

Probabilités continues

Support du cours PROC

Guillaume Euvrard, juin 2020
EPITA

Partie 2

Table des matières

1	Distribution à deux variables	3
1.1	Cas général	3
1.2	Densité jointe de deux variables aléatoires	4
1.3	Interprétation intuitive d'une densité	5
1.4	Densités déduites de la densité conjointe	6
1.5	Variables indépendantes	6
2	Suite de variables aléatoires	8
2.1	Les différents types de convergence	9
2.2	Quelques exemples	10
2.3	Relations entre ces types de convergence	13
2.4	Théorèmes de convergence	13
2.5	Exemples d'applications du théorème central limite	15

1 Distribution à deux variables

1.1 Cas général

Considérons deux variables aléatoires réelles X et Y . Pour définir la distribution du couple (X, Y) , on procède d'une manière analogue à ce qu'on a fait pour une variable :

1. On définit la tribu borélienne sur $\mathbb{R}^2$, qui est la tribu engendrée par les ensembles de la forme $] -\infty, x] \times] -\infty, y]$ ou $(x, y) \in \mathbb{R}^2$.
2. On définit la fonction de répartition du couple (X, Y) comme la fonction

$$F : \begin{cases} \mathbb{R}^2 & \longrightarrow [0, 1] \\ (x, y) & \longmapsto P((X, Y) \in] -\infty, x] \times] -\infty, y]) = P(X \leq x \text{ et } Y \leq y) \end{cases}$$

La fonction de répartition F permet de définir la probabilité de tous les ensembles de la tribu borélienne de $\mathbb{R}^2$. Cette tribu ne contient pas tous les sous-ensembles de $\mathbb{R}^2$, donc certains d'entre eux n'ont pas de probabilité définie par F .

Néanmoins, les ensembles qui nous intéressent dans la pratique sont bien dans la tribu borélienne. C'est notamment le cas de ceux de la forme $I \times J$ où I et J sont deux intervalles de $\mathbb{R}$.

Par exemple, pour tout $(a, b, c, d) \in \mathbb{R}^4$ tel que $a < b$ et $c < d$, la probabilité que (X, Y) soit dans $]a, b] \times]c, d]$ vaut

$$P(a < X \leq b \text{ et } c < Y \leq d) = F(b, d) - F(b, c) - F(a, d) + F(a, c)$$

En effet, l'ensemble $] -\infty, b] \times] -\infty, d]$ est l'union *disjointe*

$$]-\infty, b] \times] -\infty, d] =] -\infty, b] \times] -\infty, c] \cup] -\infty, a] \times]c, d] \cup]a, b] \times]c, d]$$

On en déduit la relation entre les probabilités :

$$\underbrace{P_{XY}(]-\infty, b] \times] -\infty, d])}_{F(b, d)} = \underbrace{P_{XY}(]-\infty, b] \times] -\infty, c])}_{F(b, c)} + \underbrace{P_{XY}(]-\infty, a] \times]c, d])}_{F(a, d) - F(a, c)} + P_{XY}(]a, b] \times]c, d])$$

En isolant $P_{XY}(]a, b] \times]c, d])$ des autres termes, on en déduit la relation.

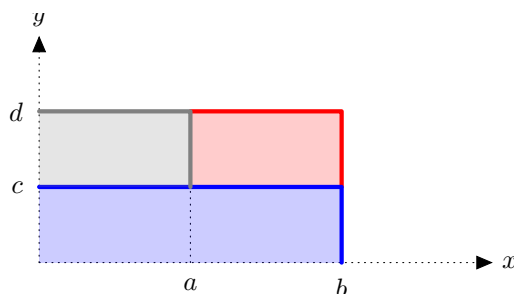


FIGURE 1 – Partition de l'ensemble $] -\infty, b] \times] -\infty, d]$. Cette partition permet de déduire la probabilité que (X, Y) soit dans $]a, b] \times]c, d]$ (l'ensemble en rouge).

Propriétés de la fonction de répartition

1. Pour tout $a \in \mathbb{R}$, la fonction $y \mapsto F(a, y)$ est croissante et $\lim_{y \rightarrow -\infty} F(a, y) = 0$
2. Pour tout $b \in \mathbb{R}$, la fonction $x \mapsto F(x, b)$ est croissante et $\lim_{x \rightarrow -\infty} F(x, b) = 0$
3. $\lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F(x, y) = 1$

1.2 Densité jointe de deux variables aléatoires

Soit X et Y deux variables aléatoires.

Définition : si il existe une fonction positive f définie sur $\mathbb{R}^2$ telle que pour tout $(a, b) \in \mathbb{R}^2$,

$$F(a, b) = \int_{x=-\infty}^a \int_{y=-\infty}^b f(x, y) dx dy$$

alors cette fonction f est une densité jointe du couple (X, Y) .

Remarques

1. Si la fonction de répartition F est deux fois différentiable sur $\mathbb{R}^2$, alors elle admet pour densité sa dérivée partielle seconde :

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}(x, y)$$

2. Il en est de même si F est deux fois différentiable, sauf en un nombre fini de points.
3. Quand une densité f existe, alors pour tout ensemble de la forme $I = [a, b] \times [c, d]$,

$$P((X, Y) \in I) = P(a \leq X \leq b \text{ et } c \leq Y \leq d) = \iint_I f(x, y) dx dy$$

Cette intégrale est le volume délimité par l'ensemble I , entre le plan $z = 0$ et la surface du graphe de la fonction f .

