Probabilité III

STEP, MINES ParisTech*

5 décembre 2019 (`#05c9041`)

Table des matières

---

*Ce document est un des produits du projet `boisgera/CDIS`, initié par la collaboration de (S)ébastien Boisgérault (CAOR), (T)homas Romary et (E)milie Chautru (GEOSCIENCES), (P)auline Bernard (CAS), avec la contribution de Gabriel Stoltz (Ecole des Ponts ParisTech, CERMICS).

Probabilités — cadre général

Interprétation

Les éléments de théorie de la mesure donnés en calcul intégral permettent une relecture des chapitres de probabilités. Le principal avantage est que les différents cas de figures déjà évoqués : lois de probabilités discrètes, à densité, mixtes vont pouvoir être traités dans un cadre unifié. On peut déjà s'apercevoir qu'une probabilité $\mathbb{P}$ définie sur un espace probabilisable (mesurable) $(\Omega, \mathcal{A})$ est une mesure positive *finie* au sens où $\mathbb{P}(\Omega) = 1$ et hérite ainsi de ses propriétés, on parle aussi de *mesure de probabilité*.

Remarque

- Dans le cas discret (Ω au plus dénombrable), la mesure de probabilité est une somme pondérée de mesures de Dirac :

$$\mathbb{P} = \sum_{\omega \in \Omega} w_\omega \delta_\omega,$$

  où $\sum_{\omega \in \Omega} w_\omega = 1$
- Dans le cas à densité ($\Omega = \mathbb{R}^n$, $n \in \mathbb{N}^\star$), la mesure de probabilité s'écrit :

$$\mathbb{P} = f(x)\mu,$$

  où f est une densité et μ la mesure de Lebesgue sur $\mathbb{R}^n$. On dit que $\mathbb{P}$ admet une densité par rapport à la mesure de Lebesgue.

Une variable aléatoire réelle, respectivement un vecteur aléatoire, X est une application mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, respectivement dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, et sa loi $\mathbb{P}_X$ est la mesure image de $\mathbb{P}$ par X.

Le fait que la composition d'un vecteur aléatoire réel par une application $\mathcal{B}(\mathbb{R}^n)/\mathbb{R}$ mesurable est une variable aléatoire s'obtient immédiatement par le résultat du chapitre précédent de calcul intégral portant sur la composition de fonctions mesurables. On peut généraliser la définition des espaces vectoriels $\mathcal{L}^1$ et $\mathcal{L}^2$ de la manière suivante :

Définition — Espace $\mathcal{L}^1$

Soit X une variable aléatoire. X est intégrable et on note $X \in \mathcal{L}^1$, ou $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, si et seulement si $\mathbb{E}(|X|) = \int_\Omega X(\omega)\mathbb{P}(d\omega) < +\infty$.

Définition — Espace $\mathcal{L}^2$

Soit X une variable aléatoire. X est de carré intégrable et on note $X \in \mathcal{L}^2$, ou $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, si et seulement si $\mathbb{E}(X^2) = \int_\Omega X^2(\omega)\mathbb{P}(d\omega) < +\infty$.

Les propriétés des espaces $\mathcal{L}^1$ et $\mathcal{L}^2$ données au chapitre 2 du cours de probabilités sont vraies en toute généralité. On peut d'ailleurs étendre ces définitions pour un $p \in N^\star$ quelconque.

Définition — Espace $\mathcal{L}^p$

Soit X une variable aléatoire. On note $X \in \mathcal{L}^p$, ou $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, si et seulement si $\mathbb{E}(|X|^p) = \int_\Omega |X|^p(\omega)\mathbb{P}(d\omega) < +\infty$.

Si $X \in \mathcal{L}^p$, on dit qu'elle admet un moment d'ordre p. Du fait que $\mathbb{P}$ est une mesure finie, on a :

Proposition

Soit $p \in \mathbb{N}^\star$, on a l'inclusion :

$$\mathcal{L}^{p+1}(\Omega, \mathcal{A}, \mathbb{P}) \subset \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$$

Démonstration Supposons $X \in \mathcal{L}^{p+1}(\Omega, \mathcal{A}, \mathbb{P})$. On a

$$|X|^p \leq \max(1, |X|^{p+1}) = 1_{|X|<1} + 1_{|X|\geq 1}|X|^{p+1}.$$

Le terme de droite est intégrable, en effet :

$$\mathbb{E}(1_{|X|<1} + 1_{|X|\geq 1}|X|^{p+1}) \leq \int_\Omega \mathbb{P}(d\omega) + \int_\Omega |X|^{p+1}\mathbb{P}(d\omega) = 1 + \mathbb{E}(|X|^{p+1}.$$

donc $|X|^p$ est intégrable.

On donne ici dans le cas général les inégalités de Markov et de Bienaymé-Chebyshev, déjà vues en CPGE. ■

Inégalité de Markov

Soit $p \in \mathbb{N}^\star$ et $a \in \mathbb{R}^\star$. Soit $X \in \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, on a

$$\mathbb{P}(|X| > a) \leq \frac{\mathbb{E}(|X|^p)}{a^p}$$

Démonstration On a

$$|X|^p \geq a^p 1_{[a,+\infty[}(|X|)$$

prenant l'espérance, on obtient ainsi

$$\mathbb{E}(|X|^p) \geq a^p \mathbb{E}(1_{[a,+\infty[}(|X|)) = a^p \mathbb{P}(|X| > a).$$

■

Inégalité de Bienaymé-Chebyshev

Soit $X \in \mathcal{L}^2$, on a

$$\mathbb{P}(|X - \mathbb{E}(X)| > a) \leq \frac{\mathbb{V}(X)}{a^2}$$

Démonstration C'est une application immédiate de l'inégalité de Markov à $(X - \mathbb{E}(X))$ avec $p = 2$. ■

Remarque

L'inégalité de Bienaymé-Chebyshev est très utile en pratique. Elle permet de mesurer la probabilité des grands écarts entre X et sa moyenne. Par exemple, avec $a = 10\sigma X$, il en résulte qu'il est improbable qu'une variable aléatoire X dévie de son espérance $\mathbb{E}(X)$ de plus de 10 fois son écart-type (probabilité inférieure à 0.01). Cette inégalité, tout à fait générale, n'est cependant pas très précise, et surestime très souvent en pratique le membre de gauche. On préférera, quand c'est possible, calculer directement ces probabilités à partir de la loi de X.

On peut également réécrire la proposition portant sur l'espérance de la composée d'une variable aléatoire et d'une fonction mesurable avec l'intégrale de Lebesgue :

Proposition

Soit X une variable aléatoire réelle de loi $\mathbb{P}_X$ et g une fonction $\mathcal{B}(\mathbb{R})/\mathcal{B}(\mathbb{R})$-mesurable (borélienne). Alors $g(X)$ est intégrable si et seulement si l'intégrale

$$\int_{\mathbb{R}} |g(x)|\mathbb{P}_X(dx) = \int_{\Omega} |g(X(\omega))|\mathbb{P}(d\omega),$$

est finie et dans ce cas

$$\mathbb{E}(g(X)) = \int_{R} g(x)\mathbb{P}(dx) = \int_{\Omega} g(X(\omega))\mathbb{P}(d\omega).$$

ce résultat a été démontré en exercice au chapitre précédent.

On notera que $\mathbb{P}$ étant finie, elle est nécessairement σ-finie. On peut ainsi caractériser l'indépendance de deux variables aléatoires quelconques au moyen du théorème de Fubini.

TODO indépendance : besoin de l'unicité de la mesure produit dans le cas $\sigma - fini$ (à inclure dans Fubini cf CI IV?)

puis cas infini

Lemme de Borel-Cantelli

Soit A_n une suite d'événements sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

1. Si $\sum_{i=1}^{n} \mathbb{P}(A_n) < \infty$, alors $\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}\left(\bigcap_{n\geq 1} \bigcup_{k\geq n} A_n\right) = 0$.
2. Si $\sum_{i=1}^{n} \mathbb{P}(A_n) = \infty$ et si les événements A_n sont mutuellement indépendants, alors on a $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

Démonstration Exercice.

On présente maintenant l'un des résultats essentiels de la théorie des probabilités. Ce résultat montre rigoureusement que, quand le nombre de répétitions de l'expérience tend vers l'infini, la fréquence de réalisation d'un événement converge vers la probabilité de réalisation de cet événement. Ce résultat, appelé **Loi des grands nombres**, a d'autres portées fondamentales. Il est en particulier à l'origine de méthodes de calcul numérique appelées Méthodes de Monte-Carlo, qui sont extrêmement puissantes et robustes. Elles sont par exemple très utilisées en Physique, en Mathématiques Financières, dans les méthodes de quantification d'incertitudes.

Plus généralement, on va étudier les suites de variables aléatoires indépendantes [1] et les notions de convergence de variables aléatoires. On verra que plusieurs notions sont possibles, non équivalentes, ce qui enrichit mais complique aussi la description des comportements asymptotiques.

