AMS 207 **Mary Silva**
Assignment 2 **8 May 2018**

Write the Bayes factor, BIC, DIC and Gelfand and Ghost criterion to compare a model where n observations are assumed to be sampled with a poisson distribution with a gamma prior, to a model where the observations are sampled from a binomial distribution, with a fixed, large, number of trials and beta prior for the probability of success.

1. Consider the data on deaths by horse kicks in the Prussian army. Fit the data using the two different models.

   Using a poisson distribution with a gamma prior we have
   The Likelihood:

$$f_1(\boldsymbol{x}|\lambda) = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!}$$

   The prior (model 1):

$$p_1(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp -\lambda\beta$$

   The posterior:

$$\begin{aligned} p_1(\lambda|\boldsymbol{x}) &\propto f(\boldsymbol{x}|\lambda)p(\lambda) & (1)\\ &= \frac{\beta^\alpha}{\Gamma(\alpha)\prod x_i!}\exp\Big\{-\lambda(n+\beta)\Big\}\lambda^{\sum x_i+\alpha-1} & (2)\\ &\sim Gamma\left(\alpha^* = \sum x_i + \alpha, \beta^* = n+\beta\right) & (3)\end{aligned}$$

   Using a binomial distribution, with a fixed, large number of trials and a beta prior for the probability of success.
   The Likelihood:

$$f_2(\boldsymbol{x}|\theta) = \prod_1^{20}\binom{N}{x_i}\theta^{\sum x_i}(1-\theta)^{N*n-\sum x_i}$$

   The prior:

$$p_2(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

   Throughout this paper, I use large $N = 10^6$
   The posterior:

$$\begin{aligned} p_2(\lambda|\boldsymbol{x}) &\propto \left[\prod_1^{20}\binom{N}{x_i}\right]\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\sum x_i\alpha-1}(1-\theta)^{20N+\beta-\sum x_i-1} & (4)\\ &\sim Beta\left(\alpha^* = \sum x_i + \alpha, \beta^* = 20N+\beta-\sum x_i\right) & (5)\end{aligned}$$

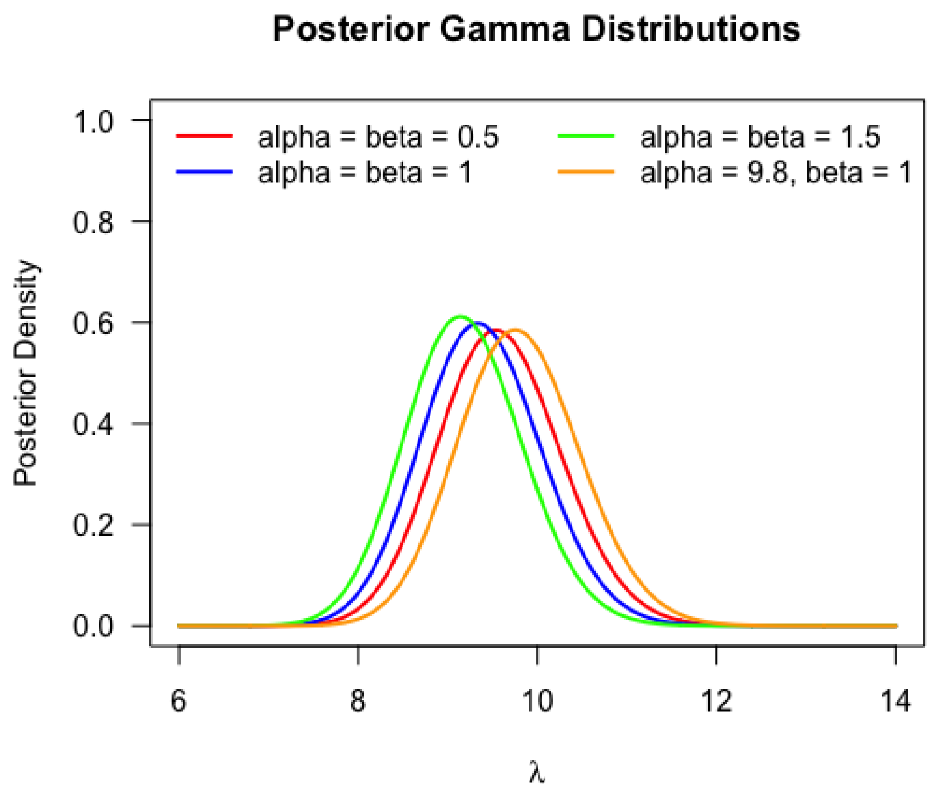
2. Perform a prior sensitivity analysis.

