# Probabilités pour les bioconcours

Pierre-Loïc Méliot

# Table des matières

Dans ce qui suit, on donne toutes les *définitions*, *propositions* et *théorèmes* importants et qu'il convient de connaître (par cœur). Des *exemples* illustrent systématiquement ces notions et résultats. Les *remarques*, pour leur part, peuvent être omises en première lecture, et font le plus souvent le lien entre les notions mathématiques probabilistes présentées, et l'intuition non mathématique que l'on peut avoir d'expériences mettant en jeu le hasard. On a également mis en remarque des résultats supplémentaires qu'il n'est pas forcément nécessaire de connaître et retenir.

Chapitre 1

# Dénombrement et probabilités discrètes

**Résumé.** Dans ce premier chapitre, on détaille la modélisation mathématique d'*expériences aléatoires* simples telles que le lancer d'un dé, le tirage de numéros au loto, celui de mains pour un jeu de cartes, *etc.* Dans chacun de ces cas, le *calcul des probabilités* repose sur l'énumération de choix possibles lors de l'expérience aléatoire. Ainsi, le *dénombrement* d'ensembles finis apparaît comme un prérequis au calcul des probabilités.

Dans tout ce qui suit, l'ensemble des entiers compris entre $a$ et $b$ sera noté $[\![a, b]\!]$. Ainsi, par exemple, $[\![2, 6]\!] = \{2, 3, 4, 5, 6\}$. L'ensemble de tous les entiers positifs (respectivement, positifs ou nuls) sera noté $\mathbb{N}^*$ (resp., $\mathbb{N}$). D'autre part, on rappelle que $a \in A$ veut dire "$a$ appartient à l'ensemble $A$", et que $A \subset B$ veut dire que "$A$ est une partie de $B$".

## 1. Cardinal d'un ensemble

De nombreuses expériences aléatoires (c'est-à-dire, dont le résultat "est pris au hasard") rentrent dans le cadre suivant : étant donné un ensemble $\Omega$ de possibilités, on choisit "au hasard" un élément $\omega \in \Omega$, et on se demande si cet élément $\omega$ satisfait un certain ensemble de conditions données par une partie $A \subset \Omega$.

Exemple 1.1. Supposons qu'on lance un dé à 6 faces, et que le résultat fasse gagner le joueur (à un jeu de société, à la roulette russe, *etc.*) s'il obtient 5 ou 6. L'ensemble des résultats possibles est

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

et on se demande quelle est la "chance" d'obtenir 5 ou 6, c'est-à-dire que le résultat $\omega$, pris au hasard dans $\Omega$, appartienne à la partie $A = \{5, 6\}$.

Exemple 1.2. On tire au hasard 6 boules parmi 49 boules numérotées (tirage du loto). Un résultat possible est, par exemple, $\{4, 8, 15, 16, 23, 42\}$. Plus généralement, un résultat du tirage du loto est un élément $\omega$ de

$$\Omega = \{\text{ensembles de 6 entiers distincts compris entre 1 et 49}\}.$$

Étant donné un élément $\omega_0 \in \Omega$, par exemple, $\{4, 8, 15, 16, 23, 42\}$, on peut se demander quelle chance un tirage aléatoire $\omega$ a d'avoir 5 entiers en commun avec $\omega_0$. Ceci revient à évaluer la probabilité de la partie

$$\begin{aligned} A = &\big\{\{4,8,15,16,23,n\},\ n \in [\![1,49]\!]\big\} \cup \big\{\{4,8,15,16,n,42\},\ n \in [\![1,49]\!]\big\} \\ &\cup \big\{\{4,8,15,n,23,42\},\ n \in [\![1,49]\!]\big\} \cup \big\{\{4,8,n,16,23,42\},\ n \in [\![1,49]\!]\big\} \\ &\cup \big\{\{4,n,15,16,23,42\},\ n \in [\![1,49]\!]\big\} \cup \big\{\{n,8,15,16,23,42\},\ n \in [\![1,49]\!]\big\}. \end{aligned}$$

Dans les cas précités, il est raisonnable de penser que la chance pour qu'un élément $\omega$ pris au hasard dans $\Omega$ tombe dans une partie $A$ est proportionnel à la taille de $A$. Autrement dit, la probabilité pour que $\omega \in A$ est donnée par la proportion

$$\frac{\text{nombre d'éléments de } A}{\text{nombre total d'éléments de } \Omega}.$$

Ainsi, et bien que toutes les expériences aléatoires ne rentrent pas dans le cadre précédemment décrit, il est naturel en probabilités d'avoir à compter le nombre d'éléments d'un ensemble $A$. Ce nombre d'éléments est appelé le **cardinal** de l'ensemble $A$, et noté $\operatorname{card} A$, ou parfois $|A|$. Un ensemble $A$ est dit fini s'il a un nombre fini d'éléments, donc, si son cardinal est un entier dans $\mathbb{N}$. Sinon, on notera $\operatorname{card} A = +\infty$. On ne donnera pas de définition plus abstraite et précise du cardinal d'un ensemble, mais on fera la remarque suivante : si $A$ et $B$ sont deux ensembles tels qu'existe entre eux une application $f : A \to B$ qui est bijective (tout élément de $B$ est atteint une et une seule fois par $f$), alors $A$ et $B$ ont le même nombre d'éléments, c'est-à-dire que $\operatorname{card} A = \operatorname{card} B$.

Exemple 1.3. Revenons à l'exemple 1.1. Le nombre de lancers possibles du dé, qui est le cardinal de $\Omega$, est $\operatorname{card} \Omega = \operatorname{card}([\![1,6]\!]) = 6$. Plus généralement, pour tout intervalle d'entiers $[\![a,b]\!]$,

$$\operatorname{card}([\![a,b]\!]) = b - a + 1.$$

Attention au $+1$ : par exemple, l'ensemble $[\![2,6]\!] = \{2,3,4,5,6\}$ contient bien 5 éléments, et non $6-2=4$. Poursuivant l'exemple, la partie $A = \{5,6\}$ contient 2 éléments, donc, la probabilité pour que le lancer de dé donne un résultat égal à 5 ou 6 est

$$\frac{\operatorname{card} A}{\operatorname{card} \Omega} = \frac{2}{6} = \frac{1}{3}.$$

## 2. Configurations classiques

Les éléments d'ensembles finis peuvent toujours être encodés par des entiers, des suites d'entiers, des ensembles d'entiers, *etc.* Dans cette section, on explique comment énumérer ces configurations, ce qui permet *in fine* de calculer le cardinal de tout ensemble fini. La discussion menée sera en lien avec les diverses constructions ensemblistes classiques : réunion d'ensembles, ensembles produits, ensembles des parties, *etc.*

**2.1. Union et intersection d'ensembles.** Soit $A$ et $B$ deux ensembles finis ; on rappelle que $A \cup B$ (la **réunion** de $A$ et de $B$) est l'ensemble des éléments qui sont dans $A$ ou dans $B$, et que $A \cap B$ (l'**intersection** de $A$ et de $B$) est l'ensemble des éléments qui sont dans $A$ et dans $B$. On suppose les cardinaux de $A$ et de $B$ connus. Une façon de compter les éléments de $A \cup B$ est lister d'abord ceux de $A$, puis ceux de $B$ ; on a alors compté au moins une fois tout élément dans $A \cup B$. Ainsi, de façon générale :

$$\operatorname{card}(A \cup B) \leq \operatorname{card} A + \operatorname{card} B.$$

En général, il n'y a pas égalité, car si $x$ appartient à $A \cap B$, alors il est compté deux fois lorsqu'on liste les objets dans $A$ puis dans $B$.

Si $A \cap B = \emptyset$, on dit que les ensembles $A$ et $B$ sont **disjoints** ; autrement dit, ils n'ont pas d'élément en commun. Souvent, on note alors leur réunion avec un symbole $\sqcup$ au lieu de $\cup$. Dans ce cas, dans l'énumération précédente, aucun élément n'est compté deux fois, donc :

PROPOSITION 1.4. *Si $A$ et $B$ sont deux ensembles disjoints, alors* $\operatorname{card}(A \sqcup B) = \operatorname{card} A + \operatorname{card} B$. *Plus généralement, si $A_1, \ldots, A_r$ sont des ensembles deux à deux disjoints, alors*

$$\operatorname{card}\left(\bigsqcup_{i=1}^{r} A_i\right) = \sum_{i=1}^{r} \operatorname{card} A_i.$$

EXEMPLE 1.5. Revenons à l'exemple 1.2. L'ensemble $A$ des tirages de 6 numéros dont 5 au moins sont dans $\{4, 8, 15, 16, 23, 42\}$ peut s'écrire comme la réunion disjointe suivante (notons que précédemment, on avait écrit $A$ comme une réunion non disjointe) :

$$\begin{aligned} A = & \big\{\{4, 8, 15, 16, 23, n\},\ n \in [\![1, 49]\!] \text{ et } n \neq 42\big\} \\ & \sqcup \big\{\{4, 8, 15, 16, n, 42\},\ n \in [\![1, 49]\!] \text{ et } n \neq 23\big\} \\ & \sqcup \big\{\{4, 8, 15, n, 23, 42\},\ n \in [\![1, 49]\!] \text{ et } n \neq 16\big\} \\ & \sqcup \big\{\{4, 8, n, 16, 23, 42\},\ n \in [\![1, 49]\!] \text{ et } n \neq 15\big\} \\ & \sqcup \big\{\{4, n, 15, 16, 23, 42\},\ n \in [\![1, 49]\!] \text{ et } n \neq 8\big\} \\ & \sqcup \big\{\{n, 8, 15, 16, 23, 42\},\ n \in [\![1, 49]\!] \text{ et } n \neq 4\big\} \\ & \sqcup \big\{\{4, 8, 15, 16, 23, 42\}\big\}. \end{aligned}$$

Le cardinal de $A$ est donc $48 \times 6 + 1 = 289$.

Si $A$ et $B$ sont deux ensembles non disjoints, il existe tout de même une formule pour calculer $\operatorname{card}(A \cup B)$. En effet, lorsqu'on énumère les éléments de $A$ puis ceux de $B$, ceux qui sont comptés deux fois sont exactement ceux dans l'intersection $A \cap B$. Donc, si l'on retranche cette quantité, on a compté tous les éléments de $A \cup B$ exactement une fois, et ainsi, la proposition 1.4 est généralisée par :

THÉORÈME 1.6. *Pour tous ensembles $A$ et $B$ finis,*

$$\operatorname{card}(A \cup B) = \operatorname{card} A + \operatorname{card} B - \operatorname{card}(A \cap B).$$

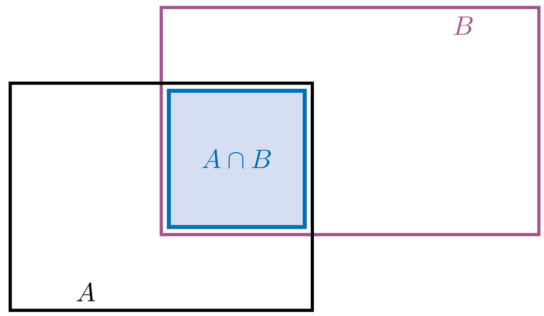


FIGURE 1. Le cardinal d'une union est $|A \cup B| = |A| + |B| - |A \cap B|$.

EXEMPLE 1.7. Soit $\Omega = [\![1, 6]\!]$, $A = \{5, 6\}$, et $B = \{\omega \in \Omega,\ \omega \text{ est pair}\} = \{2, 4, 6\}$. Alors, $A \cup B = \{2, 4, 5, 6\}$ et $A \cap B = \{6\}$, et l'on a bien

$$\operatorname{card}(A \cup B) = 4 = 2 + 3 - 1 = \operatorname{card} A + \operatorname{card} B - \operatorname{card}(A \cap B).$$

Il existe aussi des formules pour le cardinal de la réunion de plus de deux ensembles finis. On donnera juste la formule pour trois ensembles :

$$\begin{aligned}\operatorname{card}(A \cup B \cup C) = \operatorname{card} A + \operatorname{card} B + \operatorname{card} C \\ &- \operatorname{card}(A \cap B) - \operatorname{card}(A \cap C) - \operatorname{card}(B \cap C) \\ &+ \operatorname{card}(A \cap B \cap C).\end{aligned}$$

Cette formule peut se déduire de celle du théorème 1.6, appliquée plusieurs fois de suite à $A \cup B \cup C = (A \cup B) \cup C$.

**2.2. Produit d'ensembles.** Si $A$ et $B$ sont deux ensembles, on note $A \times B$ leur **produit cartésien**, qui est l'ensemble des paires ordonnées $(a, b)$ avec $a \in A$ et $b \in B$. Par exemple, avec $A = [\![1, 2]\!]$ et $B = [\![1, 3]\!]$, le produit $A \times B$ est l'ensemble des paires

$$\{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\}.$$

Notons qu'une paire $(a, b)$ n'est pas le même objet que la paire $(b, a)$ (en revanche, c'est le cas pour les ensembles sous-jacents, notés avec des accolades : $\{a, b\} = \{b, a\}$). Il existe une formule très simple pour le cardinal d'un produit :

THÉORÈME 1.8. *Pour tous ensembles $A$ et $B$ finis,*

$$\operatorname{card}(A \times B) = \operatorname{card} A \times \operatorname{card} B.$$

EXEMPLE 1.9. Dans l'exemple précédent, on a bien $6 = 2 \times 3$ paires dans $[\![1, 2]\!] \times [\![1, 3]\!]$.

On peut esquisser une preuve du théorème 1.8. Pour compter les paires dans $A \times B$, si $A = \{a_1, a_2, \ldots, a_n\}$ avec $n = \operatorname{card} A$, alors on peut répartir les paires $(a, b)$ suivant la valeur du premier élément $a = a_i$ :

$$A \times B = \bigsqcup_{i=1}^{n} \big\{(a_i, b),\ b \in B\big\}.$$

Par exemple, $[\![1,2]\!] \times [\![1,3]\!] = \{(1,1),(1,2),(1,3)\} \sqcup \{(2,1),(2,2),(2,3)\}$. On déduit alors de la proposition 1.4 que

$$\operatorname{card}(A \times B) = \sum_{i=1}^{n} \operatorname{card} B = n \operatorname{card} B = \operatorname{card} A \times \operatorname{card} B.$$

Ce résultat se généralise au cas d'un produit cartésien de plus de deux ensembles. Si $A, B, \ldots, Z$ sont des ensembles, on note $A \times B \times \cdots \times Z$ leur produit cartésien, qui est l'ensemble des suites ordonnées $(a, b, \ldots, z)$ avec $a \in A$, $b \in B$, *etc.*, $z \in Z$. Le cardinal d'un produit cartésien d'un nombre quelconque d'ensembles est le produit des cardinaux :

$$\operatorname{card}(A \times B \times \cdots \times Z) = \operatorname{card} A \times \operatorname{card} B \times \cdots \times \operatorname{card} Z.$$

Un cas particulier est celui où tous les ensembles $A, B, \ldots, Z$ sont les mêmes. On note $A^r = A \times A \times \cdots A$ le produit cartésien de $A$ avec lui-même $r$ fois ; c'est l'ensemble des suites ordonnées $(a_1, a_2, \ldots, a_r)$ avec chaque $a_i \in A$. Le cardinal de $A^r$ est $(\operatorname{card} A)^r$.

Exemple 1.10. Considérons tous les mots constitués de lettres 0 et 1, de longueur $r$ (code binaire). Par exemple, 011101010 est un tel mot de longueur $r = 9$. Le nombre de mots de longueur $n$ est $2^r$, puisqu'un mot de longueur $r$ avec les lettres 0 ou 1 peut être vu comme une suite dans $\{0,1\}^n$, avec $\operatorname{card}(\{0,1\}) = 2$.

Remarque. Le théorème 1.8, et sa généralisation à plus de deux ensembles, peuvent être associé à la règle informelle suivante, qu'il est utile de retenir : lorsqu'un objet peut être construit en faisant consécutivement plusieurs choix indépendants, le nombre de tels objets est obtenu en *multipliant* les nombres de choix. Ainsi, par exemple, lorsqu'on veut construire une suite dans $A^r$, il faut choisir d'abord le premier élément $a_1$ de la suite ($\operatorname{card} A$ choix possibles), puis le second élément $a_2$ (de nouveau $\operatorname{card} A$ choix possibles), et ainsi de suite jusqu'au $r$-ième élément, pour lequel il y a de nouveau $\operatorname{card} A$ choix possibles. Ces choix consécutifs doivent être multipliés, pour obtenir

$$\operatorname{card} A^r = (\operatorname{card} A)^r.$$

À l'inverse, si un objet peut être construit à partir d'une alternative entre plusieurs choix, alors le nombre de tels objets est obtenu en *additionnant* les nombres de choix. Ainsi, pour choisir un élément d'une union disjointe $A \sqcup B$, on peut soit le choisir dans $A$ ($\operatorname{card} A$ choix possibles), soit le choisir dans $B$ ($\operatorname{card} B$ choix possibles). Ces choix alternatifs doivent être additionnés, pour obtenir

$$\operatorname{card}(A \sqcup B) = \operatorname{card} A + \operatorname{card} B.$$

On retiendra donc que des choix consécutifs ("puis") correspondent à un produit, tandis que des choix alternatifs ("ou") correspondent à une somme.

**2.3. Arrangements et permutations.** Fixons un ensemble $A$. On rappelle que $A^r$ est l'ensemble des suites d'éléments de $A$ de longueur $r$. Un **arrangement** de taille $r$ dans $A$ est une telle suite $(a_1, a_2, \ldots, a_r)$, qui ne contient pas deux fois le même élément. Par exemple, si $A = [\![0, 9]\!]$, alors $(5, 8, 2, 4, 1)$ est un arrangement dans $A$ de taille 5, mais $(5, 8, 2, 4, 5)$ n'en est pas un, car 5 apparaît deux fois. On note $\mathfrak{A}(A, r)$ l'ensemble des arrangements de taille $r$ dans $A$. Il existe comme précédemment une formule pour le cardinal de $\mathfrak{A}(A, r)$ en fonction de celui de $A$ :

Théorème 1.11. *Si $A$ est un ensemble fini de cardinal $n$, alors le cardinal de $\mathfrak{A}(A, r)$ est*

$$n^{\downarrow r} = n(n-1)\cdots(n-r+1),$$

*c'est-à-dire le produit de $r$ entiers consécutifs, en partant de $n$ et en allant en décroissant. Cette formule donne $0$ si $r > n$.*

En effet, pour construire une suite $(a_1, a_2, \ldots, a_r)$ d'éléments de $A$ sans répétition, il y a $n$ choix possibles pour le premier élément $a_1$ ; puis, $n-1$ choix pour le second élément $a_2$ (tous les éléments de $A$, sauf $a_1$) ; puis, $n-2$ choix pour le troisième élément $a_3$ (tous les éléments de $A$, sauf $a_1$ et $a_2$) ; *etc.* jusqu'au choix du $r$-ième élément.

Exemple 1.12. Si $A = [\![1, 4]\!]$, alors il y a bien $4^{\downarrow 2} = 4 \times 3 = 12$ arrangements de taille 2 dans $A$, à savoir, $(1, 2)$, $(1, 3)$, $(1, 4)$, $(2, 1)$, $(2, 3)$, $(2, 4)$, $(3, 1)$, $(3, 2)$, $(3, 4)$, $(4, 1)$, $(4, 2)$ et $(4, 3)$. Par contre, il n'y a pas d'arrangements de taille 5 dans $A$, car il n'est pas possible de faire une liste de 5 éléments distincts à partir des 4 éléments de $A$.

Exemple 1.13. Lors du tirage des numéros du loto, le présentateur télé donne à la suite 6 numéros distincts et compris entre 1 et 49. Le nombre de telles suites possibles est

$$49^{\downarrow 6} = 49 \times 48 \times 47 \times 46 \times 45 \times 44 \simeq 1 \times 10^{10}.$$

Attention, ce n'est pas le nombre d'ensembles de 6 entiers distincts compris entre 1 et 49 (*i.e.*, le cardinal de $\Omega$ dans l'exemple 1.2), car on fait ici attention à l'ordre des numéros.

Un cas particulier d'arrangement est une liste ordonnée de *tous* les éléments de $A$, c'est-à-dire, un arrangement de taille $n = \operatorname{card} A$. On parle alors de **permutation** de $A$, et si $n = \operatorname{card} A$, on note

$$n^{\downarrow n} = n! = n(n-1)(n-2)\cdots 1$$

le nombre de permutations de $A$, que l'on lit "factorielle $n$". Par exemple, il y a $6 = 3 \times 2 \times 1$ permutations de $[\![1, 3]\!]$, à savoir, les suites $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$ et $(3, 2, 1)$.

**2.4. Parties d'un ensemble.** Un dernier type de configuration que l'on peut compter est l'**ensemble des parties** de taille donnée dans un ensemble. Soit $\Omega$ un ensemble de cardinal $n$ ; on note $\mathfrak{P}(\Omega)$ l'ensemble des parties $A \subset \Omega$, et $\mathfrak{P}(\Omega, r)$ l'ensemble des parties $A$ telles que $\operatorname{card} A = r$. Par exemple, si $\Omega = [\![1, 3]\!]$, alors

$$\begin{aligned}
\mathfrak{P}(\Omega, 0) &= \big\{\emptyset\big\} \\
\mathfrak{P}(\Omega, 1) &= \big\{\{1\}, \{2\}, \{3\}\big\} \\
\mathfrak{P}(\Omega, 2) &= \big\{\{1, 2\}, \{1, 3\}, \{2, 3\}\big\} \\
\mathfrak{P}(\Omega, 3) &= \big\{\{1, 2, 3\}\big\}
\end{aligned}$$

et $\mathfrak{P}(\Omega)$ est la réunion de ces ensembles, donc, compte 8 éléments. Notons que dans une partie (à l'inverse de ce qui se passe pour les arrangements), l'ordre n'est pas pris en compte : ainsi, $\{1, 3\}$ et $\{3, 1\}$ sont la même partie de taille 2 de $[\![1, 3]\!]$.

