Assignment: Bayes Theorem and Estimating a Population Proportion

Problem 1 (10 points)

Assume that you flip a fair coin 30 times. You observe the sequence of heads and tails in your experiment. A *run* is a sequence of all heads or tails.

1. You're interested in finding the information about the longest run. Simulate in R, the experiment described above 10,000 times and each time note the length of the longest run. Find the average of the longest run in 10,000 simulations and report it. Hint: The *rle* command in R will be helpful. (5 points)

2. What is the probability of getting a run of 7 or more in 30 coin flips based on your simulation? Does your calculation for this probability follow the frequentists' or Bayesian's view of probability? (5 points)

Problem 2 (10 points)

Alex claims to have a biased coin. To test this claim, you propose to toss his coin 20 times and count the number of heads. Let π denote the probability of getting a head. You believe that the coin is fair ($\pi = 0.5$), but there is a small chance that π is either larger or smaller than 0.5. After some thought, you place the following prior distribution on π:

π	0	.125	.250	.375	.500	.625	.750	.875	1
$p(\pi)$	.001	.002	.008	.014	.95	.014	.008	.002	.001

Suppose that you got 15 heads in 20 coin tosses.

1. Clearly state the null and alternative hypotheses. You don't need to perform the hypothesis test for this problem. (2 points)

2. Use R to find the posterior probability distribution on π. (8 points)

Problem 3 (30 points)

You are the assistant coach of the women's softball team at a California college. The head coach has asked you to assess a new first year player who is joining the team. As a high school student, she was at bat 120 times and got 40 hits. Team's coach asks you to estimate θ, her underlying true probability of getting a hit in any at bat as a college-level player and if the estimate is more than 0.2 allow her to play in the first game.

1. Clearly state the null and alternative hypotheses. (2 points)

2. Specify a beta prior that seems appropriate to capture your knowledge or uncertainty about θ before the new player plays in any college-level games. Must show your work and explain in a few sentences how you chose the values of α and β. (5 points)

3. Suppose the player now plays eight college-level games, has thirty at bats, and gets 5 hits. We will use a binomial likelihood for these data. What's the likelihood? (3 points)

4. Suppose you're asked to use a beta (33, 66) for your prior. Name the posterior distribution and specify its parameter values. (3 points)

5. Plot the posterior density.(3 points)

6. Create a plot showing the prior density, the likelihood, and the posterior density, all on the same axes. (4 points)

7. Based on your analysis, will she play in the first game or not at 0.05 threshold level? Why? (2 points)

8. What's the 95% credible interval for θ? (2 points)

9. What are the posterior mean, median, and mode? (6 points)