Imperial College London

Course: M5MF38
Setter: Dr Antoine Jacquier

MSc EXAMINATIONS IN MATHEMATICS AND FINANCE
DEPARTMENT OF MATHEMATICS

January 2019

M5MF38

Statistical Methods in Finance

Setter's signature

. . . .

Imperial College London

MSc EXAMINATIONS IN MATHEMATICS AND FINANCE
DEPARTMENT OF MATHEMATICS

January 2019

M5MF38

Statistical Methods in Finance

Date: January 2019 Time:

Answer all questions.

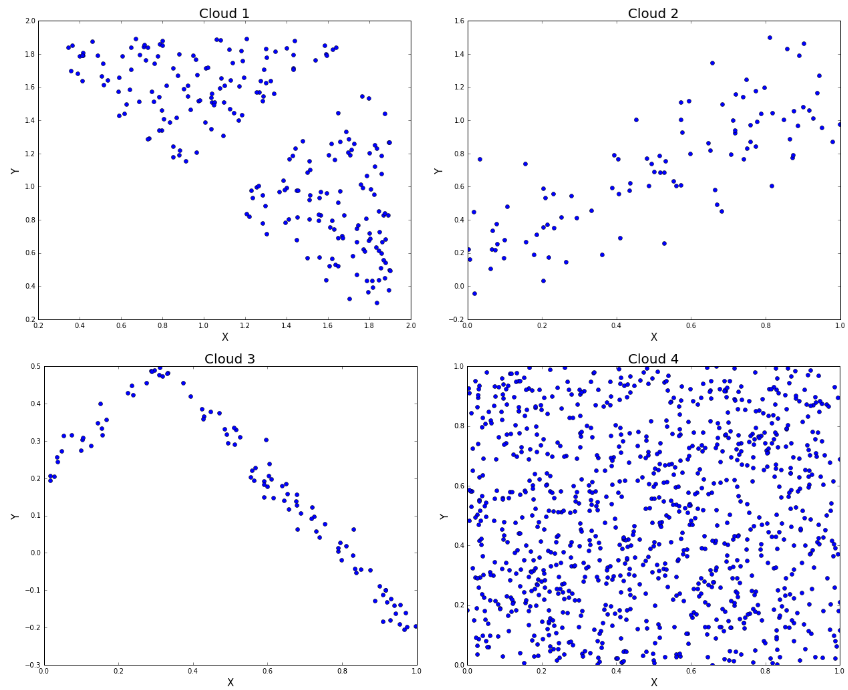
The total number of points is 100, and the precise grading is indicated in the text.

The rigour and clarity of your answers will be taken into account in the final grade.

Each problem is independent of the others.

1 [15 points] Warm-up: Preliminary questions

(i) [3 points] Consider the following four clouds of points $(x_i, y_i)_{1\leq i\leq n}$ generated from the two-dimensional distribution (X, Y): Which one(s) seem to correspond to independent X and Y? In the other case(s), what can you say about the (value of the) correlation between X and Y?



(ii) Let $X \sim \mathcal{N}(0,1)$ and, for some fixed $c > 0$, define

$$Y := \begin{cases} X, & \text{if } |X| > c, \\ -X, & \text{otherwise.} \end{cases}$$

(a) [3 points] Are X and Y independent? Justify your answer.

(b) [3 points] What is the law of Y?

(iii) [3 points] Consider the daily returns of the S&P 500. Explain in one paragraph how Principal Component Analysis can help explain their variations using the returns of all the constituents of the index.

(iv) [3 points] Consider the matrix $\mathbf{A} = (a_{i,j})_{1\leq,i,j,2} \in \mathcal{M}_{2,2}(\mathbb{R})$. Determine a sufficient and necessary condition on the coefficients ensuring that $\mathbf{A}$ is invertible, and compute $\mathbf{A}^{-1}$.