On se place désormais dans le cadre général des variables aléatoires réelles n'admettant pas nécessairement de densité (elles peuvent être discrètes, mixtes,. . .).
■

Suites de variables aléatoires indépendantes

On considère une suite **infinie** $(X_n)_{n\in\mathbb{N}^\star}$ de variables aléatoires définies sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition

La suite $(X_n)_{n\in\mathbb{N}^\star}$ de variables aléatoires est dite *indépendante* si pour tout n, la famille finie $X_1, \ldots, X_n$ est indépendante.

Il est facile de vérifier que l'indépendance est préservée par certaines transformations.

Proposition

L'indépendance de la suite $(X_n)_{n\in\mathbb{N}^\star}$ entraîne celle de

1. toute sous-suite $(X_{i_k})_{k\in\mathbb{N}^\star}$,
2. toute suite de vecteurs issus de X_n,
3. toute suite de la forme $(f_n(X_n))_{n\in\mathbb{N}^\star}$, où les fonctions f_n sont des fonctions mesurables.

Exemple

Nous considérons l'ensemble $\Omega = [0, 1[$ muni de la tribu borélienne restreinte à cet ensemble, et de la mesure de Lebesgue. A chaque réel ω, nous associons son développement dyadique (unique si l'on impose que les ω_i ne soient pas tous égaux à 1 à partir d'un certain rang) :

$$\omega = \sum_{i\in\mathbb{N}^\star} \frac{\omega_i}{2^i}, \omega_i \in \{0, 1\}.$$

1. Dans ce chapitre, sauf mention contraire, **indépendant·e·s** signifie **mutuellement indépendant·e·s**.

L'application $X_i : \Omega \to \{0,1\}$, qui à ω associe $X_i(\omega) = \omega_i$ est une variable aléatoire sur Ω. En effet, pour $x_i \in \{0,1\}$, $1 \le i \le n$,

$$\{X_i = x_i\} = \bigcup_{x_1,\dots,x_{i-1}\in\{0,1\}} \left[\sum_{j=1}^{i} \frac{x_j}{2^j}, \sum_{j=1}^{i} \frac{x_j}{2^j} + \frac{1}{2^i} \right[,$$

qui est bien un élément de la tribu borélienne de $\Omega = [0,1[$, et

$$\mathbb{P}(\{X_i = x_i\}) = \frac{1}{2^i} \sum_{x_1,\dots,x_{i-1}\in\{0,1\}} 1 = \frac{1}{2}.$$

Montrons l'indépendance des variables aléatoires $(X_i)_{1\le i\le n}$. Nous avons

$$\bigcap_{1\le i\le n} \{X_i = x_i\} = \left[\sum_{i=1}^{n} \frac{x_i}{2^i}, \sum_{i=1}^{n} \frac{x_i}{2^i} + \frac{1}{2^n} \right[$$

si bien que

$$\mathbb{P}\left(\bigcap_{1\le i\le n} \{X_i = x_i\} \right) = \frac{1}{2^n} = \prod_{1\le i\le n} \mathbb{P}(\{X_i = x_i\}).$$

Cela démontre que les variables aléatoires X_i sont indépendantes et de loi de Bernoulli de paramètre $\frac{1}{2}$. Ainsi, nous venons de construire sur Ω une suite de variables aléatoires telle que toute sous-suite finie est constituée de variables aléatoires indépendantes. C'est donc une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre $\frac{1}{2}$.

Convergences et loi des grands nombres

Convergences de variables aléatoires

Dans ce paragraphe, on étudie des modes de convergence impliquant la proximité des variables aléatoires elles-mêmes.

On considère une suite $(X_n)_{n\in\mathbb{N}^\star}$ de vecteurs aléatoires, définis sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, et à valeurs dans $\mathbb{R}^d$. On considère également sur le même espace un vecteur “limite” X. On notera $|\cdot|$ la valeurs absolue dans $\mathbb{R}$ ou la norme euclidienne dans $\mathbb{R}^d$.

Définition

1. La suite $(X_n)_{n\in\mathbb{N}^\star}$ converge *presque sûrement* vers X, ce qui s'écrit $X_n \to X$ p.s., s'il existe un ensemble $N \in \mathcal{A}$ de probabilité nulle tel que

$$X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega), \ \ \forall \omega \notin N.$$

2. La suite $(X_n)_{n\in\mathbb{N}^\star}$ converge *en probabilité* vers X, ce qui s'écrit $X_n \xrightarrow{\mathbb{P}} X$, si $\forall\epsilon > 0$ on a
$$\mathbb{P}(|X_n - X| \geq \epsilon) \xrightarrow[n\to\infty]{} 0.$$
3. La suite $(X_n)_{n\in\mathbb{N}^\star}$ converge *en moyenne* (ou dans $\mathcal{L}^1$) vers X, ce qui s'écrit $X_n \xrightarrow{\mathcal{L}^1} X$, si X_n et X sont dans $\mathcal{L}^1$ et si
$$\mathbb{E}(|X_n - X|) \xrightarrow[n\to\infty]{} 0.$$

Remarque

— La convergence presque sûre est la plus proche de la convergence simple des fonctions. Mais ici, nous permettons à certains ω de ne pas vérifier $X_n(\omega) \to X(\omega)$, si toutefois la probabilité de réalisation de l'ensemble de ces ω est nulle. C'est l'équivalent probabiliste du *presque partout* vu en calcul intégral.

— La définition de la convergence dans $\mathcal{L}^1$ se généralise aux ordres supérieurs, pour $p \in \mathbb{N}\star$, on parle de convergence dans $\mathcal{L}^p$ (en moyenne quadratique si $p = 2$) ce qui s'écrit $X_n \xrightarrow{\mathcal{L}^p} X$, si X_n et X sont dans $\mathcal{L}^p$ et si
$$\mathbb{E}(|X_n - X|^p) \xrightarrow[n\to\infty]{} 0.$$

Ces convergences ne sont pas équivalentes comme le montrent les exemples suivants.

Exemples

— Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires de Bernoulli à valeurs dans $\{0,1\}$ telles que
$$\mathbb{P}(X_n = 1) = \frac{1}{n}; \;\; \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$
Pour tout $\epsilon \in]0,1[$, la probabilité $\mathbb{P}(|X_n| \geq \epsilon) = \frac{1}{n}$ tend vers 0 quand n tend vers l'infini. Ainsi, la suite $(X_n)_{n\in\mathbb{N}^\star}$ tend vers $X = 0$ en probabilité. Comme $\mathbb{E}(X_n) = \frac{1}{n}$, elle tend également en moyenne vers 0.
Mais si on considère maintenant une suite $(Y_n)_{n\in\mathbb{N}^\star}$ de variables aléatoires de Bernoulli à valeurs dans $\{0, n^2\}$ telles que
$$\mathbb{P}(Y_n = n^2) = \frac{1}{n}; \;\; \mathbb{P}(Y_n = 0) = 1 - \frac{1}{n}.$$
Par le même argument que ci-dessus, nous voyons que la suite $(Y_n)_{n\in\mathbb{N}^\star}$ converge en probabilité vers 0, mais comme $\mathbb{E}(Y_n) = n$, la suite ne converge pas en moyenne vers 0 (ni vers aucune autre limite finie).

— Soit U une variable aléatoire uniforme sur $[0,1]$. Posons $Z_n = 1_{\{U \leq \frac{1}{n}\}}$. Alors
$$\mathbb{P}(Z_n = 1) = \frac{1}{n}; \quad \mathbb{P}(Z_n = 0) = 1 - \frac{1}{n}$$
Si $\omega \in \{U > 0\}$ est fixé, alors $\exists n_0 \in \mathbb{N}$ tel que $U(\omega) > \frac{1}{n_0}$, et donc tel que $Z_n(\omega) = 0$ pour tout $n \geq n_0$. Comme $\mathbb{P}(U > 0) = 1$, ceci montre que la suite $(Z_n)_{n \in \mathbb{N}^\star}$ converge presque sûrement vers 0.

On étudie maintenant les liens entre ces différentes convergences.

Proposition

La convergence presque sûre et la convergence en moyenne entraînent la convergence en probabilité.

Démonstration Soit $A_{n,\epsilon} = \{|X_n - X| > \epsilon\}$.

— Supposons que $X_n \to X$ p.s. et soit N l'ensemble de probabilité nulle en dehors duquel on a $X_n(\omega) \to X(\omega)$. Si $\omega \notin N$, on a $\omega \notin A_{n,\epsilon}$ pour tout $n \geq n_0$, où n_0 dépend de ω et de ϵ, ce qui implique que les variables aléatoires $Y_{n,\epsilon} = 1_{N^c \cap A_{n,\epsilon}}$ tendent simplement vers 0 lorsque $n \to \infty$. Comme on a aussi $0 \leq Y_{n,\epsilon} \leq 1$ le théorème de convergence dominée implique que $\mathbb{E}(Y_{n,\epsilon}) \xrightarrow[n\to\infty]{} 0$. Mais
$$\mathbb{P}(A_{n,\epsilon}) \leq \mathbb{P}(N^c \cap A_{n,\epsilon}) + \mathbb{P}(N) = \mathbb{P}(N^c \cap A_{n,\epsilon}) = \mathbb{E}(Y_{n,\epsilon}) \xrightarrow[n\to\infty]{} 0.$$

— Supposons que $X_n \xrightarrow{\mathcal{L}^1} X$. Pour $\epsilon > 0$, on a $1_{A_{n,\epsilon}} \leq \frac{1}{\epsilon}|X_n - X|$, donc
$$\mathbb{P}(A_{n,\epsilon}) \leq \frac{1}{\epsilon}\mathbb{E}(|X_n - X|) \to 0.$$

■

La convergence en probabilité n'entraîne pas la convergence en moyenne, comme nous l'avons vu dans l'exemple ci-dessus, ne serait-ce que parce qu'elle n'implique pas l'appartenance de X_n et X à $\mathcal{L}^1$. Si les X_n ne sont "pas trop grands", il y a cependant équivalence entre les deux modes de convergence. En voici un exemple :

Proposition

S'il existe une constante a telle que $|X_n| \leq a$ presque sûrement, il y a équivalence entre $X_n \xrightarrow{\mathbb{P}} X$ et $X_n \xrightarrow{\mathcal{L}^1} X$.

