UNIVERSITÉ DE TOULON
UFR Sciences et Techniques

Statistiques

02/12/2013

W. Aschbacher (*http://aschbacher.univ-tln.fr/*)

MB31 L2 Cours du 1er semestre 2013–2014 (6×2h CM et 8×1:30h TD)

Licence Biologie

Table des matières

1 Eléments de combinatoire

Nous commençons par étudier quelques éléments de l'analyse combinatoire. Elle nous fournira des méthodes de *dénombrement* particulièrement utiles en théorie des probabilités. D'abord, nous allons énoncer quelques principes très simples.

Pour tout ensemble A composé d'un nombre fini d'éléments, nous allons utiliser la notation $\boxed{\mathrm{card}(A)}$ pour désigner le nombre d'éléments dans A.

Enoncé 1.1 *Soient A et B des ensembles composés d'un nombre fini d'éléments.*

(a) **(Règle du produit)** *Pour l'ensemble des paires ordonnées*

$$A \times B := \{(a,b) \,|\, a \in A,\, b \in B\},$$

appelé le **produit cartésien** *de A et B, on a*

$$\mathrm{card}(A \times B) = \mathrm{card}(A) \cdot \mathrm{card}(B),$$

c.-à-d., le nombre de possibilités de former une paire ordonnée (a,b), où $a \in A$ et $b \in B$, est égal au produit des nombres d'éléments dans A et B.

(b) **(Règle de la somme)** *Si $A \cap B = \emptyset$, alors*

$$\mathrm{card}(A \cup B) = \mathrm{card}(A) + \mathrm{card}(B),$$

c.-à-d., si A et B sont des ensembles disjoints, le nombre d'éléments dans leur union est égal à la somme des nombres d'éléments dans A et dans B.

(c) **(Règle d'inclusion-exclusion)** *Si, de manière plus générale, l'intersection de A et B n'est pas nécessairement vide, c.-à-d., si $\mathrm{card}(A \cap B) \geq 0$, on a*

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

1.1 Exemples

Dans cette section, nous allons faire quelques exemples introductoires.

Exemple 1.2 *Sur un rosier (du Jardin des Plantes à Paris) se trouvent 75 coccinelles asiatiques ayant chacune entre 2 et 7 points noirs. Alors, au moins 13 coccinelles ont le même nombre de points.*

Solution Divisons les coccinelles en les 6 groupes de coccinelles ayant entre 2 et 7 points (cf. Fig. 1.2). Si on suppose que chaque groupe consiste en moins de 13 coccinelles, il y aurait au plus

$$6 \cdot 12 = 72 < 75$$

coccinelles sur le rosier ce qui est en contradiction avec l'hypothèse selon laquelle il y a 75 coccinelles sur le rosier. Il en résulte qu'au moins un groupe consiste en au moins 13 coccinelles. $\square$

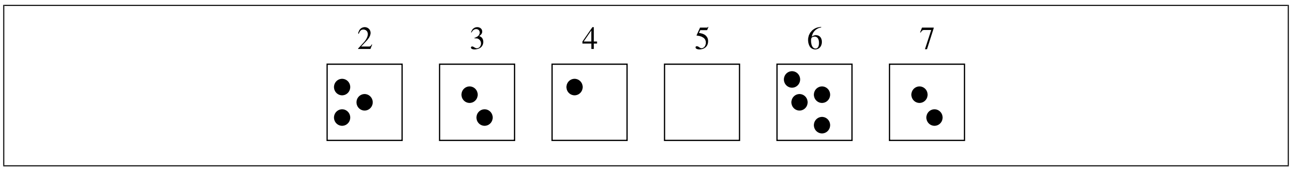

Figure 1.2

Exemple 1.3 *Combien de séquences d'ADN de longueur 10 commencent par ACG et se terminent par A ou T ?*

<u>Solution</u> Le nucléotide de la position 4 peut être choisi parmi 4 nucléotides.

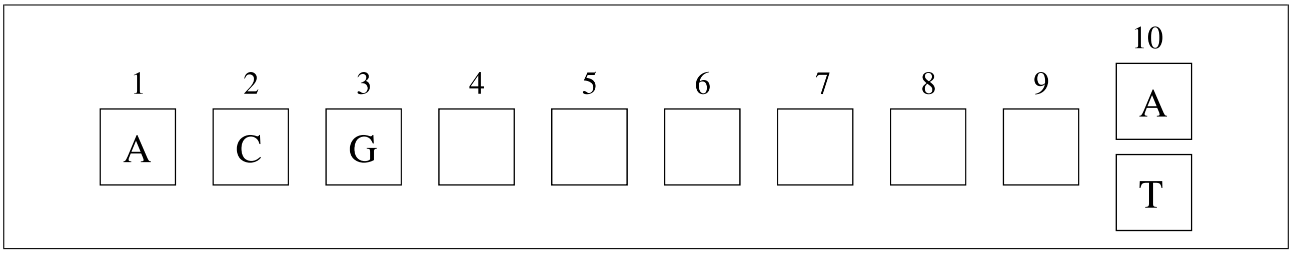

Figure 1.3

La même chose est vraie pour les positions 5 à 9, pendant que pour la dernière position, on a 2 possibilités par hypothèse. Alors, on trouve qu'il existe

$$4^6 \cdot 2 = 8192$$

séquences de ce type. □

Exemple 1.4 *Soient n points dans le plan t.q. aucun triplet de points n'est colinéaire (c.-à-d., se trouve sur la même droite). Combien de droites sont définies par ces points ?*

<u>Solution</u> Soit $\{p_1, \ldots, p_n\}$ l'ensemble des n points différents dans le plan. Comme toute paire de points (p_i, p_j) avec $i \neq j$ définit une droite qui passe par les points p_i et p_j, on s'intéresse au dénombrement des paires de points différents parmi les n points $\{p_1, \ldots, p_n\}$. Pour choisir le premier point du couple, on a n possibilités. Pour choisir le deuxième point du couple, on a $n-1$ possibilités. Alors, on obtient $n(n-1)$ paires.

Mais, comme la paire (p_i, p_j) pour $i \neq j$ définit la même droite que la paire (p_j, p_i), on trouve que ces points définissent

$$\frac{n(n-1)}{2}$$

droites dans le plan. □

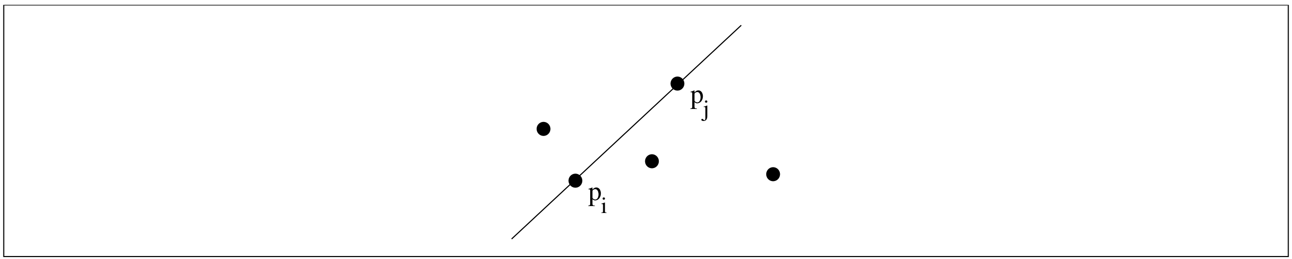

Figure 1.4

Remarque 1.5 *Dans la solution de l'Ex. 1.4, nous avons réduit le problème posé au problème de dénombrement des sous-ensembles de* $\{p_1, \ldots, p_n\}$ *contenant deux éléments. Comme l'ordre des éléments d'un ensemble est insignifiant, l'application qui envoie le sous-ensemble* $\{p_i, p_j\}$ *sur la droite qui passe par* p_i *et* p_j *est une bijection (cf. Rap. 1.13).*

Exemple 1.6 *Le code d'un coffre-fort est composé de cinq chiffres (entre* 0 *et* 9*). Combien y a-t-il de codes qui forment une suite de chiffres croissante ?*

Solution Soit N le nombre recherché de ces codes croissants. Considérons d'abord des codes composés de chiffres différents mais sans la condition de croissance et supposons qu'on fixe 5 chiffres parmi les 10 possibles. Alors, on peut former $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ codes à partir de ces 5 chiffres. Ensuite, nous remarquons que, parmi ces 120 codes, il n'y a qu'un seul code qui est croissant. Alors, comme, d'une part, le nombre de codes à chiffres différents est égal à $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30240$,

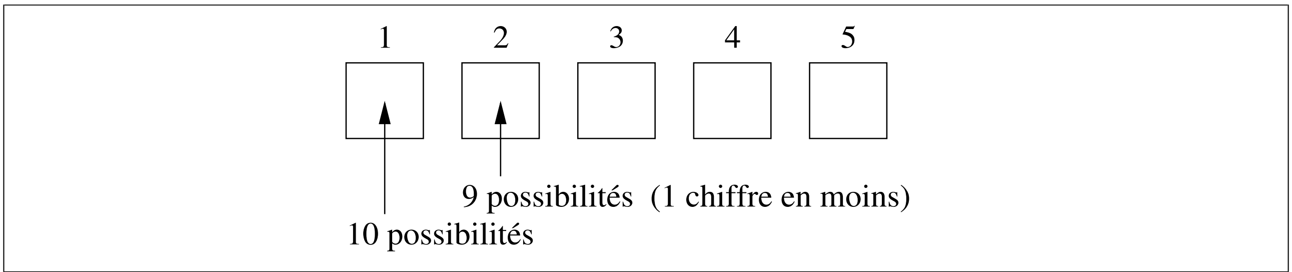

Figure 1.6

et comme, d'autre part, le nombre de codes à chiffres différents est aussi égal à $120 \cdot N$, on obtient

$$N = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 252.$$

$\square$

1.2 Définitions

Nous avons vu dans les exemples précédents qu'une partie du problème ou parfois même le problème entier peut être réduit à un problème de dénombrement de certains objets.

Deux questions importantes se posent alors :

(a) Est-ce que ces objets se répètent ?

(b) Est-ce que ces objets sont ordonnés ?

Pour capter et souligner ces deux propriétés, nous allons faire les définitions suivantes.

Définition 1.7 *Soit A un ensemble avec $\mathrm{card}(A) = n$ et soit $k \in \mathbb{N}^* = \{1, 2, \ldots\}$ avec $k \leq n$. Un* **arrangement (de k éléments de A)** *est un k-uplet*

$$(a_1, a_2, \ldots, a_k) \in A^k := \underbrace{A \times \ldots \times A}_{k \text{ fois}} \text{ t.q. } a_i \neq a_j \text{ si } i \neq j.$$

Nous rappelons que les entrées d'un k-uplet sont ordonnées (cf. En. 1.1 (a) et Rés. 1.12 ci-dessous).

