# Chapitre 1

# Probabilités

## 1.1 Dénombrement : motivation

### 1.1.1 Calculer une probabilité par dénombrement

**Exemple 1** *Je tire au hasard 2 étudiants parmi les 57 présents dans la salle. Il y a*

| | *Groupe 3* | *Groupe 4* | *TOTAL* |
|---|---|---|---|
| *Filles* | *3* | *10* | *13* |
| *Garçons* | *25* | *19* | *44* |
| *TOTAL* | *28* | *29* | *57* |

*Quelle est la probabilité que j'aie tiré 1 étudiant(e) de chaque genre et 1 étudiant(e) de chaque groupe.*

Au total, il y a $57 \times 56 = 3192$ *arrangements* de 2 étudiant(e)s parmi 57. On considère que tous ces arrangements sont équiprobables.

Les arrangements favorables sont de la forme $(F3, G4)$, $(F4, G3)$, $(G3, F4)$ et $(G4, F3)$. Donc l'ensemble $E$ des arrangements favorables est la réunion de 4 sous-ensembles disjoints $A$, $B$, $C$ et $D$.

$$\begin{aligned}
\mathrm{card}A &= \mathrm{card}\{\text{filles du groupe 3}\} \times \mathrm{card}\{\text{garçons du groupe 4}\} = 3 \times 19 = 357,\\
\mathrm{card}B &= \mathrm{card}\{\text{filles du groupe 4}\} \times \mathrm{card}\{\text{garçons du groupe 3}\} = 10 \times 25 = 250,\\
\mathrm{card}C &= \mathrm{card}\{\text{garçons du groupe 3}\} \times \mathrm{card}\{\text{filles du groupe 4}\} = 25 \times 10 = 250,\\
\mathrm{card}D &= \mathrm{card}\{\text{garçons du groupe 4}\} \times \mathrm{card}\{\text{filles du groupe 3}\} = 19 \times 3 = 357.
\end{aligned}$$

D'où $\mathrm{card}E = 357 + 250 + 250 + 357 = 1214$. La probabilité de tirer 1 étudiant(e) de chaque genre et 1 étudiant(e) de chaque groupe vaut

$$P(E) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{1214}{3192} = 0.38.$$

**Remarque 2** *Autre méthode.*

On aurait pu écrire $E = H \cup I$ où $H = \{$la fille tirée est du groupe 3$\}$ et $I =\{$la fille tirée est du groupe 4$\}$. Il y a deux ordres possibles (fille en premier, garçon en premier). On compte les arrangements qui commencent avec la fille en premier (il y en a 357 dans $H$, 250 dans $I$) donc le nombre total est $\mathrm{card}H = 2 \times 357$, $\mathrm{card}I = 2 \times 250$

**Exemple 3** *Généralisation. Probabilité pour qu'en tirant 3 étudiants au hasard, on tire exactement 1 fille et exactement 1 étudiant(e) du groupe 3.*

Le nombre d'arrangements de 3 étudiant(e)s parmi 57 est $57\times 56\times 55 = 175560$. Soit $E$ ={tirages avec exactement 1 fille et exactement 1 étudiant(e) du groupe 3}. Soit $H$ le sous-ensemble de $E$ formé des tirages dans lesquels la fille est du groupe 3. Former un élément de $H$, c'est choisir d'abord la fille dans le groupe 3, puis choisir un garçon dans le groupe 4, puis choisir un autre garçon (distinct du premier) dans le groupe 4, puis choisir l'ordre dans lequel ils vont sortir (il y en a autant que de *permutations* de 3 personnes, soit 6). Par conséquent,

$$P(H) = 3 \times 19 \times 18 \times 6 = 6156.$$

Soit $I$ le sous-ensemble de $E$ formé des tirages dans lesquels la fille est du groupe 4. Former un élément de $I$, c'est choisir d'abord la fille dans le groupe 4, puis choisir un garçon dans le groupe 3, puis choisir un garçon dans le groupe 4, puis choisir l'ordre dans lequel ils vont sortir (il y en a autant que de *permutations* de 3 personnes, soit 6). Par conséquent,

$$P(I) = 10 \times 25 \times 19 \times 6 = 28500.$$

Comme $H$ et $I$ sont disjoints et $E = H \cup I$,

$$\mathrm{card}E = \mathrm{card}H + \mathrm{card}I = 6156 + 28500 = 34656.$$

D'où

$$P(E) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{34656}{175560} = 0.197.$$

**Remarque 4** *Autre méthode.*

Au lieu des compter des arrangements, on compte des paires ou des triplets non ordonnés (on parle plus généralement de *combinaisons*), puisque l'ordre dans lequel les individus sont tirés n'a pas d'importance.

Posons $O'$ ={combinaisons de 3 individus parmi 57}, i.e. $\mathrm{card}O' = \frac{57\times 56\times 55}{6} = 29260$. L'ensemble qui nous intéresse est $E'$ ={combinaisons de 3 individus parmi 57 comportant exactement 1 fille et 1 groupe 3}. Alors $H'$ ={combinaisons de 3 individus dont 1 fille du groupe 3 et 2 garçons du groupe 4} a $3 \times 19 \times 18 = 1026$ éléments. De même, $I'$ ={combinaisons de 3 individus dont 1 fille du groupe 4 et 1 garçon de chaque groupe} a $10 \times 25 \times 19 = 4750$ éléments. Il vient

$$\mathrm{card}E' = \mathrm{card}H' + \mathrm{card}I' = 1026 + 4750 = 5776,$$

$$P(E) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{5776}{29260} = 0.197.$$

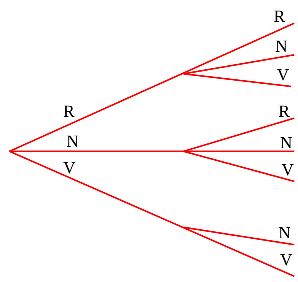
## 1.2 Techniques de dénombrement

### 1.2.1 Diagrammes arborescents ou arbres

**Exemple 5** *On considère une urne qui contient deux boules rouges, deux noires et une verte. On tire deux boules sans remise. Il s'agit d'une expérience à deux étapes où les différentes possibilités qui peuvent survenir sont représentées par un arbre horizontal.*

On obtient trois branches principales et trois branches secondaires pour chaque étape sauf pour le cas où une verte a été tirée en premier.

Le nombre de branches terminales de cet arbre donne le nombre d'éléments de l'univers.



Lorsqu'on rencontre beaucoup d'étapes dans une expérience et de nombreuses possibilités à chaque étape, l'arbre associé à l'expérience devient trop complexe pour être analysé. Ces problèmes se simplifient à l'aide de formules algébriques, comme on va le voir.

La démonstration de ces formules repose sur le fait que dans le cas d'une expérience à deux étapes, par exemple, un arbre qui aurait $r$ branches principales et $s$ branches secondaires commençant à partir des $r$ branches principales aura $rs$ branches terminales.

### 1.2.2 Arrangements et permutations

Envisageons un ensemble de $n$ objets différents. Choisissons maintenant $r$ de ces $n$ objets et ordonnons les.

**Définition 6** *Une disposition ordonnée de $r$ objets distincts pris parmi $n$ est appelée* arrangement *de $r$ objets pris parmi $n$ (on a obligatoirement $r \leq n$).*

**Combien y en a-t-il** ?

Pour compter le nombre total d'arrangements de $r$ objets pris parmi $n$, il suffit de considérer les $r$ positions comme fixées et de compter le nombre de façons dont on peut choisir les objets pour les placer dans ces $r$ positions. C'est une expérience à $r$ étapes où l'on applique la technique du paragraphe précédent. Pour la première position, on a $n$ choix possibles. Pour la deuxième position, on a $n-1$ choix possibles... Pour la $r$-ième position, on a $n-r+1$ choix possibles. Si on désigne par $A_n^r$ le nombre total d'arrangements cherchés, l'arbre aura $A_n^r$ branches terminales. On conclut

**Proposition 7**

$$A_n^r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}.$$

**Rappel 8** *$n!$ (lire "factorielle $n$") est le produit de tous les entiers jusqu'à $n$, $n! = n(n-1)(n-2)\cdots 3.2.1$. Par convention, $0! = 1$.*

**Exemple 9** *Les arrangements de deux lettres prises parmi 4 lettres $\{a,b,c,d\}$ sont au nombre de $A_4^2 = \frac{4!}{2!} = 12$. Ce sont : $(a,b),(a,c),(a,d),(b,a),(b,c),(b,d)$, $(c,a),(c,b),(c,d),(d,a),(d,b),(d,c)$.*

**Cas particulier** : $r = n$ Il s'agit d'ordonner $n$ objets entre eux, c'est-à-dire d'effectuer une permutation de ces n objets.

