

Probabilités continues

Support du cours PROC

Guillaume Euvrard, Février 2021
EPITA

Table des matières

1	Motivation	3
2	Dénombrable ou pas, qu'est-ce que ça change ?	4
2.1	Retour sur le cas fini ou dénombrable	4
2.2	Cas d'un intervalle de $\mathbb{R}$	5
3	Tribu borélienne et fonction de répartition	5
3.1	Un peu de formalisme : espace probabilisé	5
3.2	Tribu borélienne	6
3.3	Conséquence pratique	8
4	Variable continue à densité	8
4.1	Définition d'une densité	8
4.2	Propriétés d'une densité	9
4.3	Compréhension intuitive de la densité	10
4.4	Densité de $\varphi(X)$	11
5	Exemples de loi à densité	12
5.1	Loi uniforme	13
5.2	Loi exponentielle	13
5.3	Loi normale	14
6	Espérance et variance d'une variable aléatoire à densité	16
6.1	Espérance de X	16
6.2	Variance et écart-type de X	17
7	Exemples de calculs de $E(X)$ et de $\text{Var}(X)$	18
7.1	Loi uniforme	18
7.2	Loi exponentielle	19
7.3	Loi normale	21
8	Lois à deux variables	22
8.1	Cas général	22
8.2	Densité jointe de deux variables aléatoires	23
8.3	Interprétation intuitive d'une densité	24
8.4	Densités déduites de la densité conjointe	25
8.5	Variables indépendantes	25
9	Suite de variables aléatoires	27
9.1	Les différents types de convergence	28
9.2	Quelques exemples	29
9.3	Relations entre ces types de convergence	32
9.4	Théorèmes de convergence	32
9.5	Exemples d'applications du théorème central limite	34

Rappels :

1. Étant donnée une expérience aléatoire, on appelle Ω l'ensemble de ses résultats possibles.
2. Une variable aléatoire X est une application $X : \Omega \longrightarrow \mathbb{R}$.
3. L'ensemble $X(\Omega)$ est l'image de Ω par l'application X . C'est l'ensemble de toutes les valeurs possibles prises par $X(\omega)$.
4. La loi de X est une fonction

$$\begin{aligned} P_X : \mathcal{P}(\mathbb{R}) &\longrightarrow \mathbb{R} \\ A \subset \mathbb{R} &\longmapsto P(X \in A) \end{aligned}$$

Pour tout $A \subset \mathbb{R}$, on note $P_X(A)$ ou $P(X \in A)$ la probabilité qu'une réalisation de X soit dans A .

5. Dans certains cas, on connaît explicitement Ω et la fonction de probabilités sur $\mathcal{P}(\Omega)$. On en déduit alors la loi de X .
6. Très souvent cependant, on ne connaît ni Ω ni la fonction de probabilité sur $\mathcal{P}(\Omega)$. L'ensemble Ω contient toutes les sources d'aléas possibles et une observation de X n'est qu'une valeur réelle, fonction de tous ces aléas. On peut néanmoins définir la loi P_X .

Convention de notation : on note $X, Y, Z \dots$ avec des lettres majuscules les variables aléatoires. Les lettres minuscules $x, y, z \dots$ sont réservées à des valeurs particulières que peuvent prendre ces variables, elles désignent donc des nombres réels.

1 Motivation

Les variables aléatoires étudiées jusqu'ici sont des variables «discrètes». C'est-à-dire que l'ensemble des valeurs possibles est soit fini :

$$X(\Omega) = \{x_1, x_2, \dots, x_N\} = \{x_n, n \in \llbracket 1, N \rrbracket\}$$

soit infini, mais dénombrable :

$$X(\Omega) = \{x_0, x_1, x_2, \dots\} = \{x_n, n \in \mathbb{N}\}$$

Cependant, de nombreuses variables apparaissant dans la vraie vie ne sont pas discrètes. Souvent, l'ensemble des valeurs possibles est un intervalle de $\mathbb{R}$, voire $\mathbb{R}$ tout entier. Or un intervalle n'est pas dénombrable : il contient beaucoup plus d'éléments que $\mathbb{N}$. Ces variables aléatoires ne peuvent donc pas être décrites de la même manière. Citons par exemple :

- Temps d'attente dans une file d'attente.
- Durée de vie d'un appareil avant usure.
- Taux de fécondité futur dans une population.
- Mesure d'une variable sanitaire (rythme cardiaque, taux de globules dans le sang, ...) sur un individu tiré au hasard.
- Niveau de radioactivité dans l'air à un instant donné.
- etc.

Il est pourtant nécessaire de pouvoir définir et manipuler les lois de telles variables. Dans les situations évoquées ci-dessus, cela permettrait par exemple de calibrer une file d'attente, de conduire une politique d'achats d'appareils, de faire des projections démographiques, d'établir un diagnostic médical ou encore de détecter un niveau de radioactivité anormalement élevé.

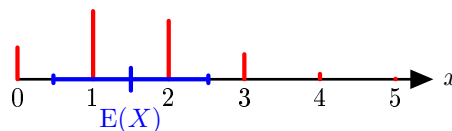


FIGURE 1 – loi d’une variable aléatoire finie. Ici, $X(\Omega) = \llbracket 0, 5 \rrbracket$. L’axe des abscisses représente les valeurs possibles de X , l’axe des ordonnées leurs probabilités. À chaque valeur possible, on place un trait rouge dont la hauteur est la probabilité que X prenne cette valeur. On ajoute l’espérance de X , ainsi qu’un segment horizontal reliant les points d’abscisses $E(X) - \sigma$ et $E(X) + \sigma$ (σ est l’écart-type de X). La longueur de ce segment (2σ) représente la dispersion de la loi autour de l’espérance.

2 Dénombrable ou pas, qu’est-ce que ça change ?

Une difficulté pratique survient quand on veut définir la loi d’une variable aléatoire X : il faut définir $P_X(A)$ pour tout $A \subset \mathbb{R}$. Autrement dit, il faut un procédé qui, à partir d’un ensemble A , permette de calculer numériquement la probabilité. Comment définir une «formule numérique» dont la variable est non pas un réel $x \in \mathbb{R}$ mais un ensemble $A \subset \mathbb{R}$?

2.1 Retour sur le cas fini ou dénombrable

Supposons que $X(\Omega)$ est un ensemble fini ou infini dénombrable :

$$X(\Omega) = \{x_n, n \in K\} \quad \text{avec} \quad K = \llbracket 1, N \rrbracket \text{ (cas fini)} \quad \text{ou} \quad K = \mathbb{N} \text{ (cas dénombrable)}$$

Alors il suffit de définir les nombres $P_X(\{x_n\}) = P(X=x_n)$ pour tout $n \in K$.

En effet, pour tout $A \subset \mathbb{R}$, on peut calculer numériquement $P_X(A)$ de la façon suivante :

1. On cherche l’ensemble $A' = A \cap X(\Omega)$
2. On définit $P_X(A) = \sum_{x_n \in A'} P_X(\{x_n\}) = \sum_{x_n \in A'} P(X=x_n)$

Par exemple, si $X(\Omega) = \{-1, 0, 1\}$ et qu’on doit calculer $P_X([0, 2])$:

1. On définit $A' = [0, 2] \cap \{-1, 0, 1\} = \{0, 1\}$
2. On déduit $P_X([0, 2]) = P(X=0) + P(X=1)$

Si l’ensemble A' est infini, la somme ci-dessus est une série numérique et la probabilité est la limite de la série.

Notons toutefois que les nombres $P(X=x_n)$ ne peuvent pas être choisis n’importe comment : il faut d’une part qu’ils soient positifs, d’autre part que

$$\sum_{n \in K} P(X=x_n) = 1$$

Dans le cas où $X(\Omega)$ est infini, alors $K = \mathbb{N}$ et la somme ci-dessus est la limite d’une série numérique. Cette dernière doit donc converger et sa limite doit valoir 1. Quand c’est la cas, la série qui définit $P_X(A)$ est toujours convergente, car constituée de termes positifs et majorée par 1.

On dit que les ensembles $\{x_n\}$ engendrent la loi. La connaissance de leurs probabilités permet de définir toute la loi.

On peut ensuite faire des constructions plus complexes, comme définir l'espérance et la variance de X :

$$E(X) = \sum_{n \in K} x_n P(X=x_n) \quad \text{et} \quad \text{Var}(X) = E\left((X - E(X))^2\right) = \sum_{n \in K} (x_n - E(X))^2 P(X=x_n)$$

Si l'ensemble $X(\Omega)$ est infini, les sommes ci-dessus sont des séries numériques. Si elles divergent, la variable X n'a pas d'espérance ou pas de variance.

