

Probabilités et statistique

Université Paris 1 Panthéon-Sorbonne

Cours de deuxième année de licence de sciences économiques

FABRICE ROSSI

Cette œuvre est mise à disposition selon les termes de la [licence Creative Commons Paternité - Partage à l'Identique 3.0 non transposé](#).

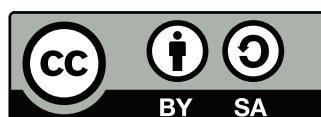


Table des matières

Table des matières	iii
1 Expérience aléatoire et probabilités	1
1.1 Expérience aléatoire	1
1.2 Évènements	4
1.3 Vocabulaire probabiliste	6
1.4 Probabilité	7
1.5 Probabilités sur un univers fini	12
2 Probabilités conditionnelles	21
2.1 Évènement réalisé	21
2.2 Probabilité conditionnelle	21
2.3 Expériences aléatoires composées	24
2.4 Règle des probabilités totales	26
2.5 Règle de Bayes	28
2.6 Indépendance	30
2.7 Indépendance conditionnelle	33
3 Variables aléatoires	37
3.1 Introduction	37
3.2 Notions générales	37
3.3 Variable aléatoire numérique	42
3.4 Variable aléatoire fonction d'une autre variable aléatoire	49
4 Variables aléatoires discrètes	53
4.1 Définition	53
4.2 Entropie et mode	54
4.3 Variable aléatoire discrète numérique	56
4.4 Moments	57
4.5 Lois classiques	64
A Théorie des ensembles	69
A.1 Notations et opérations	69
A.2 Propriétés des opérations ensemblistes	72
A.3 Produit cartésien	74
A.4 Partition	74

B Fonctions	77
B.1 Définition	77
B.2 Cas particuliers	77
B.3 Composition de fonctions	78
B.4 Fonction réciproque	78
C Dénombrement	81
C.1 Ensembles finis et ensembles dénombrables	81
C.2 Cardinaux et opérations ensemblistes	81
C.3 Listes d'éléments	82
C.4 Sous-ensembles	83
Évolutions de ce document	85
Index	87

Chapitre 1

Expérience aléatoire et probabilités

1.1 Expérience aléatoire

Définition 1.1 On appelle *expérience aléatoire* ou *épreuve aléatoire* une expérience dont on connaît l'ensemble des résultats possibles mais dont on ne peut prévoir le résultat effectif avec certitude.

☞ **Exemple 1.1** Le lancer d'un dé est une expérience aléatoire : le résultat est un entier compris entre 1 et 6 dont la valeur ne peut être connue avant le lancer. ☞

Définition 1.2 L'ensemble des résultats possibles pour une expérience aléatoire, généralement noté Ω , est appelé l'**univers** de l'expérience, aussi connu sous le nom d'**ensemble fondamental** ou d'**espace des possibles** (ou encore l'**espace des états**).

☞ **Exemple 1.2** L'univers de l'expérience du lancer de dé de l'exemple 1.1 est l'ensemble

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

On peut aussi avoir un point de vue plus proche des objets physiques et représente l'univers comme l'ensemble des faces obtenues, soit

$$\Omega = \{\square, \square, \square, \square, \square, \square\}. \quad \text{☞}$$

☞ **Exemple 1.3** Le lancer d'une pièce de monnaie est une expérience aléatoire dont l'univers est l'ensemble $\Omega = \{\text{pile}, \text{face}\}$. ☞

L'univers d'une expérience aléatoire étant un ensemble, sa description s'appuie sur la théorie des ensembles dont les éléments nécessaires pour ce cours sont rappelés dans l'annexe A. On utilise en particulier la notion de produit cartésien, comme l'illustrent les exemples suivants.

☞ **Exemple 1.4** On choisit une carte à jouer au hasard dans un jeu de 32 cartes. Une façon basique de décrire l'univers consiste à numéroter les cartes de 1 à 32 dans un ordre arbitraire, puis à utiliser $\Omega = \{1, 2, \dots, 32\}$.

On peut aussi considérer l'ensemble des couleurs $C = \{\clubsuit, \spadesuit, \heartsuit, \diamondsuit\}$ et l'ensemble des valeurs et des figures $VF = \{7, 8, 9, 10, Valet, Dame, Roi, As\}$. Ω est alors défini comme le produit cartésien entre VF et C , soit $\Omega = VF \times C$. Une carte est ainsi représentée par un couple de la forme $(8, \clubsuit)$ pour le 8 de trèfle, par exemple.

Notons que l'ordre choisi $VF \times C$ est arbitraire et qu'on pourrait utiliser $\Omega = C \times VF$ sans que cela n'influence les résultats. Une carte serait alors de la forme $(\heartsuit, 4)$ pour le 4 de cœur, par exemple. ☞

☞ **Exemple 1.5** On lance deux fois de suite un dé. Comme pour un seul dé, l'espace des possibles associé à un lancer peut être proche des objets physiques eux-mêmes, soit par exemple $\mathcal{L} = \{\square, \square, \square, \square, \square, \square\}$. On peut aussi considérer seulement la valeur du dé et donc prendre $\mathcal{L} = \{1, 2, 3, 4, 5, 6\}$. Comme on réalise deux lancers, l'expérience produit un couple de résultats, c'est-à-dire un élément du produit cartésien $\mathcal{L} \times \mathcal{L}$. L'univers de l'expérience est donc $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ pour la version numérique, ou $\Omega = \{\square, \square, \square, \square, \square, \square\}^2$ pour une version plus imagée.

Notons que dans cette expérience, l'ordre dans un couple de résultats est significatif puisqu'on lance d'abord un dé, puis après avoir observé son résultat, on le relance. On a donc bien un premier résultat, puis un second. De ce fait la paire (\square, \square) est bien différente de la paire (\square, \square) . ☞

La nature de l'expérience peut conduire à des univers relativement complexes, comme le montrent les exemples suivants.

☞ **Exemple 1.6** Considérons une urne contenant 3 jetons numérotés de 1 à 3. On tire au hasard un premier jeton puis un second jeton. Le résultat de l'expérience est donc une paire (ordonnée) d'éléments choisis dans $J = \{1, 2, 3\}$. L'univers n'est pas cependant J^2 . En effet, si on tire d'abord le jeton 1, par exemple, celui n'est plus disponible dans l'urne. Les résultats de l'expérience sont donc des paires d'éléments distincts. L'univers est alors

$$\begin{aligned}\Omega &= \{(a, b) \in J^2 \mid a \neq b\}, \\ &= \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}.\end{aligned}$$

Cet ensemble ne peut pas se formuler plus simplement sous forme d'un produit cartésien, par exemple. ☞

Une des difficultés de modélisation est de décider si les résultats d'une expérience sont discernables. Les exemples suivants illustrent cette difficulté.

☞ **Exemple 1.7** On lance deux dés **simultanément**. Si on considère les deux dés comme **indiscernables**, on ne doit pas faire de différence entre les paires (\square, \square) et (\square, \square) : rien ne permet d'ordonner les dés (contrairement à la situation de l'exemple 1.5). Pour représenter

l'univers, on introduit l'ensemble des résultats d'un seul dé, $R = \{\square, \square, \square, \square, \square, \square\}$. On peut alors représenter l'univers sous la forme suivante :

$$U = \{r \in \mathcal{P}(R) \mid 1 \leq |r| \leq 2\}.$$

Dans cette définition, les éléments de U sont des sous-ensembles de R (d'où le $r \in \mathcal{P}(R)$) contraints à contenir au plus deux éléments (d'où la condition $1 \leq |r| \leq 2$). En d'autres termes, les éléments de U s'écrivent soit $\{a, b\}$ avec $a \in R$ et $b \in R$, soit $\{c\}$ avec $c \in R$. Ceci s'explique par le fait que par nature, un ensemble n'est pas ordonné et donc que $\{a, b\} = \{b, a\}$, ce qui semble bien adapté pour l'expérience considérée. Cependant, un ensemble ne contient qu'une seule fois chacun de ses éléments. De ce fait, l'ensemble $\{\square, \square\}$ correspond en réalité à l'ensemble $\{\square\}$, ce qui explique la nécessité d'inclure des sous-ensembles de R réduits à un unique élément. Techniquement, ce détail ne change rien à la qualité de la modélisation : les seuls singletons de U correspondent aux lancers pour lesquels on obtient deux fois le même chiffre.

Il est parfois cependant plus clair de supposer les objets étudiés comme discernables et donc de considérer comme dans l'exemple 1.5 des n -uplets plutôt que des sous-ensembles de n éléments. En considérant les dés comme discernables, on obtient ici le même univers que dans l'exemple 1.5, c'est-à-dire :

$$\Omega = \{\square, \square, \square, \square, \square, \square\} \times \{\square, \square, \square, \square, \square, \square\}.$$

Une façon de visualiser le caractère discernable des dés est de faire comme s'ils avaient chacun une couleur différente. Avec un dé noir et un dé blanc, ceci donne par exemple :

$$\Omega = \{\blacksquare, \blacksquare, \blacksquare, \blacksquare, \blacksquare, \blacksquare\} \times \{\square, \square, \square, \square, \square, \square\}.$$

Ici, le fait que les dés soient lancés simultanément n'empêche pas de donner les résultats dans un ordre fixé, par exemple le dé noir en premier comme ci-dessus.

Nous verrons notamment dans l'exemple 1.17 que le modèle discernable est souvent plus simple à utiliser que le modèle indiscernable. ∞

∞ **Exemple 1.8** On considère une urne contenant cinq jetons, trois rouges et deux bleus. On tire au hasard un jeton dans l'urne. Comme dans l'exemple 1.7, on peut être tenté de modéliser de façon naturelle l'expérience en supposant que les jetons ne sont discernables que par leur couleur. L'univers naturel est alors

$$U = \{\bullet, \bullet\}.$$

Cet univers respecte une symétrie des couleurs qui n'est pas très satisfaisante intuitivement : on se doute que si on prend un jeton au hasard dans l'urne, on tombera plus fréquemment sur un rouge que sur un bleu. Pour représenter cette intuition au niveau de l'univers, on rend les jetons complètement discernables en les numérotant de 1 à 5. Comme dans l'exemple 1.7, ceci est artificiel mais est très utile pour faciliter l'analyse de l'expérience. L'univers devient alors par exemple

$$\Omega = \{\color{red}{1}, \color{red}{2}, \color{red}{3}, \color{blue}{4}, \color{blue}{5}\}.$$

La numérotation étant arbitraire, on peut très bien mettre les jetons bleus en premier ou choisir n'importe quel ordre. L'intérêt du modèle est qu'il n'est plus symétrique : on respecte dans l'univers la structure de l'expérience. En ce sens, Ω est un meilleur modèle que U , mais ce dernier reste un modèle exact. ∞

1.2 Évènements

Définition 1.3 Soit une expérience aléatoire et son univers Ω . On appelle **évènement**¹ un sous-ensemble de Ω , c'est-à-dire un sous-ensemble de tous les résultats possibles. Un évènement est donc un élément de $\mathcal{P}(\Omega)$.

On appelle **évènement élémentaire** les singletons de $\mathcal{P}(\Omega)$, c'est-à-dire les ensembles réduits à un seul élément. Un évènement élémentaire est donc de la forme $\{\omega\}$ pour tout $\omega \in \Omega$.

⚠ Remarque 1.1 Dans les situations complexes, en particulier quand Ω n'est pas dénombrable (par exemple $\Omega = \mathbb{R}$), on ne peut pas considérer toute partie de Ω comme un évènement. On se donne alors un sous-ensemble (strict) de $\mathcal{P}(\Omega)$, celui des évènements observables. Ce sous-ensemble est une tribu, c'est-à-dire qu'il vérifie des propriétés de stabilité qu'on ne détaillera pas ici car la notion n'est pas au programme de ce cours.

Mathématiquement, un évènement s'exprime sous forme d'un ensemble. On peut aussi le décrire grâce à une formule logique s'appliquant aux résultats possibles de l'expérience : l'évènement ensemble est alors l'ensemble des résultats qui rendent la formule vraie.

∞ Exemple 1.9 On considère de nouveau l'exemple 1.2 du lancer d'un dé. Voici quelques exemples d'évènements :

- l'évènement $A = \{\square, \boxplus, \boxtimes\}$ correspond à l'obtention d'un chiffre pair ;
- l'évènement $B = \{\square, \boxplus, \boxtimes\}$ correspond à l'obtention d'un chiffre impair ;
- l'évènement $C = \{\square, \square, \square\}$ correspond à l'obtention d'un chiffre plus petit ou égal à trois ;
- l'évènement $D = \{\boxtimes\}$ correspond à l'obtention d'un six. C'est un évènement élémentaire. ∞

∞ Exemple 1.10 Dans l'expérience des deux lancers successifs d'un dé de l'exemple 1.5 :

- l'obtention de deux fois de suite la même valeur est l'évènement

$$A = \{(\square, \square), (\square, \square), (\square, \square), (\boxplus, \boxplus), (\boxplus, \boxplus), (\boxtimes, \boxtimes)\};$$

- l'obtention d'un tirage dont le premier lancer est pair est l'évènement

$$B = \{\square, \boxplus, \boxtimes\} \times \{\square, \square, \square, \square, \square, \boxtimes\}. \quad \infty$$

Une formule logique relativement simple peut conduire à un évènement complexe, comme le montre l'exemple suivant.

∞ Exemple 1.11 On considère l'exemple 1.2 des deux lancers d'un dé et l'évènement « obtenir une somme de 6 en ajoutant le résultat des deux lancers ». Une façon simple d'écrire l'évènement de façon ensembliste est la suivante

$$A = \{(u, v) \in \{\square, \square, \square, \boxplus, \boxplus, \boxtimes\}^2 \mid u + v = 6\}.$$

1. On rencontre encore l'orthographe événement malgré les rectifications orthographiques de 1990...

Bien que parfaitement correcte mathématiquement, cette formulation ne renseigne pas beaucoup sur l'évènement. En outre, elle s'appuie sur l'identité entre le dé et sa valeur numérique. Malheureusement, il n'existe pas d'autre formulation simple et on doit donc se contenter soit de la formule précédente, soit d'une description exhaustive de A (en extension)

$$A = \{(\square, \boxtimes), (\square, \boxplus), (\boxtimes, \square), (\boxplus, \square), (\boxtimes, \boxtimes)\}.$$

L'expérience étant très simple, la description exhaustive n'est pas trop grande. Dans le cas d'une expérience plus complexe, on peut rapidement arriver à des formulations ensemblistes assez lourdes. ∞

La traduction d'une formule logique en évènement dépend de la modélisation effectuée, c'est-à-dire essentiellement de l'univers, comme le montrent les exemples suivants.

∞ **Exemple 1.12** On reprend l'exemple 1.8 et on considère l'évènement « obtenir un jeton rouge ». Dans le modèle indiscernable, l'univers étant réduit à $\Omega_i = \{\bullet, \bullet\}$, l'évènement s'écrit simplement $A_i = \{\bullet\}$. Dans le modèle discernable, l'univers devient $\Omega_d = \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}\}$, ce qui conduit naturellement à une nouvelle description pour l'évènement, à savoir $A_d = \{\mathbf{1}, \mathbf{2}, \mathbf{3}\}$. ∞

∞ **Exemple 1.13** On considère l'expérience du lancer simultané de deux dés (exemple 1.7) et l'évènement « obtenir une somme de 6 ». Si on suppose les dés discernables, on se retrouve exactement dans la situation de l'exemple 1.11. En reprenant l'idée des dés colorés, on peut même formuler A d'une façon un peu différente, soit

$$A = \{(\blacksquare, \boxtimes), (\blacksquare, \boxplus), (\boxtimes, \square), (\boxplus, \square), (\boxtimes, \square)\}.$$

Cette représentation exacerbe les différences entre les deux dés.

Si on suppose au contraire que les deux ne sont pas discernables, l'évènement est assez difficile à écrire en compréhension. En effet, comme nous l'avons vu dans l'exemple 1.7, l'univers U est constitué de sous-ensembles de $\{\square, \square, \boxtimes, \boxtimes, \boxplus, \boxplus\}$ de cardinal 1 ou 2. On ne peut donc pas se contenter de considérer des sous-ensembles de la forme $\{a, b\}$. Une formulation correcte possible est

$$A = \{\{a, b\} \in U \mid a \neq b \text{ et } a + b = 6\} \cup \{\{\boxtimes\}\},$$

ce qui isole le cas du double \boxtimes qui est représenté par l'ensemble $\{\boxtimes\}$. Il est en fait plus simple d'écrire directement A en extension, soit

$$A = \{\{\square, \boxtimes\}, \{\square, \boxplus\}, \{\boxtimes\}\}.$$

Ce résultat est parfaitement correct mathématiquement, mais il peut surprendre, ce qui conduit souvent à préférer la solution dans laquelle on impose aux dés d'être discernables, même s'ils ne le sont pas. Le seul point à bien vérifier dans ce cas est que la modélisation choisie est la même pour l'univers et pour les évènements : supposer par exemple les dés discernables dans l'univers mais indiscernables dans les évènements conduit nécessairement à des erreurs. ∞

1.3 Vocabulaire probabiliste

Bien que l'univers d'une expérience et les événements associés soient décrits grâce aux concepts de la théorie des ensembles, la théorie des probabilités utilise un vocabulaire différent détaillé ci-dessous (ce vocabulaire est en partie partagé avec la logique mathématique).

Vocabulaire probabiliste	Vocabulaire ensembliste	Notation
événement de l'univers Ω	sous-ensemble de Ω	$A \subset \Omega$
événement impossible	ensemble vide	\emptyset
événement certain	ensemble plein Ω	Ω
événement contraire de A	complémentaire de A dans Ω	\bar{A}
A et B sont incompatibles	A et B sont disjoints	$A \cap B = \emptyset$
A implique B	A est inclus dans B	$A \subset B$
A et B	intersection de A et B	$A \cap B$
A ou B	union de A et B	$A \cup B$

Pour bien utiliser les notations ensemblistes et le vocabulaire probabiliste, il faut conserver à l'esprit le sens d'un événement : c'est un sous-ensemble de tous les résultats possibles pour une expérience. Chacun de ces résultats **réalise** l'événement. En ce sens, A et B doit bien se traduire par l'intersection des ensembles A et B : en effet, pour réaliser à la fois A et B , il faut qu'un résultat de l'expérience appartienne aux deux ensembles, ce qui est exactement la définition de $A \cap B$. On raisonne de la même façon pour A ou B et pour A implique B .

La traduction des connecteurs logiques (et, ou, implique) en opérations ensemblistes facilite le passage d'une formule logique à une description ensembliste pour un événement². Le processus est illustré par l'exemple suivant.

♣ **Exemple 1.14** On considère un jeu de 32 cartes (cf l'exemple 1.4) dans lequel on choisit deux cartes successivement sans remise. Si on note P le paquet de cartes, l'univers de l'expérience est

$$\Omega = \{(p_1, p_2) \in P^2 \mid p_1 \neq p_2\}.$$

En effet, le tirage étant successif, on observe bien une paire ordonnée de cartes. De plus, l'absence de remise fait que les cartes sont nécessairement distinctes.

On s'intéresse à l'événement « obtenir au moins un roi ». Cette formulation logique est clairement équivalente à « obtenir exactement un roi ou obtenir exactement deux rois ». L'utilisation du connecteur logique « ou » montre que l'événement est obtenu comme union ensembliste des événements « obtenir exactement un roi » et « obtenir exactement deux rois » (événements qui sont d'ailleurs clairement incompatibles). Le deuxième événement peut être décrit directement comme suit

$$A_{2 \text{ rois}} = \{((\text{Roi}, c_1), (\text{Roi}, c_2)) \mid (c_1, c_2) \in \{\clubsuit, \spadesuit, \heartsuit, \diamondsuit\}^2 \text{ et } c_1 \neq c_2\}.$$

On utilise une deuxième décomposition pour le premier événement en s'appuyant sur le fait qu'« obtenir exactement un roi » est logiquement équivalent à « obtenir exactement

2. On pourra se reporter à l'annexe A pour obtenir une liste de propriétés importantes des opérations ensemblistes.

un roi en premier ou obtenir exactement un roi en deuxième ». De nouveau, on doit donc passer par l'union de deux évènements. On a

$$\begin{aligned} A_{\text{premier roi}} &= \{((\text{Roi}, c_1), (r, c_2)) \mid (c_1, c_2) \in \{\clubsuit, \spadesuit, \heartsuit, \diamondsuit\}^2 \text{ et } r \neq \text{Roi}\}, \\ A_{\text{second roi}} &= \{((r, c_1), (\text{Roi}, c_2)) \mid (c_1, c_2) \in \{\clubsuit, \spadesuit, \heartsuit, \diamondsuit\}^2 \text{ et } r \neq \text{Roi}\}. \end{aligned}$$

Finalement, l'évènement « obtenir au moins un roi » s'écrit donc

$$A = A_{\text{premier roi}} \cup A_{\text{second roi}} \cup A_{2 \text{ rois}}.$$

∞

1.4 Probabilité

Définition 1.4 Soit une expérience aléatoire et son univers Ω . On appelle **probabilité** sur Ω (ou mesure de probabilité) une fonction \mathbb{P} de $\mathcal{P}(\Omega)$ dans $[0, 1]$ telle que :

1. $\mathbb{P}(\Omega) = 1$
2. pour toute suite (dénombrable) de sous-ensembles de Ω $(A_i)_{i \geq 0}$ disjoints deux à deux (c'est-à-dire tels que pour tout $j \neq k$, $A_j \cap A_k = \emptyset$),

$$\mathbb{P}\left(\bigcup_{i \geq 0} A_i\right) = \sum_{i \geq 0} \mathbb{P}(A_i).$$

Cette seconde propriété est la **sigma additivité** de la mesure de probabilité.

⚠ Remarque 1.2 Comme nous l'avons indiqué dans la remarque 1.1, on s'autorise dans le cas général à ne considérer que certains sous-ensembles de Ω comme observables. Dans ces situations, la probabilité n'est définie que pour les ensembles considérés.

⚠ Remarque 1.3 Nous numérotions les suites (finies ou dénombrables) de façon arbitraire à partir de 0 ou de 1, où même d'un entier quelconque. Si elle est utilisée de façon cohérente, la numérotation n'a pas d'influence sur les résultats. Il faut simplement s'assurer qu'on utilise une numérotation fixe pour une suite donnée.

Classiquement, les probabilités s'interprètent comme des limites de fréquences. Considérons par exemple le lancer d'une pièce parfaitement symétrique. Intuitivement, nous avons « autant de chance » d'obtenir pile ou face en lançant la pièce. Cependant, nous savons tout aussi intuitivement que si nous lançons dix fois de suite la pièce, nous obtiendrons seulement « environ » 5 fois pile. Par « environ », nous entendons qu'en général, nous nous attendons à obtenir 5 fois pile, mais qu'il n'est pas impossible d'obtenir 4 ou 6 fois pile et qu'il n'est pas impossible, mais plus rare, d'obtenir 1 ou 9 fois pile. Si nous répétons n fois l'expérience et que nous calculons la fréquence des piles, nous nous attendons de ce fait à obtenir environ 0,5. L'intuition suggère que la fréquence sera d'autant plus proche de 0,5 que n est grand, ce que nous pouvons formaliser en disant que la probabilité d'obtenir pile est la limite de la fréquence quand n tend vers l'infini.

♣ **Exemple 1.15** On lance une pièce de monnaie. L'univers de l'expérience est alors $\Omega = \{\text{pile}, \text{face}\}$. Décrivons l'ensemble $\mathcal{P}(\Omega)$. On a clairement

$$\mathcal{P}(\Omega) = \{\emptyset, \{\text{pile}\}, \{\text{face}\}, \{\text{pile}, \text{face}\}\}.$$

Pour définir une probabilité, on doit donner quatre réels dans $[0, 1]$ vérifiant les propriétés de la définition 1.4. On a bien sûr

$$\mathbb{P}(\{\text{pile}, \text{face}\}) = \mathbb{P}(\Omega) = 1.$$

Si la pièce est parfaitement régulière, l'interprétation ci-dessus conduit à poser

$$\mathbb{P}(\{\text{pile}\}) = \mathbb{P}(\{\text{face}\}) = 0,5.$$

Comme l'évènement \emptyset ne se produit jamais (l'expérience a toujours un résultat), il semble naturel de poser enfin

$$\mathbb{P}(\emptyset) = 0.$$

Nous verrons plus loin que \mathbb{P} ainsi définie est bien une probabilité. Il nous faut en particulier vérifier que la deuxième propriété de la définition est bien satisfaite. C'est le rôle de la proposition 1.1. ♣

Propriétés 1.1 *Soit une expérience aléatoire, son univers Ω et une probabilité \mathbb{P} sur Ω . Alors \mathbb{P} vérifie les propriétés suivantes :*

1. $\mathbb{P}(\emptyset) = 0$;
2. pour toute suite finie d'évènements deux à deux disjoints $(A_i)_{1 \leq i \leq n}$, on a $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$;
3. pour tout évènement A , $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$;
4. pour tous évènements A et B ,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),$$

et en particulier quand $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$;

5. \mathbb{P} est croissante, c'est-à-dire que pour tous A et B tels que $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$;
6. pour toute suite croissante de sous-ensembles de Ω , les $(A_i)_{i \geq 0}$ (avec donc $i \leq j \Rightarrow A_i \subset A_j$), on a

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{i \geq 0} A_i\right) ;$$

7. pour toute suite décroissante de sous-ensembles de Ω , les $(A_i)_{i \geq 0}$ (avec donc $i \leq j \Rightarrow A_j \subset A_i$), on a

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{i \geq 0} A_i\right).$$

Preuve Ces propriétés sont des conséquences des caractéristiques des probabilités :

1. Considérons la suite d'ensembles définie par $A_0 = \Omega$ et $A_i = \emptyset$ pour tout $i \geq 1$. On constate que pour tout $j \neq k$, $A_j \cap A_k = \emptyset$. En effet, comme $j \neq k$, au moins un des indices est supérieur ou égal à 1, et l'ensemble correspondant est alors l'ensemble vide. Son intersection avec n'importe quel autre ensemble étant vide, on en déduit $A_j \cap A_k = \emptyset$. On a donc une suite d'ensembles disjoints deux à deux à laquelle on peut appliquer la propriété de sigma additivité de la définition 1.4. On a donc

$$\mathbb{P}\left(\bigcup_{i \geq 0} A_i\right) = \sum_{i \geq 0} \mathbb{P}(A_i) = \mathbb{P}(\Omega) + \sum_{i \geq 1} \mathbb{P}(\emptyset).$$

Or, $\mathbb{P}(\emptyset)$ est une valeur numérique fixée de $[0, 1]$. Supposons par l'absurde que $\mathbb{P}(\emptyset) > 0$. On a alors $\sum_{i \geq 1} \mathbb{P}(\emptyset) = \infty$. Mais on a aussi $\bigcup_{i \geq 0} A_i = \Omega$ et donc que $\mathbb{P}\left(\bigcup_{i \geq 0} A_i\right) = \mathbb{P}(\Omega) = 1$ (la deuxième égalité venant de la première propriété de \mathbb{P} selon la définition 1.4). On a donc $1 = \infty$, ce qui est impossible et donne donc $\mathbb{P}(\emptyset) = 0$.

2. Soit donc une suite finie d'évènements deux à deux disjoints $(A_i)_{1 \leq i \leq n}$. On complète cette suite en une suite infinie en prenant $A_0 = \emptyset$ et $A_i = \emptyset$ pour $i > n$. Cette suite est constituée d'ensembles deux à deux disjoints. En effet, si on considère l'intersection $A_j \cap A_k$ pour $k \neq j$, soit les deux ensembles sont dans la suite finie d'origine et leur intersection est donc vide par hypothèse. Dans le cas contraire, au moins un des deux ensembles est vide (par construction) et l'intersection est donc vide. On peut alors appliquer la sigma additivité qui donne

$$\mathbb{P}\left(\bigcup_{i \geq 0} A_i\right) = \sum_{i \geq 0} \mathbb{P}(A_i) = \sum_{i=1}^n \mathbb{P}(A_i),$$

la deuxième égalité venant de $\mathbb{P}(\emptyset) = 0$ que nous venons de démontrer. Comme en outre nous avons $\bigcup_{i \geq 0} A_i = \bigcup_{i=1}^n A_i$, ce qui conduit à la conclusion recherchée.

3. Soit un évènement A . Considérons la suite finie d'ensembles définie par $A_1 = A$, $A_2 = \bar{A}$. Cette suite est constituée d'ensembles deux à deux disjoints car $A_1 \cap A_2 = \emptyset$, par définition du complémentaire. On peut donc appliquer la propriété 2, ce qui donne

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = P(A) + P(\bar{A}).$$

Or, on a $A \cup \bar{A} = \Omega$ par définition du complémentaire. On a donc

$$\mathbb{P}(\Omega) = 1 = P(A) + P(\bar{A}),$$

ce qui conduit donc à $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$.