Enoncé 1.8 *Le nombre d'arrangements est égal à*

$$A_n^k := \frac{n!}{(n-k)!},$$

où, pour tout $n \in \mathbb{N} = \{0, 1, 2, \ldots\}$, la **factorielle** *est définie par*

$$n! := \begin{cases} 1, & n = 0, \\ 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n, & n \geq 1. \end{cases}$$

Solution Pour la première entrée du k-uplet, il y a n choix possibles, pour la deuxième, il n'y en a plus que $n-1$, et pour la k-ième, il n'en reste plus que $n-k+1$. Alors, on obtient

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

□

Remarque 1.9 *Si $k = n$, un arrangement s'appelle également une* **permutation** *(au sens habituel du terme), c.-à-d., une bijection $A \to A$ (cf. Rap. 1.13). Alors, on a $A_n^n = n!$.*

Définition 1.10 *Soit A un ensemble avec $\mathrm{card}(A) = n$ et soit $k \in \mathbb{N}^*$ avec $k \leq n$. Une* **combinaison (de k éléments de A)** *est un sous-ensemble $\{a_1, a_2, \ldots, a_k\} \subseteq A$.*

Enoncé 1.11 *Le nombre de combinaisons est égal au* **coéfficient binomial**

$$\boxed{C_n^k := \binom{n}{k} := \frac{n!}{k!(n-k)!}.}$$

<u>Solution</u> Soit C_n^k le nombre de sous-ensemble contenant k éléments d'un ensemble de n éléments. Comme dans Ex. 1.6, on voit que ce nombre est égal au nombre de suites de chiffres croissantes (entre 1 et n) de longueur k. A partir de chacune de ces C_n^k suites, on peut former $k!$ k-uplet. Alors, le nombre de tous les k-uplets d'éléments différents de A, c.-à-d., des arrangements, est égal, d'une part, à $k!C_n^k$, et, d'autre part, d'après En. 1.8, à A_n^k. Il en résulte que

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}.$$

□

A présent, nous allons résumer les propriétés d'une collection d'objets et introduire des définitions et une notation supplémentaire.

Résumé 1.12 *Une collection de k objets $a_1, \ldots, a_k$ (d'un ensemble de n éléments) dont les objets sont :*

(a) <u>ordonnés, sans répétition</u> est un arrangement : $\boxed{(a_1, \ldots, a_k) \text{ avec } a_i \neq a_j \text{ pour } i \neq j}$ *(dont le nombre est égal à A_n^k).*

(b) <u>non ordonnés, sans répétition</u> est une combinaison : $\boxed{\{a_1, \ldots, a_k\}}$ *(dont le nombre est égal à C_n^k).*

(c) <u>non ordonnés, avec répétition</u> est une **combinaison avec répétition** *:* $\boxed{\langle a_1, \ldots, a_k \rangle}$ *(dont le nombre est égal à C_{n+k-1}^k).*

(d) <u>ordonnés, avec répétition</u> est une **permutation avec répétition** *:* $\boxed{(a_1, \ldots, a_k)}$ *(dont le nombre est égal à n^k).*

Rappels

Rappel 1.13 *Soient A et B deux ensembles. Une application $f : A \to B$ s'appelle* **injective** *si, pour tout élément b dans l'***ensemble d'arrivée** *B, il existe <u>au plus</u> un* **antécédent** *a dans l'***ensemble de définition** *A, c.-à-d., un $a \in A$ t.q. $b = f(a)$. L'application f s'appelle* **surjective** *si, pour tout $b \in B$, il existe <u>au moins</u> un $a \in A$ t.q. $b = f(a)$. L'application f s'appelle* **bijective** *(ou une* **bijection***) si, pour tout $b \in B$, il existe <u>un et un seul</u> $a \in A$ t.q. $b = f(a)$,c.-à-d., f est bijectif ssi f est injecif et surjectif, cf. Fig. 1.13 :*

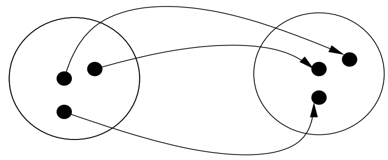

Figure 1.13

2 Probabilités

Une expérience aléatoire est une expérience qui, répétée dans des conditions qui semblent identiques, peut conduire à des résultats différents. Le résultat d'une telle expérience s'appellera un événement.
La quantification de la possibilité qu'un tel événement a de se réaliser correspond à la notion intuitive de probabilité.

2.1 Modèle probabiliste

Définition 2.1 *L'ensemble des résultats possibles d'une expérience aléatoire s'appelle* **univers** *et sera noté* $\boxed{\Omega}$.

Exemple 2.2

(a) *A l'expérience aléatoire du jet d'un dé à six faces numerotées, on peut associer l'univers fini*

$$\Omega := \{1, 2, \dots, 6\}.$$

Mais on pourrait également s'intéresser aux résultats pairs et impairs de cette expérience auquel cas on rentiendrait l'univers fini

$$\Omega := \{\text{pair}, \text{impair}\}.$$

(b) *Si on tire une carte d'un jeu de* 32 *cartes, en fonction de la question posée, on peut retenir les univers finis*

$$\Omega := \{7, 8, 9, 10, \text{V}, \text{D}, \text{R}, \text{As}\},$$
$$\Omega := \{\text{trèfle}, \text{carreau}, \text{coeur}, \text{pique}\},$$
$$\Omega := \{\text{rouge}, \text{noir}\}.$$

(c) *En général, l'univers* Ω *ne doit pas être fini. Il peut aussi être infini dénombrable ou continu.*

A présent, nous introduisons le cadre pour décrire les résultats d'une expérience aléatoire. Pour tout ensemble $\mathcal{M}$, nous allons noter $\boxed{\mathcal{P}(\mathcal{M})}$ l'**ensemble des parties** de $\mathcal{M}$, c.-à-d., la famille qui contient tous les sous-ensembles de $\mathcal{M}$.

Définition 2.3 *Soit Ω un univers donné. Une famille $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ de parties de Ω s'appelle* **tribu** *sur Ω si elle satisfait les conditions suivantes :*

(T1) $\Omega \in \mathcal{A}$

(T2) Si $A \in \mathcal{A}$*, alors* $A^c \in \mathcal{A}$*.*

(T3) Si $A_n \in \mathcal{A}$ *pour tout* $n \in \mathbb{N}$*, alors* $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$*.*

Nous avons utilisé la notation A^c *pour le complémentaire de* A *dans* Ω*, cf. Rap. 2.22.*
Les éléments de la tribu $\mathcal{A}$ *s'appellent les* **événements** *de l'expérience aléatoire. Les événements qui ne contiennent qu'un seul élément de* Ω *s'appellent les* **événements élémentaires***. Le couple* $\boxed{(\Omega, \mathcal{A})}$ *s'appelle un* **espace probabilisable***.*

Remarque 2.4

(a) *Notons d'abord que, même si* Ω *est fini, le nombre*

$$\mathrm{card}(\mathcal{P}(\Omega)) = 2^{\mathrm{card}(\Omega)}$$

est rapidement très grand. Dans ce cas, il peut être souhaitable de ne considérer qu'une famille $\mathcal{A}$ *restreinte de parties de* Ω*, c.-à-d.,* $\mathcal{A} \subset \mathcal{P}(\Omega)$*.*
Pour que le résultat des opérations ensemblistes comme l'union, l'intersection et le complémentaire soit encore un événement, il est nécessaire que $\mathcal{A}$ *soit* <u>*fermé*</u> *p.r. à ces opérations ce qui est garanti si* $\mathcal{A}$ *est une* <u>*tribu*</u>*.*

(b) *Néanmoins, dans la plupart des applications, on choisira* $\mathcal{P}(\Omega)$ *comme ensemble des événements si* Ω *est fini et même si* Ω *est* <u>*infini dénombrable*</u> *(par contre, si* Ω *est non dénombrable, nous ne choisirons jamais cette tribu).*

Exemple 2.5

(a) *La tribu* $\mathcal{A} = \mathcal{P}(\Omega)$ *s'appelle la* **tribu triviale***.*

(b) *La tribu* $\mathcal{A} = \{\emptyset, \Omega\}$ *s'appelle la* **tribu grossière***.*

(c) *Soit* $A \in \mathcal{P}(\Omega)$ *fixé. Alors,* $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$ *est une tribu.*

La terminologie d'espace probabilisable suppose que nous pouvons associer une <u>probabilité</u> au couple $(\Omega, \mathcal{A})$ traduisant par un nombre les possibilités de réalisation des événements de l'expérience aléatoire à laquelle nous nous intéressons.

Définition 2.6 *Soit* $(\Omega, \mathcal{A})$ *un espace probabilisable. Une application*

$$P : \mathcal{A} \to \mathbb{R}$$

s'appelle une **probabilité** *sur* $(\Omega, \mathcal{A})$ *si elle satisfait les conditions suivantes :*

(P1) $P(A) \in [0, 1]$ *pour tout* $A \in \mathcal{A}$

(P2) $P(\Omega) = 1$

(P3) P *est* σ**-additif***, c.-à-d., si* $A_n \in \mathcal{A}$ *pour tout* $n \in \mathbb{N}$ *t.q.*

$$A_n \cap A_m = \emptyset \quad \textit{pour tout } n \neq m,$$

la probabilité de l'union est égale à la somme des probabilités, c.-à-d., on a

$$\boxed{P\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} P(A_n).}$$

Le triplet $\boxed{(\Omega, \mathcal{A}, P)}$ *s'appelle un* **espace probabilisé**.

Comme conséquence de Déf. 2.6, on a les propriétés suivantes.

Enoncé 2.7 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé. Alors :*

(a) $P(\emptyset) = 0$

(b) $P(A^c) = 1 - P(A)$ *pour tout* $A \in \mathcal{A}$

(c) $P(A) \leq P(B)$ *pour tout* $A, B \in \mathcal{A}$ *t.q.* $A \subseteq B$

(d) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *pour tout* $A, B \in \mathcal{A}$

Solution Cf. Exr. 5 □

En ce qui concerne En. 2.7 *(d)*, nous notons que $A \cap B \in \mathcal{A}$ parce que, d'après la loi de De Morgan, on peut écrire

$$A \cap B = (A^c \cup B^c)^c \in \mathcal{A}.$$

Dans l'énoncé suivant, nous allons introduire un cas particulier important d'une probabilité dite l'équiprobabilité.

Enoncé 2.8 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé, où*

$$\begin{aligned}\Omega &:= \{\omega_1, \ldots, \omega_n\},\\ \mathcal{A} &:= \mathcal{P}(\Omega).\end{aligned}$$

(a) Si P *est connu sur les événements élémentaires (tous les sous-ensembles de* Ω *contenant un seul élément sont des événements élémentaires si* $\mathcal{A} = \mathcal{P}(\Omega)$*), on peut déterminer la probabilité de tout événement* $A \in \mathcal{A}$ *par*

$$P(A) = \sum_{i\in\{1,\ldots,n\}:\ \omega_i\in A} P(\{\omega_i\}).$$

Dans cette situation, la probabilité d'un événement quelconque est donc la somme des probabilités de tous les événements élémentaires qui y sont inclus.

