MAP 311 - Aléatoire

Leçon 6

2016-2017

Méthode de Monte-Carlo

Application de la loi des grands nombres.

Proposition

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes de loi uniforme sur $[a,b]$, $a<b$. Soit f une fonction intégrable sur $[a,b]$. Alors

$$\frac{1}{b-a}\int_a^b f(x)dx = \lim_{n\to+\infty}\frac{f(X_1)+\cdots+f(X_n)}{n} \quad p.s.$$

Preuve : On applique la loi forte des grands nombres aux variables aléatoires $f(X_i)$. Or

$$\mathbb{E}\big(f(X_1)\big) = \frac{1}{b-a}\int_a^b f(x)dx.$$

D'où le résultat.

Généralisation

On veut calculer

$$I = \int_A f(\boldsymbol{z}) d\boldsymbol{z}$$

où f est intégrable sur $A \subset \mathbb{R}^d$.
On écrit

$$f(\boldsymbol{z}) = g(\boldsymbol{z}) p(\boldsymbol{z})$$

avec p densité de probabilité de support A.
Considérons des vecteurs aléatoires $\boldsymbol{Z}_i$ indépendants de densité p. Alors

$$\mathbb{E}\big(g(\boldsymbol{Z}_1)\big) = \int_A g(\boldsymbol{z}) p(\boldsymbol{z}) d\boldsymbol{z} = \int_A f(\boldsymbol{z}) d\boldsymbol{z} = I$$

D'après la Loi des Grands Nombres, une valeur approchée de I est

$$I_n = \frac{g(\boldsymbol{Z}_1) + \cdots + g(\boldsymbol{Z}_n)}{n}$$

Pour obtenir I_n : on simule les vecteurs aléatoires $\boldsymbol{Z}_i$, on évalue $g(\boldsymbol{Z}_i)$.
Remarque : Le choix de p est lié à la facilité de simulation des $\boldsymbol{Z}_i$.

Commentaires sur la méthode :

- Inconvénient : I_n est aléatoire, donc il est difficile a priori de contrôler l'erreur $I_n - I$. C'est possible grâce au théorème de la limite centrale (cf. fin du cours).
- Avantage : la fonction f peut-être irrégulière, contrairement aux hypothèses des méthodes déterministes (quadrature).
- Avantage : la vitesse de convergence ne dépend pas de la dimension (cf. fin du cours), ce qui n'est pas le cas des méthodes déterministes.

Pour conclure : Les algorithmes obtenus par méthode de Monte-Carlo sont extrêmement utilisés si on veut un temps de calcul court et en grandes dimensions.

Objectif : calculer

$$V_d = \iint_{[-1,1]^d} \mathbf{1}_{x_1^2+\cdots+x_d^2\leq 1} dx_1 \cdots dx_d$$

Ratio volume sphère/cube :

$$R_d = \frac{V_d}{2^d} = \iint_{[-1,1]^d} \mathbf{1}_{x_1^2+\cdots+x_d^2\leq 1} \frac{dx_1}{2} \cdots \frac{dx_d}{2}$$

Exemple : $R_2 = \frac{\pi}{4} \simeq 0{,}785$, $R_3 = \frac{\pi}{6} \simeq 0{,}524$, $R_4 = \frac{\pi^2}{32} \simeq 0{,}308$, etc.
Loi des Grands Nombres : Si $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ suite de v.a. indépendantes de loi uniforme sur $[-1,1]^d$, alors

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\|\boldsymbol{X}_i\|\leq 1} \overset{n\to+\infty}{\longrightarrow} R_d = \mathbb{E}\big(\mathbf{1}_{\|\boldsymbol{X}_1\|\leq 1}\big) \text{ p.s.}$$

Considérons l'erreur renormalisée :

$$E_n = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\|\boldsymbol{X}_i\| \le 1} - R_d\Big)$$

On a $\mathbb{E}[E_n] = 0$, $\mathbb{E}[E_n^2] = \mathrm{Var}\big(\mathbf{1}_{\|\boldsymbol{X}_1\| \le 1}\big)$.
Question : A-t-on convergence presque sure ou en probabilité de

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\|\boldsymbol{X}_i\| \le 1} - R_d\Big) \ ?$$

NON : Nous allons avoir une convergence beaucoup plus faible.

