Course 52558: Problem Set 2
Due April 21st, 2009

1. **Computing with a non-conjugate model**. Suppose $y_1, \ldots, y_n$ are independent samples from a Cauchy distribution with unknown center θ and known scale 1 so that

$$p(y_i \mid \theta) \propto \frac{1}{1 + (y_i - \theta)^2}$$

Assume, for simplicity, that the prior distribution for θ is uniform on $[0, 1]$. Given the observations $(y_1, \ldots, y_5) = (-2, -1, 0, 1.5, 2.5)$:

   (a) Compute (i.e. write a function for) the *unnormalized* density function $p(\theta)p(y \mid \theta)$, on a grid of points $\theta = 0, \frac{1}{m}, \frac{2}{m}, \ldots, 1$, for some large integer m. Using the grid approximation, compute and plot the *normalized* posterior density function, $p(\theta \mid y)$, as a function of θ.

   (b) Sample 1000 draws of θ from the posterior density and plot a histogram of the draws.

   (c) Use the 1000 samples of θ to obtain 1000 samples from the predictive distribution of a future observation, y_6, and plot a histogram of the predictive draws.

2. **Jeffrey's prior**. Suppose $y \mid \theta \sim \text{Poisson}(\theta)$. Find Jeffrey's prior density for θ, and then find α and β for which the Gamma(α, β) density is a close match to Jeffrey's density. Use a plot to verify that the densities are indeed similar.

3. **Binomial and Multinomial models**. Suppose data $(y_1, \ldots, y_J)$ follow a multinomial distribution with parameters $(\theta_1, \ldots, \theta_J)$. Also suppose that $\theta = (\theta_1, \ldots, \theta_J)$ has a Dirichlet distribution

$$p(\theta) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_J)}{\prod_{j=1}^{J} \Gamma(\alpha_j)} \theta_1^{\alpha_1 - 1} \ldots \theta_J^{\alpha_J - 1}$$

where $\alpha_j > 0$ are the 'prior sample sizes'. It is easy to see that the Dirichlet distribution generalizes the Beta distribution and is conjugate to the multinomial distribution that generalized the binomial distribution.

   (a) Derive the posterior $p(\theta \mid y)$.

   (b) Let $\delta = \frac{\theta_1}{\theta_1 + \theta_2}$. and derive the marginal posterior for δ.

(c) Show that $p(\delta \mid y)$ is identical to the posterior distribution for δ obtained by treating y_1 as an observation from the binomial distribution with probability δ and sample size $y_1 + y_2$, ignoring the data $y_2, \ldots, y_J$.

This result justifies the application of the binomial distribution to multinomial problems when we are only interested in two of the categories. For example, see the next problem.

4. **Comparison of two multinomial observations**. On September 25, 1988, the evening of a Presidential campaign debate, ABC News conducted a survey of registered voters in the United states; 639 person were polled before the debate and 639 different persons were polled after. The results are displayed in the following table:

Survey	Bush	Dukakis	No opinion/other	Total
pre-debate	294	307	38	639
post-debate	288	322	19	639

Assume that both surveys are independent random samples from the population of registered voters. Model the data with two different multinomial distributions. For $j = 1, 2$, let α_j be the proportion of voters who preferred Bush, out of those who had a preference for either Bush or Dukakis at the time of survey j. Plot a histogram of the posterior density $\alpha_2 - \alpha_1$. What is the posterior probability that there was a shift toward Bush?

5. **Binomial with unknown probability and sample size**. Consider data $y_1, \ldots, y_n$ modeled as iid $\text{Bin}(N, \theta)$, with both N and θ unknown. Defining a convenient family of prior distributions on (N, θ) is difficult, partly because of the discreteness of N.

Raftery (1988) considers a hierarchical approach based on assigning the parameter N a Poisson distribution with unknown mean μ. To define a prior distribution over (N, θ), Raftery defines $\lambda = \mu\theta$ and specifies a prior distribution on (λ, θ). This is done because 'it would seem easier to formulate prior information about λ, the unconditional expectation of the observations, than about μ, the mean of the unobserved quantity N'.

(a) A suggested non-informative prior distribution if $p(\lambda, \theta) \propto \lambda^{-1}$. What is a motivation for this non-informative distribution? Is the distribution improper? Transform to determine $p(N, \theta)$.

(b) The approach is applied to counts of water-bucks obtained be remote photography on five separate days in Kruger Park in South Africa. The counts were $53, 57, 66, 67, 72$. Perform the Bayesian analysis on these data and display a scatter-plot of posterior simulations of (N, θ). What is the posterior probability that $N > 100$.

(c) Why not simply use a Poisson prior distribution with a fixed μ as a prior distribution for N?