Introduction to Bayesian Modeling

Homework Assignment #1

Due 3 March 2020 by the start of class.

1. An enzyme-linked immunosorbent assay (ELISA) test is performed to determine if the human immunodeficiency virus (HIV) is present in the blood of individuals. The ELISA test is not perfect. Suppose that the ELISA test correctly indicates HIV 99% of the time, and that the proportion of the time that it correctly indicates no HIV is 99.5%. Suppose that the prevalence among blood donors is known to be 1/10,000.

   (a) What proportion of blood that is donated will test positive using the ELISA test?

   (b) What proportion of the blood that tests negative on the ELISA test is actually infected with HIV?

   (c) What is the probability that a positive ELISA outcome is truly positive, that is, what proportion of individuals with positive outcomes are actually infected with HIV?

2. An article in the July 22, 2009 edition of the International Herald Tribune indicated that 7% of British children attend special schools that cater to privileged parents and that 75% of all judges are known to have attended such schools. The implication was that underprivileged, but presumably intelligent and hardworking children were being excluded from high ranking professions like judgeships. We want to look at the relative probabilities of being a judge given that one did or did not attend an elite school.

   Specifically, let E denote attending an elite school with E^c the complement. Let J denote becoming a judge with J^c the complement. Let $p = \Pr(J)$ be the (unknown to us) proportion of judges among the populace in Great Britain. We are given that $\Pr(E|J) = 0.75$ and $\Pr(E) = 0.07$.

   (a) Use the definition of conditional probability to find $\Pr(J|E)$ and $\Pr(J|E^c)$ as functions of p. Find an actual number for the ratio $\Pr(J|E)/\Pr(J|E^c)$. What can you say about the effect of availability of elite schooling on the prospects of becoming a judge in Great Britain?

   (b) Let $q = \Pr(E)$. What value of q would correspond to no effect of E on the chances of becoming a judge later in life?

3. Show that the mode of the $\mathrm{Beta}(a, b)$ distribution is $(a-1)/(a+b-2)$ when $a, b > 1$, zero when $a < 1 \le b$ and one when $b < 1 \le a$. Discuss the behavior of the mode when both a and

b are less than one.

4. Consider the expression

$$f(y|\theta) = \prod_{i=1}^{n} f_i(y_i|\theta),$$

and define $f_i(y_i|\theta) \equiv f(y_i|\theta, x_i)$ where x_i denotes a known "covariate" variable. In particular, let the x_i's identify two groups, $x_i = 1$ for $i = 1, \ldots, k$ and $x_i = 0$ for $i = k+1, \ldots, n$. With $\theta = (\theta_1, \theta_2)'$, define $\lambda_i = \theta_1^{x_i}\theta_2^{1-x_i}$ so that λ_i equals θ_1 if $x_i = 1$ and equals θ_2 if $x_i = 0$.

Now consider $f(y_i|\theta, x_i) \equiv f(y_i|\lambda_i)$ for each of the choices:

(i) $f(y_i|\lambda_i) = \lambda_i e^{-\lambda_i y_i}$,
(ii) $f(y_i|\lambda_i) = \lambda_i^{y_i} e^{-\lambda_i}/y_i!$,
(iii) $f(y_i|\lambda_i) = 2\lambda_i y_i e^{-\lambda_i y_i^2}$.

(a) Using Table 2.1, for each of the three choices identify the distribution with density $f(y_i|\lambda_i)$ when $x_i = 1$ and also when $x_i = 0$.

(b) For each choice, simplify the product $\prod_{i=1}^{n} f_i(y_i|\theta)$ so that as a function of θ it is proportional to a function that depends only on some combination of $\sum_{i=1}^{k} g(y_i)$, $\sum_{i=k+1}^{n} g(y_i)$, $\prod_{i=1}^{k} g(y_i)$, and $\prod_{i=k+1}^{n} g(y_i)$ where $g(x)$ can be one of x, x^2, or $x!$.

5. Derive the following posterior densities and show that these densities are the densities shown in Table 2.3.

(a) An exponential sample with a Gamma (α, β) prior.

(b) A normal sample with known mean, μ, and a Gamma (α, β) prior on the precision, τ.

6. Derive the posterior distribution for the conjugate prior by using the *complete the square formula*

$$r(\mu - v)^2 + s(\mu - w)^2 = (r+s)(\mu - \hat{\mu})^2 + \frac{rs}{r+s}(v-w)^2, \quad \hat{\mu} = \frac{r}{r+s}v + \frac{s}{r+s}w,$$

and adapting the arguments illustrated in the previous subsection for the SIR prior.

(a) Derive the conditional density $p(\mu|\,\tau, y)$. Show that it is the normal density given in this subsection.

(b) Using the result in (a), obtain the marginal densities $p(\mu|y)$ and $p(\tau|y)$.

(c) Derive the predictive density for a future observation based on this model.