**Définition 10** *Une permutation de $n$ éléments est une disposition ordonnée de ces $n$ éléments.*

**Proposition 11** *Les permutations de $n$ éléments sont au nombre de $A_n^n = n!$.*

### 1.2.3 Combinaisons

**Définition 12** *Un choix de $r$ objets distincts pris parmi $n$ sans tenir compte de leur ordre est appelé combinaison de $r$ objets pris parmi $n$.*

Dans l'exemple précédent correspondant à l'ensemble des quatre lettres $\{a,b,c,d\}$, la combinaison $\{a,b\}$ est la même que la combinaison $\{b,a\}$ alors que l'arrangement $(a,b)$ est différent de l'arrangement $(b,a)$.

**Combien y en a-t-il** ? Le nombre total de combinaisons de $r$ objets pris parmi $n$ est noté $C_n^r$ ou $\binom{r}{n}$. Pour trouver l'expression de $\binom{r}{n}$, comparons le nombre d'arrangements et de combinaisons possibles de $r$ objets pris parmi $n$.

- Dans un arrangement on choisit $r$ objets, puis on tient compte de leur ordre.
- Dans une combinaison seul le choix des $r$ objets compte. Comme le nombre de façons d'ordonner les $r$ objets choisis est $r!$, on conclut qu'à chaque combinaison de $r$ objets pris parmi $n$, on peut associer $r!$ arrangements et donc qu'il y a $r!$ fois plus d'arrangements que de combinaisons.

On conclut

**Proposition 13**

$$\binom{r}{n} = \frac{A_n^r}{r!} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}.$$

**Exemple 14** *Le nombre de combinaisons de deux lettres prises parmi quatre* $\{a, b, c, d\}$ *est* $\binom{2}{4} = \frac{4!}{2!2!} = 6$. *Ce sont :* $\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}$.

### 1.2.4 Permutations lorsque certains éléments sont semblables

Dans les paragraphes précédents, on a supposé que les $n$ objets étaient tous différents. Il arrive parfois que les $n$ objets en contiennent un certain nombre qui sont indiscernables.

Supposons qu'il n'y ait que $k$ sortes d'objets distincts sur les $n$ objets. Il y a
- $n_1$ objets de la 1-ère sorte,
- $n_2$ objets de la 2-ème sorte....
- $n_k$ objets de la $k$-ème sorte.

On a bien sûr $n_1 + n_2 + \cdots + n_k = n$.

Pour déterminer le nombre total de permutations distinctes, comparons ce nombre cherché $\mathcal{P}$ avec le nombre obtenu si on supposait les objets différenciés. Plaçons nous dans le cas de l'exemple suivant : On cherche le nombre d'anagrammes du mot $PROBABILITE$.

Choisissons un de ces anagrammes : le plus simple est $PROBABILITE$.
- Si on différencie les lettres $B$, cette disposition peut provenir des deux permutations $PROB_1AB_2ILITE$ ou $PROB_2AB_1ILITE$, soit 2! possibilités.
- Si on différencie les lettres $I$, cette disposition peut provenir des deux permutations $PROBABI_1LI_2TE$ ou $PROBABI_2LI_1TE$, soit encore 2! possibilités

A un anagramme correspond donc $2! \times 2! = 4$ permutations, ce qui signifie qu'il y a 4 fois plus de permutations que d'anagrammes. Le mot $PROBABILITE$ comprend 11 lettres. Il y a 11! permutations possibles. On a donc $\frac{11!}{2!2!} = 9979200$ anagrammes possibles.

**Cas général**. La différenciation des $n_1$ premiers objets donnera $n_1!$ fois plus d'éléments que ce qu'on cherche, la différenciation des $n_2$ premiers objets donnera $n_2!$ fois plus d'éléments que ce qu'on cherche, et finalement on trouve que $n!$ est $n_1!n_2!\cdots n_k!$ fois plus grand que le nombre cherché $\mathcal{P}$. On conclut

**Proposition 15** *Le nombre d'anagrammes d'un mot de* $n$ *lettres, comportant seulement* $k < n$ *lettres distinctes, en nombres* $n_1, \ldots, n_k$ *est*

$$\mathcal{P} = \frac{n!}{n_1!n_2!\cdots n_k!}.$$

### 1.2.5 Cas ou les éléments ne sont pas obligatoirement distincts

Combien y a-t-il de manières de choisir $r$ éléments parmi $n$ de façon ordonnée en n'imposant pas qu'ils soient tous distincts les uns des autres ?

En 1ère position, il y a $n$ choix possibles. En 2ème position, il y a encore $n$ choix possibles... En $r$ème position, il y a toujours $n$ choix possibles.

Conclusion : Il y a donc $n^r$ choix pour les r éléments ($r$ peut être supérieur à $n$ dans ce cas).

### 1.2.6 Récapitulation

| | Conditions | Le nombre de tirages possibles est le nombre de : | Un exemple usuel |
|---|---|---|---|
| $p \geq n$ | les $p$ éléments ne sont pas nécessairement . tous distincts mais sont ordonnés | $p$-listes d'éléments de E, soit : $n^p$ | tirages successifs . avec remise de $p$ objets parmi $n$ |
| $p < n$ | les $p$ éléments sont tous distincts et ordonnés | arrangements de $p$ éléments de $E$, soit : $A_n^p$ | tirages successifs sans remise de $p$ objets parmi $n$. |
| $p = n$ | les $n$ éléments sont tous distincts et ordonnés | permutations des $n$ éléments de $E$, soit : $n!$ | anagrammes d'un . mot formé de lettres toutes distinctes |
| $p < n$ | les $p$ éléments sont tous distincts et non ordonnés | combinaisons de $p$ éléments de E, soit $\binom{p}{n}$ | tirages simultanés de $p$ objets parmi $n$. |

## 1.3 Probabilités : motivation

### 1.3.1 Solution de l'exemple 1 par probabilités conditionnelles

On note $\Omega$ l'ensemble des arrangements de 2 étudiants parmi 57, muni de la probabilité uniforme. La probabilité de l'évènement $A$, c'est la probabilité de tirer en premier une fille du groupe 3, soit 3/57, multipliée par la probabilité, sachant qu'on a dejà tiré une fille du groupe 3, de tirer un garçon du groupe 4, soit

$$P(A) = P(F3) \times P(G4|F3) = \frac{3}{57} \times P(G4|F3).$$

Or une fois qu'on a tiré une fille du groupe 3, il reste à tirer un étudiant dans un paquet de 56 qui compte 19 garçons du groupe 4, donc $P(G4|F3) = \frac{19}{56}$. Il vient $P(A) = \frac{3\times 19}{57\times 56} = \frac{357}{3192}$. Idem pour $B$, $C$ et $D$. Donc

$$P(E) = \frac{357}{3192} + \frac{250}{3192} + \frac{250}{3192} + \frac{357}{3192} = \frac{1214}{3192} = 0.38.$$

En fait, on a noté $F3$ l'évènement {le premier tirage est une fille du groupe 3} et $G4$ l'évènement {le second tirage est un garçon du groupe 4}. Alors $A = F3 \cap G4$, et on a utilisé le principe

$$P(F3 \cap G4) = P(F3) \times P(G4|F3).$$

**Exemple 16** *Quelle est la probabilité que tous les étudiants dans la salle aient des dates d'anniversaires distinctes ?*

On fait l'hypothèse que les dates d'anniversaires des 57 étudiants sont indépendantes et que pour chaque étudiant, toutes les dates de 1 à 365 sont équiprobables.

Soit $\Omega =${listes de 57 dates d'anniversaire}. Alors card$\Omega = 365^{57} = 10^{146}$. Soit $E =${listes de 57 dates d'anniversaires toutes distinctes}={arrangements de 57 dates parmi 365}. D'où

$$P(E) = \frac{A_{365}^{57}}{365^{57}} = 0.00988.$$

On peut aussi raisonner par probabilités conditionnelles. Soit $A_n$ l'évènement {les $n$ premières dates sont distinctes} et $B_n$ l'évènement {la $n$-ème est distincte des $n-1$ dates précédentes}. Alors

$A_n = B_n \cap A_{n-1}$, et $P(B_n|A_{n-1}) = \frac{365-n+1}{365}$, d'où

$$\begin{aligned} P(A_n) &= P(B_n \cap A_{n-1}) \\ &= P(A_{n-1}) \times P(B_n|A_{n-1}) \\ &= P(A_{n-1}) \times \frac{365-n+1}{365} \\ &= \frac{365-1}{365} \cdots \frac{365-n+2}{365} \frac{365-n+1}{365}. \end{aligned}$$

On conclut car $E = A_{57} = 0.00988$.

