MAP 311 - Aléatoire

Leçon 5

2016-2017

Justifier “les lois empiriques du hasard”

Vision intuitive (fréquentiste) d’une probabilité.

- Soit A un événement pouvant se produire lors d’une certaine expérience aléatoire.
  Exemple : $A =$“obtenir Pile” lors du lancer d’une pièce.
- On répète n fois l’expérience aléatoire. On note $N(A)$ le nombre de fois où l’événement A se produit.
  La fréquence empirique d’apparition de A, $N(A)/n$, est aléatoire.
  Mais lorsque $n \to \infty$ le rapport $N(A)/n$ se stabilise autour d’une valeur limite déterministe $f(A)$.
- Exemple : $f(A) = 1/2$ dans le cas où A est l’événement “obtenir Pile”.
  Cette limite est notre **idée intuitive de la probabilité** de l’événement A.

Justifier “les lois empiriques du hasard”

- Dans cette vision intuitive :
  - si A est l'événement certain $A = \Omega$, alors $N(\Omega) = n$ et donc $f(\Omega) = 1$.
  - si les événements A et B sont incompatibles, $A \cap B = \emptyset$, on a $N(A \cup B) = N(A) + N(B)$ et donc $f(A \cup B) = f(A) + f(B)$.

• Modèle probabiliste : les deux axiomes d'une probabilité $\mathbb{P}$ sont empruntées aux propriétés de la notion intuitive de fréquence empirique d'apparition.

• Nous allons justifier a posteriori cette définition axiomatique d'une probabilité par la Loi des Grands Nombres. On va montrer que les fréquences empiriques “convergent” vers la probabilité :

$$\frac{N(A)}{n} \stackrel{n\to\infty}{\longrightarrow} \mathbb{P}(A)$$

Mathématiquement, nous nous poserons la question suivante :

- Soient $Y_1, \dots, Y_n$ des variables aléatoires indépendantes et de même loi.
- Les Y_i modélisent n résultats d'une même expérience (comme par exemple n lancers de pièces, $Y_i = \mathbf{1}_{A_i}$ où $A_i =$"obtenir Pile au $i$ème lancer").
- Quel est le comportement de la moyenne empirique

$$\frac{Y_1 + \cdots + Y_n}{n}$$

  quand n tend vers l'infini ?
  Jacques Bernoulli (1713 Ars Conjectandi) : jeu de Pile ou Face.
  Généralisations : Chebyshev, Kolmogorov (20ème siècle).
- Comment définir la limite de la fonction

$$\omega \mapsto \frac{Y_1(\omega) + \cdots + Y_n(\omega)}{n} \quad ?$$

Motivations

Question plus générale : Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires. Comment définir la convergence de la suite $(X_n)_{n\in\mathbb{N}}$?
Question cruciale pour analyser les modèles de manière asymptotique, les estimer et les simuler numériquement.

- Etude de grands échantillons en statistiques (sondages, etc.)
- Méthodes de Monte-Carlo : méthodes de calcul par approximations aléatoires.
- Stabilisation-moyennisation en temps grand d'un système dynamique stochastique (Débruitage).
- Comportement de grandes populations (biologie) ou de systèmes de particules en interaction (mécanique statistique, réseaux).
- Et bien d'autres applications.

Différents types de convergences

Grosse difficulté : plusieurs types de convergence, de nature différente.

- Les variables aléatoires $(X_n)_{n\in\mathbb{N}}$ et X sont définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.
  On décrit la proximité des variables aléatoires X_n et X : $|X_n - X| \to 0$.
  - $|X_n - X|$ petit pour tout ω ?
  - $|X_n - X|$ petit avec grande probabilité ?
  - $|X_n - X|$ petit en moyenne ?
- Les variables aléatoires $(X_n)_{n\in\mathbb{N}}$ et X peuvent être définies sur des espaces de probabilité différents.
  On décrit la proximité des lois des variables aléatoires $(X_n)_{n\in\mathbb{N}}$ et X.
  Par exemple : A-t-on la convergence de la suite des fonctions de répartition F_{X_n} vers F_X ?

