Exercices de base - Processus Empiriques

Master 2 Probabilités et Statistiques

Exercices fondamentaux

Exercice 1 : Fonction de répartition empirique

Soit X_1, X_2, X_3 trois observations indépendantes de loi uniforme sur $[0, 1]$:

$$X_1 = 0.2, \quad X_2 = 0.5, \quad X_3 = 0.8.$$

1. Donner l'expression de la fonction de répartition empirique $F_3(t)$.
2. Tracer le graphe de $F_3(t)$ pour $t \in [-0.5, 1.5]$.
3. Calculer $F_3(0.3)$, $F_3(0.5)$, $F_3(1)$.
4. Vérifier que F_3 est une fonction en escalier, croissante, continue à droite.

Exercice 2 : Propriétés élémentaires

Soit $X_1, \ldots, X_n$ i.i.d. de fonction de répartition F. On note F_n la fonction de répartition empirique.

1. Montrer que pour t fixé, $nF_n(t)$ suit une loi binomiale $\mathcal{B}(n, F(t))$.
2. En déduire que $\mathbb{E}[F_n(t)] = F(t)$.
3. Montrer que $\mathrm{Var}(F_n(t)) = \dfrac{F(t)(1 - F(t))}{n}$.
4. Calculer $(F_n(s), F_n(t))$ pour $s \leq t$.

Exercice 3 : Convergence ponctuelle

Soit $X_1, \ldots, X_n$ i.i.d. de loi de Bernoulli de paramètre $p = 0.7$.

1. Pour $n = 10$, simuler (ou imaginer) un échantillon et calculer $F_{10}(0.5)$.
2. Quelle est la loi de $F_n(0.5)$?
3. En utilisant la loi des grands nombres, que peut-on dire de la limite de $F_n(0.5)$ quand $n \to \infty$?
4. Calculer $\P(|F_n(0.5) - 0.7| > 0.1)$ pour $n = 100$ en utilisant l'inégalité de Tchebychev.

Exercice 4 : Distance de Kolmogorov-Smirnov

On considère l'échantillon : $0.1, 0.3, 0.4, 0.7, 0.9$.

1. Tracer sur un même graphique :
   - La fonction de répartition empirique $F_5(t)$
   - La fonction de répartition théorique $F(t) = t$ (loi uniforme sur $[0, 1]$)

2. Calculer la distance de Kolmogorov-Smirnov :

$$D_5 = \sup_{t \in [0,1]} |F_5(t) - t|.$$

3. Aux points de saut de F_5, calculer les différences $|F_5(t) - t|$.

Exercice 5 : Inégalité de Dvoretzky-Kiefer-Wolfowitz (DKW)

1. Énoncer l'inégalité DKW.
2. Pour $n = 100$ et $\alpha = 0.05$, calculer la largeur des bandes de confiance :

$$\varepsilon = \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\alpha}\right)}.$$

3. Interpréter : avec probabilité au moins 0.95, on a pour tout t :

$$F_n(t) - \varepsilon \leq F(t) \leq F_n(t) + \varepsilon.$$

4. Comparer avec l'intervalle de confiance usuel pour une proportion.

Exercice 6 : Pont brownien élémentaire

Un pont brownien $B(t)$ sur $[0,1]$ vérifie :

$$B(0) = B(1) = 0, \quad \mathbb{E}[B(t)] = 0, \quad (B(s), B(t)) = \min(s,t) - st.$$

1. Calculer $\mathrm{Var}(B(0.5))$.
2. Pour $s = 0.3$ et $t = 0.7$, calculer $(B(0.3), B(0.7))$.
3. Montrer que $B(t)$ et $B(1-t)$ ont même loi.
4. Si $B(t) \sim \mathcal{N}(0, t(1-t))$, calculer $\P(|B(0.5)| > 1)$.

Exercice 7 : Bootstrap simple

On observe : $2.1, 3.4, 1.7, 2.8, 2.3$.

1. Calculer la fonction de répartition empirique F_5.
2. Tirer un échantillon bootstrap de taille 5 (par exemple : $2.8, 2.1, 2.3, 2.1, 3.4$).
3. Calculer la fonction de répartition bootstrap F_5^*.
4. Calculer $D_5^* = \sup_t |F_5^*(t) - F_5(t)|$.

Exercice 8 : Test d'adéquation simple

On veut tester si des données proviennent d'une loi uniforme sur $[0,1]$. Les données sont : $0.15, 0.45, 0.65, 0.85$.

1. Calculer la statistique de Kolmogorov-Smirnov D_4.
2. La valeur critique pour $n = 4$ au niveau 5% est environ 0.624. Conclure.
3. Proposer une procédure bootstrap pour estimer la p-valeur.

Exercice 9 : Processus empirique sur un échantillon

Soit l'échantillon trié : $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$.

1. Exprimer $F_n(t)$ en fonction des statistiques d'ordre.
2. Montrer que $F_n(x_{(k)}) = \frac{k}{n}$.
3. Pour $t \in [x_{(k)}, x_{(k+1)})$, que vaut $F_n(t)$?
4. Calculer $F_n(x_{(k)}^-)$ (limite à gauche).

Exercice 10 : Simulation et convergence

1. Simuler (ou imaginer) un échantillon de taille $n = 20$ de loi $\mathcal{N}(0,1)$.
2. Tracer $F_{20}(t)$ et $\Phi(t)$ (f.d.r. de la loi normale).
3. Calculer $D_{20} = \sup_t |F_{20}(t) - \Phi(t)|$.
4. Réfléchir à ce qui se passe quand n augmente.

Corrections guidées

Exercice 1

1. $F_3(t) = \frac{1}{3}[\mathbf{1}_{\{0.2 \leq t\}} + \mathbf{1}_{\{0.5 \leq t\}} + \mathbf{1}_{\{0.8 \leq t\}}]$.
2. Fonction en escalier avec sauts en 0.2, 0.5, 0.8.
3. $F_3(0.3) = 1/3$, $F_3(0.5) = 2/3$, $F_3(1) = 1$.

Exercice 2

1. $nF_n(t) = \sum_{i=1}^n \mathbf{1}_{\{X_i \leq t\}} \sim \mathcal{B}(n, F(t))$.
2. $\mathbb{E}[F_n(t)] = \frac{1}{n}\mathbb{E}[nF_n(t)] = \frac{1}{n}nF(t) = F(t)$.
3. $\mathrm{Var}(F_n(t)) = \frac{1}{n^2}\mathrm{Var}(nF_n(t)) = \frac{1}{n^2}nF(t)(1 - F(t))$.
4. Pour $s \leq t$: $(\mathbf{1}_{\{X_i \leq s\}}, \mathbf{1}_{\{X_i \leq t\}}) = F(s) - F(s)F(t)$.

Exercice 5

1. ¶$(\sup_t |F_n(t) - F(t)| > \varepsilon) \leq 2e^{-2n\varepsilon^2}$.
2. $\varepsilon = \sqrt{\frac{1}{200}\ln(40)} \approx 0.136$.
3. Les bandes sont : $[F_n(t) - 0.136, F_n(t) + 0.136]$.
4. Pour une proportion, l'intervalle est $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Exercice 6

1. $\mathrm{Var}(B(0.5)) = 0.5 \times 0.5 = 0.25$.
2. $(B(0.3), B(0.7)) = 0.3 \times 0.3 = 0.09$.
3. Même moyenne et même covariance.
4. $B(0.5) \sim \mathcal{N}(0, 0.25)$, ¶$(|B(0.5)| > 1) = 2(1 - \Phi(2)) \approx 0.0455$.