(b) *Si tous les événements élémentaires ont la même probabilité, c.-à-d., si*

$$P(\{\omega_i\}) = \frac{1}{n} \quad \textit{pour tout } i \in \{1, \ldots, n\},$$

la probabilité d'un événement quelconque est donnée par

$$\boxed{P(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)} \quad \textit{pour tout } A \in \mathcal{A}.} \tag{2.1}$$

La probabilité définie par la propriété (2.1) *s'appelle l'***équiprobabilité***.*

Remarque 2.9

(a) *Une propriété importante de l'équiprobabilité est le fait que, pour son calcul, on est ramené à un problème de* <u>*dénombrement*</u> *(cf. chapitre 1).*

(b) *En général, une probabilité n'est* <u>*pas*</u> *de la forme* (2.1) *mais n'est spécifiée que par les propriétés (P1)–(P3) de Déf. 2.6.*

Exemple 2.10 *Une expérience aléatoire consiste à tirer deux fois dans une urne qui contient une boule blanche et deux boules noires (identiques).*

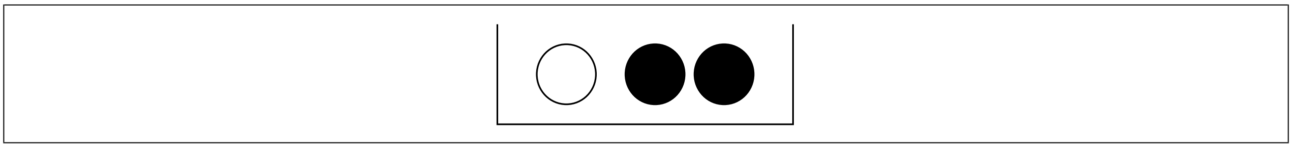

Figure 2.10

La première boule n'est pas remise dans l'urne mais est remplacée par une boule de la couleur opposée.

Nous devons faire un choix pour modéliser cette expérience aléatoire.

(a) *Quel choix de Ω est naturel ?*

(b) *Soit $\mathcal{A} := \mathcal{P}(\Omega)$ et soit la probabilité P sur $\mathcal{A}$ définie par la multiplication des équiprobabilités pour chacun des tirages.*

Quelle est la valeur de P pour chaque événement élémentaire ?

<u>Solution</u>

(a) Nous définissons un **singleton** de Ω (c.-à-d., un sous-ensemble de Ω qui ne contient qu'un seul élément de Ω) comme une paire ordonnée de couleurs dont la première entrée est la couleur de la première boule tirée et la deuxième entrée est la couleur de la boule tirée ensuite. Si la première boule tirée est blanche, la deuxième boule tirée sera forcément noire. Si la première boule tirée est noire, la deuxième boule peut être blanche ou noire. Alors, nous définissons l'univers comme

$$\Omega := \{(\mathrm{B}, \mathrm{N}), (\mathrm{N}, \mathrm{B}), (\mathrm{N}, \mathrm{N})\}.$$

(b) Comme $\mathcal{A} = \mathcal{P}(\Omega)$, tous les singletons de Ω sont des événements élémentaires. Pour leur probabilité, l'hypothèse implique que

$$\begin{aligned}
P(\{(\mathsf{B},\mathsf{N})\}) &= \frac{1}{3} \cdot 1 = \frac{1}{3}, \\
P(\{(\mathsf{N},\mathsf{B})\}) &= \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}, \\
P(\{(\mathsf{N},\mathsf{N})\}) &= \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}.
\end{aligned}$$

On verifie qu'en effet, on obtient

$$P(\{(\mathsf{B},\mathsf{N})\}) + P(\{(\mathsf{N},\mathsf{B})\}) + P(\{(\mathsf{N},\mathsf{N})\}) = 1,$$

et donc $P(\Omega) = 1$ parce que Ω est égal à l'union de ses sous-ensembles $\{(\mathsf{B},\mathsf{N})\}$, $\{(\mathsf{N},\mathsf{B})\}$ et $\{(\mathsf{N},\mathsf{N})\}$ qui sont disjoints deux à deux, c.-à-d., la propriété *(P2)* de Déf. 2.6 est satisfaite.

□

2.2 Probabilités conditionnelles

Définition 2.11 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et soit $B \in \mathcal{A}$ t.q. $P(B) > 0$. Pour tout $A \in \mathcal{A}$, la* **probabilité conditionnelle de A sachant** *B, notée $P(A|B)$, est définie par*

$$\boxed{P(A|B) := \frac{P(A \cap B)}{P(B)},}$$

où nous rappelons que $A \cap B \in \mathcal{A}$ (cf. après En. 2.7).

Enoncé 2.12 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et soit $B \in \mathcal{A}$ fixé t.q. $P(B) > 0$. Alors, l'application $P(\cdot\,|B) : \mathcal{A} \to \mathbb{R}$ est une probabilité.*

Solution Nous allons vérifier les propriétés *(P1)–(P3)* de Déf. 2.6 l'une après l'autre :

(P1) D'après En. 2.7 *(c)*, on a $P(A \cap B) \leq P(B)$ parce que $A \cap B \subseteq B$. Comme, en plus, $P(A|B) \geq 0$, on obtient

$$0 \leq P(A|B) = \frac{P(A \cap B)}{P(B)} \leq \frac{P(B)}{P(B)} = 1.$$

Alors, on trouve que $P(\cdot\,|B) : \mathcal{A} \to [0,1]$.

(P2) Sur l'univers Ω, nous obtenons

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

(P3) Si $A_n \in \mathcal{A}$ pour tout $n \in \mathbb{N}$ et si $A_n \cap A_m = \emptyset$ pour tout $n \neq m$, alors, on calcule

$$\begin{aligned} P\Big(\bigcup\nolimits_{n\in\mathbb{N}} A_n \Big| B\Big) &= \frac{P\big(\big[\bigcup_{n\in\mathbb{N}} A_n\big] \cap B\big)}{P(B)} = \frac{P\big(\bigcup_{n\in\mathbb{N}} \big[A_n \cap B\big]\big)}{P(B)} = \frac{\sum_{n\in\mathbb{N}} P\big(A_n \cap B\big)}{P(B)} \\ &= \sum_{n\in\mathbb{N}} \frac{P\big(A_n \cap B\big)}{P(B)} = \sum_{n\in\mathbb{N}} P(A_n|B), \end{aligned}$$

où, dans la troisième égalité, nous avons utilisé la propriété *(P3)* de la probabilité P.

□

Exemple 2.13 *On lance trois fois une pièce de monnaie et on s'intéresse aux deux résultats suivants :*

<u>*Résultat 1 :*</u> *Obtenir au moins deux fois face.*

<u>*Résultat 2 :*</u> *Obtenir face au premier lancer.*

(a) Donner un modèle probabiliste pour décrire cette expérience aléatoire.

(b) Dans le modèle retenu, quelle est la probabilité d'obtenir au moins deux fois face sachant qu'on a obtenu face au premier lancer ?

<u>Solution</u>

(a) Afin de mettre en place un modèle probabiliste pour cette expérience aléatoire, nous choisissons d'abord l'univers

$$\Omega := \{\mathsf{p}, \mathsf{f}\}^3,$$

où p et f représentent respectivement pile et face. Un élément $\omega \in \Omega$ est donc un triplet $\omega = (\alpha_1, \alpha_2, \alpha_3)$ (ordonné, les entrées pouvant se répéter, c.-à-d., une permutation avec répétition, cf. Rés. 1.12 *(d)*), où $\alpha_i \in \{\mathsf{p}, \mathsf{f}\}$ pour $i \in \{1, 2, 3\}$. Ensuite, comme tribu sur Ω, nous retenons la tribu triviale

$$\mathcal{A} := \mathcal{P}(\Omega).$$

Finalement, pour compléter les données nécessaires pour définir un espace probabilisé, nous choisissons l'équiprobabilité sur $(\Omega, \mathcal{A})$, c.-à-d., d'après En. 2.8, on définit

$$P(A) := \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)} = \frac{\mathrm{card}(A)}{\mathrm{card}(\{\mathsf{p}, \mathsf{f}\}^3)} = \frac{1}{2^3}\,\mathrm{card}(A) \quad \text{pour tout } A \in \mathcal{A}.$$

Les événements $A_1, A_2 \in \mathcal{A}$ qui représentent *Résultat 1* et *Résultat 2* dans ce modèle probabiliste s'écrivent comme

$$\begin{aligned} A_1 &:= \{(\mathsf{f}, \mathsf{f}, \mathsf{p}), (\mathsf{f}, \mathsf{p}, \mathsf{f}), (\mathsf{p}, \mathsf{f}, \mathsf{f}), (\mathsf{f}, \mathsf{f}, \mathsf{f})\}, \\ A_2 &:= \{(\mathsf{f}, \mathsf{p}, \mathsf{p}), (\mathsf{f}, \mathsf{f}, \mathsf{p}), (\mathsf{f}, \mathsf{p}, \mathsf{f}), (\mathsf{f}, \mathsf{f}, \mathsf{f})\}. \end{aligned}$$

Dans ce modèle, ces événements ont donc la probabilité

$$
\begin{aligned}
P(A_1) &= \frac{1}{8}\,\mathrm{card}(A_1) = \frac{4}{8} = \frac{1}{2},\\
P(A_2) &= \frac{1}{8}\,\mathrm{card}(A_2) = \frac{1}{2}.
\end{aligned}
$$

(b) Calculons maintenant la probabilité conditionnelle de A_1 sachant A_2. Comme

$$A_1 \cap A_2 = \{(\mathsf{f},\mathsf{f},\mathsf{p}), (\mathsf{f},\mathsf{p},\mathsf{f}), (\mathsf{f},\mathsf{f},\mathsf{f})\},$$

nous pouvons en déduire que

$$P(A_1|A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = \frac{\frac{1}{8}\,\mathrm{card}(A_1 \cap A_2)}{\frac{1}{8}\,\mathrm{card}(A_2)} = \frac{3}{4},$$

et donc, $P(A_1|A_2) = \frac{3}{4} > \frac{1}{2} = P(A_1)$.

□

La formule suivante est une conséquence directe de la définition de la probabilité conditionnelle (cf. Déf. 2.11).

Enoncé 2.14 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et soient* $A_1, \ldots, A_n \in \mathcal{A}$ *des événements t.q.* $P(A_1 \cap \ldots \cap A_{n-1}) > 0$*. Alors, on a la* **formule des probabilités composées***,*

$$\boxed{P(A_1 \cap \ldots \cap A_n) = P(A_1) \prod_{i=2}^{n} P(A_i | A_1 \cap \ldots \cap A_{i-1}).}$$

Solution Nous remarquons d'abord que, d'après En. 2.7 *(c)*, on a

$$P(A_1) \geq P(A_1 \cap A_2) \geq \ldots \geq P(A_1 \cap \ldots \cap A_{n-1}) > 0,$$

ce qui implique que $P(A_i|A_1 \cap \ldots \cap A_{i-1})$ est bien défini parce que nous pouvons diviser par $P(A_1 \cap \ldots \cap A_{i-1}) > 0$ pour tout $i \in \{2, \ldots, n\}$ (cf. Déf. 2.11). Alors, comme naturellement $A_1 \cap A_2 = A_2 \cap A_1$, on peut écrire

$$P(A_1 \cap A_2) = P(A_1)\,\frac{P(A_2 \cap A_1)}{P(A_1)} = P(A_1)P(A_2|A_1).$$

En itérant cette formule, on obtient pour les trois événements $A_1, A_2, A_3 \in \mathcal{A}$ que

$$
\begin{aligned}
P(A_1 \cap A_2 \cap A_3) &= P((A_1 \cap A_2) \cap A_3) = P(A_1 \cap A_2)P(A_3|A_1 \cap A_2)\\
&= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2),
\end{aligned}
$$

et on arrive à la formule générale par induction complète. □

La décomposition suivante d'un univers Ω sera utilisée ci-dessous.

