Statistics 568 Bayesian Analysis Spring 2021

Homework 5

Due: Wed 02/24/21 @ 11:59pm rutgers.instructure.com/courses/120689

Problem 1. Let $y = (y_1, \ldots, y_{100})$ be independent samples from a $N(\theta, 1)$ distribution with a truncated uniform prior distribution: $p(\theta) \propto 1$ for $\theta \in [-10^5, 10^5]$, and 0 otherwise. We wish to check the model using the test statistic that is the maximum absolute observed value:

$$T(y) = \max_i |y_i|.$$

Consider a dataset in which $\bar{y} = 5.1$ and $T(y) = 8.1$.

1. What is the posterior predictive distribution for y^{rep}? Make a histogram for the posterior predictive distribution of $T(y^{rep})$, and give the posterior predictive p-value for the observation $T(y) = 8.1$.

2. The **prior predictive distribution** of y^{rep} is defined as

$$p(y^{rep}) = \int p(y^{rep} \mid \theta) p(\theta) d\theta.$$

   (Note: compare to the definition of the posterior predictive distribution.) What is the prior predictive distribution for y^{rep} in this example? Roughly sketch the prior predictive distribution of $T(y^{rep})$ and give the approximate prior predictive p-value for the observation $T(y) = 8.1$.

3. Your answers from the previous questions should show that the data are consistent with the posterior predictive but not the prior predictive distribution. Does this make sense? Explain.

Problem 2. Perform posterior predictive check for the hierarchical model for the rat tumor data discussed in Chapter 5. Define three test quantities that might be of scientific interest, and compare them to their posterior predictive distributions. (One of the three should be a marginal posterior predictive check.)

Problem 3. In a study of "traumatic avoidance learning", Solomon and Wynne describe an experiment in which dogs learn to jump a barrier to avoid an intense electric shock. The subjects were 30 "mongrel dogs of medium size" weighing 9 to 13 kilograms. The apparatus was a variation of the Miller-Mowrer shuttle box used for avoidance training of rats. The box consisted of two compartments separated by a barrier and a "guillotine-type gate", which could be raised or lowered. The barrier was adjusted to the height of each dog's back. The floor of the apparatus consisted of steel bars which were wired to the shock circuit.

The condition stimulus (CS) consisted of turning out the lights above the compartment the dog was in and simultaneously raising the gate. The other compartment was still illuminated. In ten pretest trials none of the 30 dogs jumped the barrier during a **2-minute** exposure of the CS. During training the CS was presented for **10 seconds** and was then followed by an intense electric shock applied through the floor to the dogs' feet. The voltage was the "highest possible without producing tetany of the dogs' leg muscles". The current was about 100 to 125 milliamperes for most dogs. The shock was left on until the dog escaped over the barrier into the illuminated compartment, where no shock was administered. The gate was closed as soon as the dog jumped. If a dog **jumped before the shock** was turned on (i.e. during the 10-second period), the trial was recorded as an **avoidance** trial, whereas if the dog **did not jump until it was shocked**, the trial was recorded as an **escape** or shock trial. The experiment was designed so that shock could be escaped or avoided only by jumping the barrier.

The file `dogs.txt` contains the data for the first 25 trials on each of the 30 dogs in the experiment. Occurrence of shock is indicated by S, non-occurrence by a blank. (The number assignments to the dogs were made by Solomon and Wynne and do not imply that a selection has been made.)

1. Create a probabilistic learning model for these data. The model should reflect the increases in probability of avoidance with practice as jumping is reinforced by shock and avoidance. The model would give the probability of avoidance as it depends on trial number and previous experience. Also, to keep computation simple, give your model only two or three parameters.

2. Set up a weakly informative prior distribution.
3. Compute the posterior distribution and discuss the fit of the model to the data (weaknesses and strengths).
4. Simulate a set of fake data from your model (that is, create a table of simulated data from 30 new dogs). Display, on a single page, the data from the real and the fake dogs. Discuss the similarities and differences between the real data and the simulated data.

Hint to the model specification. Let T_{ij} be a latent variable indicating the waiting time for dog j in trial i to jump when presented the CS. The observed data of whether the dog received the shock can be seen as censored waiting times. Suppose $T_{ij} \sim Exp(\lambda_{ij})$, where $\log \lambda_{ij}$ is linear in trial number as well as the dog's previous experience in shock avoidance.