Exam 2

550.433

You are not permitted to use your notes or any textbook while taking this exam. You are permitted to use the internet, but **only** to look up commands in `R`, `Matlab`, or any other language that might help you in your coding.

For all of the problems that require coding, you **must** supply your code, a histogram or scatterplot of random numbers (which ever is appropriate), and a written explanation of your code. This all must be on one attachment as well. Do not send me multiple files.

The exam is due before 11:59 pm of May 10, 2020. Take as much time as you like.

1. (15 pts.) Student's T_ν density with ν degrees of freedom is given by

$$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}.$$

   Calculate the mean of a t distribution with 4 degrees of freedom using a Metropolis-Hastings algorithm with candidate density

   (a) N(0,1)

   (b) t with 2 degrees of freedom.

   (c) Which of the candidate densities works more efficiently? Defend your answer in as much detail as possible.

2. (15 pts.) Assume X has the density $f_X(x)$ such that

$$f_X(x) = \frac{4}{\beta^3\sqrt{\pi}} x^2 \exp\left\{-\, x^2 / \beta^2\right\} \quad 0 < x < \infty,$$

   and $\beta = 2$. Use Monte Carlo methods to estimate the values of $E(X^2)$ and $P(1 \leq X \leq 4)$ as efficiently as possible. Use importance sampling.

3. (15 pts.) Suppose the random variables X and Y both take on values in the interval $(0, B)$. Suppose that the joint density of X given $Y = y$ is

$$f(x|y) = C(y)\exp(-xy) \qquad 0 < x < B,$$

   and the joint density of Y given $X = x$ is

$$f(y|x) = C(x)\exp(-xy) \qquad 0 < y < B,$$

   where $C(y)$ is some constant that depends on y and $C(x)$ is some constant that depends on X. Give a method (write it down.....no need to code it up) on how you would simulate values of X and Y. **BE AS SPECIFIC AS POSSIBLE**. If you plan to simulate from a distribution, tell me what exactly the distribution is and exactly how you will simulate from it.

4. (15 pts.) This problem has to do with Rao-Blackwellization.

   (a) In as much detail as you can, explain the concept of Rao-Blackwellization and why it's helpful.

   (b) X is said to follow a log-normal distribution with parameters μ and σ^2 if $\log(X) \sim N(\mu, \sigma^2)$. Consider the model given by $X \sim \text{lognormal}(0, 1)$ and $\log(Y) = 9 + 3 \times \log(X) + \epsilon$, where $\epsilon \sim N(0, 1)$. We wish to estimate $E\left(\frac{Y}{X}\right)$. Compare the performance of the standard Monte Carlo estimator and the Rao-Blackwellized estimator of this expectation. It helps to know that if $\log(W) \sim N(\mu, \sigma^2)$, then $E(W) = \exp\left(\mu + \sigma^2/2\right)$.

5. (15 pts.) My dad's name is James Botts. He's a medical internist in my hometown, and he's a reasonably smart guy, but he doesn't know as much statistics as we do. In fact, to brush up on his statistical knowledge, he recently took an online introductory statistics course which covers the basics of hypothesis testing, confidence intervals, and random variables. Write my dad a few paragraphs explaining to him what the Bootstrap and Jackknife methods are. Tell him why these methods are sometimes needed, and give some details about the methods themselves. And when discussing the Jackknife, explain why (in some cases) it might be preferred to the Bootstrap.

6. (15 pts.) One of the homework problems I had you do was illustrate that the bootstrap isn't a very good way of estimating θ, for the $\text{Unif}(0, \theta)$ distribution. Show the same thing for the mean of a Cauchy distribution. In other words, I would like you to perform a simulation study which illustrates that the bootstrap estimate of the mean of a Cauchy distribution is not ideal (and can fail). The command to simulate Cauchy random variables in `R` is `rcauchy`.

7. (10 pts.) Consider the stochastic differential equation

$$dX(t) = f[X(t)]dt + g[X(t)]dW(t).$$

   Tell me what the Euler-Maruyama solution to such an equation is, and provide a one-paragraph discussion as to why the Euler-Maruyama approximate solution to SDEs "makes sense." That is, comment as to why it is a natural discrete representation of a stochastic differential equation. Also discuss (in detail) the two types of convergence for solutions to SDEs.