STATISTICAL METHODS IN FINANCE, PROBLEM SHEET 4
MSC IN MATHEMATICS AND FINANCE, 2023-2024

CRIS SALVI, WILL TURNER

Exercise 1 (Performance of estimators)**.** We consider the statistical model $\mathcal{F} = \left\{\mathcal{N}(\mu, \sigma^2), \sigma^2 > 0\right\}$, with $\mu \in \mathbb{R}$ known. We wish to determine an optimal estimator of the variance σ^2. Consider a given iid sequence $(X_i)_{i=1,\dots,n}$, sampled from $\mathcal{N}(\mu, \sigma^2)$, and define, for any strictly positive sequence $(\alpha_n)_{n\geq 0}$, the estimator

$$\widehat{\theta}_n := \alpha_n \sum_{i=1}^{n} (X_i - \mu)^2 .$$

(i) For which values of $(\alpha_n)_{n\geq 0}$ is the sequence $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ consistent?

(ii) By considering the quadratic error $R_n\left(\widehat{\theta}_n, \theta\right) := \mathbb{E}\left[\left|\widehat{\theta}_n - \theta\right|^2\right]$, determine the value of α_n providing the most efficient estimator.

(iii) Consider now the so-called Stein's loss function $L\left(\widehat{\theta}_n, \theta\right) := \dfrac{\widehat{\theta}_n}{\theta} - 1 - \log\left(\dfrac{\widehat{\theta}_n}{\theta}\right)$. Find the value of α_n minimising $L\left(\widehat{\theta}_n, \theta\right)$. How does this estimator compare to the previous one?

Exercise 2 (Empirical CDF)**.** Let $X_i \overset{\text{iid}}{\sim} \text{Exp}(1)$ for $i = 1, \dots, n$, with $F_{X_i} = (1 - \mathrm{e}^{-x})\mathbb{1}_{(0,\infty)}(x)$. Write a Python script to plot the empirical CDF (ECDF) of $n \in \{10, 50, 100, 1000\}$ samples from Exp(1). Plot an approximate asymptotic 95% confidence intervals and the true CDF. What happens with the ECDF as n tends to infinity? What happens to the confidence intervals? Why is that?

Date: May 9, 2024.

Exercise 3 (SPX model)**.** In this exercise we will build a model for the daily returns of S&P 500 index (SPX) and calculate some of its statistics. We assume its underlying distribution to be the mixed normal distribution

$$F_{\mathrm{NMIX}}(x) = \sum_{i=1}^{k} w_i F_{\mathcal{N}}(\mu_i, \sigma_i^2), \quad w_i \geq 0, \ \sum_{i=1}^{k} w_i = 1 \text{ for some } k \in \mathbb{N},$$

where $F_{\mathcal{N}}$ is the CDF of the Normal distribution and $\boldsymbol{w} = (w_1, \dots, w_k)$ are the weights.

(i) Let $X \sim \mathrm{NMIX}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{w})$ and denote by $X_1, \dots, X_k$ the component distributions. Develop a general expression for the central moments of X.

(ii) Simulate $n = 10000$ samples from $\mathrm{NMIX}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{w})$ for

$$\boldsymbol{\mu} = [-0.2, 0.1, 0.1]^\top, \quad \boldsymbol{\sigma}^2 = [0.4, 0.1, 0.2]^\top, \quad \boldsymbol{w} = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]^\top$$

Hint: Draw the weights from the multinomial distribution with the vector of event probabilities $\boldsymbol{w}$*.* Furthermore, compute $\mathbb{E}[X]$ and $\mathbb{V}[X]$ using the expression in (i).

(iii) Estimate the mean and the standard variance of the generated samples with the simple unbiased estimators. What do you notice?

(iv) The Value-at-Risk (VaR) is defined as the $(1-\alpha)$-th quantile of X:

$$\mathrm{VaR}_\alpha(X) := -\inf\left\{x \in \mathbb{R} : F_X(x) > \alpha\right\}.$$

Estimate the VaR_α at $\alpha = 99\%, 95\%, 90\%$ from the generated data.

(v) Fit a Gaussian Kernel density estimator (KDE) to the generated data, and plot it. Does it fit the data well? *Hint: You can use `scipy.stats` library for kernel density estimation.*

Department of Mathematics, Imperial College London
Email address: wft17@ic.ac.uk