La “proximité des lois” de variables aléatoires

Exemple (cas discret) : Si X_n suit une loi binomiale $\mathcal{B}(n, a_n)$ avec $a_n = \theta/n$, alors pour tout $j \in \mathbb{N}$,

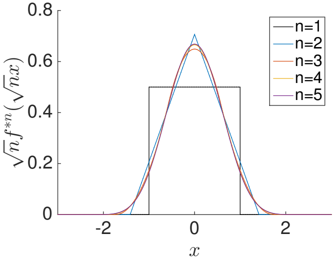
$$\mathbb{P}(X_n = j) \overset{n\to+\infty}{\longrightarrow} \mathbb{P}(X = j)$$

où X suit une loi de Poisson de paramètre θ.
Preuve :

$$\begin{aligned}
\mathbb{P}(X_n = j) &= \binom{n}{j} a_n^j (1-a_n)^{n-j} = \frac{1}{j!}\frac{n!}{(n-j)!}\left(\frac{\theta}{n}\right)^j\left(1-\frac{\theta}{n}\right)^{n-j} \\
&= \frac{1}{j!}\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-j+1}{n}\theta^j\left(1-\frac{\theta}{n}\right)^{n-j} \overset{n\to+\infty}{\longrightarrow} \frac{1}{j!}\theta^j e^{-\theta}
\end{aligned}$$

La "proximité des lois" de variables aléatoires

Exemple (cas à densité) :
Soient $(U_i)_{i\geq 1}$ indépendantes de loi uniforme sur $[-1,1]$, de densité $f(x) = \frac{1}{2}\mathbf{1}_{[-1,1]}(x)$.
La densité de $U_1 + U_2$ est $f^{*2}(x) = f * f(x) = \frac{1}{4}(2 - |x|)\mathbf{1}_{[-2,2]}(x)$.
La densité de $\sum_{i=1}^{n} U_i$ est $f^{*n}(x)$.
La densité de $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i$ est $\sqrt{n}f^{*n}(\sqrt{n}x)$. Elle se rapproche très vite de la densité d'une loi normale :



Dans ces exemples, ce sont les probabilités/les lois qui convergent, on ne s'intéresse pas aux variables aléatoires elles-mêmes.

Une nouvelle notion de convergence : la convergence en Loi

Définition

La suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X, et on écrit $X_n \xrightarrow{\mathcal{L}} X$, si pour toute fonction continue bornée f,

$$\mathbb{E}(f(X_n)) \overset{n\to+\infty}{\longrightarrow} \mathbb{E}(f(X))$$

Remarque : on ne demande pas aux v.a. d'être définies sur le même espace de probabilité.

Exemple. Un cas discret : si les v.a. X_n et X prennent un nombre fini de valeurs $\{a_k, 1 \leq k \leq K\}$, alors la suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X ssi $\forall k \in \{1, \ldots, K\}$,

$$\lim_{n\to+\infty} \mathbb{P}(X_n = a_k) = \mathbb{P}(X = a_k)$$

Exemple : Un cas à densité. Soit $(X_n)_{n\in\mathbb{N}}$ et X des variables aléatoires de lois respectives $\mathcal{N}(0,\sigma_n^2)$ et $\mathcal{N}(0,\sigma^2)$, où $\sigma_n \to \sigma \in]0,+\infty[$ quand $n\to\infty$, alors la suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X.
Preuve : Soit f continue bornée sur $\mathbb{R}$.

$$\begin{aligned}
\mathbb{E}\big(f(X_n)\big) &= \int_{\mathbb{R}} f(x)\frac{1}{\sqrt{2\pi}\sigma_n}\exp\Big(-\frac{x^2}{2\sigma_n^2}\Big)dx \\
&\overset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}} f(x)\frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{x^2}{2\sigma^2}\Big)dx = \mathbb{E}\big(f(X)\big)
\end{aligned}$$

par convergence dominée.

La convergence en loi est très faible !

Exemple : Un cas à densité. Soit X une v.a. gaussienne centrée réduite.

On pose pour tout n : $X_n = -X$. Alors $X_n \xrightarrow{\mathcal{L}} X$!

Preuve : Soit f continue bornée sur $\mathbb{R}$.

$$\begin{aligned}\mathbb{E}\big(f(X_n)\big) &= \mathbb{E}\big(f(-X)\big) = \int_{\mathbb{R}} f(-x)\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{x^2}{2}\Big)dx \\ &= \int_{\mathbb{R}} f(x)\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{x^2}{2}\Big)dx = \mathbb{E}\big(f(X)\big)\end{aligned}$$

Proposition

Si $X_n \xrightarrow{\mathbb{P}} X$ quand $n \to \infty$, alors $X_n \xrightarrow{\mathcal{L}} X$.

Preuve : Soit f fonction continue bornée. Soient $\varepsilon > 0$ et $N > 0$.
f est uniformément continue sur $[-N, N]$: Il existe $\alpha > 0$ tel que $|x - y| \leq \alpha,\ |x| \leq N \Longrightarrow |f(x) - f(y)| \leq \varepsilon$.
En notant $\Delta_n = f(X_n) - f(X)$:

$$\mathbb{E}(\Delta_n) = \mathbb{E}(\Delta_n \mathbf{1}_{|X_n|>N}) + \mathbb{E}(\Delta_n \mathbf{1}_{|X_n|\leq N, |X_n - X|\leq \alpha}) + \mathbb{E}(\Delta_n \mathbf{1}_{|X_n|\leq N, |X_n - X|>\alpha})$$