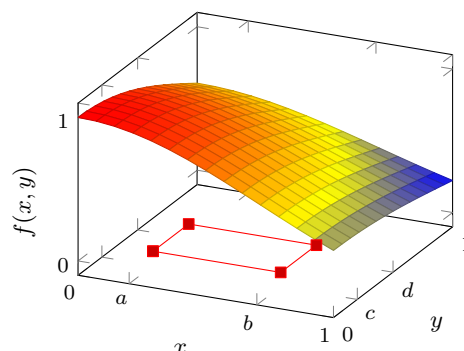


FIGURE 2 – Interprétation géométrique de l'intégrale de f sur $[a, b] \times [c, d]$. L'intégrale est le volume délimité par l'ensemble $[a, b] \times [c, d]$ (dans le plan $z = 0$) et le graphe de la fonction.

1.3 Interprétation intuitive d'une densité

Si on choisit deux nombres δx et δy petits dans $\mathbb{R}_+^*$, on peut approximer les variables X et Y par deux variables discrètes X' et Y' prenant leurs valeurs dans

$$X'(\Omega) = \{p\delta x, p \in \mathbb{Z}\} \quad \text{et} \quad Y'(\Omega) = \{q\delta y, q \in \mathbb{Z}\}$$

Pour tout $x = p\delta x \in X'(\Omega)$ ($p \in \mathbb{Z}$) et tout $y = q\delta y \in Y'(\Omega)$ ($q \in \mathbb{Z}$), la probabilité que $(X', Y') = (x, y)$ est

$$P(X' = x \text{ et } Y' = y) = f(x, y) \delta x \delta y$$

On a alors, pour tout $(a, b) = (p_1\delta x, p_2\delta x)$, $(p_1, p_2) \in \mathbb{Z}^2$, $p_1 < p_2$, et pour tout $(c, d) = (q_1\delta y, q_2\delta y)$, $(q_1, q_2) \in \mathbb{Z}^2$, $q_1 < q_2$,

$$P((X', Y') \in [a, b] \times [c, d]) = \sum_{i=p_1}^{p_2} \sum_{j=q_1}^{q_2} f(i\delta x, j\delta y) \delta x \delta y \simeq \iint_{[a, b] \times [c, d]} f(x, y) dx dy$$

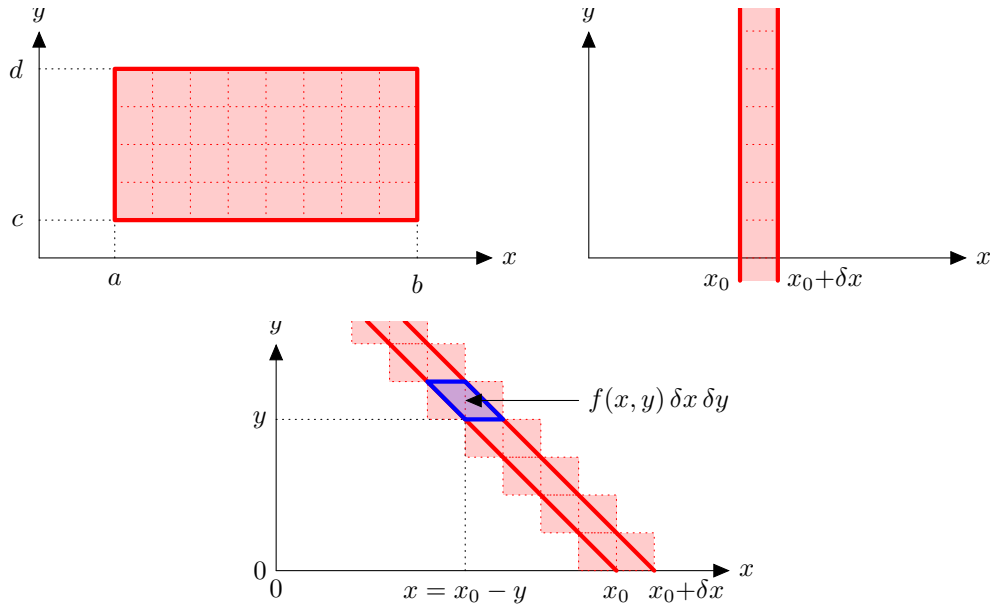


FIGURE 3 – Approximation de la distribution jointe de X et Y . Cette distribution est approximée par celle d'un couple (X', Y') prenant ses valeurs sur la grille $\{(p\delta x, q\delta y), (p, q) \in \mathbb{Z}^2\}$. **En haut à gauche :** la probabilité que (X', Y') soit sur un point de la grille à l'intérieur du rectangle $[a, b] \times [c, d]$ est une approximation de celle que (X, Y) soit dans ce rectangle. **En haut à droite :** la densité de X s'obtient en considérant la probabilité que X soit entre x_0 et $x_0 + \delta x$. C'est la somme des probabilités des événements « $(X, Y) \in [x_0, x_0 + \delta x] \times [q\delta y, (q+1)\delta y]$ ». **En bas :** de même, la densité de $X + Y$ en x_0 s'obtient en considérant la probabilité de l'évènement $X + Y \in [x_0, x_0 + \delta x]$. C'est la somme des probabilités élémentaires situées autour la droite d'équation $x + y = x_0$.

1.4 Densités déduites de la densité conjointe

Il ne faut pas chercher une démonstration mathématique solide des propriétés suivantes. Une telle démonstration fait appel à des outils mathématiques complexes et peu utiles dans les situations pratiques. En revanche, il est important de comprendre leurs significations géométriques (figure 3). Considérons un couple de variables aléatoires (X, Y) , de densité conjointe $f(x, y)$.

Définition : densité marginale de X et de Y .

La densité marginale de X est la fonction $f_X : x \mapsto \int_{-\infty}^{+\infty} f(x, y) dy$.