Démonstration Etant donnée la proposition précédente, dont on reprend les notations, il suffit de montrer que la convergence en probabilité implique la convergence en moyenne lorsque $|X_n| \leq a$.

Comme $|X_n| \leq a$, on a $\{|X| > a + \epsilon\} \subset A_{n,\epsilon}$, et donc $\mathbb{P}(|X| > a + \epsilon) \leq \mathbb{P}(A_{n,\epsilon})$. En faisant $n \to \infty$, on en déduit que $\mathbb{P}(|X| > a + \epsilon) = 0$. Ceci est vrai pour tout ϵ et donc

$$\mathbb{P}(|X| > a) = 0.$$

Comme $|X_n| \leq a$, on a aussi

$$|X_n - X| \leq \epsilon + (X_n + X)1_{A_{n,\epsilon}} \leq \epsilon + 2a1_{A_{n,\epsilon}}$$

sur l'ensemble $\{|X| \leq a\}$ qui est de probabilité 1. On a ainsi

$$\mathbb{E}(|X_n - X|) \leq \epsilon + 2a\mathbb{P}(A_{n,\epsilon})$$

On en déduit que $\limsup_n \mathbb{E}(|X_n - X|) \leq \epsilon$, et comme ϵ est arbitrairement proche de 0, on a le résultat souhaité. ■

Les rapports entre convergence presque-sûre et convergence en probabilité sont plus subtils. La première de ces deux convergences est plus forte que la seconde d'après la proposition, mais “à peine plus”, comme le montre le résultat suivant.

Proposition

Si $X_n \xrightarrow{\mathbb{P}} X$, il existe une sous-suite (n_k) telle que $X_{n_k} \to X$ p.s. quand $k \to \infty$.

Démonstration Comme la suite $(X_n)_{n\in\mathbb{N}^\star}$ converge en probabilité vers X, on peut définir une sous-suite de la manière suivante : posons $n_1 = 1$, et soit

$$n_j = \inf\left\{n > n_{j-1}; \mathbb{P}\left(|X_r - X_s| > \frac{1}{2^j}\right) < \frac{1}{3^j},\ \text{pour } r, s \geq n\right\}.$$

Il résulte alors de

$$\sum_j = \mathbb{P}\left(|X_{n_{j+1}} - X_{n_j}| > \frac{1}{2^j}\right) < \sum_j \frac{1}{3^j} < \infty$$

que la suite $(X_{n_j})_{j\in\mathbb{N}^\star}$ converge presque-sûrement. C'est en effet une conséquence du lemme de Borel-Cantelli appliqué aux ensembles

$$A_j = \left\{|X_{n_{j+1}} - X_{n_j}| > \frac{1}{2^j}\right\}.$$

■

Exemple

Soient $\Omega = \mathbb{R}$ muni de sa tribu borélienne et $\mathbb{P}$ la probabilité uniforme sur $[0, 1]$. Soit $X_n = 1_{A_n}$, où A_n est un intervalle de $[0, 1]$ de longueur $\frac{1}{n}$. Ainsi, $\mathbb{E}(X_n) = \frac{1}{n}$, et la suite X_n tend vers $X = 0$ en moyenne, et donc en probabilité. Supposons que les A_n soient placés bout-à-bout, en recommençant en 0 chaque fois qu'on arrive au point 1. Il est clair que l'on parcourt indéfiniment l'intervalle $[0, 1]$ (car la série de terme général 1/n diverge). Ainsi la suite numérique $X_n(\omega)$ ne converge pour aucun ω, et on n'a pas $X_n \to X$ presque-sûrement ; cependant comme la série $\sum_n 1/n^2$ converge, il s'en suit que $X_{n^2} \to X = 0$ presque-sûrement. Nous avons donc la convergence presque-sûre de la sous-suite $(X_{n^2})_{n\in\mathbb{N}^\star}$.

Proposition

Soit f une fonction continue de $\mathbb{R}^d$ dans $\mathbb{R}$.

1. Si $X_n \to X$ presque-sûrement, alors $f(X_n) \to f(X)$ presque-sûrement.
2. Si $X_n \xrightarrow{\mathbb{P}} X$, alors $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

Démonstration

1. Soit N l'ensemble de probabilité nulle en dehors duquel on a $X_n(\omega) \to X(\omega)$. Si $\omega \notin N$, il vient
$$\lim_{n\to\infty} f(X_n(\omega)) = f(\lim_{n\to\infty} X_n(\omega)) = f(X(\omega))$$
par continuité de f, d'où le résultat.
2. On remarqe d'abord que si $K > 0$ et $\epsilon > 0$,
$$\{|f(X_n) - f(X)| \geq \epsilon\} \subset \{|X| > K\} \cup \{|X| \leq K, |f(X_n) - f(X)| \geq \epsilon\}.$$
La fonction f est uniformément continue sur $\{x : |x| \leq K\}$, donc il existe $\eta > 0$ tel que $|x-y| < \epsilon$ et $|x| \leq K$ et $y \leq K$ impliquent $|f(x)-f(y)| < \epsilon$. On a donc
$$\{|f(X_n) - f(X)| \geq \epsilon\} \subset \{|X| > K\} \cup \{|X_n - X| \geq \eta\}$$
d'où
$$\mathbb{P}(|f(X_n) - f(X)| \geq \epsilon) \leq \mathbb{P}(|X| > K) + \mathbb{P}(|X_n - X| \geq \eta).$$
Par hypothèse, il vient
$$\limsup_n \mathbb{P}(|f(X_n) - f(X)| \geq \epsilon) \leq \mathbb{P}(|X| > K).$$
Enfin, $\lim_{K\to+\infty} \mathbb{P}(|X| > K) = 0$ (par convergence dominée) et donc la lim sup ci-dessus est nulle. On a ainsi le résultat.

■

La loi des grands nombres

Dans ce paragraphe, on considère une suite $(X_n)_{n\in\mathbb{N}^\star}$ de variables aléatoires **indépendantes et de même loi** (ou indépendantes et identiquement distribuées, i.i.d. en abrégé). On considère la "moyenne" des n premières variables aléatoires :

$$M_n = \frac{X_1 + \ldots + X_n}{n},$$

et notre objectif est de montrer que M_n converge vers l'espérance des X_n lorsque celle-ci existe (comme les X_n ont même loi, cette espérance est la même pour tout n). Il s'agit là d'un des résultats essentiels de toute la théorie des probabilités, connu sous le nom de **loi des grands nombres**.

Nous allons démontrer dans un premier temps la loi des grands nombres pour des variables aléatoires de carré intégrable.

Théorème

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires indépendantes, de même loi et de **carré intégrable**, et $m = \mathbb{E}(X_n)$ leur moyenne. Alors la suite $(M_n)_{n\in\mathbb{N}^\star}$ définie par

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

converge vers m, **presque sûrement et en moyenne**, quand n tend vers l'infini. Elle converge donc aussi en probabilité. On a même convergence en *moyenne quadratique*, à savoir que :

$$\mathbb{E}((M_n - m)^2) \xrightarrow[n\to\infty]{} 0.$$

Le résultat sur la convergence en probabilité est appelé *loi faible des grands nombres*. Sa preuve est presque immédiate. Elle résulte de l'inégalité de Bienaymé-Chebyshev (exercice). Le résultat est peu informatif et permet d'obtenir certains contrôles d'erreurs. Le résultat prouvant la convergence presque-sûre est appelé *loi forte des grands nombres*. Sa preuve est plus délicate et utilise le lemme de Borel-Cantelli.

Démonstration

Notons σ^2 la variance des variables X_n, bien définie puisqu'on les a supposées de carré intégrable. En vertu de la linéarité de l'espérance, on a

$$\mathbb{E}(M_n) = m, \quad \mathbb{E}((M_n - m)^2) = \mathbb{V}(M_n) = \frac{\sigma^2}{n},$$

d'où la convergence en moyenne quadratique.

Comme $\mathbb{E}(Y)^2 \leq \mathbb{E}(Y^2)$, on en déduit que $\mathbb{E}(|M_n - m|) \to 0$, donc on a aussi la convergence en moyenne.

La preuve de la convergence presque-sûre est plus délicate.

Quitte à remplacer X_n par $X_n - m$ (et donc M_n par $M_n - m$), nous pouvons supposer que $m = 0$.