## 1.4 Probabilité sur un ensemble fini

### 1.4.1 Evènement aléatoire

Historiquement, la notion de probabilité s'est dégagée à partir d'exemples simples empruntés aux jeux de hasard (le mot hasard vient de l'arabe *az-zahr* : le dé).

Nous allons introduire cette notion en l'associant à un exemple : le jeu de dé.

| DÉFINITIONS | EXEMPLE |
|---|---|
| Une expérience aléatoire est une expérience dont on ne peut prévoir le résultat. | L'expérience est le jet d'un dé cubique ordinaire. Le résultat de l'expérience est le nombre indiqué sur la face supérieure du dé. |
| On peut alors lui associer alors un univers appelé aussi ensemble fondamental de l'expérience qui est l'ensemble de tous les résultats possibles de l'expérience aléatoire. On le note $\Omega$. | $\Omega = \{1, 2, 3, 4, 5, 6\}$. |
| Un événement aléatoire est un sous-ensemble de $\Omega$. | L'événement *obtenir un nombre pair* est le sous-ensemble $A = \{2, 4, 5\}$ de $\Omega$. |
| On dit que l'événement $A$ est réalisé si le résultat de l'expérience appartient à $A$. | Si la face supérieure du dé indique 5, $A$ n'est pas réalisé. Si elle indique 4, $A$ est réalisé. |
| Si un événement ne contient qu'un seul élément, on dit que c'est un événement élémentaire. | $B = \{1\}$ est un des 5 . événements élémentaires de $\Omega$ |

### 1.4.2 De la fréquence à la probabilité de réalisation d'un évènement aléatoire

La fréquence théorique d'un événement est la limite de la fréquence de réalisation de cet événement lorsque le nombre de répétitions d'une même expérience tend vers l'infini (c'est ce qu'exprime une des lois de la théorie des probabilités appelée la loi faible des grands nombres).

Cela signifie que si l'on veut connaître la fréquence théorique d'apparition du nombre 6 dans notre jet de dé, il suffit de le lancer un grand nombre de fois, 10000 par exemple. La fréquence théorique cherchée, que l'on appellera probabilité de réalisation de l'événement élémentaire $\{6\}$, sera très voisine de la fréquence expérimentale d'apparition du nombre 6 au cours de nos 10000 lancers. Elle sera encore plus voisine de la fréquence expérimentale obtenue lors de 100000 lancers. Le grand nombre de répétitions de l'expérience aléatoire efface la notion de "chance".

La notion de fréquence théorique ou de probabilité va permettre d'indiquer si, lors d'une expérience aléatoire, un événement donné est plus ou moins susceptible d'être réalisé.

Les probabilités peuvent être classées suivant trois critères :

- Une probabilité à priori est une probabilité déterminée à l'avance, sans effectuer aucune expérience.

**Exemple**. On peut à priori accorder une probabilité de 0.5 à événement qui consiste à obtenir le côté face d'une pièce de monnaie non truquée.

- La probabilité empirique d'un événement est déterminée à l'aide de l'observation et de l'expérimentation. C'est la valeur limite de la fréquence de réalisation de événement lorsque l'expérience est réalisée un très grand nombre de fois.
  **Exemple**. Si lorsqu'on a lancé la pièce de monnaie 10000 fois on constate que la fréquence du côté face se stabilise autour de 0.65, il faut envisager de réviser notre probabilité à priori et conclure que la pièce est truquée. Ce type de probabilité joue un rôle important pour les prévisions d'articles en stock chez un détaillant, pour le calcul des primes des compagnies d'assurances, etc...
- La probabilité subjective est le dernier type de probabilité. Elle intervient lorsqu'il est impossible d'établir une probabilité à priori ou une probabilité empirique.
  **Exemple**. Le directeur d'une entreprise peut en se fiant à son expérience affirmer qu'il y a une probabilité de 0.6 que ses employés déclenchent une grève.

Nous nous intéresserons principalement aux deux premiers types de probabilité.

Vu qu'une probabilité peut être considérée comme une fréquence idéale, on lui connaît d'avance certaines propriétés.

- Une probabilité est une quantité sans dimension.
- Elle est toujours comprise entre 0 et 1.
- L'univers $\Omega$ a la probabilité maximum d'être réalisé, car c'est l'événement certain. Sa probabilité de réalisation est donc égale à 1.
- Si $A$ et $B$ sont *incompatibles* (ensembles disjoints), la fréquence de réalisation de l'événement $A$ ou $B$ est la somme de la fréquence de réalisation de $A$ et de la fréquence de réalisation de $B$.

### 1.4.3 Propriétés des probabilités d'un évènement aléatoire

**Définition 17** *Si l'univers $\Omega$ est constitué de $n$ événements élémentaires $\{e_i\}$, une* mesure de probabilité *sur $\Omega$ consiste à se donner $n$ nombres $p_i \in [0,1]$, les probabilités des événements élémentaires, tels que*

$$\sum_{i=1}^{n} p_i = 1.$$

*Si l'événement $A$ est la réunion disjointe de $k$ événements élémentaires $\{e_i\}$, avec $0 < k < n$, la probabilité de $A$ vaut, par définition,*

$$p(A) = p(\bigcup_{i=1}^{k}\{e_i\}) = \sum_{i=1}^{k} p(e_i) = \sum_{i=1}^{k} p_i.$$

*Par suite, $0 \leq p(A) \leq 1$.*

La signification concrète de la probabilité d'un événement $A$ est la suivante. Dans une expérience aléatoire, plus $p(A)$ est proche de 1, plus $A$ a de chances d'être réalisé ; plus $p(A)$ est proche de 0, moins il a de chances d'être réalisé.

**Exemple 18** *Probabilité* uniforme *ou* équiprobabilité *: tous les $P_i$ valent $1/n$. La probabilité d'un sous-ensemble à $k$ éléments vaut alors $p(A) = \dfrac{k}{n} = \dfrac{cardA}{card\Omega}$.*

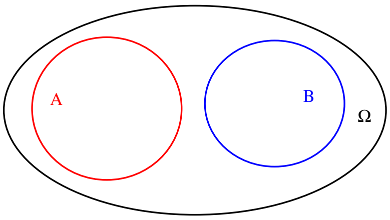
On exprime aussi cette propriété par la formule

$$p(A) = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}.$$

Les propriétés suivantes découlent de la définition.

**Propriété 19** *Si $A$ et $B$ sont incompatibles, i.e., si leur intersection $A \cap B$ est vide, alors*

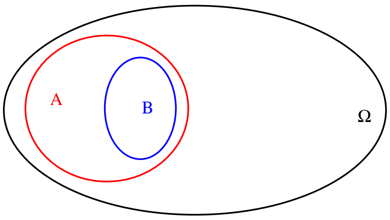
$$p(A \cup B) = p(A) + p(B).$$




**Propriété 20** *Si $B$ est un sous-ensemble de $A$,*

$$B \subseteq A \Rightarrow p(B) \leq p(A).$$

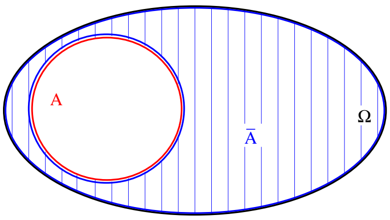
En effet, $B = A \cup (B \setminus A)$. Or $A$ et $B \setminus A$ sont incompatibles ($B \setminus A$ est l'ensemble des éléments de $B$ qui ne sont pas éléments de $A$). Donc $p(B) = p(A) + p(B \setminus A)$. Comme $p(B \setminus A)$ est positive, on obtient le résultat annoncé.