4. La démonstration de la troisième propriété s'appuie sur la décomposition de $A \cup B$ en trois sous-ensembles. En effet, un élément de $A \cup B$ est soit élément de A mais pas de B (donc élément de $A \cap \bar{B}$), soit élément de B mais pas de A (donc élément de $B \cap \bar{A}$), soit enfin élément de A et de B (donc élément de $A \cap B$). Il est clair que ces trois ensembles sont disjoints et qu'on a bien (cf l'annexe A)

$$A \cup B = (A \cap \bar{B}) \cup (B \cap \bar{A}) \cup (A \cap B).$$

On applique alors la propriété 2, ce qui donne

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}((A \cap \overline{B}) \cup (B \cap \overline{A}) \cup (A \cap B)), \\ &= \mathbb{P}(A \cap \overline{B}) + \mathbb{P}(B \cap \overline{A}) + \mathbb{P}(A \cap B).\end{aligned}$$

D'autre part, on sait aussi que $A = (A \cap B) \cup (A \cap \overline{B})$, relation dans laquelle l'union est disjointe. De ce fait, en appliquant de nouveau la propriété 2, on a

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}).$$

De la même façon, on a

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap \overline{A}).$$

En ajoutant les deux égalités, on obtient

$$\mathbb{P}(A) + \mathbb{P}(B) = 2\mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) + \mathbb{P}(B \cap \overline{A}).$$

En combinant avec l'égalité démontrée pour $\mathbb{P}(A \cup B)$ on obtient bien

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Le cas particulier $A \cap B = \emptyset$ s'obtient soit directement comme conséquence de la propriété 2, soit en utilisant la relation qui vient d'être démontrée en conjonction avec la propriété 1.

5. Soit A et B tels que $A \subset B$. Alors $B = A \cup (B \cap \overline{A})$. Comme l'union est disjointe, on a donc par la propriété 2

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap \overline{A}).$$

Comme une probabilité est positive ou nulle, on en déduit que $\mathbb{P}(A) \leq \mathbb{P}(B)$.

6. Pour prouver la propriété 6, il faut s'affranchir du fait que les A_i ne sont pas disjoints. Comme la suite $(A_i)_{i \geq 0}$ est croissante, on peut construire une suite des différences en posant $B_i = A_i \setminus A_{i-1}$ pour $i \geq 1$ et $B_0 = A_0$.

On montre par récurrence que $\bigcup_{i=0}^n B_i = \bigcup_{i=0}^n A_i = A_n$. Notons que $\bigcup_{i=0}^n A_i = A_n$ est une conséquence évidente de la croissance de la suite des A_i . Pour l'autre égalité, le cas $n = 0$ est évident par définition de $B_0 = A_0$. En supposant l'hypothèse vraie au rang n , on étudie $\bigcup_{i=0}^{n+1} B_i = (\bigcup_{i=0}^n B_i) \cup B_{n+1}$. Par hypothèse de récurrence, on a donc $\bigcup_{i=0}^{n+1} B_i = A_n \cup B_{n+1}$. Par définition de $B_{n+1} = A_{n+1} \setminus A_n$, $A_n \cup B_{n+1} = A_{n+1}$, ce qui montre que l'hypothèse est vraie au rang $n + 1$.

D'autre part, il est clair que les B_i sont disjoints deux à deux. Soit en effet $0 \leq i < j$, alors par propriété précédente, $B_i \subset A_{j-1}$ car $A_{j-1} = \bigcup_{k=0}^{j-1} B_k$. Or $B_j = A_j \setminus A_{j-1}$ et donc $B_j \cap A_{j-1} = \emptyset$ ce qui implique $B_j \cap B_i = \emptyset$.

Les B_i étant disjoints, on peut appliquer l'additivité pour conclure que

$$\mathbb{P}(A_n) = \sum_{i=0}^n \mathbb{P}(B_i).$$

La somme de droite converge vers $\sum_{i \geq 0} \mathbb{P}(B_i) = \mathbb{P}\left(\bigcup_{i \geq 0} B_i\right)$ par sigma additivité, et on a donc

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{i \geq 0} B_i\right) = \mathbb{P}\left(\bigcup_{i \geq 0} A_i\right).$$

7. La propriété 7 s'obtient simplement en passant au complémentaire. En effet, si la suite $(A_i)_{i \geq 0}$ est décroissante, alors la suite $(\bar{A}_i)_{i \geq 0}$ est croissante. On lui applique donc la propriété 6 qui donne

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\bar{A}_n) = \mathbb{P}\left(\bigcup_{i \geq 0} \bar{A}_i\right).$$

Or

$$\bigcup_{i \geq 0} \bar{A}_i = \overline{\bigcap_{i \geq 0} A_i},$$

et donc

$$\mathbb{P}\left(\bigcup_{i \geq 0} \bar{A}_i\right) = 1 - \mathbb{P}\left(\bigcap_{i \geq 0} A_i\right),$$

ce qui permet de conclure car $\lim_{n \rightarrow +\infty} \mathbb{P}(\bar{A}_n) = 1 - \lim_{n \rightarrow +\infty} \mathbb{P}(A_n)$. \square

▲ Remarque 1.4 La définition et les propriétés des probabilités font intervenir des unions et intersections infinies (dénombrables). Intuitivement $\omega \in \bigcup_{i \geq 1} A_i$ si au moins un des événements de la suite $(A_i)_{i \geq 1}$ contient le résultat ω . De la même façon, $\omega \in \bigcap_{i \geq 1} A_i$ si tous les événements de la suite $(A_i)_{i \geq 1}$ contiennent le résultat ω .

En combinant de telles unions et intersections infinies, on peut exprimer des propriétés logiques intéressantes. Par exemple,

$$\omega \in \bigcap_{n \geq 1} \left(\bigcup_{k \geq n} A_k \right),$$

s'interprète de la façon suivante : le résultat ω est réalisé infiniment souvent dans la suite $(A_i)_{i \geq 1}$. En d'autres termes, il existe une infinité d'indices i tels que $\omega \in A_i$. On remarque en effet que pour tout n , $\omega \in \bigcup_{k \geq n} A_k$ (en raison de l'interprétation de l'intersection infinie). Mais $\omega \in \bigcup_{k \geq n} A_k$ signifie qu'il existe un indice $k \geq n$ tel que $\omega \in A_k$ (en raison de l'interprétation de l'union infinie). Supposons que le nombre d'indices i tels que $\omega \in A_i$ soit fini : alors il existe un plus grand indice i_{\max} vérifiant cette propriété. Mais $\omega \in \bigcup_{k \geq i_{\max}+1} A_k$ et il existe donc $k > i_{\max}$ tel que $\omega \in A_k$ ce qui contredit le caractère maximal de i_{\max} . On en déduit qu'il y a bien une infinité d'indices i tels que $\omega \in A_i$.

De la même façon, la propriété

$$\omega \in \bigcup_{n \geq 1} \left(\bigcap_{k \geq n} A_k \right),$$

s'interprète de la façon suivante : il existe un indice k tel que pour tout $i \geq k$, $\omega \in A_i$.

Ces deux constructions complexes permettent de définir des limites ensemblistes.

1.5 Probabilités sur un univers fini

La définition d'une mesure de probabilité peut sembler complexe car on doit *a priori* définir une fonction sur l'ensemble de tous les sous-ensembles de l'univers considéré. Dans le cas où l'univers est un ensemble fini³, la situation est beaucoup plus simple.

Cas général

On a tout d'abord la proposition suivante.

Proposition 1.1 *Soit une expérience aléatoire dont l'univers Ω est fini et s'écrit donc $\Omega = \{\omega_1, \dots, \omega_n\}$. Une probabilité \mathbb{P} sur Ω est caractérisée de façon unique par les n réels de $[0, 1]$, $(p_i)_{1 \leq i \leq n}$ tels que $p_i = \mathbb{P}(\{\omega_i\})$. Réciproquement, tout ensemble de n réels de $[0, 1]$ $(p_i)_{1 \leq i \leq n}$ tels que $\sum_{i=1}^n p_i = 1$ définit une (unique) probabilité \mathbb{P} sur Ω par $\mathbb{P}(A) = \sum_{i=1}^n \delta_{\omega_i \in A} p_i$, où $\delta_{\omega_i \in A}$ vaut 1 si $\omega_i \in A$, zéro sinon.*

En termes moins formels, se donner une probabilité sur un univers fini à n éléments revient à se donner n nombres réels de $[0, 1]$ correspondant aux probabilités de n événements élémentaires de l'univers. La somme des n réels doit être de 1.

Preuve (hors programme) Soit donc $\Omega = \{\omega_1, \dots, \omega_n\}$ et deux probabilités \mathbb{P}_1 et \mathbb{P}_2 qui coïncident sur les n événements élémentaires, c'est-à-dire pour tout $i \in \{1, \dots, n\}$, $\mathbb{P}_1(\{\omega_i\}) = \mathbb{P}_2(\{\omega_i\})$. On veut montrer que \mathbb{P}_1 et \mathbb{P}_2 sont en fait la même probabilité.

Soit donc un événement $A \subset \Omega$. Si $A = \emptyset$, alors $\mathbb{P}_1(A) = 0 = \mathbb{P}_2(A)$. Sinon, il existe un sous-ensemble non vide J de $\{1, \dots, n\}$ tel que $A = \bigcup_{j \in J} \{\omega_j\}$. D'après les propriétés 1.1, on a donc

$$\mathbb{P}_1(A) = \sum_{j \in J} \mathbb{P}_1(\{\omega_j\}) = \sum_{j \in J} \mathbb{P}_2(\{\omega_j\}) = \mathbb{P}_2(A).$$

en appliquant pour l'égalité centrale l'hypothèse sur les deux probabilités. On a donc montré que pour tout A , $\mathbb{P}_1(A) = \mathbb{P}_2(A)$, ce qui montre que $\mathbb{P}_1 = \mathbb{P}_2$. Une probabilité est donc uniquement caractérisée les $p_i = \mathbb{P}(\{\omega_i\})$.

Soit maintenant réciproquement. n nombres réels de $[0, 1]$ $(p_i)_{1 \leq i \leq n}$ tels que $\sum_{i=1}^n p_i = 1$. Montrons que \mathbb{P} telle que définie dans la proposition est bien une mesure de probabilité. Il est clair tout d'abord que $\mathbb{P}(A) \geq 0$ pour tout A . De plus, comme $\mathbb{P}(A) \leq \sum_{i=1}^n p_i = 1$, \mathbb{P} est bien à valeurs dans $[0, 1]$. De plus, il est aussi clair que $\mathbb{P}(\Omega) = \sum_{i=1}^n p_i$ et donc que $\mathbb{P}(\Omega) = 1$.

Considérons maintenant la propriété de sigma additivité et soit donc une suite d'événements de Ω , $(A_i)_{i \geq 0}$ deux à deux disjoints. Notons $B = \bigcup_{i \geq 0} A_i$. Comme $B \subset \Omega$, il existe J un sous-ensemble de $\{1, \dots, n\}$ tel que $B = \bigcup_{j \in J} \{\omega_j\}$. Soit $J_i = \{j \in J | \omega_j \in A_i\}$. Comme les A_i sont disjoints deux à deux, les J_i le sont aussi. Par définition de \mathbb{P} puis par définition des J_i , on a

$$\mathbb{P}(A_i) = \sum_{k=1}^n \delta_{\omega_k \in A_i} p_k = \sum_{j \in J_i} p_j,$$

3. La définition d'un ensemble fini et de son cardinal sont rappelées dans l'annexe C

et donc

$$\sum_{i \geq 0} \mathbb{P}(A_i) = \sum_{i \geq 0} \sum_{j \in J_i} p_j.$$

Par définition $J = \bigcup_{i \geq 0} J_i$. Comme l'union est disjointe, on a

$$\sum_{i \geq 0} \sum_{j \in J_i} p_j = \sum_{j \in J} p_j = \sum_{k=1}^n \delta_{\omega_k \in B} p_k = \mathbb{P}(B).$$

On a donc bien $\sum_{i \geq 0} \mathbb{P}(A_i) = \mathbb{P}(\bigcup_{i \geq 0} A_i)$. \square

Exemple 1.16 On considère $\Omega = \{0, 1, 2, 3, 4, 5\}$, l'ensemble des notes possibles pour des films sur un site d'agrégation de critiques. L'expérience aléatoire consiste à fixer un film et à choisir au hasard une des notes qui lui ont été attribuées par les spectateurs et critiques. En fonction de la perception du film, la probabilité sur Ω change. Par exemple, un film bien critiqué peut être caractérisé par la probabilité suivante :

ω	0	1	2	3	4	5
$\mathbb{P}_{bon}(\{\omega\})$	0	0	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{5}$	$\frac{1}{5}$

alors qu'un film considéré comme très mauvais pourrait avoir la probabilité suivante :

ω	0	1	2	3	4	5
$\mathbb{P}_{nul}(\{\omega\})$	$\frac{2}{5}$	$\frac{1}{2}$	$\frac{1}{10}$	0	0	0

Dans les deux, nous avons bien des probabilités car la somme des valeurs numériques indiquées dans les tableaux est de 1. De plus, toutes ces valeurs sont éléments de $[0, 1]$.

La proposition 1.1 permet en outre de calculer la probabilité d'un évènement pour les deux films. Par exemple, la probabilité de tomber sur une note inférieure ou égale à 3 pour le bon film est donnée par

$$\begin{aligned} \mathbb{P}_{bon}(\{0, 1, 2, 3\}) &= \mathbb{P}_{bon}(\{0\}) + \mathbb{P}_{bon}(\{1\}) + \mathbb{P}_{bon}(\{2\}) + \mathbb{P}_{bon}(\{3\}), \\ &= 0 + 0 + \frac{1}{10} + \frac{3}{10}, \\ &= \frac{2}{5}. \end{aligned}$$

La première égalité est une simple application de la propriété 2 des propriétés 1.1. On utilise ensuite le tableau ci-dessous pour réaliser les calculs. \heartsuit

Cas équiprobable

La situation se simplifie encore s'il est naturel de faire une hypothèse de symétrie sur l'expérience aléatoire. Considérons en effet le lancer d'un dé non truqué. Par symétrie matérielle de l'objet, on s'attend à tomber avec autant de chance sur chacun des faces. En termes de fréquences, on s'attend, lors de lancers répétés, à obtenir environ autant de fois chacune des six faces. Ceci se traduit naturellement en supposant que les probabilités de chacun des évènements élémentaires sont égales.

Définition 1.5 Soit une expérience aléatoire et son univers **fini** Ω . La probabilité **uniforme** sur Ω est celle qui associe à chaque évènement élémentaire la même probabilité. Elle est définie par

$$\forall \omega \in \Omega, \mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}.$$

On parle alors d'**équiprobabilité** pour l'expérience concernée.

△ Remarque 1.5 Le choix d'utiliser la probabilité uniforme pour modéliser le hasard dans une expérience aléatoire n'est pas anodin. Considérons de nouveau l'exemple 1.12 dans lequel on choisit au hasard un jeton dans une urne qui en compte cinq, trois rouges et deux bleus.

Si on modélise les jetons comme indiscernables, l'univers de l'expérience est réduit aux deux couleurs. Dans cette situation, l'hypothèse d'équiprobabilité ne semble pas pertinente : intuitivement, le fait d'avoir plus de jetons rouges que de bleus laisse penser qu'on tirera plus souvent un jeton rouge et donc que la probabilité de l'évènement correspondant devrait être plus grande que celle de l'évènement « tirer un jeton bleu ». Bien que l'univers soit ici symétrique (deux couleurs), l'expérience sous-jacente ne l'est pas : il y a plus de jetons rouges que de jetons bleus.

Si on modélise au contraire les jetons comme discernables, l'univers contient cinq résultats possibles, un par jeton. Il est alors naturel de faire une hypothèse d'équiprobabilité, c'est même d'ailleurs le sens qu'on accorde en général à l'expression « choisir au hasard » dans cette situation : aucun des jetons n'est privilégié à priori et on a donc une chance sur cinq de choisir l'un d'entre eux. Ce choix conduit à des valeurs intuitivement correctes pour les probabilités (en particulier, avoir deux chances sur cinq de tomber sur un jeton bleu).

On voit donc que le choix de l'univers comme modélisation de l'expérience aléatoire peut avoir des conséquences très importantes par la suite, en particulier au niveau de la construction d'une probabilité.

∞ Exemple 1.17 Reprenons l'exemple 1.7 dans lequel on lance simultanément deux dés. Considérons tout d'abord la modélisation dans laquelle les dés sont discernables. On a donc

$$\Omega_d = \{\blacksquare, \blacklozenge, \blacktriangle, \blacklozenge, \blacktriangle, \blacklozenge\} \times \{\square, \square, \square, \square, \square, \square\},$$

en matérialisant la différence entre les dés par leur couleur. On sait que $|\Omega_d| = 36$ et donc que dans l'hypothèse d'une probabilité uniforme, la probabilité d'un évènement élémentaire est de $\frac{1}{36}$.

Considérons maintenant la modélisation dans laquelle les dés ne sont pas discernables. On a alors

$$\Omega_u = \{\{a, b\} \mid a \in \{\square, \square, \square, \square, \square, \square\} \text{ et } b \in \{\square, \square, \square, \square, \square, \square\}\}.$$

On peut détailler le contenu de Ω_u en recensant toutes les possibilités, ce qui donne

◻◻	◻◻	◻◻	◻◻	◻◻	◻◻
	◻◻	◻◻	◻◻	◻◻	◻◻
		◻◻	◻◻	◻◻	◻◻
			◻◻	◻◻	◻◻
				◻◻	◻◻
					◻◻

Dans ce tableau, on a toujours mis le dé donnant le plus petit chiffre en premier, mais cet ordre est arbitraire car l'indiscernabilité des dés empêche toute construction d'ordre. On constate d'après ce recensement exhaustif que $|\Omega_u| = 21$. Dans l'hypothèse d'une probabilité uniforme, la probabilité d'un évènement élémentaire est de $\frac{1}{21}$.

On constate donc que les deux modèles correspondent à des univers différents et donc des probabilités uniformes différentes. Considérons maintenant un évènement simple, comme celui d'obtenir la somme de 12 en lançant les deux dés. Dans les deux modèles, il s'agit d'un évènement élémentaire, soit $(\text{Ⓜ}, \text{Ⓜ})$ pour le modèle discernable et $\{\text{Ⓜ}, \text{Ⓜ}\} = \{\text{Ⓜ}\}$ pour le modèle indiscernable. Dans le modèle discernable, cet évènement a donc une probabilité de $\frac{1}{36}$ alors que dans le modèle indiscernable, on obtient une probabilité de $\frac{1}{21}$.

Bien entendu, ces résultats sont contradictoires, puisqu'il s'agit d'une même expérience aléatoire et d'un même évènement. Un des deux modèles ne correspond donc pas à la réalité. Dans le cas présent, c'est le choix de la probabilité uniforme qui n'est pas adaptée au modèle indiscernable. Intuitivement, cela vient du fait qu'en pratique, pour obtenir par exemple $\{\square, \square\}$, on devrait tenir compte du fait qu'il existe deux possibilités non observables, correspondant à (\square, \square) et (\square, \square) . Au contraire, il n'y a qu'une façon d'obtenir $\{\square, \square\} = \{\square\}$. Intuitivement, il faudrait donc que ce deuxième type d'évènements soit moins probable que le premier type. Ce n'est pas le cas pour la probabilité uniforme.

Cet exemple montre qu'il est général plus simple de choisir un modèle discernable car celui-ci est plus souvent compatible avec une probabilité uniforme que le modèle indiscernable. ∞

Les propriétés classiques des probabilités simplifient grandement le calcul de la probabilité d'un évènement dans le cas d'équiprobabilité.

Propriété 1.2 *Soit une expérience aléatoire et son univers fini Ω , muni de la probabilité \mathbb{P} uniforme. Pour tout évènement $A \subset \Omega$, on a*

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

Preuve On écrit $\Omega = \{\omega_1, \dots, \omega_n\}$, avec $n = |\Omega|$. Tout évènement A s'écrit $A = \cup_{j \in J} \{\omega_j\}$ pour un ensemble $J \subset \{1, \dots, n\}$. On a donc

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{j \in J} \{\omega_j\}\right) = \sum_{j \in J} \mathbb{P}(\{\omega_j\}) = |J| \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}.$$

On s'appuie d'abord sur les propriétés des probabilités, puis sur le fait que $|A| = |J|$. □

Le calcul de la probabilité d'un évènement dans le cas d'équiprobabilité se ramène donc au calcul du cardinal d'un ensemble, c'est-à-dire à du dénombrement. L'annexe C rappelle des résultats importants de dénombrement utiles en calcul des probabilités. Nous illustrons ces résultats dans les exemples qui concluent ce chapitre.

∞ **Exemple 1.18** On lance deux fois de suite un dé (cf exemple 1.5). L'univers de l'expérience est donc $\Omega = \{\square, \blacksquare, \boxtimes, \boxplus, \boxminus, \boxdot\}^2$. Le dé n'étant pas truqué, il est naturel de considérer une probabilité uniforme sur Ω . D'après l'annexe C, on a donc $|\Omega| = |\{\square, \blacksquare, \boxtimes, \boxplus, \boxminus, \boxdot\}|^2 = 36$. Considérons maintenant les évènements de l'exemple 1.10.

L'évènement « obtention de deux fois le même résultat » (évènement A) par un raisonnement simple : le nombre de paires dont les deux éléments sont égaux est identique au nombre de choix possible pour le premier élément, c'est-à-dire à 6. Pour la probabilité uniforme, on a donc

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}.$$

Considérons l'évènement B , l'« obtention d'un tirage dont le premier lancer est pair ». On a donc

$$B = \{\square, \boxtimes, \boxdot\} \times \{\square, \blacksquare, \boxtimes, \boxplus, \boxminus, \boxdot\}.$$

D'après l'annexe C, on a donc

$$|B| = |\{\square, \boxtimes, \boxdot\}| \times |\{\square, \blacksquare, \boxtimes, \boxplus, \boxminus, \boxdot\}| = 3 \times 6 = 18.$$

On en déduit que pour la probabilité uniforme, $\mathbb{P}(B) = \frac{1}{2}$. ∞

Tirages multiples

Un cas particulier important d'expérience aléatoire illustre parfaitement les liens entre dénombrement et probabilité uniforme. Considérons l'expérience suivante : on place n jetons numérotés de 1 à n dans une urne, puis on choisit au hasard p jetons dans cette urne. On note $J = \{1, \dots, n\}$ l'ensemble des jetons. En fonction du mode de tirage, l'expérience est modélisée par des univers différents.

Tirages successifs avec remise : il s'agit du mode de tirage qui conduit au modèle le plus simple. On choisit les jetons un par un, en remettant le jeton choisi dans l'urne après chaque tirage. On obtient donc un p -uplet dont les éléments sont choisis dans J , sans contrainte particulière. On peut notamment tomber plusieurs fois sur le même jeton. Formellement, l'univers de l'expérience est donc $\Omega = J^p$ et on a donc $|J^p| = |J|^p = n^p$. L'expérience est caractérisée par la prise en compte de l'ordre (on obtient une liste de p jetons) et par la remise qui autorise à avoir plusieurs fois le même jeton.

Tirages successifs sans remise : dans cette situation, on choisit les jetons un par un, en ne remettant pas les jetons après tirage. On obtient ainsi un p -uplet (une liste de p jetons) dont les éléments sont tous distincts. Formellement, l'univers de l'expérience est donc

$$\Omega = \{(j_1, \dots, j_p) \in J^p \mid \forall (k, l) \ k \neq l \Rightarrow j_k \neq j_l\}.$$

D'après l'annexe C, Ω est en fait l'ensemble des arrangements de p éléments parmi n et donc $|\Omega| = A_n^p$. Les deux aspects importants de l'expérience qui conduisent à

ce résultat sont la prise en compte de l'ordre des tirages (et donc l'obtention d'un p -uplet) et de l'absence de remise dans l'urne (et donc le fait que les éléments du p -uplet sont distincts).

Tirage simultané : le dernier mode classique de tirage consiste à prendre en une seule fois un paquet de p jetons. Comme dans le mode précédent, chaque tirage ne peut contenir qu'une seule fois un jeton. Cependant, on ne peut pas déduire du tirage un ordre sur les p jetons choisis. L'univers de l'expérience est donc le suivant

$$\Omega = \{K \subset J \mid |K| = p\} = \{\{j_1, \dots, j_p\} \subset J \mid \forall (k, l) \ k \neq l \Rightarrow j_k \neq j_l\}.$$

D'après l'annexe C, Ω est donc l'ensemble des combinaisons de p éléments choisis parmi n et donc $|\Omega| = C_n^p$. Ce mode de tirage est caractérisé par son aspect simultané qui conduit à la fois à l'absence d'ordre et de répétition.

Comme indiqué dans l'exemple 1.7, il est parfois plus clair ou plus simple de ne pas modéliser directement l'expérience étudiée mais plutôt une version dans laquelle on aurait plus d'information. Les exemples 1.7 et 1.8 montrent qu'on peut considérer comme discernables résultats qui ne le sont pas dans l'expérience réelle. Dans le présent contexte, on peut modéliser un tirage simultané par une série de tirages successifs : on fait comme si on pouvait observer un ordre dans le paquet de jetons tirés de l'urne.

Le modèle général du tirage dans une urne se retrouve dans de nombreuses situations concrètes illustrées dans les exemples suivants.

☛ **Exemple 1.19** Beaucoup de pays possèdent une loterie nationale qui propose une variante de jeu de loto. Un exemple d'un tel jeu est celui dans lequel chaque joueur choisit 6 numéros distincts parmi 42. Le tirage du lot est aussi de 6 numéros (sans remise) dans les 42. Le joueur gagne s'il a découvert au moins 3 des nombres, le gain augmentant avec le nombre de numéros obtenus.

Le tirage se déroule de façon séquentielle, mais il est plus simple de le considérer comme simultané car l'ordre n'est pas pris en compte dans les événements correspondants à des gains. On se retrouve donc dans la situation d'un tirage simultané de 6 éléments parmi 42, donc l'univers est

$$\Omega = \{K \subset \{1, 2, \dots, 42\} \mid |K| = 6\}.$$

On a donc $|\Omega| = C_{42}^6$.

Considérons maintenant les différents événements correspondants à des gains :

6 numéros : cette situation est la plus facile à gérer. En effet, chaque combinaison de 6 nombres est unique. De ce fait, l'évènement « la combinaison du joueur a été obtenue » est de cardinal 1 et donc, sous l'hypothèse de probabilité uniforme, sa probabilité est de $\frac{1}{C_{42}^6}$ (soit environ $1.9 \cdot 10^{-7}$).

moins de 6 numéros : la situation se complique quand le tirage ne donne pas la combinaison du joueur. Considérons l'évènement A_p « le tirage comporte p numéros contenus dans la combinaison du joueur », avec $0 < p < 6$. Formellement, on a

$$A_p = \{K \subset \{1, \dots, 42\} \mid |K| = 6, |K \cap T| = p\},$$

où T désigne la combinaison du joueur. Un tirage de A_p se décompose en p éléments choisis dans T et en $6 - p$ éléments choisis dans $\{1, \dots, 42\} \setminus T$. On peut donc réécrire A_p sous la forme suivante

$$A_p = \{K \subset \{1, \dots, 42\} \mid K = B \cup C, C \subset T, |C| = p, \\ B \subset \{1, \dots, 42\} \setminus T, |B| = 6 - p\}.$$

Or, il y a C_6^p sous-ensembles distincts à p éléments de T et C_{36}^{6-p} sous-ensembles distincts à $6 - p$ éléments de $\{1, \dots, 42\} \setminus T$. D'autre part, il est clair que pour un sous-ensemble B fixé dans la décomposition ci-dessus, on obtient un K différent pour chacun des C possibles. De ce fait, on a

$$|A_p| = C_6^p C_{36}^{6-p},$$

et donc

$$\mathbb{P}(A_p) = \frac{C_6^p C_{36}^{6-p}}{C_{42}^6}.$$

On obtient finalement les probabilités suivantes :

p	1	2	3	4	5
$\mathbb{P}(A_p)$	0,43	0,17	0,027	$1,8 \cdot 10^{-3}$	$4,1 \cdot 10^{-5}$

☞ **Exemple 1.20** L'ancien Loto de la Française des jeux (avant 2008) utilisait les règles suivantes. Un joueur choisit 6 numéros parmi 49. Un tirage du loto consiste en 7 numéros parmi 49, repartis en deux groupes : les 6 numéros principaux et un numéro complémentaire. Pour gagner, le joueur doit avoir obtenu au moins 3 numéros parmi les 6 principaux. Les gains augmentent en fonction du nombre de numéros obtenus dans l'ensemble principal, en tenant compte de l'éventuelle obtention du numéro complémentaire.