Théorème 1.14. *Si $\Omega$ est de taille $n$ et si $r \in [\![0, n]\!]$, alors l'ensemble $\mathfrak{P}(\Omega, r)$ de ses parties de taille $r$ a son cardinal donné par le coefficient binomial*

$$\binom{n}{r} = \frac{n^{\downarrow r}}{r!} = \frac{n!}{r!\,(n-r)!}.$$

*L'ensemble de toutes les parties $\mathfrak{P}(\Omega)$ a pour cardinal $2^n$.*

Pour la première formule, considérons l'application

$$\begin{aligned}
\mathfrak{A}(\Omega, r) &\to \mathfrak{P}(\Omega, r) \\
(a_1, a_2, \ldots, a_r) &\mapsto \{a_1, a_2, \ldots, a_r\}
\end{aligned}$$

qui oublie l'ordre dans un arrangement, et lui associe la partie sous-jacente. Chaque partie de taille $r$ est atteinte par cette application, et ce plusieurs fois, puisqu'une même partie $\{a_1, \ldots, a_r\}$ est atteinte par toutes les permutations possibles de la suite $(a_1, \ldots, a_r)$. Plus précisément, le nombre de suites qui sont envoyées sur une même partie est ce nombre de permutations, c'est-à-dire $r!$. On en déduit que

$$\operatorname{card} \mathfrak{A}(\Omega, r) = r! \times \operatorname{card} \mathfrak{P}(\Omega, r),$$

ce qui mène à la première formule. Pour la seconde formule, une preuve est donnée par l'argument suivant : pour construire une partie $A \subset \Omega$, on doit choisir pour chaque élément $\omega \in \Omega$ si $\omega$ appartient ou non à $A$ (2 choix possibles). Il faut répéter ce choix $n$ fois consécutivement (le nombre d'objets dans $\Omega$), donc, $\operatorname{card} \mathfrak{P}(\Omega) = 2^n$.

Exemple 1.15. Les coefficients binomiaux $\binom{3}{0}$, $\binom{3}{1}$, $\binom{3}{2}$ et $\binom{3}{3}$ sont respectivement 1, 3, 3 et 1. Ils correspondent bien aux cardinaux des ensembles de parties précédemment présentés lorsque $\Omega = [\![1, 3]\!]$. De plus, on a bien $\operatorname{card} \mathfrak{P}(\Omega) = 1 + 3 + 3 + 1 = 8 = 2^3$.

Il y a deux propriétés des coefficients binomiaux qu'il est très utile de retenir. La première est la formule du binôme de Newton :

Proposition 1.16. *Pour toutes variables $x, y$,*

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r\, y^{n-r}.$$

Ainsi, les coefficients binomiaux rentrent en jeu dans le développement des puissances d'une somme. Notons que la formule du binôme implique la seconde partie du théorème 1.14 : si $\operatorname{card}\Omega = n$, alors

$$\begin{aligned}\operatorname{card}\mathfrak{P}(\Omega) &= \operatorname{card}\left(\bigsqcup_{r=0}^{n} \mathfrak{P}(\Omega, r)\right) = \sum_{r=0}^{n} \operatorname{card}\mathfrak{P}(\Omega, r) \\ &= \sum_{r=0}^{n} \binom{n}{r} = \sum_{r=0}^{n} \binom{n}{r} 1^r\, 1^{n-r} = (1+1)^n = 2^n.\end{aligned}$$

D'autre part, les coefficients binomiaux vérifient la formule de récurrence suivante :

Proposition 1.17. *Pour tous $n, r$,*

$$\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1}.$$

On peut démontrer cette formule en réduisant au même dénominateur les fractions mises en jeu, mais on peut aussi donner une preuve sans calcul. Si $A$ est une partie de $[\![1, n+1]\!]$ de taille $r+1$, alors soit $A$ ne contient pas $n+1$, auquel cas c'est une partie de $[\![1, n]\!]$ de taille $r+1$ ($\binom{n}{r+1}$ choix possibles) ; soit $A$ contient $n+1$, auquel cas $A \setminus \{n+1\}$ est une partie de $[\![1, n]\!]$ de taille $r$ ($\binom{n}{r}$ choix possibles, et ce choix détermine entièrement $A$ en lui rajoutant l'élément $n+1$). De cette alternative, on en déduit que le nombre de parties de taille $r+1$ dans $[\![1, n+1]\!]$ est $\binom{n+1}{r+1} = \binom{n}{r+1} + \binom{n}{r}$.

La proposition 1.17 permet de calculer les coefficients binomiaux à l'aide du **triangle de Pascal** :

$$\begin{array}{c} 1 \\ 1\ \ 1 \\ 1\ \ 2\ \ 1 \\ 1\ \ 3\ \ 3\ \ 1 \\ 1\ \ 4\ \ 6\ \ 4\ \ 1 \\ 1\ 5\ 10\ 10\ 5\ 1 \end{array}$$

Les lignes de ce triangle étant numérotées à partir de 0, les nombres de la $n$-ième ligne de ce triangle sont les coefficients binomiaux $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$. La règle de construction du triangle est très simple : les lignes commencent et finissent par 1, et chaque nombre est la somme des deux nombres directement aussi de lui, conformément à la règle donnée par la proposition 1.17.

Exemple 1.18. On peut finalement calculer la probabilité de gagner au loto. Le nombre de parties à 6 éléments dans $\Omega = [\![1, 49]\!]$ est

$$\binom{49}{6} = \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13\,983\,816.$$

Par conséquent, une grille de loto étant justement un choix d'une de ces combinaisons de 6 chiffres, la probabilité pour qu'une grille corresponde au tirage d'une partie aléatoire est $\frac{1}{13\,983\,816}$.

## 3. Espaces de probabilité discrets

Il est maintenant temps de formaliser mathématiquement les expériences aléatoires précédemment décrites, ce qui va nous mener à la notion d'espace de probabilité.

**3.1. Mesures de probabilité.** De façon générale, une expérience aléatoire peut être décrite comme le choix d'un élément $\omega$ dans un ensemble $\Omega$, chaque élément ayant une certaine chance $\mathbb{P}[\omega]$ d'être choisi. Ces quantités $\mathbb{P}[\omega]$ ne sont pas forcément toutes égales, et d'autre part, cette approche n'est pas tout à fait la bonne lorsque l'ensemble $\Omega$ des résultats possibles est "trop grand" et contient une infinité non dénombrable d'éléments. Pour cette raison, dans ce qui suit, on supposera que l'ensemble $\Omega$ est **dénombrable**, c'est-à-dire qu'il existe une bijection entre $\Omega$ et une partie de l'ensemble des entiers $\mathbb{N}$. Cette propriété est en particulier vérifiée si $\Omega$ est fini de cardinal $n$ (dans ce cas, il est en bijection avec la partie $[\![1, n]\!] \subset \mathbb{N}$). L'hypothèse de dénombrabilité permet également de gérer des ensembles infinis "pas trop grands".

On note comme précédemment $\mathfrak{P}(\Omega)$ l'ensemble des parties de $\Omega$.

Définition 1.19. *Une **mesure de probabilité** sur $\Omega$ est une fonction*

$$\mathbb{P} : \mathfrak{P}(\Omega) \to [0, 1]$$

*qui vérifie les hypothèses suivantes :*

*(1) On a $\mathbb{P}[\emptyset] = 0$ et $\mathbb{P}[\Omega] = 1$.*

*(2) Si $(A_i)_{i \in I}$ est une famille de parties disjointes de $\Omega$, alors*

$$\mathbb{P}\left[\bigsqcup_{i \in I} A_i\right] = \sum_{i \in I} \mathbb{P}[A_i].$$

On appelle **espace de probabilité (discret)** la donnée d'un ensemble dénombrable $\Omega$ et d'une probabilité $\mathbb{P}$ sur cet ensemble. À une telle donnée correspond l'expérience aléatoire qui consiste à tirer un élément $\omega \in \Omega$, de telle façon que pour toute partie $A \subset \Omega$, la probabilité pour que $\omega$ tombe dans $A$ est $\mathbb{P}[A]$.

Exemple 1.20. Le lancer de dé de l'exemple 1.1 correspond à l'espace de probabilité $\Omega = [\![1, 6]\!]$, et à la probabilité $\mathbb{P}[A] = \frac{\operatorname{card} A}{\operatorname{card} \Omega} = \frac{\operatorname{card} A}{6}$.

Exemple 1.21. Le tirage du loto de l'exemple 1.2 correspond à l'espace de probabilité $\Omega = \mathfrak{P}(\Omega, 6)$, et à la probabilité $\mathbb{P}[A] = \frac{\mathrm{card}\, A}{\mathrm{card}\, \Omega} = \frac{\mathrm{card}\, A}{\binom{49}{6}}$.

De façon plus générale, pour tout ensemble fini $\Omega$, on dispose d'une mesure de probabilité sur $\Omega$ donnée par la formule

$$\mathbb{P}[A] = \frac{\mathrm{card}\, A}{\mathrm{card}\, \Omega},$$

et appelée **probabilité uniforme** sur $\Omega$. Attention, ce n'est pas du tout la seule probabilité possible sur un ensemble fini.

Exemple 1.22. On considère une expérience aléatoire qui a deux résultats possibles, 0 ("échec") et 1 ("succès"). On parle alors d'**expérience de Bernoulli**. Toute mesure de probabilité sur $\Omega = \{0, 1\}$ prend la forme suivante :

$$\mathbb{P}[\emptyset] = 0 \quad ; \quad \mathbb{P}[\{1\}] = p \quad ; \quad \mathbb{P}[\{0\}] = 1 - p \quad ; \quad \mathbb{P}[\Omega] = 1$$

pour un certain paramètre $p \in [0, 1]$. En effet, si $\mathbb{P}$ est une mesure de probabilité sur un ensemble fini $\Omega$, alors pour toute partie $A$, si $A^{\mathrm{c}} = \{\omega \in \Omega,\ \omega \notin A\}$ désigne le complémentaire de $A$, alors

$$\mathbb{P}[A] + \mathbb{P}[A^{\mathrm{c}}] = \mathbb{P}[\Omega] = 1,$$

donc $\mathbb{P}[A^{\mathrm{c}}] = 1 - \mathbb{P}[A]$. Appliquant ce résultat à $A = \{1\}$ et $A^{\mathrm{c}} = \{0\}$, on en déduit la forme d'une mesure de probabilité pour une expérience de Bernoulli. Cette mesure de probabilité est uniforme si et seulement si $p = \frac{1}{2}$.

Si $(\Omega, \mathbb{P})$ est un espace de probabilité discret, les parties $A \in \mathfrak{P}(\Omega)$ sont appelées **événements** pour l'expérience aléatoire, et on peut avec la fonction $\mathbb{P}$ calculer toutes les probabilités des événements. Compte tenu de l'additivité des probabilités, la fonction $\mathbb{P}$ est entièrement déterminée par les probabilités des singletons :

$$\forall A \subset \Omega,\ \mathbb{P}[A] = \sum_{\omega \in A} \mathbb{P}[\{\omega\}].$$

Ainsi, parfois, on décrira une probabilité sur un ensemble on donnant uniquement sa valeur sur les singletons. Les valeurs de $\mathbb{P}$ sur les événements généraux s'en déduisent à l'aide de la formule précédente.

Exemple 1.23. La probabilité uniforme sur un ensemble fini $\Omega$ est donnée par ses valeurs sur les singletons :

$$\forall \omega \in \Omega,\ \mathbb{P}[\omega] = \frac{1}{\mathrm{card}\, \Omega}.$$

**3.2. Événements indépendants.** Dans ce qui suit, on s'intéresse au calcul des probabilités d'événements d'un espace de probabilité quelconque. Les notions d'**indépendance** et de **conditionnement** simplifient ce calcul lorsqu'on considère des combinaisons de plusieurs événements (réunion, intersection, *etc.*). Pour commencer, donnons l'analogue du théorème 1.6 pour des probabilités :

PROPOSITION 1.24. *Soit* $(\Omega, \mathbb{P})$ *un espace de probabilité, et* $A$ *et* $B$ *deux événements. On a toujours :*

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

En effet, on peut écrire $A \cup B$ comme la réunion disjointe de $A \cap B$ et de

$$A \setminus B = \{\omega \in A,\ \omega \notin B\} \quad ; \quad B \setminus A = \{\omega \in B,\ \omega \notin A\}.$$

Par conséquent,

$$\begin{aligned}\mathbb{P}[A \cup B] &= \mathbb{P}[A \setminus B] + \mathbb{P}[B \setminus A] + \mathbb{P}[A \cap B] \\ &= (\mathbb{P}[A] - \mathbb{P}[A \cap B]) + (\mathbb{P}[B] - \mathbb{P}[A \cap B]) + \mathbb{P}[A \cap B] \\ &= \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]\end{aligned}$$

car $A = (A \setminus B) \sqcup (A \cap B)$ et $B = (B \setminus A) \sqcup (A \cap B)$.

On peut alors s'intéresser à la probabilité d'une intersection $A \cap B$ d'événements. Pour tous événements $A$ et $B$, on a *a priori*

$$0 \leq \mathbb{P}[A \cap B] \leq \min(\mathbb{P}[A], \mathbb{P}[B])$$

puisque $A \cap B \subset A$ et $A \cap B \subset B$.

DÉFINITION 1.25. *On dit que deux événements* $A$ *et* $B$ *sont* ***indépendants*** *si*

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B].$$

EXEMPLE 1.26. Pour le lancer de dé, considérons les trois événements

$$\begin{aligned}A &= \{\text{le lancer est pair}\} = \{2, 4, 6\} \\ B &= \{\text{le lancer est plus grand que } 5\} = \{5, 6\} \\ C &= \{\text{le lancer est plus petit que } 5\} = \{1, 2, 3, 4, 5\}.\end{aligned}$$

Les événements $A$ et $B$ sont indépendants, car

$$\mathbb{P}[A \cap B] = \mathbb{P}[\{6\}] = \frac{1}{6} = \frac{1}{2} \times \frac{1}{3} = \mathbb{P}[A]\,\mathbb{P}[B].$$

En revanche, $A$ et $C$ ne sont pas indépendants, car

$$\mathbb{P}[A \cap C] = \mathbb{P}[\{2, 4\}] = \frac{1}{3} \neq \frac{1}{2} \times \frac{5}{6}.$$

REMARQUE. La notion mathématique d'événements indépendants formalise l'intuition suivante. Soit $A$ et $B$ deux événements d'un espace de probabilité $(\Omega, \mathbb{P})$. On suppose qu'il existe une suite de choix consécutifs indépendants $c_1, \ldots, c_r, d_1, \ldots, d_s$ tels que :

(1) Cette suite de choix produit un élément aléatoire $\omega \in \Omega$ qui suit la mesure de probabilité $\mathbb{P}$ ;

(2) L'appartenance $\omega \in A$ peut être décidée uniquement à partir des choix $c_1, \ldots, c_r$, et l'appartenance $\omega \in B$ peut être décidée uniquement à partir des choix $d_1, \ldots, d_s$.

Alors, $A$ et $B$ sont indépendants au sens de la définition 1.25. Illustrons ceci, en gardant les mêmes notations que dans l'exemple précédent. Il existe une façon de choisir aléatoirement $\omega \in [\![1, 6]\!]$ qui est particulièrement adaptée au calcul des probabilités de $A$, $B$ et $A \cap B$ :

$c_1$ : On choisit aléatoirement $a \in \{0, 1\}$, avec probabilité $\frac{1}{2}$ pour chaque résultat.

$d_1$ : On retranche $a$ à $2b$, où $b$ est choisi aléatoirement dans $\{1, 2, 3\}$, avec probabilité $\frac{1}{3}$ pour chaque résultat.

La quantité $\omega = 2b - a$ est comprise entre 1 et 6, et on vérifie aisément que par cette construction, tous les entiers de $[\![1, 6]\!]$ sont atteints avec la même probabilité $\frac{1}{6}$. Donc, la suite de choix $(c_1, d_1)$ donne $\omega \in [\![1, 6]\!]$ suivant la probabilité uniforme (lancer de dé à 6 faces). Or,

(1) Le résultat $\omega$ est pair (dans $A$) si et seulement si $a = 0$ ;

(2) Le résultat $\omega$ est plus grand que 5 (dans $B$) si et seulement si $b = 3$.

Comme $c_1$ et $d_1$ sont des choix consécutifs indépendants, $A$ et $B$ sont des événements indépendants, et on multiplie leurs probabilités pour calculer $\mathbb{P}[A \cap B]$.

Remarque. Il existe une notion d'indépendance pour plus de deux événements. Si $A_1, \ldots, A_n$ sont $n$ événements dans un espace de probabilité $\Omega$, on dit qu'ils sont indépendants si, pour tout choix de certains de ces événements $A_{j_1}, A_{j_2}, \ldots, A_{j_r}$ avec $2 \leq r \leq n$, on a

$$\mathbb{P}\left[A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_r}\right] = \mathbb{P}[A_{j_1}]\, \mathbb{P}[A_{j_2}] \cdots \mathbb{P}[A_{j_r}].$$

Attention, c'est plus fort que de demander simplement que $\mathbb{P}[A_1 \cap A_2 \cap \cdots \cap A_n] = \mathbb{P}[A_1]\, \mathbb{P}[A_2] \cdots \mathbb{P}[A_n]$. Par exemple, avec 3 événements $A, B, C$, l'indépendance revient à demander que

$$\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A]\, \mathbb{P}[B]\, \mathbb{P}[C],$$

mais aussi que $\mathbb{P}[A \cap B] = \mathbb{P}[A]\, \mathbb{P}[B]$, $\mathbb{P}[A \cap C] = \mathbb{P}[A]\, \mathbb{P}[C]$ et $\mathbb{P}[B \cap C] = \mathbb{P}[B]\, \mathbb{P}[C]$.

**3.3. Probabilités conditionnelles.** Les probabilités conditionnelles permettent d'aller au-delà de la notion d'indépendance. On fixe comme précédemment un espace de probabilité discret $(\Omega, \mathbb{P})$ et deux événements $A$ et $B$.

Définition 1.27. *Supposons $\mathbb{P}[B] \neq 0$. Alors, la* ***probabilité conditionnelle*** *de $A$ sachant $B$ est la quantité*

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Comme $A \cap B \subset B$, ce rapport de probabilités est toujours compris entre 0 et 1. En fait, la fonction $\mathbb{P}_B : A \in \mathfrak{P}(\Omega) \to \mathbb{P}[A|B] \in [0, 1]$ est une nouvelle mesure de probabilité pour $\Omega$, en général différente de $\mathbb{P}$. Notons que $A$ et $B$ sont indépendants si et seulement si $\mathbb{P}[A|B] = \mathbb{P}[A]$. Ainsi, la notion de conditionnement englobe celle d'indépendance. Elle formalise l'intuition que l'on peut avoir du calcul des probabilités lorsqu'on a des informations partielles sur le résultat de l'expérience aléatoire : savoir que le résultat tombe dans $B$ modifie la probabilité initiale $\mathbb{P}$ en la probabilité "sachant $B$" $\mathbb{P}_B$.

EXEMPLE 1.28. On garde les mêmes notations que dans l'exemple 1.26. La probabilité conditionnelle de $A$ sachant $C$ est

$$\mathbb{P}[A|C] = \frac{\mathbb{P}[\{2,4\}]}{\mathbb{P}[\{1,2,3,4,5\}]} = \frac{\frac{1}{3}}{\frac{5}{6}} = \frac{2}{5}.$$

Sachant $C$, il y a donc moins d'une chance sur 2 pour que le résultat soit pair.

Pour conclure, il existe deux formules importantes pour le calcul des probabilités conditionnelles, qui servent parfois à calculer des probabilités d'événements en passant par un conditionnement. On suppose dans ce qui suit que $(\Omega, \mathbb{P})$ est un espace de probabilité discret et que $(A_1, \ldots, A_r)$ est un **système complet d'événements**, c'est-à-dire que $\Omega = A_1 \sqcup A_2 \sqcup \cdots \sqcup A_r$ (union disjointe), et que chaque $A_i$ a une probabilité non nulle.

THÉORÈME 1.29. *On a la formule des **probabilités totales** :*

$$\forall B,\ \mathbb{P}[B] = \sum_{i=1}^{r} \mathbb{P}[B|A_i]\,\mathbb{P}[A_i].$$

En effet, on peut écrire $B$ comme la réunion disjointe $(B\cap A_1)\sqcup(B\cap A_2)\sqcup\cdots\sqcup(B\cap A_r)$, et chaque $B \cap A_i$ a pour probabilité $\mathbb{P}[B \cap A_i] = \mathbb{P}[B|A_i]\,\mathbb{P}[A_i]$.

EXEMPLE 1.30. Poursuivant l'exemple 1.26, on a bien

$$\mathbb{P}[A] = \frac{1}{2} = \frac{2}{5} \times \frac{5}{6} + 1 \times \frac{1}{6} = \mathbb{P}[A|B]\,\mathbb{P}[B] + \mathbb{P}[A|B^c]\,\mathbb{P}[B^c].$$

THÉORÈME 1.31. *Dans le même contexte, on a la **formule de Bayes** :*

$$\forall B,\ \mathbb{P}[A_j|B] = \frac{\mathbb{P}[B|A_j]\,\mathbb{P}[A_j]}{\sum_{i=1}^{r} \mathbb{P}[B|A_i]\,\mathbb{P}[A_i]}.$$

En effet, $\mathbb{P}[A_j|B] = \frac{\mathbb{P}[A_j \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A_j]\,\mathbb{P}[A_j]}{\mathbb{P}[B]}$, et le dénominateur est donné par la formule des probabilités totales. La formule de Bayes montre en particulier que les deux probabilités conditionnelles $\mathbb{P}[A_j|B]$ et $\mathbb{P}[B|A_j]$ peuvent être très différentes.

EXEMPLE 1.32. On réalise un test de dépistage sur une population de 100 personnes. On suppose que 3 personnes sont malades, et que :

(1) Si une personne est malade, alors le test est positif avec probabilité 99.5%.

(2) Si une personne n'est pas malade, alors le test est tout de même positif avec probabilité 3% ("faux positif").

Une personne fait le test et obtient un résultat positif. Sa probabilité d'être malade est donnée par la formule de Bayes :

$$\begin{aligned}\mathbb{P}[\text{malade}|+] &= \frac{\mathbb{P}[+|\text{malade}]\,\mathbb{P}[\text{malade}]}{\mathbb{P}[+|\text{malade}]\,\mathbb{P}[\text{malade}] + \mathbb{P}[+|\text{sain}]\,\mathbb{P}[\text{sain}]} \\ &= \frac{\frac{199}{200}\frac{3}{100}}{\frac{199}{200}\frac{3}{100} + \frac{3}{100}\frac{97}{100}} = \frac{597}{1179} \simeq 0.51.\end{aligned}$$

Lorsqu'une personne a un résultat positif, elle n'a donc en fait qu'environ une chance sur 2 d'être réellement malade. Ça ne veut pas dire que le test est mauvais : son intêret est d'avoir un très faible taux de "faux négatif" (personne malade qui a un test négatif), et donc de ne pas laisser passer de personnes malades. En pratique, un second test plus coûteux / précis permet ensuite de vérifier si les personnes déjà repérées par le premier test sont effectivement malades.

---

Exercice. Calculer la probabilité au loto d'avoir les 6 bons numéros ; 5 bons numéros ; 4 bons numéros ; et 3 bons numéros.

