
Feuille de TP n°3

Loi des Grand Nombres et Théorème de la Limite Centrale

1 Loi des Grands Nombres

La loi des grands nombres (LGN) est un résultat fondamental en Probabilités. Elle affirme que si $(X_n)_n$ est une suite de variables aléatoires indépendantes et de même loi que X et si $S_n = \sum_{k=1}^n X_k$, alors $\lim_{n \rightarrow +\infty} S_n/n = \mathbb{E}(X)$ p.s. si et seulement si X est intégrable. On peut trouver divers raffinements de la LGN. 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, $S_n/n \rightarrow 0$ 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, $S_n/n \rightarrow 0$ p.s.

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, \dots, X_n) , définie $\forall x \in \mathbb{R}$ par

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

On a par le Théorème de Glivenko-Cantelli que $\forall x \in \mathbb{R}$, $F_n(x) \rightarrow F(x)$ p.s. et

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

Créer un code Matlab 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 $\mathcal{G}(p)$ où les paramètres de ces lois sont affectés par l'utilisateur. 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 variables aléatoires i.i.d. de loi uniforme sur $\{1, \dots, \theta\}$. Montrer que $\hat{\theta}_n = 2\bar{X}_n - 1$ est un estimateur sans biais et fortement consistant de θ . Vérifier-le par 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 $\hat{\theta}_n = (\sum_{k=1}^n X_k X_{k-1}) / (\sum_{k=0}^{n-1} X_k^2)$. Si $S_n = \sum_{k=0}^n X_k^2$, montrer que $S_n/n \rightarrow \sigma^2 / (1 - \theta^2)$ p.s. En déduire que $\hat{\theta}_n \rightarrow \theta$ p.s. Vérifier-le par simulation pour divers choix de $(\varepsilon_n)_n$. Montrer également que $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi vers une loi $\mathcal{N}(0, 1 - \theta^2)$ et vérifier-le par simulation.

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

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

avec $X_0 = 1$. La suite $(X_n)_n$ représente le nombre d'individus à la n^e génération. La suite $(Y_{n,k})$, 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 X_n/m^n converge p.s. vers une v.a. L . En déduire que $\hat{m}_n = X_n/X_{n-1}$ et $\hat{m}_n = \sum_{k=2}^n X_k / \sum_{k=1}^{n-1} X_k$ sont deux estimateurs fortement consistants de m . Proposer un estimateur fortement consistant $\hat{\sigma}_n^2$ de σ^2 . Vérifier par simulation la consistance de ces estimateurs lorsque $(Y_{n,k})$ est associée à la loi binomiale $\mathcal{B}(4, 1/2)$. Essayer d'autres lois.

2 Théorème Limite Centrale

Le théorème de la limite centrale (TLC) est le second résultat fondamental en Probabilités. 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 variables aléatoires i.i.d. de loi de Bernoulli $\mathcal{B}(p)$ avec $0 < p < 1$, alors $\forall a, b \in \mathbb{R}$ avec $a < b$

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x^2}{2}\right) dx$$

où $S_n = \sum_{k=1}^n X_k$ et $q = 1 - p$. Plus généralement, si $(X_n)_n$ est une suite de variables aléatoires i.i.d. de carré intégrable, d'espérance m et de variance $\sigma^2 > 0$, alors

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

Exercice 2.1. Créer un code Matlab permettant d'illustrer le TLC pour un N -échantillon de loi Uniforme $\mathcal{U}([0, 1])$.

Exercice 2.2 (Test de Kolmogorov-Smirnov). On se place dans le cadre de l'exercice 1.1. Si F est continue, on a

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

où L_F est une v.a. de fonction de répartition $\Phi_F(x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k \exp(-2k^2 x^2)$ si $x > 0$ et $\Phi_F(x) = 0$ sinon. La bibliothèque Stixbox fournit la fonction Matlab `pks` qui permet de calculer Φ_F . Grâce à cette convergence, on peut construire un test d'adéquation sur F , appelé test de Kolmogorov-Smirnov. Créer un code Matlab permettant d'illustrer le test de Kolmogorov-Smirnov pour un N -échantillon de loi normale $\mathcal{N}(0, 1)$.

Exercice 2.3 (Estimateur à noyau d'une densité). Soit $(X_n)_n$ une suite de v.a. indépendantes et de même loi, de densité de probabilité f . A partir de F_n , on estime f par \hat{f}_n définie $\forall x \in \mathbb{R}$ par $\hat{f}_n(x) = (F_n(x + h_n) - F_n(x - h_n))/2h_n$ où $(h_n)_n$ est une suite positive décroissante avec $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$. On peut par exemple choisir $h_n = n^{-\alpha}$ avec $0 < \alpha < 1$. Montrer que \hat{f}_n converge en probabilité vers f et que $\forall x \in \mathbb{R}$,

$$\sqrt{2nh_n} \frac{(\hat{f}_n(x) - \mathbb{E}(\hat{f}_n(x)))}{\sqrt{\hat{f}_n(x)}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Créer un code Matlab permettant de vérifier ce résultat pour la loi normale $\mathcal{N}(0, 1)$ et la loi exponentielle $\mathcal{E}(1)$.

Exercice 2.4 (Sommes partielles de la série exponentielle et loi de Poisson). Pour tout $x, \lambda > 0$, soit $L_n(x) = \exp(-n\lambda) \sum_{k=0}^{\lfloor nx \rfloor} (n\lambda)^k / k!$. 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.