Homework 3: 2/6 - 2/10

STA 336 Due 11:59pm Monday, February 13 Name:

Instructions: Write-up complete solutions to the following problems and submit answers on Gradescope. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A rubric for homework problems appears on the final page of this assignment.

- Unless otherwise noted, problem numbers are taken from the 4th edition of DeGroot and Schervish's *Probability and Statistics*.

Monday 2/6

Section 7.4: (2, 3), 6, 12, 15
Note: Problems enclosed in parentheses will be graded as a single problem.

Wednesday 2/8

You are statisticians employed by the consulting firm *BayesBall*. A veteran major-league baseball scout seeks your advice regarding the probability an amateur baseball player Phil Hatt will get a base hit against a major-league pitcher. The scout has arranged for P. Hatt to have at least 10 at bats against a major-league pitcher.

The traditional batting average estimator $\hat{\theta}_f = \frac{X}{n}$ (i.e. proportion of hits in n at bats) is a frequentist estimator that uses observed data, but ignores prior information. Assuming each of the n at bats constitute an independent Bernoulli trial with probability θ of a base hit, then

$$X \sim \text{Bin}(n, \theta)$$

Suppose we have the following additional prior information:

- P. Hatt appears to be a good but not great player. He is one of the better batters on a somewhat above average high-school team.
- The few major-league scouts who have watched him play do not believe his batting ability is at the professional level.
- A barely adequate major-league hitter has a batting average of 0.2.
- A very good major-league batter has a batting average of 0.3.
- Ty Cobb has the all-time best major-league batting average of 0.366.

Problem AP1

a. Explain why it may be a good idea to use a Beta prior.

b. Determine 'reasonable' values of the hyperparameters α and β in the prior distribution for θ based on the facts listed above and properties of the Beta distribution. Justify your choices by computing appropriate means, variances, probabilities, and graphs.

c. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be a sample of n at bats, where $x_i = 1$ if the ith at bat resulted in a base hit, and $x_i = 0$ otherwise. Find a formula for posterior distribution of θ in terms of a generic sample $\mathbf{x}$ and generic values of α and β. And then write down the formula for the values of α and β you specified in the previous part.

d. In R, plot the prior distribution for θ, along with posterior distributions for several values of x, using the values of α and β you selected previously.

Problem AP2

a. Find the general formula for the Bayes Estimator $\hat{\theta}_b$ for θ in terms x (the number of base hits) and generic α and β. Then write down the Bayes estimator for the value of α and β you specified in Problem 1.

b. Complete the following table to compare the frequentist estimator $\hat{\theta}_f$ and your Bayes estimator $\hat{\theta}_b$:

x	$\hat{\theta}_f$	$\hat{\theta}_b$
0	0.0	
1	0.1	
2	0.2	
3	0.3	
4	0.4	
5	0.5	
6	0.6	
7	0.7	
8	0.8	
9	0.9	
10	1.0	

c. Show that $\hat{\theta}_b$ is a weighted average of $\hat{\theta}_f$ and the prior mean $\frac{\alpha}{\alpha+\beta}$.

d. Suppose you actually had no prior knowledge about typical batting averages in baseball, or of P. Hatt's talents. What would be a reasonable prior to use in this case? Compute the Bayes estimator for this prior, and compare to the frequentist estimator.

e. Show how to obtain the frequentist estimator using an improper prior.

Friday 2/10

Section 7.5: (2, 3), 5, 6, 11
Note: Problems enclosed in parentheses will be graded as a single problem.

Homework 3: 2/6 - 2/10

STA 336 Due 11:59pm Monday, February 13 Name:

General Rubric

Points	Criteria
5	The solution is correct **and** well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information)
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.