Chapitre 2

# Variables aléatoires discrètes et continues

**Résumé.** Une *variable aléatoire* est un nombre aléatoire, qui peut être entier (dans $\mathbb{N}$ ou $\mathbb{Z}$) ou réel (dans $\mathbb{R}$). Cette distinction mène à des calculs de nature différente (sommes *vs.* intégrales), mais les mêmes notions peuvent être développées dans les deux cas : *loi*, *fonction de répartition*, *moyenne*, *variance*. Dans ce chapitre, on présente également des variables aléatoires suivant des lois dites classiques, qui apparaissent naturellement lors d'expériences aléatoires.

## 1. Variables aléatoires discrètes

**1.1. Notion de variable aléatoire.** Il y a deux façons différentes de présenter la notion de **variable aléatoire**. Dans un premier temps, considérons un espace de probabilité (discret) $(\Omega, \mathbb{P})$, et une fonction $X : \Omega \to \mathbb{Z}$, où $\mathbb{Z}$ désigne l'ensemble des entiers relatifs. Pour toute partie $A \subset \mathbb{Z}$, on peut calculer :

$$\mathbb{P}_X[A] = \mathbb{P}[X \in A] = \mathbb{P}[\{\omega \in \Omega \,|\, X(\omega) \in A\}].$$

La fonction

$$\begin{aligned} \mathbb{P}_X : \mathfrak{P}(\mathbb{Z}) &\to [0,1] \\ A &\mapsto \mathbb{P}[X \in A] \end{aligned}$$

est une mesure de probabilité sur $\mathbb{Z}$, au sens de la définition 1.19. On dit que $X$ est une variable aléatoire discrète, et que $\mathbb{P}_X$ est sa **loi**. Par additivité d'une mesure de probabilité, cette loi est entièrement déterminée par les valeurs $\mathbb{P}_X[k]$, $k \in \mathbb{Z}$, et l'on a $\sum_{k\in\mathbb{Z}} \mathbb{P}_X[k] = 1$.

Exemple 2.1. On considère le lancer de deux dés indépendants; il est donné par l'espace de probabilité $\Omega = [\![1,6]\!] \times [\![1,6]\!] = [\![1,6]\!]^2$, muni de la probabilité uniforme $\mathbb{P}[(\omega_1,\omega_2)] = \frac{1}{36}$ pour tous $\omega_1, \omega_2$. On considère la variable aléatoire $X(\omega_1,\omega_2) = \omega_1 + \omega_2$, c'est-à-dire, la somme des deux lancers. Sa loi est donnée par :

| $k$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbb{P}_X[k]$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |

En effet, pour calculer par exemple $\mathbb{P}[X = 5]$, on regarde toutes les possibilités d'obtenir pour somme $\omega_1 + \omega_2 = 5$ :

$$\mathbb{P}[X = 5] = \mathbb{P}[(\omega_1,\omega_2) \in \{(1,4),(2,3),(3,2),(4,1)\}] = \frac{4}{36}.$$

Exemple 2.2. On considère le lancer d'un seul dé à 6 faces. On peut le voir comme une variable aléatoire, qui serait issue d'un espace de probabilités très grand. Soit

$$\Omega = \{\text{façons "physiques" de lancer le dé (inclinaison, vitesse, distance à la table, } \textit{etc.})\},$$

et $\mathbb{P}$ une mesure de probabilité sur cet ensemble, qui modélise cette expérience aléatoire. Pour un certain lancer $\omega \in \Omega$, on note $X(\omega)$ le résultat de ce lancer, c'est-à-dire, le nombre dans $[\![1,6]\!]$ que l'on obtient avec ce lancer. C'est une variable aléatoire, et souvent, on suppose que sa loi est la probabilité uniforme $\mathbb{P}_X[k] = \frac{1}{6}$ pour tout $k \in [\![1,6]\!]$. Dans cet exemple, il est très difficile de travailler avec $(\Omega, \mathbb{P})$, car c'est un ensemble très grand, compte tenu du nombre de paramètres physiques qui régissent le lancer d'un dé. En revanche, on peut tout à fait travailler avec une variable aléatoire $X$ issue de cet espace, *e.g.* le résultat du lancer.

L'exemple précédent est important, et il mène à une seconde définition plus simple des variables aléatoires discrètes. Dans de nombreux cas, étant donné un espace de probabilité $(\Omega, \mathbb{P})$ et une variable aléatoire $X$ sur $\Omega$ de loi $\mathbb{P}_X$, on peut "oublier" l'espace de départ $\Omega$, et travailler seulement avec le nombre aléatoire $X$, dont on connaît toutes les probabilités (la loi $\mathbb{P}_X$). Ainsi :

Définition 2.3. *Une variable aléatoire discrète est un nombre aléatoire $X$ qui prend ses valeurs dans une partie de $\mathbb{Z}$. Il lui est associé une mesure de probabilité sur (une partie de) $\mathbb{Z}$, notée $\mathbb{P}_X$, et entièrement déterminée par les valeurs $\mathbb{P}[X = k]$, $k \in \mathbb{Z}$.*

Ainsi, on a oublié l'espace $\Omega$, et on travaillera uniquement avec les probabilités $\mathbb{P}_X[k]$, qui vérifient $\sum_{k\in\mathbb{Z}} \mathbb{P}_X[k] = 1$. Autrement dit, l'étude des variables aléatoires discrètes est celle des mesures de probabilité sur $\mathbb{Z}$.

**1.2. Espérance et variance d'une variable aléatoire.** Une variable aléatoire étant une expérience aléatoire dont le résultat est numérique, il est possible d'en définir la valeur moyenne. Elle est donnée par la définition suivante :

Définition 2.4. *La* ***moyenne****, ou* ***espérance****, d'une variable aléatoire discrète $X$ est la quantité*

$$\mathbb{E}[X] = \sum_{k\in\mathbb{Z}} \mathbb{P}_X[k]\, k,$$

*qui est bien définie au moins dans les deux cas suivants :*

*(1) si $X$ ne prend qu'un nombre fini de valeurs :* $\mathrm{card}\,\{k \in \mathbb{Z} \,|\, \mathbb{P}_X[k] > 0\} < +\infty$.

*(2) ou, si $X$ prend des valeurs positives (dans ce cas la somme $\mathbb{E}[X]$ peut être $+\infty$).*

Remarque. Supposons que $X$ soit une variable aléatoire qui provienne d'un espace de probabilité discret $(\Omega, \mathbb{P})$, comme décrit au début du paragraphe précédent. Alors, l'espérance de $X$ est aussi donnée par la formule

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} \mathbb{P}[\omega]\, X(\omega).$$

En effet, on peut scinder cette somme suivant la valeur $k$ de $X(\omega)$, et ainsi :

$$\begin{aligned}\sum_{\omega\in\Omega}\mathbb{P}[\omega]\,X(\omega) &= \sum_{k\in\mathbb{Z}}\left(\sum_{\omega\in\Omega\,|\,X(\omega)=k}\mathbb{P}[\omega]\,k\right) = \sum_{k\in\mathbb{Z}}k\left(\sum_{\omega\in\Omega\,|\,X(\omega)=k}\mathbb{P}[\omega]\right)\\ &= \sum_{k\in\mathbb{Z}}k\,\mathbb{P}[X(\omega)=k] = \sum_{k\in\mathbb{Z}}k\,\mathbb{P}_X[k]\end{aligned}$$

ce qui est la formule de la définition.

Exemple 2.5. Considérons le lancer d'un dé, avec probabilité uniforme sur $[\![1,6]\!]$. Son espérance est

$$\mathbb{E}[X] = \frac{1}{6}\times 1+\frac{1}{6}\times 2+\frac{1}{6}\times 3+\frac{1}{6}\times 4+\frac{1}{6}\times 5+\frac{1}{6}\times 6 = \frac{21}{6} = \frac{7}{2}.$$

Exemple 2.6. Considérons comme dans l'exemple 2.1 la somme du lancer de deux dés indépendants. Son espérance est

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{36}\times 2+\frac{2}{36}\times 3+\frac{3}{36}\times 4+\frac{4}{36}\times 5+\frac{5}{36}\times 6+\frac{6}{36}\times 7\\ &\quad+\frac{5}{36}\times 8+\frac{4}{36}\times 9+\frac{3}{36}\times 10+\frac{2}{36}\times 11+\frac{1}{36}\times 12\\ &= \frac{252}{36} = 7.\end{aligned}$$

On obtient le double de la moyenne du lancer d'un dé, ce qui est une conséquence de la linéarité de l'espérance de variables aléatoires : si $Y$ et $Z$ sont deux variables aléatoires discrètes, alors on a toujours

$$\mathbb{E}[Y+Z] = \mathbb{E}[Y]+\mathbb{E}[Z].$$

En particulier, pour le lancer de deux dés, on a donc, avec les notations de l'exemple 2.1,

$$\mathbb{E}[X] = \mathbb{E}[\omega_1+\omega_2] = \mathbb{E}[\omega_1]+\mathbb{E}[\omega_2] = \frac{7}{2}+\frac{7}{2} = 7.$$

Remarque. La notion d'espérance donnée par la définition 2.4 n'est pas immédiatement reliée à la notion intuitive de moyenne. Soit $X$ une variable aléatoire discrète, et $X_1, X_2, \ldots, X_n, \ldots$ des répétitions indépendantes de cette expérience aléatoire. Lorsque $X$ est le lancer d'un dé, on peut par exemple penser à $n$ lancers de dé consécutifs, dont les résultats sont notés $X_1, \ldots, X_n$. La **moyenne empirique** des variables $X_1, \ldots, X_n$ est

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}X_i,$$

qui est de nouveau un nombre aléatoire (pas forcément entier, mais rationnel si les $X_i$ prennent des valeurs entières). Il n'est pas clair *a priori* que ce nombre aléatoire $\overline{X}_n$, qui est bien ce que l'on imagine intuitivement comme une moyenne, ait un quelconque lien avec l'espérance $\mathbb{E}[X]$, qui est une valeur déterministe (non aléatoire). La raison pour laquelle les deux notions sont liées, et la justification de l'utilisation de $\mathbb{E}[X]$ comme "moyenne" de $X$, tiennent dans le résultat suivant. Sous de bonnes hypothèses (les mêmes que celles assurant la bonne définition de $\mathbb{E}[X]$), le nombre aléatoire $\overline{X}_n$ a pour limite $\mathbb{E}[X]$ lorsque $n$ tend vers l'infini, et ce avec probabilité égale à 1 :

$$\mathbb{P}\left[\lim_{n\to\infty} \overline{X}_n = \mathbb{E}[X]\right] = 1.$$

Ce résultat difficile est connu sous le nom de **loi des grands nombres**.

DÉFINITION 2.7. *La* ***variance*** *d'une variable aléatoire discrète $X$ est, si c'est bien défini,*

$$\mathrm{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

La variance de $X$ mesure la propension de $X$ à dévier de sa valeur moyenne. En particulier, $\mathrm{var}(X) = 0$ si et seulement si $X$ est une constante. On peut donner une autre formule pour $\mathrm{var}(X)$. En effet, étant entendu que l'espérance d'une constante $m$ est $\mathbb{E}[m] = m$, on a, avec $m = \mathbb{E}[X]$ :

$$\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}[(X - m^2)] = \mathbb{E}[X^2] - 2\mathbb{E}[mX] + \mathbb{E}[m^2] \\
&= \mathbb{E}[X^2] - 2m\,\mathbb{E}[X] + m^2 = \mathbb{E}[X^2] - m^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2.
\end{aligned}$$

EXEMPLE 2.8. Calculons la variance du lancer d'un dé. On a

$$\mathbb{E}[X^2] = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6},$$

et ainsi

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

Il n'est pas vrai en général que la variance d'une somme de variables aléatoires $Y$ et $Z$ soit égale à la somme des variances de $Y$ et de $Z$. En effet, on peut calculer

$$\begin{aligned}
\mathrm{var}(Y + Z) &= \mathbb{E}[(Y + Z)^2] - \mathbb{E}[Y + Z]^2 \\
&= (\mathbb{E}[Y^2] + 2\,\mathbb{E}[YZ] + \mathbb{E}[Z^2]) - (\mathbb{E}[Y^2] + 2\,\mathbb{E}[Y]\mathbb{E}[Z] + \mathbb{E}[Z^2]) \\
&= (\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) + (\mathbb{E}[Z^2] - \mathbb{E}[Z]^2) + 2(\mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]) \\
&= \mathrm{var}(Y) + \mathrm{var}(Z) + 2(\mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]),
\end{aligned}$$

et le dernier terme n'a aucune raison *a priori* d'être nul. Dans le paragraphe suivant, on donne néanmoins une condition suffisante en ce sens.

**1.3. Variables aléatoires indépendantes.** Plus tard (Chapitre 4), on verra en détail comment travailler avec des paires de variables aléatoires $(X, Y)$. On introduit ici uniquement la notion de **variables aléatoires indépendantes**, qui est le pendant pour les variables aléatoires de la notion d'événements indépendants du précédent chapitre. Soit $X$ et $Y$ deux variables aléatoires discrètes. On dit qu'elles sont indépendantes si, pour toutes parties $A$ et $B$ de $\mathbb{Z}$,

$$\mathbb{P}[X \in A \text{ et } Y \in B] = \mathbb{P}[X \in A]\, \mathbb{P}[Y \in B].$$

Il suffit en fait de le vérifier pour des parties $A$ et $B$ qui sont des singletons :

$$\forall k, l \in \mathbb{Z},\ \ \mathbb{P}[X = k \text{ et } Y = l] = \mathbb{P}[X = k]\, \mathbb{P}[Y = l].$$

Ainsi, la loi du couple $(X, Y)$ est le produit des lois de $X$ et de $Y$.

Proposition 2.9. *Si $X$ et $Y$ sont deux variables aléatoires discrètes indépendantes, alors $\mathbb{E}[XY] = \mathbb{E}[X]\, \mathbb{E}[Y]$.*

En effet,

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{m\in\mathbb{Z}} \mathbb{P}[XY = m]\, m = \sum_{m\in\mathbb{Z}} \left( \sum_{k,l \text{ tels que } m=kl} \mathbb{P}[X = k \text{ et } Y = l] \right) m \\
&= \sum_{k\in\mathbb{Z},\, l\in\mathbb{Z}} \mathbb{P}[X = k \text{ et } Y = l]\, kl = \sum_{k\in\mathbb{Z},\, l\in\mathbb{Z}} \mathbb{P}[X = k]\, k\ \mathbb{P}[Y = l]\, l \\
&= \left( \sum_{k\in\mathbb{Z}} \mathbb{P}[X = k]\, k \right) \left( \sum_{l\in\mathbb{Z}} \mathbb{P}[Y = l]\, l \right) = \mathbb{E}[X]\, \mathbb{E}[Y].
\end{aligned}$$

Une conséquence de ce fait est :

Théorème 2.10. *Si $Y$ et $Z$ sont deux variables aléatoires indépendantes, alors on a $\mathrm{var}(Y + Z) = \mathrm{var}(Y) + \mathrm{var}(Z)$.*

En effet, on a vu que de façon générale,

$$\mathrm{var}(Y + Z) - \mathrm{var}(Y) - \mathrm{var}(Z) = 2(\mathbb{E}[YZ] - \mathbb{E}[Y]\, \mathbb{E}[Z]).$$

Dans le cas de variables indépendantes, le terme de droite dans cette égalité est nul d'après la proposition précédente.

Exemple 2.11. La variance de la somme $X$ de deux lancers de dés est donc $\mathrm{var}(X) = 2 \times \frac{35}{12} = \frac{35}{6}$, puisque la variance du lancer d'un dé est $\frac{35}{12}$ d'après l'exemple 2.8.

Exemple 2.12. Les résultats sur l'espérance d'un produit et sur la variance d'une somme deviennent faux si l'on prend deux variables non indépendantes. Prenons par exemple $X = -Y$, avec $\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = \frac{1}{2}$. L'espérance de $X$ est 0, et la variance de $X$ est 1. Or, $XY = -X^2 = -1$ avec probabilité 1, donc

$$\mathbb{E}[XY] = -1 \neq 0 \times 0 = \mathbb{E}[X]\, \mathbb{E}[Y].$$

De même, $\mathrm{var}(X + Y) = \mathrm{var}(0) = 0 \neq 1 + 1 = \mathrm{var}(X) + \mathrm{var}(Y)$.

**1.4. Lois discrètes classiques.** Pour conclure notre étude des variables aléatoires discrètes, nous donnons une liste de lois dites classiques, avec dans chaque cas le calcul de l'espérance et de la variance.

1.4.1. *Loi uniforme.* Soit $[\![a+1,b]\!]$ un intervalle d'entiers, qui est de cardinal $b-a$. Une variable aléatoire de **loi uniforme** sur cet intervalle est une variable $X$ avec

$$\mathbb{P}[X=k]=\begin{cases}\frac{1}{b-a} & \text{si } a+1\leq k\leq b,\\ 0 & \text{sinon.}\end{cases}$$

Cette variable est donc équiprobablement répartie sur $[\![a+1,b]\!]$. Par exemple, le lancer d'un dé est une variable uniforme sur $[\![1,6]\!]$. On note $X\sim\mathcal{U}([\![a+1,b]\!])$ pour dire qu'une variable suit la loi uniforme sur $[\![a+1,b]\!]$. Dans ce cas, le calcul de l'espérance et de la variance de $X$ est lié aux identités suivantes :

$$\sum_{k=1}^{n}k=\frac{n(n+1)}{2}\qquad;\qquad\sum_{k=1}^{n}k^2=\frac{n(n+1)(2n+1)}{6}.$$

On en déduit

$$\begin{aligned}\mathbb{E}[X]&=\sum_{k=a+1}^{b}\frac{k}{b-a}=\frac{1}{b-a}\left(\sum_{k=1}^{b}k-\sum_{k=1}^{a}k\right)=\frac{1}{b-a}\left(\frac{b^2+b}{2}-\frac{a^2+a}{2}\right)=\frac{a+b+1}{2};\\ \mathbb{E}[X^2]&=\sum_{k=a+1}^{b}\frac{k^2}{b-a}=\frac{1}{b-a}\left(\sum_{k=1}^{b}k^2-\sum_{k=1}^{a}k^2\right)\\ &=\frac{1}{b-a}\left(\frac{2b^3+3b^2+b}{6}-\frac{2a^3+3a^2+a}{6}\right)=\frac{2(a^2+ab+b^2)+3(a+b)+1}{6};\\ \mathrm{var}(X)&=\mathbb{E}[X^2]-\mathbb{E}[X]^2=\frac{(b-a)^2-1}{12}.\end{aligned}$$

Exemple 2.13. Avec $a=0$ et $b=6$, on retrouve bien $\frac{a+b+1}{2}=\frac{7}{2}$ pour l'espérance d'un lancer de dé, et $\frac{(b-a)^2-1}{12}=\frac{35}{12}$ pour la variance.

1.4.2. *Loi de Bernoulli.* Une variable aléatoire suit une **loi de Bernoulli** de paramètre $p$ si $X$ prend pour valeurs 0 ou 1, avec

$$\mathbb{P}[X=1]=p\qquad;\qquad\mathbb{P}[X=0]=1-p.$$

On note dans ce cas $X\sim\mathcal{B}(p)$. Une telle variable modélise une expérience à deux résultats possibles ("succès" et "échec"), et l'on compte 1 en cas de succès. Les espérance et variance de $X$ sont

$$\begin{aligned}\mathbb{E}[X]&=p\times 1+(1-p)\times 0=p;\\ \mathrm{var}(X)&=\mathbb{E}[X^2]-\mathbb{E}[X]^2=p-p^2=p(1-p).\end{aligned}$$

1.4.3. *Loi binomiale.* Considérons $n$ expériences de Bernoulli $X_1, \ldots, X_n$ indépendantes et de même loi $\mathcal{B}(p)$. On s'intéresse au nombre $X$ de succès obtenus, qui est la somme $X = X_1 + X_2 + \cdots + X_n$. On a, pour $k \in [\![0, n]\!]$,

$$\begin{aligned}
\mathbb{P}[X = k] &= \sum_{(x_1,\ldots,x_n)\in\{0,1\}^n \,|\, k \text{ entrées } x_i \text{ sont égales à } 1} \mathbb{P}[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] \\
&= \sum_{(x_1,\ldots,x_n)\in\{0,1\}^n \,|\, k \text{ entrées } x_i \text{ sont égales à } 1} p^k (1-p)^{n-k} \\
&= (\text{nombre de suites dans } \{0,1\}^n \text{ avec } k \text{ entrées égales à } 1) \times p^k (1-p)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

On dit que $X$ suit une **loi binomiale** de paramètres $n$ et $p$, ce qu'on note $X \sim \mathcal{B}(n, p)$. Comme $X = X_1 + X_2 + \cdots + X_n$ est somme de variables indépendantes, son espérance et sa variance sont obtenues en sommant celles calculées pour des variables de Bernoulli :

$$\mathbb{E}[X] = np \quad ; \quad \mathrm{var}(X) = np(1-p).$$

Exemple 2.14. Deux joueurs de même niveau s'affrontent lors de 10 parties de poker. Pour chaque partie, le joueur $A$ gagne avec probabilité $\frac{1}{2}$, et il perd avec cette même probabilité. Le nombre de parties remportées au final suit la loi binomiale $\mathcal{B}(10, \frac{1}{2})$, dont les probabilités sont (avec deux chiffres significatifs) :

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbb{P}[X = k]$ | 0.00098 | 0.0098 | 0.044 | 0.12 | 0.21 | 0.25 | 0.21 | 0.12 | 0.044 | 0.0098 | 0.00098 |

L'espérance est $\frac{10}{2} = 5$ et la variance est $\frac{10}{4} = \frac{5}{2}$.

Remarque. La loi binomiale est associée à l'expérience aléatoire dite de **tirage avec remise**. On considère une urne avec $B$ boules blanches et $N$ boules noires, et on effectue $n$ tirages dans cette urne, en remettant à chaque fois la boule tirée dans l'urne. Soit $X$ le nombre de boules blanches obtenues au cours de ces $n$ tirages. Comme chaque tirage est une expérience de Bernoulli de paramètre $p = \frac{B}{B+N}$, $X$ suit une loi binomiale $\mathcal{B}(n, \frac{B}{B+N})$. Son espérance est $\frac{nB}{B+N}$.