2.2 Cas d'un intervalle de $\mathbb{R}$

Un intervalle $I \subset \mathbb{R}$ est infini non dénombrable. Sauf dans le cas d'un singleton de la forme $[a, a]$, il n'est pas possible de trouver une suite (x_n) telle que

$$I = \{x_n, n \in \mathbb{N}\}$$

Ce résultat, établi par Cantor en 1874 (voir par exemple [ici](#) pour sa démonstration par «l'argument de la diagonale») a de nombreuses conséquences en mathématiques.

En ce qui nous concerne, si l'ensemble $X(\Omega)$ des valeurs possibles d'une variable aléatoire X est un intervalle I , on ne peut pas définir $P_X(A)$ comme on le fait dans le cas fini ou dénombrable.

Par exemple, on sait bien que $P_X(I) = P(X \in I) = 1$, mais on ne peut pas traduire cette égalité par une relation de la forme

$$\sum_{n=0}^{+\infty} P(X=x_n) = 1$$

car la partie gauche de l'égalité ne peut pas être $P(X \in I)$.

3 Tribu borélienne et fonction de répartition

On souhaite définir la loi de X dans le cas où $X(\Omega)$ est un intervalle, donc non dénombrable. L'idée de base est de définir $P_X(A)$ pour certains ensembles $A \subset \mathbb{R}$, mais pas pour tous.

3.1 Un peu de formalisme : espace probabilisé

Soit $\Sigma \subset \mathcal{P}(\mathbb{R})$. Cet ensemble Σ contient des sous-ensembles de $\mathbb{R}$. Dans la pratique, ce sera l'ensemble des $A \subset \mathbb{R}$ pour lesquels $P_X(A)$ existe.

Définition : cet ensemble Σ est une tribu sur $\mathbb{R}$ si :

1. $\mathbb{R} \in \Sigma$
2. pour tout $A \in \Sigma$, son complémentaire $\bar{A}$ est aussi dans Σ
3. pour toute famille finie ou dénombrable $(A_n)_{n \in K}$ (avec K de la forme $\llbracket 1, N \rrbracket$ ou $K = \mathbb{N}$) d'éléments de Σ , $\bigcup_{n \in K} A_n \in \Sigma$

Quand on s'est donné une tribu, on peut définir une fonction de probabilité P_X .

Définition : étant donnée une tribu Σ , une fonction de probabilité sur Σ est une fonction $P_X : \Sigma \longrightarrow [0, 1]$ telle que

1. $P_X(\mathbb{R}) = 1$

2. pour toute famille finie ou dénombrable **disjointe** $(A_n)_{n \in K}$ (avec K de la forme $\llbracket 1, N \rrbracket$ ou $K = \mathbb{N}$) d'éléments de Σ ,

$$P_X \left(\bigcup_{n \in K} A_n \right) = \sum_{n \in K} P_X(A_n)$$

On dit que le triplet $(\mathbb{R}, \Sigma, P_X)$ est un «espace probabilisé».

Exemple 1 : soit l'intervalle $A =]-\infty, 1]$. On peut définir la tribu

$$\Sigma = \{\emptyset, A, \overline{A}, \mathbb{R}\} = \{\emptyset,]-\infty, 1],]1, +\infty[, \mathbb{R}\}$$

On remarque d'ailleurs qu'il suffit de définir la seule probabilité $P_X(A)$ pour définir toute la loi sur Σ . En effet, dès lors que cette probabilité est entre 0 et 1, on a

$$P_X(\emptyset) = 0; \quad P_X(\overline{A}) = 1 - P_X(A) \quad \text{et} \quad P_X(\mathbb{R}) = 1$$

On dit que A engendre la tribu Σ .

Exemple 2 : soient les intervalles $A =]-\infty, 1]$ et $B =]-\infty, 3]$. La tribu engendrée par A et B est

$$\Sigma = \{ \emptyset, A, B,]1, 3],]1, +\infty[,]3, +\infty[,]-\infty, 1] \cup]3, +\infty[\}$$

Là aussi, on remarque que les seules probabilités $P_X(A)$ et $P_X(B)$ suffisent à définir toute la loi sur Σ . Par exemple,

$$P_X(]3, +\infty[) = 1 - P_X(B), \quad P_X(]-\infty, 1] \cup]3, +\infty[) = P_X(A) + (1 - P_X(B)), \quad P_X(]1, 3]) = P_X(B) - P_X(A)$$

Il faut néanmoins que ces deux nombres vérifient les inégalités

$$0 \leq P_X(A) \leq P_X(B) \leq 1$$

pour que chaque élément de Σ ait une probabilité positive.

3.2 Tribu borélienne

Définition : la tribu borélienne est la tribu engendrée par les intervalles de la forme $A =]-\infty, x]$ ou x décrit $\mathbb{R}$.

Définition : soit X une variable aléatoire réelle. Sa fonction de répartition est la fonction

$$F : \begin{cases} \mathbb{R} & \longrightarrow \mathbb{R} \\ x & \longmapsto P_X(]-\infty, x]) \end{cases}$$

Autrement dit, pour tout $x \in \mathbb{R}$, $F(x)$ est la probabilité de l'évènement « $X \in]-\infty, x]$ », donc de l'évènement « $X \leq x$ »¹.

La tribu borélienne ne contient pas toutes les parties A de $\mathbb{R}$. Ainsi, la fonction de répartition F ne permet pas de déduire $P_X(A)$ pour tout $A \subset \mathbb{R}$. Elle contient néanmoins les ensembles qui nous intéressent dans

1. Parfois, on note F_X cette fonction de répartition, pour spécifier qu'il s'agit de la loi de la variable X . Cependant, quand il n'y a pas d'ambiguïté sur la variable aléatoire, on note F pour alléger le texte.

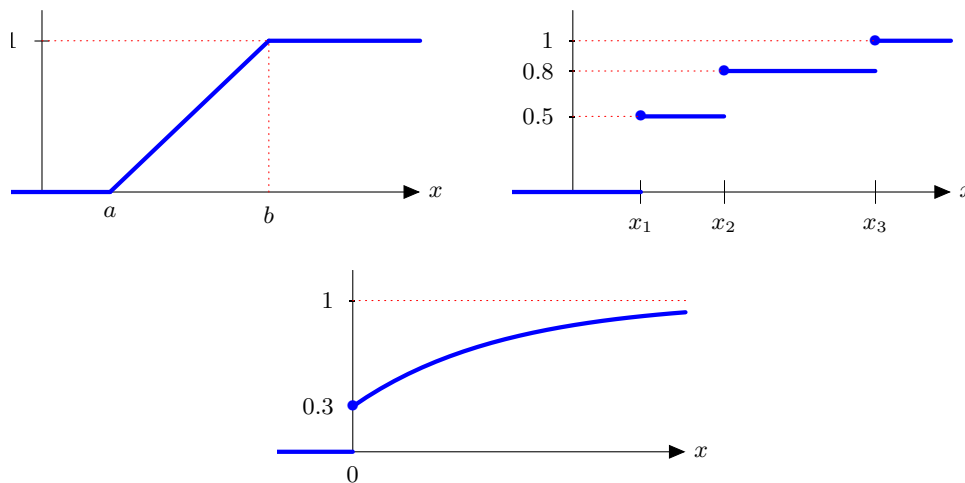


FIGURE 2 – Trois exemples de fonctions de répartition. **En haut à gauche** : cas d'une variable continue, qui prend toutes ses valeurs dans $[a, b]$. La probabilité de « $X \leq a$ » est $F(a) = 0$ et la probabilité de « $X > b$ » est $1 - F(b) = 0$. **En haut à droite** : cas d'une variable finie. Cette variable ne prend que trois valeurs : x_1 avec la probabilité 0.5, x_2 avec la probabilité 0.3 et x_3 avec la probabilité 0.2. **En bas** : fonction de répartition d'une variable «hybride». Cette variable ne peut être négative, elle vaut 0 avec la probabilité 0.3 puis a une loi continue pour les valeurs strictement positives. Imaginons par exemple que X soit la distance parcourue à pied par un individu tiré au hasard lors d'une journée tirée au hasard : la fonction F sera alors de cette forme.

la pratique, notamment tous les intervalles (ouverts ou fermés, bornés ou non, à droite comme à gauche). On a en effet la propriété suivante.

