Applied Statistics and Probability – Math 377
Fall 2020 Midterm 2

Answer all the following questions.

1. (10 points) Consider the pdf

$$f(x) = 6xe^{-2x^3}\mathbf{1}(x > 0).$$

(a) How would you transform a $uniform(0, 1)$ random variable U into a random variable X that would have the pdf f?

(b) Consider the pdf

$$g(x) = cx(2 + \sin(x))e^{-2x^3}\mathbf{1}(x > 0),$$

where c is an appropriate constant. Suppose Gen_from_f is a R function that generates one random variable X following the pdf f. Use this to write a pseudocode to generate one random variable Y following the pdf g.
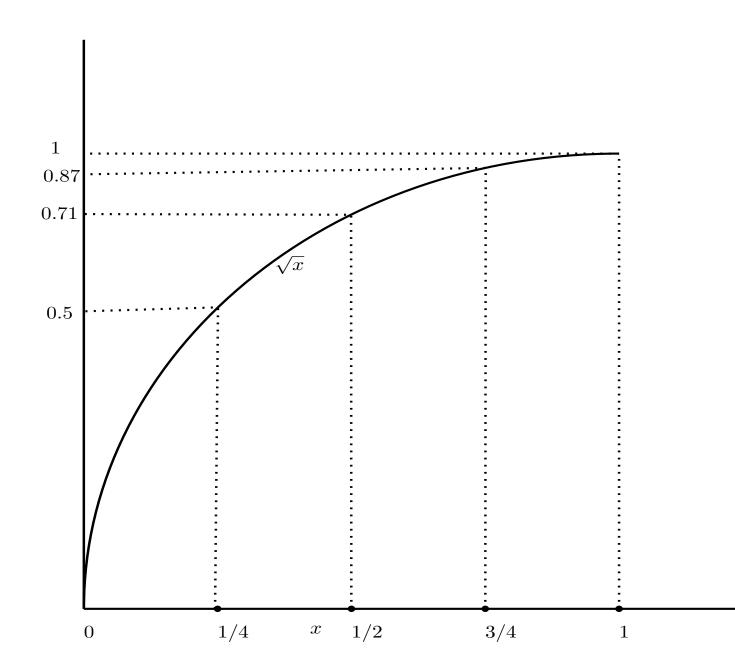
2. (6 points) For the data given in Midterm2Data1.csv, plot the following two figures as subplots in a single frame: (a) index plot without outliers (if any), and (b) the QQ plot for checking normality. Does the QQ plot suggest Normality?

3. (5 points) Suppose that the data given in Midterm2Data2.csv is a random sample from a distribution F with mean μ and variance σ^2. Estimate the p-value for the following hypotheses.

$$H_0 : \sigma^2 \leq 1 \text{ vs } H_1 : \sigma^2 > 1.$$

Use the Bootstrap resampling method with 10000 resamples.

4. (12 points) Consider the dataset given in Midterm2Data3.csv. Suppose the two columns are random samples from two distributions F and G having means μ_1, μ_2 and variances σ_1^2, σ_2^2 respectively. Use 1% as the level of significance to test whether the distributions F and G are Normal or not, and whether $\sigma_1^2 = \sigma_2^2$ or not. Use the outcomes of these analysis to decide what test should be used for testing $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$. Should H_0 be rejected at 1% level of significance?

5. (7 points) A sample of 700 cars were tracked for 1 year. Among these cars, 300 had cellular phones, 48 met at least one accident within that year, and 374 cars neither had cell phones nor met any accident. Use these data to test the hypothesis that having a cellular phone in your car and being involved in an accident are independent. Use the 5% level of significance. Do not use any approximate test.

6. (5 points) A random sample of size 100 were obtained from a distribution F having support $[0, 1]$. Among these samples, 44 are between 0 and $1/4$, 24 are between $1/4$ and $1/2$, 20 are between $1/2$ and $3/4$, and 12 are between $3/4$ and 1. Test the following hypotheses and obtain the p-value. The following figure is useful.

$$H_0 : F(x) = \sqrt{x}\mathbf{1}(0 < x < 1) \text{ vs } H_1 : H_0 \text{is false.}$$