STT1000 – Examen Intra 2

(Automne 2021)

Les calculatrices sont autorisées. Les documents sont en revanche interdits, sauf une page d'aide mémoire. L'examen dure 2 heures. Toute sortie avant la fin est autorisée, mais sera définitive.
La feuille propose 7 exercices et un barême approximatif est donné à titre indicatif. Les réponses doivent être reportées sur le cahier joint. Si vous utilisez 2 cahiers, merci de le mentionner, en indiquant 1/2 et 2/2 respectivement.
N'hésitez pas à faire des dessins pour vous aider, mais ne considérez pas un dessin comme une preuve. Si vous utilisez un résultat du cours dans votre preuve, nommez-le aussi précisément que possible.

Si $X \sim \mathcal{N}(0,1)$, on a les probabilités suivantes

$$\begin{cases} \mathbb{P}[X \leq -3] \approx 0.1350\% & \mathbb{P}[X \leq 1] \approx 84.1345\% \\ \mathbb{P}[X \leq -2] \approx 2.2750\% & \mathbb{P}[X \leq 2] \approx 97.7250\% \\ \mathbb{P}[X \leq -1] \approx 15.865\% & \mathbb{P}[X \leq 3] \approx 99.8650\% \\ \mathbb{P}[X \leq 0] = 50.0000\% & \mathbb{P}[X \leq 4] \approx 99.9968\% \end{cases}$$

Si vous pensez que des hypothèses manquent pour répondre à la question, indiquez le dans le cahier. Si vous avez besoin d'introduire des objets mathématiques non définis dans l'énoncé, définissez les clairement.

Exercice 1 – ***(Surface d'un pré)*** [10 points]

Un agriculteur possède un champ carré, dont il veut estimer la superficie. Quand il mesure un côté de son champ, il sait que l'erreur de la mesure est une variable aléatoire d'espérance nulle et de variance σ^2, et que les erreurs sont indépendantes d'une mesure à l'autre. Il réalise une première mesure d'un des côtés et trouve x_1. Il en déduit la superficie $s_1 = x_1^2$. Par précaution, il effectue une seconde mesure de ce même côté, et trouve x_2. Il en déduit la superficie $s_2 = x_2^2$.

Perplexe, il abandonne ses mesures et réfléchit pour savoir quelle est la bonne façon de procéder. Il dispose de deux estimations de la surface de son champ s_1 et s_2, mais considère 3 autres estimations :

$$s_3 = \frac{s_1 + s_2}{2},\ s_4 = \left(\frac{x_1 + x_2}{2}\right)^2 \text{ et } s_5 = x_1 x_2.$$

Parmi ces 5 statistiques, lesquels sont des estimateurs sans biais de la surface ?

Exercice 2 – ***(Moyenne)*** [20 points]

Pour estimer l'espérance μ d'une population, on utilise souvent la moyenne empirique $\widehat{\mu} = \overline{x}$, à partir d'un échantillon $\{x_1, \cdots, x_n\}$, obtenu à partir de variables i.i.d. Soit $a \in [0, 1]$. On propose l'estimateur $\widetilde{\mu} = a \cdot \overline{x}$.

1. Calculer biais($\widehat{\mu}$), Var($\widehat{\mu}$) et l'erreur quadratique moyenne de $\widehat{\mu}$.
2. Calculer biais($\widetilde{\mu}$), Var($\widetilde{\mu}$) et l'erreur quadratique moyenne de $\widetilde{\mu}$.
3. Tracer (sommairement) $a \mapsto$ biais$(\widetilde{\mu})^2$ et $a \mapsto$ Var($\widetilde{\mu}$) sur $[0, 1]$.
4. Existe-t-il a tel que l'erreur quadratique de $\widetilde{\mu}$ soit inférieure à l'erreur quadratique de $\widehat{\mu}$?

Exercice 3 – ***(Loi normale censurée)*** *[15 points]*

Considérons un échantillon $\{x_1, \cdots, x_n\}$ tiré suivant une loi $\mathcal{N}(\theta, 1)$. Malheureusement, seules des indicatrices, indiquant que les observations étaient positives, ont été observées, i.e. $\{y_1, \cdots, y_n\}$, avec $y_i = \mathbf{1}(x_i > 0)$.

1. Quelle est l'estimateur de la méthode des moments de θ (construit à partir des observations y_i) ?
2. Quelle est l'estimateur du maximum de vraisemblance de θ (construit à partir des observations y_i) ?

Exercice 4 – [15 points]

On considère des variables independentes $Y_1, \cdots, Y_n$ suivant des lois de Bernoulli $\mathcal{B}(p)$ i.i.d. On veut estimer $\theta = p(1-p)$. Assez naturellement, on pose $\widehat{\theta} = \overline{Y}(1 - \overline{Y})$.

1. Calculer biais($\widehat{\theta}$).
2. À partir de $\widehat{\theta}$, construire un estimateur sans biais de θ.

Exercice 5 – ***(Borne de Cramér-Rao)*** [15 points]

On considère des variables independentes $Y_1, \cdots, Y_n$ de densité

$$f_\theta(y) = 2^\theta \theta y^{-(1+\theta)} \mathbf{1}_{[2,\infty)}(y), \text{ avec } \theta > 1.$$

Calculer la borne de Cramér Rao pour ce modèle.

Exercice 6 – ***(Estimateurs)*** [20 points]

On considère des variables independentes $Y_1, \cdots, Y_n$ de loi

$$f_\theta(y) = \mathbb{P}[Y_i = y] = (1-\theta)^{y-1}\theta, \text{ pour } y = 1, 2, \cdots \text{ et } \theta \in (0,1).$$

1. Calculer l'estimateur de la méthode des moments de θ.
2. Calculer l'estimateur du maximum de vraisemblance de θ.

Exercice 7 – ***(Maximum de vraisemblance)*** [20 points]

On considère la fonction de répartition sur $[0, \infty)$,

$$F_\theta(y) = \mathbb{P}[Y \leq y] = \left(1 + \frac{1}{y^2}\right)^{-\theta}, \ \theta > 0.$$

On dispose d'un échantillon $\{y_1, \cdots, y_n\}$ tiré suivant des variables i.i.d. de loi F_θ.

1. Calculer l'estimateur du maximateur du maximum de vraisemblance $\widehat{\theta}$
2. Calculer l'information de Fisher de θ dans ce modèle.
3. Donner la loi de $\sqrt{n}\left(\widehat{\theta} - \theta\right)$ lorsque $n \to \infty$?