Propriété : soient X une variable aléatoire réelle, F sa fonction de répartition et Σ la tribu borélienne. Alors la fonction F est croissante avec

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

De plus, pour tout $(a, b) \in \mathbb{R}^2$ tel que $a < b$:

1. l'intervalle $]a, +\infty[$ est dans Σ et sa probabilité vaut

$$P_X([a, +\infty[) = P(X > a) = 1 - F(a)$$

2. l'intervalle $]a, b]$ est dans Σ et sa probabilité vaut

$$P_X([a, b]) = P(a < X \leq b) = F(b) - F(a)$$

3. l'intervalle ouvert à droite $] -\infty, a[$ est dans Σ et sa probabilité vaut

$$P_X([-\infty, a]) = P(X < a) = \lim_{x \rightarrow a^-} F(x)$$

4. De même, les intervalles $[a, +\infty[$, $[a, b]$, $[a, b[$ et $]a, b]$ sont tous dans Σ . Leurs probabilités sont

$$P_X([a, +\infty[) = 1 - \lim_{x \rightarrow a^-} F(x) \quad P_X([a, b]) = F(b) - \lim_{x \rightarrow a^-} F(x)$$

$$P_X([a, b]) = \lim_{x \rightarrow b^-} F(x) - \lim_{x \rightarrow a^-} F(x) \quad P_X([a, b]) = \lim_{x \rightarrow b^-} F(x) - F(a)$$

3.3 Conséquence pratique

Une conséquence de cette propriété est qu'une valeur précise a peut avoir une probabilité nulle sans pour autant être une valeur impossible. En effet, $]-\infty, a]$ est l'union disjointe $]-\infty, a[\cup \{a\}$. Ainsi,

$$P_X(\{a\}) = P_X(]-\infty, a]) - P_X(]-\infty, a[) = F(a) - \lim_{x \rightarrow a^-} F(x)$$

Si F est continue en a , on obtient $P_X(\{a\}) = P(X=a) = 0$.

De fait, les variables aléatoires qu'on étudiera dans la suite ont toutes une fonction de répartition continue. Ce cas de figure sera donc le cas habituel. Dans cette situation, la probabilité $P_X([a, b])$ ne change plus selon qu'on ouvre ou ferme l'intervalle $[a, b]$.

Interprétons ce phénomène sur un exemple : supposons qu'un utilisateur envoie une requête à un serveur et que X soit le temps de réponse (en secondes) du serveur. La probabilité que X soit *exactement* égal à 1 est nulle. Ce qui peut être non nul, c'est la probabilité que X vaille $1 \pm \delta t$ (où δt est petit), c'est-à-dire que X soit dans un intervalle $[1-\delta t, 1+\delta t]$. Mais il y a «trop» de valeurs dans cet intervalle pour que chacune d'elles, prise isolément, ait une probabilité non nulle.

En revanche, une variable aléatoire discrète a toujours une fonction de répartition discontinue (figure 2).

4 Variable continue à densité

Considérons une variable aléatoire X ayant une fonction de répartition F continue. S'il existe une fonction f telle que, pour tout $[a, b] \subset \mathbb{R}$,

$$F(b) - F(a) = \int_a^b f(x) dx$$

alors f est une densité² de X .

En fait, si la fonction de répartition F est dérivable, alors X admet une densité f qui n'est autre que la dérivée F' . Mais il arrive aussi que F soit dérivable par morceaux, c'est-à-dire qu'elle soit dérivable sur $\mathbb{R}$ sauf en un nombre fini de points (figure 3). Si elle admet en ces points une dérivée à gauche et une dérivée à droite, alors X admet là aussi une densité.

4.1 Définition d'une densité

Définition : soient une variable aléatoire X et F sa fonction de répartition. Une densité de X est une fonction f (si elle existe) telle que

$$\forall a \in \mathbb{R}, \quad F(a) = \int_{-\infty}^a f(x) dx$$

Remarque : si une densité existe, alors la fonction F est continue sur $\mathbb{R}$. De plus, elle est dérivable sauf éventuellement en un nombre fini de points. La densité f vaut alors $F'(x)$, sauf aux points où F n'est pas dérivable. En ces points-là, f peut valoir n'importe quoi. C'est pour cela qu'on parle d'*une* densité plutôt que de *la* densité.

2. Comme pour la fonction de répartition, on note parfois f_X au lieu de f , notamment quand plusieurs variables aléatoires sont en jeu, si elles ont des densités différentes.

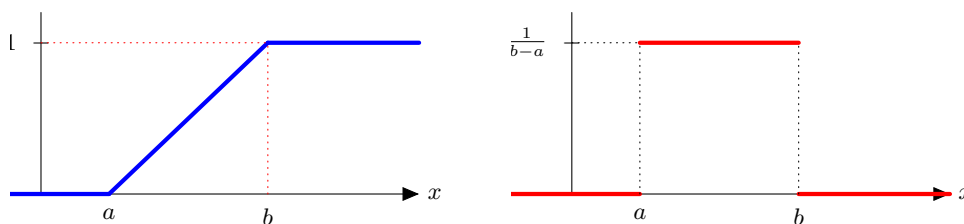


FIGURE 3 – Loi uniforme sur un intervalle $[a, b]$. **Courbe de gauche** : la fonction de répartition de X . Elle est dérivable sur $\mathbb{R}$, sauf en a et en b . Néanmoins, en ces deux points, elle admet une dérivée à gauche et une dérivée à droite. **Courbe de droite** : densité de X . Elle est égale à la dérivée F' , sauf en a et b . En ces deux points, elle peut prendre n'importe quelle valeur positive, c'est pourquoi on parle d'une densité plutôt que de la densité. Dans cet exemple, il est bien sûr naturel de lui donner en a et en b une des valeurs 0 ou $\frac{1}{b-a}$.

4.2 Propriétés d'une densité

Soit X une variable aléatoire admettant une densité f .

Loi de X : la fonction de répartition F est continue. Ainsi :

1. Pour tout $a \in \mathbb{R}$, $P(X=a) = F(a) - \lim_{x \rightarrow a^-} F(x) = 0$;
2. Pour tout $(a, b) \in \mathbb{R}^2$ tel que $a \leq b$, la probabilité que X soit dans $[a, b]$ ne dépend pas de si l'intervalle est ouvert ou fermé. Dans tous les cas, elle vaut

$$P_X([a, b]) = P(a \leq X \leq b) = \int_a^b f(x) dx$$

Propriétés vérifiées par la fonction f :

1. Pour tout $x \in \mathbb{R}$, $f(x) \geq 0$
2. L'intégrale $\int_{-\infty}^{+\infty} f(x) dx$ converge et vaut 1.

Inversement, si une fonction f vérifie ces deux dernières propriétés, elle peut être la densité d'une variable aléatoire X . Il suffit de définir la fonction de répartition

$$F : x \mapsto \int_{-\infty}^x f(t) dt$$

Puis pour construire une variable X de densité f , on peut par exemple :

- Tirer une variable Z uniformément sur $[0, 1]$. Cette variable suit la loi illustrée sur la figure 3. En particulier, pour $z \in [0, 1]$, $P(Z \leq z) = z$.
- Si la valeur observée de Z est z , on cherche $x \in \mathbb{R}$ telle que $F(x) = z$ et on affecte à X cette valeur x . En effet, en définissant X de cette sorte, comme F est croissante, on obtient

$$P(X \leq x) = P(Z \leq z) = z = F(x)$$

En particulier, si F est strictement monotone, alors elle est bijective de $\mathbb{R}$ dans $]0, 1[$, et $X = F^{-1}(Z)$.

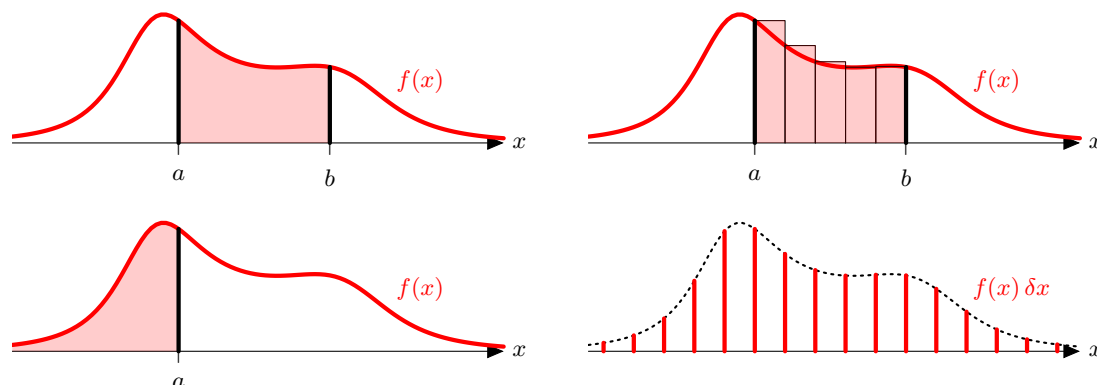


FIGURE 4 – Représentation géométrique et approximation numérique d’une intégrale. **En haut à gauche** : étant donné une fonction f (graphe en trait rouge épais), son intégrale sur un intervalle $[a, b]$ est l’aire représentée en rouge clair. **En haut à droite** : cette intégrale peut être approchée comme la somme des aires de ces rectangles. Chaque rectangle couvre en largeur un intervalle $[x_k, x_k + \delta x]$ et en hauteur $[0, f(x_k)]$. Son aire est donc $f(x_k) \delta x$. Cette approximation est d’autant meilleure que la largeur δx est petite. **En bas à gauche** : on suppose que la fonction f est une densité, c’est-à-dire qu’elle est positive ou nulle sur $\mathbb{R}$ et que son intégrale entre $-\infty$ et $+\infty$ converge et vaut 1. La fonction de répartition F d’une variable X de densité f est la fonction qui, à tout $a \in \mathbb{R}$, associe l’intégrale de f entre $-\infty$ et a . **En bas à droite** : on peut approcher la loi de X par celle d’une variable discrète X' prenant les valeurs $x_k = k \delta x$ ($k \in \mathbb{Z}$), chaque valeur ayant la probabilité $f(x_k) \delta x$. Dans ce dernier dessin, l’échelle en y a été modifiée d’un facteur δx , pour que les probabilités $f(x_k) \delta x$ apparaissent à la même hauteur que $f(x_k)$.