Définition 2.15 *Un* **système complet d'événements** *d'un espace probabilisé* $(\Omega, \mathcal{A}, P)$ *est une famille finie* $A_1, \ldots, A_n \in \mathcal{A}$ *ayant les propriétés suivantes :*
(S1) $\Omega = \bigcup_{i=1}^n A_i$
(S2) $A_i \cap A_j = \emptyset$ *pour tout* $i, j \in \{1, \ldots, n\}$ *avec* $i \neq j$
(S3) $P(A_i) > 0$ *pour tout* $i \in \{1, \ldots, n\}$
(S4) $\sum_{i=1}^n P(A_i) = 1$
Nous remarqu'ons que (S4) est une conséquence de (S1), (S2) et Déf. 2.6 (P2) et (P3).

En décomposant un événement donné p.r. à un tel système complet d'événements, on peut calculer sa probabilité de manière suivante.

Enoncé 2.16 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé,* $B \in \mathcal{A}$ *fixé et* $A_1, \ldots, A_n \in \mathcal{A}$ *un système complet d'événements. Alors, on a la* **formule de la probabilité totale***,*

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

Solution En utilisant Déf. 2.15 *(S1)*, on peut écrire

$$B = B \cap \Omega = B \cap (A_1 \cup \ldots \cup A_n) = (B \cap A_1) \cup \ldots \cup (B \cap A_n).$$

Alors, d'après Déf. 2.15 *(S2)*, *(S3)* et Déf. 2.6 *(P3)*, on obtient

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

$\square$

En utilisant la formule des probabilités composées (cf. En. 2.14) et la formule de la probabilité totale (cf. En. 2.16), nous arrivons à la formule et au théorème de Bayes.

Enoncé 2.17 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $A, B \in \mathcal{A}$ *t.q.* $P(A) > 0$ *et* $P(B) > 0$. *Alors, on a la* **formule de Bayes***,*

$$\boxed{P(A|B) = \frac{P(A)P(B|A)}{P(B)}.}$$

Si $A_1, \ldots, A_n \in \mathcal{A}$ *est un système complet d'événements, alors, on a le* **théorème de Bayes***,*

$$\boxed{P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^n P(A_j)P(B|A_j)} \quad \textit{pour tout } i \in \{1, \ldots, n\}.}$$

Solution D'après En. 2.14, on a $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$ et alors, on trouve la formule de Bayes,

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

Ensuite, soit $A_1, \ldots, A_n \in \mathcal{A}$ est un système complet d'événements (cf. Déf. 2.15). Alors, si, dans la formule de Bayes, on remplace A par A_i et $P(B)$ par la formule de la probabilité totale (cf. En. 2.16), on arrive au théorème de Bayes. □

Exemple 2.18 *Une urne contient deux pièces de monnaie. Si on lance la pièce no.* 1*, on a une probabilité* $\frac{1}{2}$ *d'obtenir face (pièce équilibrée) pendant que pour la pièce no.* 2*, on a une probabilité* $\frac{1}{3}$ *d'obtenir face (pièce biaisée).*

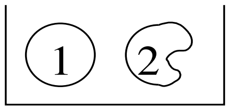

Figure 2.18

On pioche une pièce au hasard et on la lance. Supposons qu'on obtient face. Quelle est la probabilité que la pièce no. 1 *a été lancée ?*

Solution Nous commençons par mettre en place un espace probabilisé $(\Omega, \mathcal{A}, P)$. Comme notre expérience aléatoire consiste en deux étapes, "piocher" et "lancer", nous choisissons l'univers

$$\Omega := \{1, 2\} \times \{\mathsf{p}, \mathsf{f}\}.$$

Ensuite, comme tribu, nous retenons la tribu trivial $\mathcal{A} := \mathcal{P}(\Omega)$.
Finalement, la probabilité $P : \mathcal{A} \to [0, 1]$ nous est donnée dans la description de l'expérience. Comme les résultats "piocher pièce no. 1 et obtenir pile au face au lancer", "piocher pièce no. 2 et obtenir pile au face au lancer", "piocher pièce no. 1 ou 2 et obtenir pile au lancer" et "piocher pièce no. 1 ou 2 et obtenir face au lancer" se décrivent par

$$\begin{aligned} A_1 &:= \{(1, \mathsf{p}), (1, \mathsf{f})\}, \\ A_2 &:= \{(2, \mathsf{p}), (2, \mathsf{f})\}, \\ A_p &:= \{(1, \mathsf{p}), (2, \mathsf{p})\}, \\ A_f &:= \{(1, \mathsf{f}), (2, \mathsf{f})\}, \end{aligned}$$

le texte nous fournit que la probabilité admet les valeurs

$$P(A_1) := \frac{1}{2}, \quad P(A_2) := \frac{1}{2},$$

ce qui décrit l'équiprobabilité pour la premiére étape "piocher" de l'expérience. En ce qui concerne la deuxième étape "lancer", on a

$$P(A_f|A_1) := \frac{1}{2}, \quad P(A_f|A_2) := \frac{1}{3},$$

exprimant le fait que la pièce no. 2 est biaisée (noter que ces conditions nous permettent de déterminer la probabilité de tous les événements élémentaires mais nous n'en avons pas besoin pour répondre à la question posée ici).
Comme $A_1, A_2 \in \mathcal{A}$ constituent un système complet d'événements (cf. Déf. 2.15), nous pouvons appliquer le théorème de Bayes (cf. En. 2.17) pour déterminer la probabilité de l'événement qui nous intéresse. On obtient alors

$$P(A_1|A_f) = \frac{P(A_1)P(A_f|A_1)}{P(A_1)P(A_f|A_1) + P(A_2)P(A_f|A_2)} = \frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{3}} = \frac{3}{5}.$$

$\square$

2.3 Indépendance

Définition 2.19 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé. Deux événements* $A_1, A_2 \in \mathcal{A}$ *s'appellent* **indépendants** *(p.r. à la probabilité* P*) si*

$$\boxed{P(A_1 \cap A_2) = P(A_1)P(A_2),}$$

c.-à-d., la probabilité de <u>*réalisation simultanée de deux événements*</u> *indépendants est égale au produit des probabilités que chacun de ces événements se produise séparément.*
Les événements $A_1, \ldots, A_n \in \mathcal{A}$ *avec* $n \geq 2$ *s'appellent* **indépendants (deux à deux)** *(p.r. à* P*) si*

$$\boxed{P(A_i \cap A_j) = P(A_i)P(A_j) \quad \textit{pour tout } i, j \in \{1, \ldots, n\} \textit{ avec } i \neq j.}$$

Cette notion d'indépendance est à distinguer de la notion suivante qui est plus forte.

Définition 2.20 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé. Les événements* $A_1, \ldots, A_n \in \mathcal{A}$ *avec* $n \geq 2$ *s'appellent* **mutuellement indépendants** *(p.r. à* P*) si, pour tout* $k \in \{2, \ldots, n\}$ *et tout* $1 \leq i_1 < \ldots < i_k \leq n$*, on a*

$$\boxed{P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_k}).}$$

Remarque 2.21

(a) *Si les événements sont mutuellement indépendants, ils sont indépendants (deux à deux) parce qu'on peut choisir* $k = 2$ *dans Déf. 2.20.*

(b) *Si les événements sont indépendants (deux à deux), ils ne sont pas mutuellement indépendants en général.*

Exemple : Soit $\Omega := \{1,2,3,4\}$, $\mathcal{A} := \mathcal{P}(\Omega)$ *et* P *l'équiprobabilité. En plus, soient* $A_1 := \{1,2\}$, $A_2 := \{1,3\}$ *et* $A_3 := \{1,4\}$. *Alors,* A_1, A_2, A_3 *sont indépendants (deux à deux),*

$$\begin{aligned} P(A_1 \cap A_2) &= \frac{1}{4}, \quad P(A_1)P(A_2) = \frac{1}{2} \cdot \frac{1}{2}, \\ P(A_1 \cap A_3) &= \frac{1}{4}, \quad P(A_1)P(A_3) = \frac{1}{2} \cdot \frac{1}{2}, \\ P(A_2 \cap A_3) &= \frac{1}{4}, \quad P(A_2)P(A_3) = \frac{1}{2} \cdot \frac{1}{2}, \end{aligned}$$

mais ils ne sont pas mutuellements indépendants parce que

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{4}, \quad P(A_1)P(A_2)P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

(c) *Si les événements* $A_1, \ldots, A_n \in \mathcal{A}$ *satisfont* $P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot \ldots \cdot P(A_n)$, *ils ne sont pas indépendants (deux à deux) en général.*

Exemple : Soit $\Omega := \{1,\ldots,6\}^2$, $\mathcal{A} := \mathcal{P}(\Omega)$ *et* P *l'équiprobabilité. En plus, soient* $A_1 := \{(i,j) \in \Omega \,|\, j \in \{1,2,5\}\}$, $A_2 := \{(i,j) \in \Omega \,|\, j \in \{4,5,6\}\}$ *et* $A_3 := \{(i,j) \in \Omega \,|\, i+j = 9\}$. *Alors, on a*

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(\{(4,5)\}) = \frac{1}{36}, \\ P(A_1)P(A_2)P(A_3) &= \frac{3 \cdot 6}{36} \cdot \frac{3 \cdot 6}{36} \cdot \underbrace{\frac{4}{36}}_{\text{Cf. Exr. 7}} = \frac{1}{36}, \end{aligned}$$

mais les événements ne sont pas indépendants (deux à deux) parce que

$$P(A_1 \cap A_2) = \frac{6}{36} = \frac{1}{6}, \quad P(A_1)P(A_2) = \frac{3 \cdot 6}{36} \cdot \frac{3 \cdot 6}{36} = \frac{1}{4},$$

et de manière analogue,

$$\begin{aligned} P(A_1 \cap A_3) &= \frac{1}{36}, \quad P(A_1)P(A_3) = \frac{1}{18}, \\ P(A_2 \cap A_3) &= \frac{1}{12}, \quad P(A_2)P(A_3) = \frac{1}{18}. \end{aligned}$$

Rappels

Rappel 2.22 *Soit* A *un sous-ensemble de l'ensemble* Ω. *Le* **complémentaire** *de* A *(p.r. à* Ω*), noté* A^c, *est défini par* $\boxed{A^c := \{\omega \in \Omega \,|\, \omega \notin A\}}$. *Soit* A *un sous-ensemble de l'ensemble* B. *La différence ensembliste* B **moins** A, *notée* $B \setminus A$, *est définie par* $\boxed{B \setminus A := \{b \in B \,|\, b \notin A\}}$ *(nous notons que* $A^c = \Omega \setminus A$ *et* $B \setminus A = B \cap A^c$*). En plus, on a la* **loi de De Morgan**, *c.-à-d.,* $(A \cup B)^c = A^c \cap B^c$.