Donc

$$|\mathbb{E}(\Delta_n)| \leq 2\|f\|_\infty \mathbb{P}(|X_n| > N) + \varepsilon + 2\|f\|_\infty \mathbb{P}(|X_n - X| > \alpha)$$

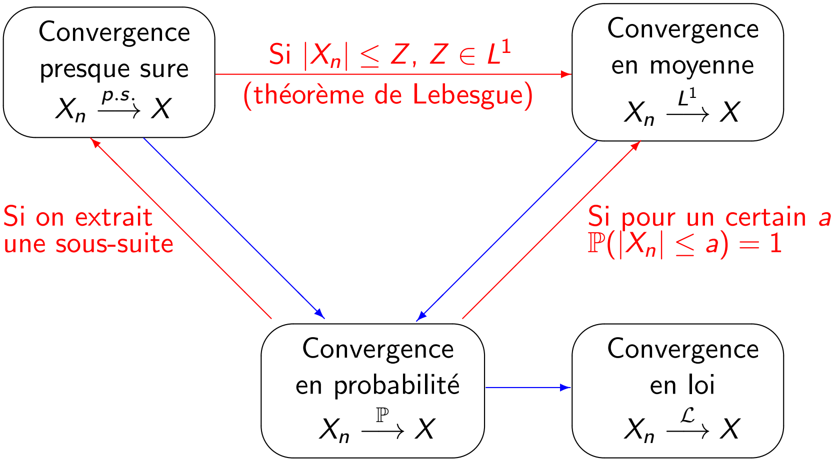
Or $\mathbb{P}(|X_n| > N) \leq \mathbb{P}(|X| > N - 1) + \mathbb{P}(|X_n - X| > 1)$, donc

$$\limsup_{n\to\infty} |\mathbb{E}(\Delta_n)| \leq 2\|f\|_\infty \mathbb{P}(|X| > N - 1) + \varepsilon$$

Comme c'est vrai pour tous $\varepsilon > 0$ et $N > 0$, on obtient le résultat.

Relations entre modes de convergences

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



Théorème

(Théorème de Slutsky) Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires réelles (avec X_n et Y_n définies sur le même espace de probabilité) telles que

$$X_n \xrightarrow{\mathcal{L}} X \text{ et } Y_n \xrightarrow{\mathbb{P}} c$$

où c est une constante. Alors

$$(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$$

En particulier, $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$ et $X_n Y_n \xrightarrow{\mathcal{L}} cX$.

Très utile en statistique !

Attention : Même en supposant que toutes les v.a. sont définies sur le même espace de probabilité, si

$$X_n \xrightarrow{\mathcal{L}} X \text{ et } Y_n \xrightarrow{\mathcal{L}} Y$$

alors on n'a pas forcément

$$(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$$

Exemple : Si $X_n = Y_n = X$ et $Y = -X$ où $X \sim \mathcal{N}(0,1)$, alors on a $X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{\mathcal{L}} Y$, $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, X)$ et non pas $(X, Y) = (X, -X)$.
Or la loi de (X, X) est différente de la loi de $(X, -X)$:

$$\mathbb{P}(X + X = 0) = 0 \neq 1 = \mathbb{P}(X - X = 0).$$

Proposition

Soient $(X_n)_{n\in\mathbb{N}}$ et X des variables aléatoires réelles de fonctions de répartition respectives F_n et F. Alors

$$X_n \xrightarrow{\mathcal{L}} X \iff F_n(t) \stackrel{n\to\infty}{\longrightarrow} F(t) \textit{ pour tout } t \textit{ en lequel } F \textit{ est continue}$$

Preuve : technique.

La fonction de répartition est un outil pour montrer la convergence en loi. Il y en a un autre.

Fonction caractéristique : Un outil pour étudier la loi et la convergence en loi

- Cadre général de vecteurs aléatoires $\boldsymbol{X}$ à valeurs dans $\mathbb{R}^d$.
- Une nouvelle fonction pour caractériser la loi.
- Fonction à valeurs complexes.
- Mathématiquement : transformée de Fourier.
- Notation : Pour $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{j=1}^d x_j y_j$.
- Si $\boldsymbol{X}$ vecteur aléatoire à valeurs dans $\mathbb{R}^d$, alors $e^{i\langle \boldsymbol{u}, \boldsymbol{X} \rangle} = \cos \langle \boldsymbol{u}, \boldsymbol{X} \rangle + i \sin \langle \boldsymbol{u}, \boldsymbol{X} \rangle$ variable aléatoire bornée (en module) par 1.

