Stat 260/CS 294 Homework Assignment 1 (due February 8)

1. A retail company monitoring the sales of a new product records the numbers sold at each of I similar retail outlets in each of n consecutive weeks. Sales are modelled using a Poisson distribution for each outlet with a constant weekly rate:

$$(x_{ij} \mid \theta_i) \sim \text{Po}(\theta_i), \qquad \text{independently,}$$

for $(i = 1 : I; j = 1 : n)$, where x_{ij} is the sales at outlet i in week j. Possible variation between outlets is modelled by assuming the means θ_i are independently drawn from the gamma prior with uncertain mean $1/\mu$ and specified shape $a > 0$, i.e.,

$$(\theta_i \mid \mu) \sim \text{Ga}(a, a\mu).$$

The x_{ij} are assumed conditionally independent of μ given θ_i. For notation, set $\Theta = \theta_{1:I} = \{\theta_i, i = 1 : I\}$ and $X = x_{1:I,1:n} = \{x_{ij}, i = 1 : I, j = 1 : n\}$.

   (a) Write down the complete form of $p(X \mid \Theta, \mu)$.

   (b) Show that, conditional on X and μ, the θ_i are independent and gamma distributed, namely $(\theta_i \mid X, \mu) \sim \text{Ga}(a + ny_i, a\mu + n)$, independently, where y_i is a function of the data X. Identify y_i.

   (c) By integrating over Θ, find the marginal likelihood $p(X \mid \mu)$. Be sure to identify all terms involving μ, but you may ignore other "constants."

   (d) If μ has the gamma prior $\mu \sim \text{Ga}(r, s)$, show that

$$p(\mu \mid X) \propto \mu^{r+aI-1} e^{-s\mu} (n + a\mu)^{-q}, \qquad \mu > 0,$$

   where $q = I(a + n\bar{x})$ and $\bar{x}$ is the overall sample mean.

   (e) Assuming the result of (4), let r and s tend to zero. Show that the resulting posterior for μ is such that $\mu = \phi/\bar{x}$, where $\phi \sim F_{k,h}$ with $k = 2aI$ and $h = 2nI\bar{x}$.

Some facts related to the F distribution. If $\phi \sim F_{k,h}$, then

- $p(\phi) \propto \phi^{k/2-1}/(h + k\phi)^{(k+h)/2}$,
- $\phi^{-1} \sim F_{h,k}$.

2. In a linear regression model the n-vector of responses y has distribution $(y \mid \beta) \sim N(X\beta, I_n)$, with mean response vector $\mu = E(y \mid \beta) = X\beta$ and identity variance matrix, where X is the $n \times p$ design matrix of rank p and β is the p-vector of regression coefficients. Suppose that the prior for β is (the so-called g-prior), $\beta \sim N(0, g^{-1}(X'X)^{-1})$ for some number $g > 0$.

   (a) What is the posterior distribution of $(\beta \mid y)$?

   (b) Show that posterior mean $E(\beta \mid y)$ can be expressed as a function of $\hat{\beta}$, the usual MLE of β.

   (c) What is the posterior mean $\tilde{\mu} = E(\mu \mid y)$?

   (d) What is the posterior variance matrix of μ?

(e) Consider the special case of an orthogonal design, so that $X'X = I_p$. Denote by μ_i the ith element of μ. Under the posterior $p(\mu \mid y)$ are μ_j and μ_k independent for $j \neq k$?

3. Paleobotanists estimate the moment in the remote past when a given species became extinct by taking cylindrical, vertical core samples well below the earth's surface and looking for the last occurrence of the species in the fossil record, measured in meters above the point P at which the species was known to have first emerged. Letting $\{y_i, i = 1, \ldots, n\}$ denote a sample of such distances above P at a random set of locations, the model

$$(y_i \mid \theta) \sim \text{Unif}(0, \theta)$$

emerges from simple and plausible assumptions. In this model the unknown $\theta > 0$ can be used, through carbon dating, to estimate the species extinction time. This problem is about Bayesian inference for θ, and it will be seen that some of our usual intuitions do not quite hold in this case.

(a) Show that the likelihood may be written as

$$l(\theta \mid y) = \theta^{-n} I(\theta \geq \max(y_1, \ldots, y_n)),$$

where $I(A) = 1$ if A is true and 0 otherwise.

(b) The Pareto distribution (written $\theta \sim \text{Pareto}(\alpha, \beta)$ has density:

$$p(\theta) = \begin{cases} \alpha\beta^\alpha\theta^{-(\alpha+1)} & \text{if } \theta \geq \beta \\ 0 & \text{otherwise} \end{cases},$$

where $\alpha, \beta > 0$.

With the likelihood viewed as a constant multiple of a density for θ, show that the likelihood corresponds to the Pareto$(n-1, m)$ distribution. Now let the prior for θ be taken to be Pareto(α, β) and derive the posterior distribution $p(\theta \mid y)$. Is the Pareto conjugate to the uniform?

(c) In an experiment conducted in the Antarctic in the 1980's to study a particular species of fossil ammonite, the following wasa linearly rescaled version of the data obtained, in ascending order: $y = (0.4, 1.0, 1.5, 1.7, 2.0, 2.1, 3.1, 3.7, 4.3, 4.9)$. Prior information equivalent to a Pareto prior with $(\alpha, \beta) = (2.5, 4)$ was available. Plot the prior, likelihood, and posterior distributions arising from this data set on the same graph, and briefly discuss what this picture implies about the updating of information from prior to posterior in this case.

(d) Make a table summarizing the mean and standard deviation for the prior, likelihood and posterior distributions, using the (α, β) choices and the data in part (d) above. In Bayesian updating the posterior mean is often a weighted average of the prior mean and the likelihood mean (with positive weights), and the posterior standard deviation is typically smaller than either the prior or likelihood standard deviations. Are each of these behaviors true in this case? Explain briefly.