AMS 597 Spring 2021
Homework 4 Due 03/30/2021 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 #4. ID: XXXXXXXXX. Name: XXX XXX

1. The leukemia gene expression dataset available in the url below consists of 72 subjects/patients and 3571 genes.
   `http://www.ams.sunysb.edu/~pfkuan/Teaching/AMS597/Data/leukemiaDataSet.txt`

   Each patient is of either type ALL (Acute lymphocytic leukemia) or type AML (Acute myelogenous leukemia). Using the genes as covariates, we will construct a model that can predict the two types of leukemia as follows:

   (a) First split the data randomly into two subsets containing 50 (`trainData`) and 22 (`testData`) subjects, respectively as follows:

   ```
   dat <- read.delim('http://www.ams.sunysb.edu
   /~pfkuan/Teaching/AMS597/Data/leukemiaDataSet.txt'
   ,header=T,sep='\t') ### please read this as a single line in R, I break this into
   ### 3 lines to avoid overflowing outside paper margin

   str(dat)

   set.seed(123)
   trainID <- sample(1:72,round(0.7*72))

   trainData <- dat[trainID,]
   testData <- dat[-trainID,]
   ```

   (b) Build your best model on `trainData` using the `Group` variable as response and the genes as predictors/covariates.

   (c) Evaluate your model from (2) on the `testData` by computing the percentage of AML correctly predicted, the percentage of ALL correctly predicted and the overall percentage of AML and ALL correctly predicted.

2. Write a function that will generate and return a random sample of size n from the two-parameter exponential distribution $Exp(\lambda, \nu)$ for arbitrary n, λ, and ν using inverse transform method. Note that the pdf of $X \sim Exp(\lambda, \nu)$ is

$$f(x) = \lambda e^{-\lambda(x-\nu)}, x \geq \nu$$

   and $\lambda > 0, \nu > 0$. Generate a random sample of size 1000 from the $Exp(2, 1)$ distribution.

3. Write a function to generate a random sample of size n from the standard Cauchy distribution with pdf $f(x) = \frac{1}{\pi(1+x^2)}, -\infty < x < \infty$ using inverse transform method. Generate a random sample of size 1000 from the Cauchy distribution.

4. Write a function to generate a random sample of size n from the $Beta(a, b)$ distribution by the acceptance-rejection method. Generate a random sample of size 1000 from the $Beta(3, 2)$ distribution.

5. Write a function to generate a random sample of size n from the $Gamma(\alpha, 1)$ distribution by the acceptance-rejection method. Generate a random sample of size 1000 from the $Gamma(3, 1)$ distribution. (Hint: you may use $g(x) \sim Exp(\lambda = 1/\alpha)$ as your proposal distribution, where λ is the rate parameter. Figure out the appropriate constant c).