1.4.4. *Loi hypergéométrique.* On considère une population de $N$ individus, avec deux types d'individus : $M$ individus de type 1, et $N - M$ individus de type 2. On tire $n \leq \min(M, N - M)$ personnes au hasard parmi les $N$ individus, et on note $X$ le nombre d'individus de type 1 parmi cette sous-population de $n$ personnes. Ce nombre peut être n'importe quel entier entre 0 et $n$. Le nombre total de choix possibles pour la sous-population de $n$ personnes est le coefficient binomial $\binom{N}{n}$. Si l'on veut choisir une sous-population avec exactement $k$ personnes de type 1, une façon de le faire est de choisir $k$ personnes parmi les $M$ individus de type 1 ($\binom{M}{k}$ possibilités), puis, de compléter cette population par $n-k$ personnes choisies parmi les $N-M$ individus de type 2 ($\binom{N-M}{n-k}$ possibilités). Donc,

$$\begin{aligned}\mathbb{P}[X=k] &= \frac{\text{nombre de choix donnant } k \text{ personnes de type 1}}{\text{nombre total de possibilités}}\\ &= \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \quad \text{pour } k \in [\![0,n]\!]\,.\end{aligned}$$

On dit que $X$ suit une **loi hypergéométrique** de paramètres $N$, $M$ et $n$, ce qu'on note $X \sim \mathcal{H}(N,M,n)$. L'identité

$$\sum_{k=0}^{n} \binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n},$$

qui découle du fait que la loi $\mathcal{H}(N,M,n)$ est une probabilité, est appelée **formule de Vandermonde**.

Calculons l'espérance de $\mathcal{H}(M,N,n)$ :

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{n} k\,\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}}\sum_{k=1}^{n} M\binom{M-1}{k-1}\binom{(N-1)-(M-1)}{(n-1)-(k-1)}\\ &= \frac{M}{\binom{N}{n}}\sum_{l=0}^{n-1}\binom{M-1}{l}\binom{(N-1)-(M-1)}{(n-1)-l} = M\frac{\binom{N-1}{n-1}}{\binom{N}{n}} = n\,\frac{M}{N}\end{aligned}$$

en faisant sur la seconde ligne le changement de variables $l=k-1$, et en utilisant ensuite la formule de Vandermonde. Le même type de manipulation donne :

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{k=0}^{n} k(k-1)\,\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\\ &= \frac{1}{\binom{N}{n}}\sum_{k=2}^{n} M(M-1)\binom{M-2}{k-2}\binom{(N-2)-(M-2)}{(n-2)-(k-2)}\\ &= \frac{M(M-1)}{\binom{N}{n}}\sum_{l=0}^{n-2}\binom{M-2}{l}\binom{(N-2)-(M-2)}{(n-2)-l}\\ &= M(M-1)\frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)\,M(M-1)}{N(N-1)}.\end{aligned}$$

On en déduit la valeur de $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X]$, et celle de

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = n\,\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right).$$

EXEMPLE 2.15. On tire deux mains de 2 cartes pour deux joueurs de poker, et 3 cartes que l'on pose face visible sur le tapis (*flop*). Le joueur $A$ a un coeur parmi ses 2 cartes, et il y a 2 coeurs parmi les cartes visibles. La suite du jeu (*turn* et *river*) révèle deux cartes en plus des 3 déjà visibles. Soit $X$ le nombre de coeurs parmi ces 2 nouvelles cartes. Ce nombre suit une loi hypergéométrique

$$\mathcal{H}(47, 10, 2).$$

En effet, il reste 2 cartes à tirer parmi les $47 = 52-2-3$ cartes que le joueur $A$ ne connaît pas, et le nombre de coeurs parmi ces cartes est $(52/4) - 3 = 13 - 3 = 10$. En particulier, la probabilité pour que les deux dernières cartes révélées soient des coeurs, et donc pour que le joueur $A$ puisse constituer une main de 5 cartes de même couleur, est

$$\frac{\binom{10}{2}\binom{37}{0}}{\binom{47}{2}} = \frac{45}{1081} = 0.042.$$

Si en revanche le joueur $A$ avait déjà 2 coeurs dans sa main, alors le nombre $X$ de coeurs supplémentaires parmi les 2 dernières cartes suivrait une loi

$$\mathcal{H}(47, 9, 2).$$

Alors, la probabilité pour que $X$ soit plus grand que 1, et donc pour que le joueur $A$ puisse constituer une main de 5 cartes de même couleur, est

$$1 - \mathbb{P}[X = 0] = 1 - \frac{\binom{9}{0}\binom{38}{2}}{\binom{47}{2}} = 1 - \frac{703}{1081} = \frac{378}{1081} = 0.35.$$

Remarque. La loi hypergéométrique est associée à l'expérience de **tirage sans remise**. On considère une urne avec $B$ boules blanches et $N$ boules noires, et on effectue $n \leq \min(B, N)$ tirages dans cette urne, en ne remettant pas les boule tirées dans l'urne. Soit $X$ le nombre de boules blanches obtenues au cours de ces $n$ tirages. La variable $X$ suit une loi hypergéométrique $\mathcal{H}(B + N, B, n)$. En particulier, son espérance est $\frac{nB}{B+N}$, ce qui est la même valeur que dans le cas avec remise. En revanche, la variance est différente par rapport au cas avec remise.

1.4.5. *Loi géométrique.* Si l'on effectue une suite $X_1, X_2, \ldots, X_n, \ldots$ d'expérience de Bernoulli indépendantes, on peut se demander combien de temps il faut attendre pour avoir le premier succès, c'est-à-dire, quel est le plus petit $k \geq 1$ tel que $X_k = 1$. Supposons que toutes les variables $X_i$ sont indépendantes et de même loi $\mathcal{B}(n, p)$. Si $X$ est la loi du temps du premier succès, alors on a $X = k$ si et seulement si $X_1 = X_2 = \cdots = X_{k-1} = 0$ et $X_k = 1$. Par conséquent,

$$\mathbb{P}[X = k] = \mathbb{P}[X_1 = X_2 = \cdots = X_{k-1} = 0 \text{ et } X_k = 1] = (1 - p)^{k-1}p.$$

On dit que $X$ suit une **loi géométrique** de paramètre $p$ notée $\mathcal{G}(p)$. Le calcul de l'espérance et la variance de $X$ sont liés à l'identité

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \text{ si } |q| < 1,$$

et aux deux premières dérivées de cette identité :

$$\sum_{k=1}^{\infty} k\, q^{k-1} = \frac{1}{(1-q)^2} \qquad ; \qquad \sum_{k=2}^{\infty} k(k-1)\, q^{k-2} = \frac{2}{(1-q)^3}.$$

Avec $q = 1 - p$, on obtient

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k\, q^{k-1} p = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

D'autre part,

$$\mathbb{E}[X(X-1)] = \sum_{k=2}^{\infty} k(k-1)\, q^{k-1} p = \frac{2qp}{(1-q)^3} = \frac{2(1-p)p}{p^3} = \frac{2}{p}\left(\frac{1}{p} - 1\right).$$

On en déduit que $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \frac{2-p}{p^2}$, et que

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1-p}{p^2} = \frac{1}{p}\left(\frac{1}{p} - 1\right).$$

Remarque. La valeur de l'espérance est facile à retenir : si une expérience de Bernoulli a une chance sur $k \geq 1$ d'être un succès, alors on s'attend à devoir attendre en moyenne $k$ expériences pour avoir un succès. C'est pourquoi l'espérance d'une loi géométrique est l'inverse du paramètre des expériences de Bernoulli sous-jacentes.

Remarque. On appelle aussi variable géométrique une variable aléatoire de loi

$$\mathbb{P}[Y = k] = (1-p)^k p \quad \text{pour tout } k \geq 0.$$

Une telle variable est décalée de $-1$ par rapport à une variable $X \sim \mathcal{G}(p)$ : $Y = X - 1$. Elle représente le nombre d'échecs avant le premier succès dans une suite d'expériences de Bernoulli indépendantes de paramètre $p$.

1.4.6. *Loi de Poisson.* Finalement, il existe une dernière loi classique discrète qu'il est utile de connaître, appelée **loi de Poisson**. Pour tout paramètre $\lambda$ réel, on rappelle l'identité

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \mathrm{e}^{\lambda},$$

qu'on peut même prendre comme définition de la fonction exponentielle. Par conséquent, si $\lambda \geq 0$, alors la formule

$$\mathbb{P}[X = k] = \mathrm{e}^{-\lambda}\, \frac{\lambda^k}{k!}$$

est une mesure de probabilité sur $\mathbb{N}$, la loi de Poisson de paramètre $\lambda$, notée $\mathcal{P}(\lambda)$. On expliquera plus tard (§3) dans quel contexte cette loi apparaît : elle est couramment utilisée pour le nombre de réalisations d'événements aléatoires rares, tels que : le nombre de pannes d'un outil, le nombre de sinistres couverts par une assurance, le nombre de catastrophes climatiques dans un pays au cours d'une année, le nombre de poissons pêchés dans un grand lac, *etc.* L'espérance et la variance de $X$ sont calculées comme suit :

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\,\mathrm{e}^{-\lambda}\,\frac{(\lambda)^k}{k!} = \mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda\mathrm{e}^{-\lambda}\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} = \lambda;$$

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)\,\mathrm{e}^{-\lambda}\,\frac{(\lambda)^k}{k!} = \mathrm{e}^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!} = \lambda^2\mathrm{e}^{-\lambda}\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} = \lambda^2;$$

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Dans le tableau qui suit, on a résumé tous les calculs sur les lois classiques discrètes :

| loi | valeurs possibles | probabilités | espérance | variance |
|---|---|---|---|---|
| $\mathcal{U}([\![a+1,b]\!])$ | $[\![a+1,b]\!]$ | $\frac{1}{b-a}$ | $\frac{a+b+1}{2}$ | $\frac{(b-a)^2-1}{12}$ |
| $\mathcal{B}(p)$ | $\{0,1\}$ | $p$ ou $1-p$ | $p$ | $p(1-p)$ |
| $\mathcal{B}(n,p)$ | $[\![0,n]\!]$ | $\binom{n}{k}\,p^k(1-p)^{n-k}$ | $np$ | $np(1-p)$ |
| $\mathcal{H}(N,M,n)$ | $[\![0,n]\!]$ | $\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$ | $n\,\frac{M}{N}$ | $n\,\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$ |
| $\mathcal{G}(p)$ | $\mathbb{N}^*$ | $p(1-p)^{k-1}$ | $\frac{1}{p}$ | $\frac{1}{p}\left(\frac{1}{p}-1\right)$ |
| $\mathcal{P}(\lambda)$ | $\mathbb{N}$ | $\mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}$ | $\lambda$ | $\lambda$ |

## 2. Variables aléatoires continues

Jusqu'ici, on a travaillé avec des variables aléatoires qui prenaient uniquement des valeurs entières, donc, dans l'ensemble dénombrable $\mathbb{Z}$. Or, on aimerait également pouvoir travailler avec des nombres aléatoires réels, qui peuvent prendre n'importe quelle valeur dans un intervalle $[a,b]$, ou même dans tout $\mathbb{R}$. Malheureusement, le cadre précédemment décrit ne permet pas de traîter ce type d'éxpérience aléatoire, car $\mathbb{R}$, ou tout intervalle $[a,b]$ avec $a<b$, est un ensemble infini non dénombrable. Sans rentrer plus dans les détails, le problème est qu'il y a trop de parties de $\mathbb{R}$ pour pouvoir définir de façon convenable une mesure de probabilité $\mathbb{P} : \mathfrak{P}(\mathbb{R}) \to [0,1]$. Néanmoins, si l'on demande seulement de pouvoir calculer la probabilité pour un nombre aléatoire de tomber dans un intervalle arbitraire de $\mathbb{R}$ (au lieu d'une partie arbitraire), alors la plupart des choses dites précédemment restent vraies.

**2.1. Densité et fonction de répartition.** Ainsi, par analogie avec la définition 2.3, on est amené à définir :

DÉFINITION 2.16. *Une variable aléatoire continue est un nombre aléatoire $X$ qui prend ses valeurs dans (un intervalle) de $\mathbb{R}$. On demande qu'il existe une fonction $f_X$ positive et continue par morceaux, la* ***densité*** *de $X$, telle que pour tout intervalle $[a,b] \subset \mathbb{R}$,*

$$\mathbb{P}[X \in [a,b]] = \int_a^b f_X(s)\,ds.$$

*Pour que $f_X$ soit la densité d'une variable aléatoire réelle, il faut et il suffit que $f_X$ soit positive et que*

$$\mathbb{P}[X \in \mathbb{R}] = \int_{\mathbb{R}} f_X(s)\,ds = 1.$$

EXEMPLE 2.17. Soit $f_X(x)$ la fonction qui vaut 1 si $x \in [0,1]$, et 0 sinon. C'est la densité de probabilité d'une variable aléatoire $X$ qui est uniforme sur l'intervalle $[0,1]$, et peut prendre n'importe quelle valeur dans cet intervalle "avec la même probabilité". Cet énoncé doit être pris avec précaution, car pour n'importe quelle valeur réelle $x \in X$, on a $\mathbb{P}[X = x] = \int_x^x 1\,ds = 0$. En revanche, si $[a,b] \subset [0,1]$ est un intervalle de taille non nulle, alors $\mathbb{P}[a \leq X \leq b] = b - a > 0$. La notion de densité est la bonne pour avoir une répartition "diffuse" des probabilités sur tout un intervalle, mais il convient de noter qu'à ce titre, toute valeur fixe $x \in \mathbb{R}$ a probabilité 0 d'être *exactement* la valeur d'une variable aléatoire continue (à densité) $X$.

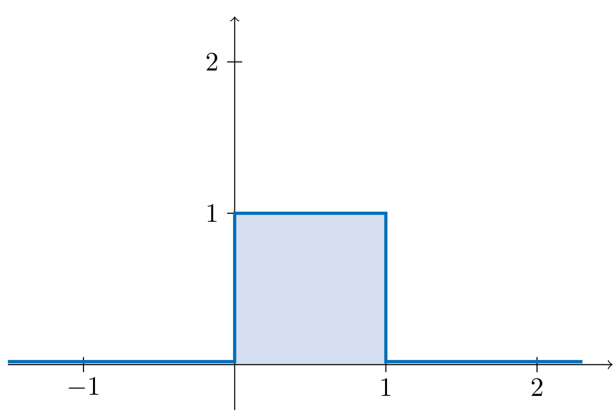


FIGURE 1. Densité d'une variable aléatoire uniforme sur $[0,1]$.

EXEMPLE 2.18. Soit $f_X(x)$ la fonction qui vaut $e^{-x}$ si $x \in \mathbb{R}_+$, et 0 sinon. Comme $\int_0^\infty e^{-s}\,ds = 1$, $f_X$ est la densité d'une variable aléatoire continue $X$, qui peut prendre n'importe quelle valeur positive. Pour tout intervalle $[a,b]$ avec $0 < a < b$,

$$\mathbb{P}[X \in [a,b]] = \int_a^b e^{-s}\,ds = e^{-a} - e^{-b}.$$

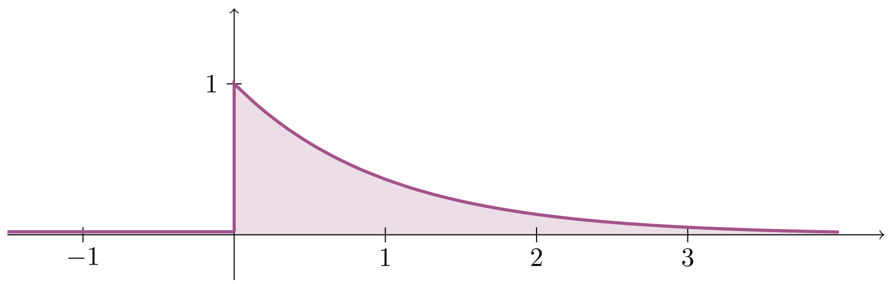


FIGURE 2. Densité d'une variable aléatoire exponentielle sur $\mathbb{R}$.

La densité $f_X$ d'une variable aléatoire continue permet de calculer toutes les probabilités d'intervalles $\mathbb{P}[X \in [a,b]]$ ; à ce titre, on dit qu'elle détermine entièrement la loi de $X$. La fonction $x \in \mathbb{R} \mapsto f_X(x)$ joue un rôle analogue pour les variables aléatoires continues à la fonction $k \in \mathbb{Z} \mapsto \mathbb{P}_X[k]$ pour les variables aléatoires discrètes. Pour les variables aléatoires continues, il existe un autre quantité qui détermine entièrement la loi : la **fonction de répartition** $F_X$, qui est définie par

$$F_X(t) = \mathbb{P}[X \leq t] = \int_{-\infty}^{t} f_X(s)\, ds.$$

PROPOSITION 2.19. *La fonction de répartition $F_X$ d'une variable aléatoire continue est une fonction croissante, dont les limites en $-\infty$ et $+\infty$ sont respectivement $0$ et $1$. C'est l'unique primitive de $f_X$ qui a ces limites, et réciproquement, la densité est obtenue à partir de la fonction de répartition par dérivation :*

$$f_X(x) = F_X'(x).$$

EXEMPLE 2.20. Soit $U$ la variable aléatoire de densité $f_U(u) = 1$ si $u \in [0,1]$, et $0$ sinon. La fonction de répartition de $U$ est

$$F_U(t) = \begin{cases} 0 & \text{si } t < 0, \\ t & \text{si } 0 \leq t \leq 1, \\ 1 & \text{si } t > 1. \end{cases}$$

En effet, si l'on veut par exemple calculer $F_U(t)$ pour $t \in [0,1]$, on a :

$$F_U(t) = \int_{-\infty}^{t} f_U(u)\, du = \int_{-\infty}^{0} 0\, du + \int_{0}^{t} 1\, du = t.$$

EXEMPLE 2.21. Soit $X$ la variable aléatoire continue de fonction de répartition $F_X(t) = \frac{\arctan t}{\pi} + \frac{1}{2}$. C'est bien une fonction croissante qui a pour limites respectivement $0$ et $1$ en $-\infty$ et $+\infty$. La densité de cette variable aléatoire est $\frac{1}{\pi}$ fois la dérivée de la fonction arctan, c'est-à-dire,

$$f_X(s) = \frac{1}{\pi(1+s^2)}.$$

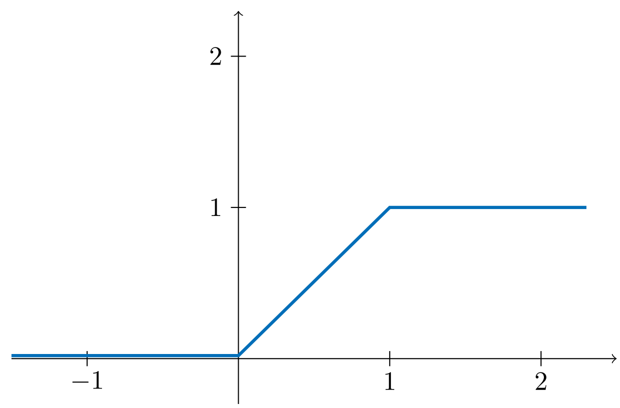


FIGURE 3. Fonction de répartition d'une variable aléatoire uniforme sur $[0,1]$.

Selon les calculs, la densité ou la fonction de répartition est la quantité la plus appropriée à manipuler. Notons que la fonction de répartition détermine aisément toutes les probabilités des intervalles : pour tous $a < b$,

$$\mathbb{P}[X \in [a,b]] = \int_a^b f_X(s)\,ds = \int_{-\infty}^b f_X(s)\,ds - \int_{-\infty}^a f_X(s)\,ds = F_X(b) - F_X(a).$$

**2.2. Espérance et variance d'une variable aléatoire continue.** Pour une variable aléatoire discrète, on a défini l'espérance de $X$ par la formule $\mathbb{E}[X] = \sum_{k\in\mathbb{Z}} \mathbb{P}_X[k]\,k$. Par analogie, on peut définir l'espérance d'une variable aléatoire continue par la formule :

$$\mathbb{E}[X] = \int_{\mathbb{R}} f_X(s)\,s\,ds.$$

Des conditions suffisantes pour que cette intégrale soit bien définie sont :

(1) si $f_X$ est non nulle uniquement sur un intervalle borné $[m,M]$,

(2) ou, si $f_X$ est non nulle uniquement sur $\mathbb{R}_+$.

Plus généralement, pour toute fonction $h$, l'espérance de la variable aléatoire $h(X)$ est donnée par la formule

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(s)\,f_X(s)\,ds.$$

En particulier, $\mathbb{E}[X^2] = \int_{\mathbb{R}} f_X(s)\,s^2\,ds$, et la variance de $X$ est définie comme dans le cas discret par

$$\mathrm{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Notons que dans la formule pour $\mathbb{E}[h(X)]$, on intègre $h(s)$ toujours contre la densité $f_X(s)\,ds$, quelque soit la fonction $h$ (cette propriété est parfois appelée théorème de transfert).

Exemple 2.22. Soit $U$ la variable aléatoire continue de densité donnée par $f_U(u) = 1$ si $u \in [0,1]$, et 0 sinon. L'espérance de $U$, ou valeur moyenne, est

$$\mathbb{E}[U] = \int_0^1 1\, u\, du = \left[\frac{u^2}{2}\right]_0^1 = \frac{1}{2}.$$

La variance de $U$ est donnée par

$$\mathbb{E}[U^2] = \int_0^1 1\, u^2\, du = \frac{1}{3} \qquad ; \qquad \mathrm{var}(U) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Exemple 2.23. Soit $X$ la variable aléatoire continue de densité donnée par $f_X(x) = \mathrm{e}^{-x}$ si $x \geq 0$, et 0 sinon. L'espérance de $X$ est

$$\mathbb{E}[X] = \int_0^\infty \mathrm{e}^{-x}\, x\, dx = \big[-\mathrm{e}^{-x}\, x\big]_0^\infty + \int_0^\infty \mathrm{e}^{-x}\, dx = 1,$$

et la variance de $X$ est donnée par

$$\mathbb{E}[X^2] = \int_0^\infty \mathrm{e}^{-x}\, x^2\, dx = \big[-\mathrm{e}^{-x}\, x^2\big]_0^\infty + 2\int_0^\infty \mathrm{e}^{-x}\, x\, dx = 2\,\mathbb{E}[X] = 2$$
$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 2 - 1 = 1.$$

Donnons une liste de propriétés de l'espérance et de la variance qui sont valables à la fois pour les variables discrètes et les variables continues :

(1) linearité de l'espérance : $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, et $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$ pour toute constante $\lambda$.
(2) positivité de l'espérance : si $X \geq 0$ avec probabilité 1, alors $\mathbb{E}[X] \geq 0$. Plus généralement, si $X \geq Y$ sont deux variables aléatoires, alors $\mathbb{E}[X] \geq \mathbb{E}[Y]$.
(3) variance : $\mathrm{var}(aX+b) = a^2\, \mathrm{var}(X)$ pour tous coefficients $a$ et $b$.

