CS7DS3 Assignment 1

February 26, 2024

To be submitted by **1 p.m.** on **Monday 11th March, 2023**. Submit solutions through blackboard. Legible scans are fine for some questions but please be aware that the Turnitin submission requires a minimum amount of printed text. All plots should be produced using standard software.

Please remember to print your **name** and **student number** on your submission.

Please show your workings. Where code has been used, an outline description, along with relevant results, is sufficient.

If you have any questions about the assignment, email me: arwhite@tcd.ie.

Problem set - 15 marks total

The Detroit Pistons are a basketball team in the National Basketball Association (NBA). Unfortunately, their start to the current 2023-24 season has not gone well. By December 27th, 2023, the Pistons had played 30 games, of which they had lost 28 and won only 2. (Games can only be won or lost, and not drawn, etc.)

In the absence of further information, we can model the probability that the Pistons win a basketball game using a binomial distribution with parameter θ. That is, let data y_i denote the outcome of a game i, with $y_i = 1$ if the Pistons win and 0 otherwise. Then $y_i \sim \mathcal{B}(1, \theta)$, and $\mathbb{P}(y_i = 1) = \theta$.

A colleague is interested in estimating θ using Bayesian methods. For the prior $p(\theta)$, she decides to use a beta distribution, $\mathcal{B}e(a, b)$, but must still decide how to choose a and b. After careful consideration, she decides that three different choices of a and b are potentially suitable. These are described as follows:

A) A so-called "non-informative" $\mathcal{B}e(1, 1)$ prior.

B) After reviewing the final overall standings from the previous season, your colleague notices that the win percentages of teams ranged from about 20% to 70%. Naturally, the win percentages were centred around 50%. She decides that a $\mathcal{B}e(8, 8)$ prior describes this behaviour accurately enough for her purposes.

C) At the beginning of the season, a well known sports pundit had predicted that the Pistons would win about 28 games over the course of the 82 game season, corresponding to about a 34% winning percentage. The pundit seemed very confident when he made this prediction. Your colleague decides to represent this opinion using a $\mathcal{B}e(7, 13)$ prior.

Questions

1) Describe and compare the key properties of each of these potential prior distributions, as well as how closely they match their descriptions above. Which prior is the most informative? Explain your reasoning. **[3 marks]**

2) Which prior is the most appropriate choice in your opinion? Explain your preference. **[1 marks]**

3) Using your preferred prior distribution, along with the data from the 30 games that the Pistons have played, compute the posterior distribution $p(\theta|y, a, b)$ for the winning probability parameter θ. Describe the key properties of this distribution, including how it differs from the prior distribution that you chose. **[3 marks]**

4) Consider the remaining 52 games of the season that the Pistons were yet to play as of December 27th, 2023. Using Monte Carlo methods or otherwise, along with your preferred posterior distribution $p(\theta|y, a, b)$, calculate the probability that the Pistons will win at least another 8 games over the course of the season. **[3 marks]**

5) A betting website offers players the chance to play two games, Game 1 and Game 2. In Game 1, for a stake of €100, the player wins €$10X$, where X equals the total number of games won during the remainder of the season. In Game 2, for a stake of €1000, the player wins €$10X^2$. What is the expected return on each game? Which game do you think it is better to play, or would you prefer not to play at all? Explain your answer. **[3 marks]**

6) Briefly describe how sensitive your calculations for the expected return for Games 1 and 2 in part 5) were to your choice of prior distribution. In other words, if you use one of the other candidate prior distributions described above, how much would this change your expected return estimate for each game? **[2 marks]**