Montrons tout d'abord que la sous-suite $(M_{n^2})_{n\in\mathbb{N}^\star}$ converge presque-sûrement vers 0.

La convergence dans $\mathcal{L}^1$ impliquant la convergence en probabilité, on sait qu'on peut extraire de $(M_n)_n$ une sous-suite convergeant p.s. vers 0. Cependant, cela ne suffit pas puisqu'on veut que la suite $(M_n)_n$ elle-même converge p.s. Pour le montrer, on construit d'abord une sous suite-particulière qui converge p.s. puis on traite les termes qui se trouvent entre deux éléments successifs de la sous-suite.

D'après l'inégalité de Bienaymé-Chebyshev et ce qui précède, on a pour $q \in \mathbb{N}^\star$

$$\mathbb{P}(|M_{n^2}| \geq \frac{1}{q}) \leq \frac{\sigma^2 q^2}{n^2}$$

Donc si $A_{n,q} = \{|M_{n^2}| \geq \frac{1}{q}\}$, nous obtenons que $\sum_{n\geq 1} \mathbb{P}(A_{n,q}) < \infty$. Posons ensuite $B_{n,q} = \cup_{m\geq n} A_{m,q}$ et $C_q = \cap_{n\geq 1} B_{n,q} = \limsup_n A_{n,q}$. En appliquant le lemme de Borel-Cantelli, on obtient que $\mathbb{P}(C_q) = 0$. En conséquence, si on pose $N = \cup_{q\in\mathbb{N}^\star} C_q$, on obtient $\mathbb{P}(N) \leq \sum_{q\in\mathbb{N}^\star} \mathbb{P}(C_q) = 0$.

Si $\omega \notin N$, alors $\omega \in \cap_{q\in\mathbb{N}^\star} (C_q)^c$. Ainsi, $\omega \notin C_q$ pour tout $q \geq 1$, et donc $\omega \notin B_{n,q}$ pour n assez grand (car $B_{n,q}$ est décroissant en n). Cela siginfie que pour tout $\omega \notin N$, pour tout $q \geq 1$, il existe un n assez grand tel que $M_{k^2} \leq \frac{1}{q}$ dès que $k \geq n$. Autrement dit, $M_{n^2} \to 0$ si $\omega \notin N$, avec $\mathbb{P}(N) = 0$, d'où

$$M_{n^2} \xrightarrow[n\to\infty]{} 0 \text{ p.s.}$$

Montrons maintenant que la suite $(M_n)_{n\in\mathbb{N}^\star}$ tend presque-sûrement vers 0.

Pour tout entier n, notons $p(n)$, l'entier tel que $p(n)^2 \leq n \leq (p(n)+1)^2$. Alors,

$$M_n - \frac{p(n)^2}{n} M_{p(n)^2} = \frac{1}{n} \sum_{p=p(n)^2+1}^{n} X_p,$$

et puique les variables aléatoires de la somme sont indépendantes, il vient

$$\begin{aligned}
\mathbb{E}\left(\left(M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right)^2\right) &= \frac{n - p(n)^2}{n^2}\sigma^2 \\
&\leq \frac{2p(n)+1}{n^2}\sigma^2 \\
&\leq \frac{2\sqrt{n}+1}{n^2}\sigma^2 \leq \frac{3}{n^{3/2}}\sigma^2
\end{aligned}$$

où on a utilisé le fait que $p(n) \leq \sqrt(n)$.

En appliquant de nouveau l'inégalité de Bienaymé-Chebyshev, on obtient

$$\mathbb{P}\left(\left|M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right| > a\right) \leq \frac{2\sqrt{n}+1}{n^2}\frac{\sigma^2}{a^2}$$

Comme la série $\sum_n \frac{2\sqrt{n}+1}{n^2}$ converge, le même raisonnement que pour M_{n^2} décrit ci-dessus, montre que

$$M_n - \frac{p(n)^2}{n} M_{p(n)^2} \to 0 \text{ p.s.}$$

Par ailleurs, on a déjà montré que $M_{p(n)^2} \to 0$ p.s. et $\frac{p(n)^2}{n} \to 1$. On en déduit que $M_n \to 0$ p.s.

Plus généralement, le théorème suivant donne les hypothèses minimales assurant la validité de la loi des grands nombres, à savoir que les X_n sont dans $\mathcal{L}_1$ (on se référera par exemple à ce document en ligne ou à Jacod and Protter (2003) pour la démonstration).

Théorème

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires indépendantes, de même loi et **intégrables**, et $m = \mathbb{E}(X_n)$ leur moyenne. Alors la suite $(M_n)_{n\in\mathbb{N}^\star}$ définie par

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

converge vers m, **presque sûrement et en moyenne**, quand n tend vers l'infini.

Convergence en loi — fonction caractéristique — théorème central limite

Nous allons introduire maintenant une nouvelle notion de convergence de suites de variables aléatoires. La convergence en loi définie dans ce paragraphe va concerner les lois des variables aléatoires. Elle signifiera que les lois sont asymptotiquement "proches", sans que les variables aléatoires elles-mêmes le soient nécessairement.

Convergence en loi

On considère des vecteurs aléatoires X_n et X, tous à valeurs dans le même espace $\mathbb{R}^d$, mais pouvant éventuellement être définis sur des espaces de probabilité différents.

Définition

On dit que la suite $(X_n)_{n\in\mathbb{N}^\star}$ *converge en loi* vers X et on écrit $X_n \xrightarrow{\mathcal{L}} X$, si pour toute fonction réelle f **continue bornée** sur $\mathbb{R}^d$,

$$\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X)).$$

Exemple

Un cas très simple est celui où toutes les variables aléatoires X_n prennent un nombre fini de valeurs $\{x_i, 1 \leq i \leq N\}$. Alors, la suite $(X_n)_{n\in\mathbb{N}^\star}$ converge en loi vers X si et seulement si

$$\lim_{n\to+\infty} \mathbb{P}(X_n = x_i) = \mathbb{P}(X = x_i), \;\; \forall 1 \leq i \leq N$$

Il suffit d'écrire pour f continue bornée

$$\mathbb{E}(f(X_n)) = \sum_{i=1}^{N} f(x_i) P(X_n = x_i)$$

Dans l'exemple ci-dessus, N est fini et fixé. Mais nous avons un résultat analogue (en faisant tendre N vers l'infini) si les variables aléatoires ont un nombre dénombrable de valeurs. En particulier, le cas de la convergence de la loi binomiale vers la loi de Poisson a été traité en CPGE.

Remarque

Dans la définition, les v.a. X_n et X peuvent être définies sur des univers distincts puisque seules leurs lois sont en cause. Il arrive même qu'une suite X_n converge vers une limite X qui ne peut pas exister sur les espaces sur lesquels sont définies les X_n, parce que ceux-ci sont trop "petits" : par exemple, si X_n est une variable binomiale à n modalités, convenablement normalisée, et la limite X est gaussienne (on pourra justifier ceci par le théorème central limite); l'espace naturel sur lequel est définie X_n contient $n+1$ points, et sur un tel espace toutes les v.a. sont discrètes. La convergence en loi permet donc une sorte de convergence pour des v.a. pour lesquelles toute autre forme de convergence serait impossible.

Exemple

Soit $(X_n)_{n\in\mathbb{N}^\star}$ et X des variables aléatoires de lois respectives $\mathcal{N}(0, \sigma_n^2)$ et $\mathcal{N}(0, \sigma^2)$. On suppose que la suite de réels positifs $(\sigma_n)_{n\in\mathbb{N}^\star}$ converge vers $\sigma > 0$

quand n tend vers l'infini. Alors la suite $(X_n)_{n\in\mathbb{N}^\star}$ converge en loi vers X. En effet, soit f une fonction continue bornée sur $\mathbb{R}$. On a

$$\begin{aligned}\mathbb{E}(f(X_n)) &= \int_{\mathbb{R}} f(y)\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left(-\frac{y^2}{2\sigma_n^2}\right)dy\\ &\to \int_{\mathbb{R}} f(y)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y^2}{2\sigma^2}\right)dy\end{aligned}$$

où l'on a utilisée le théorème de convergence dominée.

La convergence en loi est plus faible que la convergence en probabilité et donc aussi que les convergences presque-sûre et en moyenne.

Proposition

Soient X_n et X des v.a., toutes définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$.

Démonstration Soit f une fonction réelle continue bornée. D'après la proposition, on a $f(X_n) \xrightarrow{\mathbb{P}} f(X)$ et donc $f(X_n)$ converge aussi en moyenne vers $f(X)$ par la proposition. Comme $|\mathbb{E}(Y)| \leq \mathbb{E}(|Y|)$ pour toute variable aléatoire réelle Y, on en déduit $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$. ∎

Un moyen efficace de caractériser la convergence en loi des variables aléatoires réelles passe par l'étude de la suite des fonctions de répartition.

Proposition

Soient X_n et X des variables aléatoires réelles de fonctions de répartition respectives F_n et F. Pour que $X_n \xrightarrow{\mathcal{L}} X$, il faut et il suffit que $F_n(x) \xrightarrow[n\to\infty]{} F(x)$ pour tout x en lequel F est continue.