Remarque. L'espérance et la variance d'une variable aléatoire continue ont la même interprétation que dans le cas discret : l'espérance de $X$ est la valeur moyenne que l'on peut obtenir en répétant un grand nombre de fois l'expérience aléatoire de tirage de $X$, tandis que la variance est la propension qu'a $X$ à dévier de cette valeur moyenne. Un résultat mathématique qui rend concret cette intuition est l'**inégalité de Bienaymé-Chebyshev** : si l'espérance et la variance de $X$ sont bien définies, alors pour tout seuil $a > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq a] \leq \frac{\mathrm{var}(X)}{a^2}.$$

En effet, notant $m = \mathbb{E}[X]$, on a, pour une variable aléatoire continue,

$$\mathbb{P}[|X - m| \geq a] = \mathbb{P}\left[\frac{(X-m)^2}{a^2} \geq 1\right] = \mathbb{E}[h(X)],$$

où $h(X)$ est la fonction qui vaut 1 si $\frac{(X-m)^2}{a^2} \geq 1$, et 0 sinon. Notons qu'on a toujours

$$h(X) \leq k(X) = \frac{(X-m)^2}{a^2};$$

en effet, c'est vrai dans les deux cas ($\frac{(X-m)^2}{a^2} \geq 1$ et $\frac{(X-m)^2}{a^2} < 1$). Donc,

$$\mathbb{P}[|X-m| \geq a] = \mathbb{E}[h(X)] \leq \mathbb{E}[k(X)] = \int_{-\infty}^{\infty} \frac{(s-m)^2}{a^2} f_X(s)\, ds = \frac{\operatorname{var}(X)}{a^2}.$$

Le résultat est aussi vrai pour une variable aléatoire discrète, en remplaçant les intégrales par des sommes.

**2.3. Lois continues classiques.** Comme dans le cas discret, il existe une liste de lois continues dites classiques, dont nous donnons dans ce qui suit les densités, fonctions de répartition, espérances et variances.

2.3.1. *Loi uniforme.* Soit $[a,b]$ un intervalle de longueur $b-a>0$. Une variable aléatoire $X$ suit une loi uniforme continue sur ce segment si elle a densité

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

On note alors $X \sim \mathcal{U}([a,b])$. C'est l'analogue de l'exemple 2.17, mais sur un segment arbitraire $[a,b]$ au lieu de $[0,1]$. La constante $\frac{1}{b-a}$ est choisie pour que $\int_{\mathbb{R}} f_X(s)\, ds = 1$. La fonction de répartition de $X$ est

$$F_X(t) = \begin{cases} 0 & \text{si } t < a, \\ \frac{t-a}{b-a} & \text{si } a \leq t \leq b\ . \\ 1 & \text{si } t > b. \end{cases}$$

L'espérance de $X$ est

$$\mathbb{E}[X] = \frac{1}{b-a} \int_a^b s\, ds = \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \frac{a+b}{2},$$

et la variance de $X$ est

$$\operatorname{var}(X) = \frac{1}{b-a}\int_a^b \left(s - \frac{a+b}{2}\right)^2 ds = \frac{1}{3(b-a)}\left[\left(s-\frac{a+b}{2}\right)^3\right]_a^b = \frac{(b-a)^2}{12}.$$

On notera la ressemblance entre ces formules et celles pour l'espérance et la variance d'une loi uniforme discrète. De fait, la loi $\mathcal{U}([a,b])$ est l'analogue continu de la loi discrète $\mathcal{U}([\![a,b]\!])$.

2.3.2. *Loi exponentielle.* Soit $\lambda$ un réel strictement positif. On dit qu'une variable aléatoire $X$ suit une **loi exponentielle** de paramètre $\lambda$ si elle a densité

$$f_X(s) = \begin{cases} \lambda \mathrm{e}^{-\lambda s} & \text{si } s \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

On note alors $X \sim \mathcal{E}(\lambda)$. La fonction de répartition de $X$ est alors

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0, \\ 1 - \mathrm{e}^{-\lambda t} & \text{si } t \geq 0. \end{cases}$$

Par conséquent, pour $t$ positif, $\mathbb{P}[X \geq t] = 1 - \mathbb{P}[X < t] = 1 - F_X(t) = \mathrm{e}^{-\lambda t}$, ce qui justifie l'appellation de variable exponentielle. L'exemple 2.18 est une variable exponentielle de paramètre $\lambda = 1$. L'espérance d'une variable $X \sim \mathcal{E}(\lambda)$ est

$$\mathbb{E}[X] = \int_0^\infty \lambda \mathrm{e}^{-\lambda s}\, s\, ds =_{(t=\lambda s)} \frac{1}{\lambda} \int_0^\infty \mathrm{e}^{-t}\, t\, dt = \frac{1}{\lambda},$$

et la variance est donnée par

$$\mathbb{E}[X^2] = \int_0^\infty \lambda \mathrm{e}^{-\lambda s}\, s^2\, ds =_{(t=\lambda s)} \frac{1}{\lambda^2} \int_0^\infty \mathrm{e}^{-t}\, t^2\, dt = \frac{2}{\lambda^2}$$

$$\mathrm{var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Exemple 2.24. Soit $T$ le temps que l'on doit attendre avant la première panne d'une voiture. On peut imaginer que, si à un instant $t$ la voiture n'a pas encore eu de panne, alors elle est "comme neuve" à cet instant, et donc que pour tous temps $s, t \geq 0$,

$$\frac{\mathbb{P}[T \geq t+s]}{\mathbb{P}[T \geq t]} = \mathbb{P}[T \geq t+s \,|\, T \geq t] = \mathbb{P}[T \geq s],$$

puisque conditionnellement au fait que $T \geq t$, on regarde à l'instant $t$ une voiture qui a les mêmes temps de panne qu'une voiture neuve. Soit $F_T(t)$ la fonction de répartition de $T$, et $G_T(t) = 1 - F_T(t) = \mathbb{P}[T \geq t]$. On vient de voir que

$$G_T(s+t) = G_T(s)\, G_T(t)$$

pour tous $s, t \geq 0$. Fixons $t$ et dérivons cette équation par rapport à $s$ :

$$G_T'(s+t) = G_T'(s)\, G_T(t) \qquad ; \qquad G_T'(t) = G_T'(0)\, G_T(t)$$

la seconde partie étant obtenue en prenant $s = 0$. Si $-\lambda = G_T'(0)$, alors l'unique solution de cette équation différentielle avec $G_T(0) = \mathbb{P}[T \geq 0] = 1$ est $G_T(t) = \mathrm{e}^{-\lambda t}$, d'où $F_T(t) = 1 - \mathrm{e}^{-\lambda t}$. Par conséquent, $T$ suit une loi exponentielle avec un certain paramètre $\lambda > 0$, qui est relié au temps moyen de panne par l'équation $\mathbb{E}[T] = \frac{1}{\lambda}$.

2.3.3. *Loi gaussienne.* Une troisième famille de lois continues classiques est donnée par les densités

$$f_X(s) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{(s-m)^2}{2\sigma^2}},$$

où $m \in \mathbb{R}$ et $\sigma^2$ est un paramètre réel strictement positif. Pour vérifier que la formule précédente donne bien une probabilité sur $\mathbb{R}$, remarquons d'abord que

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{(s-m)^2}{2\sigma^2}}\, ds =_{\left(u=\frac{s-m}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{u^2}{2}}\, du.$$

Il suffit donc de montrer que $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{u^2}{2}}\, du = \frac{I}{\sqrt{2\pi}} = 1$. Or,

$$I^2 = \iint_{\mathbb{R}^2} \mathrm{e}^{-\frac{u^2+v^2}{2}}\, du\, dv = \int_{r=0}^\infty \int_{\theta=0}^{2\pi} r\mathrm{e}^{-\frac{r^2}{2}}\, dr\, d\theta = 2\pi \int_{r=0}^\infty r\mathrm{e}^{-\frac{r^2}{2}}\, dr = 2\pi$$

en faisant le changement de variables polaire $u = r\cos\theta$ et $v = r\sin\theta$. Ainsi, On a bien $I = \sqrt{2\pi}$, et $f_X(s)$ est une densité de probabilité pour toutes valeurs de $m$ et de $\sigma^2$.

On dit d'une variable aléatoire $X$ qui a cette densité qu'elle suit une **loi gaussienne** ou **normale** $\mathcal{N}(m,\sigma^2)$. Les paramètres $m$ et $\sigma^2$ sont reliés à la moyenne et la variance d'une telle loi :

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} s\,\mathrm{e}^{-\frac{(s-m)^2}{2\sigma^2}}\,ds = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}(s-m)\,\mathrm{e}^{-\frac{(s-m)^2}{2\sigma^2}}\,ds + \frac{m}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}\mathrm{e}^{-\frac{(s-m)^2}{2\sigma^2}}\,ds\\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} u\,\mathrm{e}^{-\frac{u^2}{2\sigma^2}}\,du\right) + m = 0 + m = m,\end{aligned}$$

l'intégrale sur la seconde ligne étant nulle car la fonction intégrée est impaire. Pour la variance, on a

$$\begin{aligned}\mathbb{E}[(X-m)^2] &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}(s-m)^2\,\mathrm{e}^{-\frac{(s-m)^2}{2\sigma^2}}\,ds =_{\left(u=\frac{s-m}{\sigma}\right)} \frac{\sigma^2}{\sqrt{2\pi}}\int_{\mathbb{R}} u^2\,\mathrm{e}^{-\frac{u^2}{2}}\,du\\ &= \frac{\sigma^2}{\sqrt{2\pi}}\left(\left[-u\,\mathrm{e}^{-\frac{u^2}{2}}\right]_{-\infty}^{\infty} + \int_{\mathbb{R}}\mathrm{e}^{-\frac{u^2}{2}}\,du\right) = \frac{\sigma^2}{\sqrt{2\pi}}\left(0+\sqrt{2\pi}\right) = \sigma^2.\end{aligned}$$

Il n'y a pas de forme simple pour la fonction de répartition d'une loi gaussienne. Notons tout de même que si $X \sim \mathcal{N}(m,\sigma^2)$ et $a \in \mathbb{R}_+^*$, $b \in \mathbb{R}$, alors

$$\begin{aligned}\mathbb{P}[X \leq t] &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{t}\mathrm{e}^{-\frac{(s-m)^2}{2\sigma^2}}\,ds\\ &=_{\left(s=\frac{r-b}{a}\right)} \frac{1}{\sqrt{2\pi a^2\sigma^2}}\int_{-\infty}^{at+b}\mathrm{e}^{-\frac{(r-m-b)^2}{2a^2\sigma^2}}\,dr = \mathbb{P}[aX+b \leq at+b],\end{aligned}$$

donc $aX+b$ suit une loi $\mathcal{N}(m+b, a^2\sigma^2)$. En particulier, toutes les variables gaussiennes sont obtenues par transformation affine d'une variable aléatoire gaussienne **standard** $X \sim \mathcal{N}(0,1)$, puisque si $X$ est une variable gaussienne avec moyenne 0 et variance 1, alors $\sigma X + m \sim \mathcal{N}(m,\sigma^2)$.

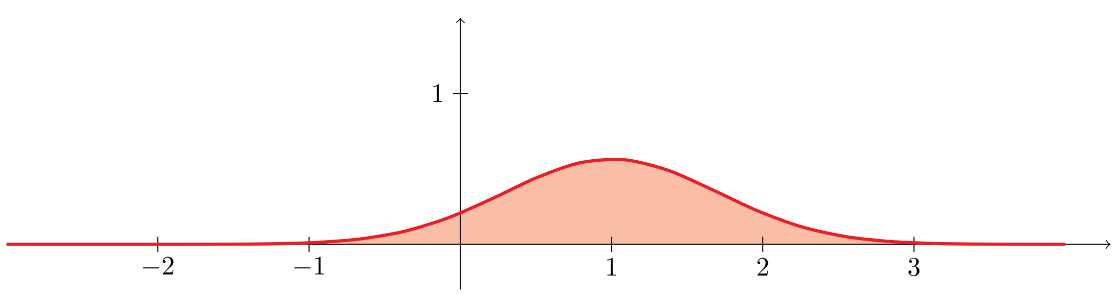


FIGURE 4. Densité d'une variable aléatoire gaussienne, ici avec $m = 1$ et $\sigma^2 = \frac{1}{2}$.

On a résumé dans le tableau suivant les propriétés des lois classiques continues :

| loi | intervalle | densité | espérance | variance |
|---|---|---|---|---|
| $\mathcal{U}([a,b])$ | $[a,b]$ | $\frac{1}{b-a}\, 1_{s\in[a,b]}$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |
| $\mathcal{E}(\lambda)$ | $\mathbb{R}_+$ | $\lambda\, \mathrm{e}^{-\lambda s}\, 1_{s\in\mathbb{R}_+}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| $\mathcal{N}(m,\sigma^2)$ | $\mathbb{R}$ | $\frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{(s-m)^2}{2\sigma^2}}$ | $m$ | $\sigma^2$ |

Dans ce tableau, la fonction $1_{s\in A}$ vaut 1 si $s \in A$, et 0 sinon.

**2.4. Transformations de variables aléatoires continues.** Pour conclure notre présentation des variables aléatoires continues, nous expliquons comment calculer la fonction de répartition et la densité d'une variable aléatoire $Y$ qui est obtenue par **transformation** d'une autre variable aléatoire continue $X$ de densité connue.

2.4.1. *Fonction d'une variable aléatoire.* Soit $X$ une variable aléatoire continue de densité $f_X$ et de fonction de répartition $F_X$ ; et $h : \mathbb{R} \to \mathbb{R}$ une fonction. On note $Y = h(X)$ ; c'est également une variable aléatoire continue, et le plus souvent, on peut exprimer la fonction de répartition $F_Y$ à partir de $F_X$. Plutôt que de donner une méthode générale, considérons les exemples suivants :

Exemple 2.25. Supposons $X \sim \mathcal{N}(0,1)$, et considérons $Y = X^2$. La fonction de répartition de $Y$ vaut 0 sur $\mathbb{R}_-$, car $Y$ est une variable positive. D'autre part, pour $t \geq 0$,

$$F_Y(t) = \mathbb{P}[Y \leq t] = \mathbb{P}[X^2 \leq t] = \mathbb{P}\Big[-\sqrt{t} \leq X \leq \sqrt{t}\Big] = F_X(\sqrt{t}) - F_X(-\sqrt{t}).$$

On en déduit la valeur de la densité de $Y$, en dérivant par rapport à $t$ :

$$f_Y(s) = \begin{cases} 0 & \text{si } s < 0, \\ \frac{1}{2\sqrt{s}}\big(f_X(\sqrt{s}) + f_X(-\sqrt{s})\big) = \frac{1}{\sqrt{2\pi s}}\,\mathrm{e}^{-\frac{s}{2}} & \text{si } s \geq 0. \end{cases}$$

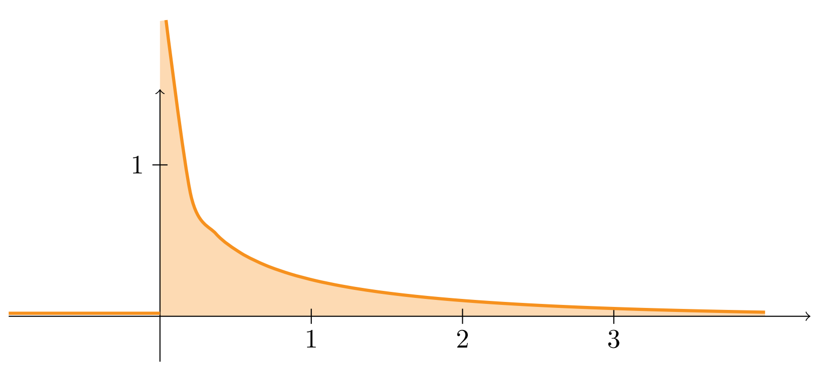


Figure 5. Densité du carré d'une variable aléatoire gaussienne standard.

Exemple 2.26. Soit $X \sim \mathcal{E}(1)$ une variable aléatoire de densité exponentielle $f_X(s) = \exp(-s)$ si $s \geq 0$, et 0 sinon. On pose $Y = \frac{1}{X}$. Rappelons que la fonction de répartition de $X$ est $F_X(t) = 1 - \mathrm{e}^{-t}$ si $t \geq 0$, et 0 sinon. La fonction de répartition de $Y$ peut être calculée comme suit : si $t \geq 0$, alors

$$F_Y(t) = \mathbb{P}\left[\frac{1}{X} \leq t\right] = \mathbb{P}\left[\frac{1}{t} \leq X\right] = 1 - \mathbb{P}\left[X \leq \frac{1}{t}\right] = \mathrm{e}^{-\frac{1}{t}}.$$

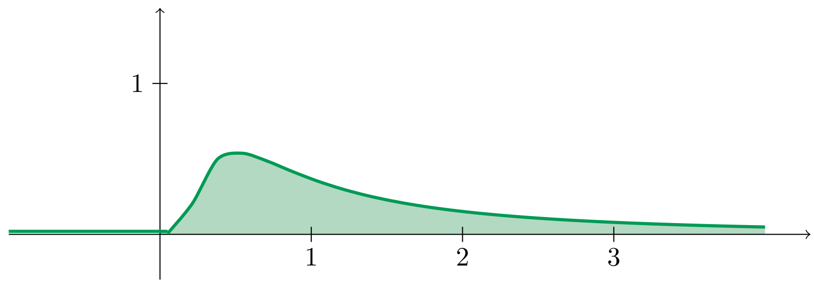


Figure 6. Densité de l'inverse d'une variable aléatoire exponentielle.

On en déduit, en dérivant par rapport à $t$, la densité de $Y$ :

$$f_Y(s) = \begin{cases} \frac{1}{s^2}\,\mathrm{e}^{-\frac{1}{s}} & \text{si } s \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Il est important de retenir que dans ces transformations de variables aléatoires, on calcule d'abord la fonction de répartition de la nouvelle variable $Y$, puis sa densité par dérivation. En utilisant le formalisme du changement de variables, il serait possible de donner une formule générale pour calculer $f_Y$ à partir de $f_X$, mais il est plus simple de retenir la méthode précédente.

2.4.2. *Maximum ou minimum de deux variables aléatoires.* Une autre transformation importante de variables aléatoires continues est

$$Z = \max(X, Y),$$

où $X$ et $Y$ sont des variables aléatoires continues. On suppose dans ce qui suit $X$ et $Y$ indépendantes : dans le cas continu, ceci signifie que pour tous intervalles $[a, b]$ et $[c, d]$, on a

$$\mathbb{P}[X \in [a, b] \text{ et } Y \in [c, d]] = \mathbb{P}[X \in [a, b]]\,\mathbb{P}[Y \in [c, d]].$$

La notion d'indépendance de variables aléatoires, déjà étudiée pour des variables discrètes, sera approfondie dans les chapitres suivants. Dans le cas continu, elle implique :

$$\mathbb{P}[X \leq s \text{ et } Y \leq t] = \mathbb{P}[X \leq s]\,\mathbb{P}[Y \leq t] = F_X(s)\,F_X(t).$$

On peut alors calculer la fonction de répartition de $Z$. En effet, dire que le maximum $Z$ de deux nombres $X$ et $Y$ est plus petit que $t$ est équivalent à demander que $X \leq t$ et $Y \leq t$. Par conséquent :

$$F_Z(t) = \mathbb{P}[Z \leq t] = \mathbb{P}[X \leq t \text{ et } Y \leq t] = F_X(t)\,F_Y(t).$$

Par dérivation, on en déduit ensuite la densité de $Z$.

Exemple 2.27. Soit $X$ et $Y$ deux variables aléatoires indépendantes et de même loi uniforme $\mathcal{U}([0,1])$. La fonction de répartition de $Z = \max(X, Y)$ est

$$F_Z(t) = (F_X(t))^2 = \begin{cases} 0 & \text{si } t < 0, \\ t^2 & \text{si } 0 \leq t \leq 1, \\ 1 & \text{si } t > 1. \end{cases}$$

La densité de $Z$ est donc

$$f_Z(s) = \begin{cases} 2s & \text{si } 0 \leq s \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

De même, si $X$ et $Y$ sont deux variables aléatoires continues indépendantes, alors

$$W = \min(X, Y)$$

est aussi une variable à densité, dont la fonction de répartition peut être calculée comme suit :

$$\begin{aligned} F_W(t) = \mathbb{P}[W \leq t] = 1 - \mathbb{P}[W \geq t] &= 1 - \mathbb{P}[X \geq t \text{ et } Y \geq t] = 1 - \mathbb{P}[X \geq t]\, \mathbb{P}[Y \geq t] \\ &= 1 - (1 - F_X(t))(1 - F_X(t)) = F_X(t) + F_Y(t) - F_X(t)\, F_Y(t). \end{aligned}$$

De nouveau, la densité de $W$ s'en déduit par dérivation. Il est important de retenir que de nouveau, les transformations de variables aléatoires $\max(\cdot)$ et $\min(\cdot)$ se lisent aisément sur les fonctions de répartition, mais pas directement sur les fonctions densité.

Exemple 2.28. Si $X$ et $Y$ sont deux variables aléatoires indépendantes et de même loi uniforme $\mathcal{U}([0,1])$, alors la fonction de répartition de $W = \min(X, Y)$ est

$$F_W(t) = F_X(t) + F_Y(t) - F_X(t)\, F_Y(t) = \begin{cases} 0 & \text{si } t < 0, \\ 2t - t^2 & \text{si } 0 \leq t \leq 1, \\ 1 & \text{si } t > 1. \end{cases}$$

La densité de $W$ est donc

$$f_W(s) = \begin{cases} 2(1-s) & \text{si } 0 \leq s \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

## 3. Approximations

Si les variables aléatoires continues sont de manipulation aisée, puisque l'on sait calculer les intégrales qui y sont reliées, ce n'est pas forcément le cas des variables discrètes, particulièrement les variables de lois binomiales et hypergéométriques. La théorie de l'**approximation** permet de remplacer dans les calculs ces variables par des variables aléatoires plus simples, sous certaines conditions. Dans ce qui suit, on détaille ces conditions et leurs applications.

