Probabilités continues

Support du cours PROC

Guillaume Euvrard, juin 2020
EPITA

Partie 1

Table des matières

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 distribution de X est une fonction

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

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

Convention de notation : on note X, Y, $Z \cdots$ avec des lettres majuscules les variables aléatoires. Les lettres minuscules x, y, $z \cdots$ 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 :

$$X(\Omega) = \{x_1, x_2, \cdots, x_N\} = \{x_n, n \in [\![1, N]\!]\}$$

soit infini, mais dénombrable :

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

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 distributions 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.

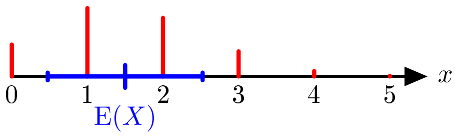

FIGURE 1 – `Distribution d'une variable aléatoire finie.` Ici, $X(\Omega) = [\![0, 5]\!]$. L'axe des abscisses représente les valeurs possibles de X, l'axe des ordonnées leurs probabilités. À chaque valeur possible, on place un trait rouge dont la hauteur est la probabilité que X prenne cette valeur.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

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

Une difficulté pratique survient quand on veut définir la distribution 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 :

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

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 ou $X(\Omega)$ 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 distribution. La connaissance de leurs probabilités permet de définir toute la distribution.

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

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

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 distribution 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 $\Sigma \subset \mathscr{P}(\mathbb{R})$. 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 $\overline{A}$ est aussi dans Σ
3. pour toute famille finie ou dénombrable $(A_n)_{n\in K}$ (avec K de la forme $[\![1, N]\!]$ ou $K = \mathbb{N}$) d'éléments de Σ, $\displaystyle\bigcup_{n\in K} A_n \in \Sigma$

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

On remarque d'ailleurs qu'il suffit de définir la seule probabilité $P_X(A)$ pour définir toute la distribution sur Σ. En effet, dès lors que cette probabilité est entre 0 et 1, on a
$$P_X(\emptyset)=0;\qquad P_X\left(\overline{A}\right)=1-P_X(A)\qquad\text{et}\qquad P_X(\mathbb{R})=1$$

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

Là aussi, on remarque que les seules probabilités $P_X(A)$ et $P_X(B)$ suffisent à définir toute la distribution sur Σ. Par exemple,
$$P_X\big(]3,+\infty[\big)=1-P(B)\,,\quad P_X\big(]-\infty,1]\cup]3,+\infty[\big)=P_X(A)+\big(1-P_X(B)\big)\,,\quad P_X\big(]1,3]\big)=P_X(B)-P_X(A)$$

Il faut néanmoins que ces deux nombres vérifient les inégalités
$$0\leqslant P_X(A)\leqslant P_X(B)\leqslant 1$$

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
$$F:\left|\begin{array}{ccc}\mathbb{R} & \longrightarrow & \mathbb{R}\\ x & \longmapsto & P_X\big(]-\infty,x]\big)\end{array}\right.$$

Autrement dit, pour tout $x\in\mathbb{R}$, $F(x)$ est la probabilité de l'évènement «$X\in]-\infty,x]$», donc de l'évènement «$X\leqslant x$» [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 distribution de la variable X. Cependant, quand il n'y a pas d'ambiguïté sur la variable aléatoire, on note F pour alléger le texte.