Le tirage est légèrement plus complexe que celui de l'exemple 1.19. Comme dans ce dernier, on ne tient pas compte de l'ordre. On note J l'ensemble des entiers de 1 à 49 (les numéros possibles). L'univers de l'expérience est alors

$$\Omega = \{(K, c) \in \mathcal{P}(J) \times J \mid |K| = 6, c \in J \setminus K\},$$

c'est-à-dire qu'un tirage est une paire composée de 6 numéros distincts choisis dans J (le sous-ensemble K) et d'un numéro (c) choisi parmi les 43 numéros restants après le tirage de K . Il y a clairement C_{49}^6 possibilités pour K car un choix de numéros est une combinaison de 6 numéros parmi 49. Pour chaque valeur de K , on a 43 possibilités pour c , puisqu'on choisit c dans les $49 - 6 = 43$ éléments restants. On a donc au total $|\Omega| = 43 \times C_{49}^6$ possibilités.

Les combinaisons gagnantes sont celles qui contiennent des numéros de K , avec éventuellement le chiffre c . Notons T la combinaison de 6 numéros distincts choisis par le joueur et étudions les différents gains possibles :

6 numéros : cette situation est proche de celle de l'exemple 1.19. Elle correspond à $K = T$ et donc plus précisément à l'évènement

$$A = \{(T, c) \in \mathcal{P}(J) \times J \mid c \in J \setminus T\}.$$

Il est clair que $|A| = 43$. Sous l'hypothèse d'une probabilité uniforme, on a donc

$$\mathbb{P}(A) = \frac{43}{43 \times C_{49}^6} = \frac{1}{C_{49}^6} \simeq 7,15 \cdot 10^{-8}.$$

On remarque que tout se passe ici comme dans l'exemple 1.19, c'est-à-dire comme si le numéro complémentaire n'existait pas.

moins de 6 numéros sans complémentaire : considérons l'évènement A_p « le tirage comporte p numéros principaux contenus dans la combinaison du joueur », avec $0 < p < 6$. Si T désigne la combinaison choisie par le joueur, une combinaison dans A_p se décompose en deux sous-ensembles B et D , et en un numéro complémentaire c . $D \subset T$, de cardinal p , est l'ensemble des numéros du joueur contenus dans le tirage. B est le reste des numéros principaux, avec $|B| = 6 - p$. Enfin, le complémentaire c est choisi dans $J \setminus (B \cup T)$ (en effet, le complémentaire n'est pas dans la combinaison choisie par le joueur). Formellement, A_p s'écrit donc

$$A_p = \{(K, c) \in \mathcal{P}(J) \times J \mid K = B \cup D, D \subset T, |D| = p, \\ B \subset J \setminus T, |B| = 6 - p, c \in J \setminus (B \cup T)\}.$$

On raisonne comme dans l'exemple 1.19 : il y a C_6^p sous-ensembles distincts à p éléments de T (les choix pour D), C_{43}^{6-p} sous-ensembles distincts à $6 - p$ éléments de $J \setminus T$ et enfin $43 - (6 - p) = 37 + p$ choix possibles pour le complémentaire. Ces choix étant tous indépendants, on en déduit donc que

$$|A_p| = (37 + p)C_6^p C_{43}^{6-p}.$$

Sous l'hypothèse de probabilité uniforme, on en déduit donc que

$$\mathbb{P}(A_p) = \frac{(37 + p)C_6^p C_{43}^{6-p}}{43 \times C_{49}^6}.$$

On obtient finalement les probabilités suivantes :

p	1	2	3	4	5
$\mathbb{P}(A_p)$	0,36	0,12	0,016	$9,2 \cdot 10^{-4}$	$1,8 \cdot 10^{-5}$

moins de 6 numéros avec complémentaire : considérons l'évènement B_p « le tirage comporte p numéros principaux ainsi que le numéro complémentaire contenus dans la combinaison du joueur », avec $0 < p < 6$. Si T désigne la combinaison choisie par le joueur, une combinaison dans B_p se décompose comme précédemment en deux sous-ensembles A et D , et en un numéro complémentaire c . $D \subset T$, de cardinal p , est l'ensemble des numéros du joueur contenus dans les numéros principaux du tirage. A est le reste des numéros principaux, avec $|A| = 6 - p$. Enfin, le complémentaire est choisi dans $T \setminus D$, puisqu'il est contenu dans la combinaison du joueur. De ce fait, il y a $6 - p$ choix possibles pour le complémentaire. Par un raisonnement similaire à celui du cas sans complémentaire, on en déduit donc que

$$|B_p| = (6 - p)C_6^p C_{43}^{6-p},$$

ce qui conduit à

$$\mathbb{P}(B_p) = \frac{(6-p)C_6^p C_{43}^{6-p}}{43 \times C_{49}^6},$$

et aux probabilités suivantes :

p	1	2	3	4	5
$\mathbb{P}(B_p)$	0,048	0,012	0,0012	$4,5 \cdot 10^{-5}$	$4,3 \cdot 10^{-7}$

»

Chapitre 2

Probabilités conditionnelles

2.1 Évènement réalisé

Les probabilités peuvent s'interpréter comme une prévision sur le résultat futur d'une expérience aléatoire. Quand on lance un dé (à six faces et non truqué), dire que la probabilité d'obtenir 1 est de $\frac{1}{6}$ revient à dire qu'on s'attend en moyenne à obtenir 1 une fois sur 6.

Intuitivement, si on obtient une information certaine (même partielle) sur le résultat d'une expérience aléatoire, il semble naturel que les probabilités associées soient modifiées. Considérons une première personne qui lance un dé sans regarder le résultat et qui demande à une deuxième personne de lui dire si ce résultat est pair. Avant d'avoir la réponse, la première personne suppose que chaque entier entre 1 et 6 a la même probabilité d'être obtenu, $\frac{1}{6}$. Après avoir eu une réponse, par exemple positive, la première personne ne connaît toujours pas le résultat effectif, mais elle modifie ses croyances (ses attentes) sur ce résultat. Elle sait maintenant que les chiffres 1, 3 et 5 sont impossibles et suppose naturellement que les trois autres chiffres sont équiprobables, et sont donc chacun de probabilité $\frac{1}{3}$. Ainsi, la probabilité d'obtenir 2 est passée de $\frac{1}{6}$ quand on ne savait rien sur le résultat de l'expérience à $\frac{1}{3}$ maintenant qu'on sait que l'évènement « obtenir un chiffre pair » est réalisé.

La notion de *probabilité conditionnelle* permet de formaliser le raisonnement développé dans l'exemple qui précède de manière à pouvoir l'appliquer simplement à toute expérience aléatoire dans laquelle on sait que certains évènements sont réalisés.

2.2 Probabilité conditionnelle

Définition 2.1 Soit une expérience aléatoire d'univers Ω et une probabilité \mathbb{P} sur Ω . Soit un évènement $A \subset \Omega$ tel que $\mathbb{P}(A) > 0$. On appelle **probabilité conditionnelle sachant A** la fonction de $\mathcal{P}(\Omega)$ dans $[0,1]$, notée $\mathbb{P}(\cdot|A)$, définie par

$$\forall B \in \mathcal{P}(\Omega), \mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

La notation $\mathbb{P}(B|A)$ se lit « probabilité de B sachant A ».

☞ **Exemple 2.1** Reprenons l'exemple de l'introduction : on lance un dé à six faces non truqué, ce qui conduit à l'univers $\Omega = \{\square, \square, \square, \square, \square, \square\}$ et à l'équiprobabilité. Soit A l'évènement « obtenir un chiffre pair ». Il est clair que $\mathbb{P}(A) = \frac{1}{2}$ car $A = \{\square, \square, \square\}$. Pour tout évènement B , on a donc $\mathbb{P}(B|A) = 2\mathbb{P}(B \cap A)$. Ceci conduit à la table suivante pour les évènements élémentaires :

d	\square	\square	\square	\square	\square	\square
$\mathbb{P}(\{d\} A)$	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0	$\frac{1}{3}$

En effet pour tout chiffre impair d l'évènement $\{d\} \cap A$ est impossible, et est donc de probabilité nulle. Au contraire, pour tout chiffre pair, $\{d\} \cap A = \{d\}$ et la probabilité de $\mathbb{P}(\{d\} \cap A)$ est donc $\frac{1}{6}$.

On retrouve ainsi les probabilités obtenues dans l'exemple introductif. ☞

⚠ **Remarque 2.1** Pour bien utiliser les probabilités conditionnelles, il faut absolument s'appuyer sur l'intuition que dans $\mathbb{P}(B|A)$, l'évènement A est réalisé de façon certaine. De ce fait, tout se passe comme si on changeait d'univers pour l'expérience aléatoire.

Considérons en effet l'exemple 2.1. Comme A est réalisé de façon certaine, tout se passe comme si on avait une nouvelle expérience aléatoire dont l'univers est justement A , puisque les seuls résultats maintenant possibles sont les éléments de A . En supposant que les évènements élémentaires de ce nouvel univers sont toujours équiprobables, la probabilité de $B \subset A$ est alors $\frac{|B|}{|A|}$. Or, dans l'univers Ω de départ, la probabilité de A était $\frac{|A|}{|\Omega|}$ et celle de B était $\frac{|B|}{|\Omega|}$. Comme $B \subset A$, $A \cap B = B$ et donc dans Ω , on avait $\mathbb{P}(B \cap A) = \frac{|B|}{|\Omega|}$. On constate alors que

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\frac{|B|}{|\Omega|}}{\frac{|A|}{|\Omega|}} = \frac{|B|}{|A|},$$

ce qui correspond bien à l'intuition du changement d'univers.

Proposition 2.1 Pour tout A tel que $\mathbb{P}(A) > 0$, la fonction $\mathbb{P}(\cdot|A)$ est une probabilité sur Ω .

Preuve Pour montrer que $\mathbb{P}(\cdot|A)$ est une probabilité, il suffit de vérifier les deux conditions de la définition 1.4 :

1. d'après la définition de $\mathbb{P}(\cdot|A)$, on a

$$\mathbb{P}(\Omega|A) = \frac{\mathbb{P}(\Omega \cap A)}{\mathbb{P}(A)}.$$

Or $\Omega \cap A = A$ et donc $\mathbb{P}(\Omega \cap A) = \mathbb{P}(A)$, soit finalement $\mathbb{P}(\Omega|A) = 1$, comme demandé par la définition.

2. Considérons maintenant une suite de sous-ensembles de Ω $(A_i)_{i \geq 0}$ disjoints deux à deux. Posons $B_i = A_i \cap A$. Comme les A_i sont disjoints deux à deux, les B_i le sont aussi. Donc, par σ additivité de la probabilité \mathbb{P} , on a

$$\mathbb{P}\left(\bigcup_{i \geq 0} B_i\right) = \sum_{i \geq 0} \mathbb{P}(B_i).$$

En divisant de chaque côté par $\mathbb{P}(A) > 0$, on obtient

$$\frac{\mathbb{P}\left(\bigcup_{i \geq 0} B_i\right)}{\mathbb{P}(A)} = \sum_{i \geq 0} \frac{\mathbb{P}(B_i)}{\mathbb{P}(A)}.$$

Par définition de $\mathbb{P}(\cdot|A)$, chaque terme $\frac{\mathbb{P}(B_i)}{\mathbb{P}(A)}$ est en fait $\mathbb{P}(A_i|A)$. D'autre part, on a

$$\bigcup_{i \geq 0} B_i = \bigcup_{i \geq 0} (A_i \cap A) = \left(\bigcup_{i \geq 0} A_i \right) \cap A,$$

par distributivité des opérations d'intersection et d'union. De ce fait,

$$\frac{\mathbb{P}\left(\bigcup_{i \geq 0} B_i\right)}{\mathbb{P}(A)} = \mathbb{P}\left(\bigcup_{i \geq 0} A_i \middle| A\right).$$

On a donc

$$\mathbb{P}\left(\bigcup_{i \geq 0} A_i \middle| A\right) = \sum_{i \geq 0} \mathbb{P}(A_i|A),$$

ce qui montre la σ additivité de $\mathbb{P}(\cdot|A)$. \square

Comme $\mathbb{P}(\cdot|A)$ est une probabilité, on peut lui appliquer les propriétés 1.1 des probabilités. Par exemple, si B et C sont deux événements disjoints, $\mathbb{P}(B \cup C|A) = \mathbb{P}(B|A) + \mathbb{P}(C|A)$.

Exemple 2.2 Reprenons l'exemple 1.8 de l'urne contenant trois jetons rouges et deux jetons bleus. On considère l'expérience dans laquelle on tire deux jetons successivement et sans remise. L'univers de l'expérience est alors

$$\Omega = \left\{ (j_1, j_2) \in \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}\}^2 \mid j_1 \neq j_2 \right\}.$$

Par symétrie, on se trouve dans un cas classique d'équiprobabilité (cf la section 1.5) dans lequel on utilisera le fait que $|\Omega| = A_5^2 = 20$.

Soit l'évènement A « obtenir un premier jeton rouge ». Cet évènement s'écrit mathématiquement

$$\begin{aligned} A &= \{(j_1, j_2) \in \{\mathbf{1}, \mathbf{2}, \mathbf{3}\} \times \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}\} \mid j_1 \neq j_2\}, \\ &= \{\mathbf{1}\} \times \{\mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}\} \cup \{\mathbf{2}\} \times \{\mathbf{1}, \mathbf{3}, \mathbf{4}, \mathbf{5}\} \cup \{\mathbf{3}\} \times \{\mathbf{1}, \mathbf{2}, \mathbf{4}, \mathbf{5}\}. \end{aligned}$$

On constate que $|A| = 3 \times 4 = 12$, et donc que $\mathbb{P}(A) = \frac{12}{20} = \frac{3}{5}$. De la même façon, on détermine facilement que l'évènement B « obtenir un deuxième jeton bleu » est de probabilité $\mathbb{P}(B) = \frac{4 \times 2}{20} = \frac{2}{5}$. Enfin, l'évènement $A \cap B$ est donné par

$$A \cap B = \{\mathbf{1}, \mathbf{2}, \mathbf{3}\} \times \{\mathbf{4}, \mathbf{5}\},$$

et donc $\mathbb{P}(A \cap B) = \frac{6}{20} = \frac{3}{10}$. On peut alors calculer des probabilités conditionnelles, par exemple :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{3}{10} \times \frac{5}{2} = \frac{3}{4},$$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{3}{10} \times \frac{5}{3} = \frac{1}{2}.$$

La deuxième probabilité conditionnelle est très simple à interpréter. Quand A est réalisé, on sait qu'il reste dans l'urne deux jetons de chaque couleur. On a donc maintenant une chance sur deux de tomber sur un jeton bleu.

La première probabilité est plus délicate à appréhender intuitivement car on donne une information sur le *deuxième* tirage. On ne peut donc pas faire le même raisonnement que pour la seconde probabilité. Il faut alors étudier les événements élémentaires qui constituent B . Si on fixe le second jeton au jeton 4, on voit que A a obligatoirement été choisi dans $\{1, 2, 3, 5\}$, et donc qu'il y a 3 chance sur 4 qu'il soit bleu. La situation est symétrique pour le cas où le second jeton est 5, ce qui conduit bien à la probabilité obtenue. ♣

2.3 Expériences aléatoires composées

Il arrive fréquemment en pratique qu'une expérience aléatoire consiste en plusieurs étapes faisant chacune intervenir une source de hasard. C'est le cas à chaque fois qu'on réalise une succession de tirages aléatoires d'objets dans un ensemble (par exemple des cartes dans un jeu de cartes). Une telle expérience aléatoire est dite **composée**. La notion de probabilité conditionnelle permet de calculer relativement facilement les probabilités de certains événements pour une expérience composée. Elle facilite aussi la définition de la probabilité sur l'univers de l'expérience.

♣ **Exemple 2.3** On dispose de deux urnes dans lesquelles sont placés des billes numérotées. La première urne contient 3 billes numérotées 1, 2 et 3, soit $U_1 = \{1, 2, 3\}$. La seconde urne contient 4 billes numérotées 2, 3, 4 et 5, soit $U_2 = \{2, 3, 4, 5\}$. On suppose que le choix d'une bille dans une urne obéit à une loi uniforme. L'expérience aléatoire composée consiste en la procédure suivante. Dans une première étape, on lance une pièce non truquée. Dans la deuxième étape, on choisit une bille dans une des deux urnes. Si on a obtenu pile dans la première étape, la bille est choisie dans U_1 , sinon elle est choisie dans U_2 .

L'univers de cette expérience est constitué de couples avec le résultat du lancer de la pièce et la bille obtenue. Il est clair qu'on a

$$\Omega = (\{\text{pile}\} \times U_1) \cup (\{\text{face}\} \times U_2).$$

D'autre part, la première étape de l'expérience est un lancer simple d'une pièce non truquée. Il est donc clair que $\mathbb{P}(\ll \text{pile} \gg) = \mathbb{P}(\ll \text{face} \gg) = \frac{1}{2}$, si « pile » (respectivement « face ») désigne ici l'évènement dans lequel la première étape produit le résultat pile (resp. face) quel que soit le résultat de la deuxième étape, c'est-à-dire si

$$\ll \text{pile} \gg = \{\text{pile}\} \times U_1,$$

$$\ll \text{face} \gg = \{\text{face}\} \times U_2.$$

En d'autres termes, \mathbb{P} est ici la probabilité définie sur l'univers de l'expérience composée.

De la même façon, il est facile de déterminer les probabilités d'obtenir certaines billes si on sait dans quelle urne la bille est choisie. Or, comme la deuxième étape a lieu *après* l'obtention du résultat de la première étape, l'évènement correspondant est *certain*. On doit donc considérer les probabilités de la seconde étape comme des probabilités conditionnelles. Ainsi on a par exemple $\mathbb{P}(B = 2 | \text{« pile »}) = \frac{1}{3}$, où B désigne le numéro porté par la bille obtenue dans la deuxième étape. En effet, si on a obtenu pile dans la première étape, on choisit dans l'urne U_1 uniformément et chaque chiffre a donc une probabilité de un tiers. De même, on a $\mathbb{P}(B = 2 | \text{« face »}) = \frac{1}{4}$ car le choix est maintenant réalisé dans l'urne U_2 qui contient 4 billes.

On peut alors calculer $\mathbb{P}(B = 2)$, soit la probabilité d'obtenir une bille numérotée 2 dans l'expérience composée (c'est-à-dire sans tenir compte de l'urne dont elle provient). On remarque que l'évènement $\{B = 2\}$ s'écrit comme l'union disjointe $\{B = 2\} = \{B = 2 \text{ et pile}\} \cup \{B = 2 \text{ et face}\}$. Or, par définition des probabilités conditionnelles,

$$\begin{aligned} \mathbb{P}(\{B = 2 \text{ et pile}\}) &= \mathbb{P}(\{B = 2\} \cap \{\text{« pile »}\}) \\ &= \mathbb{P}(B = 2 | \text{« pile »})\mathbb{P}(\text{« pile »}). \end{aligned}$$

De même, on a

$$\begin{aligned} \mathbb{P}(\{B = 2 \text{ et face}\}) &= \mathbb{P}(\{B = 2\} \cap \{\text{« face »}\}) \\ &= \mathbb{P}(B = 2 | \text{« face »})\mathbb{P}(\text{« face »}), \end{aligned}$$

et donc, puisque l'union des évènements est disjointe :

$$\begin{aligned} \mathbb{P}(B = 2) &= \mathbb{P}(B = 2 | \text{« pile »})\mathbb{P}(\text{« pile »}) + \mathbb{P}(B = 2 | \text{« face »})\mathbb{P}(\text{« face »}), \\ &= \frac{1}{3} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2}, \\ &= \frac{7}{24}. \end{aligned}$$

△ Remarque 2.2 *On pourrait croire naïvement qu'il est possible de raisonner directement sur l'univers Ω de l'exemple précédent, en faisant par exemple une hypothèse d'uniformité. On constate que $|\Omega| = 7$, ce qui conduit à supposer que la probabilité de tout $\omega \in \Omega$ est de $\frac{1}{7}$. En utilisant de nouveau la décomposition disjointe $\{B = 2\} = \{B = 2 \text{ et pile}\} \cup \{B = 2 \text{ et face}\}$, on en déduit que $\mathbb{P}(B = 2) = \frac{1}{7} + \frac{1}{7} = \frac{2}{7}$.*

Cette modélisation est cependant fautive, car elle repose sur une hypothèse implicite de symétrie entre tous les évènements élémentaires. Or, ceux-ci ne sont clairement pas équiprobables puisque les billes 2 et 3 apparaissent dans les deux urnes, contrairement aux autres billes (par exemple). Il est donc clair que la bille 1, par exemple, ne peut pas être tirée aussi souvent qu'une bille 2 : le seul moyen d'obtenir la bille 1 est d'abord d'obtenir pile (une chance sur deux), puis de tirer la bille 1 dans l'urne U_1 (une chance sur trois). Pour obtenir une bille 2, on peut soit faire pile puis tirer la bille 2 dans l'urne U_1 , ce qui arrive clairement aussi souvent que de faire pile puis de tirer la bille 1 (par hypothèse). Mais on a en plus la possibilité d'obtenir une bille 2 en faisant face, puis en tirant la bille 2 de l'urne U_2 . De ce fait, l'hypothèse d'équiprobabilité n'est pas acceptable, et le seul moyen de déterminer la probabilité sur Ω associée à l'expérience est de passer par le concept de probabilités conditionnelles.

2.4 Règle des probabilités totales

L'analyse réalisée dans l'exemple 2.3 s'appuie sur une décomposition d'un évènement en une union disjointe d'évènements sous une forme particulière qui facilite l'utilisation des probabilités conditionnelles. On dispose d'une proposition générale qui systématise le raisonnement réalisé.

Proposition 2.2 (Règle des probabilités totales) *Considérons une expérience aléatoire décrite par l'univers Ω et la probabilité \mathbb{P} . Soit une partition $\{A_1, \dots, A_n\}$ de Ω en n évènements A_1, \dots, A_n . Soit B un évènement quelconque. On a*

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B \cap A_i) \quad (2.1)$$

Si en outre les A_1, \dots, A_n sont tels que pour tout i , $\mathbb{P}(A_i) > 0$, on a

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i). \quad (2.2)$$

Preuve Comme $\{A_1, \dots, A_n\}$ est une partition, les $B \cap A_i$ sont disjoints deux à deux. En effet, on a $(B \cap A_i) \cap (B \cap A_j) = B \cap (A_i \cap A_j)$ (par associativité, commutativité et $B \cap B = B$). Or, si $i \neq j$, $A_i \cap A_j = \emptyset$ car $\{A_1, \dots, A_n\}$ est une partition. Donc si $i \neq j$, $(B \cap A_i) \cap (B \cap A_j) = \emptyset$. Donc $\bigcup_{i=1}^n (B \cap A_i)$ est une union disjointe et

$$\mathbb{P}\left(\bigcup_{i=1}^n (B \cap A_i)\right) = \sum_{i=1}^n \mathbb{P}(B \cap A_i).$$

En outre, $\bigcup_{i=1}^n (B \cap A_i) = B$. En effet, par distributivité,

$$\bigcup_{i=1}^n (B \cap A_i) = B \cap \left(\bigcup_{i=1}^n A_i\right).$$

Or, comme $\{A_1, \dots, A_n\}$ est une partition de Ω , $\bigcup_{i=1}^n A_i = \Omega$ et donc $\bigcup_{i=1}^n (B \cap A_i) = B \cap \Omega = B$. Donc $\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{i=1}^n (B \cap A_i)\right)$ ce qui achève de démontrer la règle des probabilités totales dans le premier cas (équation 2.1).

Si en outre tous les $\mathbb{P}(A_i) > 0$, on a par définition des probabilités conditionnelles : pour tout i , $\mathbb{P}(B \cap A_i) = \mathbb{P}(B|A_i)\mathbb{P}(A_i)$. Il suffit de réinjecter ces égalités dans l'équation (2.1) pour obtenir l'équation (2.2). \square

La règle des probabilités totales est très utile pour analyser les expériences composées, même quand celles-ci sont plus simples que dans l'exemple 2.3. Elle s'applique notamment dans le cas des tirages sans remise, comme dans l'exemple suivant.

Exemple 2.4 On tire deux cartes successivement et sans remise dans un jeu de 32 cartes (cf l'exemple 1.4). On cherche la probabilité d'obtenir un trèfle pour la seconde carte, soit $\mathbb{P}(C_2 \in \clubsuit)$. Or, les évènements $\{C_1 \in \clubsuit\}$ et $\{C_1 \notin \clubsuit\}$ forment une partition

de Ω : en effet, la première carte est soit un trèfle, soit une autre carte, de façon exclusive. On a donc, par la règle des probabilités totales

$$\mathbb{P}(C_2 \in \clubsuit) = \mathbb{P}(C_2 \in \clubsuit | C_1 \in \clubsuit)\mathbb{P}(C_1 \in \clubsuit) + \mathbb{P}(C_2 \in \clubsuit | C_1 \notin \clubsuit)\mathbb{P}(C_1 \notin \clubsuit).$$

Lors du premier tirage, le paquet est complet et comprend 32 cartes, dont 8 sont des trèfles. Par symétrie, on a donc clairement

$$\mathbb{P}(C_1 \in \clubsuit) = \frac{8}{32} = \frac{1}{4},$$

et donc

$$\mathbb{P}(C_1 \notin \clubsuit) = 1 - \frac{1}{4} = \frac{3}{4}.$$

De plus, si $\{C_1 \in \clubsuit\}$ est réalisé, le paquet de cartes restantes contient maintenant 31 cartes dont 7 sont des trèfles. On a donc

$$\mathbb{P}(C_2 \in \clubsuit | C_1 \in \clubsuit) = \frac{7}{31}.$$

Au contraire, si $\{C_1 \notin \clubsuit\}$ est réalisé, le paquet de cartes restantes contient maintenant 31 cartes dont 8 sont des trèfles. On a donc

$$\mathbb{P}(C_2 \in \clubsuit | C_1 \notin \clubsuit) = \frac{8}{31}.$$

On obtient ainsi

$$\mathbb{P}(C_2 \in \clubsuit) = \frac{7}{31} \times \frac{1}{4} + \frac{8}{31} \times \frac{3}{4} = \frac{1}{4}.$$

L'analyse directe de ce problème se fait de la façon suivante. D'après l'exemple 1.14, l'univers de l'expérience est

$$\Omega = \{(C_1, C_2) \in P^2 \mid C_1 \neq C_2\},$$

où P désigne le paquet de cartes. Il s'agit donc d'arrangements et on a $|\Omega| = A_{32}^2 = 32 \times 31 = 992$. Par symétrie, il est naturel d'utiliser sur cet univers une probabilité uniforme. Il nous faut donc calculer le cardinal de l'évènement $\{C_2 \in \clubsuit\}$. On peut décomposer cet évènement en deux évènements disjoints (un peu de la même façon que dans l'analyse par probabilités conditionnelles), en écrivant :

$$\{C_2 \in \clubsuit\} = \{C_2 \in \clubsuit, C_1 \in \clubsuit\} \cup \{C_2 \in \clubsuit, C_1 \notin \clubsuit\}.$$

Le premier sous-évènement correspond à un arrangement restreint aux cartes trèfles, c'est-à-dire

$$\{C_2 \in \clubsuit, C_1 \in \clubsuit\} = \{(C_1, C_2) \in \clubsuit^2 \mid C_1 \neq C_2\},$$

et on a donc clairement

$$|\{C_2 \in \clubsuit, C_1 \in \clubsuit\}| = A_8^2 = 8 \times 7 = 56.$$

Le second sous-évènement correspond à choisir une carte non trèfle parmi 24 possibilités, puis une carte trèfle parmi 8 possibilité, ce qui donne

$$\{C_2 \in \clubsuit, C_1 \notin \clubsuit\} = \clubsuit \times (P \setminus \clubsuit),$$

et donc

$$|\{C_2 \in \clubsuit, C_1 \notin \clubsuit\}| = 24 \times 8 = 192.$$

Finalement, on a donc

$$|\{C_2 \in \clubsuit\}| = 56 + 192 = 248.$$

Par uniformité de la probabilité, on obtient ainsi

$$\mathbb{P}(C_2 \in \clubsuit) = \frac{248}{992} = \frac{1}{4}.$$

Bien entendu, les deux méthodes de résolution conduisent au même résultat. En pratique, la première est souvent plus simple que la seconde, mais elle ne s'applique naturellement que quand l'expérience est séquentielle. Les tirages simultanés, par exemple, ne peuvent pas être étudiés facilement au moyen des probabilités conditionnelles. \heartsuit

Notons que l'exemple précédent applique une version simple de la règle des probabilités totales dans laquelle on étudie un événement et son complémentaire. On l'énonce de façon générale dans la proposition suivante.

