Week 4 Problems

1. (Adapted from Gelman 2.10) Suppose there are N cable cars in San Francisco, numbered sequentially from 1 to N. You see a cable car at random; it is numbered 203. You wish to estimate N.

   Assume your prior distribution on N is geometric with mean 100; that is,

   $$p(N) = (1/100)(99/100)^{N-1}, \text{ for } N = 1, 2, \dots$$

   a) What is your posterior distribution for N up to a constant of proportionality?

   $$\begin{aligned} p(N|X) &\propto p(X|N)p(N) \\ &= \frac{1}{N}\left(\frac{1}{100}\right)\left(\frac{99}{100}\right)^{N-1} \text{ for } N \geq 203 \\ &\propto \frac{1}{N}\left(\frac{99}{100}\right)^{N-1} \text{ for } N \geq 203 \end{aligned}$$

   b) Find the posterior mean and standard deviation by approximating the normalizing constant in R (without simulating).

   The posterior is

   $$\begin{aligned} p(N|X) &= \frac{p(X|N)p(N)}{p(X)} \\ &= \frac{1}{p(X)}\frac{1}{N}\left(\frac{99}{100}\right)^{N-1} \text{ for } N \geq 203 \end{aligned}$$

   The normalizing constant can be approximated by

   $$p(X) \approx \sum_{N=203}^{10000} \frac{1}{N}\left(\frac{99}{100}\right)^{N-1}$$

   The posterior mean and standard deviation are then approximated by

   $$\begin{aligned} E[p(N|X)] &\approx \sum_{N=203}^{10000} N\frac{\frac{1}{N}\left(\frac{99}{100}\right)^{N-1}}{p(X)} = \sum_{N=203}^{10000} \frac{\left(\frac{99}{100}\right)^{N-1}}{p(X)} \\ sd[p(N|X)] &\approx \sqrt{\sum_{N=203}^{10000} (N - E[p(N|X)])\frac{\frac{1}{N}\left(\frac{99}{100}\right)^{N-1}}{p(X)}} \end{aligned}$$

```
> N <- 203:10000
> p.x <- sum((1/N) * (99/100)^(N - 1))
> E.N <- sum(((99/100)^(N - 1))/p.x)
> E.N
[1] 279.0885
> sd.N <- sqrt(sum((N - E.N)^2 * ((1/N) * (99/100)^(N - 1))/p.x))
> sd.N
[1] 79.96458
```

c) Now find the posterior mean and standard deviation by simulation in R. Are your answers similar to those in b)?

```
> N <- 203:10000
> n.sim <- 10000
> unnormal.post <- (1/N) * (99/100)^(N - 1)
> post.draws <- sample(N, size = n.sim, prob = unnormal.post,
+     replace = T)
> mean(post.draws)
[1] 280.2904
> sd(post.draws)
[1] 81.43097
```

2. (adapted from Gelman 3.2) On September 25, 1988, the evening of a Presidential campaign debate, ABC News conducted a survey of registered voters in the United States; 639 persons were polled before the debate, and 639 different persons were polled after.

Survey	Bush	Dukakis	No opinion/other	Total
pre-debate	294	307	38	639
post-debate	288	332	19	639

Assume the surveys are independent simple random samples from the population of registered voters. Model the data with two different multinomial distributions.

a) What is the posterior probability that support for Bush increased between the two surveys?

```
> library(MCMCpack)
> n.sim <- 10000
> y.1 <- c(294, 307, 38)
> y.2 <- c(288, 332, 19)
> alpha.0 <- c(1, 1, 1)
> post.1 <- rdirichlet(n.sim, alpha = y.1 + alpha.0)
> post.2 <- rdirichlet(n.sim, alpha = y.2 + alpha.0)
> mean(post.2[, 1] > post.1[, 1])
[1] 0.3594
```

b) Of the voters who had a preference for either Bush or Dukakis, what is the posterior probability that there was a shift toward Bush between the two surveys?

```
> a <- b <- 1
> y.bd.1 <- c(294, 307)
> y.bd.2 <- c(288, 332)
> post.bd.1 <- rbeta(n.sim, y.bd.1[1] + a, y.bd.1[2] + b)
> post.bd.2 <- rbeta(n.sim, y.bd.2[1] + a, y.bd.2[2] + b)
> mean(post.bd.2 > post.bd.1)
[1] 0.1936
```

3. Suppose we have n observations that follow a Normal distribution with a common mean μ and variance σ^2. Also, suppose that we know the mean of the data, but want to learn about the

variance of the data. Find the posterior distribution of the variance σ^2 given an Inverse-gamma prior. Specifically, find $p(\sigma^2|\mathbf{y})$ given

$$\begin{aligned} Y_i &\sim \mathrm{N}(\mu, \sigma^2) \\ \sigma^2 &\sim \text{Inv-gamma}(\alpha, \beta) \end{aligned}$$

$$\begin{aligned} p(\sigma^2|\mathbf{y}) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right) \times \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} e^{-\beta/\sigma^2} \\ &\propto \prod_{i=1}^{n} (\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right) (\sigma^2)^{-(\alpha+1)} e^{-\beta/\sigma^2} \\ &= (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right) (\sigma^2)^{-(\alpha+1)} e^{-\beta/\sigma^2} \\ &= (\sigma^2)^{-(\alpha+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta}{\sigma^2} + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right)\right] \\ &= (\sigma^2)^{-(\alpha+\frac{n}{2}+1)} \exp\left[-\left(\frac{2\beta + 2\left(\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\right)}{2\sigma^2}\right)\right] \\ &= (\sigma^2)^{-(\alpha+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}}{\sigma^2}\right)\right] \end{aligned}$$

The posterior is an Inv-gamma$\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\right)$ distribution.