TD10. Loi des grands nombres, théorème central limite

1. Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes toutes de loi uniforme sur l'intervalle $[0,1]$. Soit $f : [0,1] \to \mathbb{R}$ une fonction continue. Que peut-dire de

$$\frac{f(U_1) + \ldots + f(U_n)}{n}$$

lorsque n tend vers $+\infty$?

Solution de l'exercice 1. La fonction f, continue sur le segment $[0,1]$, y est bornée. Ainsi, les variables aléatoires $(f(U_n))_{n\geq 1}$ sont indépendantes, de même loi et admettent un moment d'ordre 1. On peut leur appliquer la loi forte des grands nombres (puisque la fonction f est bornée, les variables admettent un moment d'ordre 2 et on est même dans le cas dont on a étudié la démonstration en cours) pour trouver que $\frac{f(U_1)+\ldots+f(U_n)}{n}$ converge presque sûrement, lorsque n tend vers l'infini, vers

$$\mathbb{E}[f(U_1)] = \int_0^1 f(t)\, dt.$$

Cette méthode est parfois utilisée pour calculer des valeurs approchées d'intégrales et s'appelle la *méthode de Monte-Carlo.*

2. On considère une suite de jets indépendants d'un dé équilibré. On désigne par X_k le résultat du k-ème jet et par Y_n le plus grand résultat observé au cours des n premiers jets.

$$Y_n = \max_{1\leq k\leq n} X_k$$

a) Que peut-on dire des propriétés de convergence de la suite $(Y_n)_n$?

b) On pose : $N_n = Card\{k \leq n : X_k = 6\}$ pour tout $n \in \mathbb{N}^*$. Etablir la convergence presque sûre de la suite $\left(\frac{N_n}{n}\right)_{n\in\mathbb{N}^*}$.

Solution de l'exercice 2.

a) Il est assez évident que la suite va converger presque sûrement vers 6 : Y_n vaut 6 à partir de la première observation d'un 6, et la probabilité qu'on ne fasse aucun 6 est nulle. Formalisons ceci

$$\mathbb{P}(\exists k \in \mathbb{N} : X_k = 6) = 1$$

Or $(\omega : \exists k \in \mathbb{N} : X_k(\omega) = 6) \subset (\omega : \lim_{n\to\infty} Y_n(\omega) = 6)$, d'où

$$\mathbb{P}\left(\lim_{n\to\infty} Y_n = 6\right) = 1$$

et la convergence presque sûre de Y_n vers 6.

b) N_n la variable aléatoire qui compte le nombre de 6 peut se réécrire $N_n = \sum_{k=1}^n \mathbb{1}_{X_k=6}$. La suite de variables aléatoires $(Z_k := \mathbb{1}_{(X_k=6)})$ est indépendante identiquement distribuée. Et $\mathbb{1}_{(X_1=6)}$ suit une loi de Bernoulli de paramètre $\mathbb{P}(X_1 = 6) = \frac{1}{6}$, elle possède donc une espérance. D'après la loi forte des grands nombres, on a

$$\frac{N_n}{n} = \frac{Z_1 + ... + Z_n}{n} \overset{p.s.}{\to} \frac{1}{6}$$

3. A l'approche des élections, un institut de sondage contacte successivement des individus. Notre modèle est le suivant : les appels sont indépendants et chaque individu répond qu'il va voter pour le candidat A avec probabilité p_A (et pour le candidat B avec probabilité $p_B = 1 - p_A$). Le but est d'*estimer* le paramètre p_A du modèle.

a) Soit $N_A(n)$ le nombre de réponses en faveur du candidat A collectées en n appels. Que dire de la convergence de la suite $\frac{N_A(n)}{n}$?

b) L'application de l'inégalité de Tchebychev à la variable aléatoire $\frac{N_A(n)}{n}$ permet-elle de retrouver ce résultat ?

c) Déduire de cette même inégalité un intervalle $I(n, \delta)$ tel que

$$\mathbb{P}\left(p_A \in I(n, \delta)\right) \geq 1 - \delta$$

Un tel intervalle est appelé *intervalle de confiance* pour le paramètre p_A.