Proposition 2.3 (Cas simple de la règle des probabilités totales) *Soit une expérience aléatoire décrite par l'univers Ω et la probabilité \mathbb{P} . Pour tout événement A tel que $1 > \mathbb{P}(A) > 0$, et tout événement B , on a*

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\bar{A})\mathbb{P}(\bar{A}) \quad (2.3)$$

Preuve Il s'agit simplement d'une application de la proposition 2.2 à $\{A, \bar{A}\}$ qui forme par définition une partition de Ω . Notons que comme $\mathbb{P}(A) < 1$, on a bien $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A) > 0$ ce qui permet l'application de la règle générale. On pourrait bien sûr considérer le cas $\mathbb{P}(A) = 1$, mais cela ne présente pas grand intérêt puisqu'on se retrouve alors avec $\mathbb{P}(B|A) = \mathbb{P}(B)$. \square

2.5 Règle de Bayes

Dans l'exemple 2.2, nous avons calculé des probabilités conditionnelles dans le sens causal (l'évènement réalisé porte sur la première partie de l'expérience) et dans le sens anti-causal (l'évènement réalisé porte sur la deuxième partie de l'expérience). Ce deuxième sens est particulièrement utile pour faire du raisonnement « probabiliste » : on sait qu'un évènement est réalisé et on se demande quelle peut en être la cause. On cherche alors à établir les probabilités des différentes causes possibles, sachant la nature exacte de l'évènement réalisé.

\heartsuit **Exemple 2.5** Reprenons les résultats de l'exemple 2.2. L'évènement A consiste à obtenir au premier tirage un jeton bleu. Par symétrie, on a vu que $\mathbb{P}(A) = \frac{3}{5}$. Supposons maintenant l'expérience réalisée, c'est-à-dire les deux tirages effectués.

Tant qu'aucune information sur le résultat n'est connue, on suppose toujours que $\mathbb{P}(A) = \frac{3}{5}$. Si on nous révèle le résultat du deuxième tirage, à savoir que l'évènement B « obtenir un deuxième jeton bleu » est réalisé, il est clair que cette information apporte

indirectement une information sur le résultat du premier tirage. En effet, si on a obtenu en premier un jeton bleu, alors le nombre de jetons bleus a été diminué et il est donc devenu moins probable d'en tirer un deuxième que dans la situation contraire. Comme on a obtenu un jeton bleu, il semble naturel de supposer que le premier jeton était plus probablement bleu que rouge. La probabilité conditionnelle $\mathbb{P}(A|B)$ traduit cette intuition de façon mathématique. On a obtenu $\mathbb{P}(A|B) = \frac{3}{4}$ dans l'exemple 2.2, une valeur plus grande que $\mathbb{P}(A) = \frac{3}{5}$. Cette révision de nos attentes concernant A montre l'influence de l'information certaine sur ces attentes. ∞

Pour faciliter ce raisonnement probabiliste, on s'appuie sur la règle de Bayes énoncée ci-dessous.

Proposition 2.4 (Règle de Bayes) *Soit une expérience aléatoire décrite par l'univers Ω et la probabilité \mathbb{P} . Soit A et B , deux évènements de probabilités non nulles ($\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$). On a*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \quad (2.4)$$

Preuve La formule s'obtient en appliquant deux fois la définition des probabilités conditionnelles. En effet :

$$\begin{aligned} \mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \\ &= \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \end{aligned}$$

□

∞ **Exemple 2.6 (Urne de Pòlya)** Soit une urne contenant une bille blanche et une bille noire. À chaque étape de l'expérience composée, on tire une bille dans l'urne. L'évènement B_k correspond à obtenir une bille blanche à l'étape k , son complémentaire N_k correspondant à l'obtention d'une bille noire à l'étape k . Après le tirage, on replace dans l'urne la bille tirée et une autre de la même couleur.

Le premier tirage est très simple, on a de façon évidente $\mathbb{P}(B_1) = \mathbb{P}(N_1) = \frac{1}{2}$, par symétrie. Pour étudier la deuxième étape, on applique le cas simple de la loi des probabilités totales en conditionnant par le résultat du premier tirage. On a donc :

$$\mathbb{P}(N_2) = \mathbb{P}(N_2|N_1)\mathbb{P}(N_1) + \mathbb{P}(N_2|B_1)\mathbb{P}(B_1),$$

En utilisant $\overline{B_1} = N_1$. Or, par symétrie, il est clair que

$$\begin{aligned} \mathbb{P}(N_2|N_1) &= \frac{2}{3}, \\ \mathbb{P}(N_2|B_1) &= \frac{1}{3}. \end{aligned}$$

et donc que

$$\begin{aligned}\mathbb{P}(N_2) &= \frac{2}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2}, \\ &= \frac{1}{2}.\end{aligned}$$

On peut alors appliquer la règle de Bayes pour voir ce que révèle le résultat du deuxième tirage sur le premier tirage. Comme dans l'exemple 2.2, on s'attend à ce que l'obtention d'une bille noire (événement N_2) augmente probabilité (conditionnelle) de l'évènement N_1 . La règle de Bayes donne ici :

$$\begin{aligned}\mathbb{P}(N_1|N_2) &= \frac{\mathbb{P}(N_2|N_1)\mathbb{P}(N_1)}{\mathbb{P}(N_2)}, \\ &= \frac{\frac{2}{3} \times \frac{1}{2}}{\frac{1}{2}}, \\ &= \frac{2}{3}.\end{aligned}$$

Le résultat confirme ainsi l'intuition. ∞

2.6 Indépendance

Dans certaines situations, avoir une information sur une partie d'une expérience aléatoire sous la forme d'un évènement réalisé n'entraîne pas de révision de la probabilité d'un autre évènement, comme le montre l'exemple simple suivant.

∞ **Exemple 2.7** On lance deux dés à six faces non truqués, un dé rouge et un dé noir. Soit l'évènement A « les deux dés donnent des résultats identiques » et l'évènement B « le dé rouge donne 1 ». On étudie $\mathbb{P}(A)$ et $\mathbb{P}(A|B)$.

L'univers de l'expérience est clairement

$$\Omega = \{\color{red}\blacksquare, \color{red}\blacksquare, \color{red}\blacksquare, \color{red}\blacksquare, \color{red}\blacksquare, \color{red}\blacksquare\} \times \{\blacksquare, \blacksquare, \blacksquare, \blacksquare, \blacksquare, \blacksquare\},$$

avec comme convention que le premier résultat est celui du dé rouge. Par symétrie, on prend sur Ω la probabilité uniforme. A est alors la diagonale de Ω , soit

$$A = \{(\color{red}\blacksquare, \blacksquare), (\color{red}\blacksquare, \blacksquare), (\color{red}\blacksquare, \blacksquare), (\color{red}\blacksquare, \blacksquare), (\color{red}\blacksquare, \blacksquare), (\color{red}\blacksquare, \blacksquare)\}.$$

et donc $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$. On retrouve essentiellement l'analyse de l'exemple 1.18.

L'évènement B s'écrit simplement $\{\color{red}\blacksquare\} \times \{\blacksquare, \blacksquare, \blacksquare, \blacksquare, \blacksquare, \blacksquare\}$, ce qui montre que $\mathbb{P}(B) = \frac{1}{6}$. Enfin, $A \cap B = \{(\color{red}\blacksquare, \blacksquare)\}$ et donc $\mathbb{P}(A \cap B) = \frac{1}{36}$. En appliquant la définition des probabilités conditionnelles, on a ainsi

$$\begin{aligned}\mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \\ &= \frac{1}{36} \times 6 = \frac{1}{6}.\end{aligned}$$

On a donc $\mathbb{P}(A|B) = \mathbb{P}(A)$ ce qui montre que savoir que B est réalisé n'apporte pas de connaissance sur A . ∞

Dans une telle situation, on parle d'indépendance entre les deux évènements, selon la définition suivante.

Définition 2.2 Soit une expérience aléatoire décrite par l'univers Ω et la probabilité \mathbb{P} . Soit A et B deux évènements. On dit que A et B sont **indépendants** si et seulement si :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B). \quad (2.5)$$

On note alors $A \perp B$.

Une conséquence élémentaire de l'indépendance est donnée par la propriété ci-dessous.

Propriété 2.1 Soit une expérience aléatoire décrite par l'univers Ω et la probabilité \mathbb{P} . Soit A et B deux évènements. Si $\mathbb{P}(A) > 0$ alors $A \perp B$ si et seulement si :

$$\mathbb{P}(B|A) = \mathbb{P}(B), \quad (2.6)$$

et de même si $\mathbb{P}(B) > 0$ alors $A \perp B$ si et seulement si :

$$\mathbb{P}(A|B) = \mathbb{P}(A). \quad (2.7)$$

Preuve Pour passer de $A \perp B$ aux formules, on applique simplement la définition des probabilités conditionnelles. Par exemple, on écrit $\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$, puis la propriété d'indépendance $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$ et enfin, on simplifie par $\mathbb{P}(A)$. Pour l'autre sens de la preuve, on applique exactement la même formule pour obtenir la condition d'indépendance de la définition 2.2. \square

⚠ Remarque 2.3 Attention, il est très important de ne pas confondre des évènements incompatibles (cf la section 1.3) avec des évènements indépendants. En effet, si A et B sont incompatibles, alors $\mathbb{P}(A \cap B) = 0$. Dans cette situation, l'indépendance n'est possible que si au moins des évènements est impossible. En effet, on doit avoir $\mathbb{P}(A) \times \mathbb{P}(B) = \mathbb{P}(A \cap B) = 0$. Cette situation se produit parfois, mais ce n'est généralement pas celle qui nous intéresse. Au contraire, la notion d'indépendance prend tout son intérêt quand $\mathbb{P}(A \cap B) > 0$: on se trouve en effet confronté à des évènements compatibles (le résultat de l'expérience aléatoire peut conduire à la réalisation de A et de B simultanément) mais qui pourtant n'apportent pas d'information l'un sur l'autre, ce qui peut d'ailleurs sembler paradoxal.

∞ Exemple 2.8 Reprenons l'exemple de l'urne de Pòlya (exemple 2.6). D'après les calculs effectués précédemment, N_1 et N_2 ne sont pas des évènements indépendants. En effet, $\mathbb{P}(N_2|N_1) = \frac{2}{3}$ alors que $\mathbb{P}(N_2) = \frac{1}{2}$ (et $\mathbb{P}(N_1) > 0$).

Considérons maintenant une expérience plus simple dans laquelle on part toujours d'une urne avec une bille blanche et une bille noire. On effectue ensuite une série de tirages **avec remise** : le contenu de l'urne ne change donc jamais. L'évènement « obtenir une bille noire au tirage k » est noté ici S_k . Comme pour l'urne de Pòlya, on a clairement $\mathbb{P}(S_1) = \mathbb{P}(\overline{S_1}) = \frac{1}{2}$. Comme le contenu de l'urne n'est jamais modifié, il est aussi évident que $\mathbb{P}(S_k) = \mathbb{P}(\overline{S_k}) = \frac{1}{2}$. Quand on considère un seul tirage, on a donc des

probabilités identiques pour l'urne classique et l'urne de Pòlya. En revanche, les tirages sont indépendants dans l'urne classique.

Pour deux tirages, l'univers est en effet $\Omega = \{Blanche, Noire\}^2$. Par symétrie, il est clair qu'on doit prendre sur cet univers une probabilité uniforme. On a donc $\mathbb{P}(S_1 \text{ et } S_2) = \frac{|\{(Noire, Noire)\}|}{|\Omega|} = \frac{1}{4}$. On a donc bien $\mathbb{P}(S_1 \text{ et } S_2) = \mathbb{P}(S_1) \times \mathbb{P}(S_2)$ et donc $S_1 \perp S_2$. Notons que ce résultat peut être obtenu aussi en considérant l'expérience comme composée et en appliquant les techniques vues précédemment.

On constate que c'est avant tout la probabilité sur l'univers qui détermine si ces événements sont indépendants, et non pas simplement leur description en français. ∞

La notion d'indépendance se généralise à plus de deux événements.

Définition 2.3 Soit une expérience aléatoire décrite par l'univers Ω et la probabilité \mathbb{P} . Soit une famille d'événements A_i indexés par un ensemble quelconque d'indices, I (par exemple $I = \{1, 2, \dots, n\}$ pour n événements). On dit que les A_i sont **indépendants dans leur ensemble** si et seulement si pour tout sous-ensemble fini d'indices $J \subset I$,

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j). \quad (2.8)$$

Pour bien comprendre l'impact de cette définition, étudions le cas de trois événements, A_1 , A_2 et A_3 . L'ensemble d'indices est $I = \{1, 2, 3\}$ et on doit donc considérer tous les sous-ensembles de I . Il est clair que le sous-ensemble vide et les sous-ensembles réduits à un seul indice ne sont pas intéressants. Reste donc :

1. $J = \{1, 2\}$ ce qui donne $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \times \mathbb{P}(A_2)$ soit $A_1 \perp A_2$;
2. $J = \{1, 3\}$ ce qui donne $\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1) \times \mathbb{P}(A_3)$ soit $A_1 \perp A_3$;
3. $J = \{2, 3\}$ ce qui donne $\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2) \times \mathbb{P}(A_3)$ soit $A_2 \perp A_3$;
4. et enfin $J = I$ ce qui donne

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \times \mathbb{P}(A_2) \times \mathbb{P}(A_3).$$







En résumé, trois événements sont indépendants *dans leur ensemble* s'ils sont indépendants *deux à deux* mais aussi s'ils vérifient la quatrième condition ci-dessus.

Ceci montre que l'indépendance d'un ensemble d'événements est une condition beaucoup plus forte que l'indépendance deux à deux de tous les événements considérés. L'exemple suivant illustre cette situation.

∞ **Exemple 2.9** Reprenons l'exemple 2.7 qui introduit la notion d'indépendance et considérons l'événement C « le dé noir donne 1 ». Il est clair par symétrie du problème que $C \perp A$ (de la même façon que $B \perp A$) et qu'on a $\mathbb{P}(C) = \frac{1}{6}$. En outre, $B \perp C$. En effet, $B \cap C = \{(\blacksquare, \blacksquare)\}$ et donc par uniformité de la probabilité $\mathbb{P}(B \cap C) = \frac{1}{36}$ qui est bien égale à $\mathbb{P}(B) \times \mathbb{P}(C)$. Les trois événements A , B et C sont donc bien indépendants *deux à deux*. Cependant, l'événement $A \cap B \cap C$ est réduit à $\{(\blacksquare, \blacksquare)\} = B \cap C$ et est donc lui aussi de probabilité $\frac{1}{36}$. Or $\mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C) = \frac{1}{6^3} = \frac{1}{216} \neq \frac{1}{36}$. De ce fait, A , B et C ne sont pas indépendants *dans leur ensemble*. ∞

La notion d'indépendance est intéressante avant tout comme outil de modélisation, situation dans laquelle son emploi est plus naturel et intuitif que pour une expérience fixée. En effet, dans le langage courant, dire que deux objets sont indépendants signifie qu'ils n'ont aucun lien. Comme le montrent les exemples précédents, la situation en probabilité est beaucoup plus subtile puisque deux événements qui sont logiquement liés (obtenir un dé portant un 1 et obtenir deux dés portant des chiffres identiques) peuvent être indépendants statistiquement. Cependant, quand on souhaite modéliser une situation faisant intervenir de l'aléa, il peut être raisonnable de faire l'hypothèse que des objets sans lien conduisent à des événements indépendants, ce qui permettra de construire une probabilité sur l'univers associé. Par exemple, si on lance deux dés distincts, il est relativement raisonnable de supposer que le premier dé est indépendant du second, au sens où tout événement portant uniquement sur le premier dé est indépendant de tout événement portant sur le second dé. L'exemple suivant développe cette idée.

☞ **Exemple 2.10** Supposons donnés deux dés spéciaux. Le premier est un dé à 6 faces truqué. La probabilité de chaque face est donnée par le tableau suivant :

x						
$\mathbb{P}(D_1 = x)$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{3}{20}$

Dans ce tableau, on a noté $D_1 = x$ l'évènement « obtenir la face x en lançant le premier dé ». On notera de la même façon $D_2 = y$ l'évènement « obtenir la face y en lançant le deuxième dé »,

Le deuxième dé est un dé à 4 faces non truqué. On lance les deux dés simultanément et on cherche la probabilité de l'évènement A « obtenir un total de 5 ».

Les deux dés étant distincts, il est naturel de supposer que les événements qui les concernent exclusivement sont indépendants. À tout tirage de la forme $D_1 = x$ et $D_2 = y$, cette hypothèse associe alors la probabilité $\mathbb{P}(D_1 = x) \times \mathbb{P}(D_2 = y)$, en vertu de $\{D_1 = x\} \perp \{D_2 = y\}$. Comme le dé à quatre faces n'est pas truqué, on a en outre $\mathbb{P}(D_2 = y) = \frac{1}{4}$ pour tout y . Il ne reste alors plus qu'à exprimer A sous forme d'une union d'évènements élémentaires, soit

$$A = \{(\square, 4), (\square, 3), (\square, 2), (\square, 1)\}.$$

La probabilité de A est alors la somme des probabilités des paires, soit

$$\begin{aligned} \mathbb{P}(A) &= \frac{1}{4} \left(\frac{1}{4} + \frac{1}{10} + \frac{1}{20} + \frac{1}{5} \right), \\ &= \frac{3}{20}. \end{aligned}$$

☞

2.7 Indépendance conditionnelle

La notion d'indépendance simple est parfois trop grossière : dans certaines situations, la dépendance entre deux événements A et B n'est pas « complète » au sens où il existe un troisième événement C dont la réalisation avérée rend les deux événements A et B indépendants, alors qu'ils ne le sont pas en général. La notion d'indépendance

Considérons maintenant l'évènement $A \cap B \cap C$. En étudiant le tableau qui donne $A \cap B$, on constate que $|A \cap B \cap C| = 9$, et donc que $\mathbb{P}(A \cap B \cap C) = \frac{9}{216} = \frac{1}{24}$. On a donc $\mathbb{P}(A \cap B|C) = \frac{1}{4}$. On remarque qu'on a donc finalement

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \times \mathbb{P}(B|C).$$

Or, $\mathbb{P}(\cdot|C)$ est une probabilité, et donc, au sens de cette probabilité, A et B sont indépendants. Intuitivement, cela ne semble pas choquant. En effet, si on ne sait rien sur le résultat de l'expérience, A et B ne peuvent pas être indépendants : si on apprend que A a eu lieu, on en déduit quelque chose sur B . On s'attend à ce que B soit plus probable puisque la réalisation de A implique que le dé noir a une valeur inférieure ou égale à trois, ce qui réduit les chances d'obtenir un couple (noir, blanc) qui n'est pas compatible avec B .

En revanche, on sent bien que la dépendance entre A et B s'explique par le dé noir. Donc si on connaît le résultat du dé noir (ce qui est le cas quand C est réalisé), on ne s'attend plus à avoir une dépendance entre A et B . En d'autres termes, toute l'information sur B disponible dans A est déjà disponible dans C . D'où cette indépendance entre A et B quand C est connu. ∞

Formellement, l'indépendance conditionnelle se définit comme suit.

Définition 2.4 Soit une expérience aléatoire décrite par l'univers Ω et la probabilité \mathbb{P} . Soit A , B et C trois évènements, avec $\mathbb{P}(C) > 0$. On dit que A et B sont **conditionnellement indépendants sachant C** si et seulement si :

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \times \mathbb{P}(B|C). \quad (2.9)$$

On note alors $(A \perp B) | C$.

Comme pour l'indépendance, l'indépendance conditionnelle est très utile comme outil de modélisation. C'est une hypothèse naturelle quand deux phénomènes sont liés à un troisième. Si on observe les deux phénomènes sans connaître le troisième, on peut constater une dépendance entre eux, mais la connaissance du troisième fait disparaître la dépendance. L'exemple suivant illustre ce phénomène.

∞ **Exemple 2.12** On se donne deux urnes U_1 et U_2 contenant chacune une bille blanche et une bille noire. On dispose de plus d'une bille blanche additionnelle. On procède à l'expérience suivante : on lance une pièce équilibrée. Si on obtient face, on place la bille blanche dans l'urne U_1 , sinon on la place dans l'urne U_2 . On tire ensuite une bille dans chaque urne. On étudie les évènements

$$A_i = \text{« on tire une bille blanche dans l'urne } U_i \text{ »}.$$

Intuitivement, A_1 et A_2 ne peuvent pas être indépendants. En effet, si on sait par exemple que A_1 est réalisé, on pense naturellement que A_2 est moins probablement réalisé car l'obtention d'une bille blanche plaide pour la présence de la bille blanche additionnelle dans U_1 . En revanche, si on connaît le résultat du lancer de la pièce, alors A_1 et A_2 semblent naturellement indépendants, car les tirages dans les urnes ne s'influencent pas mutuellement.

Pour modéliser ce problème, commençons par déterminer son univers Ω . Le résultat complet de l'expérience est un triplet précisant le résultat du lancer de la pièce pris dans $\{Pile, Face\}$, ainsi que les billes obtenues. Il est indispensable ici de faire le choix d'un modèle discernable au niveau des billes. L'urne U_1 contient ainsi les billes B_1 et N_1 , alors que l'urne U_2 contient les billes B_2 et N_2 . La bille supplémentaire est notée B_3 . Ω est alors

$$\{Face\} \times \{B_1, B_3, N_1\} \times \{B_2, N_2\} \cup \{Pile\} \times \{B_1, N_1\} \times \{B_2, B_3, N_2\}.$$

Il semble assez naturel de choisir sur Ω une probabilité uniforme, vu le caractère totalement symétrique du problème. On a $|\Omega| = 12$.

L'évènement A_1 s'écrit alors

$$A_1 = \{Face\} \times \{B_1, B_3\} \times \{B_2, N_2\} \cup \{Pile\} \times \{B_1\} \times \{B_2, B_3, N_2\},$$

et donc $|A_1| = 7$. De la même façon, $|A_2| = 7$, et donc $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \frac{7}{12}$. Or

$$A_1 \cap A_2 = \{Face\} \times \{B_1, B_3\} \times \{B_2\} \cup \{Pile\} \times \{B_1\} \times \{B_2, B_3\},$$

soit $|A_1 \cap A_2| = 4$, et donc $\mathbb{P}(A_1 \cap A_2) = \frac{1}{3}$, ce qui montre que A_1 et A_2 , ne sont pas indépendants, comme prévu (notons tout de même que $\frac{1}{3}$ est numériquement très proche de $\frac{49}{144}$). On remarque en particulier, comme prévu, que $\mathbb{P}(A_2|A_1) = \frac{3}{7}$ ce qui est plus faible que $\mathbb{P}(A_2)$.

Notre dernière intuition était l'indépendance entre A_1 et A_2 sachant le résultat de la pièce. Techniquement, on pense donc avoir $(A_1 \perp A_2) | P$ et $(A_1 \perp A_2) | \bar{P}$ où P est l'évènement « la pièce donne Pile ». Or, il est clair que $\mathbb{P}(P) = \frac{1}{2}$. De plus, on a

$$A_1 \cap P = \{Pile\} \times \{B_1\} \times \{B_2, B_3, N_2\},$$

et donc $\mathbb{P}(A_1 \cap P) = \frac{3}{12}$, alors que

$$A_2 \cap P = \{Pile\} \times \{B_1, N_1\} \times \{B_2, B_3\},$$

soit donc $\mathbb{P}(A_2 \cap P) = \frac{4}{12}$. Finalement,

$$A_1 \cap A_2 \cap P = \{Pile\} \times \{B_1\} \times \{B_2, B_3\},$$

et donc $\mathbb{P}(A_1 \cap A_2 \cap P) = \frac{1}{6}$. On obtient donc finalement $\mathbb{P}(A_1|P) = \frac{1}{2}$, $\mathbb{P}(A_2|P) = \frac{2}{3}$ et $\mathbb{P}(A_1 \cap A_2|P) = \frac{1}{3}$, ce qui montre bien que $(A_1 \perp A_2) | P$. Des calculs similaires montrent que $(A_1 \perp A_2) | \bar{P}$.

Or tous ces résultats, parfaitement valides, pouvaient être obtenus en faisant un peu plus explicitement les hypothèses induites par le choix de la probabilité uniforme sur Ω . On peut d'abord appliquer le principe des expériences aléatoires composées et donc dire que $\mathbb{P}(P) = \mathbb{P}(\bar{P}) = \frac{1}{2}$ car la pièce utilisée est équilibrée. Ensuite, on peut faire l'hypothèse naturelle que les tirages dans les deux urnes sont indépendants sachant la composition des urnes, c'est-à-dire sachant le résultat du lancer de la pièce. Cela signifie que pour tout évènement B_i portant uniquement sur le résultat de l'urne i et pour tout évènement C portant uniquement sur la pièce, on a $(B_i \perp B_j) | C$. Enfin, on peut appliquer le principe du conditionnement classique des expériences composées, en calculant donc $\mathbb{P}(B_i | C)$ en fonction de la composition de l'urne U_i induite par l'évènement C .

L'énorme avantage de cette deuxième façon de procéder est qu'elle s'applique tout aussi facilement quand la pièce n'est pas équilibrée, quand les évènements secondaires sont directement spécifiés sous la forme $\mathbb{P}(B_i | C)$, etc. Au contraire, le raisonnement par dénombrement est limité au cas des expériences symétriques. ◊

Chapitre 3

Variables aléatoires

3.1 Introduction

Il est fréquent en pratique que le résultat d'une expérience aléatoire ne soit pas le point intéressant dans un problème concret. On s'intéresse plutôt à un résultat *dérivé* de celui de l'expérience aléatoire. C'est le cas notamment dans les jeux de hasard. Considérons, par exemple, deux joueurs qui s'accordent sur les règles suivantes : le premier joueur lance un dé ; si le résultat est pair, il donne 1 € au second joueur, sinon, le second joueur lui donne 1 €. Du point de vue du premier joueur, le gain est donc soit 1 €, soit -1 €, et c'est cette information qui importe. Or l'expérience aléatoire concerne simplement le lancer du dé et pas dans le transfert d'argent entre joueurs. Celui-ci est de fait *déterministe* quand le résultat du lancer est connu. Par déterministe, on entend que pour un résultat donné de l'expérience aléatoire (ici le lancer du dé), le résultat final (ici le transfert d'argent) est toujours le même.

La notion de variable aléatoire permet de formaliser ce type de situations. On dispose d'une expérience aléatoire et on transforme de façon déterministe son résultat en un autre résultat. Mathématiquement, on représente la partie aléatoire par une probabilité \mathbb{P} sur un univers Ω , et la partie déterministe par une fonction. La combinaison d'un élément aléatoire avec un élément déterministe produit au final un résultat aléatoire, mais la séparation des deux éléments simplifie en général l'analyse. Dans certaines situations, on pourra en outre donner directement la combinaison des deux opérations, ce qui simplifiera la modélisation de certains phénomènes.

3.2 Notions générales

Définition 3.1 Soit une expérience aléatoire d'univers Ω et de probabilité associée \mathbb{P} . Soit un ensemble quelconque W . Une **variable aléatoire** sur (Ω, \mathbb{P}) , X , est une fonction de Ω dans W . X est dite **à valeurs dans** W .

☛ **Exemple 3.1** Reprenons plus formellement la situation de l'introduction : le premier joueur lance un dé. L'expérience aléatoire est donc modélisée par $\Omega = \{\square, \square, \square, \square, \square, \square\}$ muni de la probabilité uniforme (car, sans information spécifique, on suppose le dé non truqué). Soit la fonction G de Ω dans $W = \{-1, 1\}$ définie par $G(\omega) = 1$ si ω est impair

et $G(\omega) = -1$ sinon. On peut représenter G par le tableau suivant :

ω	\square	\square	\square	\square	\square	\square
$G(\omega)$	1	-1	1	-1	1	-1

G représente clairement le gain du premier joueur en fonction du résultat de l'expérience et est donc un cas typique de variable aléatoire. Comme ω , le résultat de l'expérience, est aléatoire (ne peut pas être connu à l'avance), la valeur $G(\omega)$ est elle-même aléatoire, bien que pour toute valeur de ω , $G(\omega)$ soit parfaitement déterminé. ∞

∞ **Exemple 3.2** On place dans une urne des billes numérotées de 1 à 3. L'urne contient une bille portant le chiffre 1, deux portant le chiffre 2 et 3 portant le chiffre 3. L'expérience consiste à tirer une bille au hasard dans l'urne. En considérant les billes discernables, l'univers est

$$\Omega = \{B_1^1, B_2^1, B_2^2, B_3^1, B_3^2, B_3^3\},$$

où la notation B_i^j désigne la j -ième bille portant le chiffre i . Par symétrie, on utilise pour cette expérience la probabilité uniforme \mathbb{P} .