**Propriété 21** *On appelle $\emptyset$ l'évènement impossible, puisqu'il n'est jamais réalisé. Sa probabilité vaut $p(\emptyset) = 0$.*

**Propriété 22** *On note $\bar{A}$ l'évènement contraire de $A$. C'est le complémentaire de $A$ dans $\Omega$. Sa probabilité vaut*

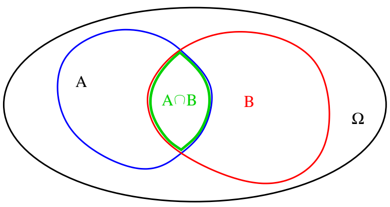
$$p(\bar{A}) = 1 - p(A).$$




**Propriété 23** (Théorème des probabilités totales). *Si $A$ et $B$ sont deux sous-ensembles de $\Omega$,*

$$p(A \cup B) = p(A) + p(B) - p(A \cap B).$$

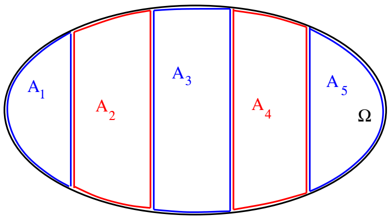
**Preuve.** $p(A) = p(A \setminus B) + p(A \cap B)$ car $A \setminus B$ et $A \cap B$ sont incompatibles. De même $p(B) = p(B \setminus A) + p(A \cap B)$ car $B \setminus A$ et $A \cap B$ sont incompatibles. De plus, $p(A \cup B) = p(A \setminus B) + p(B \setminus A) + p(A \cap B)$, car $A \setminus B$ , $B \setminus A$ et $A \cap B$ sont incompatibles. En additionnant, il vient $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.■



**Propriété 24** (Généralisation du théorème des probabilités totales ou règle de l'addition). *Si $A_1, \dots, A_k$ forment une* partition *de $\Omega$, i.e. ils sont deux à deux disjoints ($i \neq j \Rightarrow A_i \cap A_j = \emptyset$), et $\Omega = \bigcup_{j=1}^{k} A_j$, alors*

$$\sum_{j=1}^{k} p(A_j) = 1.$$

Dans cette situation, on dit parfois que les $A_j$ forment un *système complet d'évènements.*




## 1.5 Probabilité conjointe

La probabilité que deux événements $A$ et $B$ se réalisent est appelée probabilité conjointe de $A$ et $B$, notée $p(A \cap B)$ et s'énonçant probabilité de $A$ et $B$. Le calcul de cette probabilité s'effectue de manière différente selon que $A$ et $B$ sont dépendants ou indépendants, c'est-à-dire selon que la réalisation de l'un influence ou non celle de l'autre.

### 1.5.1 Evénements indépendants

**Exemple 25** *Je lance un dé rouge et un dé vert et je cherche la probabilité d'obtenir un total de 2. Je dois donc obtenir 1 avec chacun des deux dés. La probabilité d'obtenir 1 avec le dé rouge est* $1/6$ *et demeurera* $1/6$ *quelque soit le résultat du dé vert. Les deux événements "obtenir 1 avec le dé rouge" et "obtenir 1 avec le dé vert" sont indépendants.*

**Propriété 26** *Si deux événements sont indépendants, la probabilité qu'ils se réalisent tous les deux est égale au produit de leurs probabilités respectives. On peut donc écrire :*

$$p(A \cap B) = p(A) \times p(B).$$

Dans notre exemple : $p(\text{total} = 2) = p(\text{dé vert} = 1) \times p(\text{dé rouge} = 1) = 1/36$.

**Remarque 27** *Les tirages avec remise constituent une bonne illustration d'événements indépendants.*

### 1.5.2 Evénements dépendants - probabilité conditionnelle

Si deux événements sont dépendants plutôt qu'indépendants, comment calculer la probabilité que les deux se réalisent, puisque la probabilité de réalisation de l'un dépend de la réalisation de l'autre ? Il nous faut connaître pour cela le degré de dépendance des deux événements qui est indiqué par la notion de probabilité conditionnelle.

**Définition 28** *Soient $A$ et $B$ deux événements, $A$ étant supposé de probabilité non nulle. On appelle* probabilité conditionnelle *de $B$ par rapport à $A$, la probabilité de réalisation de l'événement $B$ sachant que $A$ est réalisé. On la note*

$$p(B|A) = \frac{p(A \cap B)}{p(A)}.$$

$p(B|A)$ se lit *p de B si A* ou *p de B sachant A*.

**Remarque 29** *L'application :* $p_B : A \mapsto p_B(A) = p(A|B)$, $\Omega \to [0,1]$, *est une probabilité sur* $\Omega$ *et vérifie toutes les propriétés d'une probabilité.*

**Théorème 1** (Théorème des probabilités composées ou règle de la multiplication).

$$p(A \cap B) = p(B|A)p(A) = p(A|B)p(B).$$

En voici une généralisation. Soit $A_1, \ldots, A_k$ un système complet d'évènements. Alors

$$p(B) = \sum_{j=1}^{k} p(B \cap A_j) = \sum_{j=1}^{k} p(A_j)p(B|A_j).$$

**Théorème 2** (Formule de Bayes). *Soit* $A_1, \ldots, A_k$ *un système complet d'évènements. Soit* $E$ *un évènement de probabilité non nulle. Alors*

$$p(A_j|E) = \frac{p(A_j \cap E)}{p(E)} = \frac{p(A_j)p(E|A_j)}{\sum_{i=1}^{k} p(A_i)p(E|A_i)}.$$

**Remarque 30** *Les tirages sans remise constituent une bonne illustration d'événements dépendants.*

**Exercice 31** *Une urne contient 5 boules noires et 3 boules blanches. Quelle est la probabilité d'extraire 2 boules blanches en 2 tirages ?*

**Solution de l'exercice 31.** *Tirage sans remise.*
Appelons $B_1$, l'événement : obtenir une boule blanche au premier tirage.
Appelons $B_2$, l'événement : obtenir une boule blanche au deuxième tirage.
La probabilité cherchée $p(B_1 \cap B_2)$ est égale à $p(B_1) \times p(B_2|B_1)$. Or $p(B_1)$ vaut $3/8$ et $p(B_2|B_1)$ est égale à $2/7$ puisque lorsqu'une boule blanche est sortie au premier tirage, il ne reste plus que 7 boules au total, dont 2 seulement sont blanches. On conclut que $p(B_1 \cap B_2) = \frac{3}{8} \times \frac{2}{7} = \frac{3}{28}$.

## 1.6 Comment aborder un exercice de probabilités ?

Dans de nombreux problèmes, la recherche des solutions peut être facilitée par la démarche suivante.

1. Déterminer la liste des événements élémentaires ou décrire le contenu de l'univers $\Omega$.
2. Rechercher la mesure de probabilité associée à cet univers.
   – Soit la probabilité est uniforme et dans ce cas, la probabilité d'un événement $A$ est donnée par $p(A) = \frac{\mathrm{card} A}{\mathrm{card} \Omega}$.
   – Soit on détermine la probabilité de chaque événement élémentaire en n'oubliant pas que la somme de toutes les probabilités de ces événements élémentaires est égale à 1.
3. Identifier correctement le ou les événements dont on cherche à évaluer la probabilité.
4. Utiliser la formule appropriée permettant de calculer la probabilité demandée. On pourra se poser la question suivante : Doit-on calculer la probabilité ?
   – D'un événement élémentaire ?
   – D'un événement contraire ?
   – Événements compatibles ou incompatibles (probabilités totales) ?
   – Événements dépendants ou indépendants (probabilités composées) ?

**Exercice 32**
1. *On jette deux dés non pipés. Quelle est la probabilité d'obtenir un total de 7 points ?*
2. *Cette fois-ci les dés sont pipés : les numéros pairs sont deux fois plus probables que les numéros impairs. Quelle est la probabilité d'obtenir un total différent de 8 ?*