**3.1. Approximation binomiale de la loi hypergéométrique.** On considère $X \sim \mathcal{H}(N, M, n)$, variable aléatoire discrète suivant une loi hypergéométrique. Rappelons que cette variable représente le nombre d'individus de type 1 dans un tirage sans remise de $n$ individus issus d'une population avec $M$ individus de type 1, et $N - M$ individus de type 2. Si ces deux nombres $M$ et $N - M$ sont grands devant $n$, alors il est raisonnable de penser que faire un tirage *sans remise* doit donner un résultat pour les probabilités $\mathbb{P}[X = k]$ assez semblable au tirage *avec remise*, car après chaque individu tiré, on modifie assez peu les proportions $\frac{M}{N}$ et $\frac{N-M}{N}$. Ainsi :

THÉORÈME 2.29. *Supposons $M$ et $N - M$ grands devant $n$, par exemple tous deux plus grands que $10 \times n$. Alors, si $X \sim \mathcal{H}(N, M, n)$,*

$$\mathbb{P}[X = k] \simeq \binom{n}{k} p^k (1-p)^{n-k},$$

*où $p = \frac{M}{N}$. Autrement dit, la loi de $X$ est approchée par une loi binomiale $\mathcal{B}(n, \frac{M}{N})$.*

En effet, dans la formule de la loi hypergéométrique,

$$\begin{aligned}
\mathbb{P}[X = k] &= \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \\
&= \frac{\frac{M(M-1)\cdots(M-k+1)}{k(k-1)\cdots 1}\,\frac{(N-M)(N-M-1)\cdots(N-M-(n-k)+1)}{(n-k)(n-k-1)\cdots 1}}{\frac{N(N-1)\cdots(N-n+1)}{n(n-1)\cdots 1}} \\
&= \binom{n}{k}\left(\frac{M}{N} \times \cdots \times \frac{M-k+1}{N-k+1}\right)\left(\frac{N-M}{N-k} \times \cdots \times \frac{N-M-(n-k)+1}{N-n+1}\right).
\end{aligned}$$

Comme $M \gg k$, le premier produit de la dernière ligne est équivalent à $\left(\frac{M}{N}\right)^k = p^k$. Comme $N - M \gg n - k$, le second produit de la la dernière ligne est équivalent à $\left(\frac{N-M}{N}\right)^{n-k} = (1-p)^{n-k}$. Ainsi, dans de nombreux cas, on peut remplacer une loi hypergéométrique par une loi binomiale.

**3.2. Approximation poissonienne de la loi binomiale.** Dans l'asymptotique $n \to \infty$, une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$ peut à son tour être approchée par une variable d'une loi plus simple. Le choix de cette loi dépend en revanche du comportement de $p$ en fonction de $n$. Fixons un paramètre $\lambda > 0$, et considérons une variable $X$ suivant une loi binomiale $\mathcal{B}(n, p = \frac{\lambda}{n})$. On peut alors faire les approximations suivantes dans le calcul de $\mathbb{P}[X = k]$, lorsque $n$ est très grand :

$$\begin{aligned}
\mathbb{P}[X = k] &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{\lambda^k}{k!}\,\frac{n(n-1)\cdots(n-k+1)}{n^k}\left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&\simeq \frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^n \simeq \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}.
\end{aligned}$$

Ainsi :

THÉORÈME 2.30. *Supposons $n$ grand (disons plus grand que 100) et $p = \frac{\lambda}{n}$, avec $\lambda$ pas trop grand (disons entre 0 et 5). Alors, si $X \sim \mathcal{B}(n,p)$,*

$$\mathbb{P}[X = k] \simeq \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!},$$

*où $\lambda = np$. La loi de $X$ est donc approchée par une loi de Poisson $\mathcal{P}(\lambda)$.*

EXEMPLE 2.31. On considère un conducteur qui, chaque jour, a une infime chance d'avoir un accident de voiture, par exemple une chance sur mille. Pendant une année, le nombre d'accidents qu'il aura suit une loi binomiale $\mathcal{B}(n = 365, p = \frac{1}{1000})$. On a dans ce contexte $n$ grand et $np = \frac{365}{1000}$ entre 0 et 5, donc on peut légitimement utiliser l'approximation poissonienne de la loi binomiale. Ainsi, le nombre d'accidents de voiture pendant une année est approché par une loi de Poisson $\mathcal{P}(\lambda = \frac{365}{1000})$.

**3.3. Approximation gaussienne de la loi binomiale.** L'approximation poissonienne est valable pour des expériences de Bernoulli nombreuses mais de probabilité $p = \frac{\lambda}{n}$ très faible ; à ce titre, on dit que c'est la loi des événements rares. Lorsqu'on considère des expériences de Bernoulli nombreuses mais dont les probabilités ne tendent pas vers 0 avec $n$, un autre schéma d'approximation doit être employé. Soit $X \sim \mathcal{B}(n,p)$, et

$$Y = \frac{X - np}{\sqrt{np(1-p)}},$$

qui est une variable aléatoire discrète mais à valeurs dans (une sous-partie dénombrable de) $\mathbb{R}$. Cette renormalisation est choisie de sorte que

$$\mathbb{E}[Y] = \frac{\mathbb{E}[X] - np}{\sqrt{np(1-p)}} = 0;$$
$$\mathrm{var}(Y) = \frac{\mathrm{var}(X)}{np(1-p)} = 1.$$

THÉORÈME 2.32. *Si $n$ est grand, disons plus grand que 100, et si $p$ n'est pas trop proche de 0 et de 1 (disons entre $\frac{1}{10}$ et $\frac{9}{10}$), alors pour tous $a < b$,*

$$\mathbb{P}[a \leq Y \leq b] \simeq \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-\frac{s^2}{2}}\, ds.$$

*Autrement dit, la loi de $Y$ est approchée par une gaussienne standard $\mathcal{N}(0,1)$.*

EXEMPLE 2.33. Si deux joueurs de même niveau s'affrontent lors de 400 parties de poker, alors le nombre $X$ de parties remportées par le premier joueur suit une loi binomiale $\mathcal{B}(400, \frac{1}{2})$. L'approximation gaussienne est valable, et

$$Y = \frac{X - 200}{10}$$

peut être approchée par une loi gaussienne $\mathcal{N}(0,1)$. Ainsi, la probabilité pour que le premier joueur remporte au moins 20 parties de plus que le second est

$$\mathbb{P}[X \geq 210] = \mathbb{P}[X - 200 \geq 10] = \mathbb{P}[Y \geq 1] \simeq \frac{1}{\sqrt{2\pi}} \int_1^{\infty} \mathrm{e}^{-\frac{s^2}{2}}\, ds \simeq 0.159.$$

Remarque. L'approximation gaussienne fournit une justification à l'introduction et à l'étude des lois normales $\mathcal{N}(m,\sigma^2)$ : elles apparaissent comme approximations de lois de sommes de variables aléatoires indépendantes.

---

Exercice. Soit $r$ un entier plus grand que 1, et $p \in (0,1)$. On considère une suite d'expériences de Bernoulli indépendantes de même paramètre de succès $p$, et on note $X$ le nombre aléatoire de succès avant le $r$-ième échec. Par exemple, si $r = 3$ et la suite d'expériences de Bernoulli a pour résultat

$$(1,0,1,1,1,0,1,1,0,\dots)$$

alors $X = 6$, puisqu'il y a 6 succès avant le troisième échec. Montrer que la loi de $X$ est

$$\mathbb{P}[X = k] = \binom{r+k-1}{r-1} p^k\,(1-p)^r$$

pour tout $k \geq 0$ (avec $\mathbb{P}[X = k] = 0$ si $k < 0$). Calculer en fonction de $r$ l'espérance et la variance de $X$.

Exercice. Soit $\lambda > 0$, et $f(s)$ la fonction

$$f_\lambda(s) = \begin{cases} C_\lambda\, \mathrm{e}^{-s} s^{\lambda-1} & \text{si } s > 0, \\ 0 & \text{si } s \leq 0 \end{cases},$$

où $C_\lambda$ est l'unique constante telle que $f_\lambda(s)$ soit une densité de probabilité. Si $\lambda = k$ est un entier, calculer $C_k$ (on pourra montrer que $C_\lambda = \frac{C_{\lambda-1}}{\lambda-1}$ pour tout $\lambda > 1$, et calculer d'autre part $C_1$). Calculer en fonction de $\lambda$ l'espérance et la variance d'une variable aléatoire continue de densité $f_\lambda(s)$.

Chapitre 3

# Sommes de variables aléatoires

**Résumé.** Si $X$ et $Y$ sont deux variables aléatoires indépendantes (discrètes ou continues), alors il existe une formule qui permet de calculer la loi de la somme $Z = X + Y$ de ces deux variables : la formule du *produit de convolution*. On donne cette formule dans ce chapitre, en traitant séparément les cas discrets et continus. À titre d'exemple, on applique cette formule à plusieurs lois classiques qui ont une propriété de *stabilité* par somme.

## 1. Convolution de deux lois discrètes

Soit $X$ et $Y$ deux variables aléatoires discrètes, à valeurs dans $\mathbb{Z}$, qu'on suppose indépendantes :

$$\forall (k,l) \in \mathbb{Z}^2,\ \ \mathbb{P}[X = k \text{ et } Y = l] = \mathbb{P}[X = k]\,\mathbb{P}[Y = l].$$

On note $Z = X + Y$, et on s'intéresse à la loi de $Z$, qui est elle aussi une variable aléatoire discrète à valeurs dans $\mathbb{Z}$. Cette loi est donnée par le calcul suivant :

$$\begin{aligned}\mathbb{P}[Z = k] &= \sum_{l \in \mathbb{Z}} \mathbb{P}[X = l \text{ et } Z = k] \\ &= \sum_{l \in \mathbb{Z}} \mathbb{P}[X = l \text{ et } Z - X = k - l] \\ &= \sum_{l \in \mathbb{Z}} \mathbb{P}[X = l \text{ et } Y = k - l] \\ &= \sum_{l \in \mathbb{Z}} \mathbb{P}[X = l]\,\mathbb{P}[Y = k - l],\end{aligned}$$

la dernière ligne utilisant l'indépendance de $X$ et $Y$. Ainsi :

THÉORÈME 3.1. *La loi de $Z = X + Y$ avec $X$ et $Y$ indépendantes est donnée par la formule dite du* ***produit de convolution***

$$\mathbb{P}_Z[k] = \sum_{l \in \mathbb{Z}} \mathbb{P}_X[l]\,\mathbb{P}_Y[k - l].$$

Un cas particulier de cette formule est celui où $X$ et $Y$ sont deux variables aléatoires positives ou nulles. Dans le cas où $X \geq 0$ et $Y \geq 0$, dans la formule du produit de convolution, tous les termes avec $l < 0$ sont nuls, car $\mathbb{P}[X = l] = 0$ si $l < 0$. De même, tous les termes avec $l > k$ sont nuls, car $\mathbb{P}[Y = k - l] = 0$ si $k - l < 0$, donc si $k < l$. Ainsi :

Proposition 3.2. *Si $X$ et $Y$ sont indépendantes et positives, alors la formule du produit de convolution s'écrit*

$$\mathbb{P}_Z[k] = \sum_{l=0}^{k} \mathbb{P}_X[l]\, \mathbb{P}_Y[k-l]$$

*pour $k \geq 0$ (pour $k < 0$, $\mathbb{P}_Z[k] = 0$ car $X \geq 0$ et $Y \geq 0$ implique $Z \geq 0$).*

Exemple 3.3. Soit $X$ et $Y$ deux variables indépendantes suivant des lois de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. La loi de $Z = X + Y$ est calculée comme suit : pour $k \geq 0$,

$$\begin{aligned} \mathbb{P}_Z[k] &= \sum_{l=0}^{k} \mathrm{e}^{-\lambda}\, \frac{\lambda^l}{l!}\; \mathrm{e}^{-\mu}\, \frac{\mu^{k-l}}{(k-l)!} \\ &= \frac{\mathrm{e}^{-\lambda-\mu}}{k!} \sum_{l=0}^{k} \frac{k!}{l!\,(k-l)!}\, \lambda^l\, \mu^{k-l} \\ &= \mathrm{e}^{-\lambda-\mu}\, \frac{(\lambda+\mu)^k}{k!} \end{aligned}$$

en utilisant à la dernière ligne la formule du binôme de Newton. On obtient pour $Z$ une loi de Poisson de paramètre $\lambda + \mu$.

Exemple 3.4. Soit $X$ et $Y$ deux variables indépendantes suivant des lois binomiales $\mathcal{B}(m, p)$ et $\mathcal{B}(n, p)$, avec le même paramètre $p$ pour les deux variables. La loi de $Z = X+Y$ porte sur $[\![0, m+n]\!]$, et est donnée par la formule

$$\begin{aligned} \mathbb{P}_Z[k] &= \sum_{l=0}^{k} \binom{m}{l} p^l (1-p)^{m-l} \binom{n}{k-l} p^{k-l} (1-p)^{n-k+l} \\ &= p^k (1-p)^{m+n-k} \sum_{l=0}^{k} \binom{m}{l} \binom{n}{k-l} \\ &= \binom{m+n}{k} p^k (1-p)^{m+n-k} \end{aligned}$$

en utilisant à la dernière ligne la formule de Vandermonde, introduite lors de l'étude des lois hypergéométriques. On obtient pour $Z$ une loi binomiale de paramètres $(m+n, p)$.

En dehors de la loi de la somme $Z$ de deux variables aléatoires discrètes, on peut calculer d'autres quantités telles que l'espérance et la variance de $Z$. Elles sont données par la proposition suivante :

Proposition 3.5. *Soit $Z = X+Y$ somme de deux variables aléatoires discrètes. Dans tous les cas,*

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

*Si de plus $X$ et $Y$ sont indépendantes, alors*

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y).$$

Ces règles ont déjà été énoncées dans le paragraphe 1.2.

## 2. Convolution de deux lois continues

Si $X$ et $Y$ sont des variables aléatoires continues, leurs densités $f_X(s)$ et $f_Y(t)$ jouent le rôle des fonctions $\mathbb{P}_X[k]$ et $\mathbb{P}_Y[l]$ pour des variables discrètes. Supposons $X$ et $Y$ indépendantes, au sens donné au paragraphe 2.4.2. Alors,

Théorème 3.6. *La variable $Z$ a une densité sur $\mathbb{R}$ donnée par la formule*

$$f_Z(s) = \int_{t\in\mathbb{R}} f_X(t)\, f_Y(s-t)\, dt.$$

De nouveau, un cas particulier de cette formule est celui où $X \geq 0$ et $Y \geq 0$, ce qui revient à demander que $f_X(s) = 0$ pour tout $s < 0$, et $f_Y(s) = 0$ pour tout $s < 0$. Dans ce cas, dans l'intégrale ci-dessus, la partie de l'intégrale correspondant à l'intervalle $(-\infty, 0]$ est nulle, car $f_X(t) = 0$ pour tout $t < 0$. De même, la partie de l'intégrale correspondant à l'intervalle $[s, +\infty)$ est nulle, car $f_Y(s-t) = 0$ lorsque $s - t < 0$, donc lorsque $t > s$. Ainsi :

Proposition 3.7. *Si $X$ et $Y$ sont continues, indépendantes et positives, alors la formule du produit de convolution s'écrit*

$$f_Z(s) = \int_{t=0}^{s} f_X(t)\, f_Y(s-t)\, dt$$

*pour $s \geq 0$ (pour $s < 0$, $f_Z(s) = 0$ car $X \geq 0$ et $Y \geq 0$ implique $Z \geq 0$).*

Exemple 3.8. Soit $X$ et $Y$ deux variables aléatoires indépendantes, de même loi exponentielle $\mathcal{E}(1)$. La loi de la somme $Z = X + Y$ a une densité qui s'écrit, pour $s \geq 0$,

$$f_Z(s) = \int_{t=0}^{s} \mathrm{e}^{-t}\, \mathrm{e}^{-(s-t)}\, dt = \mathrm{e}^{-s} \int_{t=0}^{s} 1\, dt = s\, \mathrm{e}^{-s}.$$

Pour $s \leq 0$, $f_Z(s) = 0$, donc,

$$f_Z(s) = \begin{cases} s\,\mathrm{e}^{-s} & \text{si } s \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

On peut ensuite, en intégrant cette densité, calculer la fonction de répartition de $Z$ :

$$F_Z(t) = \begin{cases} 1 - (1+t)\,\mathrm{e}^{-t} & \text{si } t \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Exemple 3.9. Soit $X$ et $Y$ deux variables aléatoires indépendantes, de lois normales $\mathcal{N}(m_1,(\sigma_1)^2)$ et $\mathcal{N}(m_2,(\sigma_2)^2)$. La loi de $Z = X+Y$ a sa densité donnée par le calcul suivant :

$$\begin{aligned}
f_Z(s) &= \frac{1}{\sqrt{2\pi(\sigma_1)^2}\,\sqrt{2\pi(\sigma_2)^2}} \int_{t=-\infty}^{\infty} \mathrm{e}^{-\frac{(t-m_1)^2}{2(\sigma_1)^2}}\,\mathrm{e}^{-\frac{(s-t-m_2)^2}{2(\sigma_2)^2}}\,dt \\
&=_{(u=t-m_1)} \frac{1}{2\pi\sigma_1\sigma_2} \int_{u=-\infty}^{\infty} \mathrm{e}^{-\frac{u^2}{2(\sigma_1)^2}}\,\mathrm{e}^{-\frac{(s-u-m_1-m_2)^2}{2(\sigma_2)^2}}\,du \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\,\mathrm{e}^{-\frac{(s-m_1-m_2)^2}{2(\sigma_2)^2}} \int_{u=-\infty}^{\infty} \mathrm{e}^{-\frac{u^2}{2(\sigma_1)^2}}\,\mathrm{e}^{-\frac{u^2-2(s-m_1-m_2)u}{2(\sigma_2)^2}}\,du \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\,\mathrm{e}^{-\frac{(s-m_1-m_2)^2}{2(\sigma_2)^2}+\frac{(s-m_1-m_2)^2}{2(\sigma_2)^2\left(1+\frac{(\sigma_2)^2}{(\sigma_1)^2}\right)}} \int_{u=-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\left(u\sqrt{\frac{1}{(\sigma_1)^2}+\frac{1}{(\sigma_2)^2}}-\frac{s-m_1-m_2}{\sigma_2\sqrt{1+\frac{(\sigma_2)^2}{(\sigma_1)^2}}}\right)^2}\,du \\
&=_{\left(v=u\sqrt{\frac{1}{(\sigma_1)^2}+\frac{1}{(\sigma_2)^2}}-\frac{s-m_1-m_2}{\sigma_2\sqrt{1+\frac{(\sigma_2)^2}{(\sigma_1)^2}}}\right)} \frac{1}{2\pi\sqrt{(\sigma_1)^2+(\sigma_2)^2}}\,\mathrm{e}^{-\frac{(s-m_1-m_2)^2}{2((\sigma_1)^2+(\sigma_2)^2)}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{v^2}{2}}\,dv \\
&= \frac{1}{\sqrt{2\pi((\sigma_1)^2+(\sigma_2)^2)}}\,\mathrm{e}^{-\frac{(s-m_1-m_2)^2}{2((\sigma_1)^2+(\sigma_2)^2)}}.
\end{aligned}$$

On obtient ainsi pour $Z$ une loi gaussienne de moyenne $m = m_1+m_2$ et de variance $\sigma^2 = (\sigma_1)^2+(\sigma_2)^2$.

Exemple 3.10. Considérons de nouveau deux variables aléatoires $X$ et $Y$ indépendantes, de même loi exponentielle $\mathcal{E}(1)$. On s'intéresse à la loi de $Z = X-Y$. La densité de $-Y$ est

$$f_{-Y}(s) = \begin{cases} \mathrm{e}^s & \text{si } s \le 0 \\ 0 & \text{si } s > 0 \end{cases} \;=\; f_Y(-s).$$

D'autre part, la formule du produit de convolution donne

$$f_Z(s) = \int_{t=-\infty}^{\infty} f_X(t)\,f_{-Y}(s-t)\,dt = \int_{t=-\infty}^{\infty} f_X(t)\,f_Y(t-s)\,dt.$$

Pour que $f_X(t)\,f_Y(t-s)$ soit non nul, il faut à la fois que $t$ soit positif, et que $t-s$ soit positif, c'est-à-dire que $t \ge s$. Ainsi :

$$f_Z(s) = \int_{\max(0,s)}^{\infty} \mathrm{e}^{-t}\,\mathrm{e}^{s-t}\,dt = \mathrm{e}^s \int_{\max(0,s)}^{\infty} \mathrm{e}^{-2t}\,dt = \begin{cases} \frac{\mathrm{e}^s}{2} & \text{si } s \le 0 \\ \frac{\mathrm{e}^{-s}}{2} & \text{si } s > 0 \end{cases} \;=\; \frac{\mathrm{e}^{-|s|}}{2}.$$

On obtient ainsi une variable aléatoire dont la densité décroit exponentiellement vite, de façon symétrique sur les nombres positifs et les nombres négatifs.

Pour conclure cette présentation du calcul de la loi de la somme de deux variables à densité, notons que les résultats de la proposition 3.5 restent vrais dans le cadre continu :

l'espérance d'une somme de variables continues est toujours la somme des espérances, et la variance d'une somme de variables aléatoires continues indépendantes est la somme des variances.

---

Exercice. Calculer la loi de la somme de deux variables aléatoires indépendantes de lois géométriques $\mathcal{G}(p)$ et $\mathcal{G}(p')$ (on distinguera les cas $p = p'$ et $p \neq p'$).

Exercice. Calculer la loi de la somme de deux variables aléatoires indépendantes de lois à densité $\frac{1}{\pi\,\lambda(1+(\frac{s}{\lambda})^2)}$ et $\frac{1}{\pi\,\mu(1+(\frac{s}{\mu})^2)}$, où $\lambda$ et $\mu$ sont deux réels strictement positifs.

Chapitre 4

# Couples de variables aléatoires

**Résumé.** On étudie dans ce chapitre les *couples* $(X, Y)$ de variables aléatoires, discrètes ou continues, et pas nécessairement indépendantes. Dans ce contexte, la loi de $X$ et la loi de $Y$ ne déterminent pas entièrement la loi du couple, et on doit introduire de nouveaux outils pour capturer les interactions entre les deux variables : *loi jointe*, *lois marginales*, *lois conditionnelles*. On explique également comment calculer la moyenne de fonctions $h(X, Y)$ de couples de variables aléatoires, et on introduit la notion de *covariance*.

## 1. Couples de variables et loi jointe

Dans le chapitre 2, on a présenté les variables aléatoires discrètes comme des fonctions $X : \Omega \to \mathbb{Z}$ issues d'un espace de probabilité $(\Omega, \mathbb{P})$, et les variables aléatoires continues comme des fonctions $X : \Omega \to \mathbb{R}$. Ces fonctions induisent des mesures de probabilité sur $\mathbb{Z}$ ou sur $\mathbb{R}$, les lois (discrètes ou continues), et l'on a expliqué jusqu'ici comment utiliser ces lois pour étudier les nombres aléatoires $X$. Avec la même approche, on peut considérer des variables aléatoires multi-dimensionnelles $X : \Omega \to \mathbb{Z}^d$ ou $X : \Omega \to \mathbb{R}^d$, avec $d \geq 2$. Dans ce chapitre, on regardera uniquement le cas $d = 2$, qui contient déjà toutes les difficultés inhérentes au passage à des variables aléatoires en plusieurs dimensions.

