L2 Mass - Stat   Corrigé de l'examen du 13 mai 19

Q $\underset{\text{de } E(X)}{\text{estimateur}}$ = formule f en $X_1, \dots, X_n$ (qui approxime $E(X)$)   (1 pt)

convergent si pour tout $\varepsilon > 0$ $\lim\limits_{n\to+\infty} P(|f(X_1,\dots,X_n) - E(X)| > \varepsilon) = 0$ (1 pt)

ou si $E\left(\left(f(X_1,\dots,X_n) - E(X)\right)^2\right) \underset{n\to\infty}{\longrightarrow} 0$

ou si $E(f(X_1,\dots,X_n)) \underset{n\to\infty}{\longrightarrow} E(X)$ et $Var(f(X_1,\dots,X_n)) \underset{n\to\infty}{\longrightarrow} 0$

La moyenne empirique $\frac{1}{n}(X_1 + \dots + X_n)$ est un estimateur convergent de $E(X)$ d'après la loi des grands nombres (1 + pt)

Ex2 a $M_n = \frac{1}{n}(X_1 + \dots + X_n)$ est un estimateur convergent et sans biais de $E(X) = \int_0^a x \frac{4x^3}{a^4}\,dx = \left[\frac{4}{5}\frac{x^5}{a^4}\right]_0^a = \frac{4}{5}a$ (1 + 1)

donc $\frac{5}{4}M_n$ est un estimateur convergent sans biais de a : $\underbrace{E\left(\frac{5}{4}M_n\right) = \frac{5}{4}E(X) = a}_{1}$, $Var\left(\frac{5}{4}M_n\right) = \left(\frac{5}{4}\right)^2 \frac{Var(X)}{n} \to 0$

b $M = \max(X_1, \dots, X_n)$ est un estimateur convergent de a (maximum de vraisemblance) (1)

$P(|M - a| > \varepsilon) = P(X < a - \varepsilon)^n = \left(1 - \frac{\varepsilon}{a}\right)^{4n}$ si $\varepsilon \leq a$ tend vers 0 si $\varepsilon > 0$ (1)

$= 0$ sinon

Ex3 a Au pire chacun gagne 100 et on doit alors avoir $c \geq 100$

ou bien gain $= 1000c - N\,100 = 100(10c - N)$ où N = nbre de billets gagnants $\in \{0, \dots, 1000\}$ (1)

gain $\geq 0 \Leftrightarrow 10c \geq N$ ce qui n'est garanti que si $c \geq 100$

b $E(\text{gain}) = 100(10c - E(N)) \geq 0$ si $c \geq \frac{1}{10}E(N)$ (1)

$N \sim B(1000, 2/100) \leadsto E(N) = 20 \leadsto c \geq 2$ (1 + 0,5)

c On veut $P(N > 10c) < \frac{1}{100}$ (1)

$P(N > 10c) = P\left(\frac{N - E(N)}{\sigma(N)} > \frac{10c - E(N)}{\sigma(N)}\right) \underset{(1)}{\approx} 1 - F_{N(0,1)}\left(\frac{10c - E(N)}{\sigma(N)}\right) < \frac{1}{100}$ si $\frac{10c - E(N)}{\sigma(N)} \geq 2.33$ d'après table (1)

$c \geq \frac{E(N)}{10} + 2.33\,\frac{\sigma(N)}{10} = 2 + 1,03$ $\left(E(N) = 1000 \times 2/100 = 20\ ;\ \sigma(N) = \sqrt{1000 \times 2/100 \times 98/100} = \frac{14}{\sqrt{10}}\right)$ (1)

Ex4 a Intervalle de fluctuation de $\frac{\varepsilon}{\sigma_\varepsilon} \sim N(0,1)$ au seuil 80% : $[-1,28 ; 1.28]$ d'après table

$\leadsto$ intervalle de fluctuation de ε : $[-1,28\sigma ; 1.28\sigma] = [-2,56 ; 2.56]$ (2)

$\leadsto$ ——— $\ell + \varepsilon$: $[\ell - 2,56 ; \ell + 2,56]$

b $\ell = 120$ $P(\ell + \varepsilon < 118) = P(\varepsilon < -2) = P\left(\frac{\varepsilon}{2} < -1\right) = F_{N(0,1)}(-1) = 1 - F_{N(0,1)}(1) = 1 - 0,84 = 0,16$ (2)

$\uparrow$ d'après table

c $N :=$ nbre de tiges de long. < 118 $N \sim B(100 ; 0,16)$ d'après b.   (1)

$P(N \geq 20) = P\left(\frac{N - E(N)}{\sigma(N)} \geq \frac{20 - E(N)}{\sigma(N)}\right) \approx 1 - F_{N(0,1)}\Big(\underbrace{\frac{20 - E(N)}{\sigma(N)}}_{\approx 1.09}\Big) \approx 0,14$ d'après table   (1)

statistique de test : N dont on calcule sous l'hyp. $\ell = 120$ la p-valeur pour la valeur observée de N (1)

ici la p-valeur est 0,14 ; on rejette l'hyp. $\ell = 120$ au seuil 0,14

d N = nbre de tiges de long. $\ell 120$. Sous l'hyp $\ell = 120$ on a $N \sim B(100 ; 1/2)$ car la médiane de ε est 0   (1)

p-valeur de N : $1 - F_{B(100, 1/2)}(80) \approx 1 - F_{N(50, 5^2)}(80) = 1 - F_{N(0,1)}\left(\frac{80 - 50}{5}\right) < 1 - F_{N(0,1)}(4.5) \underset{(1,5)}{=} 3.10^{-6}$ d'après table

donc oui on rejette l'hyp $\ell = 120$