Notons que puisque la fonction F est continue à droite et croissante, l'ensemble des points où F est continue est l'ensemble $D = \{x : F(x-) = F(x)\}$, et son complémentaire est au plus dénombrable. Ainsi, D est dense [2] dans $\mathbb{R}$.

Démonstration

2. Soient Ω un espace topologique et A une partie de Ω. On dit que A est *dense* dans Ω si l'une des propriétés équivalentes est satisfaite : tout ouvert non vide de Ω contient des éléments de A ; l'adhérence de A est égale à Ω ; tout point de Ω est adhérent à A ; le complémentaire de A est d'intérieur vide.

1. Supposons d'abord que $X_n \xrightarrow{\mathcal{L}} X$. Soit $a \in \mathbb{R}$ tel que $F(a-) = F(a)$. Pour tout $p \in \mathbb{N}^\star$ et tout $b \in \mathbb{R}$, il existe une fonction $f_{p,b}$ continue bornée sur $\mathbb{R}$ telle que
$$1_{]-\infty,b]} \leq f_{p,b} \leq 1_{]-\infty,b+\frac{1}{p}]}.$$
Alors, par définition, $\mathbb{E}(f_{p,b}(X_n)) \to \mathbb{E}(f_{p,b}(X))$ quand n tend vers l'infini. L'inégalité ci-dessus implique que $F_n(a) = \mathbb{P}(X_n \leq a) \leq \mathbb{E}(f_{p,a}(X_n))$ et $\mathbb{E}(f_{p,a}(X)) \leq F(a+1/p)$; donc $\limsup_n F_n(a) \leq F(a+1/p)$ pour tout p et donc on a aussi $\limsup_n F_n(a) \leq F(a)$. On a également que $F_n(a) \geq \mathbb{E}(f_{p,a-1/p}(X_n))$ et $\mathbb{E}(f_{p,a-1/p}(X)) \geq F(a-1/p)$; donc $\liminf_n F_n(a) \geq F(a-1/p)$ pour tout p et donc aussi $\liminf_n F_n(a) \geq F(a)$, puique $F(a-) = F(a)$. Ces deux résultats impliquent que $F_n(a) \xrightarrow[n\to\infty]{} F(a)$.
2. Inversement, supposons $F_n(x) \xrightarrow[n\to\infty]{} F(x)$ pour tout $x \in T$, où T est une partie dense de $\mathbb{R}$. Soit f une fonction continue bornée sur $\mathbb{R}$ et $\epsilon > 0$. Soient $a, b \in T$ avec $F(a) \leq \epsilon$ et $F(b) \geq 1-\epsilon$. Il existe n_0 tel que
$$n \geq n_0 \Rightarrow \mathbb{P}(X_n \notin]a,b]) = 1 - F_n(b) + F_n(a) \leq 3\epsilon.$$
La fonction f est uniformément continue sur $[a,b]$, donc il existe un nombre fini de points $a_0 = a < a_1 < \ldots < a_k = b$ appartenant tous à T, et tels que $|f(x) - f(a_i)| \leq \epsilon$ si $a_{i-1} \leq x \leq a_i$. Donc
$$g(x) = \sum_{i=1}^{k} f(a_i) 1_{]a_{i-1},a_i]}(x)$$
vérifie $|f-g| \leq \epsilon$ sur $]a,b]$. Si $M = \sup_x |f(x)|$, il vient alors
$$|\mathbb{E}(f(X_n)) - \mathbb{E}(g(X_n))| \leq M\mathbb{P}(X_n \notin [a,b]) + \epsilon,$$
de même pour X. Enfin, $\mathbb{E}(g(X_n)) = \sum_{i=1}^{k} f(a_i)(F_n(a_i) - F_n(a_{i-1})$, et de même pour X par définition de g. Comme $(F_n(a_i))_{n\in\mathbb{N}^\star}$ converge vers $F(a_i)$ pour tout i, on en déduit l'existence de n_1 tel que
$$n \geq n_1 \Rightarrow |\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))| \leq \epsilon.$$
Finalement, on a
$$n \geq \sup(n_0, n_1) \Rightarrow |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq 3\epsilon + 5M\epsilon.$$
Vu l'arbitraire sur ϵ, on en déduit que $\mathbb{E}(f(X_n))$ converge vers $\mathbb{E}(f(X))$, d'où le résultat.

■

Corollaire

Si la suite $(X_n)_{n\in\mathbb{N}^\star}$ de variables aléatoires réelles converge en loi vers X, et si la loi de X admet une densité, alors pour tous $a < b$,

$$\mathbb{P}(X_n \in]a,b]) \xrightarrow[n\to\infty]{} \mathbb{P}(X \in]a,b]).$$

Démonstration La fonction de répartition de X est alors continue en tout point. (Mais pas nécessairement celle des variables aléatoires X_n.) ■

Fonctions caractéristiques

Dans ce paragraphe, nous introduisons un outil important en calcul des probabilités : il s'agit de ce que l'on appelle *la fonction caractéristique* d'une variable aléatoire, et qui dans d'autres branches des mathématiques s'appelle aussi *la transformée de Fourier*. Elle nous sera notamment très utile pour démontrer le théorème central limite.

On notera $< x, y >$ le produit scalaire de deux vecteurs de $\mathbb{R}^n$. Si $u \in \mathbb{R}^n$, la fonction (complexe) $x \mapsto e^{i<u,x>}$ est continue, de module 1. Donc si X est un vecteur aléatoire à valeurs dans $\mathbb{R}^n$, nous pouvons considérer $e^{i<u,X>}$ comme une variable aléatoire à valeurs complexes. Ses parties réelle $Y = \cos(< u, X >)$ et imaginaire $Z = \sin(< u, X >)$ sont des variables aléatoires réelles. Ces variables aléatoires réelles sont de plus bornées par 1, donc elles admettent une espérance. Il est alors naturel d'écrire que l'espérance de $e^{i<u,x>}$ est

$$\mathbb{E}(e^{i<u,X>}) = \mathbb{E}(Y) + i\mathbb{E}(Z) = \mathbb{E}(\cos < u, X >) + i\mathbb{E}(\sin < u, X >)$$

Définition

Si X est un vecteur aléatoire à valeurs dans $\mathbb{R}^n$, sa *fonction caractéristique* est la fonction ϕ_X de $\mathbb{R}^n$ dans $\mathbb{C}$ définie par

$$\phi_X(u) = \mathbb{E}(e^{i<u,x>}) = \int_{\mathbb{R}^n} e^{i<u,x>}\mathbb{P}_X(dx).$$

Remarque

La fonction caractéristique ne dépend en fait **que de la loi** $\mathbb{P}_X$ **de** X : c'est la "transformée de Fourier" de la loi $\mathbb{P}_X$.

Nous verrons que cette fonction porte bien son nom, au sens où elle caractérise la loi $\mathbb{P}_X$. C'est une notion qui, de ce point de vue, généralise la fonction génératrice

G, vue en CPGE dans le cas discret. Elle vérifie

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

pour une variable X à valeurs dans $\mathbb{N}$.

Proposition

ϕ_X est de module inférieur à 1, continue, avec

$$\phi(0) = 1; \;\; \phi_X(-u) = \overline{\phi_X(u)}.$$

Démonstration $|z|$ désigne le module d'un nombre complexe z.
Comme $\mathbb{E}(Y)^2 \leq \mathbb{E}(Y^2)$ pour toute variable réelle Y, on a :

$|\phi_X(u)|^2 = \mathbb{E}(\cos < u, X >)^2 + \mathbb{E}(\sin < u, X >)^2 \leq \mathbb{E}(\cos^2 < u, X > + \sin^2 < u, X >) = 1.$

Pour montrer la continuité, considérons une suite $u_p \xrightarrow[p\to\infty]{} u$. Il y a convergence simple de $e^{i<u_p,X>}$ vers $e^{i<u,X>}$. Comme ces variables aléatoires sont de module inférieur à 1, le théorème de convergence dominée assure que $\phi_X(u_p) \xrightarrow[p\to\infty]{} \phi_X(u)$. ϕ_X est donc continue. ■

Proposition

Si X est un vecteur aléatoire à valeurs dans $\mathbb{R}^n$, si $a \in \mathbb{R}^m$ et A est une matrice réelle de taille $m \times n$, alors

$$\phi_{a+AX}(u) = e^{i<u,a>}\phi_X(A^t u), \forall u \in \mathbb{R}^m$$

Démonstration Nous avons $e^{i<u,a+AX>} = e^{i<u,a>}e^{i<A^t u,X>}$. En effet, $< u, AX > = < A^t u, X >$. On prend ensuite les espérances pour obtenir le résultat.
■

Exemples

1. X suit une loi binomiale $\mathcal{B}(n,p)$: $\phi_X(u) = (pe^{iu} + 1 - p)^n$.
2. X suit une loi de Poisson $\mathcal{P}(\theta)$: $\phi_X(u) = e^{\theta(e^{iu}-1)}$.
3. X suit une loi uniforme $\mathcal{U}_{[a,b]}$:$\phi_X(u) = \frac{e^{iua} - e^{iub}}{iu(b-a)}$.
4. X suit une loi exponentielle $\mathcal{E}(\lambda)$, $\lambda > 0$: $\phi_X(u) = \frac{\lambda}{\lambda - iu}$.