C'est la densité de la variable X , considérée comme une variable aléatoire en elle-même.

De même, la densité marginale de Y est la fonction $f_Y : y \mapsto \int_{-\infty}^{+\infty} f(x, y) dx$.

Propriété : densité de $X + Y$

La variable $X + Y$ admet pour densité la fonction $f_{X+Y} : x_0 \mapsto \int_{y=-\infty}^{+\infty} f(x_0 - y, y) dy$

1.5 Variables indépendantes

Définition

Les variables X et Y sont indépendantes si, pour tout $(x, y) \in \mathbb{R}^2$, $f(x, y) = f_X(x) \times f_Y(y)$.

Propriétés : soient X et Y deux variables aléatoires indépendantes.

1. Pour tout ensemble de la forme $[a, b] \times [c, d] \subset \mathbb{R}^2$, on a

$$P(X \in [a, b] \text{ et } Y \in [c, d]) = P(X \in [a, b]) \times P(Y \in [c, d]) = \left(\int_a^b f_X(x) dx \right) \times \left(\int_c^d f_Y(y) dy \right)$$

En d'autres termes, les événements « $X \in [a, b]$ » et « $Y \in [c, d]$ » sont indépendants.

2. La densité de $X + Y$ est le produit de convolution des densités marginales f_X et f_Y .

La seconde propriété se démontre en considérant que, pour tout $x_0 \in \mathbb{R}$,

$$f_{X+Y}(x_0) = \int_{y=-\infty}^{+\infty} f(x_0 - y, y) dy = \int_{y=-\infty}^{+\infty} f_X(x_0 - y) f_Y(y) dy$$

ce qui est bien le produit de convolution $f_X * f_Y$.

Exercice : distribution uniforme.

Soit deux variables X et Y dont la densité jointe est la fonction

$$f(x, y) = \begin{cases} \frac{1}{4} & \text{si } (x, y) \in [1, 3] \times [0, 2] \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer les densités marginales de X et Y .
2. En déduire que ces variables sont indépendantes.

3. Déterminer la densité de $X + Y$.

Solution

1. La densité marginale de X est définie pour tout $x_0 \in \mathbb{R}$ par $f_X(x_0) = \int_{-\infty}^{+\infty} f(x_0, y) dy$.
Si $x_0 \notin [1, 3]$, la fonction $y \mapsto f(x_0, y)$ est nulle pour tout $y \in \mathbb{R}$ et $f_X(x_0) = 0$.
Si $x_0 \in [1, 3]$, la fonction $y \mapsto f(x_0, y)$ vaut $\frac{1}{4}$ pour $y \in [0, 2]$ et 0 ailleurs. Ainsi,

$$f_X(x_0) = \int_0^2 \frac{1}{4} dy = \left[\frac{y}{4} \right]_0^2 = \frac{1}{2}$$

Finalement, la densité f_X est définie par

$$f_X(x_0) = \begin{cases} \frac{1}{2} & \text{si } x_0 \in [1, 3] \\ 0 & \text{sinon} \end{cases}$$

On voit donc que $X \rightsquigarrow \text{Unif}(1, 3)$.

De même, on montre que

$$f_Y(y_0) = \begin{cases} \frac{1}{2} & \text{si } y_0 \in [0, 2] \\ 0 & \text{sinon} \end{cases}$$

et donc $Y \rightsquigarrow \text{Unif}(0, 2)$.

2. On a bien, pour tout $(x, y) \in \mathbb{R}^2$, $f(x, y) = f_X(x) \times f_Y(y)$.
En effet, ce dernier produit s'annule si $x \notin [1, 3]$ ou $y \notin [0, 2]$. Dans le cas contraire, c'est à dire si $(x, y) \in [1, 3] \times [0, 2]$, il vaut $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.
3. Posons $Z = X + Y$ et déterminons $f_Z(x_0)$ pour $x_0 \in \mathbb{R}$. On sait que

$$f_Z(x_0) = \int_{-\infty}^{+\infty} f_X(x_0 - y) f_Y(y) dy$$

De plus, le produit $f_X(x_0 - y) f_Y(y)$ vaut

$$f_X(x_0 - y) f_Y(y) = \begin{cases} \frac{1}{4} & \text{si } 1 \leq x_0 - y \leq 3 \quad \text{et} \quad 0 \leq y \leq 2 \\ 0 & \text{sinon} \end{cases}$$

Donc

$$f_X(x_0 - y) f_Y(y) \neq 0 \iff \begin{cases} 0 \leq y \leq 2 \\ 1 \leq x_0 - y \leq 3 \end{cases} \iff \begin{cases} 0 \leq y \leq 2 \\ x_0 - 3 \leq y \leq x_0 - 1 \end{cases}$$

Pour distinguer les cas à étudier, on voit qu'il faut que $y \in [0, 2] \cap [x_0 - 3, x_0 - 1]$. Ces deux intervalles ont même longueur. Les cas à considérer sont ceux d'une intersection vide (un des intervalles ne contient que des valeurs inférieures à l'autre) et ceux d'intersection non vide (ils peuvent alors se rencontrer par les valeurs inférieures ou supérieures). On peut aussi s'appuyer sur un raisonnement géométrique (figure 4).

- (a) Cas 1 : si $x_0 - 3 > 2$, donc $x_0 > 5$.

Alors pour tout $y \in \mathbb{R}$, le produit $f_X(x_0 - y) f_Y(y)$ est nul donc $f_Z(x_0) = 0$.

- (b) Cas 2 : si $x_0 - 1 < 0$, donc $x_0 < 1$.