La fonction X de Ω dans $\{1, 2, 3\}$ définie par $X(B_i^j) = i$ est une variable aléatoire sur (Ω, \mathbb{P}) à valeurs dans $\{1, 2, 3\}$. En termes simples, la variable aléatoire donne le chiffre porté par la bille choisie au hasard. ∞

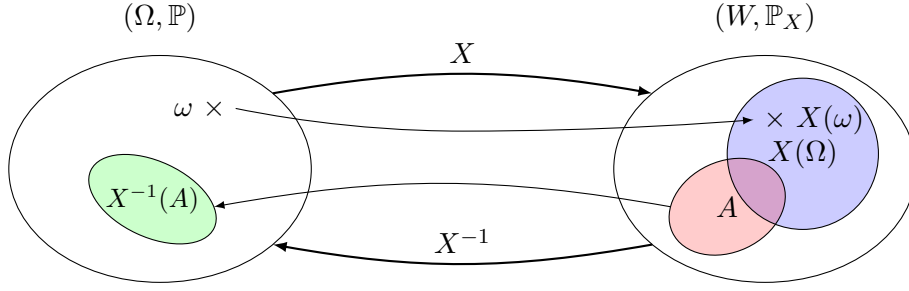
Notons qu'il est fréquent de prendre pour W ensemble « trop grand ». Dans les deux exemples précédents, on pourrait ainsi utiliser $W = \mathbb{N}$ ou $W = \mathbb{R}$. X est bien une fonction de Ω dans W , mais les valeurs réellement prises par X forment seulement un sous-ensemble de W . On a alors la définition suivante :

Définition 3.2 Soit X une variable aléatoire sur (Ω, \mathbb{P}) et à valeurs dans W . On appelle **support** de X l'ensemble image de Ω par X , noté $X(\Omega)$ et défini par

$$X(\Omega) = \{x \in W \mid \exists \omega \in \Omega, X(\omega) = x\}. \quad (3.1)$$

Le support d'une variable aléatoire X est une caractéristique propre de cette variable : on ne peut pas changer $X(\Omega)$ sans changer la variable elle-même. En revanche, on peut utiliser un ensemble pratique pour W à partir du moment où celui-ci contient $X(\Omega)$. Cependant, prendre un W vraiment « trop grand » peut poser quelques difficultés techniques décrites dans la remarque suivante.

⚠ **Remarque 3.1** Comme indiqué à plusieurs reprises, si l'ensemble Ω n'est pas fini (ou dénombrable), la probabilité \mathbb{P} ne sera définie que sur un sous ensemble de $\mathcal{P}(\Omega)$. Par cohérence, il faudra définir un sous ensemble de $\mathcal{P}(W)$ et on devra assurer que X est compatible avec ces deux sous ensembles. La notion correspondante (la mesurabilité) n'est pas au programme de ce cours. On acceptera donc ici toute fonction de Ω dans W comme une variable aléatoire potentielle.

FIGURE 3.1: Représentation graphique de la définition de \mathbb{P}_X à partir de X et de \mathbb{P} .

Définition 3.3 Soit X une variable aléatoire sur (Ω, \mathbb{P}) et à valeurs dans W . On définit une probabilité sur W , la loi de X , notée \mathbb{P}_X , par

$$\forall A \subset W, \mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in A\}).$$

On note aussi

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A),$$

et pour tout $x \in W$,

$$\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x).$$

La figure 3.1 représente graphiquement cette définition.

Preuve Pour assurer que cette définition est cohérente, il faut montrer que \mathbb{P}_X est bien une probabilité en vérifiant que les deux propriétés de la définition 1.4 sont satisfaites.

Il est clair tout d'abord que $\mathbb{P}_X(W) = 1$. En effet, $X^{-1}(W) = \Omega$ car tout élément de Ω a une image dans W , puisque X est une fonction de Ω dans W . Or comme \mathbb{P} est une probabilité $\mathbb{P}(\Omega) = 1$ ce qui permet de conclure.

Considérons maintenant une suite de sous ensembles de W , les $(A_i)_{i \geq 0}$, disjoints deux à deux. D'après les propriétés de fonctions réciproques (cf la section B.4), on a

$$X^{-1}\left(\bigcup_{i \geq 0} A_i\right) = \bigcup_{i \geq 0} X^{-1}(A_i).$$

Notons $B_i = X^{-1}(A_i)$. Les B_i sont disjoints deux à deux. Prenons en effet $i \neq j$ et supposons par l'absurde que $\omega \in B_i \cap B_j$. Alors, par définition de X^{-1} , $X(\omega) \in A_i$ et $X(\omega) \in A_j$, et donc $X(\omega) \in A_i \cap A_j$. Or les A_i sont disjoints deux à deux, ce qui est contradictoire.

Comme les B_i sont disjoints deux à deux et que \mathbb{P} est une probabilité, la sigma additivité implique que

$$\mathbb{P}\left(\bigcup_{i \geq 0} B_i\right) = \sum_{i \geq 0} \mathbb{P}(B_i).$$

Or, par définition, $\mathbb{P}(B_i) = \mathbb{P}_X(A_i)$ et donc en combinant les résultats, on obtient :

$$\mathbb{P}_X\left(\bigcup_{i \geq 0} A_i\right) = \sum_{i \geq 0} \mathbb{P}_X(A_i),$$

ce qui montre que \mathbb{P}_X est sigma additive. C'est donc bien une probabilité. \square

☞ **Exemple 3.3** Dans l'exemple 3.1, les deux valeurs possibles pour le gain, -1 et 1 , sont clairement équiprobables, car nous avons supposé Ω muni de la probabilité uniforme. De ce fait, on s'attend à ce que \mathbb{P}_G soit uniforme. On peut le vérifier en calculant $\mathbb{P}_G(A)$ pour tout $A \subset \{-1, 1\}$, soit :

$$\begin{aligned}\mathbb{P}_G(\emptyset) &= \mathbb{P}(\emptyset) = 0, \\ \mathbb{P}_G(\{-1, 1\}) &= \mathbb{P}(\Omega) = 1, \\ \mathbb{P}_G(\{-1\}) &= \mathbb{P}(\{\square, \boxtimes, \boxplus\}) = \frac{|\{\square, \boxtimes, \boxplus\}|}{|\Omega|} = \frac{1}{2}, \\ \mathbb{P}_G(\{1\}) &= \mathbb{P}(\{\square, \boxtimes, \boxplus\}) = \frac{|\{\square, \boxtimes, \boxplus\}|}{|\Omega|} = \frac{1}{2}.\end{aligned}$$

☞

☞ **Exemple 3.4** Reprenons l'exemple 3.2 et étudions la loi de X . Comme X est à valeurs dans $\{1, 2, 3\}$ qui est un ensemble fini, la proposition 1.1 indique qu'il suffit de connaître les probabilités des évènements élémentaires pour connaître la probabilité \mathbb{P}_X complètement. Or, on a

$$\begin{aligned}\mathbb{P}_X(\{1\}) &= \mathbb{P}(\{B_1^1\}), \\ &= \frac{1}{6},\end{aligned}$$

par équiprobabilité. De même

$$\mathbb{P}_X(\{2\}) = \mathbb{P}(\{B_2^1, B_2^2\}) = \frac{2}{6} = \frac{1}{3},$$

et

$$\mathbb{P}_X(\{3\}) = \mathbb{P}(\{B_3^1, B_3^2, B_3^3\}) = \frac{3}{6} = \frac{1}{2}.$$

Dans ce type de situations simples, on résume en général la loi par un tableau comme celui-ci :

x	1	2	3
$\mathbb{P}_X(\{x\})$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

☞

Justification des notations

Les notations simplifiées de la forme $\mathbb{P}(X \in A)$ et $\mathbb{P}(X = x)$ facilitent en général les raisonnements, notamment parce qu'elles autorisent les mêmes manipulations que dans le cas d'une probabilité définie directement sur un univers Ω . Ceci est une conséquence de la définition de la loi de X à partir de la fonction réciproque X^{-1} et des propriétés de cette fonction (cf la section B.4), qui font de \mathbb{P}_X une probabilité. On peut donc traduire les propriétés 1.1 de la façon suivante.

Propriétés 3.1 Soit X une variable aléatoire sur (Ω, \mathbb{P}) et à valeurs dans W . La loi \mathbb{P}_X vérifie les propriétés suivantes :

1. pour tout $A \subset W$ tel que $A \cap X(\Omega) = \emptyset$, $\mathbb{P}(X \in A) = 0$;
2. pour tout $A \subset W$ tel que $X(\Omega) \subset A$, $\mathbb{P}(X \in A) = 1$;
3. pour toute suite finie ou dénombrable de sous-ensembles de W deux à deux disjoints $(A_i)_{i \geq 1}$,

$$\mathbb{P}\left(X \in \bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mathbb{P}(X \in A_i);$$

4. pour tout $A \subset W$, $\mathbb{P}(X \notin A) = 1 - \mathbb{P}(X \in A)$;
5. pour tous sous-ensembles de W , A et B ,

$$\mathbb{P}(X \in A \cup B) = \mathbb{P}(X \in A) + \mathbb{P}(X \in B) - \mathbb{P}(X \in A \cap B);$$

6. pour tous sous-ensembles de W , A et B , tels que $A \subset B$, $\mathbb{P}(X \in A) \leq \mathbb{P}(X \in B)$.

Notons que ces propriétés s'appliquent bien sûr aussi à la notation $\mathbb{P}(X = x)$. Par exemple, si X prend ses valeurs dans $W = \mathbb{R}$, on a

$$\mathbb{P}(X \in [0,1]) = \mathbb{P}(X = 0) + \mathbb{P}(X \in]0, \frac{1}{2}]) + \mathbb{P}(X \in]\frac{1}{2}, 1]) + \mathbb{P}(X = 1),$$

pour prendre un exemple parmi d'autres de décomposition de l'ensemble $[0,1]$ en sous-ensembles disjoints.

☛ **Exemple 3.5** Reprenons l'exemple 3.2 (cf le calcul de la loi \mathbb{P}_X dans l'exemple 3.4). Cherchons la probabilité $\mathbb{P}(X \in \{2, 3\})$. On peut calculer cette probabilité de trois façons différentes :

1. une première solution consiste à traduire $X \in \{2, 3\}$ directement en un évènement de Ω . On constate en effet que pour obtenir 2 ou 3, il faut obtenir une des billes portant le numéro 2 ou une des billes portant le numéro 3. Il y a 5 billes de ce type, ce qui donne une probabilité de $\frac{5}{6}$ (par uniformité) ;
2. une deuxième solution passe par le calcul de $X^{-1}(\{2, 3\})$ afin d'appliquer la définition de \mathbb{P}_X . On trouve

$$X^{-1}(\{2, 3\}) = \{B_2^1, B_2^2, B_3^1, B_3^2, B_3^3\},$$

et donc de nouveau $\mathbb{P}(X \in \{2, 3\}) = \frac{5}{6}$;

3. enfin, on peut appliquer les propriétés 3.1, ce qui donne

$$\mathbb{P}(X \in \{2, 3\}) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3).$$

En utilisant la loi trouvée à l'exemple 3.4, on en déduit encore une fois que $\mathbb{P}(X \in \{2, 3\}) = \frac{5}{6}$.

Notons que dans les trois approches, il aurait bien sûr été plus efficace de calculer $\mathbb{P}(X = 1)$ comme complémentaire de $\mathbb{P}(X \in \{2, 3\})$, toujours en s'appuyant sur les propriétés 3.1.☛

3.3 Variable aléatoire numérique

Quand une variable aléatoire X est à valeurs dans $W = \mathbb{R}$, on parle de variable aléatoire réelle ou numérique. Le terme *réelle* est généralement utilisé même si le support de X est constitué de nombres entiers, en vertu du fait que si $X(\Omega)$ est un sous-ensemble de \mathbb{R} , on peut toujours poser $W = \mathbb{R}$.

Quand X est numérique, on peut introduire des notions supplémentaires qui caractérisent de différentes façons son comportement. Dans le présent chapitre, nous nous contenterons de la notion de fonction de répartition, après avoir introduit de nouvelles notations.

Définition 3.4 Soit X une variable aléatoire sur (Ω, \mathbb{P}) et à valeurs dans \mathbb{R} . Pour tout $t \in \mathbb{R}$, on note

$$\begin{aligned}\mathbb{P}(X \leq t) &= \mathbb{P}_X(]-\infty, t]), \\ \mathbb{P}(X < t) &= \mathbb{P}_X(]-\infty, t[), \\ \mathbb{P}(X \geq t) &= \mathbb{P}_X([t, +\infty[), \\ \mathbb{P}(X > t) &= \mathbb{P}_X(]t, +\infty[).\end{aligned}$$

Fonction de répartition

Définition 3.5 Soit X une variable aléatoire sur (Ω, \mathbb{P}) et à valeurs dans \mathbb{R} . On appelle *fonction de répartition* de X la notion F_X de \mathbb{R} dans $[0,1]$ définie par

$$\forall t \in \mathbb{R}, F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}_X(]-\infty, t]). \quad (3.2)$$

Le calcul de la fonction de répartition est en général assez mécanique, comme le montre l'exemple suivant.

∞ **Exemple 3.6** On lance deux dés à quatre faces non truqués et on appelle S la variable aléatoire donnant la somme des valeurs obtenues. On considère les deux dés comme discernables, ce qui donne pour l'univers

$$\Omega = \{1, 2, 3, 4\}^2.$$

Les dés n'étant pas truqués, on suppose que la probabilité \mathbb{P} est uniforme.

On constate de plus que

$$S(\Omega) = \{2, 3, 4, 5, 6, 7, 8\}.$$

En effet, les faces sont numérotés de 1 à 4 et la somme peut donc prendre toutes les valeurs entre 2 au minimum et 8 au maximum. Ceci permet de calculer facilement $F_S(t)$ pour certaines valeurs de t . En effet, comme la plus petite valeur de S est 2, on a clairement

$$\forall t < 2, F_S(t) = \mathbb{P}(S \leq t) = 0,$$

en appliquant la première propriété 3.1. De même, en appliquant la deuxième propriété, on constate que

$$\forall t \geq 8, F_S(t) = \mathbb{P}(S \leq t) = 1,$$

car la somme est toujours plus petite ou égale à 8.

Considérons enfin $s \in S(\Omega)$. En appliquant toujours les propriétés 3.1, on remarque que

$$\mathbb{P}(S \leq s + u) = \mathbb{P}(S \leq s) + \mathbb{P}(S \in]s, s + u]).$$

Or, si $u \in [0,1[$, $\mathbb{P}(S \in]s, s + u]) = 0$ car S ne prend que des valeurs entières et que $s \in S(\Omega)$. Comme s est entier et que $u < 1$, il n'y a aucun entier dans l'intervalle $]s, s + u]$, ce qui donne cette probabilité nulle. Cette remarque montre qu'il suffit en fait de calculer $\mathbb{P}(S \leq s) = F_S(s)$ pour tout $s \in S(\Omega)$ pour connaître la fonction F_S dans son ensemble.

Pour ce faire, on commence par calculer la loi de S . Une façon simple de le faire est d'expliciter par un tableau la définition de S . On construit le tableau suivant dans lequel chaque ligne correspond à un résultat pour le premier dé, chaque colonne à un résultat pour le second dé et chaque case à la valeur correspondante pour S :

D_1/D_2	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

Par hypothèse, \mathbb{P} est uniforme et donc chaque case à la même probabilité $\frac{1}{|\Omega|} = \frac{1}{16}$. Donc pour tout $s \in S(\Omega)$, $\mathbb{P}(S = s)$ est obtenue en comptant le nombre de cases contenant s et en le multipliant par $\frac{1}{16}$. On obtient ainsi la loi de S :

s	2	3	4	5	6	7	8
$\mathbb{P}(S = s)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Pour obtenir les valeurs de $\mathbb{P}(S \leq s)$, il suffit alors de réaliser la somme $\sum_{k \in S(\Omega), k \leq s} \mathbb{P}(S = k)$, puisque que le seul moyen d'obtenir par exemple $S \leq 3$ est d'obtenir $S = 2$ ou $S = 3$. On obtient ainsi

s	2	3	4	5	6	7	8
$\mathbb{P}(S \leq s)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{6}{16}$	$\frac{10}{16}$	$\frac{13}{16}$	$\frac{15}{16}$	$\frac{16}{16}$

En combinant ce résultat avec les remarques précédentes, on peut enfin donner F_S grâce au tableau suivant :

si $t \in$	$] - \infty, 2[$	$[2, 3[$	$[3, 4[$	$[4, 5[$	$[5, 6[$	$[6, 7[$	$[7, 8[$	$[8, + \infty[$	∞
$F_S(t)$	0	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{6}{16}$	$\frac{10}{16}$	$\frac{13}{16}$	$\frac{15}{16}$	$\frac{16}{16}$	

Le graphe de F_S est donné par la figure 3.2.

Les fonctions de répartition possèdent un ensemble de propriétés fondamentales détaillées ci-dessous.

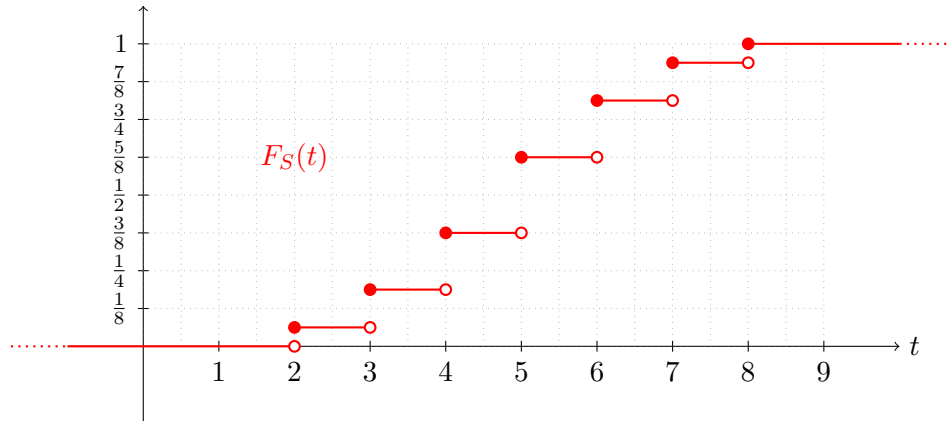


FIGURE 3.2: Graphe de la fonction de répartition F_S de l'exemple 3.6. Les couples cercle/disque indiquent les points de discontinuité de la fonction : le cercle est la limite (ici à gauche) alors que le disque est la valeur prise par la fonction.

Propriétés 3.2 Soit X une variable aléatoire sur (Ω, \mathbb{P}) à valeurs dans \mathbb{R} et F_X sa fonction de répartition. La fonction F_X vérifie les quatre propriétés fondamentales suivantes :

1. $\lim_{t \rightarrow -\infty} F_X(t) = 0$;
2. $\lim_{t \rightarrow +\infty} F_X(t) = 1$;
3. F_X est croissante :

$$s \leq t \Rightarrow F_X(s) \leq F_X(t);$$

4. F_X est continue à droite en tout point :

$$\forall t \in \mathbb{R}, \lim_{h \rightarrow 0^+} F_X(t+h) = F_X(t).$$

Preuve Ces propriétés sont des conséquences de la définition de F_X et des propriétés fondamentales des probabilités 1.1. Le point clé est le lien entre la notion de limite numérique et limite ensembliste. On constate tout d'abord que la croissance des probabilités (propriété 5), se traduit naturellement en croissance de F_X . En effet, si $s \leq t$, $] - \infty, s] \subset] - \infty, t]$, et donc

$$F_X(s) = \mathbb{P}_X(] - \infty, s]) \leq \mathbb{P}_X(] - \infty, t]) = F_X(t).$$

D'autre part, la propriété 7 des probabilités 1.1 permet de dire que

$$\lim_{i \rightarrow +\infty} \mathbb{P}_X(] - \infty, -i]) = 0.$$

En effet, la suite des $A_i =] - \infty, -i]$ est décroissante et on constate que $\bigcap_{i \geq 0} A_i = \emptyset$, ce qui permet de conclure. Par croissance de F_X , on en déduit que $\lim_{t \rightarrow -\infty} F_X(t) = 0$. De la même façon,

$$\lim_{i \rightarrow +\infty} \mathbb{P}_X(] - \infty, i]) = 1,$$

car la suite des $B_i =] - \infty, i]$ est croissante et qu'on peut donc lui appliquer la propriété 6 des probabilités 1.1. Comme $\bigcup_{i \geq 0} B_i = \mathbb{R}$, on a $\mathbb{P}_X(\bigcup B_i) = 1$.

Enfin, la continuité à droite se prouve de la même façon. Pour un $t \in \mathbb{R}$, on construit la suite des $C_i =] - \infty, t + \frac{1}{i+1}]$ qui est décroissante. On constate que $\bigcap_{i \geq 0} C_i =] - \infty, t]$. En appliquant la propriété 7, on a donc

$$\lim_{i \rightarrow +\infty} \mathbb{P}_X(] - \infty, t + \frac{1}{i+1}]) = \mathbb{P}_X(] - \infty, t]),$$

soit

$$\lim_{i \rightarrow +\infty} F_X\left(t + \frac{1}{i+1}\right) = F_X(t).$$

Par croissance de F_X ceci suffit à garantir que $\lim_{h \rightarrow 0^+} F_X(t+h) = F_X(t)$. \square

Les fonctions de répartition vérifient d'autres propriétés très utiles en pratique.

Propriétés 3.3 Soit X une variable aléatoire sur (Ω, \mathbb{P}) à valeurs dans \mathbb{R} et F_X sa fonction de répartition. La fonction F_X vérifie les propriétés suivantes :

1. $\mathbb{P}(X > x) = 1 - F_X(x)$
2. $\mathbb{P}(X \in]a, b]) = F_X(b) - F_X(a)$
3. $\mathbb{P}(X = t) = F_X(t) - \lim_{h \rightarrow 0^+} F_X(t-h)$

Preuve Les deux premières propriétés sont élémentaires et sont des conséquences immédiates de la définition de F_X . Pour la première, on a $\mathbb{P}(X > x) = \mathbb{P}(X \in]x, \infty[)$. Comme \mathbb{R} est l'union disjointe de $]x, \infty[$ et $] - \infty, x]$, les propriétés 3.1 donnent

$$1 = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(X \leq x) + \mathbb{P}(X > x).$$

On reconnaît $F_X(x)$ comme premier terme de cette somme ce qui permet de conclure.

De même $] - \infty, b] =] - \infty, a] \cup]a, b]$ et cette union est disjointe. Donc d'après les propriétés 3.1, on a

$$\mathbb{P}(X \in] - \infty, b]) = \mathbb{P}(X \in] - \infty, a]) + \mathbb{P}(X \in]a, b]),$$

soit donc, par définition de F_X ,

$$F_X(b) = F_X(a) + \mathbb{P}(X \in]a, b]).$$

La troisième propriété se prouve en utilisant une limite ensembliste. Pour un $t \in \mathbb{R}$ fixé, on a en effet :

$$\bigcap_{i \geq 0} \left] t - \frac{1}{i+1}, t \right] = \{t\},$$

et la suite des ensembles $\left] t - \frac{1}{i+1}, t \right]$ est décroissante. Donc d'après les propriétés 1.1

$$\lim_{i \rightarrow \infty} \mathbb{P}_X\left(\left] t - \frac{1}{i+1}, t \right]\right) = \mathbb{P}_X(\{t\}).$$

Or on vient de montrer que

$$\mathbb{P}_X\left(\left] t - \frac{1}{i+1}, t \right]\right) = F_X(t) - F_X\left(t - \frac{1}{i+1}\right),$$

ce qui permet de conclure que

$$\lim_{i \rightarrow \infty} F_X \left(t - \frac{1}{i+1} \right) = F_X(t) - \mathbb{P}(X = t).$$

Comme F_X est croissante, on peut en déduire que cette propriété est vraie en général pour la limite à gauche, et donc que

$$\lim_{h \rightarrow 0^+} F_X(t - h) = F_X(t) - \mathbb{P}(X = t),$$

ce qui permet de conclure. \square

⚠ Remarque 3.2 *En combinant les propriétés de F_X , on peut calculer toutes sortes de probabilités à partir simplement de cette fonction. Par exemple, en remarquant que l'union disjointe $[a, b] \cup \{b\}$ est égale à l'union disjointe $\{a\} \cup]a, b]$, on obtient*

$$\begin{aligned} \mathbb{P}(X \in [a, b]) &= \mathbb{P}(X \in]a, b]) - \mathbb{P}(X = b) + \mathbb{P}(X = a), \\ &= F_X(b) - F_X(a) - \left(F_X(b) - \lim_{h \rightarrow 0^+} F_X(b - h) \right) \\ &\quad + \left(F_X(a) - \lim_{h \rightarrow 0^+} F_X(a - h) \right), \\ &= \lim_{h \rightarrow 0^+} F_X(b - h) - \lim_{h \rightarrow 0^+} F_X(a - h). \end{aligned}$$







L'ensemble des résultats obtenus ci-dessus montrent que la fonction de répartition d'une variable aléatoire X caractérise \mathbb{P}_X . Le théorème suivant précise ce résultat (sa preuve est hors programme).

Théorème 3.1 *Soit deux variables aléatoires réelles X et Y (donc à valeurs dans \mathbb{R}). On suppose que F_X et F_Y sont identiques, c'est-à-dire que pour tout $t \in \mathbb{R}$, $F_X(t) = F_Y(t)$. Alors X et Y sont de même loi ($\mathbb{P}_X = \mathbb{P}_Y$), c'est-à-dire que pour toute partie de \mathbb{R} , A*

$$\begin{aligned} \mathbb{P}_X(A) &= \mathbb{P}_Y(A), \\ \mathbb{P}(X \in A) &= \mathbb{P}(Y \in A). \end{aligned}$$

⚠ Remarque 3.3 *Attention, il ne faut pas confondre $\mathbb{P}_X = \mathbb{P}_Y$ et $X = Y$. Il est clair bien sûr que si $X = Y$, alors $\mathbb{P}_X = \mathbb{P}_Y$. En revanche, on peut construire facilement deux variables aléatoires de même loi mais distinctes, comme le montre l'exemple suivant.*

∞ Exemple 3.7 Reprenons l'exemple 3.1 et introduisons deux variables aléatoires. La variable G , étudiée dans l'exemple 3.1, est le gain du premier joueur. La variable H est le gain du second joueur. Les deux variables sont définies par le tableau suivant :

ω						
$G(\omega)$	1	-1	1	-1	1	-1
$H(\omega)$	-1	1	-1	1	-1	1

Il est clair que $H \neq G$, car, par exemple, $H(\square) = -G(\square)$ (on a en général $H = -G$).

Nous avons obtenu la loi de G dans l'exemple 3.3. En appliquant le même raisonnement, on constate ici que

$$\mathbb{P}_H(\{1\}) = \mathbb{P}(\{\square, \boxtimes, \boxplus\}) = \frac{|\{\square, \boxtimes, \boxplus\}|}{|\Omega|} = \frac{1}{2},$$

et que

$$\mathbb{P}_H(\{-1\}) = \mathbb{P}(\{\square, \boxtimes, \boxminus\}) = \frac{|\{\square, \boxtimes, \boxminus\}|}{|\Omega|} = \frac{1}{2}.$$

On constate ainsi que $\mathbb{P}_H(\{1\}) = \mathbb{P}_G(\{1\})$ et $\mathbb{P}_H(\{-1\}) = \mathbb{P}_G(\{-1\})$, puis plus généralement que $\mathbb{P}_H = \mathbb{P}_G$. ∞

Théorème de la réciproque

Les propriétés 3.2 caractérisent totalement les fonctions de répartition. D'une part, toutes les fonctions de répartition doivent les vérifier. D'autre part, comme l'indique le théorème suivant, toute fonction qui vérifie ces quatre propriétés est la fonction de répartition d'une certaine variable aléatoire.

Théorème 3.2 *Soit F une fonction de \mathbb{R} dans $[0,1]$ vérifiant les quatre propriétés 3.2, c'est-à-dire telle que :*

1. $\lim_{t \rightarrow -\infty} F(t) = 0$;
2. $\lim_{t \rightarrow +\infty} F(t) = 1$;
3. F est croissante ;
4. F est continue à droite en tout point.

Alors il existe un univers Ω muni de la probabilité \mathbb{P} et une variable aléatoire X sur (Ω, \mathbb{P}) et à valeurs dans \mathbb{R} telle que F soit la fonction de répartition de X , soit donc $\mathbb{P}(X \leq t) = F(t)$ pour tout $t \in \mathbb{R}$.

La preuve de ce théorème est (largement) hors programme. En revanche, le théorème lui-même est très pratique pour construire des variables aléatoires, comme l'illustrent les exemples suivants.