4.3 Compréhension intuitive de la densité

Pour bien saisir ce qu’est une densité, il faut déjà comprendre ce qu’est une intégrale. En effet, une densité f d’une variable aléatoire X est définie par une relation intégrale : pour tout $(a, b) \in \mathbb{R}^2$ tel que $a \leq b$,

$$P(X \leq a) = F(a) = \int_{-\infty}^a f(x) dx \quad \text{et} \quad P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x) dx$$

Encore faut-il savoir ce qu’on entend par «intégrale» :

1. Ce n’est pas une primitive de la fonction intégrée f .
2. C’est un nombre réel, défini comme l’aire algébrique entre le graphe de la fonction, l’axe des abscisses et les droites verticales d’équations $x = a$ et $x = b$.
3. Quand la fonction f est continue, un moyen pratique pour calculer cette valeur est de trouver une primitive F de la fonction f et de calculer $F(b) - F(a)$.
4. Il existe d’autres moyens pour la calculer. L’un d’entre eux servira de fil conducteur dans les raisonnements probabilistes à venir : on se donne un nombre δx petit et on divise l’intervalle $[a, b]$ en n petits intervalles $[x_k, x_k + \delta x]$, avec $x_k = a + k \delta x$. Une approximation de l’intégrale est

$$\int_a^b f(x) dx \approx \sum_{k=0}^n f(x_k) \delta x$$

Ce n'est pas une valeur exacte mais, quand δx tend vers 0, le terme de droite tend vers l'intégrale (figure 4). Ainsi, si on choisit δx suffisamment petit, le procédé donne une bonne approximation.

Il faut avoir en tête cette dernière méthode d'approximation pour raisonner avec des densités. En effet, il devient possible d'approximer X par une variable aléatoire discrète X' prenant les valeurs possibles $\{x_k, k \in \mathbb{Z}\}$ et dont les probabilités sont $P(X'=x_k) = f(x_k)\delta x$ (figure 4).

De nombreux résultats mathématiques utilisés dans ce module seront expliqués au regard de cette interprétation.

4.4 Densité de $\varphi(X)$

On est souvent amenés à définir une variable aléatoire en fonction d'une autre. Ainsi, si X est une variable aléatoire et φ une fonction réelle, on peut définir une nouvelle variable aléatoire $Y = \varphi(X)$ et chercher sa densité.

Théorème : densité de $\alpha X + \beta$. Soient X une variable aléatoire de densité f_X , $(\alpha, \beta) \in \mathbb{R}^2$ tel que $\alpha \neq 0$ et $Y = \alpha X + \beta$. Alors la densité de Y est la fonction f_Y définie pour tout $x \in \mathbb{R}$ par

$$f_Y(x) = \frac{1}{|\alpha|} \times f_X\left(\frac{x - \beta}{\alpha}\right)$$

En fait, il existe même un théorème plus général, donnant la densité de $\varphi(X)$ quand la fonction φ est strictement monotone.

Théorème : cas où φ est strictement monotone. Soient une variable aléatoire X de densité f_X , une fonction réelle φ dérivable et strictement monotone, et la variable aléatoire $Y = \varphi(X)$. Alors la densité de Y est la fonction f_Y définie pour tout $x \in \mathbb{R}$ par

$$f_Y(x) = \left| \frac{1}{\varphi'(\varphi^{-1}(x))} \right| \times f_X(\varphi^{-1}(x))$$

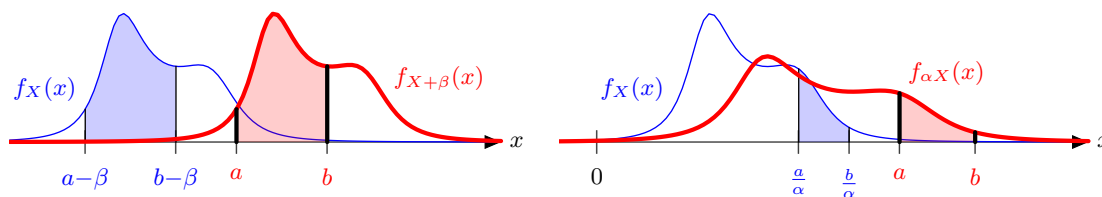


FIGURE 5 – Densité de $\alpha X + \beta$. **À gauche :** densités de X (en bleu) et de $X + \beta$ avec $\beta > 0$ (en rouge). La probabilité que $X + \beta$ soit compris entre a et b (aire en rouge) est la probabilité que X soit entre $a - \beta$ et $b - \beta$. Ainsi, $f_{X+\beta}(x) = f_X(x - \beta)$. **À droite :** densités de X (en bleu) et de αX (en rouge) avec $\alpha > 1$. La probabilité que αX soit entre a et b (aire en rouge) est la probabilité que X soit entre $\frac{a}{\alpha}$ et $\frac{b}{\alpha}$ (aire en bleue). Ainsi, $f_{\alpha X}(x) = \frac{1}{\alpha} f_X\left(\frac{x}{\alpha}\right)$. La zone en rouge est plus large que celle en bleu d'un facteur α , moins haute d'un facteur $\frac{1}{\alpha}$. Les deux zones ont la même aire.

Il est important de se souvenir du premier théorème sur la densité de $\alpha X + \beta$ et surtout de le comprendre graphiquement (figure 5). Mais dans la pratique, on retrouve ces résultats en déterminant la fonction de répartition $F_Y(x) = P(Y \leq x)$ et en la dérivant.

Exemples : soient X une variable aléatoire de densité f , et F sa fonction de répartition.

1. Retrouver la densité de $Y = -2X + 3$.

On sait que, pour tout $x \in \mathbb{R}$,

$$Y \leq x \iff -2X + 3 \leq x \iff X \geq \frac{-x + 3}{2}$$

Ainsi, la fonction de répartition de Y est

$$F_Y(x) = P(Y \leq x) = P\left(X \geq \frac{-x + 3}{2}\right) = 1 - F\left(\frac{-x + 3}{2}\right)$$

Donc la densité de Y est

$$f_Y(x) = F'_Y(x) = -F'\left(\frac{-x + 3}{2}\right) \times \left(\frac{-x + 3}{2}\right)' = \frac{1}{2} f\left(\frac{-x + 3}{2}\right)$$

2. Densité de $Y = X^2$

Pour tout $x \in \mathbb{R}$, l'évènement « $X^2 \leq x$ » est impossible si $x < 0$.

Sinon, $X^2 \leq x \iff x \in [-\sqrt{x}, \sqrt{x}]$.

Ainsi, la fonction de répartition de Y est

$$F_Y(x) = \begin{cases} 0 & \text{si } x < 0 \\ F(\sqrt{x}) - F(-\sqrt{x}) & \text{sinon} \end{cases}$$

En dérivant F_Y , on obtient la densité $f_Y(x)$: elle est nulle quand $x < 0$. Quand $x > 0$, elle vaut

$$f_Y(x) = F'_Y(x) = F'(\sqrt{x}) \times (\sqrt{x})' - F'(-\sqrt{x}) \times (-\sqrt{x})' = \frac{1}{2\sqrt{x}} [f(\sqrt{x}) + f(-\sqrt{x})]$$

On a un problème en $x = 0$ si cette dernière expression n'a pas de limite. Dans ce cas, on pourra donner à $f_Y(0)$ n'importe quelle valeur positive. La fonction f_Y ne sera pas continue, mais l'intégrale

$$\int_a^b f_Y(x) dx$$

sera de toute façon convergente pour tout $a \leq 0$ et $b \geq 0$. Elle vaudra $F_Y(b) - F_Y(a)$.