5. X suit une loi normale $\mathcal{N}(0,1)$: $\phi_X(u) = e^{-u^2/2}$.
6. X suit une loi normale $\mathcal{N}(\mu,\sigma^2)$: $\phi_X(u) = e^{iu\mu - u^2\sigma^2/2}$.

L'intérêt majeur de la fonction caractéristique réside dans le fait qu'elle caractérise la loi de la variable aléatoire (d'où son nom).

Théorème

La fonction caractéristique ϕ_X caractérise la loi du vecteur aléatoire X. Ainsi, si deux vecteurs aléatoires X et Y ont même fonction caractéristique, ils ont même loi.

Démonstration Soient les fonctions suivantes avec $\sigma > 0$:

$$f_\sigma(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{|x|^2}{2\sigma^2}\right) \text{ et } \widehat{f}_\sigma(u) = \exp\left(-\frac{|u|^2\sigma^2}{2}\right).$$

On note que f_σ est la densité d'un vecteur gaussien Z de dimension n, centré et de matrice de covariance $\sigma^2 I_n$, autrement dit dont les composantes sont indépendantes, centrées, de variances σ^2.

On a

$$\begin{aligned}
\mathbb{E}(e^{i<u,Z>}) = \int f_\sigma(x) e^{i<u,x>} dx &= \int_{\mathbb{R}^n} \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_j^2}{2\sigma^2} + iu_j x_j\right) dx_1 \dots dx_n \\
&= \prod_{j=1}^n \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2} + iu_j t\right) dt \\
&= \prod_{j=1}^n e^{-u_j^2\sigma^2/2} \\
&= \widehat{f}_\sigma(u)
\end{aligned}$$

d'après l'exemple 5 ci-dessus et en utilisant le théorème de Fubini. On remarque ainsi que

$$f_\sigma(u-v) = \frac{1}{(2\pi\sigma^2)^{n/2}} \widehat{f}_\sigma\left(\frac{u-v}{\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \int f_\sigma(x) e^{i<u-v,x>/\sigma^2} dx.$$

Supposons que X et Y admettent la même fonction caractéristique $\phi_X = \phi_Y$.

En utilisant le théorème de Fubini, on a

$$\begin{aligned}
\mathbb{E}_X(f_\sigma(X-v)) &= \int f_\sigma(u-v)\mathbb{P}_X(du) \\
&= \int_{\mathbb{R}^n} \frac{1}{(2\pi\sigma^2)^{n/2}} \left(\int_{\mathbb{R}^n} f_\sigma(x) e^{i<u-v,x>/\sigma^2} dx \right) \mathbb{P}_X(du) \\
&= \int_{\mathbb{R}^n} f_\sigma(x) \frac{1}{(2\pi\sigma^2)^{n/2}} \left(\int_{\mathbb{R}^n} e^{i<u,x/\sigma^2>} \mathbb{P}_X(du) \right) e^{-i<v,x>/\sigma^2} dx \\
&= \int_{\mathbb{R}^n} f_\sigma(x) \frac{1}{(2\pi\sigma^2)^{n/2}} \phi_X\left(\frac{x}{\sigma^2}\right) e^{-i<v,x>/\sigma^2} dx,
\end{aligned}$$

et la même égalité reste vraie pour $\mathbb{P}_Y$. On en déduit que

$$\mathbb{E}(g(X)) = \int g(u)\mathbb{P}_X(du) = \int g(u)\mathbb{P}_Y(du) = \mathbb{E}(g(X'))$$

pour toute fonction g de l'espace vectoriel de fonction engendrées par $u \mapsto f_\sigma(u-v)$.

D'après le théorème de Stone-Weiertrass [3], cet espace est dense dans l'ensemble C_0 des fonctions continues sur $\mathbb{R}^n$ et ayant une limite nulle à l'infini, pour la topologie de la convergence uniforme (la norme est le sup sur $\mathbb{R}^n$).

Par suite, $\mathbb{E}(g(X)) = \mathbb{E}(g(X'))$ pour toute fonction $g \in C_0$. Comme l'indicatrice de tout ouvert est limite croissante de fonctions de C_0, on en déduit que $\mathbb{P}_X(A) = \mathbb{E}(1_A(X))$ est égal à $\mathbb{P}_{X'}(A) = \mathbb{E}(1_A(X'))$ pour tout ouvert A, ce qui implique $P_X = P_{X'}$. ■

La fonction caractéristique offre également un moyen commode de caractériser l'indépendance.

Corollaire

Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^n$. Ses composantes X_i sont indépendantes si et seulement si pour tous $u_1, \ldots, u_n \in \mathbb{R}$,

$$\phi_X(u_1, \ldots, u_n) = \prod_{j=1}^{n} \phi_{X_j}(u_j)$$

où ϕ_X désigne la fonction caractéristique du vecteur aléatoire X, et ϕ_{X_j} celle de la composante X_j opur chaque j.

Démonstration On a $< u, X >= \sum_{j=1}^n u_j X_j$. Si les X_i sont indépendantes, et comme $e^{i<u,x>} = \prod_j e^{iu_j x_j}$, nous obtenons directement le résultat en utilisant la proposition.

3. voir par exemple Simmons (2003) p.160.

Supposons inversement qu'on ait $\phi_X(u_1,\ldots,u_n)=\prod_{j=1}^n \phi_{X_j}(u_j)$. On peut alors construire des variables aléatoires X'_j indépendantes, telles que X'_j et X_j aient même loi pour tout j et donc telles que $\phi_{X'_j}=\phi_{X_j}$. Si $X'=(X'_1,\ldots,X'_n)$, on a donc $\phi_{X'}=\phi_X$. Donc X et X' ont même loi, ce qui entraîne que pour tous boréliens A_j, on ait

$$\mathbb{P}(\bigcap_j \{X_j \in A_j\}) = \mathbb{P}(\bigcap_j \{X'_j \in A_j\}) = \prod_j \mathbb{P}(\{X'_j \in A_j\}) = \prod_j \mathbb{P}(\{X_j \in A_j\})$$

d'où l'indépendance. ■

Proposition

Si X et Y sont deux vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^n$, la fonction caractéristique de la somme $X+Y$ est donnée par

$$\phi_{X+Y} = \phi_X \phi_Y$$

Démonstration Comme $e^{i<u,X+Y>} = e^{i<u,X>}e^{i<u,Y>}$, il suffit d'appliquer la proposition. ■

Exemples :

Soient X et Y deux variables aléatoires réelles indépendantes et $Z=X+Y$:

1. Si X suit la loi normale $\mathcal{N}(m,\sigma^2)$ et Y suit $\mathcal{N}(m',\sigma'^2)$, alors Z suit une loi $\mathcal{N}(m+m',\sigma^2+\sigma'^2)$, d'après l'exemple point 6. et la proposition ci-dessus.
2. Si X et Y suivent des lois de Poisson de paramètres θ et θ', alors Z suit une loi de Poisson de paramètre $\theta+\theta'$, d'après l'exemple point 2. et la proposition ci-dessus.
3. Si X suit une loi binomiale $\mathcal{B}(n,p)$ et Y la loi biomiale $\mathcal{B}(m,p)$, alors Z suit une loi binomiale $\mathcal{B}(n+m,p)$, d'après l'exemple point 1. et la proposition.

Proposition

Soit X un vecteur aléatoire de $\mathbb{R}^n$. Si la variable $|X|^m$ (ou $|\cdot|$ désigne la norme euclidienne) est intégrable pour un entier m, la fonction ϕ_X est m fois continûment différentiable sur $\mathbb{R}^n$ et pour tout choix des indices $i_1,\ldots,i_m$,

$$\frac{\partial^m}{\partial u_{i_1}\ldots\partial u_{i_m}}\phi_X(u) = i^m\mathbb{E}(e^{i<u,X>}X_{i_1}\ldots X_{i_m}),$$

où les X_j sont les composantes de X.

Démonstration Le résultat se démontre par application itérée du théorème de dérivation sous le signe somme. $\blacksquare$

Remarque

En prenant $u = 0$ dans la proposition, la formule permet de calculer $\mathbb{E}(X_{i_1} \dots X_{i_m})$ en fonction des dérivées à l'origine de ϕ_X, autrement dit de calculer tous les moments du vecteur X, s'ils existent. Par exemple, si X est à valeurs réelles et est intégrable (respectivement de carré intégrable), on a

$$\mathbb{E}(X) = i\phi'_X(0), \ \ (\text{resp.}\mathbb{E}(X^2) = \phi"_X(0))$$

Théorème

X est un vecteur gaussien si et seulement si sa fonction caractéristique s'écrit

$$\phi_X(u) = e^{i<u,m>-\frac{1}{2}<u,Cu>}$$

où $m = \mathbb{E}(X) \in \mathbb{R}^n$ et C est la matrice de covariance de X.