∞ **Exemple 3.8** Un exemple simple est donné par la fonction F de \mathbb{R} dans $[0,1]$ définie par :

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ \frac{1}{2} & \text{si } x \in [0,1[, \\ 1 & \text{si } x \geq 1. \end{cases}$$

F vérifie clairement les propriétés du théorème 3.2 :

1. $F(x) = 0$ pour $x < 0$, donc $\lim_{x \rightarrow -\infty} F(x) = 0$;
2. $F(x) = 1$ pour $x \geq 1$, donc $\lim_{x \rightarrow +\infty} F(x) = 1$;
3. F est croissante puisqu'elle vaut 0 sur $] -\infty, 0[$, puis $\frac{1}{2}$ sur $[0,1[$ puis 1 sur $[1, \infty[$;

4. et enfin F est continue à droite en tout point puisqu'elle est constante sur les intervalles indiqués précédemment et que ces intervalles sont fermés à gauche.

Soit donc X une variable aléatoire de fonction de répartition F . D'après les propriétés 3.3, $\mathbb{P}(X = 0) = \frac{1}{2}$ et $\mathbb{P}(X = 1) = \frac{1}{2}$. En effet, comme F est nulle sur $]-\infty, 0[$, $\lim_{x \rightarrow 0+} F(-x) = 0$ et donc $\mathbb{P}(X = 0) = F(0) - \lim_{x \rightarrow 0+} F(-x) = \frac{1}{2}$. De même, $\lim_{x \rightarrow 0+} F(1-x) = \frac{1}{2}$ car F est constante et égale à 1 sur $[0, 1[$. En appliquant le même résultat avec la limite, on obtient $\mathbb{P}(X = 1) = \frac{1}{2}$.

Il est clair alors que pour tout $A \subset \mathbb{R}$ tel que $0 \notin A$ et $1 \notin A$, $\mathbb{P}(X \in A) = 0$. En effet, $\mathbb{P}(X \in A) = 1 - \mathbb{P}(X \in \overline{A})$. Mais $\{0, 1\} \subset \overline{A}$ et donc $\mathbb{P}(X \in \overline{A}) \geq \mathbb{P}(X \in \{0, 1\}) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1)$. Or cette dernière somme vaut 1 et donc $\mathbb{P}(X \in \overline{A}) = 1$ puis $\mathbb{P}(X \in A) = 0$.

En pratique, on peut donc considérer que la variable X ne prend que les valeurs 0 et 1. En outre ces valeurs sont équiprobables. Nous verrons dans le chapitre suivant que X est une variable discrète. ∞

Le deuxième exemple suivant est beaucoup plus complexe et illustre la puissance du théorème de la réciproque.

∞ **Exemple 3.9** Étudions la fonction F de \mathbb{R} dans $[0, 1]$ définie par :

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ \frac{1}{2} & \text{si } x = 0, \\ \frac{x+1}{2} & \text{si } x \in]0, 1[, \\ 1 & \text{si } x \geq 1, \end{cases}$$

et est représentée sur la figure 3.3. On constate que F vérifie bien les propriétés 1 et 2 du théorème 3.2 car F est constante en dehors de l'intervalle $[0, 1]$ et qu'elle vaut 0 sur $]-\infty, 0[$ et 1 sur $[1, \infty[$.

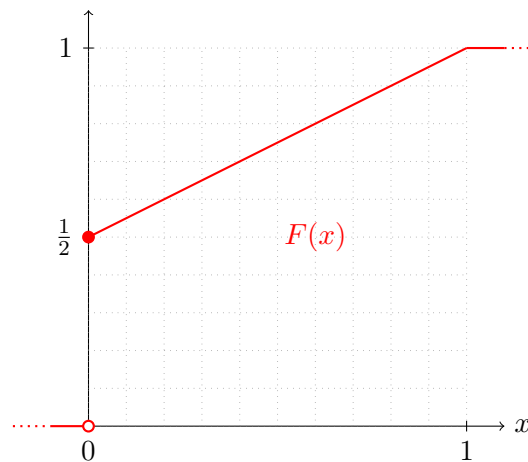


FIGURE 3.3: Graphe de la fonction F de l'exemple 3.9. On représente la discontinuité comme dans la figure 3.2.

On constate aussi que F est continue à droite. En effet, sur chacun des intervalles $]-\infty, 0[$, $]0, 1[$ et $[1, \infty[$, F est soit constante, soit affine et est donc continue. Comme $[1, \infty[$

est fermé à gauche, F est continue à droite en 1. Il reste donc à étudier la situation en 0. Or, on constate que la limite à droite de $x \mapsto \frac{x+1}{2}$ en 0 est $\frac{1}{2}$ par continuité de cette fonction (sur \mathbb{R} tout entier). Comme $F(0) = \frac{1}{2}$, F est bien continue à droite en 0, et donc finalement sur \mathbb{R} tout entier. Notons que F n'est pas continue à gauche en 0 car la limite à gauche en 0 est 0, par continuité de la fonction constante $t \mapsto 0$.

F est aussi croissante sur les intervalles $] - \infty, 0[$, $]0, 1[$ et $[1, \infty[$. En effet elle est constante sur les intervalles infinis et sur l'intervalle $]0, 1[$, sa dérivée est $\frac{1}{2} > 0$. Soit maintenant u et v tels que $u < v$. Si u et v sont dans le même intervalle, $F(u) < F(v)$. Supposons donc que cela ne soit pas le cas. Si u est dans $] - \infty, 0[$ alors $F(u) = 0$ et donc $F(v) \geq F(u)$ car F est toujours positive. Si $u = 0$, alors $F(u) = \frac{1}{2}$. Si $v \in]0, 1[$, $F(v) > \frac{1}{2}$ par croissance (stricte) de F sur $]0, 1[$. Enfin si $v \in [1, \infty[$, $F(v) = 1 > F(u)$. Enfin, si $u \in]0, 1[$, $F(u) \leq 1 = F(v)$ par croissance de F sur $]u, 1[$ et par continuité en 1. F est donc croissante sur \mathbb{R} tout entier.

L'application du théorème 3.2 permet donc de conclure à l'existence d'une variable aléatoire X dont la fonction de répartition est F . En utilisant F , on peut caractériser X :

- on constate que $\mathbb{P}(X = 0) = \frac{1}{2}$. On applique en effet les propriétés 3.3, en particulier

$$\mathbb{P}(X = 0) = F(0) - \lim_{h \rightarrow 0^+} F(-h).$$

Or, on a vu que la limite à gauche en 0 de F est 0. Comme $F(0) = \frac{1}{2}$, on obtient bien la conclusion voulue ;

- on remarque en outre que pour tout $x \in]0, 1[$, $\mathbb{P}(X = x) = 0$. En effet, en appliquant le même résultat, on doit calculer la différence entre $F(x)$ et la limite à gauche de F en x . Or sur $]0, 1[$, F est continue et donc sa limite à gauche en x est $F(x)$, ce qui permet de conclure ;
- on a aussi $\mathbb{P}(X \in]0, 1]) = \frac{1}{2}$ en appliquant de nouveau les propriétés 3.3 ;
- on en déduit que $\mathbb{P}(X \in [0, 1]) = 1$ et donc que si I est tel que $I \cap [0, 1] = \emptyset$, $\mathbb{P}(X \in I) = 0$.

Nous verrons dans les chapitres suivants que X est une variable *mixte*. Elle comporte une partie *discrète* car elle a une probabilité non nulle de valoir exactement 0. Elle comporte une partie *continue* (ou *diffuse*) répartie sur l'intervalle $]0, 1[$. Sur cet intervalle, aucune valeur n'est obtenue exactement avec une probabilité non nulle. En revanche, l'intervalle tout entier à une probabilité de $\frac{1}{2}$. ∞

3.4 Variable aléatoire fonction d'une autre variable aléatoire

Il est assez fréquent en pratique de s'intéresser à un résultat dérivé d'un autre dans une expérience aléatoire, comme le montre l'exemple suivant.

∞ **Exemple 3.10** Au jeu de la roulette, l'expérience aléatoire consiste à lancer une bille dans une roue en rotation contenant des cases. Après une phase de rotation, la bille s'arrête dans une case. Le résultat direct de l'expérience est le numéro de la case (entre 0 et 36 dans la roulette française). Il est naturel de considérer que l'univers Ω est l'ensemble des entiers compris entre 0 et 36 (inclus) et que la probabilité est uniforme sur Ω .

Trois variables aléatoires importantes sont définies sur Ω : la couleur C , la parité O et la hauteur H du résultat. Les deux dernières ont une définition simple. La parité O vaut Pair si le résultat est pair et non nul, Impair si le résultat est impair, et Zéro dans le cas particulier du tirage 0. La hauteur H vaut Manque si le résultat est entre 1 et 18, Passe s'il est entre 19 et 36, et toujours Zéro pour le cas particulier du 0. La couleur C prend trois valeurs, Vert (pour le 0), Rouge et Noir (il n'y a pas de définition simple, seule une liste donne les valeurs).

À partir de ces trois variables (et du résultat de l'expérience), on peut définir de nombreuses autres variables correspondant différents paris autorisés à la roulette. Un pari classique est celui de « chance simple » : le joueur choisit une des valeurs des trois variables, à l'exception du Zéro et de la couleur Vert (il peut donc choisir Passe, Manque, Pair, Impair, Rouge ou Noir). Si la variable aléatoire correspondante prend la valeur choisie, le joueur gagne 1 €, sinon il perd 1 €. Considérons par exemple la variable aléatoire M qui donne le gain du joueur quand celui-ci mise sur Manque. Les valeurs prises par M sont par définition $\{-1, 1\}$. Le résultat est déduit de la valeur de la variable aléatoire H . Tout l'enjeu est maintenant de donner un sens mathématique précis à cette « déduction ». ∞

Le mécanisme général de construction d'une variable à partir d'une autre est le suivant.

Définition 3.6 Soit (Ω, \mathbb{P}) une expérience aléatoire et X une variable aléatoire sur (Ω, \mathbb{P}) à valeurs dans W . V un ensemble et ϕ une fonction de W dans V . La fonction Y de Ω dans V définie par

$$\forall \omega \in \Omega, Y(\omega) = \phi(X(\omega)), \quad (3.3)$$

est une variable aléatoire sur (Ω, \mathbb{P}) à valeurs dans V . On dit que Y est une variable aléatoire **fonction** de X . Pour simplifier la définition de Y , on note $Y = \phi(X)$. On peut aussi écrire $Y = \phi \circ X$ en utilisant la notation classique \circ pour la composition des fonctions.

\triangle **Remarque 3.4** Comme rappelé lors de la remarque 3.1, si Ω n'est pas dénombrable (ou fini), on doit introduire une notion de mesurabilité dans la définition d'une variable aléatoire. Cette notion devrait aussi être utilisée ici pour s'assurer que Y est bien une variable aléatoire.

∞ **Exemple 3.11** Reprenons l'exemple 3.10 de façon un peu plus formelle. Nous avons notamment introduit la variable aléatoire H de $\Omega = \{0, \dots, 36\}$ dans $\{\text{Manque}, \text{Passe}, \text{Zéro}\}$. Nous souhaitons définir la variable aléatoire M , à valeurs dans $\{-1, 1\}$ qui donne le gain du joueur qui mise sur Manque. Il suffit pour cela de considérer la fonction ϕ de $\{\text{Manque}, \text{Passe}, \text{Zéro}\}$ dans $\{-1, 1\}$ telle que $\phi(\text{Manque}) = 1$ et $\phi(\text{Passe}) = \phi(\text{Zéro}) = -1$. Alors d'après la définition ci-dessus, M donnée par $M = \phi(H)$ est bien une variable aléatoire. ∞

La loi d'une variable aléatoire Y obtenue à partir d'une variable aléatoire X est reliée directement à la loi de X , comme le montre la proposition suivante.

3.4. VARIABLE ALÉATOIRE FONCTION D'UNE AUTRE VARIABLE ALÉATOIRE

Proposition 3.1 Soit (Ω, \mathbb{P}) une expérience aléatoire et X une variable aléatoire sur (Ω, \mathbb{P}) à valeurs dans W . V un ensemble et ϕ une fonction de W dans V . La loi de $Y = \phi(X)$ est donnée par

$$\forall A \subset V, \mathbb{P}_Y(A) = \mathbb{P}_X(\phi^{-1}(A)). \quad (3.4)$$

Preuve Soit donc $A \subset V$. Par définition $\mathbb{P}_Y(A) = \mathbb{P}(Y^{-1}(A))$. Or, $Y = \phi \circ X$, donc, d'après les propriétés des fonctions réciproques,

$$Y^{-1}(A) = (\phi \circ X)^{-1}(A) = X^{-1}(\phi^{-1}(A)),$$

et donc

$$\begin{aligned} \mathbb{P}_Y(A) &= \mathbb{P}(X^{-1}(\phi^{-1}(A))), \\ &= \mathbb{P}_X(\phi^{-1}(A)), \end{aligned}$$

par définition de \mathbb{P}_X . □

Le gros intérêt de cette proposition est qu'elle montre que si on peut définir directement une variable aléatoire X (sans passer explicitement par un univers (Ω, \mathbb{P})), on peut alors calculer la loi de toute variable obtenue à partir de X toujours sans passer par l'univers.

☞ **Exemple 3.12** Continuons l'exemple 3.10 (et 3.11). Il est facile de montrer que la loi de H est donnée par

h	Manque	Passe	Zéro
$\mathbb{P}(H = h)$	$\frac{18}{37}$	$\frac{18}{37}$	$\frac{1}{37}$

Déterminons alors la loi de $M = \phi(H)$, avec ϕ définie dans l'exemple 3.11. Comme M est à valeurs dans $\{-1, 1\}$, il suffit de calculer $\mathbb{P}_M(\{1\})$ et $\mathbb{P}_M(\{-1\})$. D'après la propriété ci-dessus et la définition de ϕ , on a

$$\begin{aligned} \mathbb{P}_M(\{1\}) &= \mathbb{P}_X(\phi^{-1}(\{1\})) & \mathbb{P}_M(\{-1\}) &= \mathbb{P}_X(\phi^{-1}(\{-1\})), \\ &= \mathbb{P}_X(\{\text{Manque}\}) & &= \mathbb{P}_X(\{\text{Passe, Zéro}\}), \\ &= \frac{18}{37} & &= \frac{19}{37}. \end{aligned} \quad \heartsuit$$

Cas numérique

Quand une variable aléatoire X est à valeurs réelles, on peut réaliser toute sorte de calculs sur les valeurs de X , ce qui revient à définir des variables aléatoires fonction de X . La convention qui consiste à écrire $Y = \phi(X)$ devient dans ce contexte très expressive, comme le montre l'exemple suivant.

☞ **Exemple 3.13** Reprenons l'exemple 3.6 du lancé de deux dés à quatre faces dans lequel on définit la variable aléatoire S , somme des deux dés. Soit maintenant la fonction ϕ de \mathbb{R} dans \mathbb{R} définie par $\phi(x) = \frac{x}{2} - 1$ et Y la variable aléatoire $Y = \phi(S)$. Comme ϕ

est donnée par une formule explicite, on peut appliquer cette formule à la variable S , ce qui revient à écrire

$$Y = \frac{S}{2} - 1.$$

On peut définir ainsi d'autres variables comme $Z = S^2 + 1$, $T = \sqrt{S}$, etc. ∞

Il faut cependant bien conserver à l'esprit que cette écriture est une convention qui masque le fait que les objets concernés (par exemple X et Y) sont des fonctions et que l'égalité est à comprendre comme une égalité entre fonctions. Si on écrit par exemple $Y = 2X + 1$, cela signifie en fait que pour tout $\omega \in \Omega$, $Y(\omega) = 2X(\omega) + 1$.

Un autre avantage des variables aléatoires numériques est qu'elles peuvent être données par une fonction de répartition, en vertu du théorème de la réciproque 3.2. Dans certains cas, on peut calculer la fonction de répartition d'une variable Y fonction d'une variable X directement à partir de la fonction de répartition de X . C'est le cas dans l'exemple suivant.

∞ **Exemple 3.14** Reprenons l'exemple 3.9 et une variable aléatoire X obtenue en appliquant le théorème de la réciproque. Soit maintenant $Y = 2X + 1$. On cherche à déterminer F_Y . D'après la définition de F_Y et la proposition 3.1, on a

$$F_Y(t) = \mathbb{P}_Y(\cdot - \infty, t] = \mathbb{P}_X(\phi^{-1}(\cdot - \infty, t]),$$

avec ϕ donnée par $\phi(x) = 2x + 1$. Il est clair que

$$\phi^{-1}(\cdot - \infty, t] = \left] -\infty, \frac{(t-1)}{2} \right],$$

car ϕ est bijective et sa fonction inverse est donnée par $\phi^{-1}(t) = \frac{(t-1)}{2}$. On a donc

$$F_Y(t) = \mathbb{P}_X \left(\left] -\infty, \frac{(t-1)}{2} \right] \right) = F_X \left(\frac{(t-1)}{2} \right),$$

ce qui permet de définir explicitement F_Y par

$$F_Y(t) = \begin{cases} 0 & \text{si } t < 1, \\ \frac{1}{2} & \text{si } t = 1, \\ \frac{t+1}{4} & \text{si } t \in]1, 3[, \\ 1 & \text{si } t \geq 3. \end{cases}$$

∞

Chapitre 4

Variables aléatoires discrètes

4.1 Définition

On étudie dans ce chapitre un cas particulier de variables aléatoires pour lesquelles certains concepts et calculs sont plus faciles à définir et réaliser que dans le cas général. Ces variables peuvent prendre un nombre « raisonnable » de valeurs, selon la définition suivante :

Définition 4.1 Soit X une variable aléatoire sur (Ω, \mathbb{P}) et à valeurs dans un ensemble quelconque W . X est dite **discrète** si son support est fini ou dénombrable.

Quand X est discrète et $|X(\Omega)| = n$, on note $(x_i)_{1 \leq i \leq n}$ les valeurs du support. Quand $X(\Omega)$ est dénombrable, on note $(x_i)_{i \geq 1}$ ces valeurs.

Le cas le plus simple est bien sur celui où $|X(\Omega)| < \infty$ car le cas dénombrable est associé à toutes les subtilités induites par le passage à l'infini. Le point important est cependant la possibilité de numéroter les valeurs du support.

▲ Remarque 4.1 La définition 4.1 sépare l'ensemble W dans lequel la variable aléatoire prend ses valeurs et le support $X(\Omega)$ de la variable. Ce point est très important car les variables aléatoires numériques sont en général à valeurs dans \mathbb{R} qui n'est pas un ensemble dénombrable. Pourtant de nombreuses variables aléatoires numériques sont discrètes ce qui montre que cette caractérisation n'est pas liée à l'ensemble « ambiant » mais bien aux valeurs réellement prises par la variable aléatoire.

∞ Exemple 4.1 Considérons le lancer d'un dé à six faces non truqué. On définit sur $\Omega = \{\square, \square, \square, \square, \square, \square\}$ la variable aléatoire X qui indique la parité du résultat du lancer et qui est donc à valeurs dans $W = \{\text{pair}, \text{impair}\}$. Comme W est fini, X est nécessairement discrète. Ici on a $X(\Omega) = W$, ce qui simplifie l'analyse. ∞

Une conséquence pratique importante du caractère discret d'une variable est qu'on peut la spécifier directement par l'intermédiaire de sa loi plutôt que comme une fonction, en s'appuyant sur la proposition 1.1. On a ainsi la proposition suivante.

Proposition 4.1 Soit $W = \{w_1, \dots, w_n\}$ un ensemble fini et n réels de $[0, 1]$, $(p_i)_{1 \leq i \leq n}$ tels que $\sum_{i=1}^n p_i = 1$. Alors il existe un univers Ω , une probabilité \mathbb{P} sur Ω et une variable aléatoire X sur (Ω, \mathbb{P}) à valeurs dans W telle que

$$\forall i, 1 \leq i \leq n, \mathbb{P}(X = w_i) = p_i.$$

Preuve Cette proposition est assez simple à prouver. Il suffit en effet d'évoquer la proposition 1.1 qui garantit l'existence d'une probabilité \mathbb{P} sur W telle que $\mathbb{P}(\{w_i\}) = p_i$ pour tout i . On pose alors $\Omega = W$ et on prend pour X la fonction identité de W dans lui-même qui à tout w associe w (et donc $X(w) = w$). Il est clair que pour tout sous-ensemble A de W , $X^{-1}(A) = A$ et donc que $\mathbb{P}_X(A) = \mathbb{P}(A)$, ce qui permet de conclure. \square

△ Remarque 4.2 La construction de la variable aléatoire dans la preuve est assez artificielle. Cela s'explique par le caractère discret du problème. Dans cette situation, il y a peu de différences entre une variable aléatoire et une probabilité.

4.2 Entropie et mode

Il semble assez clair que toutes les variables aléatoires ne correspondent pas toutes à un même « niveau de hasard ». Précisons cette idée par un exemple.

∞ Exemple 4.2 Considérons l'ensemble $W = \{a, b, c\}$. La proposition 4.1 permet de construire deux variables aléatoires X et Y dont les lois sont

x	a	b	c
$\mathbb{P}(X = x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

et

y	a	b	c
$\mathbb{P}(Y = y)$	$\frac{5}{6}$	$\frac{1}{12}$	$\frac{1}{12}$

On constate que X est « plus aléatoire » que Y dans le sens suivant : il est possible de parier sur une valeur de W avec une probabilité de gagner bien plus grande dans le cas de Y que dans le cas de X . Il suffit en effet de parier sur a . Dans ce cas, la probabilité de gagner est $\mathbb{P}(Y = a) = \frac{5}{6}$ pour Y et $\mathbb{P}(X = a) = \frac{1}{3}$. De plus, la probabilité de gagner dans le cas de la variable X ne peut pas être plus grande que $\frac{1}{3}$ quel que soit le choix de la valeur. ∞

Plusieurs mesures ont été proposées pour quantifier le niveau d'aléa que présente une variable aléatoire. La mesure la plus générale est l'entropie de Shannon décrite dans la définition suivante.

Définition 4.2 Soit X une variable aléatoire discrète à valeurs dans $W = \{w_1, \dots, w_n\}$. Pour tout réel strictement positif b , l'**entropie** (de Shannon) de X en base b , $H_b(X)$ est la quantité

$$H_b(X) = - \sum_{i=1}^n \mathbb{P}(X = w_i) \log_b \mathbb{P}(X = w_i), \quad (4.1)$$

où \log_b est la fonction logarithme en base b . Par convention, dès que $\mathbb{P}(X = w_i) = 0$ on remplace $\mathbb{P}(X = w_i) \log_b \mathbb{P}(X = w_i)$ par la valeur 0 dans la somme ci-dessus.

On utilise en général $b = 2$ et on parle alors d'entropie, sans préciser la base. On note ainsi $H(X) = H_2(X)$.

L'entropie d'une variable aléatoire mesure sa dispersion, c'est-à-dire sa propension à être le plus aléatoire possible.

☞ **Exemple 4.3** Reprenons les deux variables aléatoires de l'exemple 4.2. Un simple calcul donne

$$H(X) = -\log_2 \left(\frac{1}{3} \right) = \log_2 3 \simeq 1,585$$

$$H(Y) = -\frac{5}{6} \log_2 \left(\frac{5}{6} \right) - \frac{1}{6} \log_2 \left(\frac{1}{12} \right) \simeq 0,8167.$$

L'entropie de X est de l'ordre du double de celle de Y . La mesure numérique de dispersion est ainsi en accord avec l'analyse intuitive basée sur la probabilité de gagner un pari. ☞

Propriétés 4.1 L'entropie vérifie quelques propriétés intéressantes :

1. pour toute variable aléatoire discrète X et tout $b > 0$, $H_b(X) \geq 0$;
2. si une variable aléatoire discrète X à valeurs dans W est telle que $H_b(X) = 0$ (pour un $b > 0$) alors il existe $w \in W$ tel que $\mathbb{P}(X = w) = 1$;
3. pour toute variable aléatoire discrète X à valeurs dans W avec $n = |W|$, $H_b(X) \leq \log_b(n)$;
4. si une variable aléatoire discrète X à valeurs dans W est telle que $H_b(X) = \log_b(n)$ (pour un $b > 0$) alors pour tout $w \in W$, $\mathbb{P}(X = w) = \frac{1}{n}$ (en d'autres termes, \mathbb{P}_X est uniforme sur W).

La notion d'entropie peut être complétée par celle de **mode**.

Définition 4.3 Soit X une variable aléatoire discrète à valeurs dans $W = \{w_1, \dots, w_n\}$. On dit que w est le **mode** de X si et seulement si

$$\forall w' \in W, \mathbb{P}(X = w') < \mathbb{P}(X = w).$$

En d'autres termes, le mode de X la valeur prise de façon la plus probable par X . Si le mode existe, X est dite **unimodale**.

Alors que la notion d'entropie mesure la dispersion d'une variable aléatoire, celle de mode formalise l'idée de tendance dominante. Notons notamment qu'une variable aléatoire d'entropie maximale n'a pas de mode (sauf si elle est constante).

☞ **Exemple 4.4** Reprenons les deux variables aléatoires de l'exemple 4.2. On voit que X n'a pas de mode car les trois valeurs a , b et c sont équiprobables. Aucune valeur n'est donc dominante et la variable n'exhibe pas de tendance. Au contraire, elle a une entropie maximale.

En revanche, Y est unimodale et son mode est a . La valeur a est en effet la plus probable. Comme l'entropie de Y est faible, cela montre que Y a une forte tendance à prendre la valeur a . ☞

4.3 Variable aléatoire discrète numérique

Comme nous l'avons vu dans la section 3.4, quand une variable aléatoire est à valeurs réelles, un outil puissant est disponible, la fonction de répartition. Quand une variable aléatoire numérique est discrète, la fonction de répartition possède une forme particulière.

Propriété 4.2 Soit X une variable aléatoire *réelle et discrète* sur (Ω, \mathbb{P}) , c'est-à-dire à valeurs dans \mathbb{R} et telle que $X(\Omega)$ soit fini ou dénombrable. Alors sa fonction de répartition F_X vérifie les propriétés suivantes (en plus des propriétés générales 3.2) :

1. pour tout $x \in X(\Omega)$ tel que $\mathbb{P}(X = x) > 0$, F_X est discontinue en x ;
2. réciproquement, si F_X est discontinue en x , alors $x \in X(\Omega)$ et $\mathbb{P}(X = x) > 0$;
3. en dehors de ses points de discontinuité, F_X est constante (elle est dite constante par morceaux).

Preuve La preuve de ces propriétés est assez simple car l'essentiel fourni par les propriétés 3.2 et 3.3.

1. Soit donc $x \in X(\Omega)$ avec $\mathbb{P}(X = x) > 0$. Alors d'après la propriété 3.3

$$\mathbb{P}(X = x) = F_X(x) - \lim_{h \rightarrow 0^+} F_X(x - h) > 0.$$

On voit donc que la limite à gauche de F_X en x est différente de la valeur $F_X(x)$, ce qui montre que la fonction est discontinue en x .

2. Soit maintenant la situation inverse, c'est-à-dire un x tel que F_X soit discontinue en x . Comme F_X est une fonction de répartition, ceci n'est possible que parce que la limite à gauche de F_X en x est différente de la valeur $F_X(x)$ (car on a continuité à droite en tout point). On a donc

$$F_X(x) - \lim_{h \rightarrow 0^+} F_X(x - h) > 0,$$

par croissance de F_X . En utilisant de nouveau la propriété 3.3, on en déduit que $\mathbb{P}(X = x) > 0$, ce qui n'est possible que si $x \in X(\Omega)$.

3. Enfin, considérons s et t avec $s < t$ et $F_X(s) < F_X(t)$. Alors, par définition de F_X , $\mathbb{P}(X \in]s, t]) = F_X(t) - F_X(s) > 0$. Or $\mathbb{P}(X \in]s, t]) = \mathbb{P}(X \in]s, t] \cap X(\Omega))$. Comme $X(\Omega)$ est (au plus) dénombrable, $]s, t] \cap X(\Omega)$ est (au plus) dénombrable et peut donc s'écrire $\bigcup_{i \geq 1} \{x_i\}$. Alors par sigma additivité, on a

$$\mathbb{P}(X \in]s, t] \cap X(\Omega)) = \sum_{i \geq 1} \mathbb{P}(X = x_i).$$

Comme cette grandeur est strictement positive, au moins un des x_i est tel que $\mathbb{P}(X = x_i) > 0$. De ce fait, il y a donc au moins une discontinuité sur l'intervalle $]s, t]$. Par contra-position, s'il n'y a pas de discontinuité sur $]s, t]$, c'est que $F_X(s) = F_X(t)$ et donc que F_X est constante sur l'intervalle. \square

☞ **Exemple 4.5** L'exemple 3.6 et la figure 3.2 sont typiques du cas d'une fonction de répartition pour une variable aléatoire discrète. On voit notamment très bien sur la figure le caractère constant par morceaux de F_X . ☞

4.4 Moments

Nous avons vu dans la section 4.2 deux façons de caractériser une variable aléatoire, le mode pour la tendance centrale et l'entropie pour la dispersion. Dans le cas discret numérique, d'autres mesures sont disponibles. Elles exploitent le caractère numérique de la variable et viennent compléter le mode et l'entropie.

