III – Inférence et décision

Exercice 1

On considère lors de la tranmission de paquets IP par Wi-Fi un taux de perte de 1 % acceptable.

Quelle est la probabilité, lors de la transmission d'un fichier vidéo de 120 Mo, que 20 paquets ou moins soient perdus ?

(NB : taille maximale d'un paquet IPv4 = 65 535 octets)

Exercice 2

On dispose d'un générateur de nombres aléatoires qui renvoie des nombres uniformément distribués dans un intervalle $[0, \theta]$. Malheureusement, θ est inconnu, c'est pourquoi on cherche à l'estimer à partir d'un n-échantillon $(X_1, X_2, \ldots, X_n)$ — les X_k étant supposés indépendants.

Dix valeurs ont ainsi été reportées ci-dessous :

14,41	36,48	38,24	34,84	1,53	21,71	26,33	2,11	10,50	10,10

Premier estimateur

La loi uniforme sur $[0, \theta]$ étant centrée en $\theta/2$, nous posons comme premier estimateur

$$\Theta \;:=\; 2\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right).$$

a) Calculer l'espérance de Θ et montrer que cet estimateur est non biaisé.

b) Rappeler ou recalculer la variance d'une loi uniforme sur $[0, \theta]$.
En déduire la variance de $\displaystyle\sum_{k=1}^{n} X_k$ puis celle de Θ.

c) Plutôt que de travailler avec la loi précise de Θ, qui est pénible à expliciter lorsque $n \geqslant 3$, nous nous contenterons ici de supposer qu'il s'agit d'une loi normale ayant l'espérance et la variance calculées plus haut. Pourquoi est-ce une approximation raisonnable ?

d) Sous cette hypothèse, donner la loi de la variable aléatoire $T = \dfrac{\Theta - \theta}{\theta/\sqrt{3n}}$.

e) Étant donné $\alpha \in [0, 1]$, soit t_α le nombre réel pour lequel $\mathbb{P}[-t_\alpha \leqslant T \leqslant t_\alpha] = \alpha$. Montrer qu'il y a une probabilité α que θ soit contenu dans l'intervalle aléatoire

$$I_\alpha = \left[\frac{\Theta}{1 + \frac{t_\alpha}{\sqrt{3n}}},\; \frac{\Theta}{1 - \frac{t_\alpha}{\sqrt{3n}}}\right].$$

f) À l'aide des données fournies dans l'introduction, calculer l'intervalle de confiance pour θ de niveau 95 % obtenu à l'aide de l'estimateur Θ.

Deuxième estimateur

On se propose dans cette partie d'estimer θ à partir de la plus grande valeur observée :

$$M \;:=\; \max(X_1, X_2, \ldots, X_n).$$

a) Pour X et Y deux v.a. indépendantes, exprimer la fonction de répartition de $Z := \max(X, Y)$ grâce aux fonctions de répartition F_X et F_Y de X et Y.

b) En déduire la densité f_Z de Z en fonction de celles de X et Y.

c) Reprendre les calculs dans le cas de n variables i.i.d. $(X_1, X_2, \ldots, X_n)$. Donner la fonction de répartition du maximum M, puis montrer que sa densité est de la forme :

$$f_M = n\, f\, F^{n-1}$$

où F et f sont les fonctions de répartition et la densité communes à toutes les variables.

d) Expliciter ceci dans le cas $X_i \sim \mathcal{U}([0, \theta])$.

e) Calculer l'espérance de M et montrer qu'il s'agit d'un estimateur *biaisé* de θ.
Proposer alors un estimateur sans biais, proportionnel à M, que nous noterons Φ.

f) Calculer l'espérance de Φ^2 puis la variance de Φ.
Lequel, de Θ ou de Φ, est le meilleur estimateur ? pourquoi ?

g) À l'aide de la fonction de répartition de M calculée ci-haut, déterminez en fonction de $\alpha \in [0, 1]$ des valeurs c_α et d_α pour lesquelles $\mathbb{P}[c_\alpha \leqslant \frac{M}{\theta} \leqslant d_\alpha] = \alpha$. Serait-il raisonnable d'approximer la loi de M par une gaussienne ici ?

h) Montrer que

$$J_\alpha = \left[\frac{\Phi}{(1 + \frac{1}{n})d_\alpha}, \frac{\Phi}{(1 + \frac{1}{n})c_\alpha} \right]$$

est un intervalle de confiance de niveau α pour θ.

i) Calculez l'intervalle de niveau 95 % ainsi obtenu pour l'échantillon de la page précédente.

Exercice 3

Un étudiant d'une grande école d'ingénieurs du numérique comptabilise 15 h d'absences non excusées. Le directeur, estimant que la feuille d'appel ne passe typiquement que dans la moitié des séances, se dit que cet étudiant doit sans doute avoir été absent en réalité au moins 30 h et décide de le convoquer en conseil de discipline. Qu'en pensez-vous ?

Exercice 4

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes d'espérance μ et variance σ^2. Calculer l'espérance de

$$\Delta_n = \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

et conclure sur le caractère biaisé ou non des estimateurs $S_n^2 = \dfrac{\Delta_n}{n}$ et $\widetilde{S}_n^2 = \dfrac{\Delta_n}{n-1}$ pour σ^2.

Exercice 5

Un test servant à dépister une maladie grave vient d'être mis au point. Ses résultats, annonce l'Institut Pastèque, sont très fiables : si le sujet est malade, il y a 98% de « chances » que le test soit positif et, si l'on n'est pas malade, le test donne un résultat négatif 95% du temps. Sachant que la maladie ne frappe, heureusement, que 1% de la population, quelle est la probabilité, pour un patient dont le test est positif, de passer ses prochaines vacances au Père Lachaise ?

Exercice 6

À Francfort, les taxis sont numérotés 1, 2, $\ldots$, t, mais nous ne connaissons pas t.

Dans la rue, nous observons 4 taxis portant les numéros

$$n_1 = 512\ , \quad n_2 = 987\ , \quad n_3 = 355\ , \quad n_4 = 1200.$$

Notre chauffeur prétend que $t \geqslant 3000$. Si cette affirmation est vraie, quelle est la probabilité d'observer, parmi quatre valeurs aléatoires, des résultats aussi faibles ? Croyez-vous qu'il ait raison ?