Columbia University
Statistical Analysis and Time Series

**IEOR-4709**

A. Capponi — **Problem Set #2** — Issued: February 5, 2025
Spring 2025 — Due: **BEFORE CLASS** February 19, 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 Yisheng Jiang. The TA in charge of this homework is Jose Antonio Sidaoui.**

**Ex. 1**.
Suppose $\{X_n\}_{n\geq 1}$ is a sequence of independent and identically distributed random variables from a distribution with density $f_X(x) = \frac{1}{\beta}e^{-\frac{x-\theta}{\beta}}$, for $\theta \leq x < \infty$. Here, $\theta, \beta$ are unknown parameters. Do the following, making all your arguments complete and precise.

- Find the maximum likelihood (MLE) estimator for $\theta$ and $\beta$.
- Using the method of moments (MM), calculate an estimator for $\theta$ and $\beta$.
- Using Python or R (or your favourite programming language), randomly draw 5 values from the distribution with parameter $\theta = 3$ and $\beta = 2$. Calculate the MLE for both $\theta$ and $\beta$. Repeat this 1000 times. Calculate the mean and variance of these estimates, as well as the approximate mean squared error (i.e. the mean squared difference between the estimate and the true value).

**Ex. 2**.
Suppose $\{X_n\}_{n=1}^{N}$ is a sequence of independent and identically distributed random variables drawn from the Weibull distribution with probability density function $f_X(x) = \frac{\alpha x^{\alpha-1}}{\theta^\alpha}e^{-(x/\theta)^\alpha}$, for $x \geq 0$. Do the following, making all your arguments complete and precise.

(i) Plot the probability density function for three choices of $\alpha, \theta$, $\alpha = \theta = 2$, $\alpha = 4$ and $\theta = 2$, $\alpha = 2$ and $\theta = 5$.

(ii) Find the maximum likelihood estimator for $\theta$ assuming $\alpha$ is known.

(iii) Assume $N = 50$, $\alpha = 1$ and $\theta = 2$. Simulate $N$ independent samples from the Weibull distribution with these parameters $\alpha$ and $\theta$.

(iv) Find the maximum likelihood estimator for $\theta, \alpha$, i.e., when both are assumed to be unknown. (**a closed-form expression is unlikely to exist and you will need to solve for it numerically**). Report the numerical estimates obtained using the $N$ independent samples from part (iii).

(v) Using the method of moments (MM), calculate an estimator for $\theta, \alpha$, and evaluate it using the samples from part (iii).

**Ex. 3.**
Consider i.i.d samples $X_1, X_2, \ldots, X_n$ from a Gaussian distribution with *unknown mean* $\mu$ and unknown variance $\sigma^2$. Our objective is to construct an interval estimator with a high confidence coefficient. Let us define the sample variance estimator

$$\sigma_s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Do the following:

- Consider the statistics $T_{n-1} := \frac{\bar{X}-\mu}{\sigma_s/\sqrt{n}}$, which we can rewrite as $T_{n-1} := \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma_s}$. What is the distribution of the numerator of $T_{n-1}$? What is the distribution of the denominator of $T_{n-1}$? Assume that the numerator and denominator are independent (this is true and was an important discovery by Gossett). What is the distribution of $T_{n-1}$?
- Use the above statistics to construct an interval estimator for $\mu$. Specify the left and right extreme point of the interval.
- Suppose we want to have confidence coefficient $1-\alpha$. How do we set the constant $c$ in the interval to achieve that?

**Ex. 4.**
Consider a model for the stock price process of the form:

$$S_{t_n} = S_{t_{n-1}} e^{\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t_n + \sigma Z_n},$$

where $\Delta t_n = t_n - t_{n-1}$ is the time interval between two consecutive observations, and $Z_n$ is a Gaussian random variable with mean zero and variance $\Delta t_n$. **Hint: We can express time in years, so that a one day time interval can be chosen to be $1/365$.**
Go on Yahoo Finance. Use the stock price data of IBM from January 1, 2021 to January 1, 2022. Do the following:

- Implement the maximum-likelihood estimator method, and report the maximum likelihood estimates for $\mu$ and $\sigma$.
- Using the Fisher Information matrix, compute an approximate 95% confidence interval for the parameters $\mu$ and $\sigma$. Please, specify how you have approximated the Fisher information matrix of the true (unknown) parameters $\mu$ and $\sigma$.

**Ex. 5.**
An important quantity that investors and portfolio managers are interested in estimating is the Sharpe ratio. This is defined as $\frac{\mu - r}{\sigma}$, i.e. the excess expected return over the risk-free rate $r$, measured in units of the volatility $\sigma$. Do the following

- Using the delta method, provide an estimator for the Sharpe ratio. Determine the asymptotic distribution of the estimator.

- Go on Google or Yahoo finance, and download the time series of Citigroup stock prices from January 1, 2024 through January 1, 2025. Compute the estimate of the drift $\mu$, volatility $\sigma$, and Sharpe ratio for the log-returns of the Citigroup stock. Additionally, provide an estimate for the variance of the estimator using the data (you will need to estimate the Fisher information matrix using one of the three methods outlined in class). Assume $r = 0.01$. Use the close prices data for your analysis.