5 Exemples de loi à densité

Les lois continues usuelles sont les lois uniformes, exponentielles et normales.

5.1 Loi uniforme

Définition : soit un intervalle $[a, b] \subset \mathbb{R}$. On dit qu'une variable aléatoire X suit une loi uniforme sur $[a, b]$ si elle admet pour densité la fonction f définie par :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

On note cela : $X \rightsquigarrow \text{Unif}(a, b)$.

Cette fonction f est bien une densité : elle est positive ou nulle, et son intégrale vaut bien 1 :

$$\int_{-\infty}^{+\infty} f(x) dx = \underbrace{\int_{-\infty}^a f(x) dx}_{f(x)=0} + \underbrace{\int_a^b f(x) dx}_{f(x)=\frac{1}{b-a}} + \underbrace{\int_b^{+\infty} f(x) dx}_{f(x)=0} = \int_a^b \frac{dx}{b-a} = \left[\frac{x}{b-a} \right]_a^b = 1$$

Sa fonction de répartition F vaut, pour tout $x \in \mathbb{R}$, $F(x) = \int_{-\infty}^x f(t) dt$.

1. Si $x < a$, $F(x) = 0$
2. Si $x \in [a, b]$, $F(x) = \int_a^x \frac{1}{b-a} dt = \left[\frac{t}{b-a} \right]_a^x = \frac{x-a}{b-a}$
3. Si $x > b$, $F(x) = \int_a^b \frac{1}{b-a} dt + \int_b^x 0 dt = 1 + 0 = 1$

Les graphes de f et de F sont montrés figure 3.

Remarques :

1. Si $X \rightsquigarrow \text{Unif}(a, b)$ et si on définit $Z = \frac{X-a}{b-a}$, alors $Z \rightsquigarrow \text{Unif}(0, 1)$.
2. Inversement, si $Z \rightsquigarrow \text{Unif}(0, 1)$, alors on peut définir $X = a + (b-a)Z$. On aura $X \rightsquigarrow \text{Unif}(a, b)$.
On se sert de cela en programmation : certains environnements de programmation fournissent une fonction `rand()` qui retourne une réalisation de Z . On peut l'utiliser pour obtenir une réalisation de X .

5.2 Loi exponentielle

Définition : soit $\lambda > 0$. Une variable aléatoire X suit une loi exponentielle de paramètre λ si elle admet pour densité la fonction f définie par

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \lambda e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

On note cela : $X \rightsquigarrow \text{Exp}(\lambda)$.

La fonction f est bien une densité : elle est positive ou nulle, et son intégrale vaut bien 1 :

$$\int_{-\infty}^{+\infty} f(x) dx = \underbrace{\int_{-\infty}^0 f(x) dx}_{f(x)=0} + \underbrace{\int_0^{+\infty} f(x) dx}_{f(x)=\lambda e^{-x}} = \int_0^{+\infty} \lambda e^{-x} dx = [-e^{-\lambda x}]_0^{+\infty} = 1$$

La fonction de répartition de X vaut

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

En effet, si $x \geq 0$,

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^x = -e^{-\lambda x} + 1$$

Remarques :

1. Si $X \rightsquigarrow \text{Exp}(\lambda)$, alors $\lambda X \rightsquigarrow \text{Exp}(1)$.
2. De même si $Z \rightsquigarrow \text{Exp}(1)$, alors $\frac{Z}{\lambda} \rightsquigarrow \text{Exp}(\lambda)$.

5.3 Loi normale

Définition : soit $(m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$. Une variable aléatoire X suit une loi normal de paramètre (m, σ^2) si elle admet pour densité la fonction f définie par

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

On note cela : $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$.

Dans le cas où $m = 0$ et $\sigma^2 = 1$, on dit que X est une variable «normale centrée réduite».

Convention de notation : le second paramètre est noté σ^2 . Ainsi, σ est défini au signe près. Nous le supposons toujours positif. Nous verrons plus tard qu'il s'agit de l'écart-type de X .

La fonction f est bien une densité : elle est clairement positive et son intégrale vaut 1. Pour le montrer, il faut admettre que c'est bien le cas quand $(m, \sigma^2) = (0, 1)$, c'est-à-dire que

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

Pour d'autres valeurs de (m, σ^2) , on fait le changement de variable $u = \frac{x-m}{\sigma}$:

- Les bornes pour u sont $-\infty$ et $+\infty$
- De plus, $dx = \sigma du$

Ainsi,

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2}} \sigma du = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 1$$

Propriétés :

1. Si $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$, alors pour tout $(a, b) \in \mathbb{R}^2$ tel que $a \neq 0$, $aX + b \rightsquigarrow \mathcal{N}(am + b, a^2\sigma^2)$.
En effet, si on définit $Y = aX + b$, alors

$$\begin{aligned} f_Y(x) &= \frac{1}{|a|} f\left(\frac{x-b}{a}\right) \\ &= \frac{1}{|a|} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\frac{x-b}{a} - m\right)^2}{2\sigma^2}\right) \\ &= \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-b-am)^2}{2a^2\sigma^2}\right) \end{aligned}$$

Ce qui est la densité d'une variable $\mathcal{N}(am + b, a^2\sigma^2)$.

2. En particulier, si $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$ et qu'on définit $Z = \frac{X-m}{\sigma}$, alors Z est normale centrée réduite. C'est-à-dire que $Z \rightsquigarrow \mathcal{N}(0, 1)$.
3. Inversement, si Z est normale centrée réduite et qu'on définit $X = m + \sigma Z$, alors $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$.
4. Si X et Y sont deux variables indépendantes (la définition de l'indépendance de deux variables sera donnée plus loin) telles que

$$X \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2) \quad \text{et} \quad Y \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2)$$

Alors $(X + Y) \rightsquigarrow \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

La propriété 3 peut être utilisée en simulation informatique, quand on dispose d'une fonction retournant une réalisation aléatoire de Z et qu'on souhaite tirer X .

La fonction de répartition de X n'a pas d'expression analytique : il n'est pas possible de la calculer au moyen des fonctions classiques. Si $Z \rightsquigarrow \mathcal{N}(0, 1)$, on note parfois Φ sa fonction de répartition :

$$\Phi(x) = F_Z(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Alors pour tout variable $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$, on a

$$F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right)$$

puisque $F_X(x) = P(X \leq x) = P\left(\frac{X-m}{\sigma} \leq \frac{x-m}{\sigma}\right)$ et que $\frac{X-m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$.

La fonction Φ n'a pas d'expression analytique, mais elle est approchable numériquement. Il existe des tables de cette fonction, ou des mises en œuvre informatiques. On peut se contenter de retenir que

$$P(-1.96 \leq Z \leq 1.96) = \Phi(1.96) - \Phi(-1.96) = 95\%$$

et dans la pratique, on ne perd pas beaucoup de précision en remplaçant la valeur 1.96 par 2.

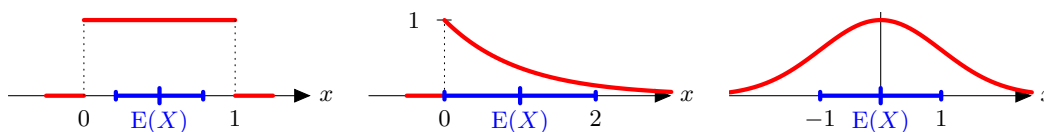


FIGURE 6 – Lois classiques. **À gauche** : densité de la loi uniforme sur $[0, 1]$. **Au milieu** : loi exponentielle ($\lambda = 1$). **À droite** : loi normale centrée réduite. Pour chaque loi, nous avons représenté en bleu, sur l'axe des abscisses, l'espérance $E(X)$ et l'écart-type (segment joignant $E(X) - \sigma(X)$ à $E(X) + \sigma(X)$).

6 Espérance et variance d'une variable aléatoire à densité

Considérons une variable aléatoire X admettant une densité f .

6.1 Espérance de X

Définition : l'espérance de X est définie par

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

Si cette intégrale diverge, la variable X n'a pas d'espérance.

Interprétation : si on se donne un nombre δx petit, on peut approcher la variable continue X par une variable discrète X' prenant les valeurs $\{x_k = k \delta x, k \in \mathbb{Z}\}$, chacune de ces valeurs ayant la probabilité $f(x_k) \delta x$.

Alors l'espérance de X' est

$$E(X') = \sum_{k=-\infty}^{+\infty} x_k P(X'=x_k) = \sum_{k=-\infty}^{+\infty} x_k f(x_k) \delta x_k$$

Cette dernière expression est une approximation de l'intégrale $\int_{-\infty}^{+\infty} x f(x) dx$, donc de $E(X)$.