3 Variables aléatoires discrètes

Afin de pouvoir effectuer des calculs numériques, comme p. ex. celui de la moyenne des différents résultats possibles d'une expérience aléatoire, on veut associer une valeur numérique à tout résultat d'une telle expérience.
Nous commençons par formaliser ce concept d'un codage numérique pour le cas où le codage n'admet qu'un ensemble dénombrable de valeurs.

3.1 Définition

Définition 3.1 *Soit* $(\Omega, \mathcal{A})$ *un espace probabilisable. Une application*

$$X : \Omega \to \mathbb{R}$$

est une **variable aléatoire discrète (vad)** *si elle satisfait les conditions suivantes :*

(VAD 1) $X(\Omega)$ *est dénombrable.*

(VAD 2) $X^{-1}(I) \in \mathcal{A}$ *pour tout* $I \in \mathcal{P}(X(\Omega))$

L'ensemble $X(\Omega)$ *s'appelle l'***univers image** *(cf. Rap. 3.24).*

Remarque 3.2

(a) Les conditions (VAD 1) et (VAD 2) expriment que l'univers image est dénombrable et que l'image réciproque (cf. Rap. 3.24) de tout $I \in \mathcal{P}(X(\Omega))$ *est un événement.*

(b) Si $\mathcal{A}$ *est la tribu triviale (cf. Ex. 2.5 (a)), toute application* $X : \Omega \to \mathbb{R}$ *à univers image dénombrable est une vad.*

3.2 Probabilité image

Si, en plus, on dispose d'une probabilité sur l'espace probabilisable $(\Omega, \mathcal{A})$, on peut définir une probabilité pour l'univers image.

Définition 3.3 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vad. Pour*

$$\mathcal{A}_X := \mathcal{P}(X(\Omega)),$$

appelé la **tribu image***, l'application* $P_X : \mathcal{A}_X \to [0, 1]$*, définie par*

$$\boxed{P_X(I) := P(X^{-1}(I)) \quad \textit{pour tout } I \in \mathcal{A}_X,}$$

s'appelle la **probabilité image (de** P **par** X**)***. On utilisera souvent les notations simplifiées*

$$\begin{aligned} P(X = x) &:= P_X(x) := P_X(\{x\}), \\ P(X < x) &:= P_X(\{y \in X(\Omega) \mid y < x\}), \\ P(X \in I) &:= P_X(I), \ \textit{etc.} \end{aligned}$$

Enoncé 3.4 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vad. Alors, le triplet* $(X(\Omega), \mathcal{A}_X, P_X)$ *est un espace probabilisé, appelé l'***espace probabilisé image***.*

<u>Solution</u> Cf. Exr. 12 □

On a le schéma suivant :

	Espace probabilisé "*abstrait*"	Espace probabilisé "*numérique*"
Univers	Ω	$X(\Omega)$
Tribu	$\mathcal{A}$	$\mathcal{A}_X$
Probabilité	P	P_X

Exemple 3.5 *Considérons l'expérience aléatoire qui consiste à jeter un dé équilibré à six faces. Comme espace probabilisé* $(\Omega, \mathcal{A}, P)$*, nous retenons*

$$\Omega := \{1, \ldots, 6\}, \quad \mathcal{A} := \mathcal{P}(\Omega), \quad P \text{ l'équiprobabilité.}$$

En plus, on s'intéresse au codage numérique $X : \Omega \to \mathbb{R}$ *défini, pour tout* $\omega \in \Omega$*, par*

$$X(\omega) := \begin{cases} 0, & \omega \text{ est impair,} \\ 1, & \omega \text{ est pair.} \end{cases}$$

Montrer que X *est une vad et déterminer l'espace probabilisé image.*

<u>Solution</u> L'univers et la tribu image sont donnés par

$$\begin{aligned} X(\Omega) &= \{0, 1\}, \\ \mathcal{A}_X &= \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}. \end{aligned}$$

D'après Rem. 3.2 *(b)*, comme l'univers image est dénombrable, X est une vad. En plus, en ce qui concerne la probabilité image, on a $P_X(\emptyset) = 0$ et $P_X(\{0, 1\}) = 1$, et on trouve

$$\begin{aligned} P_X(0) &= P(X^{-1}(\{0\})) = P(\{1, 3, 5\}) = \frac{\mathrm{card}(\{1, 3, 5\})}{\mathrm{card}(\Omega)} = \frac{3}{6} = \frac{1}{2}, \\ P_X(1) &= P(X^{-1}(\{1\})) = P(\{2, 4, 6\}) = \frac{\mathrm{card}(\{2, 4, 5\})}{\mathrm{card}(\Omega)} = \frac{3}{6} = \frac{1}{2}. \end{aligned}$$

□

A présent, nous allons définir la notion de loi d'une vad.

Définition 3.6 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vad. L'univers image étant dénombrable (cf. (VAD 1) de Déf. 3.1), nous numérotons ses éléments par*

$$X(\Omega) = \{x_0, x_1, x_2, \ldots\},$$

et nous supposons que $x_0 < x_1 < x_2 < \ldots$*.*

En utilisant la notation

$$\boxed{p_i := P_X(x_i) \quad \textit{pour tout } i \in \mathbb{N},}$$

l'ensemble des couples

$$\{(x_i, p_i)\}_{i\in\mathbb{N}}$$

s'appelle la **loi (de probabilité)** *de la vad X. De manière plus générale, on appelera également l'application P_X loi (de probabilité) de la vad X.*

Exemple 3.7

(a) Si l'univers image est fini, la loi est souvent représentée à l'aide d'une table. Pour Ex. 3.5, la table a la forme

x_i	0	1
p_i	$1/2$	$1/2$

parce que, en utilisant Ex. 3.5, on peut écrire

$$p_0 = P_X(x_0) = P(X^{-1}(\{0\})) = \frac{1}{2},$$
$$p_1 = P_X(x_1) = P(X^{-1}(\{1\})) = \frac{1}{2}.$$

Noter que la somme des entrées de la ligne "p_i" doit être égale à 1.

(b) Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $c \in \mathbb{R}$ une constante. La **variable certaine** *est l'application $X : \Omega \to \mathbb{R}$ définie, pour tout $\omega \in \Omega$, par*

$$\boxed{X(\omega) := c.}$$

Montrer que la variable certaine est une vad.
Solution Nous allons vérfier les propriétés (VAD1) et (VAD2) de Déf. 3.1 :
(VAD1) Comme l'univers image est égal à $X(\Omega) = \{c\}$, (VAD 1) est satisfait.
(VAD2) Comme $\mathcal{A}_X = \{\emptyset, \{c\}\}$ et comme les images réciproques de tous les événements dans $\mathcal{A}_X$ ont la forme $X^{-1}(\emptyset) = \emptyset$ et

$$X^{-1}(c) = \Omega,$$

la condition (VAD 2) est également satisfaite. □

En ce qui concerne la probabilité image, on a $P_X(\emptyset) = 0$ et, avec $x_0 := c$,

$$p_0 = P_X(x_0) = P(X^{-1}(c)) = P(\Omega) = 1.$$

La loi de probabilité de la variable certaine, appelée la **loi de Dirac***, est une loi qui concentre la totalité de la probabilité en un seul point.*

Elle est donnée par la table suivante :

x_i	c
p_i	1

(c) *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et soit $A_0 \in \mathcal{A}$ un événement fixé. La* **variable indicatrice** *est l'application $X : \Omega \to \mathbb{R}$ définie, pour tout $\omega \in \Omega$, par*

$$\boxed{X(\omega) := \begin{cases} 1, & \omega \in A_0, \\ 0, & \textit{sinon.} \end{cases}}$$

Montrer que la variable indicatrice est une vad.
Solution *Nous allons vérfier les propriétés (VAD1) et (VAD2) de Déf. 3.1 :*
(VAD1) Comme l'univers image est égal à $X(\Omega) = \{0, 1\}$, (VAD 1) est satisfait.
(VAD2) Comme $\mathcal{A}_X = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$ et comme les images réciproques des événements dans $\mathcal{A}_X$ ont la forme $X^{-1}(\emptyset) = \emptyset$, $X^{-1}(\{0, 1\}) = \Omega$ et

$$X^{-1}(1) = A_0,$$
$$X^{-1}(0) = A_0^c,$$

la condition (VAD 2) est également satisfaite. □

En ce qui concerne la probabilité image, on a, avec $x_0 := 0$ et $x_1 := 1$, que

$$p_0 = P_X(x_0) = P(X^{-1}(0)) = P(A_0^c) = 1 - P(A_0),$$
$$p_1 = P_X(x_1) = P(X^{-1}(1)) = P(A_0).$$

La loi de probabilité de la variable indicatrice est donc donnée par la table suivante :

x_i	0	1
p_i	$1 - P(A_0)$	$P(A_0)$

Finalement, nous allons introduire la fonction de répartition.

Définition 3.8 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une vad. La fonction*

$$F_X : \mathbb{R} \to \mathbb{R},$$

définie, pour tout $x \in \mathbb{R}$, par

$$\boxed{F_X(x) := P(X < x) \overset{\text{Déf. 3.3}}{=} P_X(\{x_i \in X(\Omega) \,|\, x_i < x\}),}$$

s'appelle la **fonction de répartition** *de la vad X.*
S'il n'y a pas de confusion possible, on écrira souvent $F := F_X$.

En général, la fonction de répartition a les propriétés suivantes.

