STT1000 – Examen Final

(Automne 2021)

Les calculatrices sont autorisées. Les documents sont en revanche interdits, sauf une page d'aide mémoire. L'examen dure 3 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 $Q \sim \chi^2(50)$, on a les probabilités suivantes

$$\begin{cases} \mathbb{P}[Q \leq 35] \approx \ 5.31\% & \mathbb{P}[Q \leq 34.764] \approx \ 5.00\% \\ \mathbb{P}[Q \leq 40] \approx 15.67\% & \mathbb{P}[Q \leq 39.754] \approx 15.00\% \\ \mathbb{P}[Q \leq 50] \approx 52.66\% & \mathbb{P}[Q \leq 49.335] \approx 50.00\% \\ \mathbb{P}[Q \leq 60] \approx 84.27\% & \mathbb{P}[Q \leq 60.346] \approx 85.00\% \\ \mathbb{P}[Q \leq 70] \approx 96.76\% & \mathbb{P}[Q \leq 67.505] \approx 95.00\% \end{cases}$$

Si $X_i \sim \mathcal{N}(0, \sigma^2)$ sont i.i.d., $Q = kS^2/\sigma^2 \sim \chi^2(k)$ où $kS^2 = X_1^2 + \cdots + X_k^2$. De plus $\mathbb{E}(Q) = k$ et $\text{Var}[Q] = 2k$.
Pour rappel, la Δ-méthode permet d'affirmer que si $g : \mathbb{R} \to \mathbb{R}$ est une fonction dérivable, telle que $g'(\mu) \neq 0$, si

$$\sqrt{n}\Big(X_n - \mu\Big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2), \text{ lorsque } n \to \infty,$$

alors

$$\sqrt{n}\Big(g(X_n) - g(\mu)\Big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, g'(\mu)^2\sigma^2), \text{ lorsque } n \to \infty.$$

Si X est une variable aléatoire de fonction de répartition F, positive ($F(x) = 0$ pour $x < 0$), d'espérance finie

$$\mathbb{E}[X] = \int_0^\infty (1 - F(x))dx.$$

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 – ***(Test Gaussien)*** [10 points]

On considère un échantillon $\{x_1, x_2, \cdots, x_n\}$ tiré suivant une loi $\mathcal{N}(\theta, 1)$. Pour tester $H_0 : \theta \leq 5$ contre $H_1 : \theta > 5$, la region de rejet suivante est considérée,

$$\left\{(x_1, x_2, \cdots, x_n) : \; \overline{x} > 5 + \frac{1}{\sqrt{n}}\right\}$$

1. Quelle est la probabilité d'erreur de première espèce de ce test ?
2. Quelle est la probabilité d'erreur de seconde espèce de ce test ?

Exercice 2 – ***(Δ-méthode)*** [20 points]

On considère un échantillon de n observations $\{y_1, y_2, \cdots, y_n\}$ provenant de n variables aléatoires indépendantes de loi $f_\theta(y) = \theta x^{\theta-1}\mathbf{1}_{[0,1]}(y)$, où $\theta \in (0, \infty)$.

1. Calculer l'estimateur par la méthode des moments $\widehat{\theta}$ de θ, et donnez g telle que $\widehat{\theta} = g(\overline{y})$
2. En utilisant le théorème central limite, donnez la distribution approchée de $\sqrt{n}(\overline{Y} - \mu)$, où $\mu = \mathbb{E}[Y]$ sera un paramètre que l'on explicitera.
3. En utilisant la Δ-méthode, donne la distribution approchée de $\sqrt{n}(\widehat{\theta} - g(\mu))$
4. En déduire un intervalle de confiance à 95% pour θ.

Exercice 3 – ***(Loi uniforme)*** [10 points]

On dispose de 5 observations, $\{0.95, \; 0.24, \; 0.83, \; 0.52, \; 0.69\}$, qu'on suppose tirées suivant une loi uniforme sur $[0, \theta]$.

1. Donner l'estimateur de la méthode des moments, $\widetilde{\theta}$, et sa valeur numérique.
2. Donner l'estimateur du maximum de vraisemblance $\widehat{\theta}$, et sa valeur numérique.
3. on veut tester $H_0 : \theta = 1$ contre $H_1 : \theta = 1.1$. Proposer une procédure de test.