Là aussi, pour tout $y \in \mathbb{R}$, le produit $f_X(x_0 - y) f_Y(y)$ est nul donc $f_Z(x_0) = 0$.

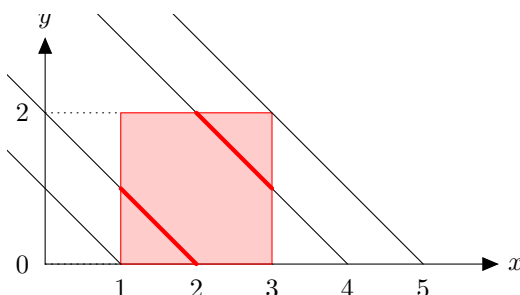


FIGURE 4 – Produit de convolution de deux densités uniformes. La zone sur fond rouge est le support du produit $f_X(x)f_Y(y)$. C'est la région dans laquelle le produit ne s'annule pas. On montre l'intersection de ce support avec la droite oblique $x + y = x_0$ pour quelques valeurs de x_0 . Si $x_0 < 1$, la droite ne rencontre pas le support. Si $x_0 \in [1, 3]$, la droite rencontre le support pour $y \in [0, x_0 - 1]$. Si $x_0 \in [3, 5]$, l'intersection se produit quand $y \in [x_0 - 3, 2]$. Enfin, quand $x_0 > 5$, l'intersection est vide.

(c) Cas 3 : si $0 \leq x_0 - 3 \leq 2 \leq x_0 - 1$, donc si $x_0 \in [3, 5]$.

Alors $f_X(x_0 - y)f_Y(y) \neq 0 \iff x_0 - 3 \leq y \leq 2$. Donc

$$f_Z(x_0) = \int_{x_0-3}^2 \frac{1}{4} dy = \left[\frac{y}{4} \right]_{x_0-3}^2 = \frac{5 - x_0}{4}$$

(d) Cas 4 : si $x_0 - 3 \leq 0 \leq x_0 - 1 \leq 2$, donc si $x_0 \in [1, 3]$.

Alors $f_X(x_0 - y)f_Y(y) \neq 0 \iff 0 \leq y \leq x_0 - 1$. Donc

$$f_Z(x_0) = \int_0^{x_0-1} \frac{1}{4} dy = \left[\frac{y}{4} \right]_0^{x_0-1} = \frac{x_0 - 1}{4}$$

$$\text{Finalement, } f_Z(x_0) = \begin{cases} \frac{x_0-1}{4} & \text{si } x_0 \in [1, 3] \\ \frac{5-x_0}{4} & \text{si } x_0 \in [3, 5] \\ 0 & \text{sinon} \end{cases}$$

2 Suite de variables aléatoires

Dans de nombreuses situations, nous avons l'intuition de la convergence d'une suite de variables aléatoires. Par exemple, si nous tirons n fois une variable, les tirages étant indépendants et suivant la même distribution, on sent bien que la moyenne des tirages converge vers l'espérance.

Formalisons cet exemple de façon plus mathématique : considérons n variables $X_1, X_2, \dots, X_n$. Supposons qu'elles soient indépendantes et identiquement distribuées (i.i.d.). Si on définit

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

on sent que $\bar{X}_n$ converge vers $E(X_n)$ (cette espérance est aussi celle de $X_1, X_2, \dots$).

Encore faut-il définir cette notion de convergence. En effet, si on se fixe une grande valeur de n , il reste possible que $\bar{X}_n$ soit loin de $E(X)$. C'est simplement peu probable.

2.1 Les différents types de convergence

Considérons une suite de variables aléatoires (X_n) et une variable aléatoire Y . Il y a plusieurs façons de définir que (X_n) converge vers Y .

Remarque : dans la définition générale, la limite Y est une variable aléatoire. Il arrive cependant souvent que cette variable soit en fait une constante réelle. C'est par exemple le cas quand on considère la moyenne

$$\bar{X}_n = \frac{X_1 + x_2 + \cdots + X_n}{n}$$

et $Y = E(X_n)$. Dans ce cas, en appelant C cette constante, Y est la variable

$$Y : \begin{cases} \Omega & \longrightarrow \mathbb{R} \\ \omega & \longmapsto C \end{cases}$$

Autrement dit, c'est la variable dont la seule valeur possible est C , qu'elle atteigne avec la probabilité $P(Y=C) = 1$.

Définition : convergence presque sûre.

On dit que la suite (X_n) converge presque sûrement vers Y si

$$P\left(\lim_{n \rightarrow +\infty} X_n = Y\right) = 1$$

Autrement dit, il existe un ensemble $\Omega' \subset \Omega$ tel que

$$\left| \begin{array}{l} P(\Omega') = 1 \\ \forall \omega \in \Omega', \quad X_n(\omega) \xrightarrow{n \rightarrow +\infty} Y(\omega) \end{array} \right.$$

On note cela : $(X_n) \xrightarrow{\text{ps}} Y$.

Remarque : la définition ne dit pas que, pour deux valeurs distinctes de ω , la suite numérique $(X_n(\omega))$ converge vers $Y(\omega)$ à la même «vitesse». Considérons l'expérience suivante : « tirer $X_1, X_2, \dots, X_n, \dots$ puis tirer Y ».

1. Si on réalise une cette expérience une première fois, on sait que (en tout cas la probabilité est 1), pour tout $\varepsilon > 0$, par exemple pour $\varepsilon = 10^{-2}$, il existe un rang n_1 à partir duquel $|X - X_n| \leq 10^{-2}$.
2. Si on recommence l'expérience une seconde fois, on sait aussi (la probabilité est 1) que, pour tout $\varepsilon > 0$, par exemple pour $\varepsilon = 10^{-2}$, il existe un rang n_2 à partir duquel $|X - X_n| \leq 10^{-2}$.

