STATISTICS 225 – Fall 2018

Homework 1 - handed out Thursday October 4, 2018

DUE DATE: Tuesday October 16, 2018

Reading:	Oct 4	prior distns (2.8, 2.9)
	Oct 9 - Oct 11	multiparameters models (3.1-3.3,3.7-3.8, Ch. 4)
	Oct 16 - Oct 18	hierarchical models (Chapter 5)

NOTE: To be fair to all, I should point out that solutions to some of the problems in the book are available on the web at *www.stat.columbia.edu/~gelman/book*. I will not assign many of these problems during the course but I may occasionally assign one. Note also that there is no guarantee that posted solutions are 100% correct.

PROBLEMS:

1. **Loss functions.** In this class we emphasize the simulation-based approach to inference. The traditional Bayesian approach to point estimation is the decision-theory approach. To provide a point estimate of a parameter θ, we introduce a loss function $L(\theta, a) \geq 0$ that gives the loss (or cost) for giving a as an estimate when θ is the true value. The Bayes estimate of θ is the value of a that minimizes the posterior expected loss $E(L(\theta, a)|y) = \int L(\theta, a)p(\theta|y)d\theta$. You can assume θ is a scalar parameter with continuous posterior density for the problem.
   (a) Show that if $L(\theta, a) = (\theta - a)^2$ (squared error loss), then the posterior mean, $E(\theta|y)$, is the unique Bayes estimate of θ. (We assume $E(\theta|y)$ exists.)
   (b) Show that if $L(\theta, a) = |\theta - a|$, then the posterior median of θ is a Bayes estimate of θ.

2. **Prior distributions for the Poisson sampling distribution**
   The Poisson distribution, with density $p(y|\lambda) = \lambda^y e^{-\lambda}/y!$ for $\lambda > 0$ and $y = 0, 1, 2, \ldots$, is commonly used to describe the distribution of the number of occurrences of some kind of event (e.g., deaths due to cancer, meteor strikes in some geographic area). Assume we observe $y_1, y_2, \ldots, y_n$ as independent (given λ) Poisson random variables. We explore the development of prior distributions for the parameter λ.
   (a) Show that the gamma family of distributions (with density $p(x|\alpha, \beta) = \beta^\alpha x^{\alpha-1} e^{-\beta x}/\Gamma(\alpha)$ for $\alpha, \beta, y > 0$) is the family of conjugate prior distributions for λ. To do this assume λ has a gamma distribution and show that the posterior distribution of λ given the data is a gamma distribution.
   (b) Consider collecting data on the number of meteors (of a given minimum size) that strike the mainland U.S. in a calendar year. A Poisson model might be appropriate here. You consult an expert who indicates they believe the average rate of meteor strikes is about 1 per year but they would not be surprised if the rate were found to differ from that by a factor of 10 (in either direction). Choose a suitable prior distribution from the conjugate prior family. Explain how you identify your choice.
   (c) Noninformative prior distributions
      i. One possible noninformative prior distribution is the flat prior distribution $p(\lambda) \propto 1$ for $\lambda > 0$ (roughly equivalent to a gamma distribution with $\alpha = 1$ and $\beta = 0$). Is this a proper distribution (i.e., is it nonnegative and does it integrate to one if suitably normalized)? Does this lead to a proper posterior distribution?
      ii. Show that Jeffreys' noninformative prior distribution for λ is $p(\lambda) \propto 1/\sqrt{\lambda}$ on the positive half of the real line.
      iii. One weakness of the argument for choosing a flat prior distribution is that we don't know on what scale we should be "flat". Show that Jeffreys' prior distribution is equivalent to a uniform prior distribution on $\phi = \sqrt{\lambda}$.
      iv. Is the Jeffreys' prior distribution a proper distribution? Does it lead to a proper posterior distribution? (Hint: It depends on the data.)