**Solution de l'exercice 32.** *Dés pipés.*

1. – L'univers est l'ensemble de tous les résultats possibles lorsqu'on jette deux dés. Imaginons que les deux dés sont reconnaissables et les résultats sont donc tous les couples $(a, b)$ où $a$ et $b$ sont des nombres compris entre 1 et 6. Il contient donc 36 éléments. On peut écrire $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$ et $\text{card}(\Omega) = 36$.
   – Tous les résultats possibles sont équiprobables. La mesure de probabilité est donc uniforme sur $\Omega$.
   – L'événement dont on cherche la probabilité est (somme $= 7$). Il est composé des événements élémentaires $(1, 6)$, $(6, 1)$, $(2, 5)$ , $(5, 2)$, $(3, 4)$, $(4, 3)$. Ils sont au nombre de 6. On peut écrire : $\text{card}(\text{somme} = 7) = 6$.
   – Finalement, étant donné que $p(A) = \frac{\text{card}A}{\text{card}\Omega}$, on obtient $p(\text{somme} = 7) = \frac{6}{36} = \frac{1}{6}$.
2. – L'univers est toujours le même.
   – On cherche à déterminer la mesure de probabilité sur $\Omega$ dans le cas où les dés sont truqués : elle n'est plus uniforme.
   Il faut répondre à la question : lorsqu'on lance un seul dé, quelle est la probabilité de chaque numéro ?
   – Tous les numéros pairs ont la même probabilité que l'on note $p^p$ ; tous les numéros impairs ont la même probabilité que l'on note $p^i$. L'énoncé nous permet d'écrire que $p^p = 2p^i$.
   – D'autre part, étant donné que les numéros 1,2,3,4,5,6 constituent l'ensemble des résultats d'un jet de dé, la somme des probabilités de ces 6 résultats vaut 1. D'où $3p^p + 3p^i = 1$, soit encore $9p^i = 1$. D'où $p^i = \frac{1}{9}$ et $p^p = \frac{2}{9}$.
   – L'événement dont on cherche la probabilité est (somme $\neq 8$). Chercher directement la probabilité de cet événement nous obligerait à considérer beaucoup de cas. Il sera donc plus rapide de déterminer d'abord la probabilité de l'événement contraire (somme $= 8$). Ce dernier est constitué des événements élémentaires $(2, 6)$, $(6, 2)$, $(3, 5)$, $(5, 3)$, $(4, 4)$.
   – Les résultats des deux dés sont indépendants. Nous pouvons donc affirmer que
   $$p(\{(2, 6)\}) = p(\{2\}) \times p(\{6\}) = \frac{2}{9} \times \frac{2}{9} = \frac{4}{81}.$$
   De même, $p(\{(6, 2)\}) = p(\{(4, 4)\}) = \frac{4}{81}$, alors que $p(\{(5, 3)\}) = p(\{(3, 5)\}) = \frac{1}{81}$
   – Finalement $p(\text{somme} = 8) = \frac{14}{81}$ et $p(\text{somme} \neq 8) = \frac{67}{81}$.

# Chapitre 2

# Variables aléatoires

## 2.1 Définition

**Exemple 33** *On jette deux fois une pièce de monnaie non truquée, et on s'intéresse au nombre de fois que le côté "face" a été obtenu. Pour calculer les probabilités des divers résultats, on introduira une variable $X$ qui désignera le nombre de "face" obtenu. $X$ peut prendre les valeurs 0,1,2.*

**Exemple 34** *On lance une fléchette vers une cible circulaire de rayon égal à 50 cm et on s'intéresse à la distance entre la fléchette et le centre de la cible. On introduira ici une variable $X$, distance entre l'impact et le centre de la cible, qui peut prendre n'importe quelle valeur entre 0 et 50.*

Dans ces deux cas, $X$ prend des valeurs réelles qui dépendent du résultat de l'expérience aléatoire. Les valeurs prises par $X$ sont donc aléatoires. $X$ est appelée variable aléatoire.

**Définition 35** *Soit un univers $\Omega$ associé à une expérience aléatoire, sur lequel on a défini une mesure de probabilité. Une* variable aléatoire *$X$ est une application de l'ensemble des événements élémentaires de l'univers $\Omega$ vers $\mathbf{R}$ (vérifiant quelques conditions mathématiques non explicitées ici).*

Une variable aléatoire est une variable (en fait une fonction !) qui associe des valeurs numériques à des événements aléatoires.

Par convention, une variable aléatoire sera représentée par une lettre majuscule $X$ alors que les valeurs particulières qu'elle peut prendre seront désignées par des lettres minuscules $x_1, x_2, \ldots, x_i, \ldots, x_n$.

Les deux variables aléatoires définies dans les exemples 33 et 34 sont de natures différentes. La première est discrète, la seconde continue.

## 2.2 Variables aléatoires discrètes

**Définition 36** *Une* variable aléatoire discrète *est une variable aléatoire qui ne prend que des valeurs entières, en nombre fini ou dénombrable.*

Pour apprécier pleinement une variable aléatoire, il est important de connaître quelles valeurs reviennent le plus fréquemment et quelles sont celles qui apparaissent plus rarement. Plus précisément, on cherche les probabilités associées aux différentes valeurs de la variable

**Définition 37** *Associer à chacune des valeurs possibles de la variable aléatoire la probabilité qui lui correspond, c'est définir la* loi de probabilité *ou la* distribution de probabilité *de la variable aléatoire.*

*Pour calculer la probabilité que la variable $X$ soit égale à $x$, valeur possible pour $X$, on cherche tous les événements élémentaires $e_i$ pour lesquels $X(e_i) = x$, et on a*

$$p(X = x) = \sum_{i=1}^{k} p(\{e_i\}),$$

*si $X = x$ sur les événements élémentaires $e_1, e_2, \ldots, e_k$.*

La *fonction de densité discrète $f$* est la fonction de $\mathbf{R}$ dans $[0, 1]$, qui à tout nombre réel $x_i$ associe $f(x_i) = p(X = x_i)$. On a bien sûr $\sum_i f(x_i) = 1$.

**Exemple 38** Cas de l'exemple 33.

*La variable $X$ = nombre de côtés "face" peut prendre les valeurs 0,1,2.*

$$\begin{aligned}
f(0) &= p(X = 0) = p((pile, pile)) = \frac{1}{4};\\
f(1) &= p(X = 1) = p((pile, face)) + p((face, pile)) = \frac{1}{2};\\
f(2) &= p(X = 2) = p((face, face)) = \frac{1}{4};\\
f(x) &= 0 \text{ si } x \notin \{0, 1, 2\}.
\end{aligned}$$

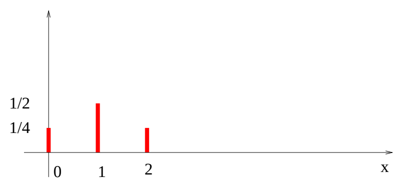
On présente sa distribution de probabilité dans un tableau.

| $x$ | 0 | 1 | 2 | total |
|---|---|---|---|---|
| $f(x) = p(X = x)$ | 1/4 | 1/2 | 1/4 | 1 |

### 2.2.1 Représentation graphique de la distribution de probabilité

Elle s'effectue à l'aide d'un diagramme en bâtons où l'on porte en abscisses les valeurs prises par la variable aléatoire et en ordonnées les valeurs des probabilités correspondantes.

Dans l'exemple du jet de pièces :




### 2.2.2 Fonction de répartition

En statistique descriptive, on a introduit la notion de fréquences cumulées croissantes. Son équivalent dans la théorie des probabilités est la fonction de répartition.

**Définition 39** *La* fonction de répartition *d'une variable aléatoire $X$ indique pour chaque valeur réelle $x$ la probabilité que $X$ prenne une valeur au plus égale à $x$. C'est la somme des probabilités des valeurs de $X$ jusqu'à $x$. On la note $F$.*

$$\forall x \in \mathbf{R}, \quad F(x) = p(X \leq x) = \sum_{x_i \leq x} p(X = x_i).$$

La fonction de répartition est toujours croissante, comprise entre 0 et 1 et se révélera un instrument très utile dans les travaux théoriques.

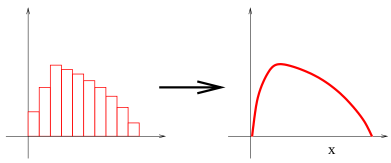
## 2.3 Variables aléatoires continues

**Définition 40** *Une variable aléatoire est dite* continue *si elle peut prendre toutes les valeurs d'un intervalle fini ou infini.*

### 2.3.1 Fonction de densité de probabilité

Dans le cours de statistique descriptive, nous avons appris à représenter la distribution d'une variable statistique continue (ou à caractère continu) à l'aide d'un histogramme de fréquences, qui est une série de rectangles. L'aire de chaque rectangle est proportionnelle à la fréquence de la classe qui sert de base au rectangle.

Si l'on augmentait indéfiniment le nombre d'observations en réduisant graduellement l'intervalle de classe jusqu'à ce qu'il soit très petit, les rectangles correspondant aux résultats vont se multiplier tout en devenant plus étroits et à la limite vont tendre à se fondre en une surface unique limitée d'une part par l'axe des abscisses et d'autre part par une courbe continue.




On abandonne alors la notion de valeur individuelle et l'on dit que la distribution de probabilité est continue. La courbe des fréquences relatives idéalisée est alors la courbe représentative d'une fonction de densité de probabilité $f$.

