Stat 624: Midterm

Inside your personal git repository, create a folder entitled "midterm" and put all files there. To ensure you will get full credit, ensure the problems are all clearly labelled, even parts a, b, c, etc.

Problem 1

Put all solutions for this problem in a file labeled Problem1.R inside the midterm folder. Let X be a random variable with density function $f(x) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$ for $x > 0$ where $\sigma > 0$.

a Create a pdf figure titled density.pdf that shows the effect of σ on the shape of the density function of X. Save the figure and put it in the midterm folder. Include code used to create the plot inside the Problem1.R code.

b This density is actually a half-normal density function. If Z is a normal distribution with mean 0 and variance σ^2, then X arises from the transformation $X = |Z|$. Create a function called "rhalfnorm" that takes n and σ and returns n random draws from the half-normal distribution with parameter σ. Built-in distributions in R can be used. (hint: don't overthink)

c The data file growth.csv shows the amount of growth in acreage of an invasive floral species between seasons. Assume this data comes from a half-normal distribution. Use Newton-Rhapson to optimize the likelihood function of the data to find the maximum likelihood estimate of σ. Do not use built in optimization functions in R.

d Using a prior of Exp(1), draw samples from the posterior distribution of σ and determine the Bayesian estimate of σ under squared error loss. Assess convergence by returning an acceptance rate between 20 and 50 % and checking diagnostic plots. Generate at least 10,000 samples and discard a burn-in of at least 1,000.

e Compare estimates and uncertainty intervals of the MLE and Bayesian estimates. The uncertainty interval for the MLE estimate must be obtained using bootstrapping.

Problem 2

a The median of the gamma distribution has no closed form solution. Write a function called "median" that takes the two parameters of a gamma distribution as inputs and uses numerical integration and root-finding to return the median of the gamma distribution up to some tolerance level. Do not use built in R functions to do this. Apply this function to two different gamma distributions:

- $\alpha = 5$ and $\theta = 2$ (i.e. mean of 10 and variance of 20)
- $\alpha = 3$ and $\theta = 3$ (i.e. mean of 9 and variance of 27)

b A certain data set with 100 observations has a median of 8. Using a simulation study, conduct a hypothesis test that this comes from a gamma distribution with $\alpha = 5$ and $\theta = 2$ (i.e. has a mean of 10 and a variance of 20). Do this by creating a sampling distribution of medians under the null hypothesis, finding a p-value based on the observed data, and interpreting the results using a size of $\alpha = 0.05$. Assume a one sided test where the rejection region is the lower 5%. Assess uncertainty of Monte Carlo estimates.

c Assume that the alternative hypothesis is that $\alpha = 3$ and $\theta = 3$. Determine the power of the test in part (b). Again Assess uncertainty of Monte Carlo estimates.