Définition

La fonction caractéristique de $\boldsymbol{X}$ est la fonction $\phi_{\boldsymbol{X}}$ définie de $\mathbb{R}^d$ dans $\mathbb{C}$ par

$$\phi_{\boldsymbol{X}}(\boldsymbol{u}) = \mathbb{E}\big(e^{i\langle \boldsymbol{u},\boldsymbol{X}\rangle}\big) = \mathbb{E}\big(\cos\langle \boldsymbol{u},\boldsymbol{X}\rangle\big) + i\mathbb{E}\big(\sin\langle \boldsymbol{u},\boldsymbol{X}\rangle\big)$$

- Si X est à valeurs réelles (cas où $d = 1$), alors $\forall u \in \mathbb{R}$,

$$\phi_X(u) = \mathbb{E}(e^{iuX})$$

- La fonction $\phi_{\boldsymbol{X}}$ ne dépend que de la loi de $\boldsymbol{X}$.
- Cas de variables aléatoires entières : Loi de $X = \{(k, p_k), k \in \mathbb{N}\}$. Alors pour tout $u \in \mathbb{R}$

$$\phi_X(u) = \sum_{k\in\mathbb{N}} p_k e^{iuk} = G_X(e^{iu})$$

où G_X =fonction génératrice de X.

- Cas d'une variable aléatoire réelle à densité f :

$$\forall u \in \mathbb{R}, \quad \phi_X(u) = \int_{\mathbb{R}} e^{iux} f(x) dx$$

$\hookrightarrow$ lien avec la transformée de Fourier de f.
Méthode générale de calcul dans le cas d'une densité : Théorème des résidus (analyse complexe).

Proposition

Soit $\boldsymbol{X}$ un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Alors la fonction $\phi_{\boldsymbol{X}}$ est continue, de module inférieur à 1, et

$$\phi_{\boldsymbol{X}}(\mathbf{0}) = 1, \quad \phi_{\boldsymbol{X}}(-\boldsymbol{u}) = \overline{\phi_{\boldsymbol{X}}(\boldsymbol{u})} \ \forall \boldsymbol{u} \in \mathbb{R}^d$$

Preuve de la continuité : Si $\boldsymbol{u}_p \to \boldsymbol{u}$, alors $e^{i\langle \boldsymbol{u}_p, \boldsymbol{X} \rangle}$ converge p.s. vers $e^{i\langle \boldsymbol{u}, \boldsymbol{X} \rangle}$ et ces variables sont bornées par 1 en module. On applique le théorème de convergence dominée pour conclure.

Fonctions caractéristiques des lois usuelles

- Si $X \sim \mathcal{E}(\lambda)$, alors
$$\phi_X(\lambda) = \frac{\lambda}{\lambda - iu}$$
Preuve : $\mathbb{E}(\exp(iuX)) = \int_0^\infty \lambda \exp(iux - \lambda x) dx$.
- Si $X \sim \mathcal{N}(0,1)$, alors
$$\phi_X(u) = \exp\left(-\frac{u^2}{2}\right)$$
- Si $X \sim \mathcal{N}(m, \sigma^2)$, alors
$$\phi_X(u) = \exp\left(ium - \frac{u^2\sigma^2}{2}\right)$$
Preuve : $X = m + \sigma Y$, où Y variable normale centrée réduite. On a $\phi_X(u) = \mathbb{E}(\exp(iuX)) = \mathbb{E}(\exp(ium + iu\sigma Y)) = \exp(ium)\phi_Y(u\sigma)$.

Fonction caractéristique de la loi normale

Comme X est de loi symétrique $\phi_X(u) = \overline{\phi_X(u)} \in \mathbb{R}$. Donc

$$\phi_X(u) = \mathbb{E}\big(\cos(uX)\big) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \cos(ux) \exp\Big(-\frac{x^2}{2}\Big) dx$$

Par le théorème de convergence dominée,

$$\begin{aligned}
\phi'_X(u) &= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \sin(ux) x \exp\Big(-\frac{x^2}{2}\Big) dx \\
&= -u \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \cos(ux) \exp\Big(-\frac{x^2}{2}\Big) dx
\end{aligned}$$

par intégration par parties. On a donc montré que pour tout $u \in \mathbb{R}$:

$$\phi'_X(u) = -u\phi_X(u)$$

Comme de plus $\phi_X(0) = 1$, on trouve que $\phi_X(u) = \exp(-u^2/2)$.

Propriété fondamentale

Proposition

La fonction caractéristique $\phi_{\boldsymbol{X}}$ caractérise la loi de la variable aléatoire $\boldsymbol{X}$: si deux variables aléatoires ont même fonction caractéristique, elles ont même loi.

Preuve : Propriété de la transformée de Fourier.

Corollaire

Soit $\boldsymbol{X} = (X_1, \ldots, X_n)$. Les composantes X_i sont indépendantes ssi pour tous $u_1, \ldots, u_n \in \mathbb{R}$, on a

$$\phi_{\boldsymbol{X}}((u_1, \ldots, u_n)) = \mathbb{E}\left(e^{i(u_1 X_1 + \cdots + u_n X_n)}\right) = \prod_{i=1}^{n} \mathbb{E}\left(e^{iu_j X_j}\right) = \prod_{i=1}^{n} \phi_{X_j}(u_j)$$

Intérêt : Condition nécessaire et suffisante pour caractériser l'indépendance de variables aléatoires.

