Columbia University
Statistical Analysis and Time Series

IEOR-4709

A. Capponi	**Problem Set #3**	Issued:	February 19, 2025
Spring 2025		Due: **BEFORE CLASS**	March 5, 2025

Note: Please put the number of hours that you spent on this homework set on top of the first page of your homework. The CA in charge of grading this homework is Boxuan Li. The TA in charge of this homework is Kishore Kuppusamy.

Ex. 1.
Let $X_1, X_2, \ldots, X_n$ be i.i.d. according to a Pareto distribution with density $f_\theta(x) = \theta c^\theta x^{-(\theta+1)}$, where $\theta > 0$ and $0 < c < x$. We want to test the hypothesis $H_0 : \theta = \theta_0$ vs the alternative hypothesis $H_1 : \theta \neq \theta_0$. Following the example shown in the class for testing whether the mean of a Gaussian distribution with unknown variance is θ_0, do the following:

- Write down the critical region of the likelihood ratio test.
- Find the threshold k in the definition of the critical region, so that the significance level of the test is α.

Ex. 2.
Go on Yahoo Finance. Download the time series of historical data of the S&P 500 from Feb 16, 2024 to Feb 16, 2025. Compute the returns of the S&P 500. Do a Q-Q plot of the empirical quantiles versus theoretical quantiles, assuming that the theoretical distribution of the return follows a Gaussian distribution with some mean μ and variance σ^2. Based on the plot, is there enough evidence to conclude that the returns are Gaussian distributed? Explain.

Ex. 3.
Consider estimating the distribution function $P(X \leq x)$ at a fixed point x based on a sample $X_1, \ldots, X_n$, where each sample X_i is independently drawn from the distribution of X. An estimator is $\frac{1}{n}\sum_i \mathbf{1}_{X_i \leq x}$. If it is known that the true underlying distribution is Gaussian with mean θ and variance 1, another possible estimator is $\Phi(x - \bar{X})$. Calculate the relative efficiency of these estimators, i.e., the ratio of their asymptotic variances. **You may find it useful to use the delta method to find the asymptotic distribution of $\Phi(x - \bar{X})$.**

Ex. 4.
A popular model in the theory of market microstructure postulates that the total number of trades, ignoring the case of no trade, within a minute for a particular stock follows a geometric distribution $p(x) = P(X = x) = p^{x-1}(1 - p)$, $x = 1, 2, \ldots$ The following table contains data of all the trade frequencies within 134 minutes with at least one trade. Do the following

- Find the MLE for p.
- Test whether the geometric distribution fits the data.

N. Trades	Frequency
1	46
2	32
3	21
4	11
5	5
6	7
7	4
8	3
9	1
10	2
11	1
12	1