Pour une variable statistique continue, l'aire des rectangles était un témoin fidèle de la fréquence de chaque classe. Il en va en de même pour une variable aléatoire continue mais il faudra raisonner à présent sur des classes infiniment petites d'amplitude $dx$. L'élément infinitésimal d'aire $f(x)\,dx$ représente la probabilité que $X$ appartienne à un intervalle d'amplitude $dx$,

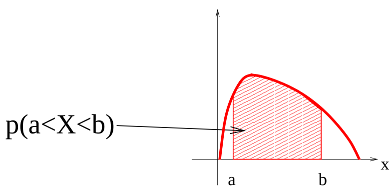
$$p(x < X \leq x + dx) = f(x)\,dx.$$

$f$ a donc les propriétés suivantes :

a) La courbe d'une fonction de densité de probabilité est toujours située au dessus de l'axe des abscisses donc $f$ est une fonction toujours positive.

b) La probabilité que la variable aléatoire $X$ soit comprise entre les limites $a$ et $b$ c'est-à-dire $p(a \leq X \leq b)$, est égale à l'aire entre l'axe des abscisses, la courbe représentative de la fonction de densité de probabilité et les droites d'équations $x = a$ et $x = b$,

$$p(a < X \leq b) = \int_a^b f(x)\,dx$$




c) L'aire totale comprise entre la courbe et l'axe des abscisses est égale à 1 :

$$\int_{\mathbf{R}} f(x)\,dx = 1.$$

d) Alors qu'une probabilité est sans dimension, une densité de probabilité a pour dimension celle de l'inverse de $X$ : $[X^{-1}]$.

e) Il résulte de a et c qu'une densité de probabilité est une fonction *intégrable* au sens de Lebesgue sur $\mathbf{R}$.

**Remarque 41** *Le cas où on a une courbe continue est un cas théorique qui supposerait :*
- *un nombre infini de mesures de la variable statistique*
- *une sensibilité très grande de l'appareil de mesure.*

Nous supposerons toutefois que nous sommes dans ce cas lorsque nous serons en présence d'un grand nombre de mesures.

### 2.3.2 Fonction de répartition

De même que pour les variables aléatoires discrètes, on peut définir la fonction de répartition $F$ de la variable continue $X$ qui permet de connaître la probabilité que $X$ soit inférieure à une valeur donnée :

$$F(x) = p(X \leq x) = \int_{-\infty}^{x} f(t)\,dt.$$

**Propriété 42** *1. $F$ est continue et croissante sur $\mathbf{R}$.*

*2. $\forall x \in \mathbf{R}, \quad F'(x) = f(x)$.*

*3. $\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 0$.*

*4. $p(a \leq X \leq b) = F(b) - F(a)$.*

**Exercice 43**

Soit $f$ la fonction définie sur $\mathbf{R}$ par $f(x) = ke^{-x}$ si $x \geq 0$, $f(x) = 0$ sinon.

1. Déterminer $k$ pour que $f$ soit la fonction de densité de probabilité d'une variable aléatoire $X$.
2. Déterminer la fonction de répartition de la variable $X$.
3. Calculer $p(1 < X < 2)$.

**Solution de l'exercice 43.** *Densité de probabilité.*

1. $f$ doit être une fonction positive, donc il nous faut impérativement trouver pour $k$ une valeur positive. Une fonction de densité de probabilité doit vérifier $\int_{\mathbf{R}} f(x)\,dx = 1$, donc $\int_0^{+\infty} ke^{-x}\,dx = 1$. Il en résulte que $k = 1$.
2. Par définition la fonction de répartition de $X$ est la fonction $F$ définie par

$$\begin{aligned} F(x) &= \int_0^x e^{-t}\,dt = 1 - e^{-x} \text{ si } x > 0, \\ &= 0 \text{ sinon.} \end{aligned}$$

3. 

$$p(1 < X < 2) = \int_1^2 e^{-x}\,dx = e^{-1} - e^{-2} \sim 0.23.$$

## 2.4 Couples de variables aléatoires

Il existe beaucoup de situations où l'intérêt se porte sur la réalisation conjointe d'événements associés à deux (ou plusieurs) variables aléatoires. Nous allons envisager deux cas : celui où les variables sont discrètes et celui où elles sont continues.

### 2.4.1 Couples de variables aléatoires discrètes

**Loi de probabilité conjointe**

Considérons deux variables aléatoires discrètes $X$ et $Y$. Il nous faut pour modéliser le problème une fonction qui nous donne la probabilité que $(X = x_i)$ en même temps que $(Y = y_j)$. C'est la loi de probabilité conjointe.

**Définition 44** *Soient $X$ et $Y$ deux variables aléatoires discrètes dont l'ensemble des valeurs possibles sont respectivement $\{x_1, x_2, \ldots, x_n\}$ et $\{y_1, y_2, \ldots, y_m\}$. Associer à chacune des valeurs possibles $(x_i, y_j)$ du couple $(X, Y)$, la probabilité $f(x_i, y_j)$, c'est définir la* loi de probabilité conjointe *ou* fonction de densité conjointe *des variables aléatoires $X$ et $Y$,*

$$f(x_i, y_j) = p((X = x_i) \text{ et } (Y = y_j)).$$

Le couple $(X, Y)$ s'appelle variable aléatoire à deux dimensions et peut prendre $m \times n$ valeurs.

**Proposition 45** *1. Pour tout $i = 1, 2, \ldots, n$ et $j = 1, 2, \ldots, m$, $0 \leq f(x_i, y_j) \leq 1$.*

*2. $\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} f(x_i, y_j) = 1$.*

On peut représenter graphiquement f sous forme d'un diagramme en bâtons en trois dimensions.

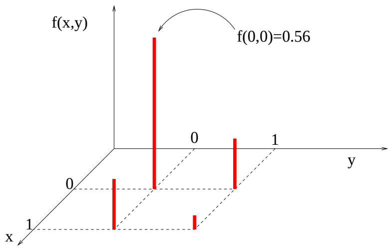
**Exemple 46** *Soit $X$ le nombre de piques obtenus lors du tirage d'une carte dans un jeu ordinaire de 52 cartes et $Y$ le nombre de piques obtenus dans un deuxième tirage, la première carte n'étant pas remise. $X$ et $Y$ ne prennent que les valeurs 0 (pas de pique) ou 1 (un pique).*

Détermination de la loi du couple $(X, Y)$.

$$\begin{aligned} f(0,0) &= \frac{39}{52} \times \frac{38}{51} = 0.56, \quad f(1,0) = \frac{13}{52} \times \frac{39}{51} = 0.19, \\ f(0,1) &= \frac{39}{52} \times \frac{13}{51} = 0.19, \quad f(1,1) = \frac{39}{52} \times \frac{12}{51} = 0.06. \end{aligned}$$

On vérifie que la somme de ces valeurs est égale à 1.

On représente $f$ sous forme d'un diagramme en bâtons en trois dimensions.




**Loi de probabilité marginale**

Lorsqu'on connaît la loi conjointe des variables aléatoires $X$ et $Y$, on peut aussi s'intéresser à la loi de probabilité de $X$ seule et de $Y$ seule. Ce sont les lois de probabilité marginales.

**Définition 47** *Soit $(X, Y)$ une variable aléatoire à deux dimensions admettant comme loi de probabilité conjointe $f(x, y)$. Alors, les lois de probabilité marginales de $X$ et $Y$ sont définies respectivement par*

$$\begin{aligned} f_X(x_i) &= p(X = x_i) = \sum_{j=1}^{m} f(x_i, y_j) \text{ pour } i = 1, 2, \ldots, n, \\ f_Y(y_j) &= p(Y = y_i) = \sum_{i=1}^{n} f(x_i, y_j) \text{ pour } j = 1, 2, \ldots, m. \end{aligned}$$

Si la loi de probabilité conjointe du couple $(X, Y)$ est présentée dans un tableau à double entrée, nous obtiendrons la loi de probabilité marginale $f_X$ de $X$ en sommant les $f(x_i, y_j)$, suivant l'indice $j$ (par colonnes) et celle de $Y$, $f_Y$, en sommant les $f(x_i, y_j)$ suivant l'indice $i$ (par lignes).

