HOMEWORK 8, STAT 251

(1) You have gathered mean arterial pressures from ten subjects currently being treated for high blood pressure with a specific drug regimen. Here are the data:

```
mapr <- c(97.6, 117.7, 126.7, 111.0, 106.7, 108.4, 118.9, 108.6, 119.1, 106.2)
```

You assume these data are (conditionally) normally distributed.

(a) Compute the log of the likelihood with $\mu = 112$ and $\sigma^2 = 8.4^2$.

(b) Compute the log of the likelihood with $\mu = \bar{y}$ and $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, where $\bar{y} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i$.

(c) Now plot the log of the likelihood for values of μ ranging from 110 to 114 in increments of .05, and with σ^2 fixed at $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, where $\bar{y} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i$.

(2) Here is the total serum cholesterol for eight urban residents of Guatemala:

```
cholest.u <- c(197, 199, 214, 217, 222, 227, 228, 234)
```

The population variance of serum cholesterol measurements for urban residents of Guatemala is assumed to be 260. The prior belief about the serum cholesterol of residents of Guatemala is that the population mean, μ_{urban}, should have a mean of 150 with a variance of 90 (and the prior will be a normal distribution).

(a) What is the posterior distribution of the mean, μ_{urban}, in this case if it is assumed that the data are conditionally iid with a normal distribution?

(b) Plot the posterior distribution from part (a).

(3) Here is the total serum cholesterol for ten rural residents of Guatemala:

```
cholest.r <- c(139, 142, 143, 144, 145, 148, 155, 162, 171, 181)
```

The population variance is assumed to be 270. The prior belief about the serum cholesterol of residents of Guatemala is that the population mean, μ_{rural}, should have a mean of 150 with a variance of 90 (and be normally distributed).

(a) What is the posterior distribution of the mean, μ_{rural}, in this case if it is assumed that the data are conditionally iid with a normal distribution?

(b) Plot the posterior distribution from part (a).

(4) Extending the previous two questions, suppose that the urban data are also assumed to be independent of the rural data. Furthermore, assume that a priori μ_{rural} is independent of μ_{urban}. With these assumptions, μ_{rural} and μ_{urban} will also be independent of each other a posteriori.

(a) What is the posterior distribution of $\mu_{urban} - \mu_{rural}$?

(b) Plot the prior and the posterior distribution of $\mu_{urban} - \mu_{rural}$ on the same graph.

(c) State and interpret in the context of this problem the (central) 95% posterior credible interval for $\mu_{urban} - \mu_{rural}$.

(5) The number of immature red tail hawks sitting on fence posts or power poles along a mile-long stretch of rural highway can reasonably be assumed to follow a Poisson distribution. However, we will be comparing data six years apart (in 2003 and then in 2009), and we suspect there might be differences in behavior between 2003 and 2009. Thus, we will assume the observations from 2003 follow the $Poisson(\theta_{2003})$ distribution, and the observations from 2009 follow the $Poisson(\theta_{2009})$ distribution. We'll also assume the 2003 data are independent of the 2009 data. Assume that *a priori* θ_{2003} and θ_{2009} are independent of each other.

(a) Along a 10 mile stretch of highway south of Saratoga Springs in 2003, the count of hawks per mile was

```
hawks.2003 <- c(0, 2, 0, 0, 0, 4, 3, 0, 0, 1)
```

What is the expected value of the posterior distribution for θ_{2003} if the prior is a gamma distribution with shape=0.2 and rate=0.2?

(b) Along that same stretch of highway in 2009, the counts were

```
hawks.2009 <- c(1,1,3,3,2,2,2,4,0,3)
```

What is the expected value of the posterior distribution for θ_{2009} if the same prior is used as in the previous problem (i.e., Gamma(0.2,0.2)?

(c) What is the expected value of the posterior distribution of $d = \theta_{2009} - \theta_{2003}$?

(d) What is the 95% posterior credible interval for d?

(e) Plot the Monte-Carlo estimated posterior distribution of $d = \theta_{2009} - \theta_{2003}$ by using the density and plot functions in R.