Solution de l'exercice 3. Introduisons une suite X_k de Bernoullis indépendantes de paramètre p_A telle que X_k prend la valeur 1 si le k-ème individu appelé se prononce pour le candidat A, et 0 dans le cas contraire.

a) On note que $N_A(n) = \sum_{k=1}^n X_k$ et donc d'après la loi forte des grands nombres,

$$\frac{N_A(n)}{n} = \frac{X_1 + X_2 + ... + X_n}{n} \overset{p.s}{\to} p_A$$

car la suite (X_k) est iid et que X_1 admet pour espérance p_A.

b) On a

$$\begin{aligned} \mathbb{E}\left[\frac{N_A(n)}{n}\right] &= p_A \\ \mathrm{Var}\left[\frac{N_A(n)}{n}\right] &= \frac{1}{n^2} n \mathrm{Var}(X_1) = \frac{p_A(1-p_A)}{n} \end{aligned}$$

donc l'inégalité de Tchebychev appliquée à la variable aléatoire $\frac{N_A(n)}{n}$ donne, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{N_A(n)}{n} - p_A\right| \geq \varepsilon\right) \leq \frac{p_A(1-p_A)}{n\varepsilon^2}$$

Si on avait $\sum_{t=0}^{\infty} \mathbb{P}\left(\left|\frac{N_A(n)}{n} - p_A\right| \geq \varepsilon\right) < \infty$ par application du lemme de Borel-Cantelli (cf exercice 5 TD 8), on obtiendrait la convergence presque sûre de $\frac{N_A(n)}{n}$ vers p_A. Mais l'inégalité obtenue par l'inégalité obtenue ne nous donne pas une majoration par le terme général d'une série convergente, donc n'est pas suffisante pour obtenir la convergence presque sûre...

c) Utiliser l'inégalité de Tchebychev va toutefois nous permettre de construire un intervalle de confiance de niveau $1-\delta$ pour p_A, c'est-à-dire une région qui contienne p_A avec probabilité au moins $1-\delta$. Pour cela on choisit dans l'inégalité de Tchebychev ci-dessus ε tel que $\frac{p_A(1-p_A)}{n\varepsilon} = \delta$. Pour cette valeur de ε on a alors

$$\mathbb{P}\left(\left|\frac{N_A(n)}{n} - p_A\right| \leq \varepsilon\right) \geq 1-\delta$$

En remarquant que $\varepsilon = \frac{1}{\sqrt{n}}\sqrt{\frac{p_A(1-p_A)}{\delta}}$, on peut poser

$$I(n,\delta) = \left[\frac{N_A(n)}{n} - \frac{1}{\sqrt{n}}\sqrt{\frac{p_A(1-p_A)}{\delta}}; \frac{N_A(n)}{n} + \frac{1}{\sqrt{n}}\sqrt{\frac{p_A(1-p_A)}{\delta}}\right]$$

4. Soit $f : [0,1] \to \mathbb{R}$ une fonction continue. Pour tout $n \geq 0$, on définit la fonction $b_n : [0,1] \to \mathbb{R}$ par la formule

$$\forall x \in [0,1],\ b_n(x) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

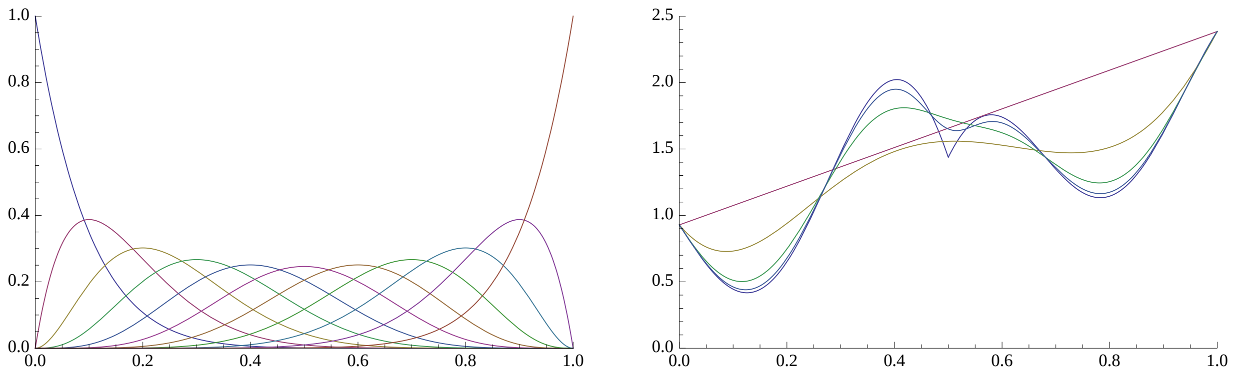
a. Montrer que la suite de fonctions $(b_n)_{n\geq 0}$ converge simplement vers f, c'est-à-dire

$$\forall x \in [0,1],\ \lim_{n\to\infty} b_n(x) = f(x).$$

b. Montrer que la convergence est uniforme, c'est-à-dire que

$$\lim_{n\to\infty} \sup\{|b_n(x) - f(x)| : x \in [0,1]\} = 0.$$

On a démontré le *théorème de Stone-Weierstrass* : toute fonction continue sur un segment y est uniformément approximable par une suite de fonction polynômiales. Plus précisément on a approximé f par une famille de *polynômes de Bernstein.*