Mais peut-être que le n_2 du second tirage est très différent du n_1 du premier tirage.

Définition : convergence en probabilité.

La suite (X_n) converge en probabilité vers Y si, pour tout $\varepsilon > 0$,

$$P(|X_n - Y| > \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$$

ou, de manière équivalente,

$$P(|X_n - Y| \leq \varepsilon) \xrightarrow{n \rightarrow +\infty} 1$$

On note cela : $(X_n) \xrightarrow{\text{prob}} Y$.

Définition : convergence L_2 .

La suite (X_n) converge vers Y au sens L_2 si $E((X_n - Y)^2) \xrightarrow{n \rightarrow +\infty} 0$

On note cela : $(X_n) \xrightarrow{L_2} Y$.

Ces trois premiers type de convergence impliquent tous, d'une façon ou d'une autre, que quand on tire les X_n et Y , les valeurs obtenue pour les X_n sont souvent proches de la valeur obtenue pour Y .

Un quatrième type de convergence est à mettre à part.

Définition : convergence en distribution.

La suite X_n converge en distribution vers Y si, pour tout intervalle I de $\mathbb{R}$,

$$P(X_n \in I) \xrightarrow{n \rightarrow +\infty} P(Y \in I)$$

Cela est équivalent à la même propriété satisfaite pour tout intervalle $I = [a, b]$:

$$\forall [a, b] \subset \mathbb{R}, \quad P(X_n \in [a, b]) \xrightarrow{n \rightarrow +\infty} P(Y \in [a, b])$$

On note cela : $(X_n) \xrightarrow{d} Y$.

Remarque : contrairement aux autres types de convergence, celle-ci n'implique pas que X_n soit proche de Y quand n est grand. Elle veut simplement dire que la distribution de X_n peut être approchée par celle de Y .

Dans la pratique, savoir que $(X_n) \xrightarrow{d} Y$ a surtout l'intérêt de simplifier des calculs : si la loi de X_n est compliquée à calculer, il devient intéressant de l'estimer par la loi de Y , du moins si cette dernière est simple. Le théorème central limite, vu plus loin, est un exemple très utilisé de convergence en distribution.

2.2 Quelques exemples

Le premier exemple permet de passer en revue tous les types de convergence.

Exemple 1 : dans ce cours, on a souvent considéré une variable aléatoire X continue de densité f , qu'on «approximait» par une variable aléatoire X_n (on notait X') définie de la façon suivante :

1. On se donne $\delta x = \frac{1}{n}$.
2. L'ensemble des valeurs possibles de X_n est $X_n(\Omega) = \{k \delta x, \quad k \in \mathbb{Z}\}$
3. Les probabilités de ces valeurs possibles sont, pour tout $x_k = k \delta x$:

$$P(X_n = x_k) = P(X \in [x_k, x_k + \delta x]) = \int_{x_k}^{x_k + \delta x} f(x) dx \simeq f(x_k) \delta x$$

Imaginons qu'on commence par tirer X , puis qu'on déduise X_n en posant :

$$X_n = x_k \quad \text{où } k \text{ est l'unique valeur telle que } x_k \leq X < x_k + \delta x$$

Dans ce cas, on a toujours $|X_n - X| < \delta x$, donc $|X_n - X| < \frac{1}{n}$. Ainsi :

1. $(X_n) \xrightarrow{\text{ps}} X$ car quelque soit le tirage de X et les valeurs de X_n définies $n \in \mathbb{N}^*$ par X ,

$$|X_n(\omega) - X(\omega)| < \frac{1}{n} \implies X_n(\omega) \xrightarrow{n \rightarrow +\infty} X(\omega)$$

(il faut voir ω comme le tirage particulier qu'on vient de faire).

2. $(X_n) \xrightarrow{\text{prob}} X$ car pour tout $\varepsilon > 0$,

$$P(|X_n - X| \leq \varepsilon) \xrightarrow{n \rightarrow +\infty} 1$$

En effet, dès que $n > \frac{1}{\varepsilon}$,

$$|X_n - X| < \frac{1}{n} < \varepsilon$$

donc

$$P(|X_n - X| \leq \varepsilon) = 1$$

3. $(X_n) \xrightarrow{L_2} X$. En effet, comme X_n s'exprime comme une fonction $\varphi(X)$, on a

$$E((X_n - X)^2) = \int_{-\infty}^{+\infty} (\varphi(x) - x)^2 f(x) dx$$

où f est la densité de X . Or, comme pour tout $x \in \mathbb{R}$, $|\varphi(x) - x| \leq \frac{1}{n}$, on déduit

$$E((X_n - X)^2) \leq \int_{-\infty}^{+\infty} \frac{1}{n^2} f(x) dx = \frac{1}{n^2} \int_{-\infty}^{+\infty} f(x) dx = \frac{1}{n^2}$$

D'où $E((X_n - X)^2) \xrightarrow{n \rightarrow +\infty} 0$.

4. $(X_n) \xrightarrow{d} X$. Nous allons au moins le montrer dans le cas où f est continue : considérons un intervalle $[a, b] \subset \mathbb{R}$ et montrons que

$$P(X_n \in [a, b]) \xrightarrow{n \rightarrow +\infty} P(X \in [a, b])$$

Le terme de droite vaut $\int_a^b f(x) dx$.

