Feuille de TP n°3
Loi des Grands Nombres et Théorème de la Limite Centrale

1 Loi des Grands Nombres

La loi des grands nombres (LGN) est un résultat fondamental en Probabilités [1]. Elle affirme que si $(X_n)_n$ est une suite de v.a.r. indépendantes et de même loi que X et si $S_n = \sum_{k=1}^n X_k$, alors

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbb{E}(X)$$

si et seulement si X est intégrable. La version multidimensionnelle s'en déduit aisément.

Exercice 1.1 (Glivenko-Cantelli). Soit $(X_n)_n$ une suite de v.a. indépendantes et de même loi, de fonction de répartition F. Soit F_n la fonction de répartition empirique de $(X_1, \ldots, X_n)$, définie $\forall x \in \mathbb{R}$ par

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{I}_{\{X_k \leqslant x\}}.$$

La LGN entraîne que $\forall x \in \mathbb{R}$, $F_n(x) \to F(x)$ p.s. Le théorème de Glivenko-Cantelli renforce cela en une convergence uniforme :

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n\to+\infty]{\text{p.s.}} 0.$$

On trouvera une preuve dans [3, Théorème 4.4.25]. Créer un programme illustrant ce résultat sur N réalisations i.i.d. de loi binomiale $\mathcal{B}(n,p)$, de loi exponentielle $\mathcal{E}(\lambda)$ et de loi géométrique $\mathcal{G}(p)$. Pour N assez grand, vérifier la LGN sur les estimateurs empiriques des deux premiers moments associés à ces lois.

Exercice 1.2 (Taille du support d'une loi uniforme). Soit $(X_n)_n$ une suite de v.a. i.i.d. de loi uniforme sur $\{1, \ldots, \theta\}$. Montrer que $\widehat{\theta}_n = 2\overline{X}_n - 1$ est un estimateur sans biais et fortement consistant de θ. Simulation !

Exercice 1.3 (Processus auto-régressif). On considère le processus autorégressif stable $X_n = \theta X_{n-1} + \varepsilon_n$ avec $|\theta| < 1$, $X_0 = 0$ et (ε_n) i.i.d. centrée, de variance σ^2. On estime θ par l'estimateur des moindres carrés $\widehat{\theta}_n = (\sum_{k=1}^n X_k X_{k-1})/(\sum_{k=0}^{n-1} X_k^2)$. Si $S_n = \sum_{k=0}^n X_k^2$, montrer que $S_n/n \to \sigma^2/(1-\theta^2)$ p.s. En déduire que $\widehat{\theta}_n \to \theta$ p.s. Vérifier-le par simulation pour divers choix de $(\varepsilon_n)_n$. Montrer également que $\sqrt{n}(\widehat{\theta}_n - \theta)$ converge en loi vers une loi $\mathcal{N}(0, 1-\theta^2)$ et vérifier-le par simulation. On pourra consulter [4, Section 4.4.4].

Exercice 1.4 (Processus de Galton-Watson explosif). On considère le processus de Galton-Watson explosif

$$Y_n = \sum_{k=1}^{Y_{n-1}} X_{n,k}$$

avec $Y_0 = 1$. La suite $(X_n)_n$ représente le nombre d'individus à la n^{e} génération. La suite $(X_{n,k})_{n,k}$, qui représente les descendants, est i.i.d. à valeurs dans $\mathbb{N}$, de moyenne $m > 1$ et de variance $\sigma^2 > 0$. Montrer que

[1] Sous des hypothèses adéquates, la LGN peut se démontrer en quelques lignes, on peut même simplifier la preuve déjà très courte qui se trouve dans [2, Exercice 5.5]. L'affaiblissement des hypothèses de la LGN conduisent à divers raffinements. Etemadi a montré qu'elle reste encore vraie si $(X_n)_n$ est une suite de v.a. deux à deux indépendantes et de même loi. Elle est également vraie si $(X_n)_n$ n'est pas constituée de v.a. de même loi. Kolmogorov a établi que si $(X_n)_n$ est une suite de v.a. indépendantes, centrées, de carré intégrable et satisfaisant $\sum_{n=1}^{\infty} n^{-2}\mathbb{E}(X_n^2) < \infty$ alors, $S_n/n \to 0$ p.s. Finalement, Rademacher et Menchov ont montré que, sans l'hypothèse d'indépendance, si $(X_n)_n$ est une suite de v.a. centrées, de carré intégrable, non corrélées et satisfaisant $\sum_{n=1}^{\infty}(n^{-1}\log n)^2\,\mathbb{E}(X_n^2) < \infty$ alors, $S_n/n \to 0$ p.s.