Solution de l'exercice 4.

a) Soit $x \in [0,1]$. On observe que $b_n(x) = \mathbb{E}[f(B/n)]$, où B suit la loi binomiale de paramètres n et x. Considérons donc $(X_n)_{n\geq 1}$ une suite de v.a.i.i.d. de loi de Bernoulli de paramètre x. On a

$$b_n(x) = \mathbb{E}\left[f\left(\frac{X_1 + \ldots + X_n}{n}\right)\right].$$

D'après la loi des grands nombres, $\frac{X_1+\ldots+X_n}{n}$ tend presque sûrement, quand n tend vers l'infini, vers $\mathbb{E}[X_1] = x$. Puisque f est continue, $f\left(\frac{X_1+\ldots+X_n}{n}\right)$ tend donc presque sûrement vers $f(x)$. Enfin, puisque f est continue sur $[0,1]$, elle est bornée. La convergence presque sûre de $f\left(\frac{X_1+\ldots+X_n}{n}\right)$ vers $f(x)$ est donc dominée par le supremum de f et le théorème de convergence dominée permet d'affirmer que

$$b_n(x) = \mathbb{E}\left[f\left(\frac{X_1 + \ldots + X_n}{n}\right)\right] \underset{n\to\infty}{\longrightarrow} f(x).$$

b) On cherche à majorer la différence $|b_n(x) - f(x)|$ indépendamment de x.

$$\begin{aligned}
|b_n(x) - f(x)| &\leq \mathbb{E}\left[\left|f\left(\frac{X_1 + ... + X_n}{n}\right) - f(x)\right|\right] \\
&\leq \mathbb{E}\left[\left|f\left(\frac{X_1 + ... + X_n}{n}\right) - f(x)\right| \mathbb{1}_{\left|\frac{X_1+...+X_n}{n} - x\right| \geq \alpha}\right] \\
&\quad + \mathbb{E}\left[\left|f\left(\frac{X_1 + ... + X_n}{n}\right) - f(x)\right| \mathbb{1}_{\left|\frac{X_1+...+X_n}{n} - x\right| < \alpha}\right]
\end{aligned}$$

où l'inégalité ci-dessus est vraie pour tout α. On se fixe $\varepsilon > 0$. Par uniforme continuité de f (f est continue sur un segment), il existe un $\alpha > 0$ tel que $|y - x| \leq \alpha \Rightarrow |f(y) - f(x)| \leq \varepsilon/2$. On a donc pour ce choix de α,

$$|b_n(x) - f(x)| \leq 2||f||_\infty \underbrace{\mathbb{P}\left(\left|\frac{X_1 + ... + X_n}{n} - x\right| \geq \alpha\right)}_{A(x)} + \varepsilon/2$$

Il suffit maintenant de majorer $A(x)$ par un petit nombre indépendamment de x : pour cela il faut appliquer l'inégalité de Tchebychev. Pour tout $x \in [0,1]$, on a : $A(x) \leq \frac{1}{n\alpha^2} Var(X_1) = \frac{1}{n\alpha^2}(x * (1-x)) \leq \frac{1}{1 * n\alpha^2}$.
Donc en prenant n assez grand, on peut rendre aussi petit que l'on veut $\sup_x A(x)$ ce qui finit la preuve.

5. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de même loi de Poisson de paramètre 1. Soit $S_n = \sum_{k=1}^n X_k$. Rappeler la loi de S_n et calculer la limite de la suite $\left(e^{-n} \sum_{k=0}^n \frac{n^k}{k!}\right)_{n\geq 1}$.

Solution de l'exercice 5. Une somme de v.a. de Poisson indépendantes est une v.a. de Poisson dont le paramètre (qui est aussi l'espérance et la variance) est la somme de ceux des v.a. qu'on a ajoutées. Ainsi S_n suit une loi de Poisson de paramètre n. En particulier, si $k \in \mathbb{N}$, alors $e^{-n}\frac{n^k}{k!} = \mathbb{P}(S_n = k)$. En sommant, on obtient

$$\left(e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}\right) = \mathbb{P}(S_n \leq n) = \mathbb{P}(S_n/n \leq 1).$$