- Ces différentes notions de convergence auront différentes utilisations pratiques : connaître l'asymptotique d'une suite de variables aléatoires, calculer une espérance, quantifier l'erreur dans une approximation, donner une fourchette d'erreur dans une estimation/un sondage.
- Ces notions de convergence ne seront pas équivalentes et pas toujours comparables.
- Ce sont des questions difficiles.
- Par exemple, si dans le jeu de Pile ou Face, on s'intéresse à la convergence de la fréquence d'apparition d'un Pile, quand le nombre de lancers tend vers l'infini, on ne peut pas espérer que la convergence vers $1/2$ ait lieu pour tout ω :
  En effet, si par exemple $\omega = PPP\cdots$, alors $N(A)/n = 1$.

1ère notion de convergence : la convergence presque sure

Définition

La suite $(X_n)_{n\in\mathbb{N}}$ *converge presque sûrement vers* X *si*

$$\mathbb{P}\Big(\lim_{n\to+\infty}|X_n - X| = 0\Big) = 1$$

L'événement où la convergence a lieu,

$$C = \{\omega\,,\ \lim_{n\to+\infty}|X_n(\omega) - X(\omega)| = 0\}$$

est de probabilité 1.

Exemple 1 : Soit U v.a. uniforme sur $[0,1]$ et $X_n = \mathbf{1}_{U<1/n}$. Alors $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers 0.
En effet, si ω est fixé, alors dès que $U(\omega) > 0$, (vrai avec probabilité 1), il existe n_0 (suffisamment grand) tel que $U(\omega) > 1/n_0$, et donc tel que $X_n(\omega) = 0\ \forall n \geq n_0$. Ici $C = \{U > 0\}$.

Exemple 2 : Soient $(Y_n)_{n\in\mathbb{N}^*}$ variables aléatoires de Bernoulli telles que $\mathbb{P}(Y_n = 1) = \frac{1}{n}$ et $\mathbb{P}(Y_n = 0) = 1 - \frac{1}{n}$. Alors

- Supposons que les Y_n soient indépendantes.
- $\mathbb{P}(Y_n = 1) = \frac{1}{n}$. La série de terme général $1/n$ est divergente.
- Si l'on pose $A_n = \{Y_n = 1\}$, alors $\sum_n \mathbb{P}(A_n) = +\infty$. Les A_n sont indépendants.
  Théorème de Borel-Cantelli $\Longrightarrow$ pour presque tout ω, une infinité de $Y_n(\omega)$ sont égaux à 1.
- La suite $(Y_n)_{n\in\mathbb{N}^*}$ ne peut pas converger presque sûrement vers 0, contrairement à l'exemple 1 où l'on avait aussi $\mathbb{P}(X_n = 1) = 1/n$ et $\mathbb{P}(X_n = 0) = 1 - 1/n$.
- Pour presque tout ω, une infinité de $Y_n(\omega)$ sont égaux à 1 et une infinité de $Y_n(\omega)$ sont égaux à 0, donc la suite ne peut pas converger presque sûrement.

Mais

Exemple 3 : Prenons $(Z_n)_{n\in\mathbb{N}^*}$ avec, pour $a > 1$,

$$\mathbb{P}(Z_n = 1) = \frac{1}{n^a} \text{ et } \mathbb{P}(Z_n = 0) = 1 - \frac{1}{n^a}$$

Soit $B_n = \{Z_n = 1\}$. Alors $\sum_n \mathbb{P}(B_n) < +\infty$.
Théorème de Borel-Cantelli $\Longrightarrow$ pour presque tout ω, un nombre fini de $Z_n(\omega)$ sont égaux à 1. Et donc $Z_n(\omega) = 0$ à partir d'un certain rang.
La suite $(Z_n)_{n\in\mathbb{N}^*}$ converge presque sûrement vers 0.

Conclusion : cette notion est délicate !