2 [30 points] Gender (in)Equalities and Pay Gap

We wish to determine the gap, if any, between the pay of male and female employees within a given company. We let Y_i denote the salary of the i-th employee, and consider the model, in $\mathbb{R}^n$,

$$\mathbf{Y} = \alpha \mathbf{1} + \beta \mathbf{X} + \boldsymbol{\varepsilon}.$$

We assume that the observation $\mathbf{X} = (X_1, \ldots, X_n)$ is deterministic, where $X_i = 1$ if the i-th employee is female, and $X_i = 0$ if he is male (we assume no other alternative). The vector $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ forms a centered Gaussian vector with covariance matrix $\sigma^2 \mathbf{I}_n$, with $\sigma > 0$ known, $\mathbf{I}_n$ denoting the identity matrix in the space of square $n \times n$ matrices. The parameters $\alpha, \beta \in \mathbb{R}$ are unknown.

(i) [3 points] What are the meaning of the parameters α and β. What happens to the model when $\beta = 0$ and when $\beta < 0$?

(ii) [10 points] Compute the maximum likelihood function of the problem, and determine a condition of the vector $\mathbf{X}$ ensuring the the maximum likelihood estimator $\widehat{\theta}_n := \left(\widehat{\alpha}_n, \widehat{\beta}_n\right)$ is well defined. We shall from now on always consider this assumption to hold. Show that

$$\widehat{\alpha}_n = \frac{\overline{Y} - \overline{XY}}{1 - \overline{X}} \qquad \text{and} \qquad \widehat{\beta}_n = \frac{\overline{XY}}{\overline{X}} - \frac{\overline{Y} - \overline{XY}}{1 - \overline{X}}$$

where $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{Y} := \frac{1}{n}\sum_{i=1}^{n} Y_i$, and $\overline{XY} := \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$.

(iii) [5 points] Determine the law of the estimator $\widehat{\beta}_n$ as a function of σ^2 and $\mathbf{X}$, and deduce an asymptotic confidence interval of level $1 - \gamma$ (for $\gamma \in (0, 1)$) for $\widehat{\beta}_n$.

(iv) [7 points] Build a statistical test of level γ for the null hypothesis $\mathcal{H}_0 : \{\text{there is no gender gap}\}$ vs the alternative $\mathcal{H}_1 : \{\text{there is a gender gap}\}$. Compute the power of the test, and discuss the influence of the parameters γ and β.

(v) [5 points] Build a statistical test of level γ for the null hypothesis $\mathcal{H}_0 : \{\text{Male salary is higher}\}$ vs the alternative $\mathcal{H}_1 : \{\text{Male salary is lower}\}$.

3 [35 points] Inverting Cochran's Theorem

Let us recall the following important theorem due to Cochran (Scottish statistician, 1909-1980):

Theorem 3.1. *For a subspace $\mathcal{M}$ of $\mathbb{R}^n$ of dimension p, denote by $\mathbf{P}$ the orthogonal projection matrix onto $\mathcal{M}$, and $\mathbf{P}^\perp := \mathbf{I}_n - \mathbf{P}$ the orthogonal projection matrix onto $\mathcal{M}^\perp$. Given a random variable $\mathbf{X} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$ with $\sigma > 0$, the vectors $\mathbf{PX}$ and $\mathbf{P}^\perp\mathbf{X}$ are independent, and*

$$\frac{\|\mathbf{PX}\|^2}{\sigma^2} \sim \chi^2_p \qquad \text{and} \qquad \frac{\|\mathbf{P}^\perp\mathbf{X}\|^2}{\sigma^2} \sim \chi^2_{n-p}.$$

We now consider an iid observation $\mathbf{X} := (X_1, \ldots, X_n)$, and define

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i \qquad \text{and} \qquad s_n^2 := \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2. \tag{1}$$

(i) [8 points] Assume (for this question only) that $X_i \sim \mathcal{N}(0,1)$ for $i = 1, \ldots, n$. Using Cochran's theorem, show that $\overline{X}_n$ and s_n^2 are independent and determine their distributions.

