Write-up of Project 1 | Kate Le | Due: Sep 28

What I found in scenario 1:

$$\ominus = \frac{k}{n}$$

The Maximum Likelihood Estimation works well in scenario 1 because the probabilities of 1-6 computed from 3 datasets are somewhat close to the loaded dice's prior distribution. Furthermore, as the dataset grows bigger from A to C, the probabilities become more correct and closer to prior distribution (probability of 6 gets closer to 50% and probabilities of 1-5 get closer to 10%) This suggests that MLE performs well with larger data.

How I compute the probability of 1-6 in scenario 1 with MLE: count how many times 1,2,3,4,5,6 appear in each dataset or their frequency, then divide their frequency by the size of the dataset.

My solution to the problem with very small values in scenarios 2,3, and 4:
In scenarios 2,3, and 4, the computed probabilities of 6 given different thetas are very small thus hard to detect. The relative answer is also what I care more about. Therefore, I multiplied these values by some constants to detect the max value among them. Specifically, I multiplied the values by 10^8 for dataset A, 10^{70} for dataset B, and 10^{300} for dataset C. While this helped to distinguish the values in dataset A and B, it did not in dataset C because the values in C are so close to zero that they cannot be distinguished from zero. Therefore, I think it's a good idea to just say Bayesian doesn't work for dataset C and I will mainly discuss my findings in three scenarios using datasets A and B.

What I found in scenario 2:

$$p(\ominus|D) = \frac{p(D|\ominus) * p(\ominus)}{p(D)}$$

In scenario 2, we assume that any theta has the same probability so I use different theta ranging from 0.1 to 0.9. It turned out that the probabilities of 6 given theta 0.4 or 0.5 are the highest. In A, 6 has the highest probability when theta=0.4. It's also interesting to notice that the probability of 6 given theta=0.5 is the second highest. In B, 6 has the highest probability when theta=0.5. It shows that as datasets grow bigger, the Bayesian method shows that the dice have a 50% chance of showing 6 like what we know before. I think the highest probability in C also follows this trend when theta=0.5, but the values are so small that we cannot see it. In summary, the Bayesian method in scenario 2 somewhat yields the same answer as the MLE in scenario 1.

How I compute the probability of 6 in scenario 2 with Bayesian: Since any $p(\ominus)$ is uniform, $p(D)$ is impossible to know, and $p(\ominus|D)$ is proportional to $p(D|\ominus)$, I will compute probability of 6 by computing $p(D|\ominus)$

What I found in scenarios 3 and 4:

$$p(\ominus|D) \;=\; \frac{p(D|\ominus) * p(\ominus)}{p(D)}$$

In scenarios 3 and 4 when the probability of theta=0.5 is high, the probability of 6 in datasets A and B are high as well. It shows that the probability of theta affects the final result in some significant way. However, the resulting probability of 6 given theta=0.5 is still far greater than when theta != 0.5. Even if other thetas have a higher probability, the probability of 6 given theta=0.5 still outweighs. It follows that the Bayesian method in these scenarios is doing well in predicting the probability of 6 because the results are close to the prior distribution of the loaded dice. However, with larger datasets, MLE is the better choice compared to Bayesian.

*How I compute the probability of 6 in scenarios 3 & 4 with Bayesian: Since any $p(\ominus)$ is given, $p(D)$ is impossible to know, and $p(\ominus|D)$ is proportional to $p(D|\ominus)$, I will compute probability of 6 by computing $p(D|\ominus)$ * $p(\ominus)$*