Problem Set #2

Due Feb. 26th

1 Analytical

1. Prove that the highest posterior density (HPD) region is the shortest interval containing $1-\alpha$ percent probability by showing that it intersects the posterior distribution at the same probability $p[b|x] - p[a|x] = 0$. Hint: set this problem up as a constrained minimization problem and solve for $\min_{\{a,b\}}(b-a)$

2. Prove the gamma distribution $p[\mu|\alpha,\beta] = \frac{\beta^\alpha}{\Gamma[\alpha]}\mu^{\alpha-1}e^{-\beta\mu}, \quad \mu,\alpha,\beta > 0$ is the conjugate prior for μ in a Poisson likelihood function,

$$p[x|\mu] = \left(\prod_{i=1}^{n} x_i!\right)^{-1} e^{log[\mu]\sum_{i=1}^{n} x_i} e^{-n\mu} \tag{1.1}$$

that is, calculate the posterior distribution of μ and show that it is also gamma distributed. (Hint: work with the kernel of the distribution)

3. The multinomial model

$$p[x|\theta] = \frac{n!}{x_1!\cdots x_k!}\theta_1^{x_1}\cdots\theta_k^{x_k} = \frac{n!}{\prod_i^k x_i!}\prod_i^k \theta_i^{x_i} \tag{1.2}$$

is a generalization of the binomial model where the number of discrete outcomes is greater than two. Prove that the conjugate prior for the generalized binomial is the generalized beta distribution, which is also called the Dirichlet distribution and is of the form:

$$p[\theta|\alpha] = \frac{\Gamma[\sum_i \alpha_i]}{\prod_i \Gamma[\alpha_i]}\prod_{i=1}^{k}\theta_i^{\alpha_i-1} \tag{1.3}$$

4. The entropy difference between two distributions is measured by the Kullback-Leibler divergence. Given any n observations of data $x \in \mathbb{R}$ we can always coarse-grain x into a K-dimensional vector of binned frequencies, $f[x_k] = \{f_1, ..., f_K\}$ where $f_j = \frac{x_j}{n}$, and view this data as a sample of a multinomial model with prior frequencies $q[x_k] = \{q_1, \ldots, q_K\}$. The likelihood function for a given sample $x = \{x_1, \ldots, x_K\}$ (where x_j is the number of observations in bin j) from a multinomial distribution, writing $n = \sum_{k=1}^{K} x_k$ for the sample size, is:

$$p[x|\theta] = \frac{n!}{x_1! \ldots x_K!} q_1^{x_1} \ldots q_K^{x_K} \tag{1.4}$$

Using Stirling's approximation to the factorial prove that

$$p[x|\theta] = e^{-n\mathrm{D}[f||q]} \tag{1.5}$$

That is, show that the log of the multinomial likelihood is equivalent to the Kullback-Leibler divergence between the empirical bin frequencies and prior probabilities.

2 R Exercises

1. Imagine an urn filled with 100 balls of unknown composition. You are told that the contents consist of red and green balls of any possible combination. Consider two priors over the distribution of red and green balls which we denoted $p[\{r, g\}]$:

   - A) a uniform prior over all possible combinations of balls, which assumes that each pair of numbers of red and green balls, $\{0, 100\}, \{1, 99\}, \{2, 98\}..., \{100, 0\}$ is equally likely to occur in the string.
   - B) a prior that each complete data string is equally likely. That is, we assign an equal probability to each possible combination of a complete draw of all balls from the urn. If we denote a draw of a red ball 1 and a draw of a green ball 0 there are 2^{100} possible strings of the form:
     $\{0, 0, 0, ..., 0\}, \{1, 0, 0, ..., 0\}, \{1, 1, 0, ..., 0\}, ..., \{0, 0, 0, ..., 1\}$

   Note: Since $100 - r = g$, where g is the number of green balls, it suffices to determine the probabilities p_r, as it follows that $p_g = 1 - p_r$.

   a) What probability does prior A assign to each composition of the urn? Plot this probability mass function in R.

   b) What prior probability does prior B assign to each of the possible data strings? Plot this probability mass function in R. (Hint: use the Binomial coefficient).

   c) Is either of these priors informative? Would you favor either one of them? Is either or both exchangeable?

d) Say we draw 10 balls from the urn with replacement and observe 7 red balls and 3 green balls. If $0 \le \theta \le 1$ is the unknown parameter representing the composition of the urn (as in the Binomial model) what is the likelihood function for this data? Plot the likelihood in R.

e) Plot the normalized posterior distribution in R for each prior. What can you say about posterior inference for the composition of the urn for each prior?

f) Calculate the max posterior from having seen 7 red balls and 3 green balls for each prior.