HW#4 for Nonparametric Part, Fall 2022

Due:

I. Do Monte Carlo simulations to compare the performances of the kernel density estimations under different settings and to make your own conclusions based on your simulations. Please do the followings:

Use the Rosenblatt-Parzen method for density function by choosing different **sample sizes** (It is up to you to make a decision to choose the values for the sample size; say 250, 1000 and 5000), **different kernels** (say the normal $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$ and Epanechnikov kernel $K(u) = 0.75(1 - u^2)_+$), **different bandwidths** ($h = d\, n^{-1/5}$ for different values of $d > 0$, and **different bandwidth selection methods** (you learned from the class, for example, cross-validation and plug-in as well as normal reference). Please make conclusions and comments based on your simulation results.

♡ To assess the performance of finite samples, for each setting, you need to compute the mean absolute deviation errors (MADE) for $\widehat{f}(\cdot)$, defined as

$$\text{MADE} = m^{-1} \sum_{k=1}^{m} \left| \widehat{f}(u_k) - f(u_k) \right|,$$

where $\widehat{f}(\cdot)$ is the nonparametric estimate of $f(\cdot)$ (**density() in R**) and $\{u_k\}_{k=1}^{m}$ are the grid points, taken to be arbitrary within the range of data. Note that you can choose any distribution to generate your samples for your simulation. Also, note that the choice of the grid points is not important so that they can be chosen arbitrarily. In general, the number of replications can be taken to be at least $n_{sim} = 1000$. The question is how to report the simulation results. There are two ways of doing so. You can display the n_{sim} values of MADE either in a boxplot form (**boxplot() in R**) or in a table by presenting the *median* and *standard deviation* of the n_{sim} values of MADE. Either one is okay but the boxplot is preferred.

II. Conduct Monte Carlo simulations as you did in **Exercise I**, to compare the performances for the kernel density approach and a series estimation for time series data. Here, you can choose a series method by yourself. Make comments on your results.

III. Please do a simulation to explore what you can obtain for the density kernel estimation $\widehat{f}(x)$ when X_t is generated from $X_t = \rho\, X_{t-1} + e_t$ for different values of $|\rho| < 1$ and $\rho = 1$. Please pay your attention to the cases when ρ is close to 1 and $\rho = 1$. Note that the density of X_t changes over time when X_t is nonstationary. Summarize your results and make comments on your results. For $\rho < 1$ but close to one, you do the same thing as you did in **Exercise II**. But for $\rho = 1$, $f(x)$ does not exist, so that you can not compute MADE. Therefore, you need to do what I said in class.

IV. Consider three real data sets for the U.S. Treasury bill (Secondary Market Rate): the **daily** 3-month Treasury bill from January 4, 1954 to present, the **weekly** 3-month Treasury bill from January 8, 1954 to present, and the **monthly** 3-month Treasury bill from January 1, 1934 to present. Please go to the data web site [see below] to download them up to today. In other words, please use the data as long as you can. Please describe the data used in your project. Of course, you can use your own dataset if you wish to do so.

1. Apply Ljung-Box test [**Box.test() in R**] to see if three series are autocorrelated or not and if they are stationary or not. Also, you might look at the autocorrelation function (ACF) [**acf() in R**]or/and partial autocorrelation function (PACF)[**pacf() in R**].
2. Apply the kernel density estimation to estimate three density functions.
3. Any conclusions and comments on three density functions?

♡ Note that the real data sets can be downloaded from the web site for Federal Reserve Bank of Saint Louis at *http://research.stlouisfed.org/fred2/categories/116*. You can use any statistical package to do your simulation. You try to use **R** since it is very simple. You need to hand in all necessary materials (tables or graphs) to support your conclusions. If you need any help, please come to see me.