(ii) Assume now that $(X_1, \ldots, X_n)$ is iid, with unknown distribution, that $\mathbb{E}(X_1^2)$ is finite and that $\overline{X}_n$ and s_n^2 defined in (1) are independent. Let $\mu := \mathbb{E}[X_1]$, $\sigma^2 := \mathbb{V}[X_1]$ and, for any $\xi \in \mathbb{R}$, $\Phi(\xi) := \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi X_1}\right]$. The goal of this question is to conclude that $\mathbf{X}$ is in fact Gaussian.

(a) [6 points] Prove that $\mathbb{E}[n s_n^2] = (n-1)\sigma^2$, and show that, for any $\xi \in \mathbb{R}$,

$$\mathbb{E}\left[s_n^2 \exp\left(\mathtt{i}\xi n \overline{X}_n\right)\right] = \Phi^n(\xi)\mathbb{E}\left[s_n^2\right]. \tag{2}$$

(b) [2 points] Prove that, for any $\xi \in \mathbb{R}$,

$$\Phi'(\xi) = \mathtt{i}\mathbb{E}\left[X_1 \mathrm{e}^{\mathrm{i}\xi X_1}\right] \qquad \text{and} \qquad \Phi''(\xi) = -\mathbb{E}\left[X_1^2 \mathrm{e}^{\mathrm{i}\xi X_1}\right]. \tag{3}$$

(c) [2 points] Show that $n s_n^2$ can be written as

$$n s_n^2 = \left(1 - \frac{1}{n}\right)\sum_{i=1}^n X_i^2 - \frac{1}{n}\sum_{j=1}^n \sum_{k=1, k\neq j}^n X_j X_k. \tag{4}$$

(d) [1 point] For a smooth function $\psi : \mathbb{R} \to \mathbb{R}$, compute $\partial^2_{xx}\log(\psi(x))$.

(e) [10 points] Using (3) and (4), find a different formulation for the term $\mathbb{E}\left[s_n^2 \exp\left(\mathtt{i}\xi n \overline{X}_n\right)\right]$ in the left-hand side of (2), as a function of $\Phi'(\xi)$ and $\Phi''(\xi)$, and deduce that the function Φ is a solution to the ordinary differential equation

$$\frac{\Phi''(\xi)}{\Phi(\xi)} - \left(\frac{\Phi'(\xi)}{\Phi(\xi)}\right)^2 + \sigma^2 = 0,$$

with boundary condition $\Phi(0) = 1$ and $\Phi'(0) = \mathtt{i}\mu$.

(f) [6 points] Deduce that $X_i \sim \mathcal{N}(\mu, \sigma^2)$ for each $i = 1, \ldots, n$.

4 [20 **points**] **Estimation**

Let $\mathfrak{D}$ denote the space of all densities on $\mathbb{R}$. For any $f, g \in \mathfrak{D}$ we define the Hellinger distance as

$$\mathbf{H}(f,g) := \frac{1}{2}\int_{\mathbb{R}} \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 \mathrm{d}x.$$

(i) [6 points] Check that $\mathbf{H}$ defines a distance on $\mathfrak{D}$ and that $\mathbf{H}(f,g) \in [0,1]$ for any $f, g \in \mathfrak{D}$.

(ii) [2 points] Let $\theta > 0$ and denote by f_θ the density of the uniform distribution on $[0,\theta]$. Prove that, if $\theta \leq \theta'$, then

$$\mathbf{H}\left(f_\theta, f_{\theta'}\right) = 1 - \sqrt{\frac{\theta}{\theta'}}.$$

(iii) [6 points] Let $\mathbf{X} = (X_1, \ldots, X_n)$ denote an iid observation with common density f_θ defined previously. Find the maximum likelihood estimator $\widehat{\theta}_n$ of θ, together with its density, and compute the value of $\mathbb{E}\left[\widehat{\theta}_n^{1/2}\right]$.

(iv) [6 points] Show that, for any $\theta > 0$,

$$\mathbb{E}_\theta\left[\mathbf{H}\left(f_{\widehat{\theta}_n}, f_\theta\right)\right] = \frac{1}{2n+1},$$

and compare this to the square root of the quadratic risk $\sqrt{\mathbb{E}_\theta\left[\left(\widehat{\theta}_n - \theta\right)^2\right]}$.