Y_n/m^n converge p.s. vers une v.a. L. En déduire que $\tilde{m}_n = Y_n/Y_{n-1}$ et $\widehat{m}_n = \sum_{k=2}^n Y_k / \sum_{k=1}^{n-1} Y_k$ sont deux estimateurs fortement consistants de m. Proposer un estimateur fortement consistant $\widehat{\sigma}_n^2$ de σ^2. Vérifier par simulation la consistance de ces estimateurs lorsque les $(X_{n,k})_{n,k}$ suivent la loi binomiale $\mathcal{B}(4, 1/2)$. Essayer d'autres lois. On pourra consulter [5, Section 12.1],

2 Théorème de la Limite Centrale

Le théorème de la limite centrale (TLC) est le second résultat fondamental en Probabilités[2]. Si $(X_n)_n$ est une suite de v.a. i.i.d. de carré intégrable, d'espérance m et de variance $\sigma^2 > 0$, alors

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Ici encore, une version multidimensionnelle s'en déduit aisément.

Exercice 2.1. Créer un programme illustrant le TLC pour un N-échantillon de loi Uniforme $\mathcal{U}([0,1])$.

Exercice 2.2 (Franchissement de barrières). Soit $a < b$ et soit $(X_n)_n$ une suite de v.a.r. i.i.d. de variance finie. Montrer en utilisant le TLC que $\inf\{n \geqslant 1; X_1 + \cdots + X_n \notin [a,b]\}$ est fini p.s., cf. [2, Exercice 8.8]. Simulations !

Exercice 2.3 (Test du χ^2). Soit $(X_n)_n$ une suite de v.a.r. i.i.d. de loi p, ne prenant que k valeurs distinctes. Si $p_n := \sum_{i=1}^n \delta_{X_i}/n$ est la loi empirique, alors le TLC entraîne que la pseudo-distance $\chi^2(p, p_n) := \sum_{i=1}^k (p_n(i) - p(i))^2/p(i)$ converge en loi vers la loi $\chi^2(k-1)$ quand n tend vers $+\infty$ (donne un test d'adéquation à p). Illustrer cela par un programme sur des données trouvées par exemple dans [6, Chapitre 15] où [3, Sections 5.4].

Exercice 2.4 (Test de Kolmogorov-Smirnov). Dans le cadre de l'exercice 1.1. Si F est *continue*, on a

$$\sqrt{n}\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n\to+\infty]{\mathcal{L}} L_F$$

où L_F est une v.a. de fonction de répartition $\Phi_F(x) = 1 + 2\sum_{k=1}^\infty (-1)^k \exp(-2k^2x^2)$ si $x > 0$ et $\Phi_F(x) = 0$ sinon. Ce résultat permet la construction d'un test d'adéquation sur F, appelé test de Kolmogorov-Smirnov[3]. La bibliothèque Stixbox fournit la fonction `pks` qui permet de calculer Φ_F. Créer un programme illustrant ce test pour un N-échantillon de loi normale $\mathcal{N}(0,1)$. Cf. par exemple [6, Section 15.4] et [3, Sections 4.5].

Exercice 2.5 (Application amusante en analyse). Pour tout $x, \lambda > 0$, soit $L_n(x) = \exp(-n\lambda)\sum_{k=0}^{[nx]} (n\lambda)^k/k!$. Montrer à l'aide de la LGN et du TLC que $L_n(x)$ tend vers 0 si $x < \lambda$, 1/2 si $x = \lambda$ et 1 sinon. Vérifier-le par simulation. Cf. [2, Exercice 4.15], et aussi [2, Exercice 5.12] dans le même esprit.

Remarque 2.6 (Autres théorèmes limites du type LGN ou TLC). Ils sont très nombreux et les exercices de cette feuille n'en donnent qu'un maigre apperçu. Cependant, les principaux concernent les martingales et les chaînes de Markov, comme nous le allons le voir au fil des feuilles.

Références

[1] P. Barbe et M. Ledoux – *Probabilités*, De la licence à l'agrégation, Belin, 1998.
[2] M. Cottrell, C. Duhamel, V. Genon-Catalot et T. Meyre – *Exercices de probabilités*, Cassini, 1999, Deuxième édition. Avec rappels de cours. Préface de Gabriel Ruget.
[3] D. Dacunha-Castelle et M. Duflo – *Probabilités et statistiques. Tome 1*, Masson, Paris, 1982, Problèmes à temps ↓xe.
[4] — , *Probabilités et statistiques. Tome 2*, Masson, Paris, 1983, Problèmes à temps mobile.
[5] D. Foata et A. Fush – *Processus stochastiques - processus de poisson, chaînes de markov et martingales*, Dunod, 2002.
[6] G. Saporta – *Probabilités, analyse des données et statistique*, Technip, 1989.

[2] On trouvera une preuve dans [1, Chapitre 5] par exemple. Le premier TLC, dû à De Moivre et connu sous le nom de Théorème de Moivre-Laplace, concerne les v.a. de Bernoulli. Il affirme que si $(X_n)_n$ est une suite de v.a. i.i.d. de loi de Bernoulli $\mathcal{B}(p)$ avec $0 < p < 1$, alors $\forall a, b \in \mathbb{R}$ avec $a < b$

$$\lim_{n\to+\infty} \mathbb{P}\left(a \leqslant \frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}} \leqslant b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{x^2}{2}\right) dx.$$

[3] Pour les tests de Kolmogorov-Smirnov et du χ^2, le TLC donne le niveau asymptotique tandis que la LGN assure que la puissance converge vers 1.