STAT 638 — Exam 1 **Name:** ________________________________
Fall 2015

INSTRUCTIONS FOR THE STUDENT:

1. You have exactly 1 hour to complete the exam.
2. There are 5 pages including this cover sheet, and 7 questions.
3. Point values are given in parentheses. **Please circle the letter of the correct answer on multiple choice questions.**
4. Do not discuss or provide any information to anyone concerning any of the questions on this exam or your solutions until I post the solutions.
5. The only materials you may use are a calculator, an 8 1/2 by 11 inch formula sheet of your own making (with writing on both sides), and a copy of the common distributions on pp. 253-258 of Hoff. **Do not use the textbook or class notes.**

I attest that I spent no more than 1 hour to complete the exam. I used only the allowed materials as described above. I did not receive assistance from anyone during the taking of this exam.

Student's Signature ______________________________

INSTRUCTIONS FOR THE PROCTOR:

(1) Record the time at which the student starts the exam: ____________________

(2) Record the time at which the student ends the exam: ____________________

(3) Immediately after the student completes the exam, please scan the exam to a .pdf file and have the student upload it to WebAssign.

(4) Collect all portions of this exam at its conclusion. Do not allow the student to take any portion of the exam.

(5) Please keep these materials until October 16, 2015, at which time you may either dispose of them or return them to the student.

I attest that the student has followed all the INSTRUCTIONS FOR THE STUDENT listed above and that the exam was scanned into a pdf and uploaded to WebAssign in my presence.

Proctor's Signature ______________________________

1. A researcher randomly selects 100 white mice from a large population of such mice, and tests each one for a fairly rare genetic mutation that makes the mice unsuitable for a certain experiment. Let θ be the proportion of mice in the population with the genetic mutation. It turned out that 4 mice in the sample had the genetic mutation. As a prior for θ the researcher decides to use a beta$(1/2, 10)$ density.

(a) (8) What feature of the researcher's prior makes it seem appropriate in this situation?

(b) (8) The beta$(1/2, 10)$ prior has an amount of information equivalent to that in a sample of how many mice?

(c) (8) What is the mean of the researcher's prior?

(d) (8) Identify the researcher's posterior distribution.

(e) (8) What are the mean and mode of the posterior distribution?

(f) (8) Describe how you would find a 95% credible interval for the proportion of all mice with the genetic mutation. **Note:** The interval need not be an HPD region.

2. Suppose that $Y_1, \ldots, Y_n$ are independent and identically distributed observations from a Poisson distribution with mean θ.

(a) (8) Show that the Jeffreys prior in this case is proportional to $\theta^{-1/2} I_{(0,\infty)}(\theta)$.

(b) (7) Is the prior proper? Why?

(c) (7) Identify the posterior corresponding to use of the Jeffreys prior. Is the posterior proper, and why?

3. (6) Lindley's paradox refers to

(a) a situation where a frequentist confidence interval for a parameter is drastically different than a Bayesian HPD interval.

(b) the fact that Bayes factors often overstate the signifcance of evidence against point null hypotheses.

(c) the fact that frequentist P-values often overstate the significance of evidence against point null hypotheses.

(d) situations where using noninformative priors is inappropriate.

(e) Dennis Lindley's son and daughter, both of whom have PhDs.

4. (6) It is of interest to test the hypotheses

$$H_0 : \theta = 0 \quad \text{vs.} \quad H_1 : \theta = 1$$

using the data $\boldsymbol{y}$. The two relevant values of the likelihood are $p(\boldsymbol{y}|0) = 1/4$ and $p(\boldsymbol{y}|1) = 1/10$, and the prior probabilities of H_0 and H_1 are $2/3$ and $1/3$, respectively. The Bayes factor and the posterior probability of H_0 are, respectively,

(a) 5 and 5/6.

(b) 5/2 and 5/6.

(c) 5 and 5/7.

(d) 5/2 and 5/7.

(e) π and e^{-1}.

5. (6) An experimenter observes data whose distribution depends on an unknown parameter θ. The posterior distribution for θ is normal with mean 10 and standard deviation 1. The experimenter wants to predict a value for Y, which is independent of the data given θ. The distribution of Y given θ is normal with mean θ and standard deviation 5. Which of the following is a valid way to generate a single value from the posterior predictive distribution of Y?

(a) Generate Y from $N(10, 1)$.

(b) Generate Y from a gamma distribution with mean 10 and standard deviation 5.

(c) Generate Y from $N(10, 25)$.

(d) Let θ be a value drawn from the prior distribution. Then Y is drawn from $N(\theta, 25)$.

(e) Let θ be a value drawn from $N(10, 1)$. Then Y is drawn from $N(\theta, 25)$.

6. (6) A good way to check whether a random sample of data fits a model $p(y|\theta)$ is to compare the

(a) empirical distribution of the data with the posterior predictive distribution.

(b) empirical distribution of the data with the posterior distribution.

(c) likelihood function with the posterior.

(d) likelihood function with the prior.

(e) prices of beef and pork.

7. (6) The data $Y_1, \ldots, Y_n$ have been observed but Y_{n+1} has not been observed. In the absence of any other information, a valid expression for the posterior predictive distribution, $m(y_{n+1}|y_1, \ldots, y_n)$, is

(a) $\int_\Theta p(y_{n+1}|\theta)p(\theta|y_1, \ldots, y_n)\, d\theta$.

(b) $\int_\Theta p(y_{n+1}|\theta)p(\theta)\, d\theta$.

(c) $\int_\Theta p(y_1, \ldots, y_n, y_{n+1}|\theta)p(\theta)\, d\theta / m(y_1, \ldots, y_n)$.

(d) $\int_\Theta p(y_1, \ldots, y_n, y_{n+1}|\theta)p(\theta|y_1, \ldots, y_n)\, d\theta$.

(e) salud!