Homework 1

Bayesian Data Analysis

Problem 1.

Prof. Guindani wants to test the efficacy of a Facebook campaign for the new professional *Master in Data Science* at UCI. For that reason, he is interested in assessing the probability that a random individual may click on the post for the campaign.

For simplicity (to avoid tedious computations), suppose he pays for a pilot study, where the post is shown to 5 individuals. Out of the five individuals, three of them click on the post.

We are interested in studying the proportion, say θ, of individuals in the population who may click on the Facebook post, which is shown to a particular sub-sample of all Facebook users.

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

For simplicity, and to understand how the prior to posterior updating works, consider the following prior distribution for the proportions:

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

This prior distribution suggests that with relatively high probability (0.8), I believe *a priori* (i.e. before seeing the data) that there is a 50% chance that a person will click on the post. Also, I do not believe that the post will receive very few (< 0.3) or very many (> 0.7) clicks.

(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) Suppose I conduct a more thorough study, and obtain data from 10 more individuals. Among them, eight clicked on the post. Based on the currently available information, how would you update your inference on the parameter θ?

(4) Based on the results from (3), would you conclude that the campaign marketing post is successful, or you'd look for a better post? [This is a thought-provoking question at this point; try your best about to answer it]

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

(matrix) multiplication]

Problem 2

A single-arm clinical trial is scheduled to assign 100 patients to an innovative treatment. At a first interim analysis, 20 out of 30 patients have had a positive outcome under the treatment. Assume a uniform prior on the probability, say θ, of a patient having a positive outcome:

(1) Based on the interim analysis, what is the posterior distribution of θ?

(2) What is a possible point estimate for θ?

(3) Find the 5% and 95% quantiles of the posterior distribution.

(4) In the continuation of the trial, 50 out of the remaining patients have had a positive outcome. What is the updated posterior of θ? What is a possible estimate for θ? Find the 5% and 95% quantiles of the distribution.

(5) Based on the results from the analysis, does the treatment appear promising?

Problem 3

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 of each subject, including the likelihood and prior.

(2) Plot the posterior distribution of the clutch success probabilities for James Harden and Le Bron James.

(3) Summarize the posterior distribution *for each player* in a table. The table 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*]