Statistics 651: Introduction to Applied Bayesian Methods

Mini-Project # 1

Due: September 19, 2008, 4pm

For each of the distributions below, do the following:

- Find the conjugate prior distribution (in symbolic notation).
- Demonstrate that it is, indeed, conjugate. That is, show that the posterior is of the same family as the prior distribution.
- Choose values for the prior distribution *and* justify your choice for those values.
- Plot the prior distribution, maximum likelihood estimate, and the posterior distribution.
- Find the posterior mean, posterior median, posterior mode, posterior variance, posterior standard deviation, and 95% central Credible interval for the parameter of interest.
- Propose three other prior distributions that you could justify (perhaps in discussion with classmates) and plot all four posterior distributions on the same graph.

The distributions are:

1. $f_*(y_i|\Theta) = (2\pi\sigma^2)^{-.5} \exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}, \qquad i = 1, \ldots, n, \Theta = \mu \text{ and } \sigma = 9$

   (data at `http://madison.byu.edu/bayes/normalmean.dat`)

   These data are 29 test scores for Statistics 221 Honors students for Exam # 1.

2. $f_*(y_i|\Theta) = (2\pi\sigma^2)^{-.5} \exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}, \qquad i = 1, \ldots, n, \Theta = \sigma^2 \text{ and } \mu = 87$

   (data at `http://madison.byu.edu/bayes/normalvariance.dat`)

   These data are 29 test scores for Statistics 221 Honors students for Exam # 1.

3. $f_*(y_i|\Theta) = \begin{pmatrix} n_i \\ y_i \end{pmatrix} \theta^{y_i}(1-\theta)^{n_i-y_i}, \qquad i = 1, \ldots, 159, \Theta = \theta$

   (data at `http://madison.byu.edu/bayes/binomial2.dat`)

   The data are the date, number of at-bats (n_i) and number of homeruns (y_i) for Barry Bonds games in the 2001 season. A "0" indicates that a homerun was *not* hit during that game.

4. $f_*(y_i|\Theta) = \frac{\lambda_i^y \exp\{-\lambda\}}{y_i}, \qquad i = 1, \ldots, 15, \Theta = \lambda$

   (data at `http://madison.byu.edu/bayes/poisson.dat`)

   The data are 15 months worth of failures of the Los Alamos National Laboratory "fastest computer in the world". The data represent the number of failures in one month which can be modelled with a Poisson distribution. After talking with an expert they could only say that they thought there "ought" to be no more than 10 failures in one month.

5. $f_*(y_i|\Theta) = \frac{1}{\lambda} \exp\left\{-\frac{y_i}{\lambda}\right\}, \qquad i = 1, \ldots, 141, \Theta = \lambda$

   (data at `http://madison.byu.edu/bayes/exponential.dat`)

   The data represent lengths (in miles) of 141 major rivers in North America, as compiled by the US Geological Survey.