Preuve :

1) Condition nécessaire : immédiat.

2) Condition suffisante : On peut construire des variables aléatoires X'_j, $j \in \{1, \dots, n\}$ indépendantes, telles que pour tout j, X_j et X'_j ont même loi.

On en déduit que $\phi_{X_j} = \phi_{X'_j}$.

En utilisant la condition nécessaire, on en déduit que

$$\phi_{\boldsymbol{X}'} = \prod_{j=1}^{n} \phi_{X'_j} = \prod_{j=1}^{n} \phi_{X_j} = \phi_{\boldsymbol{X}}$$

Ainsi les vecteurs aléatoires $\boldsymbol{X}$ et $\boldsymbol{X}'$ ont même loi.

Nous avons donc montré l'indépendance cherchée.

Exemple : Si $\boldsymbol{Y}$ est un vecteur de v.a. gaussiennes centrées réduites indépendantes, alors $(\boldsymbol{u} = (u_1, \dots, u_n))$

$$\phi_{\boldsymbol{Y}}(\boldsymbol{u}) = \prod_{j=1}^{n} \exp\Big(-\frac{u_j^2}{2} \Big) = \exp\Big(-\frac{\|\boldsymbol{u}\|^2}{2} \Big)$$

Proposition

Soit $\boldsymbol{Z}$ un vecteur aléatoire de moyenne $\boldsymbol{m}$ et de matrice de covariance définie positive $\mathbf{C}$. Si toutes les combinaisons linéaires $\sum_{i=1}^{n} b_j Z_j$ ont des lois normales, alors le vecteur $\boldsymbol{Z}$ est gaussien, i.e. sa densité est

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{C}}} \exp\Big(-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{m})^t \mathbf{C}^{-1}(\boldsymbol{z}-\boldsymbol{m})\Big)$$

Preuve : on considère la fonction caractéristique de $\boldsymbol{X} = \mathbf{C}^{-1/2}(\boldsymbol{Z}-\boldsymbol{m})$:

$$\phi_{\boldsymbol{X}}(\boldsymbol{u}) = \mathbb{E}\big(e^{i\langle \boldsymbol{u},\boldsymbol{X}\rangle}\big)$$

Comme $X_u := \langle \boldsymbol{u}, \boldsymbol{X}\rangle = \boldsymbol{u}^t\mathbf{C}^{-1/2}(\boldsymbol{Z}-\boldsymbol{m})$, elle suit une loi normale, de moyenne 0 et de variance $\boldsymbol{u}^t\mathbf{C}^{-1/2}\mathbf{C}\mathbf{C}^{-1/2}\boldsymbol{u} = \|\boldsymbol{u}\|^2$, on trouve

$$\phi_{\boldsymbol{X}}(\boldsymbol{u}) = \phi_{X_u}(1) = \exp\Big(-\frac{\|\boldsymbol{u}\|^2}{2}\Big)$$

On reconnaît la fonction caractéristique d'un vecteur gaussien de v.a. normales centrées réduites indépendantes, et $\boldsymbol{Z} = \mathbf{C}^{1/2}\boldsymbol{X} + \boldsymbol{m}$.

Fonction caractéristique d'un vecteur gaussien

Une proposition préalable :

Proposition

Si $\boldsymbol{X}$ est un vecteur aléatoire à valeurs dans $\mathbb{R}^n$, si $\boldsymbol{b} \in \mathbb{R}^l$ et $\mathbf{A}$ une matrice de taille $l \times n$, on a pour tout $\boldsymbol{u} \in \mathbb{R}^l$:

$$\phi_{\mathbf{A}\boldsymbol{X}+\boldsymbol{b}}(\boldsymbol{u}) = e^{i\langle \boldsymbol{b},\boldsymbol{u}\rangle}\phi_{\boldsymbol{X}}(\mathbf{A}^t\boldsymbol{u})$$

Preuve :

$$\mathbb{E}\left(e^{i\langle \boldsymbol{u},\mathbf{A}\boldsymbol{X}+\boldsymbol{b}\rangle}\right) = e^{i\langle \boldsymbol{b},\boldsymbol{u}\rangle}\mathbb{E}\left(e^{i\langle \boldsymbol{u},\mathbf{A}\boldsymbol{X}\rangle}\right) = e^{i\langle \boldsymbol{b},\boldsymbol{u}\rangle}\mathbb{E}\left(e^{i\langle \mathbf{A}^t\boldsymbol{u},\boldsymbol{X}\rangle}\right)$$

Théorème

$\boldsymbol{X}$ est un vecteur gaussien à valeurs dans $\mathbb{R}^n$ ssi pour tout $\boldsymbol{u} \in \mathbb{R}^n$:

$$\phi_{\boldsymbol{X}}(\boldsymbol{u}) = e^{i\langle \boldsymbol{u},\boldsymbol{m}\rangle - \frac{1}{2}\langle \boldsymbol{u},\mathbf{C}\boldsymbol{u}\rangle}$$

où $\boldsymbol{m} \in \mathbb{R}^n$ et $\mathbf{C}$ matrice de taille $n \times n$ symétrique positive.
Alors $\boldsymbol{m} = \mathbb{E}(\boldsymbol{X})$ et $\mathbf{C}$ matrice de covariance de $\boldsymbol{X}$.

