Homework 1

Bayesian Data Analysis

Problem 1.

In class, I have asked 5 students if they liked more a picture of a cat or of a dog. Three students liked the picture of the dog.

We are interested in studying the proportion, say θ, of invividuals in the population who may like the picture of the dog [to make this more interesting, you can subsitute the pictures of the cat and the dog with a Facebook survey, which is shown to a particular subsample of all Facebook users].

We assume that the students' responses are exchangeable (for now, we can consider them as i.i.d.).

Similarly as in class, consider the following prior distribution for the proportion:

θ	0.3	0.4	.5	.6	.7
$p(\theta)$	.05	.05	.8	.05	.05

(1) Compute the posterior distribution $p(\theta|x = 3)$.

(2) Compute the posterior mean $E(\theta|x = 3)$ which can be assumed as a possible estimator of the population proportion, based on the observed data.

(3) I then asked seven more students about their preferences, and among them six liked the dog picture more. Based on the currently available information, how would you update your inference on the parameter θ?

[*Hint: You can use R to compute the relevant probabilities. The operator %*% allows to do vector (matrix) multiplication*]

Problem 2.

Bob claims to have ESP. To test this claim, you propose the following experiment. You will select one card from four large cards with different geometric figures, and Bob will try to identify it. Let θ denote the probability that Bob is correct in identifying the figure for a single card. You believe that Bob has no ESP ability, but there is a small chance that θ may be somehow slightly larger or smaller than .25, out of chance.

After some thought, you place the following prior distribution on θ:

θ	0	.125	.250	.375	.500	.625	.750	.875	1
$p(\theta)$	.001	.001	.950	.008	.008	.008	.008	.008	.008

(1) According to your prior beliefs, what is the **prior** probability that $\theta > 0.5$?

(2) Suppose that the experiment is repeated ten times. Compute the marginal (aka prior predictive) probability that Bob is correct in identifying the correct card 7 times out of 10 times.

Now suppose that the experiment is indeed conducted 10 times, and Bob is correct six times and incorrect four times.

(3) Compute the **posterior** probability that Bob has ESP ability (say, $p(\theta > 0.5|\text{Data})$).

Problem 3.

Every year, Professor Snape tests the ability of some students at Hogwarts on a very difficult spell. Based on the 5 previous years he expects that only 20% of the students would be able to successfully work on the spell. [*Hint: One way to see intepret this is that we base our information on a prior sample size of 5*] \ In the Winter 2020, the professor notices that 14 out of 30 students are able to practice the spell correctly. He takes this as an indication of effectiveness of his training practices.

(1) What likelihood function would you consider for this problem?

(2) What (conjugate) prior distribution would you use based on the available prior information?

(3) Provide a mathematical expression of your posterior distribution after observing the actual experiment (i.e., the data).

(4) Compute the posterior expectation (aka posterior mean) and compare with the prior expectation (aka prior mean). What can you conclude?

(5) Based on the observed data, what is the posterior probability that the probability of success in practicing the spell is more than 25%?

(6) Comment on the difference between prior and posterior inference.

Now, suppose you want to know how - if Snapes were to test 20 more students - the probability that the number of successes is greater than 5.

(7) Compute the corresponding posterior predictive probability.

Now, suppose that Snapes is actually able to test 30 more students, and finds that 10 students are able to work on the spell.

(8) What (conjugate) prior distribution would you choose for conducting this second inference analysis? What would be the posterior distribution and the posterior expectation after seeing the results of the second batch of tests?

Problem 4

The table below has the overall free throw proportion and results of free throws taken in pressure situations, defined as "clutch" for ten National Basketball Association players (those that received the most votes for the Most Valuable Player Award) for the $2016 - 2017$ season. \

Since the overall proportion is computed using a large sample size, assume it is fixed and analyze the clutch data for each player separately using Bayesian methods. Assume a uniform prior throughout this problem.

	Overall	Clutch	Clutch
Player	proportion	makes	attempts
Russell Westbrook	0.845	64	75
James Harden	0.847	72	95
Kawhi Leonard	0.880	55	63
LeBron James	0.674	27	39
Isaiah Thomas	0.909	75	83
Stephen Curry	0.898	24	26
Giannis Antetokounmpo	0.770	28	41
John Wall	0.801	66	82
Anthony Davis	0.802	40	54
Kevin Durant	0.875	13	16

(1) Describe your model for studying the clutch success probability including the likelihood and prior.
(2) Plot the posteriors of the clutch success probabilities.
(3) Summarize the posteriors in a table, which should include (at least) the posterior mean and the (5%, 25%, 50%, 75%, 95%) posterior quantiles
(4) Do you find evidence that any of the players have a different clutch percentage than overall percentage? [*Hint: test the hypothesis that the probability of "clutch" for a player is greater than their corresponding overall proportion*]