La loi des grands nombres nous dit que $S_n/n \to \mathbb{E}[X_1] = 1$ p.s. Ainsi S_n/n est proche de 1, mais ce résultat ne nous dit pas s'il est un peu plus grand ou un peu plus petit. Pour avoir des informations sur $S_n/n - 1$, (et en particulier son signe), on applique le théorème de la limite centrale :

$$\frac{S_n - n}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1).$$

D'après le corrolaire 4.2.1 du cours, on en déduit que, si Y suit une loi normale standard, alors, lorsque $n \to \infty$,

$$\mathbb{P}(S_n/n \leq 1) = \mathbb{P}(\frac{S_n - n}{\sqrt{n}} \leq 0) \to \mathbb{P}(Y \leq 0) = 1/2.$$

L'égalité $\mathbb{P}(Y \leq 0) = 1/2$ découle du fait que Y est symétrique.

Ainsi, on a montré que

$$\left(e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}\right) \to 1/2.$$

6.

a) Soit $(p_n)_{n\geq 0}$ une suite de réels dans $]0,1[$ telle que $\lim_{n\to+\infty} np_n = \lambda > 0$. Soit $(X_n)_{n\geq 0}$ une suite de v.a. telles que pour tout n : $X_n \sim \mathcal{B}(n,p_n)$ et X une v.a. de loi de Poisson paramètre λ. Montrer que $(X_n)_{n\geq 0}$ converge en loi vers X.

b) Soit $(X_n)_{n\geq 0}$ une suite de v.a. réelles indépendantes de même loi $\mathcal{N}(0,1)$. Etudier le comportement asymptotique en loi de la suite $Y_n = \frac{1}{n}\sum_{k=1}^{n}\sqrt{k}X_k$

Solution de l'exercice 6.

a) Afin de montrer la convergence en loi de $(X_n)_{n\geq 0}$ vers X. Montrons que $\phi_{X_n}(t) \xrightarrow[n\to+\infty]{} \phi_X(t)$ pour tout $t \in \mathbb{R}$, où ϕ_X designe la fonction caractéristique d'une variable aléatoire X.

$$
\begin{aligned}
\phi_{X_n}(t) &= \left(1 - p_n + p_n e^i t\right)^n \\
&= \exp\left(n \ln(1 - p_n + p_n e^i t)\right) \\
&= \exp\left(-np_n(1 - e^i t) + np_n\varepsilon_n\right) \quad \text{où } \varepsilon_n \xrightarrow[n\to+\infty]{} 0 \\
&\sim_n \exp\left(-np_n(1 - e^i t)\right) \\
&\xrightarrow[n\to+\infty]{} e^{\lambda(e^{it}-1)} = \phi_X(t).
\end{aligned}
$$

b) Vu la forme de la variable de Y_n, on est tenté d'appliquer la loi des grands nombres. Cependant on ne peut pas l'appliquer car les variables $\sqrt{k}X_k$ ne sont pas identiquement distribuées. On va donc à nouveau utiliser les fonctions caractéristiques afin de démontrer un résultat de convergence en loi. Les variables aléatoires X_n sont indépendantes, donc Y_n est une variable gaussienne de moyenne 0 et de variance $\frac{n(n+1)}{2n^2}$ donc :

$$
\phi_{Y_n}(t) = e^{-\frac{n(n+1)}{2n^2}t^2} \xrightarrow[n\to+\infty]{} e^{-t^2/2}.
$$

On a donc montré que $(Y_n)_{n\geq 1}$ converge en loi vers une loi normale centrée réduite $\mathcal{N}(0, \frac{1}{2})$.

7. On suppose que l'intervalle de temps entre deux voitures successives à un passage à niveau (peu fréquenté) suit une loi exponentielle de moyenne 30 minutes. On suppose de plus qu'il y a indépendance entre les intervalles de temps séparant les instants de passage de voitures. Calculer (une valeur approchée) de la probabilité qu'il y ait plus de 50 voitures qui empruntent le passage à niveau une journée donnée.

Solution de l'exercice 7. Soit $(X_n)_{n\geq 0}$ une suite de v.a. i.i.d suivant la loi exponentielle de paramètre $1/30$. On a alors $\mathbb{E}[X_1] = 30$ et $Var(X_1) = 900$. X_n représente le temps (en minutes) qui sépare le passage de la $(n-1)$ième voiture de la $n-$ième voiture. On considère ensuite $S_n = X_1 + \dots X_n$. S_n donne l'instant de passage de la $n-$ième voiture. On veut déterminer

$$
\mathbb{P}[S_{50} \leq 24 \times 60].
$$

Or on a

$$
\mathbb{P}[S_{50} \leq 24 \times 60] = \mathbb{P}\left[\frac{S_{50} - 50 \times 30}{30 \times \sqrt{50}} \leq \frac{24 \times 60 - 50 \times 30}{30 \times \sqrt{50}}\right] \approx F\left(\frac{-2}{\sqrt{50}}\right),
$$

où F est la fonction de répartition de la loi normale centrée réduite, l'approximation étant donnée par le TCL. De plus $F\left(\frac{-2}{\sqrt{50}}\right) \approx F(-0.283) = 1 - F(0.283)$. En utilisant une table ou un logiciel on obtient $F(0.283) \approx 0.61$. On a donc $\mathbb{P}[S_{50} \leq 24 \times 60] \approx 0.39$.

8. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. Montrer qu'on a

$$\lim_{n\to+\infty} e^{-n} \sum_{k=0}^{\infty} f\left(\frac{k-n}{\sqrt{n}}\right) \frac{n^k}{k!} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{-\frac{t^2}{2}}\, dt.$$

Solution de l'exercice 8. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de loi de Poisson de paramètre 1. On a $\mathbb{E}[X_1] = 1$ et $\mathrm{Var}(X_1) = 1$.

D'une part, le théorème central limite assure que la suite $\left(\frac{X_1+\ldots+X_n-n}{\sqrt{n}}\right)_{n\geq 1}$ converge en loi vers la loi normale centrée réduite. Ainsi, puisque f est continue et bornée,

$$\lim_{n\to\infty} \mathbb{E}\left[f\left(\frac{X_1+\ldots+X_n-n}{\sqrt{n}}\right)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{-\frac{t^2}{2}}\, dt.$$

D'autre part, pour tout $n \geq 1$, la variable aléatoire $X_1+\ldots+X_n$ suit la loi de Poisson de paramètre n. Ainsi,

$$\mathbb{E}\left[f\left(\frac{X_1+\ldots+X_n-n}{\sqrt{n}}\right)\right] = e^{-n} \sum_{k=0}^{\infty} f\left(\frac{k-n}{\sqrt{n}}\right) \frac{n^k}{k!}.$$

En comparant ces deux égalités, on a la convergence voulue.

9. Soient $(X_n)_{n\geq 1}$ des variables aléatoires indépendantes identiquement distribuées et de carré intégrable. On suppose que leur loi a la propriété suivante : $X_1 + X_2$ a même loi que $\sqrt{2}X_1$.

a) Exprimer, pour tout $n \geq 0$, la loi de $X_1 + \ldots + X_{2^n}$ en fonction de celle de X_1.

b) Déterminer la loi commune des variables aléatoires $(X_n)_{n\geq 1}$.

Solution de l'exercice 9.

a) On va montrer par récurrence que $X_1+\ldots+X_{2^n}$ a même loi que $\sqrt{2}^n X_1$, pour tout $n \geq 1$. Pour $n = 1$, c'est l'hypothèse. Supposons le résultat démontré jusqu'au rang n et démontrons le au rang $n+1$. On écrit

$$X_1 + \ldots + X_{2^{n+1}} = (X_1 + \ldots + X_{2^n}) + (X_{2^n+1} + \ldots + X_{2^{n+1}}).$$

La v.a. $X_{2^n+1} + \ldots + X_{2^{n+1}}$ est indépendante de $X_1 + \ldots + X_{2^n}$ (car les X_n sont indépendantes et on utilise des paquets d'indices disjoints) et de même loi. Par l'hypothèse de récurrence, ces deux v.a. ont même loi que $\sqrt{2}^n X_1$. Mais alors leur somme a même loi que $\sqrt{2}^n X_1 + \sqrt{2}^n X_2 = \sqrt{2}^n (X_1 + X_2)$, qui a bien même loi (d'après l'hypothèse pour $n = 1$) que $\sqrt{2}^{n+1} X_1$. Ce qui achève la preuve par récurrence.

b) Comme X_1 est de carré intégrable, X_1 est intégrable. L'hypothèse entraine alors que $\mathbb{E}[X_1] + \mathbb{E}[X_1] = \sqrt{2}\mathbb{E}[X_1]$, et donc $\mathbb{E}[X_1] = 0$. Soit $\sigma^2 = \mathbb{E}[X_1^2]$. On applique le théorème de la limite centrale :

$$\frac{X_1+\ldots+X_n}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \sigma^2).$$

On a la même convergence pour les sous-suites, et en particulier :

$$\frac{X_1 + \ldots + X_{2^n}}{\sqrt{2^n}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \sigma^2).$$

Or, d'après la question précédente, le terme de gauche a même loi que X_1, qui suit donc la loi $\mathcal{N}(0, \sigma^2)$.