Preuve :
Sens $\Longrightarrow$: Si $\boldsymbol{X}$ est un vecteur gaussien de moyenne $\boldsymbol{m}$ et de matrice de covariance $\mathbf{C}$, alors $\boldsymbol{X} = \mathbf{C}^{1/2}\boldsymbol{Y} + \boldsymbol{m}$ avec $\boldsymbol{Y}$ vecteur de v.a. gaussiennes centrées réduites indépendantes. Donc

$$\phi_{\boldsymbol{X}}(\boldsymbol{u}) = e^{i\langle \boldsymbol{m},\boldsymbol{u}\rangle}\phi_{\boldsymbol{Y}}(\mathbf{C}^{1/2}\boldsymbol{u}) = e^{i\langle \boldsymbol{m},\boldsymbol{u}\rangle}\exp\Big(-\frac{\|\mathbf{C}^{1/2}\boldsymbol{u}\|^2}{2}\Big)$$

et $\|\mathbf{C}^{1/2}\boldsymbol{u}\|^2 = \big\langle \mathbf{C}^{1/2}\boldsymbol{u}, \mathbf{C}^{1/2}\boldsymbol{u}\big\rangle = \langle \boldsymbol{u}, \mathbf{C}\boldsymbol{u}\rangle$.
Sens $\Longleftarrow$: la fonction caractéristique caractérise la loi.

Somme de variables aléatoires indépendantes

Proposition

Si $\boldsymbol{X}$ et $\boldsymbol{Y}$ sont deux v.a. indépendantes à valeurs dans $\mathbb{R}^n$, alors

$$\phi_{\boldsymbol{X}+\boldsymbol{Y}}(\boldsymbol{u}) = \mathbb{E}\left(e^{i\langle \boldsymbol{u}, \boldsymbol{X}+\boldsymbol{Y}\rangle}\right) = \mathbb{E}\left(e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle}\right)\mathbb{E}\left(e^{i\langle \boldsymbol{u}, \boldsymbol{Y}\rangle}\right) = \phi_{\boldsymbol{X}}(\boldsymbol{u})\phi_{\boldsymbol{Y}}(\boldsymbol{u})$$

Remarque : très utilisée dans la pratique pour trouver la loi d'une somme de variables aléatoires.
Exemple : Si X a une loi $\Gamma(n, \lambda)$, alors

$$\phi_X(u) = \left(\frac{\lambda}{\lambda - iu}\right)^n$$

en utilisant le fait que X a même loi qu'une somme de n variables aléatoires indépendantes de loi $\mathcal{E}(\lambda)$.

Fonction caractéristique et moments

Proposition

Soit $\boldsymbol{X} = (X_1, \ldots, X_n)$. *Si* $|\boldsymbol{X}|^m$ *est intégrable pour un entier* m, *alors la fonction* $\phi_{\boldsymbol{X}}$ *est* m *fois continûment différentiable sur* $\mathbb{R}^n$, *et on a pour tout choix des indices* $i_1, \ldots, i_m \in \{1, \ldots, n\}$:

$$\frac{\partial^m}{\partial x_{i_1} \cdots \partial x_{i_m}} \phi_{\boldsymbol{X}}(\boldsymbol{u}) = i^m \mathbb{E}\big(e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle} X_{i_1} \cdots X_{i_m}\big)$$

Preuve (cas $m = 1$). Soit $\boldsymbol{v}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ le $j$ème vecteur de la base canonique de $\mathbb{R}^n$. On a

$$\frac{\phi_X(\boldsymbol{u} + t\boldsymbol{v}_j) - \phi_X(\boldsymbol{u})}{t} = \mathbb{E}\Big(e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle} \frac{e^{itX_j} - 1}{t}\Big)$$

Soit une suite $t_p \to 0$. Les v.a. $\frac{e^{it_pX_j} - 1}{t_p}$ convergent p.s. vers iX_j quand $p \to \infty$ et sont bornées en module par $2|X_j| \in L^1$. Par convergence dominée,

$$\frac{\phi_X(\boldsymbol{u} + t_p\boldsymbol{v}_j) - \phi_X(\boldsymbol{u})}{t_p} \overset{p\to\infty}{\longrightarrow} i\mathbb{E}\big(e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle} X_j\big)$$