**Exemple 48** *Lois marginales de l'exemple 46.*

| | $Y = y_1 = 0$ | $Y = y_2 = 1$ | $f_X(x_i)$ |
|---|---|---|---|
| $X = x_1 = 0$ | 0.56 | 0.19 | 0.75 |
| $X = x_2 = 1$ | 0.19 | 0.06 | 0.25 |
| $f_Y(y_j)$ | 0.75 | 0.25 | 1.00 |

**Remarque 49** *Le couple $(X, Y)$ et les deux variables $X$ et $Y$ constituent trois variables aléatoires distinctes. La première est à deux dimensions, les deux autres à une dimension.*

### Loi de probabilité conditionnelle

Nous avons vu dans le paragraphe précédent comment déterminer la probabilité de réalisation de deux événements lorsqu'ils sont dépendants l'un de l'autre. Pour cela, nous avons introduit la notion de probabilité conditionnelle en posant

$$p(B|A) = \frac{p(B \cap A)}{p(A)}.$$

La notion équivalente dans le cas d'un couple de variables aléatoires est celle de loi de probabilité conditionnelle permettant de mesurer la probabilité que $X$ soit égale à une valeur donnée lorsqu'on connaît déjà la valeur que prend $Y$.

**Définition 50** *Soit la variable aléatoire $(X, Y)$ à deux dimensions admettant comme loi conjointe $f(x, y)$ et comme lois marginales $f_X(x)$ et $f_Y(y)$. Si l'on suppose que la probabilité que $X$ prenne la valeur $x_i$ n'est pas nulle, alors la probabilité conditionnelle de $(Y = y_j)$ sachant que $(X = x_i)$ s'est réalisé est définie par*

$$f(y_j|x_i) = \frac{f(x_i, y_j)}{f_X(x_i)}.$$

*Les probabilités $f(y_j|x_i)$ associées aux différentes valeurs possibles $y_j$ de $Y$ constituent la loi de probabilité conditionnelle de $Y$.*

De même, si l'on suppose que la probabilité que $Y$ prenne la valeur $y_j$ n'est pas nulle, alors la probabilité conditionnelle de $(X = x_i)$ sachant que $(Y = y_j)$ s'est réalisé est définie par

$$f(x_i|y_j) = \frac{f(x_i, y_j)}{f_Y(y_j)}.$$

Les probabilités $f(x_i|y_j)$ associées aux différentes valeurs possibles $x_i$ de $X$ constituent la loi de probabilité conditionnelle de $X$.

### Cas de variables aléatoires indépendantes

Lorsque deux variables aléatoires $X$ et $Y$ sont indépendantes, la loi conditionnelle de $X$, pour toute valeur de $Y$, est identique à la loi marginale de $X$ et lorsque la loi conditionnelle de $Y$, pour toute valeur de $X$, est identique à la loi marginale de $Y$. Autrement dit,

**Proposition 51** *Soit $(X, Y)$ une variable aléatoire à deux dimensions admettant comme loi de probabilité conjointe la fonction $f(x, y)$ et comme lois de probabilité marginales $f_X(x)$ et $f_Y(y)$. Les variables aléatoires $X$ et $Y$ sont indépendantes si et seulement si les probabilités conjointes sont égales au produit des probabilités marginales, $f(x_i, y_j) = f_X(x_i) \times f_Y(y_j)$ pour toutes les valeurs $(x_i, y_j)$.*

Pour conclure que deux variables ne sont pas indépendantes, il suffit de trouver une valeur du couple $(X, Y)$ pour laquelle la relation précédente n'est pas satisfaite.

**Exemple 52** *Dans l'exemple du tirage de cartes, nous savons, par exemple que* $p((X = 0)$ *et* $(Y = 1)) = f(0, 1) = 0.19$*, alors que* $p(X = 0) \times p(Y = 1) = f_X(0) \times f_Y(1) = 0.75 \times 0.25 = 0.188$.

Conclusion : les variables $X$ et $Y$ sont dépendantes, ce qui paraît cohérent étant donné que le tirage était effectué sans remise.

### 2.4.2 Couple de variables aléatoires continues

Dans le cas de deux variables continues $X$ et $Y$, le couple $(X, Y)$ est dit continu.

#### Fonction de densité de probabilité conjointe

La distribution de probabilité conjointe de $X$ et de $Y$ est décrite par une *fonction de densité de probabilité conjointe* $f(x, y)$ définie pour chaque valeur $(x, y)$ du couple $(X, Y)$. La fonction $f$ détermine une surface au-dessus de l'ensemble des valeurs $(x, y)$.

On a $p((X, Y) \in D) =$ volume sous la surface représentative de $f$ et au-dessus du domaine $D$ du plan $(xOy)$.

Dans le cas où $D$ est un rectangle $[c, d] \times [u, v]$,

$$p((c < X \leq d) \text{ et } (u < Y \leq v)) = \int_{[c,d]\times[u,v]} f(x, y)\, dx\, dy.$$

**Propriété 53**
1. *Pour tout couple* $(x, y) \in \mathbf{R}^2$*,* $f(x, y) \geq 0$.
2. $\int_{\mathbf{R}^2} f(x, y)\, dx\, dy = 1$.
3. *Il en résulte qu'une densité de probabilité conjointe est une fonction* intégrable *au sens de Lebesgue sur* $\mathbf{R}^2$.

#### Densité de probabilité marginale

De même que pour les couples de variables aléatoires discrètes, on définit les fonctions densités de probabilité marginales et conditionnelles. Pour les définir dans le cas continu, il suffit de remplacer les sommes du cas discret par des intégrales.

**Définition 54** *Soit* $(X, Y)$ *une variable aléatoire continue à deux dimensions admettant comme densité de probabilité conjointe la fonction* $f(x, y)$*. Alors, les* densités de probabilité marginales *de* $X$ *et* $Y$ *sont définies respectivement par*

$$\begin{aligned} f_X(x) &= \int_{\mathbf{R}} f(x, y)\, dy \text{ pour } x \in \mathbf{R}, \\ f_Y(y) &= \int_{\mathbf{R}} f(x, y)\, dx \text{ pour } y \in \mathbf{R}. \end{aligned}$$

#### Variables indépendantes

**Proposition 55** *Deux variables continues* $X$ *et* $Y$ *sont indépendantes si et seulement si la fonction de densité de probabilité conjointe est égale au produit des fonctions de densité marginales.*

Autrement dit, pour tout couple $(x, y) \in \mathbf{R}^2$,

$$f(x, y) = f_X(x) f_Y(y).$$

Dans ce cas, le théorème de Fubini-Tonelli donne

$$\begin{aligned} p((c < X \leq d) \text{ et } (u < Y \leq v)) &= \int_{[c,d]\times[u,v]} f(x,y)\,dx\,dy \\ &= \left(\int_c^d f_X(x)\,dx\right)\left(\int_u^v f_Y(y)\,dy\right). \end{aligned}$$

## 2.5 Paramètres descriptifs d'une distribution

En statistique descriptive, nous avons caractérisé les distributions statistiques des valeurs observées par certains nombres représentatifs qui résumaient de façon commode et assez complète la distribution. Pour apprécier la tendance centrale d'une série d'observations, nous avons employé, entre autres, la moyenne arithmétique et pour caractériser la dispersion de la série autour de la moyenne, nous avons introduit la variance ou l'écart quadratique moyen.

### 2.5.1 Espérance mathématique d'une distribution de probabilité

Si l'on s'imagine que le nombre d'observations croît indéfiniment (on passe d'un échantillon de taille n à la population toute entière), les fréquences observées vont tendre vers les probabilités théoriques et on admet que la moyenne calculée sur l'échantillon de taille $n$ va tendre vers une valeur limite qui sera la moyenne de l'ensemble des valeurs de la population entière. On l'appelle espérance mathématique de la variable aléatoire $X$, car c'est la valeur moyenne que l'on s'attend à avoir dans un échantillon de grande taille.

**Définition 56** *1. Cas d'une variable discrète*

– *Soit $X$ une variable aléatoire discrète qui prend un nombre fini de valeurs $x_1, x_2, \ldots, x_n$ et dont la loi de probabilité est $f : f(x_i) = p(X = x_i)$. L'*espérance mathématique *de $X$, notée $E(X)$, est définie par*

$$E(X) = \sum_{i=1}^{n} x_i f(x_i).$$

– *Si la variable aléatoire $X$ prend un nombre dénombrable de valeurs $x_1, x_2, \ldots, x_n, \ldots$, son* espérance mathématique *est alors définie par $E(X) = \sum_{i=1}^{\infty} x_i f(x_i)$, à condition que la série converge absolument.*

*2. Cas d'une variable continue. Si la variable aléatoire $X$ est continue et a pour fonction de densité de probabilité $f$,* son espérance mathématique *est*

$$E(X) = \int_{\mathbf{R}} x f(x)\,dx,$$

*pourvu que la fonction $x \mapsto x f(x)$ soit intégrable sur $\mathbf{R}$.*

### 2.5.2 Variance d'une distribution de probabilités

En effectuant le même raisonnement que précédemment pour passer d'un échantillon de taille $n$ à la population totale, on suppose que la variance calculée sur l'échantillon tend vers une limite lorsque le nombre d'observations tend vers l'infini. Cette limite est appelée variance de la variable aléatoire $X$.