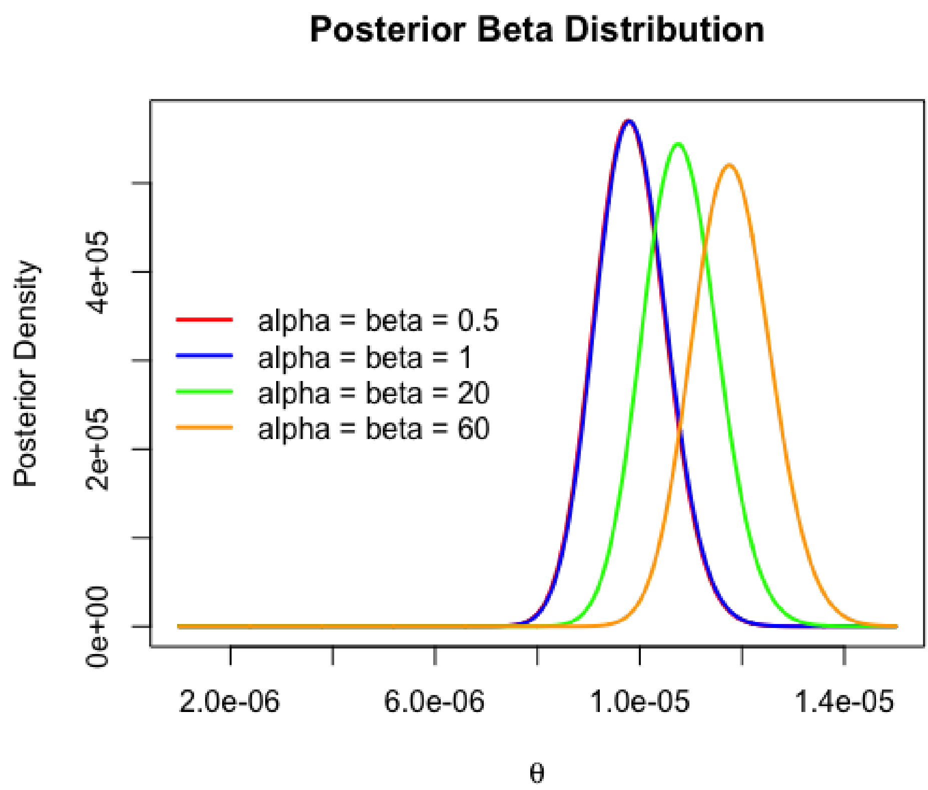
For our poisson-gamma model, we want to choose the appropriate α and β. To do this we use the mean of the data:

$$\text{mean(horsekick data)} = 9.8$$

We want to select α and β such that the mean for the gamma distribution, (α/β), is proportional to the mean for our poisson distribution, 9.8. So we have $\alpha = 9.8$ and $\beta = 1$.

We plot the posterior distributions using different α and β values and attaining the α^* and β^* values obtained in equation (3) and (5) respectively.





So for model 1: we decide that $\alpha = 9.8$ and $\beta = 1$ is appropriate.

For model 2: Based on the plots, the differences are small so any option would not effect the posterior distribution so greatly. We choose $\alpha = \beta = 1$ for convenience.

3. With n being the number of horse kick observations, we define the following criterion:

$$BIC = -2 \times \text{Likelihood} - \log(n)$$

For the Deviance Information Criterion (DIC), the deviance statistic is given by

$$D(\theta) = -2 \log \text{Likelihoods} + 2 \log h(x)$$

where $h(x)$ is the standardizing function. Then the DIC is given by

$$DIC = \bar{D} + (\bar{D} - D(\bar{\theta})) = 2\bar{D} - D(\bar{\theta})$$

Where $D(\bar{\theta})$ is the mean of posterior samples.

For the Gelfand and Ghosh criterion, denote the observed data as $\mathbf{x}$ and z as replicates of the observed data. Let $\mu_l = E(z_l|\boldsymbol{x})$ and $\sigma_l^2 = Var(z_l|\boldsymbol{x})$

$D = G + P$, where $G = \sum_l (\mu_l - x_l)^2$ and $P = \sum_l \sigma_l^2$

D seeks to reward goodness of fit penalizing complexity. So the smaller the D the better.

We provide a table specifying our R calculations

	BIC	DIC	Gelfand and Ghosh
model 1	122.15	121.17	573.20
model 2	122.15	121.18	575.86

From the table, we see that the BIC, DIC, and G&G criterion are slightly smaller for model 1 (the gamma posterior model). This normally implies that model 1 is the "better" model. But in this case since the difference is slight, we conclude that for model 2 (the beta posterior model) when we use a fixed, large, number of trials for the binomial prior as we have done, both model 1 and model 2 are essentially equally accurate (this language might be wrong but i think you get the point).

Next we look at the Bayes factor, which is simply a ratio of the marginals for model 1 and model 2.

Here we compute the marginals:

$$\begin{aligned} m_1(x) &= \int f_1(x|\lambda)p_1(\lambda)d\lambda && (6)\\ &= \frac{\beta^\alpha}{\Gamma(\alpha)\prod(x_i!)}\int e^{-\lambda(n+\beta)}\lambda^{\sum x_i+\alpha-1}d\lambda && (7)\\ &= \frac{\beta^\alpha}{\Gamma(\alpha)\prod(x_i!)}\frac{\Gamma(\sum x_i+\alpha)}{(n+\beta)^{\sum x_i+\alpha}} && (8) \end{aligned}$$

$$\begin{aligned} m_2(x) &= \int f_2(x|\lambda)p_2(\lambda)d\lambda && (9)\\ &= \left[\prod_1^{20}\binom{N}{x_i}\right]\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int \theta^{\sum x_i \alpha-1}(1-\theta)^{20N+\beta-\sum x_i-1} && (10)\\ &= \left[\prod_1^{20}\binom{N}{x_i}\right]\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\sum x_i+\alpha)\Gamma(20N+\beta-\sum x_i)}{\Gamma(\alpha+20N+\beta)} && (11) \end{aligned}$$

To calculate the Bayes factor, we use our previously defined α and β parameters from part 2, and compute the ratio of the log marginals and then exponentiate the value.

$$\text{Bayes Factor} = \exp\{\log m_1 - \log m_2\} = 123322.2$$

Our Bayes factor is extremely large. Again, this criterion favors model 1.