Enoncé 3.9 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une vad. Alors, la fonction der répartition F_X est une fonction en escalier continue à gauche. En plus, pour tout $i \in \mathbb{N}$, la valeur p_i se calcule à partir de F_X comme*

$$p_i = F_X(x_{i+1}) - F_X(x_i).$$

Solution Comme, pour tout $x \in \mathbb{R}$, nous pouvons écrire

$$F_X(x) = P_X(\{x_i \in X(\Omega) \,|\, x_i < x\}) = P_X\Big(\bigcup\nolimits_{x_i<x}\{x_i\}\Big) \overset{\text{Déf. 2.6 (P3)}}{=} \sum_{x_i<x} P_X(x_i) \overset{\text{Déf. 3.6}}{=} \sum_{x_i<x} p_i,$$

on obtient que F_X est une fonction en escalier qui est continue à gauche (cf. Fig. 3.9). Ici, nous avons utilisé la notation simplifiée $x_i < x$ à la place de $x_i \in X(\Omega) : x_i < x$.
En plus, pour tout $j \in \mathbb{N}$, on trouve

$$F_X(x_{j+1}) - F_X(x_j) = \sum_{x_i<x_{j+1}} p_i - \sum_{x_i<x_j} p_i = p_j.$$

$\square$

Pour un univers image fini $X(\Omega) = \{x_0, x_1, \ldots, x_n\}$, la fonction de répartition $x \mapsto F_X(x)$ est illustrée dans la figure suivante :

Exemple 3.10 *Pour Ex. 3.5 (cf. également Ex. 3.7 (a)), on a $x_0 = 0$ et $x_1 = 1$ et donc, pour tout $x \in \mathbb{R}$, que*

$$F(x) = \sum_{x_i<x} \frac{1}{2} = \begin{cases} 0, & x \leq 0, \\ \frac{1}{2}, & 0 < x \leq 1, \\ 1, & x > 1. \end{cases}$$

La fonction $x \mapsto F(x)$ a donc le graphe suivant :

3.3 Couples de variables aléatoires discrètes

Si l'univers Ω de notre modèle probabiliste contient des éléments ayant plusieurs propriétés intéressantes que nous voudrions décrire par un codage numérique (comme, p.ex., la production et la consommation d'énergie [dans uns sens large] d'un individu d'une population donnée), nous sommes amenés à associer plusieurs variables aléatoires aux résultats des expériences aléatoires de ce type.

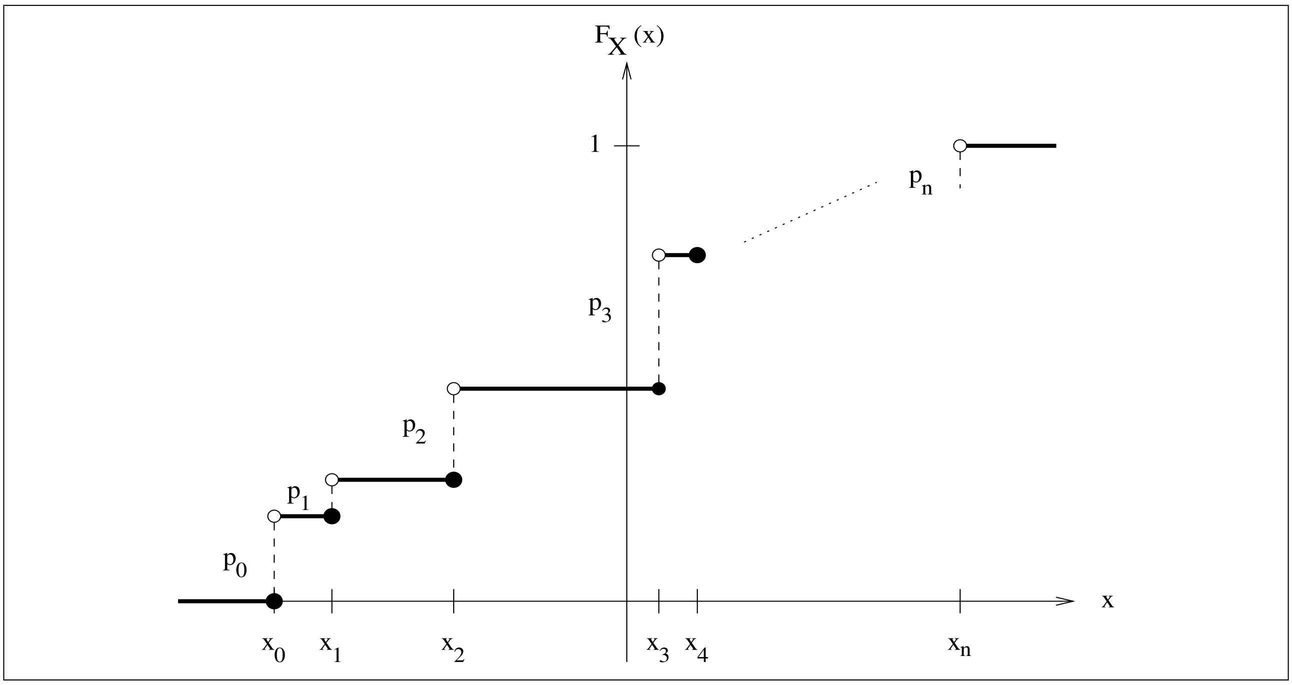

Figure 3.9

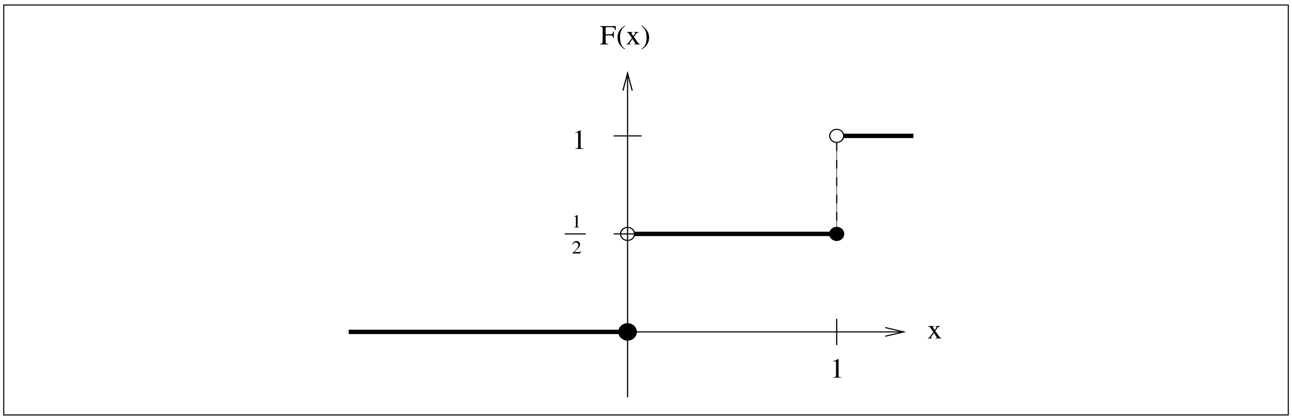

Figure 3.10

Définition 3.11 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et soient* $X : \Omega \to \mathbb{R}$ *et* $Y : \Omega \to \mathbb{R}$ *deux vad. Le* **couple de vad** X **et** Y *est l'application* $(X, Y) : \Omega \to \mathbb{R}^2$*,*

$$\Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2.$$

En outre, pour $\mathcal{A}_{(X,Y)} := \mathcal{P}(X(\Omega) \times Y(\Omega))$*, l'application* $P_{(X,Y)} : \mathcal{A}_{(X,Y)} \to [0, 1]$*, définie par*

$$\boxed{P_{(X,Y)}(\Lambda) := P((X,Y)^{-1}(\Lambda)) \quad \textit{pour tout } \Lambda \in \mathcal{A}_{(X,Y)},}$$

s'appelle la **probabilité image (de P par (X,Y))**. *On utilisera les notations $X(\Omega)=\{x_0,x_1,\ldots\}$, $Y(\Omega)=\{y_0,y_1,\ldots\}$ et, pour tout $i,j\in\mathbb{N}$,*

$$\boxed{\begin{aligned} p_{ij} &:= P_{(X,Y)}(\{x_i\}\times\{y_j\}),\\ p_{i.} &:= P_X(x_i),\\ p_{.j} &:= P_Y(y_j).\end{aligned}}$$

Les ensembles

$$\{((x_i,y_j),p_{ij})\}_{i,j\in\mathbb{N}},\quad \{(x_i,p_{i.})\}_{i\in\mathbb{N}},\quad \{(y_j,p_{.j})\}_{j\in\mathbb{N}}$$

s'appellent respectivement la **loi (de probabilité) du couple** *et les* **lois (de probabilité) marginales**. *En plus, on utilisera souvent les notations simplifiées*

$$P(X=x,Y=y):=P_{(X,Y)}(\{x\}\times\{y\}),$$
$$P((X,Y)\in\Lambda):=P_{(X,Y)}(\Lambda),\ \textit{etc.}$$

Remarque 3.12

(a) *Soient $I\in\mathcal{P}(X(\Omega))$ et $J\in\mathcal{P}(Y(\Omega))$ et soit $\Lambda\in\mathcal{A}_{(X,Y)}$ de la forme*

$$\Lambda=I\times J.$$

Montrer la probabilité image de Λ est donnée par

$$\boxed{P((X,Y)\in\Lambda)=P(X^{-1}(I)\cap Y^{-1}(J)).}$$

<u>*Solution*</u> *Comme $P((X,Y)\in\Lambda)=P_{(X,Y)}(\Lambda)=P((X,Y)^{-1}(I\times J))$ et comme*

$$\begin{aligned}(X,Y)^{-1}(I\times J)&=\{\omega\in\Omega\,|\,(X,Y)(\omega)\in I\times J\}=\{\omega\in\Omega\,|\,(X(\omega),Y(\omega))\in I\times J\}\\ &=\{\omega\in\Omega\,|\,X(\omega)\in I\}\cap\{\omega\in\Omega\,|\,Y(\omega))\in J\}\\ &=X^{-1}(I)\cap Y^{-1}(J),\end{aligned}$$

on obtient la formule désirée. □

Noter que $X^{-1}(I)$ et $Y^{-1}(J)$ sont dans $\mathcal{A}$ parce que X et Y sont des vad. Noter également qu'en général, Λ n'est pas de la forme $I\times J$ mais est seulement une union d'ensembles de cette forme (cf. Figure 3.12 pour l'exemple de $\Lambda\subseteq X(\Omega)\times Y(\Omega)=\mathbb{N}^2$, où $\Lambda=(\{2,3\}\times\{1,2\})\cup(\{4\}\times\{2\})\cup(\{5\}\times\{2,3,4\})\cup(\{6\}\times\{2,3\})$).

(b) *Si on considère un* **vecteur aléatoire** *de longueur $n\in\mathbb{N}^*$, c.-à-d., si on dispose de n vad $X_i:\Omega\to\mathbb{R}$ pour $i\in\{1,\ldots,n\}$ et si on forme*

$$(X_1,\ldots,X_n):\Omega\to\mathbb{R}^n,$$

les notions de Déf. 3.11 se généralisent de manière naturelle en partant du cas du couple (c.-à-d., de $n=2$).

