Stat 591 – Homework 04 **Due: Wednesday 10/30**

Your group should submit a write-up that includes solutions the problems stated below, along with any <u>*relevant*</u> *pictures/graphs or computer code/output.*

1. Problem 5.1 in [GDS].

2. Problem 5.2 in [GDS].

3. (Based on Problem 5.5 in [GDS].) Consider the multinomial setup you've seen in previous homework. That is, let $X = (X_1, X_2, X_3, X_4)$ and suppose that, for a given $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$, X has a four-dimensional multinomial distribution, $\mathsf{Mult}_4(n; \theta)$ with parameters n and θ. Recall, the Dirichlet prior, $\mathsf{Dir}_4(a)$, is a distribution on the four-dimensional probability simplex, parametrized by $a = (a_1, a_2, a_3, a_4)$, where each a_i is a non-negative number. This time, write the Dirichlet distribution in terms of a precision parameter m and a mean parameter $p = (p_1, p_2, p_3, p_4)$. The connection is $a = mp$, so that $m = \sum_{i=1}^4 a_i$ and $p_i = a_i/m$. Write $\mathsf{Dir}_4(m, p)$ for this "new" Dirichlet distribution.

   (a) Go to Appendix A.1 in [GDS]. Rewrite the mean vector and dispersion matrix formulas provided there in terms of the new parameters (m, p). Explain why m is called the precision parameter by looking at how the dispersion matrix changes as $m \to 0$ and as $m \to \infty$.

   (b) An idea towards defining a "non-informative" prior for the multinomial problem is to let the precision parameter $m \to 0$.

      i. Use your answer to part (a) to explain the intuition behind this choice.

      ii. There is a problem with this approach, however. Despite the intuition, the limiting prior, as $m \to 0$, is actually very informative—it's a discrete distribution, with masses at the corners of the simplex.[1] To convince yourself of this, consider the marginal distribution of θ_i under the $\mathsf{Dir}_4(m, p)$ prior. First show that, for each $i = 1, \ldots, 4$,

$$\mathsf{E}_{(m,p)}(\theta_i) \to p_i \quad \text{and} \quad \mathsf{V}_{(m,p)}(\theta_i) \to p_i(1 - p_i), \quad \text{as } m \to 0.$$

      For this, use the fact that a Dirichlet distribution has beta marginals.[2] Now argue that, since each θ_i is $\mathsf{Ber}(p_i)$ in the limit, the limiting prior must be such that one of the four corners[3] is selected at random.

   (c) Alternatively, one can define a "non-informative posterior" by taking $m \to 0$ after updating prior to posterior. Go back to your work on this multinomial–Dirichlet problem from Homework 2 and write down the posterior distribution for θ, given X, when $m \to 0$.

[1] For a precise statement of this result, see Ghosh and Ramamoorthi, *Bayesian Nonparametrics*, Springer 2003, page 93.

[2] `http://en.wikipedia.org/wiki/Dirichlet_distribution#Marginal_beta_distributions`

[3] In this case, the corners are $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, and $(0, 0, 0, 1)$.

(d) Re-do the simulations in the previous homework using this "non-informative" posterior. Plot, side by side, histograms of the posterior distribution for κ, one for this new "non-informative" posterior and one for your previous "informative" posterior. How do these plots compare?

4. (Based on Problem 5.10 in [GDS].) Let $(X_1, \ldots, X_n) \mid (\mu, \sigma^2) \overset{\text{iid}}{\sim} \mathsf{N}(\mu, \sigma^2)$.

   (a) Consider the prior $\pi(\mu, \sigma^2) \propto 1/\sigma^2$. This is the reference prior (see Example 5.4 in [GDS]) and also the right invariant prior, but not Jeffreys prior. Find the posterior distribution for (μ, σ^2).

   (b) Find the marginal posterior distribution for μ. That is, integrate out σ^2 from the joint posterior distribution for (μ, σ^2). Do you know this distribution?

   (c) Find a $100(1-\alpha)\%$ credible upper bound for μ using the marginal posterior. That is, find $\bar{\mu}_\alpha$ such that $\Pi(\mu \leq \bar{\mu}_\alpha \mid X_1, \ldots, X_n) = 1 - \alpha$.

   (d) Think of $\bar{\mu}_\alpha = \bar{\mu}_\alpha(X_1, \ldots, X_n)$ as a function of data. Show that the above credible upper bound is exactly probability matching. That is, show that

$$\mathsf{P}_{(\mu,\sigma^2)}\{\bar{\mu}_\alpha(X_1, \ldots, X_n) \geq \mu\} = 1 - \alpha, \quad \forall\ (\mu, \sigma^2).$$

5. (Based on Problem 5.11 in [GDS].) Given θ, let $X \sim \mathsf{Bin}(n, \theta)$.

   (a) Find Jeffreys prior and the corresponding posterior. (Jeffreys prior is the same whether you work with binomial directly or with iid Bernoulli's.)

   (b) Given $\alpha \in (0, 1)$, let $\bar{\theta}_\alpha(X)$ be such that $\Pi(\theta \leq \bar{\theta}_\alpha(X) \mid X) = 1 - \alpha$. For $\alpha = 0.05$, how do you compute $\bar{\theta}_\alpha(X)$ in R?

   (c) For an upper 95% confidence limit for θ, a Bayesian might use $\bar{\theta}_{0.05}(X)$ as defined above. A frequentist might use the bound $\hat{\theta} + 1.65\{\hat{\theta}(1-\hat{\theta})/n\}^{1/2}$, where $\hat{\theta} = X/n$ is the MLE. Perform simulations to compare the coverage probability of these two 95% upper confidence limits.

      — Consider $n \in \{50, 100\}$ and θ on a grid $\{0.05, \cdots, 0.95\}$ of length 20. For each (n, θ) pair, simulate 1000 $\mathsf{Bin}(n, \theta)$ and compute the Bayesian and frequentist upper limits and record the coverage proportion for each.

      — Draw two figures, one for each n. In each figure, plot the coverage probabilities for the two methods as functions of θ.

   Explain what you see in these plots. In particular, do you think the Bayes or the frequentist method is better? Explain your answer.[4]

[4] There is a really nice paper that discusses this simple but important problem; see Brown, Cai, and DasGupta, "Interval estimation for a binomial proportion," *Statistical Science*, 2001.