Quelques propriétés faciles

Proposition

- *Soit f une fonction continue. Si la suite $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers X, la suite $(f(X_n))_{n\in\mathbb{N}}$ converge p.s. vers $f(X)$.*
- *Si $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers X et si $(Y_n)_{n\in\mathbb{N}}$ converge presque sûrement vers Y, alors $((X_n, Y_n))_{n\in\mathbb{N}}$ converge p.s. vers (X, Y).*
- *Si $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers X et si $(Y_n)_{n\in\mathbb{N}}$ converge presque sûrement vers Y, alors $(X_n + Y_n)_{n\in\mathbb{N}}$ converge p.s. vers $X + Y$, $(X_n Y_n)_{n\in\mathbb{N}}$ converge p.s. vers XY, etc.*

Théorème de convergence dominée

Un lien entre convergence presque sure et convergence des espérances.

Proposition

(appelée aussi Théorème de Lebesgue): Si la suite $(X_n)_{n\in\mathbb{N}}$ *converge presque sûrement vers* X *et si*

$$\forall n, \quad |X_n| \leq Z$$

avec $Z \in L^1$*, alors* X_n *et* X *sont dans* L^1*, et*

$$\mathbb{E}\big(|X_n - X|\big) \overset{n\to+\infty}{\longrightarrow} 0$$

En particulier, $\mathbb{E}(X_n) \overset{n\to+\infty}{\longrightarrow} \mathbb{E}(X)$.

Attention, réciproque fausse : Dans l'Exemple 2, la suite $\mathbb{E}(|Y_n - 0|) = \mathbb{E}(Y_n) = 1/n$ tend vers 0 et Y_n est bornée par 1, mais $(Y_n)_{n\in\mathbb{N}}$ ne converge pas presque sûrement vers 0.

L'hypothèse de domination est cruciale.
Exemple 4 : Soit $(T_n)_n$ une suite de v.a. telle que

$$\mathbb{P}(T_n = n^2) = \frac{1}{n^{3/2}} \text{ et } \mathbb{P}(T_n = 0) = 1 - \frac{1}{n^{3/2}}$$

On a $\mathbb{P}(T_n \geq 1) = \frac{1}{n^{3/2}}$, la série de terme général $\frac{1}{n^{3/2}}$ converge, donc, pour presque tout ω, T_n est nulle à partir d'un certain rang d'après Borel-Cantelli. Ceci montre que $(T_n)_{n\in\mathbb{N}}$ converge p.s. vers 0.
Mais $\mathbb{E}(T_n) = \sqrt{n}$ ne converge pas.

2ème notion de convergence : la convergence en moyenne ou convergence en norme L^1

Dans le théorème de convergence dominée, nous avons mis en évidence la convergence suivante.

Définition

La suite $(X_n)_{n\in\mathbb{N}}$ converge en moyenne vers X (ou dans L^1) si

$$\mathbb{E}\big(|X_n - X|\big) \stackrel{n\to+\infty}{\longrightarrow} 0$$

Exemple 2 : $(Y_n)_{n\in\mathbb{N}}$ variables aléatoires de Bernoulli telles que $\mathbb{P}(Y_n = 1) = 1/n$ et $\mathbb{P}(Y_n = 0) = 1 - 1/n$.
$(Y_n)_{n\in\mathbb{N}}$ tend en moyenne vers $Y = 0$:

$$\mathbb{E}(|Y_n - Y|) = \mathbb{E}(Y_n) = \frac{1}{n} \stackrel{n\to+\infty}{\longrightarrow} 0$$

Rappel : $(Y_n)_{n\in\mathbb{N}}$ ne converge pas presque sûrement !

Remarques

Considérons des variables aléatoires $X_n \in L^1$ et $X \in L^1$.