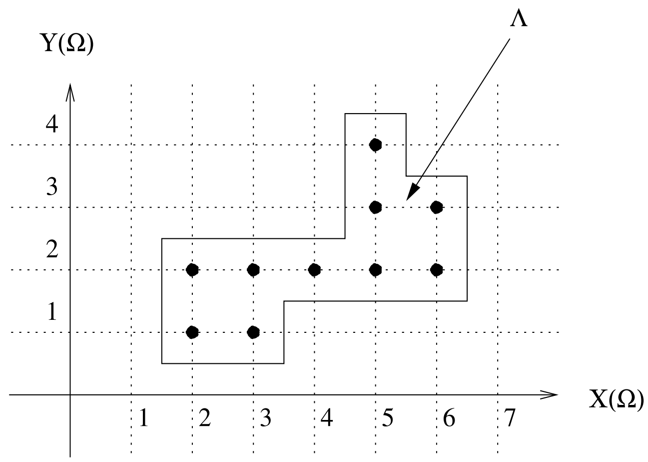

Figure 3.12

Enoncé 3.13 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $(X, Y) : \Omega \to \mathbb{R}^2$ un couple de vad. Alors, on a*

$$P_X(I) = P_{(X,Y)}(I \times Y(\Omega)) \quad \textit{pour tout } I \in \mathcal{A}_X,$$
$$P_Y(J) = P_{(X,Y)}(X(\Omega) \times J) \quad \textit{pour tout } J \in \mathcal{A}_Y.$$

Notamment, on obtient les lois mariginales à partir de la loi du couple,

$$\boxed{\begin{aligned} p_{i.} &= \sum_{j \in \mathbb{N}} p_{ij} \quad \textit{pour tout } i \in \mathbb{N}, \\ p_{.j} &= \sum_{i \in \mathbb{N}} p_{ij} \quad \textit{pour tout } j \in \mathbb{N}. \end{aligned}}$$

Solution Cf. Exr. 15 □

A l'aide de la loi du couple, nous pouvons faire le définition suivante qui est importante dans la pratique.

Définition 3.14 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $(X, Y) : \Omega \to \mathbb{R}^2$ un couple de vad. Les vad X et Y s'appellent* **indépendants** *si*

$$\boxed{p_{ij} = p_{i.}p_{.j} \quad \textit{pour tout } i, j \in \mathbb{N}.}$$

Remarque 3.15 *Noter que Déf. 3.14 exprime le fait que les événements $X^{-1}(x_i)$ et $Y^{-1}(y_j)$ sont indépentants p.r. à P pour tout $i, j \in \mathbb{N}$ (cf. Déf. 2.19), c.-à-d., qu'on a*

$$\begin{aligned} p_{ij} &= P_{(X,Y)}(\{x_i\} \times \{y_j\}) = P((X,Y)^{-1}(\{x_i\} \times \{y_j\})) \overset{\text{Rem. 3.12 (a)}}{=} P(X^{-1}(x_i) \cap Y^{-1}(y_j)) \\ &\overset{\text{Déf. 2.19}}{=} P(X^{-1}(x_i))P(Y^{-1}(y_j)) = P_X(x_i)P_Y(y_j) = p_{i.}p_{.j}. \end{aligned}$$

3.4 Espérance et variance

A présent, nous arrivons aux **moments** d'une vad. Le **moment du 1^{er} ordre** est le suivant.

Définition 3.16 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une vad. L'***espérance (mathématique)** *de X est définie (si la série converge absolument) par*

$$\boxed{E(X) := \sum_{i \in \mathbb{N}} p_i x_i,}$$

où $X(\Omega) = \{x_i\}_{i\in\mathbb{N}}$ et $\{p_i\}_{i\in\mathbb{N}}$ est la loi de X (cf. Déf. 3.6).
Il s'agit d'une moyenne pondérée des valeurs x_i que peut prendre la vad X par les probabilités correspondantes p_i. Elle peut aussi s'interpréter physiquement comme centre de gravité ($\equiv$ barycentre) de particules ponctuelles de masses p_i et de positions x_i.

Exemple 3.17

(a) Dans Ex. 3.5, on a

$$E(X) = \sum_{i=0}^{1} p_i x_i = p_0 x_0 + p_1 x_1 \overset{\text{Ex. 3.7 (a)}}{=} \frac{1}{2}.$$

(b) Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $X : \Omega \to \mathbb{R}$ une vad t.q., pour un $n \in \mathbb{N}^*$,

$$X(\Omega) = \{x_0, x_1, \ldots, x_{n-1}\},$$

et soit P_X l'équiprobabilité. Alors, on a

$$E(X) = \sum_{i=0}^{n-1} p_i x_i = \sum_{i=0}^{n-1} P_X(x_i) x_i = \sum_{i=0}^{n-1} \frac{\operatorname{card}(\{x_i\})}{\operatorname{card}(X(\Omega))}\, x_i = \frac{1}{n}\sum_{i=0}^{n-1} x_i.$$

On utilisera souvent la notation

$$\bar{X} := \frac{1}{n}\sum_{i=0}^{n-1} x_i,$$

et $\bar{X}$, étant la moyenne arithmétique de toutes les valeurs dans l'univers image, sera appelé la **moyenne empirique***. Alors, dans le cas présent, on a $E(X) = \bar{X}$.*

On peut montrer que l'espérance a la propriété utile suivante.

Enoncé 3.18 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $(X, Y) : \Omega \to \mathbb{R}^2$ un couple de vad et $h : \mathbb{R}^2 \to \mathbb{R}$ une fonction continue. Alors, $h(X, Y)$ est une vad et son espérance (si elle existe) est donnée par ce qu'on appelle le* **théorème de transfert***, c.-à-d., elle a la forme*

$$\boxed{E(h(X,Y)) = \sum_{i,j \in \mathbb{N}} p_{ij}\, h(x_i, y_j),}$$

où $X(\Omega) = \{x_i\}_{i\in\mathbb{N}}$, $Y(\Omega) = \{y_j\}_{j\in\mathbb{N}}$ et $\{p_{ij}\}_{i,j\in\mathbb{N}}$ est la loi du couple (X, Y) (cf. Déf. 3.11).

Notamment, en utilisant En. 3.18, on trouve que l'espérance est linéaire comme suit.

Enoncé 3.19 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $(X, Y) : \Omega \to \mathbb{R}^2$ *un couple de vad. Alors, pour tout* $a, b \in \mathbb{R}$*, on a*

$$\boxed{E(aX + bY) = aE(X) + bE(Y).}$$

Solution Pour $h(x, y) := ax + by$ (ce qui est une fonction continue), les formules de En. 3.18 et de En. 3.13 impliquent que

$$\begin{aligned} E(h(X,Y)) = E(aX + bY) &\overset{\text{En. 3.18}}{=} \sum_{i,j\in\mathbb{N}} p_{ij}(ax_i + by_j) \\ &= a\sum_{i\in\mathbb{N}}\Big(\sum_{j\in\mathbb{N}} p_{ij}\Big)x_i + b\sum_{j\in\mathbb{N}}\Big(\sum_{i\in\mathbb{N}} p_{ij}\Big)y_j \overset{\text{En. 3.13}}{=} a\sum_{i\in\mathbb{N}} p_{i.}x_i + b\sum_{j\in\mathbb{N}} p_{.j}y_j \\ &= aE(X) + bE(Y). \end{aligned}$$

□

En spécialisant En. 3.18, on obtient le **théorème de transfert** pour une seule vad X.

Enoncé 3.20 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé,* $X : \Omega \to \mathbb{R}$ *une vad et* $g : \mathbb{R} \to \mathbb{R}$ *une fonction continue. Alors, on a*

$$\boxed{E(g(X)) = \sum_{i\in\mathbb{N}} p_i\, g(x_i),}$$

où $X(\Omega) = \{x_i\}_{i\in\mathbb{N}}$ *et* $\{p_i\}_{i\in\mathbb{N}}$ *est la loi de* X*.*

Solution Cf. Exr. 17 □

Le **moment centré du** $2^{\text{ième}}$ **ordre** est un indicateur mesurant la dispersion des valeurs x_i que peut prendre la vad X, autour de la moyenne pondérée $E(X)$.

Définition 3.21 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vad. La* **variance** *de* X *est définie par*

$$\boxed{V(X) := E([X - E(X)]^2) = \sum_{i\in\mathbb{N}} p_i\,(x_i - E(X))^2,}$$

où $X(\Omega) = \{x_i\}_{i\in\mathbb{N}}$ *et* $\{p_i\}_{i\in\mathbb{N}}$ *est la loi de* X*. En plus, l'***écart type** *de* X *est définie par*

$$\boxed{\sigma(X) := \sqrt{V(X)}.}$$

Exemple 3.22

(a) *Pour Ex. 3.5, nous pouvons calculer que*

$$V(X) = p_0(x_0 - E(X))^2 + p_1(x_1 - E(X))^2 \overset{\text{Ex. 3.7 (a), Ex. 3.17 (a)}}{=} \frac{1}{2}\left(0 - \frac{1}{2}\right)^2 + \frac{1}{2}\left(1 - \frac{1}{2}\right)^2 = \frac{1}{4},$$

$$\sigma(X) = \frac{1}{2}.$$

(b) *Pour Ex. 3.17 (b), nous obtenons*

$$V(X) = \sum_{i=0}^{n-1} p_i(x_i - E(X))^2 = \frac{1}{n}\sum_{i=0}^{n-1}(x_i - \bar{X})^2,$$

où $\bar{X}$ est la moyenne empirique, c.-à-d., dans ce cas, la variance de X est égale à ce qu'on appelle la **variance empirique** *des valeurs possibles de X. En plus, l'écart type*

$$\sigma(X) = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(x_i - \bar{X})^2}$$

*s'appelle l'***écart type empirique***.*

(c) *Soient X et Y deux vad dont les lois sont données dans la table suivante :*

x_i	0	1	2	3	4	5	6
$P_X(x_i)$	0	0	1/8	3/4	1/8	0	0
y_i	0	1	2	3	4	5	6
$P_Y(y_i)$	1/8	1/8	1/8	1/4	1/8	1/8	1/8

Alors, on a $E(X) = E(Y) = 3$, mais $V(X) = \frac{1}{4} \ll \frac{7}{2} = V(Y)$, c.-à-d., les valeurs de Y sont beaucoup plus dispersées autour de sa moyenne pondérée que celles de X, cf. Figure 3.22.

On peut montrer que la variance possède les propriétés suivantes.

Enoncé 3.23 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une vad. Alors :*

(a) $V(X) \geq 0$

(b) $V(X) = 0$ *ssi il existe une constante $c \in \mathbb{R}$ t.q. $P(X = c) = 1$ (dans ce cas, on dit que X est* **presque sûrement** *la variable certaine)*

(c) $V(X + c) = V(X)$ *pour tout $c \in \mathbb{R}$*

(d) $V(cX) = c^2V(X)$ *pour tout $c \in \mathbb{R}$*

(e) *La variance peut s'exprimer par la* **formule de König***, c.-à-d., on a*

$$\boxed{V(X) = E(X^2) - E(X)^2.}$$

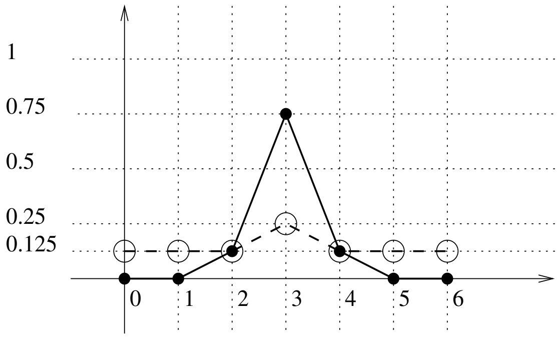

Figure 3.22

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $(X, Y) : \Omega \to \mathbb{R}^2$ un couple de vad. Si X et Y sont indépendants (cf. Déf. 3.14), on a :

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

(g) $V(X + Y) = V(X) + V(Y)$

Si on définit la **covariance** *du couple (X, Y) par*

$$C(X, Y) := E(XY) - E(X)E(Y),$$

on a, si X et Y ne sont pas nécessairement indépendants :

(h) $V(X + Y) = V(X) + V(Y) + 2C(X, Y)$

Rappels

Rappel 3.24 *Soient A et B deux ensembles et soit $f : A \to B$ une application. L'***image (directe)** *par f d'un sous-ensemble $A' \subseteq A$, notée $f(A')$, est définie par*

$$\boxed{f(A') := \{f(a) \,|\, a \in A'\} \subseteq B}.$$

*L'***image réciproque** *par f d'un sous-ensemble $B' \subseteq B$, notée $f^{-1}(B')$ (à ne pas confondre avec l'inverse de f), est définie par*

$$\boxed{f^{-1}(B') := \{a \in A \,|\, f(a) \in B'\} \subseteq A}.$$

4 Variables aléatoires continues

Si le codage se fait par un ensemble de valeurs réels qui n'est pas nécessairement dénombrable (comme, p.ex., par un intervalle dans $\mathbb{R}$), on introduit le concept suivant.

