part 1

rows = 2
cols = 4
alpha = 1
beta = 1
xs = np.arange(0.0,1.0,0.001)
outcomes = [1,0,0,1,0,0,0]
plot_num = 100*rows + 10*cols + 1
plt.subplots(rows,cols,figsize=(15,10))
plt.subplot(plot_num)
plt.plot(xs,stats.beta.pdf(xs,alpha,beta))
plt.xlabel("model")
plt.title("priors")
plt.ylim([0,3])
plt.xlim([0,1])

for trail,outcome in enumerate(outcomes):

    alpha += outcome
    beta += 1-outcome
    plot_num += 1
    plt.subplot(plot_num)
    plt.plot(xs,stats.beta.pdf(xs,alpha,beta))
    plt.xlabel("model")
    plt.ylim([0,3])
    plt.xlim([0,1])

    if outcome:
        plt.title(str("trail " + str(trail + 1) + ": H"))
    else:
        plt.title(str("trail " + str(trail + 1) + ": T"))

plt.savefig('task1.png')
plt.show()
```

Part 2

Task 1

For each of the examples, I assumed that there was independent Gaussian noise in the intensity of each pixel with a variance of $\sigma^2 = 1$. Therefore:

$$p(x, y|v) = \exp\left(-\frac{(I_1(x, y) - I_2(x + v_x, y + v_y))^2}{2\sigma^2}\right)$$

$$log(p(x, y|v)) = -\frac{(I_1(x, y) - I_2(x + v_x, y + v_y))^2}{2\sigma^2}$$

where x and y represent the intensity of pixel (x, y). To get the total likelihood, we must sum over all pixels:

$$log(p(I_1, I_2|v)) = -\sum_{x,y \in S} \frac{(I_1(x, y) - I_2(x + v_x, y + v_y))^2}{2\sigma^2} \tag{1}$$

Where I_1 and I_2 are the intensities of images 1 and 2, respectively, and S is the set of all pixels such that pixel $(x+v_x, y+v_y)$ can still be observed in the image. This means that if a pixel's velocity carries it out of the frame, or that a new pixel from out of frame comes into frame, we treat its likelihood as 1.

Using equation (1) with $\sigma^2 = 1$ and velocities $v_x \in \{-2, -1, 0, 1, 2\}$ and $v_y \in \{-2, -1, 0, 1, 2\}$, we get the results in Figure 2. ***Please note the scale of the color bar for each plot.***

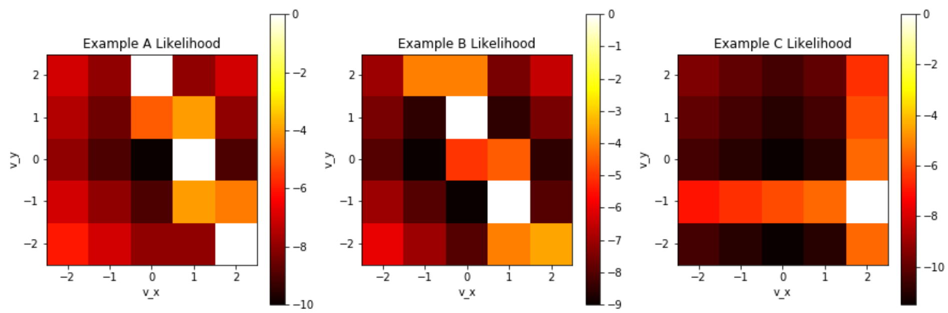

Figure 2: Likelihoods for Several Different Velocities

Task 2

Now, we incorporate the small-motion-bias prior:

$$P(v) \propto \exp\left(-||v||^2/2\sigma_p^2\right)$$
$$log(P(v)) = -||v||^2/2\sigma_p^2 + Z$$

Where Z is some normalizing constant.

Now, we can find $P(v|I_1, I_2) \propto P(v)P(I_1, I_2|v)$ by simply adding the log-prior to the log-likelihood. When doing this, we reduce the probability of velocities with larger magnitudes. Using $\sigma_p^2 = 2$ (since $\sigma_p^2 = 2$ makes the ratio $\sigma^2/\sigma_p^2 = 0.5$), we get the results shown in Figure 3. ***Please note the scale of the color bar for each plot.***

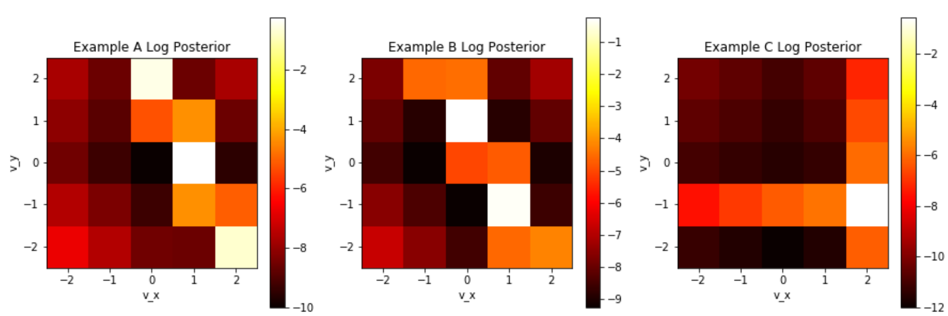

Figure 3: Unscaled Log Posteriors for Several Different Velocities

Task 3

Now, we can take all of the log-posterior values, exponentiate them, and then normalize so they add up to one. After doing this, we get the results shown in Figure 4. Both natural probabilities and log-probabilities are shown. ***Please note the scale of the color bar for each plot.***

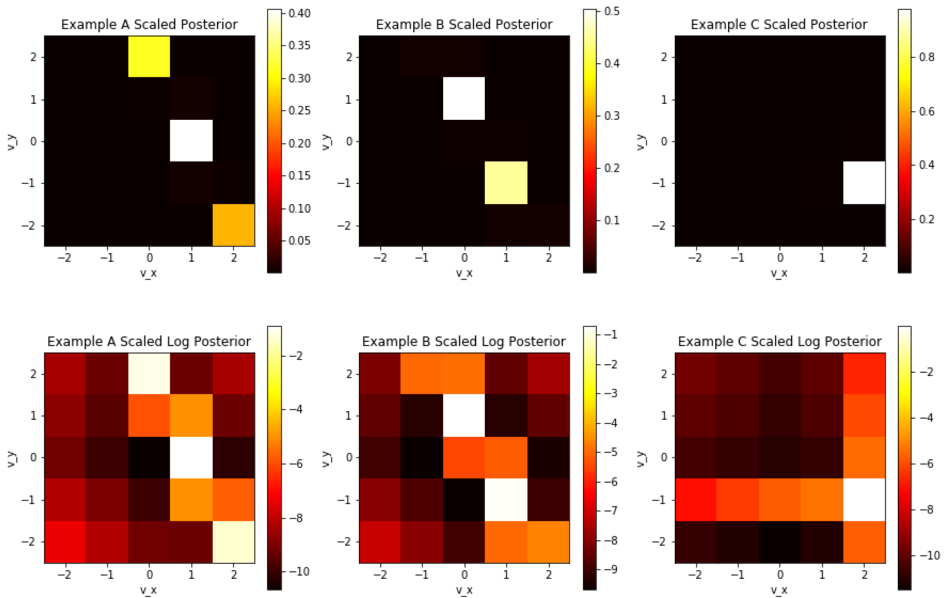

Figure 4: Scaled Posteriors for Several Different Velocities

Task 4