Exercice 4 – ***(Variance dans un modèle Gaussien)*** [25 points]

On dispose de n observations, $\{x_1, x_2, \cdots, x_n\}$, qu'on suppose tirées suivant une loi normale centrée, de variance θ, autrement dit, de loi $\mathcal{N}(0, \theta)$.

1. Donner l'estimateur de la méthode des moments, $\widetilde{\theta}$, tel que $\widetilde{\theta}$ soit un estimateur sans biais de θ.
2. Donner l'estimateur du maximum de vraisemblance $\widehat{\theta}$
3. Montrer que l'information de Fisher vérifie $I_1(\theta) = 1/(2\theta)$.
4. Montrer que $\widetilde{\theta}$ est un estimateur efficace de θ.
5. Donner la loi de $n\widetilde{\theta}/\theta$.
6. Proposer deux intervalles de confiance à 95% pour θ, en supposant $n = 50$: le premier de la forme $[0, a]$ et le second de la forme $[b, \infty)$.

Exercice 5 – ***(Observations partielles de variables uniformes)*** [25 points]

0. Soit U une variable aléatoire, avec la fonction de répartition suivante

$$F(u) = \begin{cases} 0 & \text{si } u \leq 1 \\ u/2 & \text{si } u \in (1, 2) \\ 1 & \text{si } u \geq 2 \end{cases}$$

   Montrer que

$$\mathbb{E}[U] = \frac{5}{4}.$$

On dispose de n observations, $\{x_1, x_2, \cdots, x_n\}$, qu'on suppose tirées suivant une loi uniforme sur $[0, \theta]$, avec $\theta > 1$.

1. On suppose qu'on n'observe pas les x_i mais les y_i, où

$$y_i = \underbrace{\max\{1, x_i\}}_{=g(x_i)} = \begin{cases} 1 \text{ si } x_i \leq 1 \\ x_i \text{ si } x_i > 1 \end{cases}$$

2. Si X suit une loi uniforme sur $[0, \theta]$, et si $Y = g(X)$, donner la fonction de répartition de Y. En déduire la loi du maximum, $\max\{Y_1, \cdots, Y_n\}$.
3. En déduire un estimateur sans biais de θ construit à partir de $\{y_1, y_2, \cdots, y_n\}$
4. On suppose qu'on n'observe pas les x_i mais les z_i, où

$$z_i = \underbrace{\min\{1, x_i\}}_{=h(x_i)} = \begin{cases} x_i \text{ si } x_i \leq 1 \\ 1 \text{ si } x_i > 1 \end{cases}$$

5. Soit R le nombre d'observations Z plus petites (strictement) que 1. Donner la loi de R.
6. Proposez un estimateur sans biais de $\mathbb{P}[X > 1]$, construit à partir de R

Exercice 6 – ***(Échantillon de 1 et de 2)*** [15 points]

On dispose de n observations, $\{x_1, x_2, \cdots, x_n\}$, qu'on suppose obtenues comme des réalisations de variables i.i.d. et prennant les valeurs 1 et 2, avec

$$\mathbb{P}[X_i = 1] = \frac{2-2\theta}{2-\theta} \text{ et } \mathbb{P}[X_i = 2] = \frac{\theta}{2-\theta}$$

où $\theta \in (0,1)$.

1. Calculer l'estimateur $\widehat{\theta}$ obtenu par la méthode des moments du paramètre θ
2. Montrez que

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(a, \frac{(2-\theta)^2(2\theta - \theta^2 - 2)}{2}\right), \text{ quand } n \to \infty$$

   où a sera une constante qu'on déterminera.

Exercice 7 – ***(Accidents de la route)*** [10 points]

Dans une ville on donne la répartition du nombre de jours sans accident, avec un accident, deux accidents, etc. parmi 50 jours d'observation au cours d'une même année :

Nombre d'accidents	0	1	2	3	4
Nombre de jours	21	18	7	3	1

On suppose que le nombre d'accidents par jour, X suit une loi de Poisson. Donner un intervalle de confiance de niveau 95% pour le nombre moyen d'accidents par jour (on utilisera une approximation asymptotique). Expliquer rapidement quel test vous proposeriez pour tester $H_0 : X \sim \mathcal{P}(1)$ (avec l'hypothèse alternative $H_1 : X \not\sim \mathcal{P}(1)$).