Démonstration

1. Supposons $\phi_X(u) = e^{i<u,m>-\frac{1}{2}<u,Cu>}$. Pour toute combinaison linéaire $Y = \sum_j a_j X_j =< a, X >$, et pour tout $v \in \mathbb{R}$, on a

$$\phi_Y(v) = \phi_X(va) = e^{iv<a,m>-\frac{v^2}{2}<a,Ca>}$$

   donc Y suit la loi $\mathcal{N}(< a, m >, < a, Ca >)$.
2. Inversement, soit C la matrice de covariance de X et m son vecteur moyenne. Si $Y =< a, X >= \sum_{j=1}^n a_j X_j$ avec $a \in \mathbb{R}^n$, un calcul simple montre

$$\mathbb{E}(Y) =< a, m >, \mathbb{V}(Y) =< a, Ca >$$

   Par hypothèse, Y suit une loi normale donc vu le point 6. de l'exemple plus haut, sa fonction caractéristique est

$$\phi_Y(v) = e^{iv<a,m>-\frac{v^2}{2}<a,Ca>}$$

   Mais $\phi_Y(1) = \phi_{<a,X>}(1) = \mathbb{E}(e^{i<a,X>}) = \phi_X(a)$, d'où le résultat.

Le théorème suivant caractérise la convergence en loi à l'aide des fonctions caractéristiques. C'est un critère extrêmement utile dans la pratique. $\blacksquare$

Théorème de Lévy

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$.

1. Si la suite $(X_n)_{n\in\mathbb{N}^\star}$ converge en loi vers X, alors ϕ_{X_n} converge simplement vers ϕ_X.
2. Si les ϕ_{X_n} convergent simplement vers une fonction (complexe) ϕ sur $\mathbb{R}^d$ et si cette fonction est **continue** en 0, alors c'est la fonction caractéristique d'une variable aléatoire X et $X_n \xrightarrow{\mathcal{L}} X$.

Démonstration

1. On remarque que $\phi_{X_n}(u) = \mathbb{E}(g_u(X_n))$ et $\phi_X(u) = \mathbb{E}(g_u(X))$ où g_u est la fonction continue bornée $g_u(x) = e^{i<u,x>}$. On applique alors la définition.
2. Se reporter à Jacod and Protter (2003).

■

Théorème central limite

Ce théorème est aussi connu sous le nom de théorème de la limite centrale. Plus simplement, il apparaît souvent sous l'abréviation TCL.

On considère une suite de variables aléatoire $(X_n)_{n\in\mathbb{N}^\star}$ indépendantes, de même loi et de carré intégrable. On note m et σ^2 la moyenne et la variance commune aux variables X_n, et

$$S_n = X_1 + \ldots + X_n$$

ainsi $(S_n = nM_n)$. On a vu que la loi des grands nombres assure que M_n converge vers m presque-sûrement et en moyenne. On va s'intéresser a la vitesse à laquelle cette convergence a lieu.

Pour évaluer cette vitesse, c'est-à-dire trouver un équivalent de $\frac{S_n}{n} - m$, on est amenés à étudier la limite éventuelle de la suite $n^\alpha(\frac{S_n}{n} - m)$ pour différentes valeurs de α : si α est "petit" cette suite va encore tendre vers 0, et elle va "exploser" si α est "grand". On peut espérer que pour une (et alors nécessairement une seule) valeur de α, cette suite converge vers une limite qui n'est ni infinie ni nulle.

Il se trouve que la réponse à cette question a un aspect "négatif" : la suite $n^\alpha(\frac{S_n}{n} - m)$ ne converge au sens presque-sûr, ou même en probabilité, pour aucune valeur de α. Elle a aussi un aspect "positif" : cette suite converge, au sens de la convergence en loi, pour la même valeur $\alpha = 1/2$ quelle que soit la loi des X_n, et toujours vers une loi normale.

Ce résultat, qui peut sembler miraculeux, a été énoncé par Laplace (1749-1827) et démontré beaucoup plus tard par Lyapounov (1901). Il montre le caractère universel de la loi normale en probabilités (d'où son nom).

Théorème central limite

Si les X_n sont des variables aléatoires réelles, indépendantes et de même loi, de carré intégrable, de moyenne m et de variance $\sigma^2 > 0$, alors les variables

$$\frac{S_n - nm}{\sigma\sqrt{n}}$$

convergent en loi vers une variable aléatoire de loi $\mathcal{N}(0,1)$.

En d'autres termes, $\sqrt{n}(M_n - m)$ converge vers une variable normale de loi $\mathcal{N}(0,\sigma^2)$.

Démonstration Soit ϕ la fonction caractéristique de $X_n - m$, et $Y_n = \frac{S_n - nm}{\sigma\sqrt{n}}$. Comme les X_n sont indépendantes, la proposition vue plus haut entraîne que la fonction caractéristique de Y_n est

$$\begin{aligned}
\phi_{Y_n}(u) &= \phi_{\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n}(X_j - m)}(u) \\
&= \phi_{\sum_{j=1}^{n}(X_j - m)}\left(\frac{u}{\sigma\sqrt{n}}\right) \\
&= \prod_{j=1}^{n}\phi\left(\frac{u}{\sigma\sqrt{n}}\right) \\
&= \phi\left(\frac{u}{\sigma\sqrt{n}}\right)^n
\end{aligned}$$

Comme $\mathbb{E}(X_n - m) = 0$ et $\mathbb{E}((X_n - m)^2) = \sigma^2$, la proposition portant sur la dérivée des entraîne que

$$\phi'(0) = 0 \text{ et } \phi''(0) = -\sigma^2$$

Si on fait le développement de Taylor à l'ordre 2 au voisinage de zéro, on obtient

$$\phi(u) = 1 - \frac{u^2\sigma^2}{2} + u^2 h(|u|),$$

où $h(u) \to 0$ quand $u \to 0$. On a ainsi

$$\begin{aligned}
\phi_{Y_n}(u) &= \phi\left(\frac{u}{\sigma\sqrt{n}}\right)^n \\
&= e^{n\log\phi\left(\frac{u}{\sigma\sqrt{n}}\right)} \\
&= e^{n\log(1 - \frac{u^2}{2n} + \frac{u^2}{n\sigma^2}h(\frac{u}{\sigma\sqrt{n}}))}
\end{aligned}$$

où log désigne la valeur principale du logarithme complexe [4] (elle vaut 0 au point 1 et est continue dans le cercle complexe de centre 1 et de rayon 1/2 et admet le même développement limité au voisinage de $z = 1$ que le logarithme réel). Comme $\phi(0) = 1$ et que ϕ est continue en 0, on a que pour u fixé et n assez grand,

$$\left|\phi\left(\frac{u}{\sigma\sqrt{n}}\right) - 1\right| \leq 1/2.$$

En faisant $n \to \infty$, on obtient

$$\lim_{n\to\infty} \phi_n(u) = \exp -\frac{u^2}{2}$$

Le théorème de Lévy implique alors que Y_n converge en loi vers Z de fonction caractéristique $\phi_Z(u) = e^{-u^2/2}$ où l'on reconnaît la fonction caractéristique de la loi $\mathcal{N}(0, 1)$. ■

Remarque

On peut déduire de ce résultat que $n^\alpha(\frac{S_n}{n} - m)$ converge vers 0 (resp. $+\infty$) en probabilité lorsque $\alpha < 1/2$ (resp. $\alpha > 1/2$).

Exemple : convergence des lois binomiales

Supposons que S_n suive une loi binomiale $\mathcal{B}(p, n)$. Cela revient à dire que S_n a la même loi qu'une somme $X_1 + \ldots + X_n$ de n variables aléatoires X_i indépendantes de loi $\mathcal{B}(p, 1)$, i.e. $\mathbb{P}(X_i = 1) = p$ et $\mathbb{P}(X_i = 0) = 1 - p$. On a alors $m = p$ et $\sigma^2 = p(1 - p)$.

On veut calculer $\mathbb{P}(S_n \leq x)$ pour x fixé et n grand.

Si p est petit de sorte que $\theta = np$ ne soit pas trop grand (en pratique, $\theta \leq 5$ convient), on peut utiliser l'approximation par une loi de Poisson, vue en CPGE. Si p est très proche de 1, de sorte que $\theta = n(1 - p)$ soit comme ci-dessus, alors $n - S_n$ suit à son tour une loi proche de la loi de Poisson de paramètre θ.

Dans les autres cas, on utilise la loi des grands nombres et le théorème central limite :

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} p,$$
$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

4. voir par exemple https://fr.wikipedia.org/wiki/Logarithme_complexe

Si on désigne par Φ la fonction de répartition de la loi $\mathcal{N}(0,1)$, il vient

$$\mathbb{P}(S_n \leq x) \approx \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$$

Imaginons que l'on lance 1000 fois une pièce (non truquée). On cherche la probabilité d'obtenir plus de 545 fois le côté Face. Le calcul exact utilisant les lois binomiales est extrêmement lourd. Le résultat ci-dessus nous donne une très bonne approximation. On a

$$\mathbb{P}(S_{1000} > 545) = \mathbb{P}\left(\frac{S_{1000}-500}{\sqrt{250}} > \frac{45}{\sqrt{250}}\right) \approx \int_{\frac{45}{\sqrt{250}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Cette dernière intégrale se calcule numériquement (on trouve encore des abaques où les valeurs de Φ sont tabulées) et on obtient

$$\mathbb{P}(S_{1000} > 545) \approx 1 - \Phi(2,84) \approx 0,0023.$$

Le théorème admet une version multidimensionnelle, de preuve similaire. On considère des vecteurs aléatoires X_n à valeurs dans $\mathbb{R}^d$, indépendants et de même loi, dont les composantes sont de carré intégrable. On a un vecteur moyenne $m = E(X_n)$, et une matrice de covariance $C = (c_{ij})_{i,j=1,\dots,d}$ avec $c_i j = \text{Cov}(X_i, X_j)$. On peut alors énoncer le TCL multi-dimensionnel.