- Théorème de convergence dominée : si les variables aléatoires X_n sont dominées par une variable aléatoire intégrable, alors si la suite $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers X, elle converge en moyenne vers X.
- Si $X_n \in L^2$ et $X \in L^2$ et si $\mathbb{E}(|X_n - X|^2) \overset{n\to+\infty}{\longrightarrow} 0$, alors la suite $(X_n)_{n\in\mathbb{N}}$ converge en norme L^1 vers X.
  (On applique l'inégalité de Cauchy-Schwarz).
- Soit $\varepsilon > 0$. Supposons que $(X_n)_{n\in\mathbb{N}}$ converge en norme L^1 vers X, alors par l'inégalité de Markov,

$$\mathbb{P}\big(|X_n - X| \geq \varepsilon\big) \leq \frac{\mathbb{E}(|X_n - X|)}{\varepsilon} \overset{n\to+\infty}{\longrightarrow} 0$$

  Cela va nous permettre de définir une notion de convergence plus faible.

3ème notion de convergence : la convergence en probabilité

Définition

La suite $(X_n)_{n\in\mathbb{N}}$ *converge en probabilité vers* X *si*

$$\mathbb{P}\big(|X_n - X| \geq \varepsilon\big) \overset{n\to+\infty}{\longrightarrow} 0$$

pour tout $\varepsilon > 0$.

Exemple 2 : $(Y_n)_{n\in\mathbb{N}}$ v.a. de Bernoulli telles que

$$\mathbb{P}(Y_n = 1) = \frac{1}{n} \text{ et } \mathbb{P}(Y_n = 0) = 1 - \frac{1}{n}$$

$(Y_n)_{n\in\mathbb{N}}$ tend en probabilité vers $Y = 0$. Pour tout $\varepsilon > 0$:

$$\mathbb{P}\big(|Y_n| \geq \varepsilon\big) = \frac{1}{n} \overset{n\to+\infty}{\longrightarrow} 0$$

On a donc (par l'inégalité de Markov) :

Proposition

Supposons que les X_n et X soient intégrables. Alors

$$\text{Convergence en moyenne} \Longrightarrow \text{convergence en probabilité.}$$

La réciproque est fausse.

Exemple 5 : $(Y_n)_{n\in\mathbb{N}}$ variables aléatoires telles que

$$\mathbb{P}(Y_n = n^2) = \frac{1}{n} \text{ et } \mathbb{P}(Y_n = 0) = 1 - \frac{1}{n}$$

- $(Y_n)_{n\in\mathbb{N}}$ tend en probabilité vers 0 :

$$\mathbb{P}(|Y_n| \geq \varepsilon) = \mathbb{P}(Y_n^{1/4} \geq \varepsilon^{1/4}) \leq \frac{\mathbb{E}(Y_n^{1/4})}{\varepsilon^{1/4}} = \frac{1}{n^{1/2}\varepsilon^{1/4}} \overset{n\to+\infty}{\longrightarrow} 0$$

- $\mathbb{E}(Y_n) = n$: la suite $(Y_n)_{n\in\mathbb{N}}$ ne tend pas en moyenne vers 0.

Exercice : Si il existe une constante $a > 0$ telle que $\mathbb{P}(|X_n| \leq a) = 1$, alors la convergence en probabilité entraîne la convergence en moyenne.

Proposition

Convergence presque sure $\Longrightarrow$ convergence en probabilité.

Preuve : Soit $(X_n)_{n\in\mathbb{N}}$ une suite qui converge p.s. vers X. Soit $\varepsilon > 0$.

$$\mathbb{P}\big(|X_n - X| \geq \varepsilon\big) = \mathbb{E}\big(\mathbf{1}_{|X_n - X| \geq \varepsilon}\big)$$

La suite $(\mathbf{1}_{|X_n - X| \geq \varepsilon})_{n\in\mathbb{N}}$ converge p.s. vers 0, et est dominée. Donc elle converge dans L^1.

La réciproque est fausse. Voir Exemple 2 avec $(Y_n)_{n\in\mathbb{N}}$ indépendantes.

Nous avons une réciproque partielle.

