Stat 624: Midterm

Inside your personal git repository, submit all files to the "midterm" directory. To ensure you will get full credit, ensure the problems are all clearly labelled, even parts a, b, c, etc. You have access to all class files, such as the ones in your repository and links from the course website.

Problem 1

Consider data coming from a geometric distribution,

$$f(x|p) = (1-p)^x p. \tag{1}$$

The parameter p is restricted to between 0 and 1. However, there is an alternative parametrization using $\beta = \log\left(\frac{p}{1-p}\right)$. In this case the density function is

$$f(x|\beta) = \frac{e^{\beta}}{(1+e^{\beta})^{x+1}} \tag{2}$$

Complete the following problems in a script that is called Problem1.*ext*.

(a) Write a function or functions that will input a data set of size n and returns the negative log likelihood, first derivative, and second derivative for the data in terms of β.

(b) Write a function that inputs a data set and a starting value for β and performs Newton-Raphson to find the maximum likelihood estimate of β. Write your own algorithm, do not use built-in R functions.

(c) Use the data set

$$1 \;\; 2 \;\; 2 \;\; 2 \;\; 3 \;\; 4 \;\; 4 \;\; 5 \;\; 5 \;\; 5 \;\; 5 \;\; 7 \;\; 8 \;\; 10$$

to find the maximum likelihood estimate for β. Assess the uncertainty of the Maximum Likelihood estimate using a bootstrap confidence interval.

(d) Plot the log-likelihood function as a function of β using the data in part (c). Determine a good range that shows the optimum. Mark the optimal value in the figure.

(e) Conduct a hypothesis test on the maximum likelihood estimate of β for the above data. Use the null hypothesis $H_0 : \beta = -1$ and the alternative hypothesis is $H_a : \beta < -1$. Determine the p-value when using a size of $\alpha = .05$. Assess uncertainty for the p-value.

(f) Suppose that the alternative hypothesis is $H_a : \beta = -1.5$. Determine the power of the hypothesis test in part (g) and assess uncertainty.

Problem 2

Consider the integral $I = \int_0^c e^{ax} \sin(x + b)dx$. Put the answers to the following responses in a script called Problem2.*ext*. Clearly label all parts.

(a) Write a function that will evaluate the integral using numerical integration with the trapezoidal rule for any inputs of a,b, and c.

(b) Write a function that will evaluate the integral using Monte Carlo integration and assess Monte Carlo uncertainty for any value of a, b and c.

(c) Write a function that will evaluate the integral using the hit-and-miss method and assess Monte Carlo uncertainty for any value of a, b, and c.

(d) Suppose that a computer algorithm approximates the integral several times using the same unknown values for a and b each time while varying the value for c. The outputs are also subject to some rounding and uncertainty. The integral I is observed for the following values of c.

c	0.5	1	1.5	2
I	0.5	2	5	10

Use non-linear least squares to determine the best fit values are for a and b that support this output. . Built in optimization and integration functions are okay to use.