

STATISTIQUE MATHÉMATIQUE – PARTIEL

M1 Mathématiques Fondamentales, Université Paris-Saclay
2024-2025

Lundi 17 février 2025
9h – 12h

Une feuille A4 recto manuscrite est autorisée en tant que support. La calculatrice n'est pas autorisée.

Avant de commencer :

- *Le sujet comporte quatre exercices indépendants. Il est demandé de les traiter dans l'ordre sur la copie.*
- *Des résultats pourront être admis d'une question sur l'autre, à la condition de l'écrire clairement.*
- *Une attention particulière sera portée aux questions d'interprétation, à la rigueur et à la précision de la rédaction.*
- *Le sujet est long et n'est pas nécessairement pensé pour être terminé dans son intégralité dans le temps imparti.*
- *Un barème indicatif est donné :*
 - Exercice 1 : 4 points,*
 - Exercice 2 : 8 points,*
 - Exercice 3 : 4 points.*
 - Exercice 4 : 4 points.*

Bon courage !

Exercice 1 – Fonction caractéristique des variables gaussiennes

Nous voulons montrer que, comme indiqué dans le cours, si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors pour tout $t \in \mathbb{R}$,

$$\Phi_X(t) = \exp\left(i\mu t - \frac{\sigma^2}{2}t^2\right).$$

0. Montrer le résultat lorsque X est dégénérée ($\sigma = 0$).

On suppose dans toute la suite que $\sigma > 0$.

1. Montrer que, pour tout $t \in \mathbb{R}$, on a

$$\Phi_X(t) = e^{it\mu} f_\sigma(t),$$

où

$$f_\sigma(t) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\sigma ty - y^2/2} dy.$$

2. Montrer que, pour tout $t \in \mathbb{R}$, $f_\sigma(t) \in \mathbb{R}$, que f_σ est différentiable, et trouver une équation différentielle qu'elle satisfait.
3. Conclure.
4. En déduire que si $X \sim \mathcal{N}(\mu, \sigma^2)$, et si $a, b \in \mathbb{R}$, alors

$$aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2).$$

5. En déduire que si $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ et que X_1 et X_2 sont indépendantes, alors

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Exercice 2 – Estimation dans un modèle quasi-géométrique

On considère la variable aléatoire X prenant des valeurs dans $\mathbb{N} \cup \{-1\} = \{-1, 0, 1, 2, \dots\}$ avec la distribution suivante, paramétrée par $p \in]0, 1[$:

$$\mathbb{P}_p(X = -1) = p, \quad \text{et} \quad \forall k \geq 0, \mathbb{P}_p(X = k) = (1 - p)^2 p^k.$$

Nous considérons un modèle où $(X_1, \dots, X_n)$ sont n variables aléatoires indépendantes de même loi que X .

Suffisance dans le modèle

1. Calculer $\mathbb{E}_p[X_1]$ pour tout $p \in]0, 1[$.
2. Montrer que la vraisemblance du modèle se met sous la forme

$$L(p, x_1, \dots, x_n) = (1 - p)^{2n} \left(\frac{p}{(1 - p)^2} \right)^{nU(x_1, \dots, x_n)} p^{nV(x_1, \dots, x_n)},$$

où

$$U(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_i = -1} \quad \text{et} \quad V(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n x_i \mathbb{1}_{x_i \geq 0}.$$

3. En déduire une statistique suffisante du modèle.
4. Cette statistique est-elle complète ? Justifier.

Estimation par maximum de vraisemblance

5. Montrer que pour cette vraisemblance, il existe un unique estimateur du maximum de vraisemblance, noté $\hat{p}$, dont on donnera l'expression.
6. Montrer que $\hat{p}$ est fortement consistant.
7. Montrer que $\hat{p}$ est asymptotiquement normal. On donnera les paramètres de la loi normale correspondante en fonction de p . On pourra d'abord considérer la statistique

$$S = \frac{1}{2}(U + V)$$

et montrer qu'elle est asymptotiquement normale.

Cas d'une observation : estimation UVMB

Dans tout le reste de l'exercice, nous nous focalisons sur le cas d'une seule observation ($n = 1$), et l'on note $X_1 = X$ pour alléger les notations.

8. Montrer que l'ensemble des estimateurs $T : \mathbb{N} \cup \{-1\} \rightarrow \mathbb{R}$ sans biais de p est un espace affine de dimension 1, que vous décrirez. On pourra commencer par caractériser l'ensemble

$$\mathcal{U}_0 := \{T : \mathbb{N} \cup \{-1\} \rightarrow \mathbb{R} \mid \forall p \in]0, 1[, \mathbb{E}_p[T(X)] = 0\}.$$

9. Que dire d'un estimateur UVMB de p ?
10. Voyez-vous un lien entre la réponse à la question 9 et celle de la question 4 ? Interpréter.

Exercice 3 – Information de Fisher dans un modèle linéaire gaussien

On considère le modèle statistique où $X_1, \dots, X_n$ sont des variables réelles indépendantes, avec pour tout $1 \leq i \leq n$, $X_i \sim \mathcal{N}(\alpha + \beta t_i, 1)$, où les constantes $(t_i)_{1 \leq i \leq n}$ sont connues et $\alpha, \beta \in \mathbb{R}$ sont des paramètres inconnus.

1. Donner une condition nécessaire et suffisante sur les t_i pour que le modèle soit identifiable. Interpréter.

Dans toute la suite, on se placera sous les conditions d'identifiabilité de la question 1.

2. Déterminer la matrice d'information de Fisher $I(\alpha, \beta)$. Comment se traduit sur $I(\alpha, \beta)$ la condition de la question 1 ? Interpréter.
3. Donner une borne inférieure sur la variance de tout estimateur non biaisé de α , suffisamment lisse.
4. Supposons que l'on connaît β . Donner dans ce cas une borne inférieure sur la variance de tout estimateur non biaisé de α suffisamment lisse.
5. Comment les bornes inférieures des questions 3 et 4 se comparent-elles ? Interpréter.

Exercice 4 – Estimation UVMB avec deux observations

On considère un modèle où X_1 et X_2 sont deux variables i.i.d. de même loi, ayant une densité par rapport à la mesure de comptage sur $\mathbb{N}$ donnée par

$$\forall x \in \mathbb{N}, f_\theta(x) = (x+1)\theta^2(1-\theta)^x,$$

avec θ inconnu dans $\Theta :=]0, 1[$.

-
1. Déterminer $\mathbb{E}_\theta \left[\frac{1}{X_1+1} \right]$ pour tout $\theta \in \Theta$.
 2. Montrer que $X_1 + X_2$ est une statistique suffisante.
 3. Déterminer la loi de $X_1 + X_2$ et montrer que $X_1 + X_2$ est complète.
 4. Montrer qu'il existe un unique estimateur UVMB de θ , et donner son expression. *On commencera par déterminer la densité jointe de $(X_1, X_1 + X_2)$.*

Fin du sujet.