When estimating the true velocity for each of these examples, we could use a maximum likelihood solution or a maximum a posteriori solution. If we used a maximum likelihood solution, we would make no assumptions about which velocities are more likely beforehand. Also, there are **multiple** maximum likelihood solutions for examples A and B. However, if we use the MAP solution for each example, we utilize some prior knowledge in our estimate. Therefore, we pick a **unique** solution that has a high likelihood **and** is slow (since we believe that slow velocities are more likely).

Part 3

Now, let us look at an object labeled with a blue square and red triangle. It is moving one unit at a time- either up, down, left or right. However, our observations are noisy, and we observe the x and y components of velocity for the blue square and red triangle, plus some Gaussian noise.

It is our job to compute $P(v|R_x, R_y, B_x, B_y)$. Note, then:

$$P(v|R_x, R_y, B_x, B_y) \propto P(R_x, R_y, B_x, B_y|v) * P(v)$$

But since R_x, R_y, B_x, and B_y are independent GIVEN v:

$$P(v|R_x, R_y, B_x, B_y) \propto P(R_x|v)P(R_y|v)P(B_x|v)P(B_y|v)P(v)$$

and R_x, R_y, B_x, and B_y given the direction of travel are all normal random variables with $\mu = v_x$ for R_x and B_x and $\mu = v_y$ for R_y and B_y. σ^2 is to be determined in the tasks below.

So all we need to do is define some prior and σ^2, and we can calculate $P(v|R_x, R_y, B_x, B_y)$ up to a constant. Then, we simply need to normalize, and we are done.

For all of the following tasks, assume $R_x = 0.75, R_y = -0.6, B_x = 1.4$, and $B_y = -0.2$.

Task 1

Suppose we have a uniform prior over all directions ($P(v) = 0.25$), and we set $\sigma = 1$. Then we get:

$P(up)$	$P(down)$	$P(left)$	$P(right)$
0.0395	0.1956	0.0102	0.7546

This makes sense. We have an uninformative prior, and it is clear that R_x and B_x are indicating a move the right while R_y and B_y are indicating a move down. However, R_x and B_x are larger in absolute value, so it is more likely that we have a move to the right.

Task 2

Suppose we have a uniform prior over all directions ($P(v) = 0.25$), and we set $\sigma = 5$. Then we get:

$P(up)$	$P(down)$	$P(left)$	$P(right)$
0.2416	0.2576	0.2289	0.2719

This also makes sense. We see the same trends as those for Task 1, but since σ is so large, we do not trust the data nearly as much, so all probabilities are much closer to those of the prior.

Task 3

Suppose now that our prior says that down is 5 times more likely than any other direction. Then, $P(down) = 0.625$, while $P(v) = 0.125$ for all other directions. Let us also set $\sigma = 1$. Then we get:

$P(up)$	$P(down)$	$P(left)$	$P(right)$
0.0222	0.5487	0.0057	0.4234

For this table, we see that the prior makes the probability of "down" much higher, but the likelihood still makes the probability of "right" relatively high. As a result, we have about a 50/50 shot of either direction.

Task 4

Suppose now that our prior says that down is 5 times more likely than any other direction. Then, $P(down) = 0.625$, while $P(v) = 0.125$ for all other directions. Let us also set $\sigma = 5$. Then we get:

$P(up)$	$P(down)$	$P(left)$	$P(right)$
0.1190	0.6343	0.1127	0.1339

For this table, we see that the prior makes the probability of "down" much higher, and the high variance in our likelihood function means that we don't trust the data to tell us much. Therefore, we will continue to trust our prior for the most part, and our posterior probabilities are close the the prior probabilities.