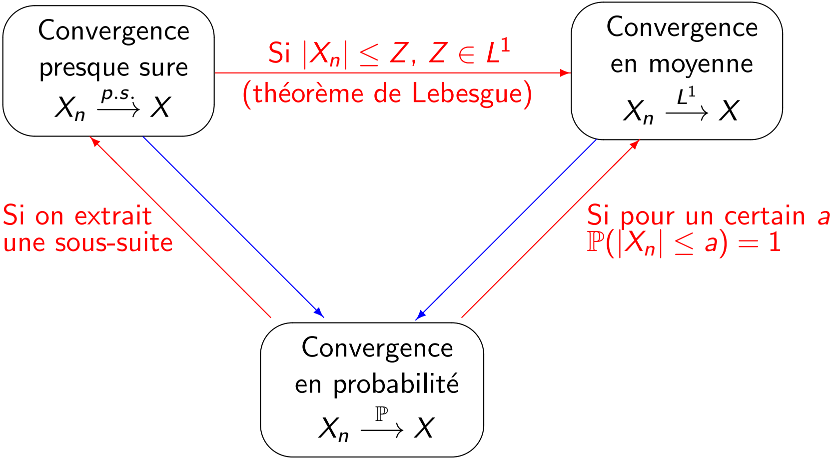
Proposition

Si $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers X, alors il existe une sous-suite $(X_{\sigma(n)})_{n\in\mathbb{N}}$ qui converge presque sûrement vers X.

Preuve : Nous allons construire la sous-suite de manière itérative.
Choisissons $\sigma(n) = i > \sigma(n-1)$ tel que $\mathbb{P}(|X_i - X| \geq 1/n) \leq 1/n^2$.
Alors $A_n = \{|X_{\sigma(n)} - X| \geq 1/n\}$ vérifie $\sum_n \mathbb{P}(A_n) < +\infty$.
Nous pouvons appliquer le théorème de Borel-Cantelli.
Avec probabilité 1, il n'y a qu'un nombre fini d'événements A_n réalisés.
Cela entraîne le résultat (presque sûrement, à partir d'un certain rang, on a $|X_{\sigma(n)} - X| < 1/n$).

Relations entre modes de convergences

Soient $(X_n)_{n\in\mathbb{N}}$ et X définies sur le même espace de probabilité.



Loi des Grands Nombres

Convergence de

$$M_n = \frac{X_1 + \cdots + X_n}{n}$$

Deux types de résultats :

- Loi faible des grands nombres : convergence en probabilité. Facile à obtenir, résultat peu informatif (Contrôle d'erreur, estimation de la limite).
- Loi forte des grands nombres : convergence presque sure, résultat plus fort, preuve plus délicate (Presque une convergence ponctuelle).
- Ces théorèmes sont vrais pour une suite de variables aléatoires indépendantes.

Loi faible des Grands Nombres

Théorème

Soit une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes, de même loi, intégrables. On pose $\mathbb{E}(X_1) = m$. Alors

$$M_n = \frac{X_1 + \cdots + X_n}{n} \stackrel{n\to+\infty}{\longrightarrow} m$$

en probabilité et en moyenne.

Preuve dans le cas où $\mathbb{E}(X_1^2) < +\infty$. Notons $\sigma^2 = \mathrm{Var}(X_1)$.
Comme les X_i sont indépendantes et de même loi, de moyenne m et de variance σ^2,

$$\mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i\ ) = m$$

$$\mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big) = \frac{1}{n^2}\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{\sigma^2}{n}$$

Alors par l'inégalité de Bienaymé-Chebyshev,

$$\mathbb{P}\big(|M_n - m| \geq \varepsilon\big) \leq \frac{\mathrm{Var}(M_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \overset{n\to+\infty}{\longrightarrow} 0$$

On a aussi

$$\mathbb{E}\big(|M_n - m|\big) \leq \sqrt{\mathbb{E}\big((M_n - m)^2\big)} = \frac{\sigma}{\sqrt{n}}$$

Loi forte des Grands Nombres

Théorème

Soit une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes, de même loi, intégrables. On pose $\mathbb{E}(X_1) = m$. Alors

$$M_n = \frac{X_1 + \cdots + X_n}{n} \overset{n\to+\infty}{\longrightarrow} m \text{ presque sûrement.}$$

Application : Justification a posteriori de l'approche fréquentiste.
n lancers de pièces indépendants, $A_n =$"obtenir Pile au $n$ème lancer". On pose $X_n = \mathbf{1}_{A_n}$. Alors $\frac{X_1+\cdots+X_n}{n}$ =fréquence empirique d'apparition de Pile. La Loi de Grands Nombres affirme que $\frac{X_1+\cdots+X_n}{n} \to 1/2$ p.s..

