Homework 1 (Due on 03/21/2022)

1. Draw 100 i.i.d. observations from the standard normal distribution. (i) Plot the CDF function; (ii) Plot the empirical CDF function; (iii) Plot the following estimate as a function of x

$$\tilde{F}(x) = \frac{1}{n}\sum_{i=1}^{n}\Phi(\frac{x - X_i}{h}). \tag{1}$$

   Use different h for the above estimator, e.g. $h = n^{-1/5}, n^{-1/4}, n^{-1/3}, n^{-1/2}$...

2. Use WAGE1.DTA data to estimate the CDF of wage by (i) empirical distribution function with randomly picking 10 observations from the sample; (ii) repeating (i), but randomly picking 100 observations; (iii) repeating (i), but using the full sample; (iv) using the estimator defined in eq. (1). Plot estimates in the same picture.

3. Use WAGE1.DTA data to estimate the density of wage by choosing different $K(\cdot)$ and bandwidth h:

   (i) the Uniform kernel $K(u) = 1(0.5 <= u \leq 0.5)$ and $h = \lambda \times \hat{\sigma} \times n^{-1/5}$ where $\lambda = 0.5, 1, 1.5, 2, 2.5$, respectively, and $\hat{\sigma}$ is the sample standard deviation of wage;

   (ii) the Gaussian kernel $K(u) = \phi(u)$ and $h = \lambda \times \hat{\sigma} \times n^{-1/5}$ where $\lambda = 0.5, 1, 1.5, 2, 2.5$, respectively;

   (iii) the Epanechnikov kernel kernel $K(u) = \frac{3}{4}(1 - u^2)\mathbb{1}(|u| \leq 1)$ and $h = \lambda \times \hat{\sigma} \times n^{-1/5}$ where $\lambda = 0.5, 1, 1.5, 2, 2.5$, respectively

   (iv) the Gaussian kernel $K(u) = \phi(u)$ and $h = 1.06 \times \hat{\sigma} \times n^{-r}$ where $r = 1/2, 1/3, 1/4, 1/5, 1/6, 1/7$, respectively

   (v) the Gaussian kernel $K(u) = \phi(u)$ and h is chosen by the cross validation method. (hint: you can either split the sample or use the leave-one out approach, please describe your CV approach in details)