**Définition 57** – *On appelle* variance *de la variable aléatoire $X$ la valeur moyenne des carrés des écarts à la moyenne,*

$$Var(X) = E\left((X - E(X))^2\right).$$

*Le calcul de la variance se simplifie en utilisant l'expression :*

$$Var(X) = E(X^2) - E(X)^2.$$

– *On appelle* écart-type *de la variable aléatoire $X$ la racine carrée de sa variance.*

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

*Dans le cas d'une variable aléatoire continue,*

Dans le cas d'une variable aléatoire discrète finie,

$$Var(X) = \sum_{i=1}^{n}(x_i - E(X))^2 f(x_i) = \left(\sum_{i=1}^{n} x_i^2 f(x_i)\right) - E(X)^2.$$

Dans le cas d'une variable aléatoire continue,

$$Var(X) = \int_{\mathbf{R}} (x - E(X))^2 f(x)\, dx = \left(\int_{\mathbf{R}} x^2 f(x)\, dx\right) - E(X)^2.$$

### 2.5.3 Fonction caractéristique d'une distribution de probabilité

**Définition 58** *On appelle* fonction caractéristique *de la variable aléatoire $X$ la fonction $\xi_X$ définie sur* $\mathbf{R}$ *par*

$$\xi_X(u) = E(e^{-2i\pi uX}).$$

Dans le cas d'une variable aléatoire discrète finie,

$$\xi_X(u) = \sum_{i=1}^{n} f(x_i) e^{-2i\pi u x_i}.$$

Dans le cas d'une variable aléatoire continue,

$$\xi_X(u) = \int_{\mathbf{R}} e^{-2i\pi ux} f(x)\, dx.$$

Dans le cas continu, on constate que $\xi_X = Ff$ est la transformée de Fourier de la densité de probabilité. Elle existe donc toujours puisque la densité de probabilité est intégrable au sens de Lebesgue.

**Proposition 59** *Soit $X$ une variable aléatoire qui possède une espérance et une variance. Alors*

$$\begin{aligned} E(X) &= -\frac{1}{2i\pi}\frac{d}{du}\xi_X(u)_{|u=0}, \\ Var(X) &= -\frac{1}{4\pi^2}\left(\frac{d^2}{du^2}\xi_X(u)_{|u=0} - (\frac{d}{du}\xi_X(u)_{|u=0})^2\right). \end{aligned}$$

**Preuve.** Dans le cas discret,

$$\begin{aligned} \frac{d}{du}\xi_X(u) &= \sum_{i=1}^{n} -2i\pi x_i f(x_i) e^{-2i\pi u x_i} = -2i\pi E(X e^{-2i\pi uX}), \\ \frac{d^2}{du^2}\xi_X(u) &= \sum_{i=1}^{n} -4\pi^2 x_i^2 f(x_i) e^{-2i\pi u x_i} = -4\pi^2 E(X^2 e^{-2i\pi uX}), \end{aligned}$$

d'où

$$\begin{aligned}\xi''_X(0) - (\xi'_X(0))^2 &= -4\pi^2 E(X^2) + 4\pi^2 E(X)^2 \\ &= -4\pi^2 (E(X^2) - E(X)^2) \\ &= -4\pi^2 Var(X).\end{aligned}$$

Dans le cas continu, le calcul est le même, au moyen de la formule pour la dérivée de la transformée de Fourier.■

**Remarque 60** *Une loi de probabilité régit le comportement d'une variable aléatoire. Cette notion abstraite est associée à la population, c'est-à-dire à l'ensemble de tous les résultats possibles d'un phénomène particulier. C'est pour cette raison que l'espérance et la variance de la loi de probabilité, qui n'ont aucun caractère aléatoire, sont appelés paramètres de la distribution de probabilité.*

### 2.5.4 Propriétés de l'espérance mathématique et de la variance

Résumons les principales propriétés de ces deux paramètres dans un tableau.

| Changement d'origine | Changement d'échelle | Transformation affine |
|---|---|---|
| $E(X+c) = E(X) + c$ | $E(aX) = aE(X)$ | $E(aX+c) = aE(X) + c$ |
| $Var(X+c) = Var(X)$ | $Var(aX) = a^2 Var(X)$ | $Var(aX+c) = a^2 Var(X)$ |
| $\sigma(X+c) = \sigma(X)$ | $\sigma(aX) = \|a\|\sigma(X)$ | $\sigma(aX+c) = \|a\|\sigma(X)$ |
| $\xi_{X+c}(u) = e^{-2i\pi cu}\xi_X(u)$ | $\xi_{aX}(u) = \xi_X(au)$ | $\xi_{aX+c}(u) = e^{-2i\pi cu}\xi_X(au)$ |

**Définition 61**
- *Une variable aléatoire $X$ est dite* centrée *si son espérance mathématique est nulle.*
- *Une variable aléatoire $X$ est dite* réduite *si son écart-type est égal à 1.*
- *Une variable aléatoire centrée réduite est dite* standardisée.

A n'importe quelle variable aléatoire X, on peut associer la variable standardisée

$$Z = \frac{X - E(X)}{\sigma(X)}.$$

En divisant la variable centrée par son écart-type, une valeur située à un écart-type de la moyenne sera ramenée à 1, une autre située à deux écarts-types sera ramenée à 2 : l'échelle de référence, ou unité de mesure, d'une variable centrée-réduite est l'écart-type.

Les valeurs des variables centrées-réduites sont complètement indépendantes des unités de départ. Une mesure exprimée en mètres ou en centimètres donne exactement la même variable centrée-réduite. On peut ainsi faire des comparaisons entre variables de natures différentes. Si un enfant est à $+3$ écarts-types de la moyenne pour sa taille et $+1$ écart-type pour son poids, on sait qu'il est plus remarquable par sa taille que par son poids.

L'examen des variables centrées-réduites est très pratique pour déceler les valeurs "anormalement" grandes ou "anormalement" petites.

Le passage d'une variable aléatoire X à une variable standardisée est requis pour l'utilisation de certaines tables de probabilité. C'est le cas pour l'utilisation de la table de la loi normale que nous traiterons dans le prochain chapitre.

#### Combinaisons de plusieurs variables aléatoires

1. Somme et différence.

   Dans tous les cas,

$$\begin{aligned}E(X+Y) &= E(X) + E(Y), \\ E(X-Y) &= E(X) - E(Y).\end{aligned}$$

Dans le cas de variables *indépendantes* :

$$\begin{aligned} Var(X+Y) &= Var(X)+Var(Y), \\ Var(X-Y) &= Var(X)+Var(Y). \end{aligned}$$

2. Produit. Dans le cas de variables *indépendantes*,

$$E(XY) = E(X)E(Y).$$

3. Conséquence. Dans le cas de variables *indépendantes*,

$$\xi_{X+Y} = \xi_X \xi_Y.$$

Ceci montre que, dans le cas de variables continues indépendantes, la densité de probabilité de $X+Y$ est le *produit de convolution* des densités de probabilité de $X$ et de $Y$.

### Covariance de deux variables aléatoires

Lorsque deux variables aléatoires ne sont pas indépendantes, il existe une caractéristique qui permet de déterminer l'intensité de leur dépendance. C'est la covariance.

**Définition 62** *La covariance de deux variables aléatoires $X$ et $Y$ est définie par*

$$Cov(X,Y) = E[(X - E(X)(Y - E(Y)] = E(XY) - E(X)E(Y).$$

**Proposition 63** *1. Si deux variables aléatoires sont indépendantes, leur covariance est nulle.*

*2. Attention : La réciproque n'est pas vraie. Deux variables de covariance nulle ne sont pas obligatoirement indépendantes.*

*3. Si deux variables aléatoires sont dépendantes,*

$$\begin{aligned} E(XY) &= E(X)E(Y) + Cov(X,Y), \\ Var(X+Y) &= Var(X)+Var(Y)+2Cov(X,Y) \end{aligned}$$