CSE 515T (Fall 2019) Assignment 1

Due Monday, 23 September 2019

1. Your doctor supposes that you may have a rare disease, Bayesian syndrome (also known as BS), which occurs in every one out of 100 members of the population. The doctor proposes you take a super-accurate (and expensive!) test to be sure. The scanner is fairly reliable; 95% of all Bayesians are identified as Bayesians, and 95% of non-Bayesians are identified as such.

   The test comes back positive and the doctor tells you it's nearly certain you're a Bayesian given that result, due to the accuracy of the test. How do you respond?

2. Suppose k has a Poisson distribution with unknown rate parameter θ

$$\Pr(k \mid \theta) = \frac{\theta^k e^{-\theta}}{k!}, \qquad k = 0, 1, 2, \ldots$$

   Let the prior for θ be a gamma distribution:

$$p(\theta \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \qquad \theta > 0$$

   where Γ is the gamma function. Show that, given an observation k, the posterior $p(\theta \mid k, \alpha, \beta)$ is a gamma distribution with updated parameters (α', β').

   Hints:

   - For nonnegative integer k, $k! = \Gamma(k+1)$.
   - $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$.
   - For common distributions (such as the gamma distribution), Wikipedia has useful properties in an infobox.
   - Beware there are two common parameterizations of the gamma distribution.

3. Suppose that in the last question, we received a sample of n observations $\{k_1, k_2, \ldots, k_n\}$. What is the posterior $p(\theta \mid k_1, k_2, \ldots, k_n, \alpha, \beta)$? What is the posterior mean? The posterior mode?

   In light of this and the previous question, can you give an interpretation of the prior parameters α and β? What happens in the limit as $n \to \infty$?

4. (Scenario quoted from Morey, et al.) A 10-meter-long research submersible with several people on board has lost contact with its surface support vessel. The submersible has a rescue hatch exactly halfway along its length, to which the support vessel will drop a rescue line. Because the rescuers only get one rescue attempt, it is crucial that when the line is dropped to the craft in the deep water that the line be as close as possible to this hatch. The researchers on the support vessel do not know where the submersible is, but they do know that it forms distinctive bubbles. These bubbles could form anywhere along the craft's length, independently, with equal probability, and float to the surface where they can be seen by the support vessel.

   We wish to perform inference about the location of the rescue hatch given observed bubbles; call this location θ.

A common "trick" when wishing to express absolute prior ignorance of a parameter is to use a so-called *uninformative* prior. In this case, we will consider the uninformative "prior" $p(\theta) = 1$. This prior does not normalize, but we will see that it does not lead to major problems.

(a) Suppose the researchers observe the locations of exactly two bubbles, x_1 and x_2. Write down an appropriate likelihood for these data given θ and derive the posterior distribution for the location of the hatch, $p(\theta \mid x_1, x_2)$, using the uninformative prior described above.

(b) Now find a 50% Bayesian credible interval for θ given (x_1, x_2). Plot the width of this interval as a function of $|x_1 - x_2|$. Is this the relationship you would expect?

5. We are going to consider a Bayesian hypothesis test that a coin is exactly fair. Let θ be the "heads probability" of a coin
$$\Pr(H) = \theta.$$

Let us place a somewhat atypical prior on θ that includes nontrivial prior probability that the coin is exactly fair (that is, $\theta = 1/2$):

$$\Pr(\theta = 1/2) = 0.9 \qquad p(\theta \mid \theta \neq 1/2) = \mathcal{U}[0, 1].$$

This can be interpreted as a prior that is a mixture between a uniform prior on the interval $[0, 1]$ with weight 0.1 (this is also a special of the beta distribution with $\alpha = \beta = 1$) and a "point mass" (Dirac delta function) at $\theta = 1/2$ with weight 0.9.

Suppose we flip a coin $n = 10$ times and observe $x = 8$ "heads." What is the posterior distribution $p(\theta \mid x, n)$? What is the posterior probability that the coin is exactly fair?

Note there is no probability density function corresponding to the prior. It will help to work in cases.

6. (Effect of weird priors.) Let us consider the following set of observations. We flip a coin independently $n = 1\,000$ times and observe $x = 900$ successes. Call the unknown bias of the coin $\theta \in (0, 1)$.

For each of the prior distributions $p(\theta)$ below, please:

- plot the prior distribution $p(\theta)$ over the range $0 < \theta < 1$
- plot the posterior distribution given the above data, $p(\theta \mid \mathcal{D})$, over the range $0 < \theta < 1$
- report the posterior mean, $\mathbb{E}[\theta \mid \mathcal{D}] = \int \theta\, p(\theta \mid \mathcal{D})\, \mathrm{d}\theta$.

(a) A uniform prior on θ, which can be realized by selecting the beta distribution with $\alpha = \beta = 1$:
$$p(\theta) = \mathcal{B}(\theta; \alpha = 1, \beta = 1).$$

(b) A prior with extreme bias toward small values of θ:

$$p(\theta) = \mathcal{B}(\theta; \alpha = 1, \beta = 100).$$

(c) A prior that has no support on values greater than $\theta = 1/2$:

$$p(\theta) = \begin{cases} 2 & \theta < 1/2; \\ 0 & \theta \geq 1/2. \end{cases}$$

7. (Optimal *Price is Right* bidding.) Suppose you have a standard normal belief about an unknown parameter θ, $p(\theta) = \mathcal{N}(\theta; 0, 1^2)$. You are asked to give a point estimate $\hat{\theta}$ of θ, but are told that there is a heavy penalty for guessing too high. The loss function is

$$\ell(\hat{\theta}, \theta; c) = \begin{cases} (\theta - \hat{\theta})^2 & \hat{\theta} < \theta; \\ c & \hat{\theta} \geq \theta \end{cases},$$

where $c > 0$ is a constant cost for overestimating. What is the Bayesian estimator in this case? How does it change as a function of c? Plot the optimal action as a function of c for $0 < c < 10$. Hint: you may need to minimize certain expressions you encounter numerically as an analytic solution may not be available.