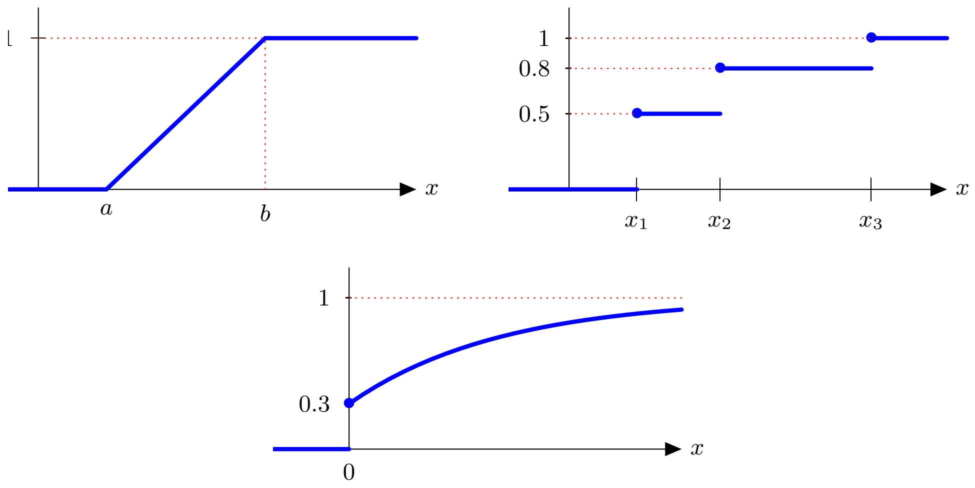

FIGURE 2 – `Trois exemples de fonctions de répartition.` **En haut à gauche :** cas d'une variable continue, qui prend toutes ses valeurs dans $[a, b]$. La probabilité de «$X \leqslant a$» est $F(a) = 0$ et la probabilité de «$X > b$» est $1 - F(b) = 0$. **En haut à droite :** cas d'une variable finie. 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 distribution 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 fonctions 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

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

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
$$P_X\big(]a, +\infty[\big) = P(X > a) = 1 - F(a)$$
2. l'intervalle $]a, b]$ est dans Σ et sa probabilité vaut
$$P_X(]a, b]) = P(a < X \leqslant b) = F(b) - F(a)$$
3. l'intervalle ouvert à droite $]-\infty, a[$ est dans Σ et sa probabilité vaut
$$P_X\big(]-\infty, a[\big) = P(X < a) = \lim_{x \to a^-} F(x)$$
4. De même, les intervalles $[a, +\infty[$, $[a, b]$, $[a, b[$ et $]a, b[$ sont tous dans Σ. Leurs probabilités sont
$$P_X\big([a, +\infty[\big) = 1 - \lim_{x \to a^-} F(x) \qquad P_X\big([a, b]\big) = F(b) - \lim_{x \to a^-} F(x)$$
$$P_X\big([a, b[\big) = \lim_{x \to b^-} F(x) - \lim_{x \to a^-} F(x) \qquad P_X\big(]a, b[\big) = \lim_{x \to b^-} F(x) - F(a)$$

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,

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

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

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 $P_X\big([a,b]\big)$ 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 $[a,b] \subset \mathbb{R}$,

$$F(b) - F(a) = \int_a^b f(x)\,\mathrm{d}x$$

alors f est une densité [2] 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

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

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 ou 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.

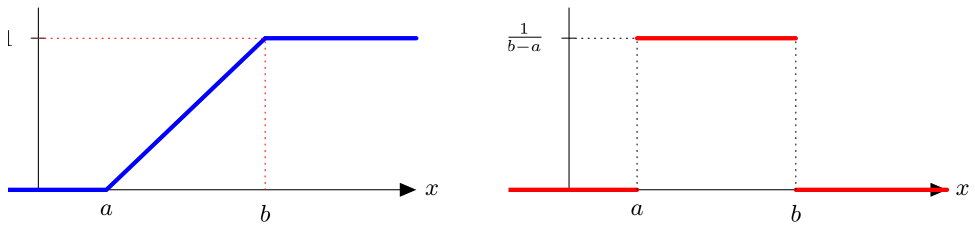

FIGURE 3 – `Distribution uniforme sur un intervalle` $[a, b]$. **Courbe de gauche :** la fonction de répartition de X. Elle est dérivable sur $\mathbb{R}$, sauf en a et en b. Néanmoins, en ces deux points, elle admet une dérivée à gauche et une dérivée à droite. **Courbe de droite :** densité de X. Elle 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.

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

1. Pour tout $a \in \mathbb{R}$, $P(X{=}a) = F(a) - \lim\limits_{x \to a^-} F(x) = 0$;
2. Pour tout $(a, b) \in \mathbb{R}^2$ tel que $a \leqslant b$, 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

$$P_X\big([a, b]\big) = P(a \leqslant X \leqslant b) = \int_a^b f(x)\, \mathrm{d}x$$

Propriétés vérifiées par la fonction f :

1. Pour tout $x \in \mathbb{R}$, $f(x) \geqslant 0$
2. L'intégrale $\displaystyle\int_{-\infty}^{+\infty} f(x)\, \mathrm{d}x$ 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

$$F : x \longmapsto \int_{-\infty}^{x} f(t)\, \mathrm{d}t$$

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

— Tirer une variable Z uniformément sur $[0, 1]$. Cette variable a la distribution illustrée sur la figure 3. En particulier, pour $z \in [0, 1]$, $P(Z{\leqslant}z) = z$.

