AM 166: Homework # 7 (due Thursday, April 26)

1. Opinion and political polls taken by the Gallup, Harris, and other organizations use random samples [typical sample size for a national poll are 600 to 1500]. On January 21-22, 1993, the *Newsweek* poll asked several questions to 663 adults. One question was with respect to gays in military. To the question, "Can gays serve effectively in the military if they keep their sexual orientation private?", 477 responded "Yes" and 186 responded "No". Let θ be the population proportion that would say "Yes". Use normal approximation to obtain the 95% posterior confidence interval for θ, assuming a uniform prior.

2. (Comparing two population proportions). A researcher was interested in whether and how the transmittal letter affected response in a mail survey. He sent out 2040 letters to people, randomly selecting between a semipersonal letter and a form letter. Of the 1018 people who received a semipersonal letter, 325 responded, while of the 1022 people who received a form letter, 225 responded. Let θ_1 be the population proportion who would respond if a semipersonal letter was received, and θ_2 the population proportion who would respond if a form letter was received.

   (a) Use normal approximation to determine the posterior distribution of $\theta_1 - \theta_2$, assuming independent uniform priors for both θ_1 and θ_2.

   (b) Find the 95% posterior confidence interval for $\theta_1 - \theta_2$.

   (c) What is the posterior probability of $\theta_1 - \theta_2 > 0$?

   [*Hint:* The posterior distribution of θ_1 and θ_2 are *independent*. You may want to use normal approximation to θ_1 and θ_2 separately. Then $\theta_1 - \theta_2$ is just the difference of two independent normal distributions.]

3. Geologists and climatologists are interested in the age of glacial moraines. One of the methods they use to date such moraines involves measuring the accumulation of chlorine. In one study published in *Science* (1990), researchers reported on the amounts of chlorine in several moraines in Bloody Canyon, California. These were the amount of chlorine (in parts per million) in five samples taken from a moraine called Older Tahoe:

   $$73 \quad 75 \quad 49 \quad 76 \quad 115.$$

   Assuming that the amount of chlorine is distributed as $N(\theta, \sigma^2)$ and set the prior of (θ, σ) as

   $$\pi(\theta, \sigma^2) \propto 1/\sigma^2.$$

   (a) Find the posterior (marginal) distribution of θ, and give a 95% posterior confidence interval for θ.

   (b) Use simulation to draw 5000 sample points to find the (predictive) probability that the next sample of moraine will contain more than 130 ppm chlorine.

   Hint: For part (b), note that the amount of chlorine that will be contained in the next sample of moraine, say $\tilde{Y}$, has distribution $N(\theta, \sigma^2)$ where (θ, σ^2) has the posterior distribution determined by the original 5 samples. Therefore, in order to draw samples for $\tilde{Y}$, you can follow these steps.

(a) Draw a sample (θ, σ^2) from the posterior distribution. To do this, you can first draw a sample of σ^2 from the marginal posterior distribution of σ^2. Once σ^2 is obtained, draw a sample of θ from the posterior conditional distribution $p(\theta|\sigma^2, y)$ which is a normal distribution.

(b) Once (θ, σ^2) is obtained, draw a sample of $\tilde{Y}$ from the distribution $N(\theta, \sigma^2)$.

(c) Repeat (a)(b) 5000 times to get 5000 samples of $\tilde{Y}$. Compute the sample proportion of $\tilde{Y} > 130$.

4. (Simulation) Suppose $Y \sim B(n; \theta)$ with θ unknown. Assume $n = 10$ and the observation is $Y = 7$. Assume the prior is

$$\pi(\theta) \propto \sin(\pi\theta).$$

Use simulation to draw 5000 samples from the posterior distribution of θ, and give the 95% posterior interval for θ. [*Hint:* You can follow these steps to run the simulation.

(a) Divide the interval [0,1] into, say 1000, equal-length subintervals, and the value of posterior density (up to a constant) at each end points of the subintervals.

(b) Run a normalization for the value you obtained in (a) so that you end up with a probability distribution. [That is, you have obtained an approximation of the posterior distribution using a discrete multinomial distribution].

(c) Draw 5000 samples from this multinomial approximation. You may want to call help from uniform distribution to draw samples from multinomial distribution.

From these 5000 samples, find the 95% posterior interval.]