Application : Calcul des moments $\mathbb{E}\big(X_{i_1}\cdots X_{i_m}\big)$ en fonction des dérivées de $\phi_{\boldsymbol{X}}$ en $\mathbf{0}$.
Si X est à valeurs réelles avec $|X|^k$ intégrable, on a

$$\phi_X^{(k)}(u) = i^k \mathbb{E}\big(X^k e^{iuX}\big) \text{ pour } k = 0, 1, 2, \ldots$$

et donc

$$\phi_X^{(k)}(0) = i^k \mathbb{E}\big(X^k\big) \text{ pour } k = 0, 1, 2, \ldots$$

En particulier

$$\mathbb{E}(X) = -i\phi'_X(0), \quad \mathbb{E}(X^2) = -\phi''_X(0)$$

Exemple : Si $X \sim \mathcal{N}(0,1)$, alors (on a $\phi_X(u) = \exp(-u^2/2)$)

$$\mathbb{E}(X^{2n+1}) = 0, \quad \mathbb{E}(X^{2n}) = \frac{(2n)!}{2^n n!}$$

Théorème de Paul Lévy

Théorème

Soient $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ *des vecteurs aléatoires à valeurs dans* $\mathbb{R}^d$.

1. *Si la suite* $\boldsymbol{X}_n \xrightarrow{\mathcal{L}} \boldsymbol{X}$, *alors* $\phi_{\boldsymbol{X}_n}$ *converge simplement vers* $\phi_{\boldsymbol{X}}$.
2. *Si les* $\phi_{\boldsymbol{X}_n}$ *convergent simplement vers une fonction* ϕ *continue en* $\mathbf{0}$, *alors* ϕ *est la fonction caractéristique d'un vecteur aléatoire* $\boldsymbol{X}$ *et* $\boldsymbol{X}_n \xrightarrow{\mathcal{L}} \boldsymbol{X}$.

Preuve :

(1) : $\boldsymbol{x} \to e^{i\langle \boldsymbol{u},\boldsymbol{x}\rangle}$ est une fonction continue bornée de $\boldsymbol{x}$.

(2) : démonstration difficile.

Exemple : $(X_n)_{n\in\mathbb{N}}$ de loi $\Gamma(\alpha_n, \lambda_n)$:

$$\phi_{X_n}(u) = \left(\frac{\lambda_n}{\lambda_n - iu}\right)^{\alpha_n}$$

Si $(\alpha_n, \lambda_n) \to (\alpha, \lambda)$, alors $X_n \xrightarrow{\mathcal{L}} X$ de loi $\Gamma(\alpha, \lambda)$.

Paul Lévy (1886-1971), X1904, professeur à l'X à partir de 1920.

Intérêt fondamental : Le théorème de la limite centrale

• Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes, de même loi, de carré intégrable, avec $\mathbb{E}(X_1) = m$ et $\mathrm{Var}(X_1) = \sigma^2$.
Posons

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

M_n converge vers m, p.s. et en moyenne.
Questions : Quelle est la vitesse de convergence ? Comment décrire l'erreur $M_n - m$?

Théorème

Soit $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires indépendantes, de même loi, de carré intégrable, avec* $\mathbb{E}(X_1) = m$ *et* $\mathrm{Var}(X_1) = \sigma^2$. *On pose*

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

Alors les variables

$$\sqrt{n}(M_n - m)$$

convergent en loi vers une variable aléatoire normale $\mathcal{N}(0, \sigma^2)$.

Donc

$$\mathbb{P}\Big(a \le \sqrt{n}\big(M_n - m\big) \le b\Big) \xrightarrow{n\to+\infty} \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\Big(-\frac{x^2}{2\sigma^2}\Big)dx$$

Caractère universel du théorème : La somme d'un grand nombre de variables aléatoires indépendantes de même loi, centrées et de carré intégrable, se comporte approximativement comme une variable normale. Théorème énoncé par Laplace (1749-1827) et prouvé rigoureusement par Lyapounov (1901).

Preuve de Paul Lévy : Avec $\sigma > 0$.

$$Y_i = \frac{X_i - m}{\sigma}, \quad Z_n = \frac{\sqrt{n}}{\sigma}(M_n - m) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

Par indépendance,

$$\phi_{Z_n}(u) = \left(\phi_{Y_1/\sqrt{n}}(u)\right)^n$$

Or

$$\begin{aligned}
\phi_{Y_1/\sqrt{n}}(u) &= \mathbb{E}\Big(\exp\big(i \frac{u}{\sqrt{n}} Y_1\big)\Big) \\
&= \mathbb{E}\Big(1 + i\frac{u}{\sqrt{n}} Y_1 - \frac{1}{2}\frac{u^2}{n} Y_1^2 + o\big(\frac{1}{n}\big)\Big) \\
&= 1 - \frac{u^2}{2n} + o\big(\frac{1}{n}\big)
\end{aligned}$$

Donc

$$\lim_{n \to +\infty} \phi_{Z_n}(u) = \exp\Big(-\frac{u^2}{2}\Big) = \phi_Z(u)$$

où $Z \sim \mathcal{N}(0, 1)$.