Exercice : Si les variables aléatoires X_n sont positives et $\mathbb{E}(X_n) = +\infty$. Alors

$$M_n = \frac{X_1 + \cdots + X_n}{n} \stackrel{n \to +\infty}{\longrightarrow} +\infty$$

presque sûrement.
Indication : considérer $\min(X_n, L)$ pour $L > 0$, puis faire tendre L vers l'infini.

Application "statistique"

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles indépendantes et de même fonction de répartition F. Fixons $x \in \mathbb{R}$ et considérons $Z_i = \mathbf{1}_{X_i \le x}$. Les variables aléatoires Z_i sont indépendantes et de même loi (de Bernoulli), et

$$\mathbb{E}(Z_i) = \mathbb{P}(X_i \le x) = F(x)$$

D'après la Loi des Grands Nombres

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \le x} \stackrel{n\to+\infty}{\longrightarrow} F(x) \text{ p.s.}$$

Corollaire

$$\frac{1}{n}\mathrm{Card}\big(\{i = 1, \ldots, n\,,\ a < X_i \le b\}\big) = F_n(b) - F_n(a) \stackrel{n\to+\infty}{\longrightarrow} F(b) - F(a)$$

Attention ! le terme de gauche est aléatoire.
Cela justifie l'histogramme que l'on construit pour simuler la densité de X.

Application en analyse : Polynômes de Bernstein

Soit $(X_n)_{n\in\mathbb{N}}$ suite de v.a. de Bernoulli indépendantes de paramètre $x \in [0,1]$.
Soit f une fonction continue sur $[0,1]$. Alors la Loi des Grands Nombres dit que $\frac{X_1+\dots+X_n}{n} \overset{n\to+\infty}{\longrightarrow} x$ p.s. et donc

$$f\Big(\frac{X_1+\dots+X_n}{n}\Big) \overset{n\to+\infty}{\longrightarrow} f(x) \text{ p.s.}$$

Théorème de convergence dominée donne

$$\mathbb{E}_x\Big(f\Big(\frac{X_1+\dots+X_n}{n}\Big)\Big) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big)\binom{n}{k} x^k(1-x)^{n-k} \overset{n\to+\infty}{\longrightarrow} f(x)$$

(on écrit $\mathbb{E}_x$ pour se souvenir que X_i est de loi de Bernoulli de paramètre x)
Donc f est limite simple de polynômes.
La convergence est-elle uniforme ?

f uniformément continue sur $[0,1]$: $\forall \varepsilon > 0$, $\exists \alpha$ tel que

$$\forall x, y \in [0,1],\ |x-y| < \alpha \Longrightarrow |f(x) - f(y)| < \varepsilon$$

Pour tout $\varepsilon > 0$, nous avons

$$\begin{aligned}
&\big|\mathbb{E}_x\big(f(M_n)\big) - f(x)\big| \\
&\leq \mathbb{E}_x\big(|f(M_n) - f(x)|\mathbf{1}_{|M_n - x| \geq \alpha}\big) + \mathbb{E}_x\big(|f(M_n) - f(x)|\mathbf{1}_{|M_n - x| < \alpha}\big) \\
&\leq 2\|f\|_\infty \mathbb{P}\big(|M_n - x| \geq \alpha\big) + \varepsilon \mathbb{P}\big(|M_n - x| < \alpha\big) \\
&\leq 2\|f\|_\infty \frac{\mathrm{Var}(X_1)}{n\alpha^2} + \varepsilon \\
&\leq \|f\|_\infty \frac{1}{2n\alpha^2} + \varepsilon
\end{aligned}$$

car $\mathrm{Var}(X_1) = x(1-x) \leq 1/4$.
Donc

$$\sup_{x \in [0,1]} \left| \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} - f(x) \right| \overset{n \to +\infty}{\longrightarrow} 0$$

On a donc prouvé : Toute fonction continue sur $[0,1]$ est approchée uniformément par une suite de polynômes (théorème de Weierstrass).