On commence par le cas discret :

Définition 4.1. *La* ***loi jointe*** *d'un couple de variables aléatoires $(X, Y)$ à valeurs dans $\mathbb{Z}^2$ est la fonction*

$$\mathbb{P}_{(X,Y)} : (k, l) \in \mathbb{Z}^2 \mapsto \mathbb{P}[X = k \text{ et } Y = l].$$

Cette fonction permet de calculer, pour toute partie $A \subset \mathbb{Z}^2$, la probabilité

$$\mathbb{P}[(X, Y) \in A] = \sum_{(k,l) \in A} \mathbb{P}_{(X,Y)}[k, l].$$

La loi jointe est une fonction à valeurs dans $[0, 1]$, qui vérifie la condition de normalisation

$$\sum_{(k,l) \in \mathbb{Z}^2} \mathbb{P}_{(X,Y)}[k, l] = 1.$$

EXEMPLE 4.2. Soit $\mathbb{P}_{(X,Y)}$ la fonction sur $\mathbb{Z}^2$ dont les valeurs non nulles sont données par le tableau suivant :

| $k \backslash l$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | $\frac{1}{12}$ | $\frac{1}{6}$ |
| 2 | $\frac{1}{6}$ | $\frac{1}{4}$ | $\frac{1}{3}$ |

Comme $0+\frac{1}{12}+\frac{1}{6}+\frac{1}{6}+\frac{1}{4}+\frac{1}{3}=1$, $\mathbb{P}_{(X,Y)}$ est la loi jointe d'un couple de variables aléatoires discrètes $(X,Y)$, avec, par exemple, $\mathbb{P}[X=1 \text{ et } Y=2]=\frac{1}{12}$.

EXEMPLE 4.3. Soit $X$ une variable aléatoire qui suit une loi géométrique décalée, de paramètre $p$ :

$$\mathbb{P}[X=k]=(1-p)^k p \quad \text{pour tout } k \geq 0.$$

On considère alors $Y$, somme de $X$ variables aléatoires indépendantes de Bernoulli de paramètre $q$, ces variables étant indépendantes de $X$ :

$$Y=\sum_{j=1}^{X} \xi_j, \quad \text{avec les } \xi_j \text{ indépendantes entre elles et indépendantes de } X.$$

Sachant $X=k$, la probabilité pour que $Y=l$ avec $l \in [\![0,k]\!]$ est donnée par une loi binomiale :

$$\mathbb{P}[Y=l|X=k]=\binom{k}{l} q^l (1-q)^{k-l}.$$

Par conséquent, la loi jointe de $(X,Y)$ porte sur les couples d'entiers $(k,l)$ avec $0 \leq k \leq l$, et elle est donnée par la formule

$$\mathbb{P}[X=k \text{ et } Y=l]=\mathbb{P}[X=k]\,\mathbb{P}[Y=l|X=k]=\binom{k}{l}(1-p)^k p\, q^l (1-q)^{k-l}.$$

Dans le cas continu, les probabilités $\mathbb{P}_{(X,Y)}[k,l]$ sont remplacées par des densités $f_{(X,Y)}(s,t)$.

DÉFINITION 4.4. *On dit d'un couple de variables aléatoires réelles $(X,Y)$ qu'il a pour* ***densité jointe*** *$f_{(X,Y)}(s,t)$ si, pour tous intervalles $[a,b]$ et $[c,d]$ de $\mathbb{R}$,*

$$\mathbb{P}[X \in [a,b] \textit{ et } Y \in [c,d]]=\int_{s=a}^{b}\int_{t=c}^{d} f_{(X,Y)}(s,t)\,ds\,dt$$

Pour la plupart des parties de $\mathbb{R}^2$, la densité jointe permet alors de calculer la probabilité

$$\mathbb{P}[(X,Y) \in A]=\iint_{(s,t)\in A} f_{(X,Y)}(s,t)\,ds\,dt,$$

la formule de la définition correspondant au cas où $A$ est le rectangle $[a,b]\times[c,d]$. D'autre part, comme dans le cas discret, la densité jointe est une fonction positive qui vérifie la condition de normalisation

$$\iint_{\mathbb{R}^2} f_{(X,Y)}(s,t)\,ds\,dt=1.$$

Remarque. La formule qui exprime $\mathbb{P}[(X, Y) \in A]$ comme l'intégrale de la densité jointe sur $A$ est vraie pour toute partie $A$ qui peut être "approchée" par une réunion de rectangles, par exemple, l'intérieur d'une courbe continue fermée :

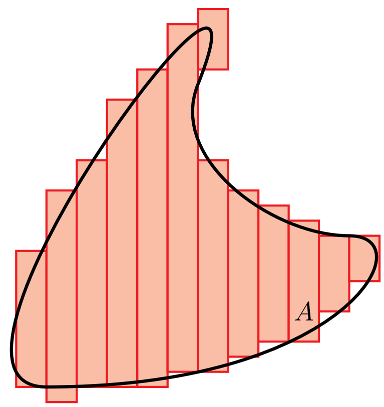


Figure 1. Partie $A \subset \mathbb{R}^2$ qui peut être approchée par une réunion de rectangles.

Nous ne préciserons pas plus quelles parties sont concernées par cette formule ; notons seulement qu'il est particulièrement difficile d'y trouver un contre-exemple.

Exemple 4.5. Soit $f(s, t)$ la fonction qui vaut 0 en dehors de $[0, 1] \times [0, 1]$, et qui est donnée par la formule

$$f(s, t) = \frac{2(s^2 + t^2)}{1 + st}$$

pour $(s, t) \in [0, 1]^2$.

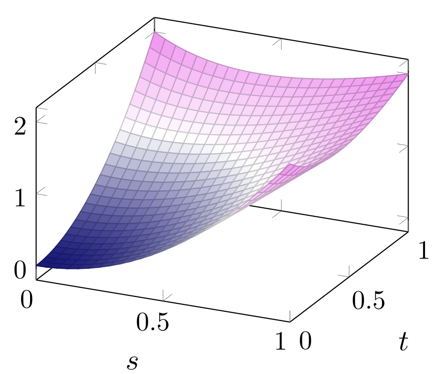


Figure 2. Une densité de probabilité pour un couple de variables $(X, Y)$ à valeurs dans $[0, 1]^2$.

C'est la densité de probabilité d'un couple $(X, Y)$ de variables aléatoires continues à valeurs dans $[0,1]^2$, car $\iint_{[0,1]^2} f(s,t)\,ds\,dt = 1$. En effet,

$$I = \int_{s=0}^{1}\int_{t=0}^{1} \frac{s^2+t^2}{1+st}\,ds\,dt = \int_{s=0}^{1}\left(\int_{t=0}^{1} \frac{st-1}{s^2} + \left(s^2+\frac{1}{s^2}\right)\frac{1}{1+st}\,dt\right)ds$$
$$= \int_{s=0}^{1}\left(\frac{1}{2s} - \frac{1}{s^2} + s\,\log(1+s) + \frac{\log(1+s)}{s^3}\right)ds.$$

Pour résoudre la singularité de la fonction en $s = 0$, on utilise la formule de Taylor avec reste intégral à l'ordre 3 pour $f(s) = \log(1+s)$ :

$$\log(1+s) = s - \frac{s^2}{2} + \frac{s^3}{2}\int_0^1 (1-u^2)\,f^{(3)}(us)\,du = s - \frac{s^2}{2} + s^3\int_0^1 \frac{(1-u)^2}{(1+us)^3}\,du.$$

On a donc bien :

$$I = \int_{s=0}^{1} s\,\log(1+s)\,ds + \int_{u=0}^{1}\int_{s=0}^{1}\frac{(1-u)^2}{(1+us)^3}\,du\,ds$$
$$= \left[\frac{s^2-1}{2}\,\log(1+s) + \frac{2s-s^2}{4}\right]_{s=0}^{1} + \frac{1}{2}\int_{u=0}^{1}\frac{(2+u)(1-u)^2}{(1+u)^2}\,du$$
$$= \frac{1}{4} + \frac{1}{2}\left[\frac{u^2}{2} - 2u - \frac{4}{1+u}\right]_{u=0}^{1} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Exemple 4.6. Soit $f(s,t) = \frac{1}{2\pi}\,\mathrm{e}^{-\frac{s^2+t^2}{2}}$. L'intégrale de $f(s,t)$ sur $\mathbb{R}^2$ est

$$\iint_{\mathbb{R}^2} \frac{1}{2\pi}\,\mathrm{e}^{-\frac{s^2+t^2}{2}}\,ds\,dt = \left(\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{e}^{-\frac{s^2}{2}}\,ds\right)\left(\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{e}^{-\frac{t^2}{2}}\,dt\right) = 1 \times 1 = 1,$$

car les deux intégrales sont celles des densités de variables gaussiennes de paramètres $m = 0$ et $\sigma^2 = 1$. Par conséquent, $f(s,t)$ est la densité jointe d'un couple de variables dans $\mathbb{R}^2$, et nous verrons plus loin que $f(s,t)$ est justement la densité d'un couple de variables gaussiennes indépendantes $X \sim \mathcal{N}(0,1)$ et $Y \sim \mathcal{N}(0,1)$.

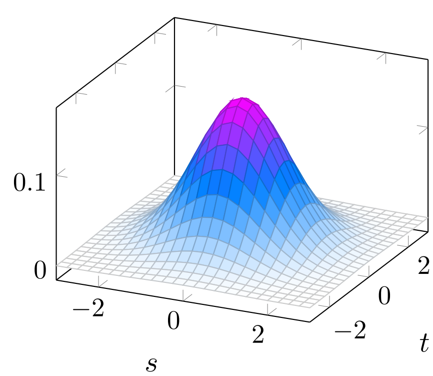


Figure 3. Densité de probabilité d'un couple de variables $(X, Y)$ avec $X \sim \mathcal{N}(0,1)$ et $Y = \mathcal{N}(0,1)$.

Soit $A = D_{(0,R)}$ le disque de rayon $R$ et de centre l'origine $(0,0)$. La probabilité pour que $(X,Y)$ appartienne à $A$ est

$$\begin{aligned}\mathbb{P}[(X,Y)\in A] &= \frac{1}{2\pi}\iint_A \mathrm{e}^{-\frac{x^2+y^2}{2}}\,dx\,dy\\ &= \frac{1}{2\pi}\int_{r=0}^{R}\int_{\theta=0}^{2\pi}\mathrm{e}^{-\frac{r^2}{2}}\,r\,dr\,d\theta\\ &= \left[-\mathrm{e}^{-\frac{r^2}{2}}\right]_0^R = 1-\mathrm{e}^{-\frac{R^2}{2}}\end{aligned}$$

en effectuant un changement de variables polaire à la seconde ligne.

Remarque. On pourrait sur le même modèle introduire la notion de loi jointe d'un $d$-uplet $(X_1, X_2, \ldots, X_d)$ de variables aléatoires discrètes ou continues, avec $d \geq 3$. Tout ce qui suit s'adapte aisément à ce cadre, en remplaçant les sommes doubles par des sommes multiples sur $d$ indices, et les intégrales doubles par des intégrales multiples $d$-dimensionnelles.

## 2. Lois marginales, lois conditionnelles

Si $(X,Y)$ est un couple de variables aléatoires discrètes (dans $\mathbb{Z}^2$) ou continues (dans $\mathbb{R}^2$), alors $X$ et $Y$ sont elles-mêmes des variables aléatoires, donc, admettent des lois, données par des fonctions $k \mapsto \mathbb{P}[X=k]$ et $l \mapsto \mathbb{P}[Y=l]$ dans le cas discret, et par des densités $s \mapsto f_X(s)$ et $t \mapsto f_Y(t)$ dans le cas continu.

Définition 4.7. *On dit que les lois $\mathbb{P}_X$ et $\mathbb{P}_Y$ (respectivement, les densités $f_X$ et $f_Y$) pour les deux variables aléatoires $X$ et $Y$ sont les **lois marginales** du couple $(X,Y)$.*

On peut calculer les lois marginales de $X$ et de $Y$ à partir de la loi jointe du couple $(X,Y)$. Dans le cas discret :

$$\mathbb{P}_X[k] = \mathbb{P}[X=k] = \sum_{l\in\mathbb{Z}} \mathbb{P}[X=k \text{ et } Y=l] = \sum_{l\in\mathbb{Z}} \mathbb{P}_{(X,Y)}[k,l].$$

Ainsi :

Théorème 4.8. *Soit $(X,Y)$ un couple de variables aléatoires discrètes, de loi jointe $\mathbb{P}_{(X,Y)}$. Les lois marginales sont données par les sommes*

$$\mathbb{P}_X[k] = \sum_{l\in\mathbb{Z}} \mathbb{P}_{(X,Y)}[k,l];$$
$$\mathbb{P}_Y[l] = \sum_{k\in\mathbb{Z}} \mathbb{P}_{(X,Y)}[k,l].$$

Par analogie, dans le cas continu :

Théorème 4.9. *Soit $(X,Y)$ un couple de variables aléatoires continues, de densité jointe $f_{(X,Y)}$. Les lois marginales de $X$ et de $Y$ ont des densités qui sont données par les intégrales*

$$f_X(s) = \int_{t\in\mathbb{R}} f_{(X,Y)}(s,t)\,dt;$$
$$f_Y(t) = \int_{s\in\mathbb{R}} f_{(X,Y)}(s,t)\,ds.$$

Ces formules donnent bien des densités de variables aléatoires, car

$$\int_{s\in\mathbb{R}} f_X(s)\,ds = \iint_{(s,t)\in\mathbb{R}^2} f_{(X,Y)}(s,t)\,ds\,dt = 1$$

et de même pour $f_Y$.

Exemple 4.10. Soit $(X,Y)$ le couple de variables aléatoires à valeurs dans $\{1,2\}\times\{1,2,3\}$ et de loi jointe donnée par le tableau de l'exemple 4.2. La loi marginale de $X$ est calculée en faisant une somme sur chaque ligne, et la loi marginale de $Y$ est calculée en faisant une somme sur chaque colonne :

| $k\backslash l$ | 1 | 2 | 3 | $\mathbb{P}_X$ |
|---|---|---|---|---|
| 1 | 0 | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{4}$ |
| 2 | $\frac{1}{6}$ | $\frac{1}{4}$ | $\frac{1}{3}$ | $\frac{3}{4}$ |
| $\mathbb{P}_Y$ | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{2}$ | 1 |

Exemple 4.11. Soit $(X,Y)$ le couple de variables aléatoires de l'exemple 4.3, de loi jointe

$$\mathbb{P}[X=k \text{ et } Y=l] = \binom{k}{l}(1-p)^k p\, q^l (1-q)^{k-l}$$

pour $k\geq 0$ et $l\in [\![0,k]\!]$. La loi marginale de $X$ est

$$\mathbb{P}_X[k] = \sum_{l=0}^{k}\binom{k}{l}(1-p)^k p\, q^l(1-q)^{k-l} = (1-p)^k p,$$

c'est-à-dire une loi géométrique de paramètre $p$, décalée de $-1$. La loi marginale de $Y$ est donnée par

$$\begin{aligned}\mathbb{P}_Y[l] &= \sum_{k=l}^{\infty}\binom{k}{l} p(1-p)^k\, q^l(1-q)^{k-l}\\ &= p(1-p)^l\, q^l \sum_{k=l}^{\infty}\binom{k}{l}[(1-p)(1-q)]^{k-l}\\ &= p(1-p)^l\, q^l \sum_{j=0}^{\infty}\binom{j+l}{l}[(1-p)(1-q)]^{j}\end{aligned}$$

pour tout $l \geq 0$. Remarquons que le développement de Taylor de $\frac{1}{(1-x)^{l+1}}$ en $x = 0$ est

$$\begin{aligned}\frac{1}{(1-x)^{l+1}} &= 1 + (l+1)\,x + \frac{(l+1)(l+2)}{2}\,x^2 + \frac{(l+1)(l+2)(l+3)}{6}\,x^3 + \cdots \\ &= \sum_{j=0}^{\infty} \binom{j+l}{l} x^j.\end{aligned}$$

Par conséquent,

$$\mathbb{P}_Y[l] = \frac{p(1-p)^l q^l}{(1-(1-p)(1-q))^{l+1}} = \left(1 - \frac{p}{p+q-pq}\right)^l \frac{p}{p+q-pq},$$

ce qui est de nouveau une loi géométrique décalée.

Exemple 4.12. Dans l'exemple 4.5, on a été amené à calculer la loi marginale de $X$ :

$$f_X(s) = \begin{cases} \int_0^1 \frac{2(s^2+t^2)}{1+st}\,dt = \frac{1}{s} - \frac{2}{s^2} + 2s\,\log(1+s) + \frac{2\,\log(1+s)}{s^3} & \text{si } s \in [0,1], \\ 0 & \text{sinon.} \end{cases}$$

Par symétrie, on a aussi

$$f_Y(t) = \begin{cases} \frac{1}{t} - \frac{2}{t^2} + 2t\,\log(1+t) + \frac{2\,\log(1+t)}{t^3} & \text{si } t \in [0,1], \\ 0 & \text{sinon.} \end{cases}$$

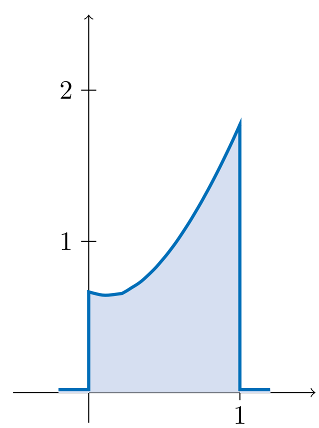


Figure 4. Loi marginale de $X$ lorsque $f_{(X,Y)}(s,t) = 1_{(s,t)\in[0,1]^2}\, \frac{2(s^2+t^2)}{1+st}$.

Exemple 4.13. Dans l'exemple 4.6, les variables $X$ et $Y$ ont toutes deux pour loi marginale une gaussienne standard $\mathcal{N}(0,1)$ :

$$f_X(s) = \frac{1}{2\pi} \int_{t=-\infty}^{\infty} \mathrm{e}^{-\frac{s^2+t^2}{2}}\,dt = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{s^2}{2}},$$

et symétriquement pour la variable $Y$.

Remarque. Attention, connaître les lois marginales de $X$ et de $Y$ ne suffit pas à déterminer la loi jointe du couple $(X, Y)$. Considérons par exemple le cas où $(X, Y)$ est un couple de variables continues, à valeurs dans $[0, 1]^2$, de densité jointe

$$f_{(X,Y)}(s, t) = g(s)\, g(t),$$

où

$$g(s) = \frac{1}{s} - \frac{2}{s^2} + 2s \log(1+s) + \frac{2\log(1+s)}{s^3}$$

si $s \in [0, 1]$. Cette loi jointe donne pour lois marginales $f_X(s) = g(s)$ et $f_Y(t) = g(t)$, c'est-à-dire les mêmes marginales que pour la loi jointe $f_{(X,Y)} = \frac{2(s^2+t^2)}{1+st}$ de l'exemple 4.5. Ainsi, deux lois jointes différentes peuvent donner exactement les mêmes lois marginales. Nous verrons plus loin qu'on peut reconstruire la loi jointe à partir des marginales seulement si $X$ et $Y$ sont des variables indépendantes.

Il existe une troisième notion de loi pour un couple de variables aléatoires : la notion de **loi conditionnelle**. Commençons par le cas discret. Soit $(X, Y)$ un couple de variables aléatoires à valeurs dans $\mathbb{Z}^2$, et $l \in \mathbb{Z}$ tel que $\mathbb{P}_Y[l] \neq 0$. On peut alors considérer les quantités

$$\mathbb{P}[X = k | Y = l] = \frac{\mathbb{P}[X = k \text{ et } Y = l]}{\mathbb{P}[Y = l]} = \frac{\mathbb{P}_{(X,Y)}[k, l]}{\mathbb{P}_Y[l]}.$$

La somme de ces quantités lorsque $k$ parcourt $\mathbb{Z}$ est égale à 1, car $\mathbb{P}_Y[l] = \sum_{k\in\mathbb{Z}} \mathbb{P}_{(X,Y)}[k, l]$. On obtient donc une nouvelle probabilité, la loi conditionnelle de $X$ sachant $Y = l$. Elle est notée

$$\mathbb{P}_{X|Y=l}[k] = \mathbb{P}[X = k|Y = l].$$

Exemple 4.14. Poursuivant l'exemple 4.2, la loi conditionnelle de $X$ sachant $Y = 3$ est

$$\mathbb{P}_{X|Y=3}[1] = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3} \qquad ; \qquad \mathbb{P}_{X|Y=3}[2] = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Ce n'est pas la même loi que la loi conditionnelle de $X$ sachant $Y = 2$ :

$$\mathbb{P}_{X|Y=2}[1] = \frac{\frac{1}{12}}{\frac{1}{3}} = \frac{1}{4} \qquad ; \qquad \mathbb{P}_{X|Y=2}[2] = \frac{\frac{1}{4}}{\frac{1}{3}} = \frac{3}{4}.$$

Ainsi, la loi conditionnelle de $X$ est modifiée en fonction des informations que l'on a sur la valeur de $Y$ ($Y = 3$ ou $Y = 2$). Les lois conditionnelles capturent donc de façon fine les interactions possibles entre deux variables aléatoires $X$ et $Y$.

Exemple 4.15. Dans l'exemple 4.3, la loi conditionnelle de $Y$ sachant $X = k$ est

$$\mathbb{P}_{Y|X=k}[l] = \binom{k}{l} q^l (1-q)^{k-l} \quad \text{pour } l \in [\![0, k]\!],$$

c'est-à-dire une loi binomiale $\mathcal{B}(k,q)$. D'autre part, la loi conditionnelle de $X$ sachant $Y=l$ est

$$\mathbb{P}_{X|Y=l}[k] = \frac{\binom{k}{l}\, p(1-p)^k\, q^l (1-q)^{k-l}}{\frac{p}{p+q-pq}\left(1-\frac{p}{p+q-pq}\right)^l} = \binom{k}{l}\,(1-(p+q-pq))^{k-l}\,(p+q-pq)^{l+1}$$

pour $k \geq l$.

Par analogie, pour des variables continues $X$ et $Y$ de densité jointe $f_{(X,Y)}(s,t)$, si $f_Y(t) \neq 0$, alors on appelle loi conditionnelle de $X$ sachant $Y=t$ la loi de probabilité sur $\mathbb{R}$ dont la densité est

$$f_{X|Y=t}(s) = \frac{f_{(X,Y)}(s,t)}{f_Y(t)} = \frac{f_{(X,Y)}(s,t)}{\int_{s\in\mathbb{R}} f_{(X,Y)}(s,t)\,ds}.$$

Intuitivement, c'est la répartition de probabilité pour $X$ si l'on sait que $Y=t$ (bien que $\mathbb{P}[Y=t]=0$). La densité ci-dessus a bien pour masse totale 1, compte tenu de la deuxième partie de la formule.