Théorème central limite multi-dimensionnel

Les vecteurs aléatoires $\frac{S_n - nm}{\sqrt{n}}$ convergent en loi vers un vecteur aléatoire gaussien centré (i.e. de moyenne nulle), de matrice C.

Remarque

Il est important de noter ici que la vitesse de convergence ne dépend pas de la dimension des vecteurs X_n.

Exercices

Inégalités de concentration

Question 1 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées d'espérance $\mathbb{E}(X_i) = m$ et de variance $\mathbb{V}(X_i) = \sigma^2 \leq 1$.

Montrer que pour tout $\delta \in]0,1[$, avec probabilité au moins $1-\delta$ on a

$$\left| \frac{1}{n} \sum_{i=1}^{n} X_i - m \right| \leq \sqrt{\frac{1}{\delta n}}$$

On peut trouver des bornes à décroissance beaucoup plus rapide dans des cas particuliers. **(?)**

Question 2 On suppose désormais que les X_i sont de loi $\mathcal{N}(m, \sigma^2)$.

i) On pose pour $u \in \mathbb{R} : M(u) = \mathbb{E}(e^{u(X_1-m)})$. Calculer $M(u)$.

ii) On pose $S_n = X_1 + \ldots + X_n$. Montrer que $\forall a \in \mathbb{R}, \forall u \in \mathbb{R}_+$,

$$\mathbb{P}(S_n - nm > a) \leq e^{-ua}(M(u))^n,$$

et que $\forall a \in \mathbb{R}, \forall u \in \mathbb{R}_-$,

$$\mathbb{P}(S_n - nm < -a) \leq e^{ua}(M(u))^n.$$

iii) Soit $Y_n = \frac{S_n}{n}$. Montrer que $\forall \epsilon > 0$,

$$\mathbb{P}(|Y_n - m| > \epsilon) \leq 2 \exp\left(\frac{-n\epsilon^2}{2\sigma^2}\right).$$

Cette inégalité est appelée inégalité de Chernov.

On peut dériver ce type d'inégalités pour différentes lois de probabilité. On voit qu'ici la concentration auprès de la moyenne se fait à vitesse exponentielle. On a le même type de résultats pour la loi de Bernoulli par exemple ce qui est très utile en apprentissage statistique dans les problèmes de classification.

iv) On suppose que $m = 1$ et que $\sigma^2 = 10$. Quelle taille d'échantillon doit-on choisir pour obtenir

$$\mathbb{P}(|Y_n - m| < \epsilon) \geq \alpha,$$

avec $\alpha = 0,95$ et $\epsilon = 0,05$,

— en utilisant l'inégalité de Bienaymé-Chebyshev ?
— en utilisant l'inégalité de Chernov ?

(?)

Lemme de Borel-Cantelli

Soit A_n une suite d'événements sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Question 1 On suppose que $\sum_{i=1}^{n} \mathbb{P}(A_n) < \infty$. Montrer que $\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}\left(\bigcap_{n\geq 1} \bigcup_{k\geq n} A_n\right) = 0$. **(?)**

Question 2 On suppose maintenant que les événements A_n sont mutuellement indépendants. Montrer que si $\sum_{i=1}^{n} \mathbb{P}(A_n) = \infty$, alors on a $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$. **(?)**

Question 3 Donner un exemple où $\sum_{i=1}^{n} \mathbb{P}(A_n) = \infty$ et $\mathbb{P}(\limsup_{n\to\infty} A_n) < 1$ quand les A_n ne sont pas indépendants. **(?)**

Loi faible des grands nombres

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires indépendantes, de même loi et de **carré intégrable**, et $m = \mathbb{E}(X_n)$ leur moyenne. Montrer que la suite $(M_n)_{n\in\mathbb{N}^\star}$ définie par

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

converge vers m en probabilité quand n tend vers l'infini.

Théorème de Slutsky

Solutions

Inégalités de concentration

Question 1 L'inégalité de Chebyshev nous donne pour tout $a > 0$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| > a\right) \leq \frac{\mathbb{V}(X_1)}{na^2}$$

Prenant, $\frac{\mathbb{V}(X_1)}{na^2} = \delta$, on obtient $a = \frac{1}{\sqrt{n\delta}}$, d'où le résultat. On retrouve au passage la même vitesse de convergence que celle donnée par le TCL.

Question 2

i) $M(u) = \mathbb{E}(e^{u(X_1 - m)}) = e^{u^2/\sigma^2}$.

ii) On a $e^{u(M_n - nm)} \geq e^{ua} 1_{S_n - nm \geq a}$, d'où

$$\mathbb{P}(S_n - nm \geq a) \leq \frac{\mathbb{E}(e^{u(M_n - nm)})}{e^{ua}} = e^{-ua} M(u)^n$$

par indépendance des X_i.

iii)

$$\begin{aligned}\mathbb{P}(|Y_n - m| \geq \epsilon) &= \mathbb{P}(S_n - nm \geq n\epsilon) + P(S_n - nm \leq -n\epsilon)\\ &\leq e^{-nu\epsilon}(M(u))^n + e^{-n(-u)(-\epsilon)}(M(-u))^n = 2e^{-nu\epsilon}(M(u))^n\\ &\leq 2e^{-nu\epsilon} e^{u^2/\sigma^2}, \ \ \forall u \geq 0\end{aligned}$$

La meilleure majoration va être obtenue en minimisant l'exposant, c'est-à-dire pour $u = \epsilon$. Nous en déduisons l'inégalité de Chernov.

iv) Par l'inégalité de Bienaymé-Chebyshev, $\mathbb{P}(|Yn - m| > \epsilon) \leq \frac{\mathbb{V}(Y_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$.

Ainsi, $\mathbb{P}(|Y_n - m| \leq \epsilon) \leq \alpha$ dès que $1 - \frac{\sigma^2}{n\epsilon^2}$. Avec $\epsilon = 0,05$, il vient $n \geq 80000$.

Par l'inégalité de Chernov, nous obtenons $2e^{-n\epsilon^2/20} \leq 0,05$. Il vient que $n \geq 29512$. Pour avoir une évaluation du même ordre de la probabilité cherchée, nous pouvons donc prendre un échantillon beaucoup plus petit si nous utilisons l'inégalité de Chernov. A taille d'échantillon fixée, nous aurons une meilleure évaluation avec cette inégalité.

Lemme de Borel-Cantelli

Question 1 On voit dans un premier temps que $\bigcap_{n\geq 0} \bigcup_{k \geq n} A_n \in \mathcal{A}$ par unions et intersections dénombrables. On a

$$\mathbb{P}(\limsup_n A_n) = \lim_{p \to \infty} P(\cup_{n \geq p} An) \leq \lim \mathbb{P}(A_n),$$

où on remarque que les deux suites sont décroissantes.

Si la série $\sum_n \mathbb{P}(A_n)$ est convergente, le reste de cette série tend vers 0 et l'inégalité implique que $\mathbb{P}(\limsup_n An) = 0$.

Question 2 Supposons maintenant que les A_n soient indépendants et que la série diverge. Soit m un nombre entier. Nous avons

$$\begin{aligned}\mathbb{P}(\cup_{i=p}^{m} Ai) = 1 - \mathbb{P}(\cap_{i=p}^{m} A_i^c) &= 1 - \prod_{i=p}^{m} \mathbb{P}(A_i^c) \text{ par indépendance} \\ &= 1 - \prod_{i=p}^{m} (1 - \mathbb{P}(A_i)) \geq 1 - e^{-\sum_{i=p}^{m} \mathbb{P}(A_i)}\end{aligned}$$

du fait de l'inégalité $1 - x \leq e^{-x}$ pour $x \geq 0$. Ainsi,

$$\mathbb{P}(\cup_{i=p}^{\infty} A_i) \geq 1 - e^{-\sum_{i=p}^{\infty} \mathbb{P}(A_i)} = 1$$

et l'on conclut finalement que pour tout p, $\mathbb{P}(\cup_{i=p}^{\infty} A_i) = 1$, ce qui implique finalement que $\mathbb{P}(\limsup_{n \to \infty} A_n) = 1$.

Question 3 Prendre tous les A_n égaux à un même événement A de probabilité $\mathbb{P}(A) \in]0, 1[$.

Loi faible des grands nombres

Appliquer L'inégalité de Bienaymé-Chebyshev à la variable aléatoire M_n.

Références

Jacod, J., and P. Protter. 2003. *L'essentiel En Théorie Des Probabilités*. Cassini. `https://hal.archives-ouvertes.fr/hal-00104956`.

Simmons, George F. 2003. *Introduction to Topology and Modern Analysis*. Reprint edition. Malabar, Fla: Krieger Publishing Company.