— 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

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

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

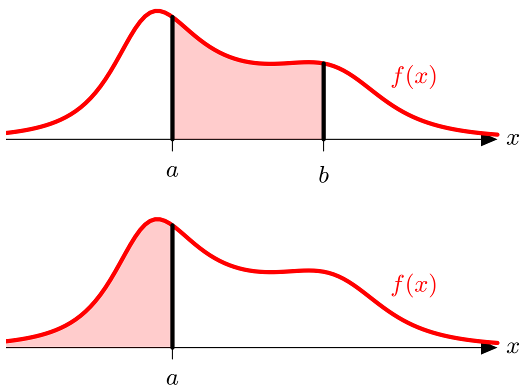

FIGURE 4 – **Représentation géométrique et approximation numérique d'une intégrale. En haut à gauche :** étant donné une fonction f (graphe en trait rouge épais), son intégrale sur un intervalle $[a,b]$ est l'aire représentée en rouge clair. **En haut à droite :** cette intégrale peut être approchée 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 le 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 distribution 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 $a \leqslant b$,

$$P(X{\leqslant}a) = F(a) = \int_{-\infty}^{a} f(x)\,\mathrm{d}x \qquad \text{et} \qquad P(a{\leqslant}X{\leqslant}b) = F(b) - F(a) = \int_a^b f(x)_d dx$$

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

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

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 ou 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

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

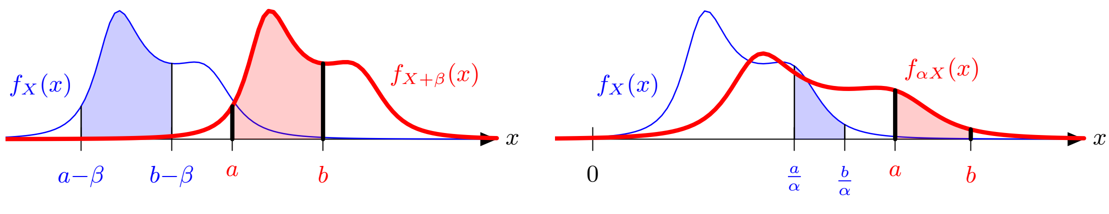

FIGURE 5 – **Densité de $\alpha X + \beta$. À gauche :** densités de X (en bleu) et de $X + \beta$ avec $\beta > 0$ (en rouge). La probabilité que $X + \beta$ soit compris entre a et b (aire en rouge) est la probabilité que X soit entre $a{-}\beta$ et $b{-}\beta$. Ainsi, $f_{X+\beta}(x) = f_X(x - \beta)$. **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 $F_Y(x)=P(Y\leqslant x)$ 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 $x\in\mathbb{R}$,
   $$Y\leqslant x \iff -2X+3\leqslant x \iff X\geqslant \frac{-x+3}{2}$$
   Ainsi, la fonction de répartition de Y est
   $$F_Y(x)=P(Y\leqslant x)=P\left(X\geqslant\frac{-x+3}{2}\right)=1-F\left(\frac{-x+3}{2}\right)$$
   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 «$X^2\leqslant x$» est impossible si $x<0$.
   Sinon, $X^2\leqslant x\iff x\in[-\sqrt{x},\sqrt{x}]$.
   Ainsi, la fonction de répartition de Y est
   $$F_Y(x)=\left|\begin{array}{cl} 0 & \text{si } x<0\\ F\left(\sqrt{x}\right)-F\left(-\sqrt{x}\right) & \text{sinon}\end{array}\right.$$
   En dérivant F_Y, on obtient la densité $f_Y(x)$: elle est nulle quand $x<0$. Quand $x>0$, elle vaut
   $$f_Y(x)=F_Y'(x)=F'\left(\sqrt{x}\right)\times\left(\sqrt{x}\right)' - F'\left(-\sqrt{x}\right)\times\left(-\sqrt{x}\right)'=\frac{1}{2\sqrt{x}}\left[f\left(\sqrt{x}\right)+f\left(-\sqrt{x}\right)\right]$$
   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
   $$\int_a^b f_Y(x)\,\mathrm{d}x$$
   sera de toute façon convergente pour tout $a\leqslant 0$ et $b\geqslant 0$. Elle vaudra $F_Y(b)-F_Y(a)$.