3. **Weibull distribution with a conjugate prior distribution**
   The Weibull distribution is a distribution that is often used for lifetimes of equipment/parts. It actually has two parameters but for the moment I'm assuming one of those is fixed (at two). The Weibull(2) density is $f(y|\theta) = 2\theta y e^{-\theta y^2}$ for $0 < y < \infty$. The parameter θ is something like the "inverse lifetime" parameter (large θ means short lifetimes; small θ means long lifetimes). The mean of the distribution is $.886\theta^{-0.5}$. Suppose we observe data $y_1, y_2, \ldots, y_n$ as independent (given θ) samples from the Weibull(2) distribution.

(a) Show that the gamma family of distributions is the conjugate family and derive the posterior distribution of θ assuming that $\theta \sim Gamma(\alpha, \beta)$.

(b) Derive the marginal distribution $p(y_1, \dots, y_n)$. (Hint: There are two related ways to obtain the marginal. (1) Use $p(y_1, \dots, y_n) = \int p(y_1, \dots, y_n|\theta)p(\theta)d\theta$; or (2) Use $p(y_1, \dots, y_n) = p(y_1, \dots, y_n|\theta)p(\theta)/p(\theta|y_1, \dots, y_n)$.) Explain why we would say $y_1, \dots, y_n$ are independent conditionally but not independent in their marginal distribution.

(c) In one application the lifetime of a kind of gear was measured in 100,000's of hours (so $y = 1$ means the part lasted 100,000 hours and $y = 0.5$ means the part lasted 50,000 hours). Gears tend to last between 50,000 and 500,000 hours (depending on the size and manufacturer), with 100,000 being a typical lifetime. This suggests a gamma prior distribution for θ with $\alpha = 1.4, \beta = 2$. Obtain a graph of this prior density and argue that this choice of prior is reasonable given the information provided.

(d) Suppose that we observe $n = 10$ with $y = (0.25, 0.52, 0.60, 0.91, 0.97, 1.00, 1.07, 1.09, 1.18, 1.48)$. (Evidently this gear is fairly typical with lifetimes around 100,000). Find and graph the posterior distribution. Give the posterior mean and variance. Give a 95% posterior interval for θ.

(e) What would be your prediction for the lifetime of a new part? Give a 95% predictive interval for the lifetime of a new part.

4. **Mixtures of conjugate prior distributions.** The use of conjugate families for prior distributions is often criticized as being too restrictive. However, we can increase the flexibility of these families by considering mixtures of conjugate prior distributions. Let's suppose that the data model is $p(y|\theta)$ and that $\pi_m(\theta), m = 1, \dots, k$ are k members of the conjugate family for $p(y|\theta)$. Recall that this means that $p_m(\theta|y) = p(y|\theta)\pi_m(\theta)/g_m(y)$ is is in the same parametric family as $\pi_m(\theta)$ (where $g_m(y) = \int p(y|\theta)\pi_m(\theta)d\theta$). Consider a prior distribution which is a mixture of distributions from the conjugate family, $\pi(\theta) = \sum_{m=1}^{k} \lambda_m \pi_m(\theta)$ where $\sum_{m=1}^{k} \lambda_m = 1$.

(a) Show that the posterior distribution is also a mixture of distributions from the conjugate family.

(b) Identify the new mixture proportions.

(c) Application: You visit another country where the coins tend to land heads more often than tails (approx 75/25 split). However there is a significant counterfeiting problem – about 20% of the coins are counterfeit and these tend to land tails more often than heads. You are visiting and given a coin by someone there.

i. Based on your reading you decide that your prior distribution for the probability of a head (call this θ) on your new coin is $\pi(\theta) = 0.8 * \text{Beta}(\theta|6, 2) + 0.2 * \text{Beta}(\theta|2, 6)$. Plot this prior distribution and explain why it makes sense given the information above.

ii. You flip the coin 10 times. You get 4 heads and 6 tails. Obtain the posterior distribution and plot it.

iii. What would you estimate as the probability of heads on the next toss?