Espérance

Définition 4.4 Soit X une variable aléatoire réelle et discrète sur (Ω, \mathbb{P}) . On appelle *espérance* (mathématique) de X , la valeur numérique notée $\mathbb{E}(X)$ définie par

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x), \quad (4.2)$$

quand elle existe.

⚠ **Remarque 4.3** La série qui apparaît à droite dans l'équation (4.2) est bien définie car $X(\Omega)$ est au plus dénombrable. On peut toujours l'écrire sous la forme suivante :

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i),$$

quand la suite $(x_i)_{i \geq 1}$ donne $X(\Omega)$.

Le problème est que cette série ne converge pas toujours et donc que l'espérance n'est pas toujours définie, d'où l'expression « quand elle existe » utilisée à la fin de la définition. Dans le présent cours, on travaille beaucoup dans le cas fini pour lequel l'espérance existe toujours. Dans les exemples du cas dénombrable, on supposera toujours que l'espérance existe.

☞ **Exemple 4.6** Considérons le lancer d'un dé à six faces non truqué. On définit sur $\Omega = \{\square, \square, \square, \square, \square, \square\}$ la variable aléatoire X qui indique la valeur de la face. On a donc

$X(\Omega) = \{1, 2, 3, 4, 5, 6\}$. Il est évident que $\mathbb{P}(X = x) = \frac{1}{6}$ pour tout $x \in X(\Omega)$. On a donc

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x=1}^6 x\mathbb{P}(X = x), \\ &= \frac{1}{6} \sum_{x=1}^6 x = \frac{1}{6} \frac{7 \times 6}{2}, \\ &= \frac{7}{2}.\end{aligned}$$

L'exemple précédent montre que l'espérance s'apparente à une valeur *moyenne*. Intuitivement, dans cet exemple, on s'attend à avoir aussi souvent chaque valeur entre 1 et 6. La valeur obtenue « en moyenne » correspond donc intuitivement à la moyenne de ces valeurs. De façon plus générale, la formule (4.2) peut être interprétée comme une moyenne pondérée des valeurs prises par la variable aléatoire, les poids étant les probabilités d'obtenir ces valeurs. En ce sens, l'espérance mathématique est une version théorique de la notion de moyenne. Elle caractérise une forme de « position » pour une variable aléatoire. Elle donne aussi une idée de la valeur moyenne obtenue en prenant plusieurs valeurs de la variable aléatoire (de façon indépendante entre chaque répétition) au sous où cette valeur moyenne s'approche de plus en plus de l'espérance quand le nombre de répétitions augmente.

Il est intéressante de comparer la notion d'espérance et celle de mode (dans le cas d'une variable aléatoire numérique). Dans certaines situations, l'espérance d'une variable aléatoire unimodale peut être proche de son mode (voir égale au mode), mais rien ne l'oblige, comme le montre l'exemple suivant.

☞ **Exemple 4.7** Soit la variable aléatoire X à valeurs dans $\{-1, 0, 1, 2\}$ de loi

x	-1	0	1	2
$\mathbb{P}(X = x)$	$\frac{1}{6}$	$\frac{3}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Il est clair que X est unimodale et que son mode est 0. Son espérance est donnée par

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x \in \{-1, 0, 1, 2\}} x\mathbb{P}(X = x), \\ &= -1 \times \frac{1}{6} + 0 \times \frac{3}{6} + 1 \times \frac{1}{6} + 2 \times \frac{1}{6}, \\ &= \frac{1}{3}.\end{aligned}$$

On constate ainsi que l'espérance et le mode ne sont pas identiques. Le mode correspond à la valeur la plus fréquente alors que l'espérance tient compte aussi des autres valeurs.☞

L'espérance a des propriétés intéressantes, notamment la suivante.

Propriété 4.3 Soit (Ω, \mathbb{P}) une expérience aléatoire et soit X une variable aléatoire réelle et discrète sur (Ω, \mathbb{P}) . Alors pour tous nombres réels a et b , on a

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b. \tag{4.3}$$

Rappelons que dans l'équation (4.3), l'expression $aX + b$ désigne la variable aléatoire $\phi(X)$ obtenue grâce à la fonction ϕ donnée par $\phi(x) = ax + b$. Quand ϕ est quelconque, on dispose du théorème de transport (aussi appelé théorème de transfert).

Théorème 4.1 *Soit X une variable aléatoire **réelle et discrète** sur (Ω, \mathbb{P}) et soit ϕ une fonction de \mathbb{R} dans \mathbb{R} . On a*

$$\mathbb{E}(\phi(X)) = \sum_{x \in X(\Omega)} \phi(x) \mathbb{P}(X = x). \quad (4.4)$$

L'aspect remarquable du théorème apparaît quand on applique la formule (4.2) à la variable $\phi(X)$. On obtient en effet

$$\mathbb{E}(\phi(X)) = \sum_{t \in \phi(X(\Omega))} t \mathbb{P}(\phi(X) = t).$$

Pour appliquer cette formule, il faut donc calculer la loi de $\phi(X)$ (par exemple en utilisant la proposition 3.1, puis faire la somme pondérée des valeurs prises par $\phi(X)$). Le théorème de transport montre qu'on peut se contenter de faire la somme pondérée des $\phi(x)$ pour $x \in X(\Omega)$, en utilisant la loi de X . Tout se passe comme si ϕ n'intervenait que sur les valeurs de X , pas sur sa loi. La preuve du théorème éclaire ce point mais est assez abstraite. Nous proposons donc à la place un exemple qui illustre cette preuve.

♣ **Exemple 4.8** Reprenons l'exemple 4.6 et la variable aléatoire X qui donne la valeur de la face du dé. Soit $Y = |X - 3|$, la variable aléatoire obtenue à partir de la fonction ϕ donnée par $\phi(x) = |x - 3|$. Calculons la loi de Y . Pour ce faire, considérons le tableau suivant qui associe aux valeurs de X celles de Y :

X	1	2	3	4	5	6
Y	2	1	0	1	2	3

On constate par exemple que

$$\begin{aligned} \mathbb{P}_Y(2) &= \mathbb{P}_X(\phi^{-1}(2)) \\ &= \mathbb{P}_X(\{1, 5\}) \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{3}. \end{aligned}$$

En conduisant ce type de calculs pour les autres valeurs, on constate que $Y(\Omega) = \{0, 1, 2, 3\}$ et que la loi de Y est donnée par

y	0	1	2	3
$\mathbb{P}(Y = y)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

On peut alors calculer $\mathbb{E}(Y)$:

$$\begin{aligned}\mathbb{E}(Y) &= \sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y), \\ &= 0 \times \frac{1}{6} + 1 \times \frac{1}{3} + 2 \times \frac{1}{3} + 3 \times \frac{1}{6}, \\ &= \frac{3}{2}.\end{aligned}$$

Or, en appliquant le théorème de transport, on obtient

$$\begin{aligned}\mathbb{E}(Y) &= \sum_{x \in X(\Omega)} \phi(x) \mathbb{P}(X = x), \\ &= \phi(1) \times \frac{1}{6} + \phi(2) \times \frac{1}{6} + \phi(3) \times \frac{1}{6} + \phi(4) \times \frac{1}{6} + \phi(5) \times \frac{1}{6} + \phi(6) \times \frac{1}{6}.\end{aligned}$$

On remarque que $\phi(1) = \phi(3)$ et donc qu'on peut regrouper les termes correspondants. De façon plus générale, si $\phi(a) = \phi(b)$, on peut remplacer $\phi(a)\mathbb{P}(X = a) + \phi(b)\mathbb{P}(X = b)$ par $\phi(a)\mathbb{P}(X \in \{a, b\})$. Ici, on obtient

$$\begin{aligned}\mathbb{E}(Y) &= 0 \times \frac{1}{6} + 1 \times \left(\frac{1}{6} + \frac{1}{6}\right) + 2 \times \left(\frac{1}{6} + \frac{1}{6}\right) + 3 \times \frac{1}{6} \\ &= 0 \times \frac{1}{6} + 1 \times \frac{1}{3} + 2 \times \frac{1}{3} + 3 \times \frac{1}{6}, \\ &= \frac{3}{2}.\end{aligned}$$

Toute l'astuce pour voir l'égalité entre les deux formules (directe et par le théorème de transport) réside donc dans le regroupement des valeurs de $x \in X(\Omega)$ qui donnent la même valeur de $\phi(x)$. Or, on sait que

$$\begin{aligned}\mathbb{P}(Y = y) &= \mathbb{P}_X(\phi^{-1}(y)), \\ &= \sum_{x \in \phi^{-1}(y)} \mathbb{P}(X = x),\end{aligned}$$

en appliquant les propriétés élémentaires des probabilités. On constate donc que

$$y \mathbb{P}(Y = y) = \phi(x) \sum_{x \in \phi^{-1}(y)} \mathbb{P}(X = x),$$

ce qui est l'étape principale de la preuve du théorème de transport. ◻

Variance

De la même façon que l'espérance complète le mode dans la détermination de la tendance centrale d'une variable aléatoire, la variance complète l'entropie dans l'estimation de sa dispersion.

Définition 4.5 Soit X une variable aléatoire réelle et discrète sur (Ω, \mathbb{P}) dont l'espérance $\mathbb{E}(X)$ existe. On appelle **variance** de X , la valeur numérique notée $\mathbb{V}(X)$ définie par

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2), \quad (4.5)$$

quand elle existe. Il s'agit donc de l'espérance de la variable aléatoire Y définie par $Y = (X - \mathbb{E}(X))^2$.

Quand la variance de X , son **écart type** est donné par

$$\sigma(X) = \sqrt{\mathbb{V}(X)}. \quad (4.6)$$

△ Remarque 4.4 Comme nous l'avons déjà indiqué dans la remarque 4.3, l'espérance d'une variable aléatoire n'existe pas toujours. Il en est de même pour la variance, celle-ci étant une espérance.

Comme l'entropie, la variance mesure la dispersion d'une variable aléatoire.

Propriétés 4.4 Soit X une variable aléatoire réelle et discrète dont l'espérance et la variance sont bien définies. On a alors

1. $\mathbb{V}(X) = \sum_{x \in X(\omega)} (x - \mathbb{E}(X))^2 \mathbb{P}(X = x)$.
2. $\mathbb{V}(X) \geq 0$ et $\mathbb{V}(X) = 0$ si et seulement si X est constante (c'est-à-dire qu'il existe un unique $x \in X(\Omega)$ tel que $\mathbb{P}(X = x) = 1$).
3. $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
4. pour tous nombres réels a et b $\mathbb{V}(aX + b) = a^2 \mathbb{V}(X)$.

Preuve La première propriété correspond à l'application directe du théorème de transport à la fonction ϕ définie par $\phi(x) = (x - \mathbb{E}(X))^2$.

Pour la deuxième propriété, on utilise la première. On constate ainsi que la variance est une somme de termes tous positifs (car une probabilité est toujours positive et qu'un carré aussi). Donc $\mathbb{V}(X) \geq 0$.

Supposons que $\mathbb{V}(X) = 0$. Cela signifie que chaque terme de la somme est nul et donc que pour tout $x \in X(\Omega)$, $(x - \mathbb{E}(X))^2 \mathbb{P}(X = x) = 0$. Considérons un $x \in X(\Omega)$, tel que $\mathbb{P}(X = x) > 0$. Il en existe au moins 1 puisque que $\sum_{x \in X(\Omega)} \mathbb{P}(X = x) = 1$. Alors $x = \mathbb{E}(X)$. De ce fait, il existe un unique x tel que $\mathbb{P}(X = x) = 1$.

Nous montrerons les deux autres propriétés après avoir introduit la notion d'ensemble de variables aléatoires. □

En pratique, le calcul de la variance d'une variable aléatoire se fait soit à partir du théorème de transport (propriété 1 ci-dessus) soit à partir de la propriété 3 ci-dessus.

☞ Exemple 4.9 Reprenons l'exemple 4.6 et la variable aléatoire X qui donne la valeur de la face du dé. Nous avons vu que $\mathbb{E}(X) = \frac{7}{2}$. Calculons $\mathbb{V}(X)$ par les trois méthodes possibles : à partir de la variable aléatoire $Y = (X - \mathbb{E}(X))^2$, en appliquant le théorème de transport à cette variable, ou avec la formule $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Par la première méthode, on commence par calculer les valeurs possibles de Y en fonction de X , ce qui donne le tableau suivant :

X	1	2	3	4	5	6
Y	$\frac{25}{4}$	$\frac{9}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{9}{4}$	$\frac{25}{4}$

On a donc $Y(\Omega) = \{\frac{1}{4}, \frac{9}{4}, \frac{25}{4}\}$. Le calcul de la loi de Y est simple, on trouve

y	$\frac{1}{4}$	$\frac{9}{4}$	$\frac{25}{4}$
$\mathbb{P}(Y = y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

On calcule alors $\mathbb{E}(Y) = \mathbb{V}(X)$ par

$$\begin{aligned}\mathbb{V}(X) &= \sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y), \\ &= \frac{1}{3} \times \left(\frac{1}{4} + \frac{9}{4} + \frac{25}{4} \right), \\ &= \frac{35}{12}.\end{aligned}$$

Le calcul par la deuxième méthode est plus direct mais correspond à une formule plus longue, soit

$$\begin{aligned}\mathbb{V}(X) &= \sum_{x \in X(\Omega)} \left(x - \frac{7}{2} \right)^2 \mathbb{P}(X = x), \\ &= \frac{1}{6} \times \left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4} \right), \\ &= \frac{35}{12}.\end{aligned}$$

Pour la troisième méthode, on calcule d'abord par le théorème de transfert $\mathbb{E}(X^2)$, ce qui donne

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{x \in X(\Omega)} x^2 \mathbb{P}(X = x), \\ &= \frac{1}{6} \times (1 + 4 + 9 + 16 + 25 + 36), \\ &= \frac{91}{6}.\end{aligned}$$

Puis on applique la formule $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$, ce qui donne

$$\begin{aligned}\mathbb{V}(X) &= \frac{91}{6} - \frac{49}{4}, \\ &= \frac{35}{12}.\end{aligned}$$

Vocabulaire

L'espérance et la variance étant deux mesures très utiles pour résumer numériquement le comportement d'une variable aléatoire, il existe un vocabulaire spécifique construit à partir d'elles.

Définition 4.6 Soit X une variable aléatoire numérique discrète dont l'espérance et la variance sont bien définies.

1. Si $\mathbb{E}(X) = 0$, X est dite **centrée**.
2. La variable aléatoire $X - \mathbb{E}(X)$ est obtenue à partir de X par **centrage** et elle est centrée.
3. Si $\mathbb{V}(X) = 1$, X est dite **réduite**.
4. Si $\mathbb{V}(X) > 0$, la variable aléatoire $\frac{X}{\sigma(X)}$ est obtenue à partir de X par **réduction** et elle est réduite.
5. Si $\mathbb{E}(X) = 0$ et $\mathbb{V}(X) = 1$, X est dite **centrée-réduite**.
6. Si $\mathbb{V}(X) > 0$, la variable aléatoire $\frac{X - \mathbb{E}(X)}{\sigma(X)}$ est obtenue à partir de X par **centrage et réduction** et elle est centrée-réduite.

On parle aussi de **versions** de X pour les différentes opérations. Par exemple $\frac{X - \mathbb{E}(X)}{\sigma(X)}$ est la version centrée-réduite de X .

Autres moments

L'espérance et la variance d'une variable aléatoire sont des **moments** de cette variable. On a plus généralement toute une collection de moments.

Définition 4.7 Soit X une variable aléatoire numérique discrète. Soit r un entier strictement positif. Le **moment** (dit aussi **moment ordinaire**) d'ordre r de X est donné par

$$\mathbb{M}_r(X) = \mathbb{E}(X^r), \quad (4.7)$$

quand cette quantité est bien définie. Dans ce cas, on a

$$\mathbb{M}_r(X) = \sum_{x \in X(\Omega)} x^r \mathbb{P}(X = x). \quad (4.8)$$

Le **moment centré** d'ordre r est donné par

$$\mu_r(X) = \mathbb{E}((X - \mathbb{E}(X))^r), \quad (4.9)$$

quand cette quantité est bien définie. Dans ce cas, on a

$$\mu_r(X) = \sum_{x \in X(\Omega)} (x - \mathbb{E}(X))^r \mathbb{P}(X = x). \quad (4.10)$$

L'espérance est donc le moment ordinaire d'ordre 1 alors que la variance est le moment centré d'ordre 2.

4.5 Lois classiques

Un des intérêts majeurs du concept de variable aléatoire est qu'il permet de définir directement des variables aléatoires sans passer par une expérience aléatoire, comme nous l'avons vu notamment à la proposition 4.1. En pratique, il existe toute une collection de lois classiques pour les variables aléatoires discrètes, très utiles pour modéliser divers phénomènes.

Loi uniforme discrète

Définition 4.8 Soit U un ensemble fini. On dit qu'une variable aléatoire X à valeurs dans U suit la **loi uniforme** sur U et on note $X \sim \mathcal{U}(U)$, si et seulement si sa loi est donnée par

$$\mathbb{P}(X = x) = \frac{1}{|U|}, \quad (4.11)$$

pour tout élément x de U .

Propriétés 4.5 Soit X une variable aléatoire de loi uniforme sur l'ensemble $\{1, \dots, n\}$. Alors

1. $\mathbb{P}(X = k) = \frac{1}{n}$ pour tout $k \in \{1, \dots, n\}$;
2. $\mathbb{E}(X) = \frac{n+1}{2}$;
3. $\mathbb{V}(X) = \frac{n^2-1}{12}$.

⚠ Remarque 4.5 Il faut bien être attentif au fait que la loi uniforme est définie pour tout ensemble U mais que les propriétés indiquées ci-dessus ne sont valables que dans le cas particulier de $U = \{1, \dots, n\}$. Si U n'est pas numérique, les notions d'espérance et de variances ne s'appliquent pas. Si U est numérique mais pas exactement égal à $\{1, \dots, n\}$, les valeurs de ces moments ne sont pas les mêmes. Par exemple si $U = \{-2, -1, 0, 1, 2\}$ et que $X \sim \mathcal{U}(U)$, alors $\mathbb{E}(X) = 0$.

Loi de Bernoulli

Définition 4.9 Toute variable aléatoire à valeurs dans $\{0, 1\}$ est dite suivre une **loi de Bernoulli**. La valeur 1 représente le **succès** alors que le 0 représente l'**échec**.

La notation $X \sim \mathcal{B}(p)$ indique que X suit une loi de Bernoulli de paramètre p , c'est-à-dire que $X(\Omega) = \{0, 1\}$ et que $\mathbb{P}(X = 1) = p$.

⚠ Remarque 4.6 Une expérience aléatoire d'univers $\Omega = \{\text{Échec}, \text{Succès}\}$ est souvent appelée une **épreuve de Bernoulli**. La variable aléatoire de Bernoulli associée est définie par $X(\text{Échec}) = 0$ et $X(\text{Succès}) = 1$.

Propriétés 4.6 Soit X une variable de Bernoulli de paramètre p ($X \sim \mathcal{B}(p)$). Alors

1. $\mathbb{E}(X) = p$;
2. $\mathbb{V}(X) = p(1 - p)$.

Loi binomiale

Définition 4.10 Soit n un entier strictement positif et $p \in [0,1]$. On dit qu'une variable aléatoire X à valeurs dans $\{0, 1, \dots, n\}$ suit une **loi binomiale** de paramètres n et p , et on note $X \sim \mathcal{B}(n, p)$, si et seulement si sa loi est donnée par

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \quad (4.12)$$

pour tout $k \in \{0, 1, \dots, n\}$.

Propriétés 4.7 Soit X une variable de loi binomiale de paramètres n et p ($X \sim \mathcal{B}(n, p)$). Alors

1. $\mathbb{E}(X) = np$;
2. $\mathbb{V}(X) = np(1-p)$.

La loi binomiale est fortement liée à la loi de Bernoulli comme le montre la proposition suivante qui donne aussi un « manuel d'utilisation » de la loi binomiale.

Proposition 4.2 Soit une expérience aléatoire consistant en n épreuves de Bernoulli **indépendantes et de même paramètre** p . Soit X la variable aléatoire comptant le nombre de succès obtenu sur l'ensemble de ces n épreuves. Alors X suit une loi binomiale de paramètres n et p .

⚠ Remarque 4.7 La notion d'indépendance utilisée ici est celle des évènements indépendants dans leur ensemble (cf définition 2.3). Plus précisément, on suppose que les n évènements $(A_i)_{1 \leq i \leq n}$, s'ils sont tels que chaque A_i ne concerne que l'épreuve numéro i , sont indépendants dans leur ensemble.

Loi géométrique

Définition 4.11 Soit $p \in]0,1[$. On dit qu'une variable aléatoire X à valeurs dans \mathbb{N}^* (l'ensemble des entiers strictement positifs) suit une **loi géométrique** de paramètre p , et on note $X \sim \mathcal{G}(p)$, si et seulement si sa loi est donnée par

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad (4.13)$$

pour tout $k \in \mathbb{N}^*$.

⚠ Remarque 4.8 La loi géométrique a un support infini dénombrable (contrairement aux lois de Bernoulli et binomiale qui ont des supports finis).

Propriétés 4.8 Soit X une variable de loi géométrique de paramètre p ($X \sim \mathcal{G}(p)$). Alors

1. $\mathbb{E}(X) = \frac{1}{p}$;
2. $\mathbb{V}(X) = \frac{1-p}{p^2}$;
3. pour tout $k \in \mathbb{N}^*$, $F_X(k) = 1 - (1-p)^k$.

Comme la loi binomiale, la loi géométrique est fortement liée à la loi de Bernoulli.

Proposition 4.3 *Soit l'expérience aléatoire suivante : on effectue une série d'épreuves de Bernoulli **indépendantes et de même paramètre** p , et on s'arrête à l'obtention du premier succès. On considère la variable aléatoire X donnant le nombre d'épreuves de Bernoulli réalisées (si $X = k$, on a obtenu $k - 1$ échecs et 1 dernier succès). Alors X suit une loi géométrique de paramètre p .*

Loi de Poisson

Définition 4.12 *Soit λ un nombre réel strictement positif. On dit qu'une variable aléatoire X à valeurs dans \mathbb{N} suit une **loi de Poisson** de paramètre λ , et on note $X \sim \mathcal{P}(\lambda)$, si et seulement si sa loi est donnée par*

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (4.14)$$

pour tout $k \in \mathbb{N}$.

Propriétés 4.9 *Soit X une variable de loi de Poisson de paramètre λ ($X \sim \mathcal{P}(\lambda)$). Alors*

1. $\mathbb{E}(X) = \lambda$;
2. $\mathbb{V}(X) = \lambda$.

Annexes

Annexe A

Théorie des ensembles

La théorie des ensembles est à la base des mathématiques. Elle est utilisée pour construire la plupart des concepts de plus haut niveau, comme par exemple les probabilités. Il s'agit d'une partie des mathématiques très formelle dont la maîtrise est largement hors du programme du présent cours. Il faut cependant connaître un minimum de notions qui sont rappelées dans ce chapitre, en général de façon assez informelle.

A.1 Notations et opérations

De façon informelle, nous considérerons un ensemble comme une *collection d'objets*. Ces objets sont les **éléments** de l'ensemble. Il n'y a pas de restriction sur le contenu d'un ensemble. Il peut notamment contenir d'autres ensembles et des objets de natures différentes.

Nous rappelons ci-dessous des notations et concepts important en théorie des ensembles.

1. la notation \emptyset désigne l'**ensemble vide**, c'est-à-dire l'ensemble ne contenant aucun élément.
2. un ensemble fini (cf la définition C.2) est souvent donné en **extension**, c'est-à-dire sous la forme de la liste des ses éléments.

∞ Exemple A.1

$$A = \{1, 2, 3\} \qquad B = \{a, b, \{0\}, 4\} \qquad \infty$$

Bien que cela soit rare, on désigne parfois l'ensemble vide par $\emptyset = \{\}$.

3. la notation $x \in A$ indique que l'objet x est **élément** de A (on dit aussi que x **appartient** à A). On note $x \notin A$ pour dire que l'objet x n'est pas élément de A .

∞ Exemple A.2

$$\begin{array}{ll} 1 \in \{1, 2, 3\} & \{a, b\} \in \{\{u, v\}, 2, \{a, b\}\} \\ 2 \notin \emptyset & \{1\} \notin \{1, 2, 3\}. \end{array} \qquad \infty$$

4. un ensemble peut être défini en **compréhension** ce qui revient à sélectionner dans un ensemble existant les éléments qui vérifient une propriété.

∞ **Exemple A.3** L'ensemble des entiers pairs U est défini à partir de l'ensemble des entiers naturels \mathbb{N} par

$$U = \{x \in \mathbb{N} \mid x \text{ est divisible par } 2\}. \quad \infty$$

La forme générale de la définition en compréhension est

$$B = \{x \in A \mid P(x)\},$$

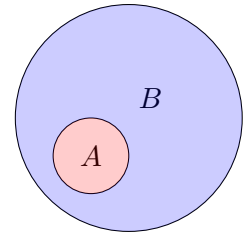
où $P(x)$ est une propriété. Les éléments de B sont ceux de A qui rendent la propriété vraie.

∞ **Exemple A.4** Si $A = \{1, 2, 3, 4\}$ et que $B = \{x \in A \mid x + 1 \leq 3\}$, alors $B = \{1, 2\}$. ∞

5. si A et B sont deux ensembles, la notation $A \subset B$ indique que A est un **sous-ensemble** de B , c'est-à-dire que tous les éléments de A sont aussi éléments de B . Plus formellement

$$(A \subset B) \Leftrightarrow (x \in A \Rightarrow x \in B).$$

Quand $A \subset B$, on dit que A est une **partie** de B . Le diagramme ci-contre, dit diagramme de Venn, illustre $A \subset B$.

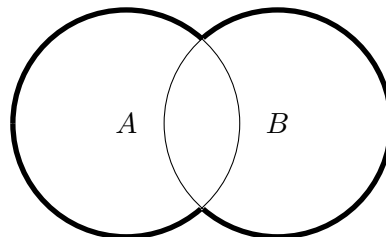


∞ **Exemple A.5**

$$\{2, 3\} \subset \{1, 2, 3\} \qquad \{2, \{a, b\}\} \subset \{\{u, v\}, 2, \{a, b\}\}$$

∞

6. si A et B sont deux ensembles, on note $A \cup B$ l'ensemble **union** de A et B , c'est-à-dire l'ensemble constitué des éléments de A et des éléments de B (et seulement eux). En d'autres termes, ce sont les objets éléments de A ou de B . Le diagramme de Venn ci-dessous représente $A \cup B$ par l'ensemble délimité par la ligne épaisse, chaque disque correspondant à un des deux ensembles A et B .



L'union se généralise à plus de deux ensembles : soit $(A_i)_{i \geq 1}$ une suite d'ensembles, on note $\bigcup_{i \geq 1} A_i$ l'union des A_i , c'est-à-dire l'ensemble constitué des objets qui appartiennent à au moins un des A_i .

Notons que l'union ne peut pas s'écrire rigoureusement en compréhension.

∞ **Exemple A.6**

$$\{a, 2, \{0\}\} \cup \{b, \emptyset, 3\} = \{a, 2, \{0\}, b, \emptyset, 3\}.$$

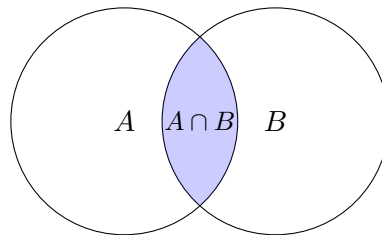
∞

7. si A et B sont deux ensembles, on note $A \cap B$ l'ensemble **intersection** de A et B , c'est-à-dire l'ensemble constitué des éléments de A qui sont aussi éléments de B (et seulement eux). De façon équivalente par symétrie, ce sont aussi les éléments de B qui sont aussi éléments de A (ou encore, les objets éléments de A et de B).

On peut définir l'intersection en compréhension par

$$A \cap B = \{x \in A \mid x \in B\} = \{x \in B \mid x \in A\}.$$

Dans le diagramme de Venn ci-dessous, l'intersection est représentée par la zone colorée au centre.



Quand $A \cap B = \emptyset$, on dit que les ensembles A et B sont **disjoints**. De plus, on dit que l'union $A \cup B$ est une **union disjointe**.

L'intersection se généralise à plus de deux ensembles : soit $(A_i)_{i \geq 1}$ une suite d'ensembles, on note $\bigcap_{i \geq 1} A_i$ l'intersection des A_i , c'est-à-dire l'ensemble constitué des objets qui appartiennent à tous les des A_i .