Pour celui de gauche : fixons $n \in \mathbb{N}^*$. Définissons

$$k_1 \text{ tel que } x_{k_1-1} \leq a < x_{k_1} \quad \text{et} \quad k_2 \text{ tel que } x_{k_2} \leq b < x_{k_2+1}$$

Alors

$$P(X_n \in [a, b]) = \underbrace{P(X_n = x_{k_1})}_{P(X \in [x_{k_1}, x_{k_1+1}])} + \underbrace{P(X_n = x_{k_1+1})}_{P(X \in [x_{k_1+1}, x_{k_1+2}])} + \cdots + \underbrace{P(X_n = x_{k_2})}_{P(X \in [x_{k_2}, x_{k_2+1}])} = P(X \in [x_{k_1}, x_{k_2+1}])$$

Ainsi, comme $a < x_{k_1}$ et $b < x_{k_2+1}$, on a

$$P(X_n \in [a, b]) - P(X \in [a, b]) = P(X \in [b, x_{k_2+1}]) - P(X \in [a, x_{k_1}])$$

Donc

$$\begin{aligned} -P(X \in [a, a + \delta x]) &\leq -P(X \in [a, x_{k_1}]) \leq P(X_n \in [a, b]) - P(X \in [a, b]) \\ &\leq P(X \in [b, x_{k_2+1}]) \leq P(X \in [b, b + \delta x]) \end{aligned}$$

On a un encadrement de $P(X_n \in [a, b]) - P(X \in [a, b])$ par deux termes.

Or chacun d'entre eux tend vers 0. Comme la densité f est continue sur $[a, b+1]$, elle y est bornée.

Appelons M un majorant. On a par alors

$$0 \leq P(X \in [b, b + \delta x]) \leq \int_b^{b+\delta x} M \, dx = M\delta x$$

(car $[b, b + \delta x] \subset [a, b + 1]$). Donc

$$0 \leq P(X \in [b, b + \delta x]) \leq \frac{M}{n} \implies P(X \in [b, b + \delta x]) \xrightarrow{n \rightarrow +\infty} 0$$

De même, on montre que

$$0 \leq P(X \in [a, a + \delta x]) \leq \frac{M}{n} \implies P(X \in [a, a + \delta x]) \xrightarrow{n \rightarrow +\infty} 0$$

Cela montre que $P(X_n \in [a, b]) - P(X \in [a, b]) \xrightarrow{n \rightarrow +\infty} 0$.

Comme ceci est vrai pour tout intervalle $[a, b] \subset \mathbb{R}$, on en déduit que $(X_n) \xrightarrow{d} X$.

Il faut absolument connaître l'exemple 2. Il apparait très fréquemment dans la vie, que ce soit la vie d'ingénieur ou la vie d'étudiant (c'est-à-dire en examen).

Exemple 2 : soient $X_1, X_2, \dots, X_n, \dots$ des variables aléatoires indépendantes, identiquement distribuées (i.i.d.), admettant une espérance et une variance. Notons

$$m = E(X_n) \quad \text{et} \quad \sigma^2 = \text{Var}(X_n)$$

Définissons maintenant la variable

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

D'une part, on a

$$E(\bar{X}_n) = \frac{1}{n} \times (E(x_1) + \dots + E(X_n)) = \frac{1}{n} \times nm = m$$

D'autre part, comme les X_i sont indépendantes, leurs variances s'ajoutent et

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \times (\text{Var}(x_1) + \dots + \text{Var}(X_n)) = \frac{1}{n^2} \times n\sigma^2 = \frac{\sigma^2}{n}$$

$$1. \quad (\bar{X}_n) \xrightarrow{L_2} m$$

En effet, comme $E(\bar{X}_n) = m$, on en déduit que

$$E((\bar{X}_n - m)^2) = \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

Donc

$$E((\bar{X}_n - m)^2) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow +\infty} 0$$

ce qui est la définition de $(\bar{X}_n) \xrightarrow{L_2} m$.

2. $(\bar{X}_n) \xrightarrow{\text{prob}} m$

Il nous faut montrer que, pour tout $\varepsilon > 0$, $P(|\bar{X}_n - m| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$.

Pour cela, nous allons utiliser l'inégalité de Tchebychev : pour une variable X admettant une espérance et une variance, on a

$$\forall \varepsilon > 0, \quad P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}$$

Dans le cas de la variable $\bar{X}_n$, dont l'espérance est m et la variance $\frac{\sigma^2}{n}$, on obtient

$$\forall \varepsilon > 0, \quad P(|\bar{X}_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n \varepsilon^2}$$

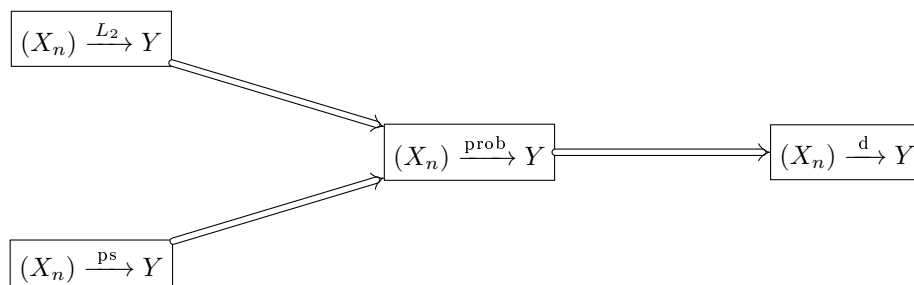
Ainsi, si on fixe $\varepsilon > 0$, on aura pour tout $n \in \mathbb{N}^*$

$$0 \leq P(|\bar{X}_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n \varepsilon^2}$$

ce qui implique que $P(|\bar{X}_n - m| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$, puisque $\frac{\sigma^2}{n \varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0$.

