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.

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 $f(x) = \frac{\alpha\lambda^\alpha}{(x+\lambda)^{\alpha+1}}$ for $x > 0$ where $\alpha > 0$ and $\lambda > 0$. This is called a Lomax distribution.

a 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 Lomax distribution. Use Newton-Rhapson to optimize the likelihood function of the data to find the maximum likelihood estimate of α and λ. Do not use built in optimization functions in R.

b Use bootstrapping to report a 95% confidence interval for both parameters. For this problem, 1000 bootstrapped samples is sufficient.

c Using a prior of Exp(3) for both parameters, draw samples from the posterior distribution of α and λ given the data in growth.csv using Metropolis-Hastings and determine the Bayesian estimate of α and λ 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. Include the code you used to check the diagnostic plots.

d Report the credible interval for each of the parameters from the Metropolis-Hastings output in part b.

Problem 2

Put all solutions for this problem in a file labeled Problem2.R inside the midterm folder. Let X be an exponential random variable with a mean of θ, Y be a Uniform random variable between 0 and 10, and let Z be equal to the sum, $Z = X + Y$.

a We know that a sum of random variables can be written as a convolution. Specifically for our case,

$$f_Z(z) = \int_0^{min(10,z)} f_X(z-\tau)f_Y(\tau)d\tau$$

Note that the bounds of the integral are from 0 to $min(10, z)$. Write a function takes two arguments, a value for z and a value for θ, which again is the *mean* of the exponentially distributed variable X. Name the function dZ and have it return the density value of $f_Z(z)$ using numerical integration to evaluate the convolution above. Do not use built in R functions for the integration.

b For this problem, assume that $\theta = 3$. Simulate values for Z using built-in R functions. In a figure saved as dZ.pdf, compare the empirical density function of the draws with values from the theoretical density function developed in part a.

c Suppose that you know that $f_Z(10) = 0.09$. Create an appropriate target function and use the bisection method to determine the value of θ that results in that density value.

d Suppose you have data that is distributed as Z, the sum of an exponential with mean θ and a Uniform on 0 to 10. The mean of 8 data points is equal to 11.7. Using a simulation study, test the hypothesis that the true value of θ is 3.5. Report a p-value for the test along with the Monte Carlo error. (hint: the integral in part a is not necessary)

e Suppose that the alternative hypothesis to the test in part d is that $\theta = 10$. Using a simulation study, determine the power of the hypothesis test in part d and report the Monte Carlo error.