∞ **Exemple A.7**

$$\{2, 3, 4\} \cap \{1, 2, 3\} = \{2, 3\}$$

$$\{a, b\} \cap \{\{a\}, \{b\}\} = \emptyset$$

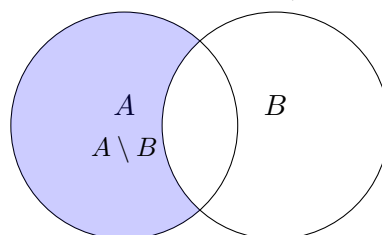
∞

8. si A et B sont deux ensembles, on note $A \setminus B$ le **complémentaire** de B dans A , c'est-à-dire les éléments de A qui ne sont pas éléments de B .

La définition du complémentaire en compréhension est

$$A \setminus B = \{x \in A \mid x \notin B\}.$$

Le diagramme de Venn ci-dessous représente $A \setminus B$ par la zone colorée.



∞ Exemple A.8

$$\{2, 3, 4\} \setminus \{1, 2, 3\} = \{4\}.$$

∞

Quand tous les ensembles considérés sont des sous-ensembles d'un ensemble fixé Ω , on note plus simplement $\bar{A} = \Omega \setminus A$ et on parle alors de **complémentaire** sans préciser relativement à quel ensemble. On remarque que si A et B sont deux sous-ensembles de Ω , on a

$$A \setminus B = A \cap \bar{B},$$

où \bar{B} désigne comme convenu le complémentaire de B dans Ω .

9. si A et B sont deux ensembles, on note $A\Delta B$ la **différence symétrique** entre A et B . Cet ensemble est constitué de l'union des éléments de A qui ne sont pas dans B et des éléments de B qui ne sont pas dans A . Par définition, on a donc

$$A\Delta B = (A \setminus B) \cup (B \setminus A).$$

Il est facile de voir qu'on a aussi

$$A\Delta B = (A \cup B) \setminus (A \cap B).$$

∞ Exemple A.9

$$\{2, 3, 4\}\Delta\{1, 2, 3\} = \{1, 4\}.$$

∞

10. si A est un ensemble, la notation $\mathcal{P}(A)$ désigne l'**ensemble des parties** de A , c'est-à-dire l'ensemble dont les éléments sont tous les sous-ensembles possibles de A (ce qui inclut notamment \emptyset et A).

L'ensemble des parties ne peut pas être défini rigoureusement en compréhension.

∞ Exemple A.10

$$\begin{aligned} \mathcal{P}(\{a, 2\}) &= \{\emptyset, \{a\}, \{2\}, \{a, 2\}\}, \\ \mathcal{P}(\{1, \{a, b\}\}) &= \{\emptyset, \{1\}, \{a, b\}, \{1, \{a, b\}\}\}. \end{aligned}$$

∞

A.2 Propriétés des opérations ensemblistes

Les opérations décrites dans la section précédente possèdent de nombreuses propriétés utiles :

commutativité l'union et l'intersection sont commutatives, c'est-à-dire que tous ensembles A et B :

$$\begin{aligned} A \cap B &= B \cap A, \\ A \cup B &= B \cup A. \end{aligned}$$

associativité l'union et l'intersection sont associatives, c'est-à-dire que l'ordre d'une série d'opérations n'importe pas. Plus précisément pour tous ensembles A , B et C :

$$(A \cap B) \cap C = A \cap (B \cap C),$$

$$(A \cup B) \cup C = A \cup (B \cup C).$$

élément neutre l'ensemble vide est un élément neutre de l'union, c'est-à-dire que pour tout ensemble A ,

$$\emptyset \cup A = A.$$

Tout un ensemble Ω est un élément neutre de l'intersection pour ses sous-ensembles, c'est-à-dire que pour tout ensemble $A \subset \Omega$,

$$\Omega \cap A = A.$$

involution dans un ensemble Ω fixé, l'opération de passage au complémentaire est involutive, c'est-à-dire que pour tout ensemble $A \subset \Omega$,

$$\overline{\overline{A}} = A.$$

distributivité l'union et l'intersection sont distributives l'une par rapport à l'autre, c'est-à-dire que pour tous ensembles A , B et C :

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

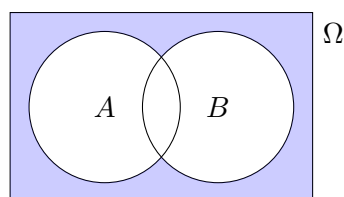
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

dualité dans un ensemble Ω fixé, l'opération de passage au complémentaire satisfait les lois de De Morgan (aussi appelées lois de dualité), c'est-à-dire que pour tous ensembles A et B :

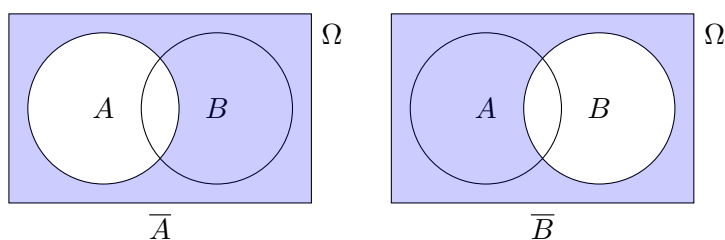
$$\overline{A \cup B} = \overline{A} \cap \overline{B},$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}.$$

On peut se convaincre de la véracité de certaines de ces propriétés en utilisant des diagrammes de Venn. Par exemple, le diagramme ci-dessous représente en couleur l'ensemble $\overline{A \cup B} = \Omega \setminus (A \cup B)$.



On constate qu'il s'agit bien de l'intersection de deux régions/ensembles représenté(e)s ci-dessous :



A.3 Produit cartésien

Définition A.1 Soit A et B deux ensembles, le **produit cartésien** de A et B (ou de A par B), noté $A \times B$ est l'ensemble de tous les couples de la forme (a, b) avec $a \in A$ et $b \in B$.

Le **carré cartésien** $A \times A$ est noté A^2 par analogie entre l'opération de multiplication entre nombres réels et le produit d'ensembles.

∞ **Exemple A.11** Si $A = \{2, a\}$, alors

$$A^2 = \{(2, 2), (2, a), (a, 2), (a, a)\}.$$

∞

La notion de produit cartésien se généralise de 2 à n ensembles.

Définition A.2 Soit A_1, \dots, A_n , n ensembles. On appelle **n-uplet** sur ces ensembles une liste (a_1, \dots, a_n) d'objets telle que $\forall i, 1 \leq i \leq n, a_i \in A_i$. L'ensemble des n -uplets est le **produit cartésien** de A_1, \dots, A_n , noté $A_1 \times A_2 \times \dots \times A_n$, ou encore $\prod_{i=1}^n A_i$.

Le produit cartésien de $\prod_{i=1}^n A$ est noté A^n .

Il est important de noter que dans un n -uplet, l'ordre des éléments est pris en compte. Par exemple, si a et b sont deux éléments distincts de A , la paire $(a, b) \in A^2$ est différente de la paire (b, a) .

A.4 Partition

Définition A.3 Soit A un ensemble et $P = \{C_1, C_2, \dots, C_k\}$ un ensemble de k sous-ensembles de A . On dit que P est une **partition** de A si et seulement si les propriétés suivantes sont vérifiées :

1. aucun des C_i n'est vide : $\forall i, 1 \leq i \leq k, C_i \neq \emptyset$;
2. les C_i ne s'intersectent pas : $\forall i, j, 1 \leq i \leq k, 1 \leq j \leq k, i \neq j \Rightarrow C_i \cap C_j = \emptyset$;
3. l'union des C_i forme A tout entier : $A = \bigcup_{i=1}^k C_i$.

∞ **Exemple A.12** Soit $A = \{1, a, z, \{u, v\}\}$. Une partition de A est donnée par les trois ensembles suivants :

$$\begin{aligned} C_1 &= \{1\} & C_2 &= \{z, \{u, v\}\}, \\ C_3 &= \{a\}. \end{aligned}$$

En revanche, les ensembles suivants ne forment pas une partition de A :

$$\begin{aligned} C_1 &= \{1\} & C_2 &= \{u, v\}, \\ C_3 &= \{a, z\}. \end{aligned}$$

En effet, leur union vaut $\{1, a, z, u, v\}$, un ensemble distinct de A .

∞

En probabilité, on utilise très fréquemment des partitions car les propriétés fondamentales des probabilités (cf propriétés 1.1) assurent que la probabilité d'un évènement A est égale

à la somme des probabilités des événements formant une partition de A . On utilise en particulier les propriétés suivantes :

Propriétés A.1 Soit A et B deux ensembles.

1. $A \cup B$ est l'union disjointe des trois ensembles $P = \{A \setminus B, A \cap B, B \setminus A\}$;
2. si A et B sont d'intersection non vide ($A \cap B \neq \emptyset$) et si $A \neq B$, alors l'ensemble $P = \{A \setminus B, A \cap B, B \setminus A\}$ est une partition de $A \cup B$;
3. si $P' = \{B_1, \dots, B_k\}$ est une partition de B et que $A \subset B$, alors A est l'union disjointe de k ensembles $A_i = A \cap B_i$ pour $1 \leq i \leq k$.

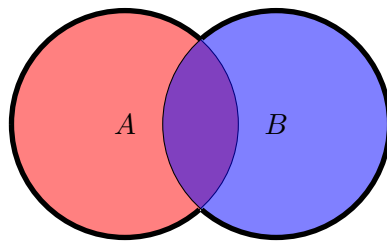


FIGURE A.1: Illustration de la décomposition de $A \cup B$ en l'union disjointe $(A \setminus B) \cup (A \cap B) \cup (B \setminus A)$ par un diagramme de Venn. $A \setminus B$ est en rouge, $B \setminus A$ en bleu et $A \cap B$ en violet.

Annexe B

Fonctions

B.1 Définition

De façon informelle, une **fonction** est un moyen d'associer à chaque élément d'un ensemble au plus élément d'un autre ensemble. Par exemple la fonction « valeur absolue » associe à un nombre quelconque soit le nombre lui-même s'il est positif, soit son opposé s'il est négatif.

Pour définir une fonction, il faut donc un **ensemble de départ**, par exemple A , et un **ensemble d'arrivée**, par exemple B . Il faut ensuite lister les associations entre certains éléments de l'ensemble de départ et les éléments correspondants dans l'ensemble de départ. Une association entre deux éléments peut être représentée comme une paire (x, y) avec $x \in A$ et $y \in B$. Mathématiquement, la fonction est alors un ensemble de ces paires. La définition formelle est alors la suivante.

Définition B.1 Soit A et B deux ensembles. Une **fonction** f de A vers B est une partie de $A \times B$, telle que si $(x, y) \in f$ et $(x, y') \in f$, alors $y = y'$.

Si $(x, y) \in f$, on note $f(x) = y$. On dit alors que y est l'**image** par f de x et que x est un **antécédent** de y par f . L'ensemble des $x \in A$ tels qu'il existe $y \in B$ avec $(x, y) \in f$ est le **domaine de définition** de f .

Pour résumer les notations pour une fonction particulière, on note

$$\begin{aligned} f &: A \rightarrow B \\ x &\mapsto y = f(x) \end{aligned} \tag{B.1}$$

On utilise souvent le concept d'image directe.

Définition B.2 Soit f une fonction de A dans B . Soit U un sous-ensemble de A . On appelle **image directe** de U par f le sous-ensemble de B défini par

$$\{y \in B \mid \exists x \in U, f(x) = y\}. \tag{B.2}$$

B.2 Cas particuliers

Définition B.3 Soit f une fonction de A vers B . f est dite

injective si pour tout x et y dans le domaine de définition de f , $x \neq y$ implique $f(x) \neq f(y)$.

surjective si pour tout y dans B , il existe au moins un x dans A tel que $f(x) = y$.

Une fonction à la fois injective et surjective est dite **bijjective**. En pratique, chaque élément de A est donc associé à un unique élément de B et vice versa.

B.3 Composition de fonctions

Définition B.4 Soit trois ensembles, A , B et C et deux fonctions, f de A dans B et g de B dans C . On suppose que $f(A)$ est inclus dans le domaine de définition de g . Alors on peut **composer** les deux fonctions en une fonction h , notée $h = g \circ f$, la **fonction composée** de f et g . h est une fonction de A dans C avec le même domaine de définition que f et donnée par

$$h(x) = g(f(x)), \quad (\text{B.3})$$

pour tout x dans le domaine de définition de f .

B.4 Fonction réciproque

Définition B.5 Soit A et B deux ensembles et f une fonction de A dans B . On appelle **fonction réciproque** de f la fonction notée f^{-1} de $\mathcal{P}(B)$ dans $\mathcal{P}(A)$ définie par

$$\forall U \subset B, f^{-1}(U) = \{a \in A \mid f(a) \in U\}.$$

⚠ Remarque B.1 La notion de fonction réciproque utilisée en probabilité ne coïncide pas exactement avec la notion classique réservée aux fonctions bijectives car on travaille ici sur une fonction ensembliste : f^{-1} associe un ensemble à un autre ensemble, plutôt qu'un élément à un autre élément.

La fonction réciproque possède quelques propriétés intéressantes résumées ci-dessous.

Propriétés B.1 Soit A et B deux ensembles et f une fonction de A dans B . f^{-1} vérifie les propriétés suivantes :

1. $f^{-1}(B) = A$.
2. pour toute suite de parties de B , les $(B_i)_{i \geq 1}$ (suite finie ou non)

$$f^{-1} \left(\bigcup_{i \geq 1} B_i \right) = \bigcup_{i \geq 1} f^{-1}(B_i),$$

et

$$f^{-1} \left(\bigcap_{i \geq 1} B_i \right) = \bigcap_{i \geq 1} f^{-1}(B_i).$$

3. pour tout sous-ensemble U de B ,

$$f^{-1}(\overline{B}) = \overline{f^{-1}(B)}.$$

Soit A , B et C trois ensembles, f une fonction de A dans B et g une fonction de B dans C , alors on a

$$(g \circ f)^{-1} = (f^{-1}) \circ (g^{-1}),$$

c'est-à-dire que pour tout U sous-ensemble de C , on a

$$(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U)).$$

Annexe C

Dénombrément

C.1 Ensembles finis et ensembles dénombrables

Définition C.1 Soit A et B deux ensembles. On dit que A est **en bijection** avec B s'il existe une fonction f bijective de A vers B dont le domaine de définition est A tout entier.

Définition C.2 Soit A un ensemble. On dit que A est **fini** si A est vide ou s'il existe un entier $n > 0$ tel que A soit en bijection avec $\{0, \dots, n-1\}$. On appelle n le **cardinal** de A qui est noté $|A|$, $\text{card}(A)$ ou encore $\#A$. C'est le nombre d'éléments de A . L'ensemble vide est de cardinal nul.

Quand A est fini, on peut numéroter ces éléments et donc écrire

$$A = \{a_1, a_2, \dots, a_n\},$$

si $n = |A|$.

Définition C.3 Soit A un ensemble. On dit que A est **dénombrable** si A est en bijection avec l'ensemble des entiers naturels \mathbb{N} . En pratique, un ensemble dénombrable se décrit comme une suite infinie d'éléments, sous la forme

$$A = \{a_0, a_1, \dots, a_n, a_{n+1}, \dots\}.$$

C.2 Cardinaux et opérations ensemblistes

On peut parfois calculer le cardinal du résultat d'une opération ensembliste à partir des cardinaux des ensembles impliquées, comme dans les cas suivants :

Union disjointe soient A et B deux ensembles finis disjoints, c'est-à-dire tels que $A \cap B = \emptyset$. On a alors

$$|A \cup B| = |A| + |B|.$$

Union soient plus généralement deux ensembles finis, on a

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Complémentaire soient A et B deux ensembles finis

$$|A \setminus B| + |A \cap B| = |A|,$$

et donc en particulier si $B \subset A$, on a

$$|\overline{B}| = |A| - |B|.$$

Attention, cette égalité n'est pas vérifiée si B n'est pas un sous-ensemble de A .

Ensemble des parties soit A un ensemble fini, on a

$$|\mathcal{P}(A)| = 2^{|A|}.$$

Produit cartésien soient A_1, \dots, A_n , n ensembles finis, on a

$$|A_1 \times A_2 \times \dots \times A_n| = |A_1| \times |A_2| \times \dots \times |A_n|.$$

C.3 Listes d'éléments

On considère un ensemble fini A de cardinal n et les produits cartésiens A^p (de cardinaux n^p). Comme rappelé plus haut, un élément d'un de ces produits cartésiens est un p -uplet constitué d'éléments de A , aussi appelé une liste à p éléments ou une **p-liste**. On s'intéresse ici aux cardinaux de certains sous-ensembles des A^p dans lesquels les p -uplets ne contiennent pas de répétition.

Permutations

Définition C.4 Soit A un ensemble. On appelle **permutation** de A une fonction f **bijjective** de A dans lui-même.

Si A est un ensemble fini de cardinal n toute permutation de A correspond à une **numérotation** (ou un **ordre**) des éléments de A qu'on peut représenter par un n -uplet particulier de A^n . D'après la définition d'un ensemble fini, A s'écrit sous la forme

$$A = \{a_1, a_2, \dots, a_n\},$$

ce qui définit une numérotation naturelle de ses éléments. Le n -uplet correspondant est

$$(a_1, a_2, \dots, a_n).$$

Une permutation de A est alors un autre n -uplet contenant exactement une fois chaque élément de A . Par exemple $(a_2, a_3, \dots, a_n, a_1)$ et $(a_n, a_{n-1}, \dots, a_2, a_1)$ sont des permutations de A . On assimile ainsi les permutations d'un ensemble A au sous-ensemble de A^n constitué des n -uplets qui contiennent exactement une fois chaque élément de A .

Théorème C.1 Un ensemble fini A de cardinal $n > 0$ possède exactement $n!$ permutations distinctes, où

$$n! = 1 \times 2 \times \dots \times (n-1) \times n,$$

est la **factorielle** de n .

Par convention, on pose $0! = 1$ et on considère que l'ensemble vide admet une unique permutation.

Ce résultat est assez clair intuitivement. En effet pour construire une liste de n éléments distincts, il faut d'abord choisir le premier élément parmi les n éléments de l'ensemble. Mais dès le deuxième élément, le choix se fait parmi seulement les $n - 1$ éléments restants, etc. On donc n choix, puis $n - 1$ choix, $n - 2$ choix, et ainsi de suite jusqu'au dernier choix qui est imposé (c'est l'élément restant). On retrouve ainsi la définition de $n!$.

Arrangements

Les permutations sont un cas particulier de listes à n éléments distincts choisis dans un ensemble de n éléments. On peut en effet se contenter de choisir $p < n$ éléments en respectant toujours le principe de n'avoir que des éléments distincts.

Définition C.5 Soit A un ensemble fini de cardinal n et soit un entier $1 \leq p \leq n$. On appelle **arrangement** de p éléments de A une fonction **injective** de $\{1, \dots, p\}$ dans A .

De façon équivalente un arrangement est une p -liste d'éléments distincts choisis parmi les n éléments de A .

Théorème C.2 Un ensemble fini A de cardinal n possède exactement A_n^p arrangements distincts (pour $1 \leq p \leq n$) avec

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

C.4 Sous-ensembles

Dans certaines situations, on s'intéresse à des « listes » d'éléments distincts d'un ensemble A dans lesquelles l'ordre n'est pas important. Il s'agit donc en fait de sous-ensembles de cardinal fixé d'un ensemble A . On les dénombre grâce au théorème suivant.

Théorème C.3 Soit A un ensemble fini de cardinal n et soit un entier p , $0 \leq p \leq n$. L'ensemble A possède exactement C_n^p sous-ensembles distincts de cardinal p , où

$$C_n^p = \frac{n!}{(n - p)!p!} = \frac{n \times (n - 1) \times \cdots \times (n - p + 1)}{p!} = \frac{A_n^p}{p!}.$$

Ces sous-ensembles sont appelés des **combinaisons** de p éléments parmi n .

▲ Remarque C.1 La notation C_n^p est peu utilisée en dehors des zones francophones. On lui préfère la notation $\binom{n}{p}$ dans le reste du monde. Il faut bien noter l'inversion des positions : le n est en indice dans C_n^p alors qu'il est situé en haut dans $\binom{n}{p}$, mais il s'agit bien de la même valeur.

La relation $C_n^p = A_n^p/p!$ s'explique assez bien intuitivement. Pour construire un sous-ensemble à p éléments de A , on peut en effet utiliser une liste de p éléments distincts de A et ne pas tenir compte de l'ordre de son contenu. Comme il y a $p!$ ordres possibles (permutations) pour p éléments, on voit qu'un ensemble correspond à $p!$ listes différentes.

Pour obtenir le nombre d'ensembles, il faut donc diviser le nombre de listes A_n^p par le nombre d'ordres possibles pour chaque liste $p!$.

Évolutions de ce document

La dernière version de ce document se trouve sur la page <http://apiacoa.org/teaching/statistics/index.fr.html>.



29/01/2017 : version 0.5.2

- correction d'une erreur de notation sur les partitions

05/01/2017 : version 0.5.1

- ajout d'exemples dans les rappels de théorie des ensembles
- modification des marges

24/02/2016 : version 0.5.0

- première version du chapitre 4 sur les variables aléatoires discrètes :
 - définition
 - notion d'entropie
 - fonction de répartition dans le cas discret
 - notion d'espérance
 - notion de variance
 - notion de moment
 - lois discrètes classiques
- chapitre 3 : ajout de la notion de mode

23/02/2016 : version 0.4.4

- réorganisation des annexes
- ajout de la définition formelle d'une fonction et des propriétés importantes des fonctions

20/02/2016 : version 0.4.3

- annexe A : ajout de la notion de différence symétrique
- chapitre 3 : ajout de la notion de variable aléatoire fonction d'une autre

26/01/2015 : version 0.4.2 corrections de fautes de frappe et clarification de certains points.

20/01/2015 : version 0.4.1

- modifications du chapitre 3 :
 - ajout de la notion de support
 - illustration graphique de la définition de la loi d'une variable aléatoire
 - ajout de graphes de fonction de répartition
 - ajout d'un exemple simple d'utilisation du théorème de la réciproque

- ajout de propriétés de la fonction de répartition
- ajout du théorème de caractérisation
- ajout d'un exemple montrant que l'égalité des lois n'implique pas l'égalité des variables aléatoires

19/01/2015 : version 0.4.0

- utilisation de symboles pour les exemples avec des dés
- modifications du chapitre 1 :
 - remarque sur les unions et intersections dénombrables
 - précision sur la numérotation des suites
 - exemples de probabilités arbitraires sur des ensembles finis
 - discussion détaillée sur les problèmes liés aux résultats discernables ou non
- modifications du chapitre 2 :
 - notion d'indépendance conditionnelle
 - deux exemples complexes d'indépendance conditionnelle
- modifications du chapitre 3 :
 - correction de fautes de frappe
 - indexation
 - introduction des notations de la forme $\mathbb{P}(X \leq t)$
- modifications des annexes :
 - rappel de la notion de définitions en extension et compréhension d'un ensemble
 - précision sur la notion d'union disjointe et de décomposition
 - ajout de diagrammes de Venn
 - calcul du cardinal d'un complémentaire

09/04/2014 : version 0.3.0

- début de la rédaction du chapitre 3 :
 - introduction élémentaire aux variables aléatoires
 - définition d'une variable aléatoire et de sa loi
 - notations simplifiées pour la loi d'une variable aléatoire
 - définition et propriété de la fonction de répartition
 - théorème de la réciproque
- ajout de propriétés des probabilités (suite croissante ou décroissante d'ensembles)

10/02/2014 : version 0.2.3 fin de la rédaction du chapitre 2 :

- expériences aléatoires composées
- règle des probabilités totales
- règle de Bayes
- indépendance

08/03/2014 : version 0.2.2 ajout de la définition des fonctions réciproques (ensemblistes)

10/01/2014 : version 0.2.1 ajout de la notation internationale $\binom{n}{p}$ et de la définition des partitions

10/01/2013 : version 0.2.0 début de rédaction du chapitre 2 :

- introduction à la notion d'évènement réalisé
- définition des probabilités conditionnelles

10/01/2013 : version 0.1.1 utilisation de la virgule pour les nombres non entier

25/02/2012 : version 0.1 première version complète du chapitre 1

Index

- Appartenance, 69
- \in , 69
- Arrangement, 83
- Associativité, 73

- Cardinal, 81
- Carré cartésien, 74
- Centrage, 63
- Combinaison, 83
- Commutativité, 72
- Complémentaire, 72
- \neg , 72
- Complémentaire relatif, 71
- \setminus , 71

- Définition en compréhension, 70
- Définition en extension, 69
- Dénombrable, 81
- Diagramme de Venn, 70
- Différence symétrique, 72
- Δ , 72
- Distributivité, 73
- Dualité, 73

- \in , 69
- Élément, 69
- Élément neutre, 73
- Ensemble dénombrable, 81
- Ensemble des parties, 72
- \mathcal{P} , 72
- Ensemble fini, 81
 - cardinal, 81
- Ensemble fondamental, *voir* Univers
- Ensemble vide, 69
- \emptyset , 69
- Ensembles disjoints, 71
- Ensembles en bijection, 81
- Entropie, 54

- Épreuve aléatoire, *voir* Expérience aléatoire
- Équiprobabilité, 14
- Espérance, 57
- \mathbb{E} , 57
- Espace des états, *voir* Univers
- Espace des possibles, *voir* Univers
- Évènement, 4
- Évènement élémentaire, 4
- Évènement certain, 6
- Évènement contraire, 6
- Évènement impossible, 6
- Évènement réalisé, 22
- Évènements incompatibles, 6
- Évènements indépendants, 31
- Expérience aléatoire, 1
- Expérience aléatoire composée, 24

- Factorielle, 82
- Fonction, 77
 - antécédent, 77
 - bijective, 78
 - composition, 78
 - \circ , 78
 - domaine de définition, 77
 - ensemble d'arrivée, 77
 - ensemble de départ, 77
 - image, 77
 - image directe, 77
 - injective, 78
 - réciproque, 78
 - surjective, 78
- Fonction de répartition, 42
 - autres propriétés, 45
 - propriétés fondamentales, 44
 - théorème de la réciproque, 47
- Inclusion, 70

- Incompatible, *voir* Évènements incompatibles
- Indépendance, 31, 32
- \perp , 31
- Indépendance conditionnelle, 35
- Intersection, 71
- \cap , 71
- Involution, 73
- Loi classique discrète
- Bernoulli, 64
 - $\mathcal{B}(p)$, 64
 - binomiale, 65
 - $\mathcal{B}(n, p)$, 65
 - géométrique, 65
 - $\mathcal{G}(p)$, 65
 - Poisson, 66
 - $\mathcal{P}(\lambda)$, 66
 - uniforme, 64
 - $\mathcal{U}(U)$, 64
- Loi d'une variable aléatoire, 39
- propriétés, 41
- Lois de De Morgan, 73
- Mesure de probabilité, 7
- Moment, 63
- centré, 63
 - ordinaire, 63
- n -uplet, 74
- Ω , 1
- Opérations ensemblistes
- Δ , 72
- Opérations ensemblistes
- élément, 69
 - \in , 69
 - élément neutre, 73
 - associativité, 73
 - commutativité, 72
 - complémentaire, 72
 - \neg , 72
 - complémentaire relatif, 71
 - \setminus , 71
 - Différence symétrique, 72
 - dualité, 73
 - inclusion, 70
 - \subset , 70
- intersection, 71
- \cap , 71
- involution, 73
- Lois de De Morgan, 73
- sous-ensemble, 70
 - union, 70
 - \cup , 70
 - union disjointe, 71
- \mathbb{P}_X , 39
- p -liste, 82
- Partie, 70
- Partition, 74
- Permutation, 82
- Probabilité, 7
- Probabilité conditionnelle, 21
- Probabilité uniforme, 14
- Produit cartésien, 74
- \times , 74
- Réduction, 63
- Règle de Bayes, 29
- Règle des probabilités totales, 26
- Sigma additivité, 7
- Sous-ensemble, 70
- \subset , 70
- Support d'une variable aléatoire, 38
- Théorème de la réciproque, 47
- Théorème de transfert
- cas discret, 59
- Union, 70
- \cup , 70
- Union disjointe, 71
- Univers, 1
- fini, 12
- Variable aléatoire, 37
- centrée, 63
 - centrée-réduite, 63
 - centrage, 63
 - centrage et réduction, 63
 - discrète, 53
 - Entropie, 54
 - Espérance, 57
 - \mathbb{E} , 57

- fonction d'une autre, 50
 - loi, 51
 - fonction de répartition, 42
 - loi, 39
 - \mathbb{P}_X , 39
 - moment, 63
 - moment centré, 63
 - moment ordinaire, 63
 - numérique, 42
 - réduction, 63
 - réduite, 63
 - réelle, 42
 - support, 38
 - $X(\Omega)$, 38
 - Variance, 61
 - \mathbb{V} , 61
- Variable aléatoire discrète, 53
- Variance, 61
- \mathbb{V} , 61
- $X(\Omega)$, 38