2.3 Relations entre ces types de convergence

Certains des modes de convergence sont plus forts que d'autres, c'est-à-dire qu'il y a des relations d'implications. Celles-ci sont résumées dans le schéma suivant.



2.4 Théorèmes de convergence

Les principaux théorèmes de convergence concernent tous une situation déjà vue ci-dessus :

1. On considère des variables $X_1, X_2, \dots, X_n, \dots$ qui ont toutes la même distribution. On suppose de plus qu'elles ont une espérance $E(X_1) = E(X_2) = \dots$
2. On définit la moyenne $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$.
3. On étudie la convergence de la suite $(\bar{X}_n)$ vers la variable constante $E(X_1)$.

Théorème : loi faible des grands nombres.

Si les deux conditions suivantes sont vérifiées :

C1 : Les variables X_i admettent une variance.

C2 : Les variables X_i sont non corrélées deux à deux, ce qui veut dire que

$$\forall (i, j \in \mathbb{N}^2, i \neq j, \quad \mathbb{E} \left((X_i - \mathbb{E}(X_i)) (X_j - \mathbb{E}(X_j)) \right) = 0$$

Alors $(\bar{X}_n) \xrightarrow{\text{prob}} \mathbb{E}(X_1)$.

Remarque : la condition C2 est toujours satisfaite si les X_i sont indépendantes deux à deux. L'hypothèse est donc moins forte que l'indépendance de ces variables et la loi faible des grands nombres peut donc être invoquée dans cas où les autres théorèmes de convergence ne sont pas applicables.

Théorème : loi forte des grands nombres.

Si les deux conditions suivantes sont vérifiées :

C1 : Les variables $|X_i|$ admettent une espérance.

C2 : Les variables X_i sont indépendantes deux à deux.

Alors $(\bar{X}_n) \xrightarrow{\text{ps}} \mathbb{E}(X_1)$.

Remarque : la condition C1 est toujours satisfaite si les X_i admettent une variance. L'hypothèse est donc moins forte que l'existence d'une variance et la loi forte des grands nombres peut donc être invoquée dans cas où les autres théorèmes de convergence ne sont pas applicables. De plus, sa conclusion est plus forte.

Enfin, le théorème central limite permet d'approcher la distribution de $\bar{X}_n$.

Théorème central limite

Si les deux conditions suivantes sont vérifiées :

C1 : Les variables X_i admettent une variance, notée σ^2 .

C2 : Les variables X_i sont indépendantes deux à deux.

On peut définir $Z_n = \frac{\bar{X}_n - \mathbb{E}(X_1)}{\sigma/\sqrt{n}}$ et $Z \rightsquigarrow \mathcal{N}(0, 1)$. On a alors :

$$(Z_n) \xrightarrow{d} Z$$

Remarques

1. On sait que $\mathbb{E}(\bar{X}_n) = \mathbb{E}(X_1)$ et que $\text{Var}(\bar{X}_n) = \frac{\text{Var}(X_1)}{n} = \frac{\sigma^2}{n}$. On peut donc réécrire S_n sous la forme :

$$Z_n = \frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sigma(\bar{X}_n)}$$

Ainsi,

$$\mathbb{E}(Z_n) = \frac{1}{\sigma(\bar{X}_n)} \left(\mathbb{E}(\bar{X}_n) - \mathbb{E}(\bar{X}_n) \right) = 0 \quad \text{et} \quad \text{Var}(Z_n) = \frac{1}{\sigma^2(\bar{X}_n)} \text{Var}(\bar{X}_n) = 1$$

Finalement, Z_n est une version «normalisée» de $\bar{X}_n$, c'est-à-dire que c'est la variable $\bar{X}_n$ transformée par homothétie-translation de façon à avoir une espérance nulle et une variance unité.

- Un aspect surprenant du théorème central limite est que sa conclusion ne dépend pas de la distribution des X_i . Ceci explique que la loi normale apparaisse dans beaucoup de variables de la vie réelle.
- Ce théorème est souvent utilisé quand on étudie la convergence en probabilité de $(\bar{X}_n)$ vers $E(X_1)$. En effet, pour tout $\varepsilon > 0$,

$$|\bar{X}_n - E(X_1)| < \varepsilon \iff -\varepsilon < \bar{X}_n - E(X_1) < \varepsilon \iff -\frac{\varepsilon\sqrt{n}}{\sigma} < Z_n < \frac{\varepsilon\sqrt{n}}{\sigma}$$

On a donc

$$P\left(|\bar{X}_n - E(X_1)| < \varepsilon\right) = P\left(-\frac{\varepsilon\sqrt{n}}{\sigma} < Z_n < \frac{\varepsilon\sqrt{n}}{\sigma}\right)$$

Le théorème permet d'approximer cette probabilité par

$$P\left(-\frac{\varepsilon\sqrt{n}}{\sigma} < Z < \frac{\varepsilon\sqrt{n}}{\sigma}\right) = \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\varepsilon\sqrt{n}}{\sigma}}^{\frac{\varepsilon\sqrt{n}}{\sigma}} e^{-\frac{x^2}{2}} dx$$

où Φ est la fonction de répartition de la loi normale centrée réduite. Cette probabilité tend bien vers 1 quand n tend vers $+\infty$, car les deux bornes tendent vers $-\infty$ et $+\infty$.