Si on tire un grand nombre de fois la variable X et qu'on fait la moyenne de ces tirages, cette dernière tend vers $E(X)$. La valeur $E(X)$ est le «centre de gravité» de la loi : si l'axe (Ox) est une tige solide dont la masse³ en chaque point x est $f(x)$, le centre de gravité de la tige est situé en $E(X)$.

Propriétés de l'espérance : soient X et Y deux variables aléatoires et $(\alpha, \beta) \in \mathbb{R}^2$. Alors

1. $E(\alpha X + \beta) = \alpha E(X) + \beta$
2. $E(X + Y) = E(X) + E(Y)$
3. Pour tout fonction φ continue,

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

si l'intégrale converge.

3. En fait, la «densité de masse».

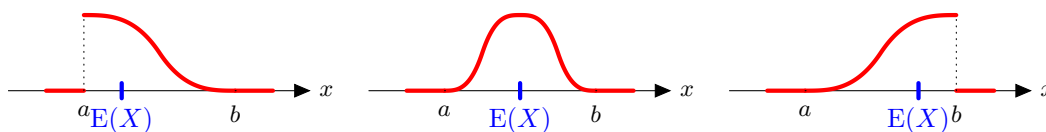


FIGURE 7 – **Espérance de X : quelques exemples.** Dans ces trois exemples, X est une variable aléatoire prenant ses valeurs sur un même intervalle $[a, b]$. **À gauche :** le poids de la distribution est plus important pour les valeurs de x proches de a . Cela attire $E(X)$ vers a . **Au milieu :** la distribution est parfaitement symétrique par rapport au centre de l'intervalle $[a, b]$. L'espérance $E(X)$ est au centre. **À droite :** le poids de la distribution est plus important pour les valeurs de x proches de b . Cela attire $E(X)$ vers b .

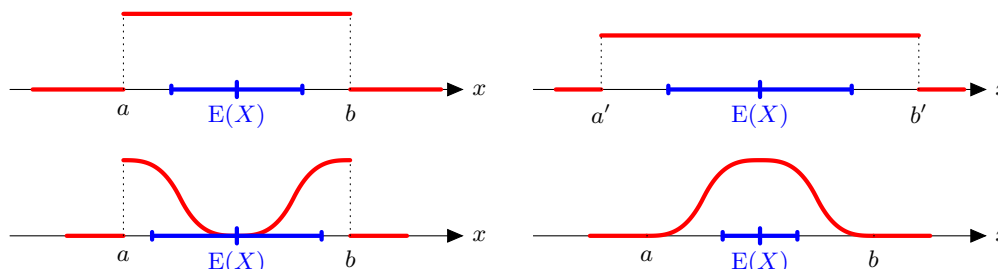


FIGURE 8 – **Variance de X : quelques exemples.** Dans ces quatre exemples, la variance est représentée par la longueur du segment en bleu, sur l'axe des abscisses : il joint les points $E(X) \pm \sigma(X)$ et sa longueur est $2\sigma(X)$. **En haut à gauche :** on considère initialement une loi uniforme sur un intervalle $[a, b]$. **En haut à droite :** supposons que X suive une loi de même forme, mais sur un intervalle $[a', b']$ plus large. Alors X est plus éloignée en moyenne de $E(X)$ et sa variance plus élevée. **En bas à gauche :** supposons à nouveau que X prend ses valeurs l'intervalle $[a, b]$. Si la loi donne une plus grande probabilité aux valeurs loin de $E(X)$, la variance est accrue. **En bas à droite :** si au contraire la loi donne une plus faible probabilité aux valeurs loin de $E(X)$, la variance décroît.

6.2 Variance et écart-type de X

Définition : la variance de X est

$$\text{Var}(X) = E\left((X - E(X))^2\right) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx$$

Si X n'a pas d'espérance ou si cette intégrale diverge, la variable X n'a pas de variance.

Dans le cas contraire, l'écart-type de X est $\sigma(X) = \sqrt{\text{Var}(X)}$.

Interprétation : donnons-nous un nombre δx petit et approchons la variable continue X par une variable discrète X' prenant les valeurs $\{x_k = k \delta x, k \in \mathbb{Z}\}$. Chacune de ces valeurs a la probabilité $f(x_k) \delta x$.

Alors la variance de X' est

$$\text{Var}(X') = \sum_{k=-\infty}^{+\infty} (x_k - E(X))^2 P(X'=x_k) = \sum_{k=-\infty}^{+\infty} (x_k - E(X))^2 f(x_k) \delta x_k$$

Cette dernière expression est une approximation de l'intégrale $\int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx$, donc de $\text{Var}(X)$.

Si on tire un grand nombre de fois la variable X et qu'on fait la moyenne des $(X - E(X))^2$ obtenus à chaque tirage, cette moyenne tend vers $\text{Var}(X)$.

L'inégalité de Tchebychev donne une relation probabiliste :

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

Ainsi, si on exprime ε sous la forme $\varepsilon = k\sigma$, on obtient

$$P(|X - E(X)| > k\sigma) \leq \frac{1}{k^2}$$

Choisissons par exemple k_0 tel que $\frac{1}{k_0^2} = 5\%$: la probabilité que X ne soit pas dans l'intervalle $E(X) \pm k_0\sigma$ est au plus 5%, donc la probabilité que X soit dans cet intervalle est au moins 95%. Cet intervalle est d'autant plus large que σ est élevée.

La variance (et l'écart-type) quantifie la distance de X à son espérance : plus $\text{Var}(X)$ est élevée, plus X est, en moyenne, loin de son espérance. La figure 8 montre quelques exemples.

Propriétés de la variance : soient X et Y deux variables aléatoires *indépendantes* et $(a, b) \in \mathbb{R}^2$. Alors

1. $\text{Var}(X) = E(X^2) - (E(X))^2$
2. $\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X)$, ce qui se traduit par : $\sigma(\alpha X + b) = |\alpha| \sigma(X)$
3. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ (si X et Y sont indépendantes)

La propriété 1 facilite souvent le calcul de la variance.

Attention à la propriété 3 : elle n'est vraie que si X et Y sont indépendantes. Il suffit pour s'en convaincre de penser au contre-exemple $Y = -X$. Dans ce cas, $X + Y$ est la variable aléatoire nulle et sa variance vaut 0.

7 Exemples de calculs de $E(X)$ et de $\text{Var}(X)$

Dans cette partie, nous menons explicitement les calculs. Leurs résultats sont montrés sur la figure 9.

7.1 Loi uniforme

Théorème : soient $(a, b) \in \mathbb{R}^2$ tels que $a < b$ et $X \rightsquigarrow \text{Unif}(a, b)$. Alors :

$$E(X) = \frac{a+b}{2} \quad \text{Var}(X) = \frac{(b-a)^2}{12} \quad \text{et donc} \quad \sigma(X) = \frac{b-a}{2\sqrt{3}}$$

Démonstration : posons $Z = \frac{X-a}{b-a} \rightsquigarrow \text{Unif}(0, 1)$ et déterminons son espérance et sa variance. Une densité de Z est

$$f(x) = \begin{cases} 1 & \text{si } x \in [0, 1] \\ 0 & \text{sinon} \end{cases}$$

1. Espérance de Z : par définition,

$$E(Z) = \int_{-\infty}^{+\infty} x f(x) dx = \underbrace{\int_{-\infty}^0 x f(x) dx}_{f(x)=0} + \underbrace{\int_0^1 x f(x) dx}_{f(x)=1} + \underbrace{\int_1^{+\infty} x f(x) dx}_{f(x)=0}$$

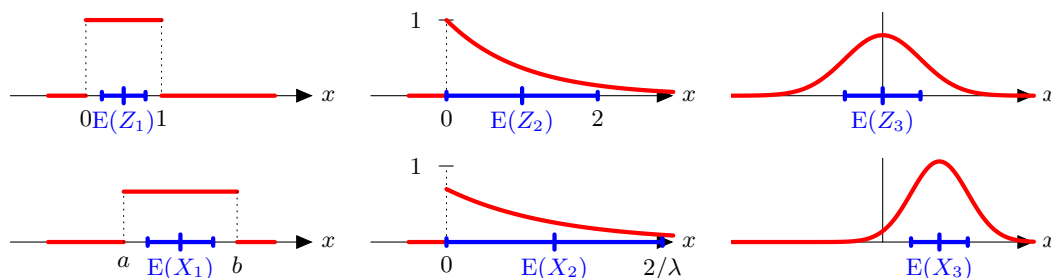


FIGURE 9 – **Espérance et variance de $\alpha X + \beta$: cas des lois classiques.** Ces exemples illustrent notamment que $E(\alpha X + \beta) = \alpha E(X) + \beta$ et $\sigma(\alpha X + \beta) = |\alpha| \sigma(X)$. **En haut** : lois de trois variables : $Z_1 \rightsquigarrow \text{Unif}(0, 1)$, $Z_2 \rightsquigarrow \text{Exp}(1)$ et $Z_3 \rightsquigarrow \mathcal{N}(0, 1)$. **En bas** : lois de variables obtenues à partir des précédentes par une relation $X_i = \alpha Z_i + \beta$. On représente à gauche $X_1 = a + (b - a)Z_1$: on a $X_1 \rightsquigarrow \text{Unif}(a, b)$. Au centre, $X_2 = \frac{1}{\lambda} Z_2$ qui vérifie $X_2 \rightsquigarrow \text{Exp}(\lambda)$ (ici, $\lambda < 1$). Enfin à droite, on définit $X_3 = m + \sigma Z_3$. On a alors $X_3 \rightsquigarrow \mathcal{N}(m, \sigma)$ (ici, $m > 0$ et $\sigma < 1$).