Exemple 4.16. Dans l'exemple 4.5, pour $t \in [0,1]$, la loi conditionnelle de $X$ sachant $Y=t$ a pour densité

$$f_{X|Y=t}(s) = \frac{2(s^2+t^2)}{(1+st)\left(\frac{1}{t}-\frac{2}{t^2}+2t\,\log(1+t)+\frac{2\,\log(1+t)}{t^3}\right)}$$

sur le segment $s \in [0,1]$. Ainsi, pour $t=0$, on obtient (en prenant la limite de l'expression pour le dénominateur)

$$f_{X|Y=0}(s) = \begin{cases} 3s^2 & \text{si } s \in [0,1], \\ 0 & \text{sinon.} \end{cases}$$

Pour $t=1$, on obtient par contre

$$f_{X|Y=1}(s) = \begin{cases} \frac{2}{4\log 2-1}\,\frac{1+s^2}{1+s} & \text{si } s \in [0,1], \\ 0 & \text{sinon.} \end{cases}$$

Ce sont deux densités de probabilité différentes (voit la figure 5) : savoir que $Y$ vaut 0

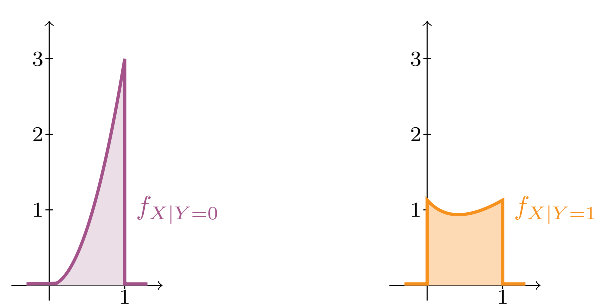


Figure 5. Lois conditionnelles de $X$ sachant $Y=0$ ou $Y=1$, lorsque $f_{(X,Y)}(s,t) = 1_{(s,t)\in[0,1]^2}\,\frac{2(s^2+t^2)}{1+st}$.

ou 1 modifie sensiblement la répartition de probabilité pour la variable $X$.

## 3. Variables aléatoires indépendantes

Avec la notion de lois conditionnelles, on peut donner une nouvelle définition plus naturelle pour la notion d'indépendance de variables aléatoires. Intuitivement, deux variables $X$ et $Y$ sont indépendantes si la connaissance de l'une des deux variables $Y$ ne donne aucune information sur $X$ ; donc, si la loi conditionnelle de $X$ sachant la valeur de $Y$ ($l \in \mathbb{Z}$ ou $t \in \mathbb{R}$) ne dépend pas de cette valeur $l$ ou $t$.

Définition 4.17. *Deux variables aléatoires discrètes $X$ et $Y$ sont dites indépendantes si la loi conditionnelle de $X$ sachant $Y = l$ ne dépend pas de $l$. Dans ce cas, pour tout $l$ tel que $\mathbb{P}_Y[l] \neq 0$,*

$$\mathbb{P}_{X|Y=l} = \mathbb{P}_X,$$

*c'est-à-dire que la loi conditionnelle de $X$ est simplement sa loi marginale.*

En effet, supposons que $\mathbb{P}_{X|Y=l}[k]$ ne dépende pas de $l$. Alors, cette valeur $f(k)$ vérifie l'équation

$$\mathbb{P}_X[k] = \sum_{l\in\mathbb{Z}} \mathbb{P}[X = k \text{ et } Y = l] = \sum_{l\in\mathbb{Z}} \mathbb{P}[Y = l]\,\mathbb{P}_{X|Y=l}[k] = f(k)\left(\sum_{l\in\mathbb{Z}} \mathbb{P}[Y = l]\right) = f(k).$$

Relions maintenant cette nouvelle définition à celle donnée dans le chapitre 2 :

Proposition 4.18. *Deux variables discrètes $X$ et $Y$ sont indépendantes si et seulement si :*

1. *Elles vérifient la condition de la définition 4.17.*
2. *Pour tout $k$ tel que $\mathbb{P}_X[k] \neq 0$, $\mathbb{P}_{Y|X=k} = \mathbb{P}_Y$.*
3. *Pour tous $k$ et $l$, $\mathbb{P}_{(X,Y)}[k, l] = \mathbb{P}_X[k]\,\mathbb{P}_Y[l]$,* i.e., *la loi jointe est le produit des lois marginales.*

En effet, on a toujours

$$\mathbb{P}_{(X,Y)}[k, l] = \mathbb{P}_{X|Y=l}[k]\,\mathbb{P}[Y = l],$$

et si la loi conditionnelle de $X$ sachant $Y = l$ ne dépend pas de $l$, alors elle est égale à la loi marginale de $X$ par le raisonnement précédent, et

$$\mathbb{P}_{(X,Y)}[k, l] = \mathbb{P}[X = k]\,\mathbb{P}[Y = l].$$

Ainsi, la définition 4.17 donne un nouveau critère pour établir l'indépendance de deux variables aléatoires discrètes.

Exemple 4.19. Dans l'exemple 4.2, les deux variables aléatoires $X$ et $Y$ ne sont pas indépendantes, puisque les lois conditionnelles $\mathbb{P}_{X|Y=2}$ et $\mathbb{P}_{X|Y=3}$ ne sont pas les mêmes.

EXEMPLE 4.20. Soit $(X, Y)$ le couple de variables aléatoires à valeurs dans $\{1, 2\} \times \{1, 2\}$, de loi jointe donnée par le tableau suivant

| $k\backslash l$ | 1 | 2 | $\mathbb{P}_X$ |
|---|---|---|---|
| 1 | $\frac{1}{4}$ | $\frac{1}{12}$ | $\frac{1}{3}$ |
| 2 | $\frac{1}{2}$ | $\frac{1}{6}$ | $\frac{2}{3}$ |
| $\mathbb{P}_Y$ | $\frac{3}{4}$ | $\frac{1}{4}$ | 1 |

où l'on a aussi fait figurer les lois marginales de $X$ et de $Y$. Chaque case $\mathbb{P}[X = k \text{ et } Y = l]$ est égale au produit des deux marginales $\mathbb{P}[X = k]$ et $\mathbb{P}[Y = l]$, donc $X$ et $Y$ sont deux variables indépendantes.

Dans le cas continu, on peut de même décider que :

DÉFINITION 4.21. *Deux variables aléatoires continues $X$ et $Y$ sont dites indépendantes si la loi conditionnelle de $X$ sachant $Y = t$ ne dépend pas de $t$. Dans ce cas, pour tout $t$ tel que $f_Y(t) \neq 0$,*

$$f_{X|Y=t} = f_X,$$

*c'est-à-dire que la densité conditionnelle de $X$ est simplement sa densité marginale.*

Des raisonnements analogues à ceux du cas discret montrent l'équivalence entre cette définition et les assertions suivantes :

PROPOSITION 4.22. *Deux variables continues $X$ et $Y$ sont indépendantes si et seulement si :*

1. *Elles vérifient la condition de la définition 4.17.*
2. *Pour tout $s$ tel que $f_X(s) \neq 0$, $f_{Y|X=s} = f_Y$.*
3. *Pour tous $s$ et $t$, $f_{(X,Y)}(s, t) = f_X(s)\, f_Y(t)$,* i.e., *la densité jointe est le produit des densités des lois marginales.*

EXEMPLE 4.23. Deux variables $X$ et $Y$ de loi jointe $f_{(X,Y)}(s, t) = \frac{2(s^2+t^2)}{1+st}\, 1_{(s,t)\in[0,1]^2}$ ne sont pas indépendantes, puisque $f_{(X,Y)}(s, t)$ n'est pas le produit des densités marginales $f_X(s)$ et $f_Y(t)$ précédemment calculées. Alternativement, on a vu que $f_{X|Y=0} \neq f_{X|Y=1}$, donc $X$ et $Y$ ne sont pas indépendantes.

EXEMPLE 4.24. Soit $(X, Y)$ le couple de variables aléatoires suivant la densité donnée dans l'exemple 4.6. On a bien

$$f_{(X,Y)}(s, t) = \frac{1}{2\pi}\, \mathrm{e}^{-\frac{s^2+t^2}{2}} = f_X(s)\, f_Y(t),$$

avec $f_X(s) = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\frac{s^2}{2}}$ et $f_Y(t) = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\frac{t^2}{2}}$. Les variables $X$ et $Y$ sont donc indépendantes (et de même loi gaussienne $\mathcal{N}(0, 1)$).

## 4. Espérances et covariances

Pour conclure, si $(X, Y)$ est un couple de variables aléatoires discrètes ou continues, on peut calculer la moyenne, ou espérance de toute fonction $h(X, Y)$ à valeurs dans $\mathbb{R}$. Cette espérance est définie par la formule

$$\mathbb{E}[h(X,Y)] = \sum_{(k,l)\in\mathbb{Z}^2} \mathbb{P}_{(X,Y)}[k,l]\, h(k,l)$$

dans le cas discret, et par la formule

$$\mathbb{E}[h(X,Y)] = \iint_{(s,t)\in\mathbb{R}^2} f_{(X,Y)}(s,t)\, h(s,t)\, ds\, dt$$

dans le cas continu. À chaque fois, on moyenne la fonction $h(\cdot,\cdot)$ en mettant des poids donnés par la loi jointe du couple $(X, Y)$.

Exemple 4.25. Calculons $\mathbb{E}[XY]$ lorsque $(X, Y)$ suit la loi discrète donnée par le tableau de l'exemple 4.2. La fonction $h$ est

$$\begin{aligned} h : \mathbb{Z}^2 &\to \mathbb{R} \\ (k,l) &\mapsto kl, \end{aligned}$$

et son espérance est

$$\begin{aligned} \mathbb{E}[XY] &= 0 \times (1\times 1) + \frac{1}{12} \times (1 \times 2) + \frac{1}{6} \times (1 \times 3) \\ &\quad + \frac{1}{6} \times (2\times 1) + \frac{1}{4} \times (2\times 2) + \frac{1}{3}\times(2\times 3) \\ &= 4. \end{aligned}$$

Exemple 4.26. Calculons $\mathbb{E}[X^2(3Y^2+1)]$ lorsque $(X, Y)$ a densité jointe $\frac{1}{2\pi}\,\mathrm{e}^{-\frac{s^2+t^2}{2}}$. La fonction $h$ est $h(s,t) = s^2(3t^2+1)$, et

$$\begin{aligned} \mathbb{E}[X^2(3Y^2+X)] &= \iint_{(s,t)\in\mathbb{R}^2} \frac{1}{2\pi}\,\mathrm{e}^{-\frac{s^2+t^2}{2}}\, s^2(3t^2+1)\, ds\, dt \\ &= \left(\frac{1}{\sqrt{2\pi}} \int_{s=-\infty}^{\infty} \mathrm{e}^{-\frac{s^2}{2}}\, s^2\, ds\right)\left(\frac{1}{\sqrt{2\pi}}\int_{t=-\infty}^{\infty} \mathrm{e}^{-\frac{t^2}{2}}\,(3t^2+1)\, dt\right) \\ &= \mathbb{E}[X^2]\,\mathbb{E}[3Y^2+1] = 1\times 4 = 4, \end{aligned}$$

car $X$ et $Y$ suivent des lois gaussiennes $\mathcal{N}(0,1)$, donc $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$.

L'exemple précédent rentre dans le cadre plus général de la proposition suivante :

Proposition 4.27. *Soit $(X, Y)$ un couple de variables aléatoires indépendantes. Pour toute fonction $h(X, Y)$ qui se factorise sous la forme $h(X,Y) = h_1(X)\,h_2(Y)$, on a*

$$\mathbb{E}[h_1(X)\,h_2(Y)] = \mathbb{E}[h_1(X)]\,\mathbb{E}[h_2(Y)].$$

En effet, dans le cas continu par exemple,

$$\begin{aligned}
\mathbb{E}[h_1(X)\,h_2(Y)] &= \iint_{(s,t)\in\mathbb{R}^2} f_{(X,Y)}(s,t)\,h_1(s)\,h_2(t)\,ds\,dt \\
&= \iint_{(s,t)\in\mathbb{R}^2} f_X(s)f_Y(t)\,h_1(s)\,h_2(t)\,ds\,dt \\
&= \left(\int_{s\in\mathbb{R}} f_X(s)\,h_1(s)\,ds\right)\left(\int_{t\in\mathbb{R}} f_Y(t)\,h_2(t)\,dt\right) = \mathbb{E}[h_1(X)]\,\mathbb{E}[h_2(Y)],
\end{aligned}$$

l'exemple précédent correspondant au cas où $h_1(s) = s^2$ et $h_2(t) = 3t^2 + 1$.

Exemple 4.28. Attention, on ne pas factoriser les espérances dans le cas de deux variables non indépendantes. Par exemple, si $(X, Y)$ a pour densité jointe $f_{(X,Y)}(s,t) = 1_{(s,t)\in[0,1]^2}\,\frac{2(s^2+t^2)}{1+st}$, alors

$$\begin{aligned}
\mathbb{E}[XY] &= \int_{s=0}^1\int_{t=0}^1 \frac{2(s^2+t^2)}{1+st}\,st\,ds\,dt = 4\int_{s=0}^1\int_{t=0}^1 \frac{s^3t}{1+st}\,ds\,dt \\
&= 4\int_{s=0}^1 s^2\left(\int_{t=0}^1 1-\frac{1}{1+st}\,dt\right)ds \\
&= 4\int_{s=0}^1 (s^2 - s\log(1+s))\,ds \\
&= 4\left[\frac{s^3}{3}+\frac{s^2}{4}-\frac{s}{2}-\frac{s^2-1}{2}\,\log(1+s)\right]_0^1 = \frac{1}{3}.
\end{aligned}$$

En revanche,

$$\begin{aligned}
\mathbb{E}[X] = \mathbb{E}[Y] &= \int_{s=0}^1 s\left(\frac{1}{s}-\frac{2}{s^2}+2s\,\log(1+s)+\frac{2\,\log(1+s)}{s^3}\right)ds \\
&= \int_{s=0}^1 (1+2s^2\,\log(1+s))\,ds + 2\int_{s=0}^1 \frac{\log(1+s)-s}{s^2}\,ds \\
&= \left[\frac{2(1+s^3)\,\log(1+s)}{3}+\frac{s+s^2}{3}-\frac{2s^3}{9}\right]_{s=0}^1 + 2\int_{s=0}^1 \frac{\log(1+s)-s}{s^2}\,ds \\
&= 4\left(\frac{\log 2}{3}+\frac{1}{9}\right) + 2\int_{s=0}^1 \frac{\log(1+s)-s}{s^2}\,ds.
\end{aligned}$$

Pour résoudre la singularité dans la deuxième intégrale, on utilise comme précédemment une formule de Taylor avec reste intégral :

$$\log(1+s) - s = -s^2\int_{u=0}^1 \frac{(1-u)}{(1+us)^2}\,du.$$

On en déduit

$$\begin{aligned}\int_{s=0}^{1} \frac{\log(1+s)-s}{s^2}\,ds &= -\int_{u=0}^{1}\int_{s=0}^{1} \frac{1-u}{(1+us)^2}\,du \\ &= \int_{u=0}^{1}\left[\frac{1-u}{u(1+us)}\right]_{s=0}^{1} du = -\int_{u=0}^{1}\frac{1-u}{1+u}\,du \\ &= -\left[2\log(1+u)-u\right]_{u=0}^{1} = 1-2\log 2,\end{aligned}$$

et ainsi,

$$\mathbb{E}[X] = \mathbb{E}[Y] = \frac{22}{9} - \frac{8\log 2}{3} \simeq 0.596.$$

On peut alors vérifier que $\mathbb{E}[XY] \neq \mathbb{E}[X]\,\mathbb{E}[Y]$, ce qui n'est pas contradictoire puisque $X$ et $Y$ ne sont pas indépendantes.

La proposition précédente peut être utilisée dans l'autre sens, pour montrer que deux variables ne sont pas indépendantes : s'il existe des fonctions $h_1$ et $h_2$ telles que $\mathbb{E}[h_1(X)h_2(Y)] \neq \mathbb{E}[h_1(X)]\,\mathbb{E}[h_2(Y)]$, alors $X$ et $Y$ ne sont pas indépendantes. Partant de cette idée, on peut construire un outil qui "mesure" la non-indépendance de deux variables aléatoires :

Définition 4.29. *La* ***covariance*** *de deux variables aléatoires discrètes ou continues est*

$$\mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])].$$

*Si $X$ et $Y$ sont deux variables indépendantes, alors* $\mathrm{cov}(X,Y) = 0$.

L'équivalence entre les deux formules se démontre de la même façon que l'équivalence entre les deux formules pour la variance d'une variable aléatoire :

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[(X-\mathbb{E}[X])^2].$$

De fait, la covariance de deux variables aléatoires $X$ et $Y$ est une généralisation de la notion de variance d'une variable aléatoire $X$, car $\mathrm{var}(X) = \mathrm{cov}(X,X)$.

Exemple 4.30. Si $(X,Y)$ est le couple de variables aléatoires de loi jointe donnée par le tableau de l'exemple 4.2, alors on a déjà calculé $\mathbb{E}[XY] = 4$, et d'autre part,

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{4}\times 1 + \frac{3}{4}\times 2 = \frac{7}{4};\\ \mathbb{E}[Y] &= \frac{1}{6}\times 1 + \frac{1}{3}\times 2 + \frac{1}{2}\times 3 = \frac{7}{3}.\end{aligned}$$

La covariance de $X$ et de $Y$ est donc $\mathrm{cov}(X,Y) = 4 - \frac{49}{12} = -\frac{1}{12}$.

Remarque. Attention, on peut avoir $\mathrm{cov}(X,Y) = 0$, et néanmoins $X$ et $Y$ non indépendantes. Par exemple, si $X$ est une variable aléatoire avec

$$\mathbb{P}[X=-1] = \mathbb{P}[X=0] = \mathbb{P}[X=1] = \frac{1}{3},$$

alors en prenant $Y = X^2$, on a

$$\mathbb{E}[X] = 0 \quad ; \quad \mathbb{E}[XY] = \mathbb{E}[X^3] = 0$$

et donc $\mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0 - 0 = 0$. Pourtant, $X$ et $Y$ ne sont pas indépendantes, car la loi conditionnelle de $Y$ sachant $X = 0$ est

$$\mathbb{P}_{Y|X=0}[0] = 1 \qquad ; \qquad \mathbb{P}_{Y|X=0}[1] = 0,$$

et elle est différente la loi conditionnelle de $Y$ sachant $X = 1$ :

$$\mathbb{P}_{Y|X=1}[0] = 0 \qquad ; \qquad \mathbb{P}_{Y|X=1}[1] = 1.$$

Ainsi, la covariance est plus un test de *dépendance* que d'*indépendance*, puisqu'elle ne permet pas de décider si deux variables sont indépendantes.

Exemple 4.31. La covariance de $(X,Y)$ de loi jointe donnée par la densité $f_{(X,Y)}(s,t) = 1_{(s,t)\in[0,1]^2}\,\frac{2(s^2+t^2)}{1+st}$ est, d'après les calculs précédents,

$$\mathrm{cov}(X,Y) = \frac{1}{3} - \left(\frac{22}{9} - \frac{8\log 2}{3}\right)^2 \simeq -0.022.$$

Le signe négatif de la covariance peut être interprété intuitivement comme suit : si $Y$ est grand par rapport à sa moyenne, alors $X$ a tendance à être plus petit que sa moyenne, et si $Y$ est plus petit que sa moyenne, alors $X$ a tendance à avoir des fluctuations positives, donc, à être plus grand que sa moyenne. Autrement dit, les **fluctuations** $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$ ont tendance à être de signes opposés.

Inversement, si $X$ et $Y$ avaient covariance positive, alors leurs fluctuations $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$ auraient tendance à être de même signe, puisque $\mathrm{cov}(X,Y)$ est l'espérance du produit de ces fluctuations.

Théorème 4.32. *Pour toutes variables aléatoires,* $|\mathrm{cov}(X,Y)| \leq \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}$. *On appelle **corrélation** entre $X$ et $Y$ le rapport*

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}}.$$

*Ce coefficient est donc toujours compris entre $-1$ et $1$.*

En effet, pour toutes variables aléatoires $X$ et $Y$ (discrètes ou continues), on obtient

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y)$$

en développant les carrés dans $\mathrm{var}(X+Y)$. Comme la variance est positive, pour tout $t \in \mathbb{R}$,

$$0 \leq \mathrm{var}(tX+Y) = t^2\,\mathrm{var}(X) + 2t\,\mathrm{cov}(X,Y) + \mathrm{var}(Y),$$

donc le discriminant

$$4(\mathrm{cov}(X,Y)^2 - \mathrm{var}(X)\mathrm{var}(Y))$$

de ce polynôme en $t$ est négatif ou nul, ce qui prouve le théorème. Le coefficient de corrélation permet de manipuler des covariances "renormalisées" de variables aléatoires, et de mesurer leurs interactions et leur dépendance indépendamment de la taille de leurs fluctuations $\mathrm{var}(X)$ et $\mathrm{var}(Y)$.

Exemple 4.33. Poursuivant toujours l'exemple 4.2, on a déjà calculé $\mathrm{cov}(X,Y) = -\frac{1}{12}$, et d'autre part,

$$\mathbb{E}[X^2] = \frac{1}{4} \times 1 + \frac{3}{4} \times 4 = \frac{13}{4} \quad ; \quad \mathrm{var}(X) = \frac{13}{4} - \left(\frac{7}{4}\right)^2 = \frac{3}{16}$$

$$\mathbb{E}[Y^2] = \frac{1}{6} \times 1 + \frac{1}{3} \times 4 + \frac{1}{2} \times 9 = 6 \quad ; \quad \mathrm{var}(Y) = 6 - \left(\frac{7}{3}\right)^2 = \frac{5}{9}.$$

On a donc

$$\rho(X,Y) = -\frac{1}{12\sqrt{\frac{3}{16} * \frac{5}{9}}} = -\frac{1}{\sqrt{15}}.$$

---

Exercice. Soit $X$ une variable aléatoire suivant une loi de Poisson de paramètre $\lambda$, et $Y$ une variable aléatoire telle que la loi conditionnelle de $Y$ sachant $X = k$ est une loi binomiale $\mathcal{B}(k,p)$, $p$ étant un paramètre fixé entre 0 et 1. Trouver la loi jointe de $X$ et $Y$, puis les lois marginales de $X$ et de $Y$. Calculer la covariance de $X$ et $Y$.

Exercice. Soit $(X,Y)$ un couple de variables aléatoires de densité

$$f_{(X,Y)}(s,t) = \begin{cases} 3s^2t + s & \text{si } 0 \le s,t \le 1, \\ 0 & \text{sinon.} \end{cases}$$

Vérifier que la formule précédente définit bien une densité de probabilité. Calculer les marginales de $X$ et de $Y$ . Calculer la loi conditionnelle de $X$ sachant $Y = t$. Les variables $X$ et $Y$ sont elles indépendantes ? Calculer la covariance de $X$ et $Y$, et leur coefficient de corrélation.