Introduction to Bayesian Data Analysis
Tutorial 2

(1) Assume $y_1, ..., y_n|\theta \overset{\text{iid}}{\sim} Pois(\theta)$. Assume a congugate prior for θ with parameters α and β. Let $\tilde{y}$ be an unobserved value of y. Derive the posterior predictive distribution $p(\tilde{y}|y_1, ..., y_n)$. Show that $Var(\tilde{Y}|y_1, ..., y_n) = E[\theta|y_1, ..., y_n] \times \frac{\beta+n+1}{\beta+n}$ and interpret this result.

(2) Problem 3.3 (Hoff).
Tumor counts: A cancer laboratory is estimating the rate of tumorigenesis in two strains of mice, A and B. They have tumor count data for 10 mice in strain A and 13 mice in strain B. Type A mice have been well studied, and information from other laboratories suggests that type A mice have tumor counts that are approximately Poisson-distributed with a mean of 12. Tumor count rates for type B mice are unknown, but type B mice are related to type A mice. The observed tumor counts are:

$$y_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6)$$

$$y_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)$$

(a) Find the posterior distributions, means, variances and 95% quantile-based confidence intervals for θ_A and θ_B, assuming a Poisson sampling distribution for each group and the following prior distribution

$\theta_A \sim Gamma(120, 10)$, $\theta_B \sim Gamma(12, 1)$, $p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$

(b) Compute and plot the posterior expectation of θ_B under the prior distributions $\theta_B \sim \text{Gamma}(12 \times n_0, n_0)$ for each value of $n_0 \in \{1, 2, ..., 50\}$. Describe what sort of prior beliefs about θ_B would be necessary in order for the posterior expectation of θ_B to be close to that of θ_A.

(c) A new mouse of type B is delivered to the lab. Predict the expected tumor counts for the new mouse assuming:

(i) Independent priors for θ_A and θ_B.

(ii) The data from mice A form a prior distribution for the posterior of θ_B.

(d) Should knowledge about population A tell us anything about population B? Discuss whether or not it makes sense to have $p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$.

(3) Problem 3.4 (Hoff) Estimate the probability θ of teen recidivism based on a study in which there were n=43 individuals released from incarceration and y=15 re-offenders within 36 months.

(a) Using a beta(2,8) prior for θ, plot $p(\theta)$, $p(y|\theta)$ and $p(\theta|y)$ as functions of θ. Find the posterior mean, mode and standard deviation of θ. Find a 95% quantile-based confidence interval.

(b) Repeat (a), but using a beta(8,2) prior for θ.

(c) Consider the following prior distribution for θ:

$$p(\theta) = \frac{1}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}[3\theta(1-\theta)^7 + \theta^7(1-\theta)]$$

which is a 75-25% mixture of a beta(2,8) and a beta (8,2) prior distribution. Plot this prior distribution and compare it to the priors in (a) and (b). Describe what sort of prior opinion this may represent.

(d) For the prior in (c):

(i) Write out mathematically $p(\theta) \times p(y|\theta)$ and simplify as much as possible.

(ii) The posterior distribution is a mixture of two distributions you know. Identify these distributions.

(iii) On a computer, calculate and plot $p(\theta) \times p(y|\theta)$ for a variety of θ values. Also find (approximately) the posterior mode, and discuss its relation to the modes in (a) and (b).

(e) Find a general formula for the weights in the mixture distribution in (d)(ii), and provide an interpretation of their values.