5 Exemples de distribution à densité

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

5.1 Distribution uniforme

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

$$f(x) = \left| \begin{array}{ll} \dfrac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{sinon} \end{array} \right.$$

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 :

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

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

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

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

Remarques :

1. Si $X \rightsquigarrow \text{Unif}(a,b)$ et si on définit $Z = \dfrac{X-a}{b-a}$, 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 Distribution exponentielle

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

$$f(x) = \left| \begin{array}{ll} 0 & \text{si } x < 0 \\ \lambda e^{-\lambda x} & \text{si } x \geqslant 0 \end{array} \right.$$

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 :

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \underbrace{\int_{-\infty}^{0} f(x)\,\mathrm{d}x}_{f(x)=0} + \underbrace{\int_{0}^{+\infty} f(x)\,\mathrm{d}x}_{f(x)=\lambda e^{-x}} = \int_{0}^{+\infty} \lambda e^{-x}\,\mathrm{d}x = \left[-e^{-\lambda x}\right]_0^{+\infty} = 1$$

La fonction de répartition de X vaut

$$F(x) = \left| \begin{array}{cl} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x} & \text{si } x \geqslant 0 \end{array} \right.$$

En effet, si $x \geqslant 0$,

$$F(x) = \int_0^x \lambda e^{-\lambda t}\,\mathrm{d}t = \left[-e^{-\lambda t}\right]_0^x = -e^{-\lambda x} + 1$$

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 $\dfrac{Z}{\lambda} \rightsquigarrow \text{Exp}(\lambda)$.

5.3 Distribution normale

Définition : soit $(m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$. Une variable aléatoire X suit une distribution 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 supposerons 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

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

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, $\mathrm{d}x = \sigma\,\mathrm{d}u$

Ainsi,

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

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 $Z = \dfrac{X-m}{\sigma}$, 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

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

   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 :

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

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

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

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

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

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

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

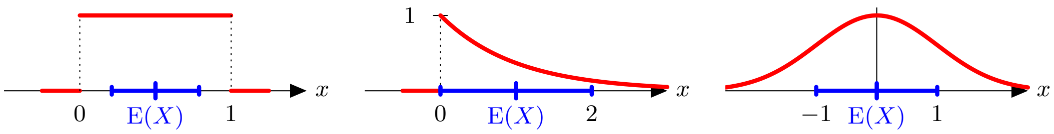

FIGURE 6 – `Distributions classiques.` **À gauche :** densité de la distribution uniforme sur $[0, 1]$. **Au milieu :** distribution exponentielle ($\lambda = 1$). **À droite :** distribution normale centrée réduite. Pour chaque distribution, nous avons représenté en bleu, sur l'axe des abscisses, l'espérance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

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

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

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

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

Cette dernière expression est une approximation de l'intégrale $\displaystyle\int_{-\infty}^{+\infty} x\, f(x)\, \mathrm{d}x$, donc de $\mathrm{E}(X)$.
Si on tire un grand nombre de fois la variable X et qu'on fait la moyenne de ces tirages, cette dernière tend vers $\mathrm{E}(X)$. La valeur $\mathrm{E}(X)$ est le «centre de gravité» de la distribution : si l'axe (Ox) est une tige solide dont la masse[3] en chaque point x est $f(x)$, le centre de gravité de la tige est situé en $\mathrm{E}(X)$.

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

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

$$\mathrm{E}\left(\varphi(X)\right) = \int_{-\infty}^{+\infty} \varphi(x)\, f(x)\, \mathrm{d}x$$

si l'intégrale converge.

3. En fait, la «densité de masse».

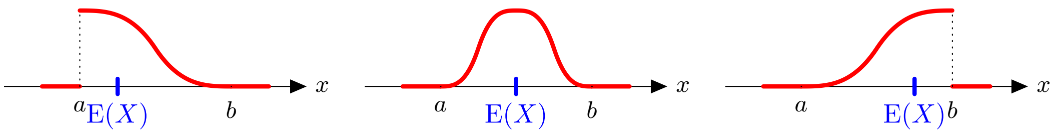

FIGURE 7 – `Espérance de` X`: quelques exemples.` Dans ces trois exemples, X est une variable aléatoire prenant ses valeurs sur un même intervalle $[a, b]$. **À gauche :** le poids de la distribution est plus important pour les valeurs de x proches de a. Cela attire $\mathrm{E}(X)$ vers a. **Au milieu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

