PROBLEMS:

1. **Loss functions.** Problem 8.1. Background: In this class we emphasize the simulation-based approach to inference and use a sample from the posterior distribution to address questions of interest. The traditional Bayesian approach to point estimation is the decision-theory approach. To provide a point estimate of a parameter $\theta$, we introduce a loss function $L(\theta, a) \geq 0$ that gives the loss (or cost) for giving $a$ as an estimate when $\theta$ is the true value. The Bayes estimate of $\theta$ 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 $\theta$ is a scalar parameter with continuous posterior density for the problem.

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 $\lambda$) Poisson random variables. We explore the development of prior distributions for the parameter $\lambda$.

   (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 $\lambda$. To do this assume $\lambda$ has a gamma distribution (i.e., use $\lambda$ in place of $x$ in the density) and show that the posterior distribution of $\lambda$ given the data is a gamma distribution.

   (b) Demonstrate that the mean of the posterior distribution for $\lambda$ is a weighted average of the maximum likelihood estimate for $\lambda$ (which is the sample mean $\bar{y}$) and the prior distribution mean (see Appendix A).

   (c) Noninformative prior distributions

      i. One possible noninformative prior distribution is the flat prior distribution $p(\lambda) \propto 1$ for $\lambda > 0$. Does this lead to a proper posterior distribution?

      ii. Show that Jeffreys’ noninformative prior distribution for $\lambda$ is $p(\lambda) \propto 1/\lambda$ on the positive half of the real line.

      iii. Show that Jeffreys’ prior distribution is equivalent to a uniform prior distribution on $\log \lambda$.

       (Hint: Assume $p(\lambda) \propto 1/\lambda$ and derive the distribution of $\phi = \log \lambda$.)

      iv. Is the Jeffreys’ prior distribution a proper distribution (i.e., is it nonnegative and does it integrate to one if suitably normalized)?

      v. If we use $p(\lambda) \propto 1/\lambda$ as the prior distribution, is the posterior distribution of $\lambda$ a proper 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

\[ f(y|\theta) = 2\theta y e^{-\theta y^2} \]

for \( 0 < y < \infty \). The parameter \( \theta \) is something like the

"inverse lifetime" parameter (large \( \theta \) means short lifetimes; small \( \theta \) 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 \( (\text{given } \theta) \) samples from the Weibull(2) distribution.

(a) Show that the gamma distribution is the conjugate prior distribution and derive the posterior distribution (i.e., identify \( \alpha \) and \( \beta \) in the posterior distribution).

(b) Derive the marginal distribution \( p(y_1, \ldots, y_n) \). (Hint: There are two related ways to obtain the marginal. (1) Use \( p(y_1, \ldots, y_n) = \int p(y_1, \ldots, y_n|\theta)p(\theta)d\theta \); or (2) Use \( p(y_1, \ldots, y_n) = p(y_1, \ldots, y_n|\theta)p(\theta)/p(\theta|y_1, \ldots, y_n) \).) Explain why we would say \( y_1, \ldots, y_n \) 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 \( \theta \) 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 \( \theta \).

4. Weibull distribution with a nonconjugate prior distribution

We continue the data analysis above by changing the prior distribution. Assume that we use a prior distribution for \( \theta \) that is uniform between 0 and 3 (\( \theta = 3 \) corresponds to a mean lifetime around 50,000 and \( \theta = .03 \) corresponds to a mean lifetime around 500,000).

(a) Compute and plot a discrete approximation to the normalized posterior density, \( p(\theta|y) \). (See 2.10a and the R computing handout for hints.)

(b) Obtain a sample of 1000 draws from the discrete approximation to the posterior distribution. (The \textit{sample} command in R will come in handy – see the handout from class.) Briefly summarize the posterior distribution (posterior mean, posterior s.d., some percentiles).

(c) The prior distribution here has some advantages over the conjugate gamma prior distribution used in the previous problem. It doesn’t favor values of \( \theta \) near 0.7 and thus can be thought of as more objective. But it also has a disadvantage in that it assigns zero probability to some part of the parameter space. This makes the computation more difficult (we lose conjugacy) but there is a more important problem with such a prior distribution. Find the posterior distribution if \( y = (.16, .16, .18, .27, .28, .36, .41, .50, .54, 1.06) \). Explain what happens. (The true \( \theta \) in this case is 5 but the posterior distribution still stops at 3. Why?)

(d) What does the posterior distribution look like for the new data when we use the Gamma(1.4,2) prior distribution? Discuss.