AMS 597 Spring 2021
Homework 5 Due 04/08/2021 10:00 AM

Instruction: Submit your homework via Blackboard. If you have difficulty submitting via Blackboard, email the TA directly. No late homework will be accepted, based on the time the email is sent out. If you are submitting via email (not recommended), put this as the subject of your email

AMS597 2021: Homework/Exam #5. ID: XXXXXXXXX. Name: XXX XXX

1. Using only random uniform generator, generate a random sample of size 1000 from the F distribution with 5 and 10 degrees of freedom. (Hint: If $X \sim \chi^2_a$ and $Y \sim \chi^2_b$, X and Y are independent, then $F = \frac{X/a}{Y/b}$ follows F distribution with a and b degrees of freedom.)

2. One disadvantage of Box-Muller algorithm is that it requires computing sine and cosine functions. By some transformation, one can show that the following closely related algorithm also generates independent standard normal random variables X and Y.

   (a) Generate $U_1 \sim U(0,1)$ and $U_2 \sim U(0,1)$

   (b) Set $V_1 = 2U_1 - 1$ and $V_2 = 2U_2 - 1$, $S = V_1^2 + V_2^2$.

   (c) If $S > 1$, return to (a).

   (d) Otherwise,

   $$X = \sqrt{\frac{-2\log S}{S}} V_1$$

   $$Y = \sqrt{\frac{-2\log S}{S}} V_2$$

   This method is called the polar method. Use this method to generate 10000 standard normal random variables.

3. A t-distributed random variable of k degrees of freedom is defined as

   $$t_k = \frac{Z}{\sqrt{W/k}}$$

   where $Z \sim N(0,1)$ and $W \sim \chi^2_k$, Z and W are independent. Using only the uniform number generator in `R` (`runif`), generate 100 random samples X, where X is a mixture t distribution, i.e., $X \sim 0.3t_3 + 0.35t_5 + 0.35t_7$.

4. Write your function `rmultivarNorm(n,mu,Sigma)` which will generate n multivariate normal random variables using only `runif`.

5. Write your own function `my.ls()` which performs the least square estimation for a continuous response variable y regressed on two predictors x_1 which is a numeric predictor and x_2 which is a categorical predictor. You may assume that your model contains an intercept. Test your function on the `ChickWeight` dataset in `R`, where y is `weight`, x_1 is

Time (assumed to be continuous) and x_2 is Diet, i.e., check if your function can reproduce the estimated beta coefficients from

```
data(ChickWeight)
fit <- lm(ChickWeight$weight~ChickWeight$Time+ChickWeight$Diet)
fit$coef
```

6. Use the mcycle data in the MASS package for this problem (see Homework 3 Qn 5). Instead of fitting polynomial regression, fit an additive model using splines to this dataset. Compare the MSE of the spline model to the polynomial regression model, where MSE $= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$, Y_i is the observed response variable and $\hat{Y}_i$ is the predicted response variable.