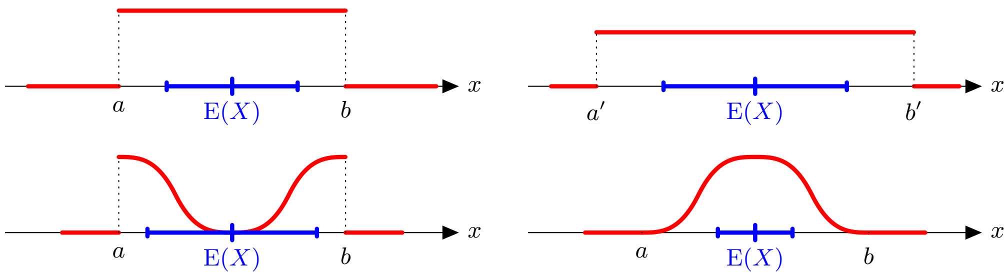

FIGURE 8 – `Variance de` X`: quelques exemples.` Dans ces quatre exemples, la variance est représentée par la longueur du segment en bleu, sur l'axe des abscisses : il joint les points $\mathrm{E}(X) \pm \sigma(X)$ et sa longueur est $2\,\sigma(X)$. **En haut à gauche :** on considère initialement une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

6.2 Variance et écart-type de X

Définition : la variance de X est

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

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 $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

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

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

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

Si on tire un grand nombre de fois la variable X et qu'on fait la moyenne des $\big(X - \mathrm{E}(X)\big)^2$ obtenus à chaque tirage, cette moyenne tend vers $\mathrm{Var}(X)$.
L'inégalité de Tchebychev donne une relation probabiliste :

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

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

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

Choisissons par exemple k_0 tel que $\frac{1}{k_0^2} = 5\%$: la probabilité que X ne soit pas dans l'intervalle $\mathrm{E}(X) \pm k_0\sigma$ 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 $\mathrm{Var}(X)$ 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. $\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \big(\mathrm{E}(X)\big)^2$
2. $\mathrm{Var}(\alpha X + \beta) = \alpha^2 \ \mathrm{Var}(X)$, ce qui se traduit par : $\sigma(\alpha X + b) = |\alpha| \ \sigma(X)$
3. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ (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 $\mathrm{E}(X)$ et de $\mathrm{Var}(X)$

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

7.1 Distribution uniforme

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

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

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

$$f(x) = \left| \begin{array}{ll} 1 & \text{si } x \in [0, 1] \\ 0 & \text{sinon} \end{array} \right.$$

1. Espérance de Z : par définition,

$$\mathrm{E}(Z) = \int_{-\infty}^{+\infty} x\, f(x)\, \mathrm{d}x = \underbrace{\int_{-\infty}^{0} x\, f(x)\, \mathrm{d}x}_{f(x)=0} + \underbrace{\int_{0}^{1} x\, f(x)\, \mathrm{d}x}_{f(x)=1} + \underbrace{\int_{1}^{+\infty} x\, f(x)\, \mathrm{d}x}_{f(x)=0}$$

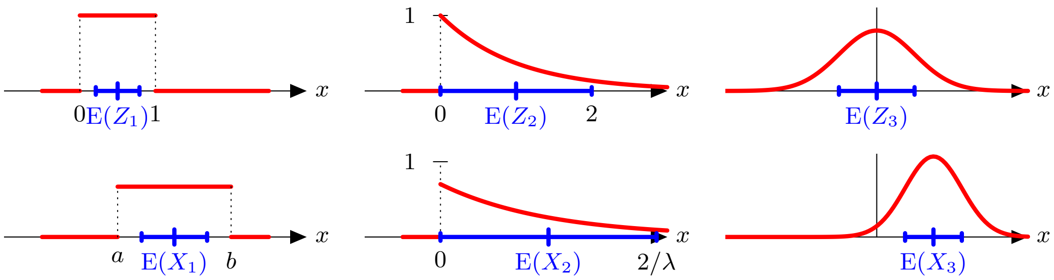

FIGURE 9 – `Espérance et variance de` $\alpha X+\beta$`: cas des distributions classiques.` Ces exemples illustrent notamment que $\mathrm{E}(\alpha X+\beta) = \alpha\,\mathrm{E}(X)+\beta$ et $\sigma(\alpha X+\beta) = |\alpha|\;\sigma(X)$. **En haut :** distributions de trois variables : $Z_1 \rightsquigarrow \mathrm{Unif}(0,1)$, $Z_2 \rightsquigarrow \mathrm{Exp}(1)$ et $Z_3 \rightsquigarrow \mathcal{N}(0,1)$. **En bas :** distributions 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 $X_1 \rightsquigarrow \mathrm{Unif}(a,b)$. Au centre, $X_2 = \frac{1}{\lambda}Z_2$ qui vérifie $X_2 \rightsquigarrow \mathrm{Exp}(\lambda)$ (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,
$$\mathrm{E}(Z) = \int_0^1 x\,\mathrm{d}x = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

