9.07 Introduction to Statistics for Brain and Cognitive Sciences
Emery N. Brown

Homework Assignment 7
November 16, 2016
Due November 28, 2016 at 5:00 PM

For each of the simulation problems, turn in both your MATLAB® code and the answers to the questions.

1. Simulation
Suppose that the background noise in magnetic resonance imaging obeys a Rayleigh probability model defined as

$$f(x \mid \tau) = \tau x \exp\{-\frac{\tau}{2} x^2\}$$

where $x > 0$ and $\tau > 0$. The cumulative distribution function is

$$F(x \mid \tau) = 1 - \exp\{-\frac{\tau}{2} x^2\}.$$

A. On the same graph, plot $f(x \mid \tau)$ for $\tau = 5, 1, \frac{1}{5}, \frac{1}{10}$ and $\frac{1}{16}$.

B. On the same graph, plot $F(x)$ for $\tau = 5, 1, \frac{1}{5}, \frac{1}{10}$ and $\frac{1}{16}$.

C. Using **Algorithm 11.1**, simulate 500 draws from $F(x \mid \tau)$ with $\tau = \frac{1}{5}$ and make a Q-Q plot to assess the agreement between simulated data and the theoretical model.

2. Method-of-Moments Analysis
The expected value of X in Problem 1 is

$$E(X) = (\frac{\pi}{2\tau})^{\frac{1}{2}}.$$

A. Assume that we have $X_1, ... X_n$ independent observations from a Rayleigh probability model. Find a method-of-moments estimate of τ.

B. Using the first 250 draws from $F(x)$ in 1C, compute a method-of-moments estimate of τ. How does it compare with the true value of $\tau = \frac{1}{5}$?

C. Write a parametric bootstrap algorithm to compute the uncertainty in $\hat{\tau}_{MM}$ using 500 bootstrap samples.

   1. Plot the histogram of the bootstrap samples.

2. Construct an approximate 95% confidence interval based on the bootstrap samples. Does the interval cover the true value of $\tau = \frac{1}{5}$?

3. Repeat 1 and 2 for a nonparametric bootstrap using the 250 samples in B to construct the empirical cumulative distribution function. Draw 500 bootstrap replicates.

4. Do the results of the parametric and nonparametric bootstraps differ? If so, how?

3. Likelihood Analysis

A. Find the maximum likelihood estimate of τ for the Rayleigh probability model.

B. Using the sample of 250 observations from 2B, compute $\hat{\tau}_{ML}$.

C. By computing the observed Fisher information, $I(\hat{\tau}_{ML})$, construct an approximate 95% confidence interval for τ. Does the interval cover the true value of $\tau = \frac{1}{5}$? How does the interval compare to the two confidence intervals computed in Problem 2?

4. Bayesian Theoretical Analysis

For the Rayleigh probability model assume that the parameter τ has a prior probability density which is a gamma density with parameters α and β. That is

$$f(\tau) = \frac{\beta^\alpha \tau^{\alpha-1}}{\Gamma(\alpha)} \exp\{-\beta\tau\}$$

for $\tau > 0$, $\alpha > 0$ and $\beta > 0$. Assume we are given a sample $X_1,...,X_n$ from $f(x|\tau)$ in Problem 1A.

A. Find the posterior density $f(\tau|x)$. What type of probability density is it? (Hint: Write $f(\tau|x) \propto f(\tau)L(\tau)$ where $L(\tau)$ is the likelihood function, and notice what class of probability models $f(\tau|x)$ belongs to. Then deduce its normalizing constant.)

B. Find $E(\tau|x)$.

C. Find $Var(\tau|x)$.

D. Find the mode of $f(\tau|x)$.

E. How do the analytic forms of $\hat{\tau}_{MM}$, $\hat{\tau}_{ML}$, $E(\tau|x)$ and the mode of $f(\tau|x)$ compare?

5. Bayesian Data Analysis

Let $\alpha = 2$ and $\beta = 0.1$ in $f(\tau)$ in Problem 4. Using the sample of 250 observations from $f(x|\tau)$ with $\tau = \dfrac{1}{5}$ in Problem 2B, compute

A. $E(\tau|x)$.

B. $Var(\tau|x)$.

C. The mode of $f(\tau|x)$.

D. Construct an approximate 95% Bayesian credibility interval for τ using a Gaussian approximation to $f(\tau|x)$.

E. Explain the difference in interpretation between the Bayesian credibility interval and the Frequentist confidence intervals computed above based on the bootstrap (Problem 2) and based on the observed Fisher information (Problem 3).

6. (Confidence Intervals and $z-$ test)
   The gross calorific value is a measure in megajoules per kilogram (MJ/kg) of the heat content of coal. Coal that meets a certain standard receives an International Organization of Standardization (ISO) certificate. If the assessment is carried out by ISO standards the measurements are known to be Gaussian with a standard deviation $\sigma = 0.1$. A sample of gross caloric values from a laboratory are in the file COAL.txt on the homework section of the course webpage.

   A. Test the null hypothesis that the true mean value for the lab is 23.73 against the alternative that it could be either greater than 23.73 or less than 23.73 at α What do you conclude?

   B. What is the $p-$ value?

   C. Construct a 95% confidence interval for the true gross calorific value based on this sample. Do your results agree with the results in A? Why or why now?

   D. Which is more informative, A and B, or C? Why?

9.07 Statistics for Brain and Cognitive Science
Fall 2016