Applications

- Permet de contrôler la précision de la loi des grands nombres.
- Applications à la méthode de Monte-Carlo : Vitesse de convergence en $1/\sqrt{n}$ indépendamment de la régularité de la fonction (Cours 7). Contrôle de l'erreur par intervalle de confiance (Cours 8).
- Applications en Statistique : observations $\rightarrow$ prédiction; quantifier l'erreur de prédiction (Cours 7).

Les hypothèses sont indispensables - un exemple

Considérons des variables aléatoires $X_n = \frac{V_n}{\sqrt{U_n}}$, avec U_n de loi uniforme sur $[0,1]$ et $\mathbb{P}(V_n = 1) = \mathbb{P}(V_n = -1) = \frac{1}{2}$, avec indépendance entre toutes les v.a..

- $(X_n)_{n\in\mathbb{N}}$ est une suite i.i.d.
- X_n est intégrable :

$$\mathbb{E}(|X_n|) = \mathbb{E}\Big(\frac{1}{\sqrt{U_n}}\Big) = \int_0^1 \frac{du}{\sqrt{u}} = 2$$

- Mais X_n n'est pas de carré intégrable car $\int_0^1 \frac{du}{u} = +\infty$.

Résultat : La suite $Z_n = \frac{1}{\sqrt{n \log n}} \sum_{i=1}^n X_i$ converge en loi vers une loi normale centrée réduite. La vitesse n'est plus la même que dans le TCL !

Preuve : Comme X_1 a une loi symétrique,

$$\begin{aligned}\phi_{X_1}(t) &= \mathbb{E}\big(\cos(tX_1)\big) = \mathbb{E}\Big(\cos\big(\frac{t}{\sqrt{U_1}}\big)\Big) = \int_0^1 \cos\big(\frac{t}{\sqrt{u}}\big)\,du \\ &= 2t^2\int_t^{+\infty}\frac{\cos x}{x^3}dx = t^2\Big(\frac{\cos t}{t^2} - \int_t^{\infty}\frac{\sin x}{x^2}dx\Big) \\ &= 1 + t^2\log t + O(t^2) \qquad [t\to 0]\end{aligned}$$

Alors la fonction caractéristique de Z_n vaut

$$\begin{aligned}\phi_{Z_n}(t) &= \Big(\phi_{X_1}\big(\frac{t}{\sqrt{n\log n}}\big)\Big)^n \\ &= \Big(1 + \frac{t^2}{n\log n}\log\big(\frac{t}{\sqrt{n\log n}}\big) + O\big(\frac{1}{n\log n}\big)\Big)^n \\ &= \Big(1 - \frac{t^2}{2n} + O\big(\frac{\log\log n}{2\log n}\big)\Big)^n \overset{n\to+\infty}{\longrightarrow} \exp\Big(-\frac{t^2}{2}\Big)\end{aligned}$$

Les hypothèses sont indispensables - un autre exemple

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi de Cauchy (de densité $\frac{1}{\pi(1+x^2)}$). X_n n'est pas intégrable.
Alors $\phi_{X_1}(t) = \exp(-|t|)$ (Voir Exercice 6.5.2).
On note $S_n = X_1 + \cdots + X_n$.
• On a

$$\phi_{S_n}(t) = \exp\big(-n|t|\big)$$

et donc

$$\phi_{S_n/\sqrt{n}}(t) = \exp\big(-\sqrt{n}|t|\big) \overset{n\to+\infty}{\longrightarrow} \mathbf{1}_{\{0\}}(t)$$

qui n'est pas continue en 0. La limite n'est pas une fonction caractéristique et on ne peut pas avoir convergence en loi. Le TCL ne s'applique pas.
• Aussi :

$$\phi_{S_n/n}(t) = \exp\big(-|t|\big)$$

donc la loi de $M_n = S_n/n$ est une loi de Cauchy pour tout n. La LGN ne s'applique pas.

Corrigé de l'exercice (leçon 5) : 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.
Preuve : Supposons que $(X_n)_n$ converge en probabilité vers X. Alors une sous-suite de $(X_n)_n$ converge p.s. vers X et comme $|X_n| \leq a$ p.s., nous en déduisons que $|X| \leq a$. Mais alors pour tout $\varepsilon > 0$,

$$\begin{aligned} \mathbb{E}\big(|X_n - X|\big) &= \mathbb{E}\big(|X_n - X|\mathbf{1}_{|X_n - X| \leq \varepsilon}\big) + \mathbb{E}\big(|X_n - X|\mathbf{1}_{|X_n - X| > \varepsilon}\big) \\ &\leq \varepsilon + 2a\mathbb{P}\big(|X_n - X| > \varepsilon\big) \end{aligned}$$

On a $2a\mathbb{P}\big(|X_n - X| > \varepsilon\big) \leq \varepsilon$ pour n assez grand par la convergence en probabilité.