Application : Filtrage d'un signal périodique bruité

Signal pur (caché) :

$$s(t), \quad t \geq 0$$

est une fonction périodique de période $T = 1$.
Signal bruité (en haut dans les simulations) : $b(t) = s(t) + B(t)$, où $B(t)$ variables indépendantes et de même loi, centrées.
Signal filtré (en bas dans les simulations) :

$$f_k(t) = \frac{b(t) + b(t-T) + \cdots + b(t-kT)}{k+1}$$

si $k = 0, 1, \ldots$ est tel que $kT \leq t < (k+1)T$.
Comme

$$f_k(t) = s(t) + \frac{B(t) + \cdots + B(t-kT)}{k+1}$$

la Loi Forte des Grands Nombres entraîne que $f_k(t) - s(t) \to 0$ lorsque $k \to +\infty$ avec probabilité 1.

Corrigé de l'exercice (leçon 4)

Soient deux variables aléatoires indépendantes X de loi $\Gamma(a,\theta)$ et Y de loi $\Gamma(b,\theta)$. Cherchons la loi du couple $(X+Y, X/(X+Y))$.
Soient $S = X+Y$ et $T = X/(X+Y)$. Posons $\boldsymbol{g}(x,y) = (x+y, x/(x+y))$. Nous avons

$$\begin{cases} s = x+y \\ t = \frac{x}{x+y} \\ x > 0,\ y > 0 \end{cases} \iff \begin{cases} x = st \\ y = s(1-t) \\ s > 0,\ 0 < t < 1 \end{cases}$$

On calcule
$$J(\boldsymbol{g}^{-1})(s,t) = \mathrm{Det}\begin{pmatrix} \partial_s st & \partial_s s(1-t) \\ \partial_t st & \partial_t s(1-t) \end{pmatrix} = \mathrm{Det}\begin{pmatrix} t & (1-t) \\ s & -s \end{pmatrix} = -s.$$

Soit $h \in \mathcal{C}_b(\mathbb{R}^2)$. Grâce à l'indépendance de X et Y, nous avons alors

$$\mathbb{E}\big(h(S,T)\big) = \int_0^\infty \int_0^1 h(s,t) \frac{\theta^{a+b}}{\Gamma(a)\Gamma(b)} s^{a+b-1} e^{-\theta s} t^{a-1} (1-t)^{b-1} ds dt$$

Nous reconnaissons un produit d'une fonction de la variable s et d'une fonction de la variable t.

Cela entraîne que les variables aléatoires S et T sont indépendantes. De plus nous reconnaissons pour S la loi $\Gamma(a+b,\theta)$. Nécessairement, par le jeu des constantes, la loi de T a la densité

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1}\mathbf{1}_{[0,1[}(t)$$

Cette loi s'appelle la loi Beta de paramètres a et b. Elle charge uniquement l'intervalle $]0,1[$.

Retour sur : Espérance conditionnelle et régression linéaire

Soit $(\boldsymbol{X}, Y)$ un vecteur aléatoire, $\boldsymbol{X}$ à valeurs dans $\mathbb{R}^n$ et Y à valeurs dans $\mathbb{R}$.

• L'espérance de Y est la constante qui approche au mieux la variable aléatoire Y :

$$\mathbb{E}(Y) = \underset{a\in\mathbb{R}}{\operatorname{argmin}}\ \mathbb{E}\big((Y-a)^2\big)$$

• Espérance conditionnelle.

On observe $\boldsymbol{X}$. Quelle est la meilleure approximation de Y sachant $\boldsymbol{X}$?

On cherche donc à résoudre :

$$\psi_0 = \underset{\psi:\mathbb{R}^n\to\mathbb{R}}{\operatorname{argmin}}\ \mathbb{E}\big((Y-\psi(\boldsymbol{X}))^2\big)$$

$\hookrightarrow$ Problème de minimisation (infini-dimensionnelle) parfois complexe.