De même, on sait que $P(-1.96 \leq Z \leq 1.96) = \Phi(1.96) - \Phi(-1.96) = 0.95$, c'est une valeur particulière de la loi normale centrée réduite. Or

$$-1.96 < Z_n < 1.96 \iff -1.96 \frac{\sigma}{\sqrt{n}} < \bar{X}_n - E(X_1) < 1.96 \frac{\sigma}{\sqrt{n}}$$

Ainsi, le théorème centrale limite permet d'affirmer que

$$P\left(E(X_1) - 1.96 \frac{\sigma}{\sqrt{n}} < \bar{X}_n < E(X_1) + 1.96 \frac{\sigma}{\sqrt{n}}\right) \simeq 0.95$$

On obtient un «intervalle de prédiction» pour $\bar{X}_n$ à 95%, centré sur $E(x_1)$ et d'autant plus étroit que n est grand.

Et dans la pratique, on pourra remplacer 1.96 par 2.

2.5 Exemples d'applications du théorème central limite

Exemple 1 : loi binomiale

Soit $p \in]0, 1[$. Considérons des variables de Bernoulli $X_1, X_2, \dots$, indépendantes de paramètres p : pour tout $n \in \mathbb{N}$, X_n vaut 0 ou 1 avec

$$P(X_n=1) = p \quad \text{et} \quad P(X_n=0) = 1 - p$$

On a pour tout $n \in \mathbb{N}^*$,

$$E(X_n) = p \quad \text{et} \quad \text{Var}(X_n) = p(1 - p)$$

Définissons les variables aléatoires

$$S_n = X_1 + \dots + X_n \quad \text{et} \quad \bar{X}_n = \frac{S_n}{n}$$

On sait que S_n suit une loi binomiale : $S_n \rightsquigarrow B(n, p)$.

De plus, on peut appliquer le théorème central limite à la suite $(\bar{X}_n)$: on pose

$$Z_n = \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{S_n - np}{\sqrt{np(1-p)}}$$

Quand n est grand (dans la pratique, $n \geq 30$), on peut approcher $P(Z_n \in [a, b])$ par $P(Z \in [a, b])$ où $Z \rightsquigarrow \mathcal{N}(0, 1)$ (voir figure 5).

Par exemple,

$$P(-1.96 < Z_n < 1.96) \simeq 0.95 \implies P\left(p - 1.96\sqrt{\frac{p(1-p)}{n}} < \bar{X}_n < p + 1.96\sqrt{\frac{p(1-p)}{n}}\right) \simeq 0.95$$

Exemple 2

La ville de Venise s'enfonce peu à peu, à moins que ce soit le niveau de la mer qui monte. On suppose que, chaque année, la baisse d'altitude de la ville exprimée en millimètres est δX telle que

$$E(\delta X) = 1 \quad \text{'et'} \quad \text{Var}(\delta X) = \frac{1}{9}$$

On suppose de plus que les changements d'altitude sont indépendants d'une année à l'autre et sont identiquement distribués.

Déterminons un intervalle de prédiction à 95% de la baisse totale au bout de 100 ans.

1. Pour une année n , on définit δX_n comme la baisse de niveau cette année-là. D'après les hypothèses, les δX_n sont i.i.d., d'espérance 1 et de variance $\frac{1}{9}$.
2. La baisse de niveau au bout de 100 ans est

$$S_{100} = \delta X_1 + \delta X_2 + \cdots + \delta X_{100}$$

On sait que

$$E(S_{100}) = 100 E(\delta X) = 100 \quad \text{et} \quad \text{Var}(S_{100}) = 100 \text{Var}(\delta X) = \frac{100}{9}$$

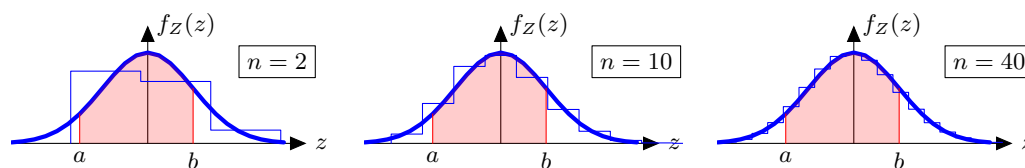


FIGURE 5 – Théorème central limite: convergence en distribution de (Z_n) vers Z . En trait bleu épais, la densité de Z . L'aire de la zone en rouge est la probabilité que $Z \in [a, b]$. Les variables X_i suivent une loi de Bernoulli de paramètre $p = 0.3$. Pour une valeur n donnée, $S_n = X_1 + \cdots + X_n$ prend les valeurs entières entre 1 et n . Donc $Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}$ prend des valeurs espacées de $\frac{1}{\sqrt{np(1-p)}}$, allant de $-\sqrt{\frac{np}{1-p}}$ à $+\sqrt{\frac{n(1-p)}{p}}$. La probabilité de $Z_n \in [a, b]$ est l'aire située entre le segment $[a, b]$ sur l'axe des abscisses et la courbe en escalier, en traits bleu fin. De gauche à droite : $n = 2$, $n = 10$ et $n = 40$.

Définissons

$$Z_{100} = \frac{S_{100} - E(S_{100})}{\sigma(S_{100})} = \frac{S_{100} - 100}{10/3}$$

D'après le théorème central limite, comme $(Z_n) \xrightarrow{d} Z$ où $Z \rightsquigarrow \mathcal{N}(0, 1)$, on peut considérer que

$$P(-1.96 \leq Z_{100} \leq 1.96) \simeq P(-1.96 \leq Z \leq 1.96) = 0.95$$

Donc

$$P\left(-1.96 \leq \frac{S_{100} - 100}{10/3} \leq +1.96\right) \simeq 0.95$$

Or

$$-1.96 \leq \frac{S_{100} - 100}{10/3} \leq +1.96 \iff 100 - 1.96 \frac{10}{3} \leq S_{100} \leq 100 + 1.96 \frac{10}{3}$$

Finalement, on obtient

$$P(S_{100} \in [93.5; 106.5]) \simeq 0.95$$