## APPM 4/5720: Computational Bayesian Statistics, Spring 2018

## Problem Set Two (Due Wednesday, February 7th)

1. Suppose that we have $n$ independent trials of an experiment where each trial can result in one of two possible outcomes labeled "success" ($S$) or "failure" ($F$). Suppose that the probability of success remains the same from trial to trial. Let $\theta = P(\text{success})$ for each trial. Let $X$ be the number of successes in $n$ trials. We know that $X$ has a binomial distribution with parameters $n$ and $\theta$ (We write $X \sim bin(n, \theta)$.) and that $X$ has pdf

$$P(X = x) = \left( \begin{array}{c} n \\ x \end{array} \right) \theta^x (1-\theta)^{n-x} \, I_{\{0,1,\ldots,n\}} = \frac{n!}{k!(n-k)!} \theta^x (1-\theta)^{n-x} \, I_{\{0,1,\ldots,n\}}.$$

If we generalize this to $k$ possible outcomes/categories with probabilities $\theta_1, \theta_2, \ldots, \theta_k$, respectively ($\sum_{i=1}^k \theta_i = 1$) and let $X_i$ be the number (in $n$ trials) of outcomes in category $i$ ($i = 1, 2, \ldots, k$), we can collect the counts as a vector $(X_1, X_2, \ldots, X_k)$. Note that we must have $\sum_{i=1}^k X_i = n$.

The vector is said to have a **multinomial distribution** and the pdf is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} \, \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_k^{x_k}.$$

Here, the $x_i$ are non-negative integers that must sum to $n$.

We have seen that the (natural) conjugate prior for the binomial distribution is the Beta distribution. The Beta distribution with parameters $a, b > 0$ has pdf

$$f(\theta) = \frac{1}{\mathcal{B}(a,b)} \, \theta^{a-1} (1-\theta)^{b-1} \, I_{(0,1)}(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \, \theta^{a-1} (1-\theta)^{b-1} \, I_{(0,1)}(\theta).$$

A generalization of the Beta distribution for the single probability $\theta$ to $k$ probabilities $\theta_1, \theta_2, \ldots, \theta_k$ that sum to 1 is the **Dirichlet distribution** which has pdf

$$f(\theta_1, \theta_2, \ldots, \theta_k) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \, \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \cdots \theta_k^{\alpha_k - 1}$$

where the $\alpha_i > 0$ are parameters of the Dirichlet distribution.

Show that the Dirichlet distribution is a conjugate prior for the multinomial distribution.

---

2. Suppose that $X_1, X_2, \ldots, X_n$ are iid random variables and that $Z$ is another random variable independent from the $X_i$. (Assume that $Z$ is not a degenerate random variable. i.e. That it is not a constant with probability 1.)

Consider the random variables $Y_i := X_i + Z$ for $i = 1, 2, \ldots, n$.

   (a) Show that the $Y_i$ are pairwise dependent. That is, for any $i \neq j$, show that $Y_i$ is not independent of $Y_j$.

   (b) Show that $Y_1, Y_2, \ldots, Y_n$ are exchangeable.

3. Simulate $10,000$ values from an exponential distribution with rate $\lambda$ for your choice of $\lambda$. Provide a histogram of your results with the exponential density superimposed. (Do not use a built-in function for exponential random variable generation. **Do** use a built-in function for uniform random variable generation!) Include your code when you turn in this HW.

---

4. Recall the Pólya Urn example from class where we started with $R_0$ red balls, $W_0$ white balls. At each step, we randomly selected a ball, replaced it, and put in an additional $c = 1$ balls of the same color. We let

$$X_i = \begin{cases} 1 & , \text{ if the } ith \text{ ball selected was red} \\ 0 & , \text{ otherwise} \end{cases}$$

The sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the proportion of red balls that have been observed by time step $n$.

In class, we saw that $\overline{X}_n \stackrel{d}{\to} X$ where $X \sim Beta(R_0, B_0)$.

Find the limiting distribution for $\overline{X}_n$ for a general value of $c$.

---

5. **[Required for 5720 Students Only:]**

   (a) Does every distribution have a conjugate prior? Prove or give a counterexample and explain.

   (b) Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with parameter $\theta$. Suppose that $\theta$ is one-dimensional.

   We say that the sample comes from a **one-parameter exponential family** if the joint pdf for $X_1, X_2, \ldots, X_n$ can be written as

$$f(\vec{X}|\theta) = a(\theta)b(\vec{x})\exp[c(\theta)d(\vec{x})]$$

   for some functions $a(\cdot)$, $b(\cdot)$, $c(\theta)$ and $d(\cdot)$ with $c$ and $d$ non-constant.

   For example, if $X_1, X_2, \ldots, X_n$ are iid with the geometric pdf

$$f(x|\theta) = (1-\theta)^x\theta \; I_{\{0,1,2,\ldots,\}}(x),$$

   the joint pdf is

$$f(\vec{x}|\theta) = (1-\theta)^{\sum x_i}\theta^n \prod_{i=1}^{n} I_{\{0,1,2,\ldots,\}}(x_i),$$

   which can be rewritten as

$$f(\vec{x}|\theta) = \underbrace{\theta^n}_{a(\theta)} \; \underbrace{I_{\{0,1,2,\ldots,\}}(x_i)}_{b(\vec{x})} \; \exp[\underbrace{\ln(1-\theta)}_{c(\theta)} \underbrace{(\textstyle\sum x_i)}_{d(\vec{x})}].$$

   Show that every exponential family has a conjugate prior.