Série d'Exercices n°3
Processus Empiriques

Master 2 Probabilités et Statistiques
Département de Mathématiques
Université Mohamed Khider - Biskra

Année Universitaire 2025/2026

Exercice 1 : Théorèmes Fondamentaux

a) **Théorème de Glivenko-Cantelli**
   i) Énoncer précisément le théorème de Glivenko-Cantelli.
   ii) Soit $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de fonction de répartition continue F. On note F_n la fonction de répartition empirique. Montrer que :

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow{p.s.} 0 \quad \text{quand } n \to \infty$$

   iii) Donner une borne sur la vitesse de convergence en utilisant l'inégalité de Dvoretzky-Kiefer-Wolfowitz.

b) **Théorème de Donsker (Théorème Fonctionnel Central Limite)**
   i) Énoncer le théorème de Donsker pour le processus empirique.
   ii) Soit $\mathbb{G}_n(t) = \sqrt{n}(F_n(t) - F(t))$. Montrer que pour F continue, $\mathbb{G}_n$ converge en loi vers un pont brownien $B \circ F$ dans l'espace $\ell^\infty(\mathbb{R})$.
   iii) Expliquer pourquoi ce résultat est fondamental pour la construction des tests d'adéquation non paramétriques.

Exercice 2 : Pont Brownien et Distribution de Kolmogorov-Smirnov

Soit $B = \{B(t), 0 \leq t \leq 1\}$ un pont brownien standard.

a) **Définition et propriétés du pont brownien**
   i) Donner la définition d'un pont brownien à partir d'un mouvement brownien.
   ii) Calculer $\mathbb{E}[B(t)]$ et $\mathrm{Cov}(B(s), B(t))$ pour $0 \leq s \leq t \leq 1$.
   iii) Montrer que $B(t)$ et $B(1-t)$ ont même loi.

b) **Maximum d'un pont brownien et distribution de KS**
   i) Soit $M = \sup_{0 \leq t \leq 1} |B(t)|$. La fonction de répartition de M est donnée par :

$$P(M \leq x) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2x^2}$$

   En déduire que sous H_0, pour n grand :

$$P(\sqrt{n} D_n \leq x) \approx 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2x^2}$$

   où $D_n = \sup_t |F_n(t) - F(t)|$.

ii) Calculer les valeurs critiques pour $\alpha = 0.05$ et $\alpha = 0.01$ à partir de cette formule. Comparer avec les valeurs tabulées pour $n = 20$.

Exercice 3 : Inégalités de Concentration et Applications

Soit $\mathcal{F} = \{\mathbb{1}_{(-\infty,t]}, t \in \mathbb{R}\}$ la classe des fonctions indicatrices des demi-droites.

a) **Inégalité de Dvoretzky-Kiefer-Wolfowitz (DKW)**

i) Énoncer l'inégalité DKW : pour tout $\epsilon > 0$,

$$P\left(\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| > \epsilon\right) \leq 2e^{-2n\epsilon^2}$$

ii) Utiliser cette inégalité pour construire des bandes de confiance non-asymptotiques pour F. Expliciter les bornes inférieure et supérieure pour un niveau de confiance $1 - \alpha$.

iii) Pour $n = 100$ et $\alpha = 0.05$, calculer la largeur des bandes de confiance. Application numérique.

b) **Comparaison avec l'approche de Kolmogorov-Smirnov**

i) Montrer que pour n fixé, l'approche DKW donne des bandes de confiance plus larges que celle basée sur la distribution exacte de Kolmogorov-Smirnov.

ii) Expliquer pourquoi on préfère souvent utiliser le test de Kolmogorov-Smirnov en pratique, malgré l'avantage de non-asymptoticité de DKW.

Exercice 4 : Processus Empirique Indexé par des Classes de Fonctions

Soit $\mathcal{F}$ une classe de fonctions mesurables bornées.

a) **Processus empirique généralisé**

i) Définir le processus empirique indexé par $\mathcal{F}$:

$$\mathbb{G}_n(f) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right), \quad f \in \mathcal{F}$$

ii) Donner deux exemples de classes $\mathcal{F}$ et expliquer leur intérêt statistique.

b) **Classes Donsker**

i) Définir ce qu'est une classe Donsker.

ii) Donner une condition suffisante sur $\mathcal{F}$ pour qu'elle soit Donsker (condition sur la dimension VC).

iii) Montrer que la classe $\mathcal{F} = \{\mathbb{1}_{(-\infty,t]}, t \in \mathbb{R}\}$ est Donsker.

Exercice 5 : Bootstrap du Processus Empirique

Soit $X_1, \ldots, X_n$ i.i.d. de loi F, et soit $X_1^*, \ldots, X_n^*$ un échantillon bootstrap tiré de F_n.

a) **Processus empirique bootstrap**

i) Définir $F_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i^* \leq t\}}$.

ii) Montrer que conditionnellement à $X_1, \ldots, X_n$, on a :

$$\sqrt{n}(F_n^* - F_n) \xrightarrow{\mathcal{L}} \mathbb{G} \quad \text{presque sûrement}$$

où $\mathbb{G}$ est le processus empirique limite.

b) **Application au test de Kolmogorov-Smirnov**

   i) Proposer une procédure bootstrap pour approximer la distribution de la statistique $D_n = \sup_t |F_n(t) - F(t)|$.

   ii) Expliquer l'intérêt de cette approche bootstrap, notamment dans le cas où les paramètres de la loi théorique sont estimés à partir des données.

Ressources bibliographiques

— Van der Vaart, A. W. (2000). *Asymptotic Statistics.*

— Shorack, G. R., & Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.*

— Lehmann, E. L., & Romano, J. P. (2005). *Testing Statistical Hypotheses.*

— Articles originaux :

  — Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione.

  — Cramér, H. (1928). On the composition of elementary errors.

  — Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit.

Bon travail !