Ainsi,

$$E(Z) = \int_0^1 x \, dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2}$$

2. Variance de Z : on sait que $\text{Var}(Z) = E(Z^2) - (E(Z))^2$.

$$\text{Or } E(Z^2) = \int_{-\infty}^{+\infty} x^2 f(x) \, dx = \int_0^1 x^2 \, dx = \left[\frac{x^3}{3} \right]_0^1 = \frac{1}{3}$$

$$\text{Ainsi, } \text{Var}(Z) = \frac{1}{3} - \left(\frac{1}{2} \right)^2 = \frac{1}{12}.$$

3. Cas de X : on a $X = a + (b - a)Z$. Ainsi,

$$E(X) = a + (b - a) E(Z) = \frac{a + b}{2} \quad \text{et} \quad \text{Var}(X) = (b - a)^2 \text{Var}(Z) = \frac{(b - a)^2}{12}$$

Remarque technique : on aurait pu bien sûr ne pas passer par la variable $Z \rightsquigarrow \text{Unif}(0, 1)$. Par exemple, pour l'espérance, on aurait écrit

$$E(X) = \int_a^b \frac{x}{b - a} \, dx$$

Le changement de variable $u = \frac{x}{b - a}$ aurait conduit au même calcul que l'on a fait pour déterminer $E(Z)$.

7.2 Loi exponentielle

Théorème : soient $\lambda > 0$ et $X \rightsquigarrow \text{Exp}(\lambda)$. Alors

$$E(X) = \frac{1}{\lambda} \quad \text{Var}(X) = \frac{1}{\lambda^2} \quad \text{et donc} \quad \sigma(X) = \frac{1}{\lambda}$$

Démonstration : là aussi, on pose $Z = \lambda X$ et on calcule son espérance et sa variance. On sait en effet que $Z \rightsquigarrow \text{Exp}(1)$ et donc qu'une densité de Z est

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ e^{-x} & \text{sinon} \end{cases}$$

1. Espérance de Z : par définition,

$$\mathbb{E}(Z) = \int_{-\infty}^{+\infty} x f(x) dx = \underbrace{\int_{-\infty}^0 x f(x) dx}_{f(x)=0} + \underbrace{\int_0^{+\infty} x f(x) dx}_{f(x)=e^{-x}}$$

Ainsi,

$$\mathbb{E}(Z) = \int_0^{+\infty} x e^{-x} dx$$

Faisons une intégration par parties, en posant

$$\begin{cases} u(x) = x \\ v'(x) = e^{-x} \end{cases} \quad \text{donc} \quad \begin{cases} u'(x) = 1 \\ v(x) = -e^{-x} \end{cases}$$

On obtient

$$\mathbb{E}(Z) = \underbrace{[-x e^{-x}]_0^{+\infty}}_0 + \int_0^{+\infty} e^{-x} dx = [-e^{-x}]_0^{+\infty} = 1$$

2. Variance de Z : on sait que $\text{Var}(Z) = \mathbb{E}(Z^2) - (\mathbb{E}(Z))^2$.

$$\text{Or} \quad \mathbb{E}(Z^2) = \int_0^{+\infty} x^2 e^{-x} dx$$

Faisons une intégration par parties, en posant

$$\begin{cases} u(x) = x^2 \\ v'(x) = e^{-x} \end{cases} \quad \text{donc} \quad \begin{cases} u'(x) = 2x \\ v(x) = -e^{-x} \end{cases}$$

On obtient

$$\mathbb{E}(Z^2) = \underbrace{[-x^2 e^{-x}]_0^{+\infty}}_0 + 2 \int_0^{+\infty} x e^{-x} dx$$

La dernière intégrale vient d'être calculée pour déterminer $\mathbb{E}(Z)$: elle vaut 1.

$$\text{Ainsi,} \quad \mathbb{E}(Z^2) = 2 \quad \text{donc} \quad \text{Var}(Z) = \mathbb{E}(Z^2) - (\mathbb{E}(Z))^2 = 1.$$

3. Cas de X : on a $X = \frac{1}{\lambda} Z$. Ainsi,

$$\mathbb{E}(X) = \frac{1}{\lambda} \mathbb{E}(Z) = \frac{1}{\lambda} \quad \text{et} \quad \text{Var}(X) = \frac{1}{\lambda^2} \text{Var}(Z) = \frac{1}{\lambda^2}$$

7.3 Loi normale

Théorème : soient $(m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$ et $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$. Alors

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

Démonstration : encore une fois, on pose $Z = \frac{X - m}{\sigma}$ et on utilise la propriété que $Z \rightsquigarrow \mathcal{N}(0, 1)$. Une densité de Z est donc

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Nous admettons sans démonstration que cette fonction f est bien une densité, c'est-à-dire (on sait déjà qu'elle est positive) que

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

1. Espérance de Z : par définition,

$$\mathbb{E}(Z) = \mathbb{E}(Z) = \int_{-\infty}^{+\infty} x f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx$$

La fonction sous le signe intégral est facile à primitiver, car on reconnaît une forme $-u'e^u$. Ainsi,

$$\mathbb{E}(Z) = \frac{1}{\sqrt{2\pi}} \left[-e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} = 0$$

puisque $\lim_{x \rightarrow -\infty} e^{-\frac{x^2}{2}} = \lim_{x \rightarrow +\infty} e^{-\frac{x^2}{2}} = 0$

2. Variance de Z : comme $\mathbb{E}(X) = 0$, on a

$$\text{Var}(Z) = \mathbb{E}(Z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx$$

Faisons une intégration par parties, en posant

$$\begin{cases} u(x) &= x \\ v'(x) &= x e^{-\frac{x^2}{2}} \end{cases} \quad \text{donc} \quad \begin{cases} u'(x) &= 1 \\ v(x) &= -e^{-\frac{x^2}{2}} \end{cases}$$

On obtient

$$\text{Var}(Z) = \frac{1}{\sqrt{2\pi}} \left[-x e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx$$

Le premier terme vaut 0 puisque $\lim_{x \rightarrow -\infty} x e^{-\frac{x^2}{2}} = \lim_{x \rightarrow +\infty} x e^{-\frac{x^2}{2}} = 0$

Quant au second terme, ce n'est autre que l'intégrale $\int_{-\infty}^{+\infty} f(x) dx$ qui vaut 1 (admis).

Ainsi, $\text{Var}(Z) = 1$.

3. Cas de X : on a $X = m + \sigma Z$. Ce qui nous mène à

$$\mathbb{E}(X) = m + \sigma \mathbb{E}(Z) = m \quad \text{et} \quad \text{Var}(X) = \sigma^2 \text{Var}(Z) = \sigma^2$$

Remarque technique : là aussi, on aurait pu ne pas passer par la variable $Z \rightsquigarrow \mathcal{N}(0, 1)$. En déterminant les intégrales, on aurait utilisé le changement de variable $u = \frac{x - m}{\sigma}$ qui aurait conduit aux mêmes calculs que l'on vient de mener.

8 Lois à deux variables

8.1 Cas général

Considérons deux variables aléatoires réelles X et Y . Pour définir la loi 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.

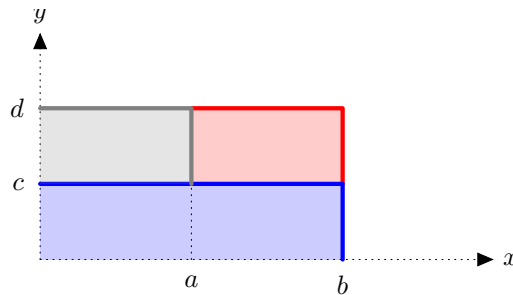


FIGURE 10 – Partition de l'ensemble $] -\infty, b] \times] -\infty, d]$. Cette partition permet de déduire la probabilité que (X, Y) soit dans $]a, b] \times]c, d]$ (l'ensemble en rouge).

