Math 457/557 Final Exam
Due by 10am March 22, 2018

The Rules: This is an open book, open notes exam. You are not to talk about this exam to anybody. Even a statement like "Problem 1 was easy" would give information, so you can't talk about the exam at all.

Turn in: Solutions to these problems. Also, email me a text file containing your R-code.

1. (10 points) A client is testing a new manufacturing process and is interested in learning about the rate at which the new process produces defective items. The client tells you that defective rates for similar manufacturing processes average about 3%, and he would be shocked if he saw a defective rate were above 10%. Let θ be the actual defective rate.

   The client samples $n = 25$ of the products produced by the new manufacturing process and finds no defective items.

   (a) Setting a prior. Explain how you can create a prior distribution for θ that reflects what the client told you about the defective rate. Tell me what the prior distribution will be (name and parameter values). Give me a plot of the prior pdf.

   (b) Show the derivation of the posterior pdf for θ. Give me the name and parameter values for the posterior distribution of θ, and a plot of the posterior pdf.

   (c) Use your posterior pdf to get an estimate for the client. This should include a point-estimate and an interval estimate. For the interval estimate, be aware that the client is interested in setting a 95% *upper bound* on the error rate.

2. (10 points) The file `GPAdata.dat` contains data on 40 graduating college seniors. The file contains three variables: y contains the students' college GPA's; $x1$ contains their college entrance verbal test scores (as percentiles); $x2$ contains their college entrance mathematics test scores (as percentiles). We will use this dataset to predict the college GPA for a student based on their entrance exam scores.

   (a) Fit a regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ using the g-prior with $g = n$, $\nu_0 = 2$ and $\sigma_0^2 = 0.4^2$ (so, σ_0^2 is about the OLS estimate for σ^2). Obtain marginal posterior means and 95% confidence intervals for $\boldsymbol{\beta}$.

   (b) Give a 95% confidence interval for the mean GPA among all students that scored 70 on their verbal entrance test and 80 on their math entrance test.

   (c) Fred is an incoming student whose verbal entrance score is 70 and whose math entrance score is 80. Give a 95% prediction interval for Fred's college GPA. Explain how you made this calculation. Give a histogram that shows the distribution for Fred's GPA.

   (d) Mary's incoming scores are 65 (verbal) and 90 (math). Based on your analysis, what is the probability that Mary's college GPA will be above Fred's college GPA? Explain how you made this calculation.

3. (Bonus: 6 homework points) Return to problem 4.8, and consider modeling the number of children of men in their 30s without bachelor's degrees (so, just the B group in problem 4.8) Ignore the men with bachelor's degrees. In problem 4.8, we saw that the Poisson model did not give us a good fit for the data.

   Suppose, instead, that we wanted to model the men as coming from two groups: Group 1 consists of unmarried/uncommitted men who will always have 0 children. Group 2 consists of married/committed men for whom the number of children will follow a Poisson distribution.

   Can you fit a model like this? Does it fit the data better? If you can work through something like this, get it to fit the data, and write it up nicely, you can have 6 bonus homework points.

   Note: I wanted this to be an exam question, but I'm running out of time and haven't worked it out. I feel like something like this should work.