Résultat : $\psi_0(\boldsymbol{x}) = \mathbb{E}[Y|\boldsymbol{X}=\boldsymbol{x}]$.

Preuve de : le ψ optimal est $\psi_0(\boldsymbol{x}) = \mathbb{E}\big(Y|\boldsymbol{X} = \boldsymbol{x}\big)$ dans le cas où $(\boldsymbol{X}, Y)$ est à densité.
En notant $\tilde{\psi}(\boldsymbol{x}) = \psi(\boldsymbol{x}) - \psi_0(\boldsymbol{x})$, on a

$$\begin{array}{rcl} \mathbb{E}\big((Y-\psi(\boldsymbol{X}))^2\big) & = & \mathbb{E}\big((Y-\psi_0(\boldsymbol{X}))^2\big) - 2\mathbb{E}\big(\tilde{\psi}(\boldsymbol{X})(Y-\psi_0(\boldsymbol{X}))\big) \\ & & +\mathbb{E}\big(\tilde{\psi}(\boldsymbol{X})^2\big) \end{array}$$

Or

$$\begin{array}{rcl} \mathbb{E}\big(\tilde{\psi}(\boldsymbol{X})\psi_0(\boldsymbol{X})\big) & = & \displaystyle\int_{\mathbb{R}^n} \tilde{\psi}(\boldsymbol{x})\psi_0(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} \\ & = & \displaystyle\int_{\mathbb{R}^n} \tilde{\psi}(\boldsymbol{x}) \Big(\int_{\mathbb{R}} y f_{Y|\boldsymbol{X}=\boldsymbol{x}}(y) dy \Big) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} \\ & = & \displaystyle\iint_{\mathbb{R}^n \times \mathbb{R}} \tilde{\psi}(\boldsymbol{x}) y f_{\boldsymbol{X},Y}(\boldsymbol{x}, y) d\boldsymbol{x} dy = \mathbb{E}\big(\tilde{\psi}(\boldsymbol{X})Y\big) \end{array}$$

Donc

$$\mathbb{E}\big((Y-\psi(\boldsymbol{X}))^2\big) = \mathbb{E}\big((Y-\psi_0(\boldsymbol{X}))^2\big) + \mathbb{E}\big(\tilde{\psi}(\boldsymbol{X})^2\big)$$

dont le minimum (en $\tilde{\psi}$) est atteint pour $\tilde{\psi} = 0$.

• Régression linéaire.

On observe $\boldsymbol{X}$. Quelle est la meilleure combinaison affine de $\boldsymbol{X}$ qui approche au mieux Y ?

On cherche donc à résoudre :

$$(\alpha_j^{reg})_{j=0}^n = \underset{(\alpha_j)_{j=0}^n \in \mathbb{R}^{n+1}}{\operatorname{argmin}} \ \mathbb{E}\Big[\Big(Y - \alpha_0 - \sum_{j=1}^n \alpha_j X_j\Big)^2\Big]$$

et on obtient $Y^{reg} = \alpha_0^{reg} + \sum_{j=1}^n \alpha_j^{reg} X_j$.

$\hookrightarrow$ Problème de minimisation (fini-dimensionnelle, quadratique) beaucoup plus simple que celui correspondant à l'espérance conditionnelle.

Exercice : Déterminer $(\alpha_j)_{j=0}^n$ en fonction de $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathrm{Cov}(X_j, Y)$, $\mathrm{Cov}(X_i, X_j)$.

Mais la régression est sous-optimale du point de vue de l'approximation : la meilleure combinaison affine de $\boldsymbol{X}$ n'est pas forcément la meilleure approximation de Y sachant $\boldsymbol{X}$. Sauf que :

• Résultat : Si $(\boldsymbol{X}, Y)$ est gaussien, alors l'espérance conditionnelle $\mathbb{E}[Y|\boldsymbol{X}]$ est la régression linéaire de Y sur $\boldsymbol{X}$.