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. $\lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F(x, y) = 1$

8.2 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 .

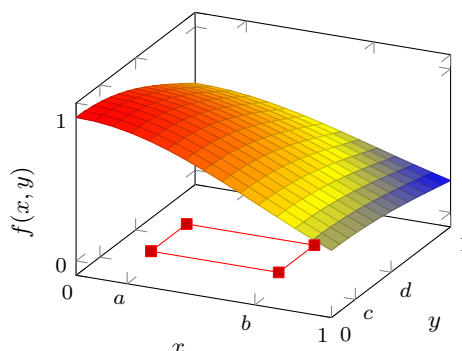


FIGURE 11 – Interprétation géométrique de l'intégrale de f sur $[a, b] \times [c, d]$. L'intégrale est le volume délimité par l'ensemble $[a, b] \times [c, d]$ (dans le plan $z = 0$) et le graphe de la fonction.

8.3 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 $(c, d) = (q_1\delta y, q_2\delta y)$, $(q_1, q_2) \in \mathbb{Z}^2$, $q_1 < q_2$,

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

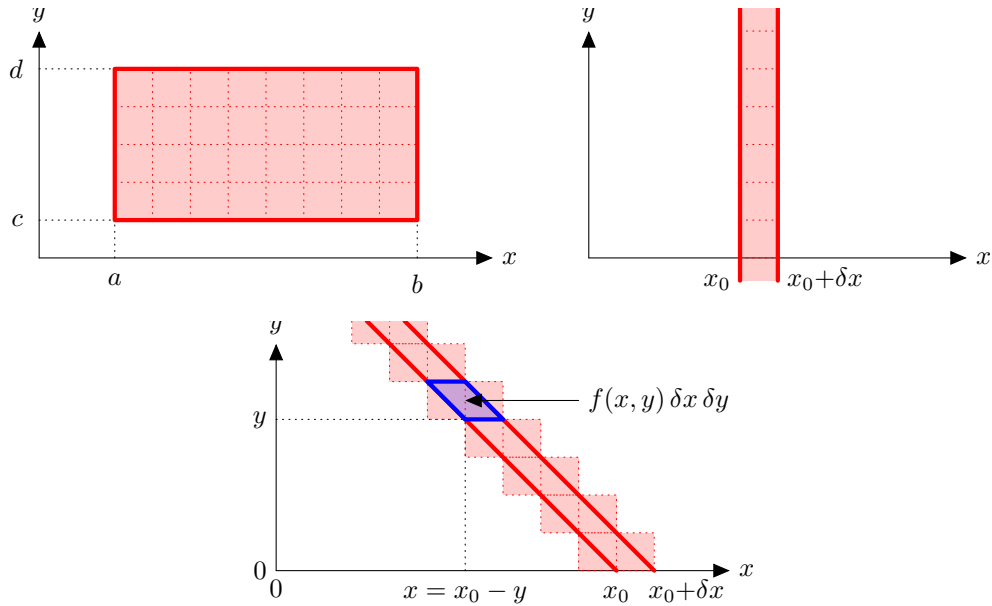


FIGURE 12 – Approximation de la loi jointe de X et Y . Cette loi est approximée par celle d'un couple (X', Y') prenant ses valeurs sur la grille $\{(p\delta x, q\delta y), (p, q) \in \mathbb{Z}^2\}$. **En haut à gauche** : la probabilité que (X', Y') soit sur un point de la grille à l'intérieur du rectangle $[a, b] \times [c, d]$ 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 $x + y = x_0$.

8.4 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 12). 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 $f_{X+Y} : x_0 \mapsto \int_{y=-\infty}^{+\infty} f(x_0 - y, y) dy$

8.5 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 : loi 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 13).

- (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 $f_Z(x_0) = 0$.

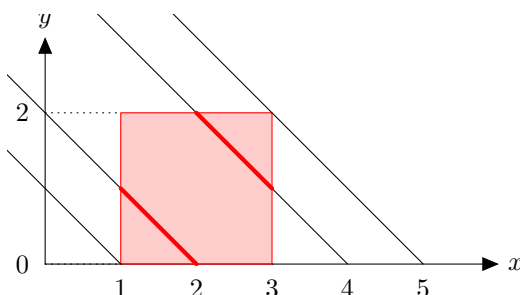


FIGURE 13 – Produit de convolution de deux densités uniformes. La zone sur fond rouge est le support du produit $f_X(x)f_Y(y)$. C'est la région dans laquelle le produit ne s'annule pas. On montre l'intersection de ce support avec la droite oblique $x + y = x_0$ pour quelques valeurs de x_0 . Si $x_0 < 1$, 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

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

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

9 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 loi, 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.

9.1 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 loi.

La suite (X_n) converge en loi 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 : $(X_n) \xrightarrow{\text{loi}} Y$.

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 loi de X_n peut être approchée par celle de Y . Dans la pratique, savoir que $(X_n) \xrightarrow{\text{loi}} Y$ 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 loi.

9.2 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. $(X_n) \xrightarrow{\text{loi}} X$. 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 $(X_n) \xrightarrow{\text{loi}} X$.

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

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

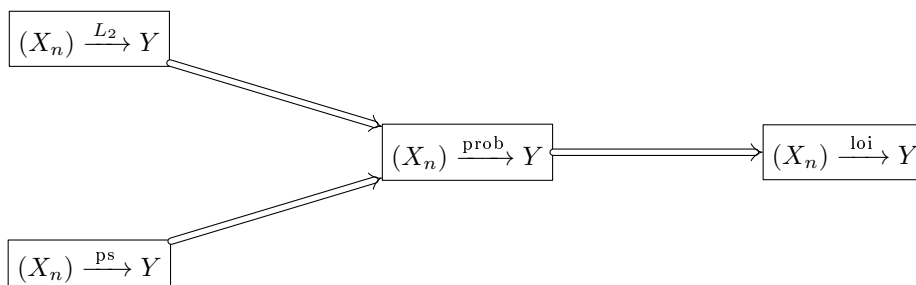
Ainsi, si on fixe $\varepsilon > 0$, on aura pour tout $n \in \mathbb{N}^*$

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

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

9.3 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.



9.4 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 loi. On suppose de plus qu'elles ont une espérance, qui est donc la même pour tous les X_i : $m = E(X_1) = E(X_2) = \dots$
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 m .

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

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

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

$$(Z_n) \xrightarrow{\text{loi}} Z$$

Remarques

1. On sait que $\mathbb{E}(\bar{X}_n) = m$ et que $\text{Var}(\bar{X}_n) = \frac{\text{Var}(X_i)}{n} = \frac{\sigma^2}{n}$. 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 loi 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 m . En effet, pour tout $\varepsilon > 0$,

$$|\bar{X}_n - m| < \varepsilon \iff -\varepsilon < \bar{X}_n - m < \varepsilon \iff -\frac{\varepsilon\sqrt{n}}{\sigma} < Z_n < \frac{\varepsilon\sqrt{n}}{\sigma}$$

On a donc

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

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 de l'intégrale 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

$$-1.96 < Z_n < 1.96 \iff -1.96 \frac{\sigma}{\sqrt{n}} < \bar{X}_n - m < 1.96 \frac{\sigma}{\sqrt{n}}$$

Ainsi, le théorème centrale limite permet d'affirmer que

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

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

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

9.5 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 14).

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

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

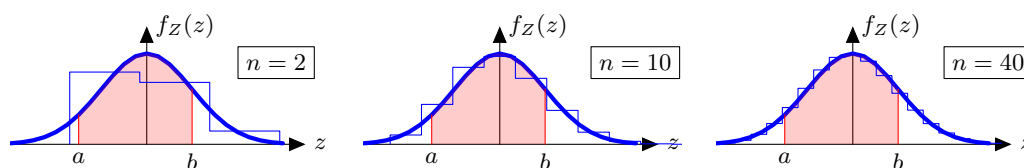


FIGURE 14 – Théorème central limite: convergence en loi de (Z_n) vers Z . En trait bleu épais, la densité de Z . L'aire de la zone en rouge est la probabilité que $Z \in [a, b]$. Les variables X_i suivent une loi de Bernoulli de paramètre $p = 0.3$. Pour une valeur n donnée, $S_n = X_1 + \cdots + X_n$ 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 $(Z_n) \xrightarrow{\text{loi}} Z$ 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

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

Finalement, on obtient

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