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

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

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

3. Cas de X : on a $X = a + (b-a)Z$. Ainsi,
$$\mathrm{E}(X) = a + (b-a)\;\mathrm{E}(Z) = \frac{a+b}{2} \qquad \text{et} \qquad \mathrm{Var}(X) = (b-a)^2\,\mathrm{Var}(Z) = \frac{(b-a)^2}{12}$$

Remarque technique : on aurait pu bien sûr ne pas passer par la variable $Z \rightsquigarrow \mathrm{Unif}(0,1)$. Par exemple, pour l'espérance, on aurait écrit
$$\mathrm{E}(X) = \int_a^b \frac{x}{b-a}\,\mathrm{d}x$$
Le changement de variable $u = \dfrac{x}{b-a}$ aurait conduit au même calcul que l'on a fait pour déterminer $\mathrm{E}(Z)$.

7.2 Distribution exponentielle

Théorème : soient $\lambda > 0$ et $X \rightsquigarrow \mathrm{Exp}(\lambda)$. Alors
$$\mathrm{E}(X) = \frac{1}{\lambda} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2} \qquad \text{et donc} \qquad \sigma(X) = \frac{1}{\lambda}$$

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

$$f(x) = \left| \begin{array}{ll} 0 & \text{si } x \leqslant 0 \\ e^{-x} & \text{sinon} \end{array} \right.$$

1. Espérance de Z : par définition,

$$\mathrm{E}(Z) = \int_{-\infty}^{+\infty} x\, f(x)\, \mathrm{d}x = \underbrace{\int_{-\infty}^{0} x\, f(x)\, \mathrm{d}x}_{f(x)=0} + \underbrace{\int_{0}^{+\infty} x\, f(x)\, \mathrm{d}x}_{f(x)=e^{-x}}$$

Ainsi,

$$\mathrm{E}(Z) = \int_{0}^{+\infty} x\, e^{-x}\, \mathrm{d}x$$

Faisons une intégration par parties, en posant

$$\left| \begin{array}{lcl} u(x) & = & x \\ v'(x) & = & e^{-x} \end{array} \right. \qquad \text{donc} \qquad \left| \begin{array}{lcl} u'(x) & = & 1 \\ v(x) & = & -e^{-x} \end{array} \right.$$

On obtient

$$\mathrm{E}(Z) = \underbrace{\left[-x\, e^{-x}\right]_0^{+\infty}}_{0} + \int_{0}^{+\infty} e^{-x}\, \mathrm{d}x = \left[-\, e^{-x}\right]_0^{+\infty} = 1$$

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

Or $\quad \mathrm{E}\left(Z^2\right) = \displaystyle\int_{0}^{+\infty} x^2\, e^{-x}\, \mathrm{d}x$

Faisons une intégration par parties, en posant

$$\left| \begin{array}{lcl} u(x) & = & x^2 \\ v'(x) & = & e^{-x} \end{array} \right. \qquad \text{donc} \qquad \left| \begin{array}{lcl} u'(x) & = & 2x \\ v(x) & = & -e^{-x} \end{array} \right.$$

On obtient

$$\mathrm{E}\left(Z^2\right) = \underbrace{\left[-x^2\, e^{-x}\right]_0^{+\infty}}_{0} + 2 \int_{0}^{+\infty} x\, e^{-x}\, \mathrm{d}x$$

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

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

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

$$\mathrm{E}(X) = \frac{1}{\lambda}\, \mathrm{E}(Z) = \frac{1}{\lambda} \qquad \text{et} \qquad \text{Var}(X) = \frac{1}{\lambda^2}\, \text{Var}(Z) = \frac{1}{\lambda^2}$$

7.3 Distribution normale

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

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

Démonstration : encore une fois, on pose $Z = \dfrac{X-m}{\sigma}$ 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

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$

1. Espérance de Z : par définition,

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

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

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

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

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

$$\mathrm{Var}(Z) = \mathrm{E}\left(Z^2\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2\, e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

Faisons une intégration par parties, en posant

$$\left|\begin{array}{lcl} u(x) & = & x \\ v'(x) & = & xe^{-\frac{x^2}{2}} \end{array}\right. \qquad \text{donc} \qquad \left|\begin{array}{lcl} u'(x) & = & 1 \\ v(x) & = & -e^{-\frac{x^2}{2}} \end{array}\right.$$

On obtient

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

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

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

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

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

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

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 $u = \dfrac{x-m}{\sigma}$ qui aurait conduit aux mêmes calculs que l'on vient de mener.