Programming (part B)

1. Try to download the data from ('link').
   a. Estimate the CDF *F(x)*.
   b. Compute and plot a 95 percent confidence envelope for *F*.
   c. Find approximate 95 percent confidence interval for *F(4.9)-F(4.3)*.

2. Generate N samples from an Exponential distribution.
   a. Estimate its parameter.
   b. Explain about bias and unbiased status of that.
   c. What are its 97% and 93% confidence interval?
   d. Repeat each process for 1000 and 10000 times. Explain how often the confidence intervals contains parameters?

3. Generate 100 samples from an Exponential distribution. Again estimate its parameter[1].
   a. Calculate and plot empirical distribution function.
   b. Find the plug-in estimator for mean, variance, skewness (for variance, use both reasonable estimators as defined in course textbook).

4. Let $X_1, \dots, X_n \sim Normal(\mu, 1)$. *Let* $\theta = e^{\mu}$ *and let* $\hat{\theta} = e^{\bar{X}}$ *be the MLE.*
   a. Generate a dataset consisting of 100 observation with the mean equal to 5.
   b. Use the bootstrap to get the standard error and 94 percent confidence interval for θ.
   c. Plot a histogram of the bootstrap replications for the nonparametric bootstrap (estimates of the distribution of $\hat{\theta}$). Compare them to the true sampling distribution of $\hat{\theta}$.

5. Let $X_1, \dots, X_n \sim Normal(\mu, 1)$.
   a. Generate a dataset consisting of 100 observations with mean equal to 5.
   b. Take $f(\mu) = 1$ and find the posterior density. Plot the density
   c. Simulate 500 draws from posterior. Plot a histogram of the simulated values and compare the histogram to the answer in part b.
   d. Let $\theta = e^{\mu}$. Find posterior density for θ analytically and by simulation.
   e. What are its 97% and 93% confidence interval for θ?

6. Let $X_1, \dots, X_n \sim Poisson(\lambda)$ and perform the Wald test with assumption of $\lambda_0 = 1, n = 50$ *and* $\alpha = 0.05$. Repeat many times and count how often you reject the null. How close is the type I error rate to 0.05.

7. In 1861, 10 essays appeared in the New Orleans Daily Crescent. They were signed "Quintus Curtuis Snodgrass" and some people suspected they were actually written by Mark Twain. To investigate this, we will consider the proportion of three letter words found in an author`s work.

From 8 Twain essays we have ::	225 .262 .217 .240 .230 .229 .235 .217
From 10 Snodgrass essays we have::	.209 .205 .196 .210 .202 .207 .224 .223 .220 .201

   a. Perform a Wald test for equality of the means. Use the nonparametric plug-in estimator. Report the p-value and a 97% confidence interval for the difference of means. What do you conclude?

[1] To get some hints, you can follow example 8.6 and its precedence examples.

b. Now use a permutation test to avoid the use of large sample methods. What is your conclusion?

8. Generate a dataset as defined at the end of this question. We want to implement a Bayesian classifier where the class-conditional probabilities are estimated using histogram. Use 50 sample from each class for train and what remains for test (150 sample for train and 150 sample for the test set). Report the confusion matrices for both training and testing subsets.

Distribution	#	Mean	cov^2
Gaussian	100	[0,0]	[1,0.5;0.5,1]
Gaussian	100	[2,2]	[1,0.4;0.4,1]
Gaussian	100	[4,1]	[1,0.3;0.3,1]

9. Reproduce the Example 9.7

[۲] As you know, Semicolon means newline in the bracket (like Matlab)