4.1 Définition

Définition 4.1 *Soit* $(\Omega, \mathcal{A})$ *un espace probabilisable. Une application*

$$X : \Omega \to \mathbb{R}$$

s'appelle une **variable aléatoire réelle (var)** *si elle satisfait la condition suivante :*

(VAR) $X^{-1}(I) \in \mathcal{A}$ *pour tout intervalle* $I \subseteq \mathbb{R}$

Remarque 4.2 *La tribu image associée à une var est la tribu dite* **borélienne***, notée* $\mathcal{B}(\mathbb{R})$*, générée par la famille d'intervalles ouverts de la forme* $]-\infty, x[$ *(on n'utilisera pas la tribu triviale dans le cas continu parce qu'elle est beaucoup trop grande). Il suffit alors de vérifier que*

$$X^{-1}(]-\infty, x[) \in \mathcal{A} \quad \text{pour tout } x \in \mathbb{R}.$$

Exemple 4.3 *La durée de vie d'une drosophile tirée au sort dans une population peut être interprétée comme une var.*

4.2 Probabilité image

Les notions de probabilité image (cf. Déf. 3.3) et de fonction de répartition (cf. Déf. 3.8) introduites dans le cas discret se généralisent de la manière suivante dans le cas continu.

Définition 4.4 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une var. La* **probabilité image (de** P **par** X**)** *est définie par*

$$\boxed{P_X(I) := P(X^{-1}(I)) \quad \text{pour tout intervalle } I \subseteq \mathbb{R}.}$$

On appelera P_X *également la* **loi (de probabilité)** *de* X*. La fonction*

$$F_X : \mathbb{R} \to \mathbb{R},$$

définie par

$$\boxed{F_X(x) := P_X(]-\infty, x[) \quad \text{pour tout } x \in \mathbb{R},}$$

s'appelle la **fonction de répartition** *de X. On utilisera souvent les notations simplifiées $F(x) := F_X(x)$ (s'il n'y a pas de confusion possible) et*

$$\begin{aligned} P(X = x) &:= P_X(x) := P_X(\{x\}), \\ P(X < x) &:= P_X(]-\infty, x[), \\ P(a \leq X < b) &:= P_X([a, b[), \text{ etc.} \end{aligned}$$

On peut montrer que la fonction de répartition a les propriétés suivantes.

Enoncé 4.5 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $X : \Omega \to \mathbb{R}$ une var et F sa fonction de répartition. Alors :*

(a) F est croissant au sens large sur $\mathbb{R}$.

(b) $F(x) \in [0, 1]$ pour tout $x \in \mathbb{R}$ et

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1.$$

(c) F est **continu à gauche***, c.-à-d., on a*

$$\lim_{h \to 0^+} F(x - h) = F(x) \quad \text{pour tout } x \in \mathbb{R}.$$

(d) En général, F n'est pas continu à droite, c.-à-d., on a

$$\lim_{h \to 0^+} F(x + h) = F(x) + P(X = x) \quad \text{pour tout } x \in \mathbb{R}.$$

Basé sur les propriétés de la fonction de répartition, on fait les définitions suivantes.

Définition 4.6 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $X : \Omega \to \mathbb{R}$ une var et F sa fonction de répartition. X s'appelle* **continu (vac)** *si F est continu. X s'appelle* **absolument continu (vaac)** *s'il existe une fonction non négative $f_X : \mathbb{R} \to \mathbb{R}$ (qu'on notera souvent $f := f_X$ s'il n'y a pas de confusion possible) t.q.*

$$\boxed{F(x) = \int_{-\infty}^{x} \mathrm{d}t\, f(t) \quad \text{pour tout } x \in \mathbb{R}.}$$

La fonction f_X s'appelle la **densité** *de la vaac X (et de sa fonction de répartition). Si X est (absolument) continu, la loi P_X est également appelée* **(absolument) continue***.*

Remarque 4.7

(a) De manière plus générale, on dit qu'une fonction $F : \mathbb{R} \to \mathbb{R}$ est une fonction de répartition si elle satisfait les propriétés (a) – (c) de En. 4.5.

(b) La propriété (d) de En. 4.5 implique qu'en un point x_0 de discontinuité de F, la graphe de F présente une saut de hauteur $P(X = x_0)$.

(c) *Aux points x où F est dérivable, on récupère la densité par*

$$\frac{\mathrm{d}}{\mathrm{d}x}F(x) = f(x).$$

(d) *La probabilité que X prenne une valeur dans l'intervalle $[a, b[$ est égale à*

$$\boxed{P(a \leq X < b) = F(b) - F(a).}$$

(e) *Une loi pour laquelle $P_X(x) = 0$ pour tout $x \in \mathbb{R}$ s'appelle* **diffuse**. *S'il existe des points $x \in \mathbb{R}$ t.q. $P_X(x) \neq 0$, la loi s'appelle* **mixte**. *On voit de En. 4.5 que la loi d'une vac est diffuse.*

(f) *Dans la pratique, les lois usuelles que nous allons rencontrer proviendront soit d'une vad soit d'une vaac (et ne pas d'une var qui n'est que continue).*

(g) *On note qu'une vaac est une vac.*

Exemple 4.8 *Soit $\theta > 0$. La* **loi exponentielle (de paramètre θ)** *est la loi absolument continue dont la fonction de répartition a la densité $f : \mathbb{R} \to \mathbb{R}$, définie par*

$$f(x) := \begin{cases} \theta \mathrm{e}^{-\theta x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Pour $x \leq 0$, on a $F(x) = 0$, et pour $x > 0$, on obtient alors

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

cf. Fig. 4.8a et 4.8b :

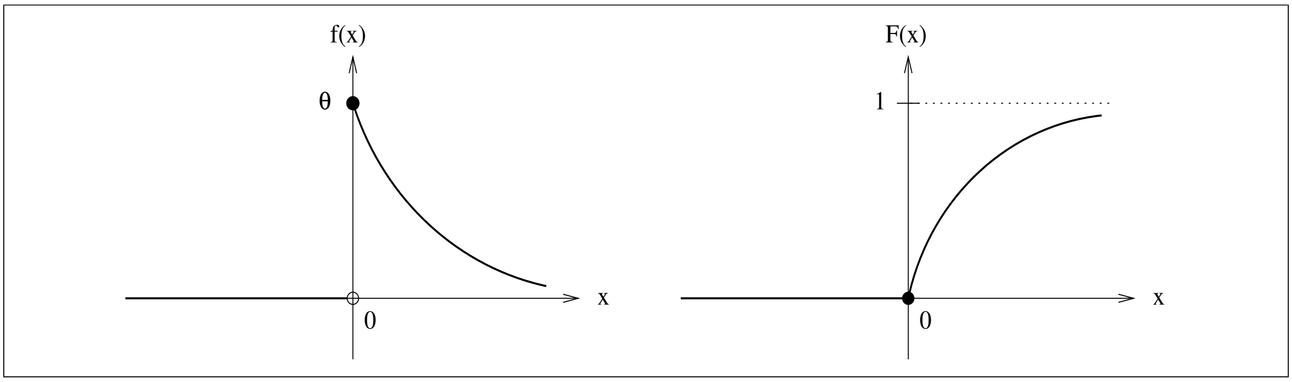

Figure 4.8a Figure 4.8b

Remarque 4.9 *On note que, pour toute fonction de répartition F (cf. Rem. 4.7 (a)), il existe un espace probabilisé $(\Omega, \mathcal{A}, P)$ et une var $X : \Omega \to \mathbb{R}$ t.q. $P(X < x) = F(x)$ pour tout $x \in \mathbb{R}$. Pour s'en convaincre, on choisit simplement $\Omega := \mathbb{R}$, $\mathcal{A} := \mathcal{B}(\mathbb{R})$, $P(]-\infty, x[) := F(x)$ pour tout $x \in \mathbb{R}$ et $X(\omega) := \omega$ pour tout $\omega \in \Omega$. Il en résulte que*

$$P(X < x) = P_X(]-\infty, x[) = P(X^{-1}(]-\infty, x[)) = P(]-\infty, x[) = F(x).$$

Nous résumons les propriétés d'une densité comme suit.

Enoncé 4.10 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une vaac de densité f. Alors :*

(a) La densité est non négative, c.-à-d., $f(x) \geq 0$ pour tout $x \in \mathbb{R}$.

(b) La densité est intégrable sur $\mathbb{R}$ et

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

(c) La probabilité que X prenne une valeur dans un intervalle donné s'obtient en intégrant la densité sur cet intervalle, c.-à-d.,

$$\boxed{P(X \in [a, b]) = \int_a^b \mathrm{d}x\, f(x).}$$

Solution Cf. Exr. 22 □

4.3 Couples de variables aléatoires continues

Comme dans le cas discret, on s'intéresse également aux modèles probabilistes dotés de plusieurs var (cf. Déf. 3.11).

Définition 4.11 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $(X, Y) : \Omega \to \mathbb{R}^2$ un couple de var. La* **probabilité image** *ou la* **loi (de probabilité)** *du couple (X, Y) est définie comme dans Déf. 3.11 par*

$$P_{(X,Y)}(\Lambda) := P((X, Y)^{-1}(\Lambda))$$

pour tout Λ (dans la tribu borélienne de $\mathbb{R}^2$, cf. Rem. 4.2) et notamment pour $\Lambda = I \times J$, où I et J sont des intervalles dans $\mathbb{R}$. La **fonction de répartition**

$$F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$$

est définie par

$$F_{(X,Y)}(x,y) := P(X < x, Y < y) := P_{(X,Y)}(]-\infty, x[\times]-\infty, y[)$$
$$= P((X,Y)^{-1}(]-\infty, x[\times]-\infty, y[)).$$

Si $F_{(X,Y)}$ est deux fois dérivable dans $\mathbb{R}^2$, le couple (X,Y) (et aussi la loi du couple) est appelé **absolument continu** *ayant la* **densité** $f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$ *définie par*

$$f_{(X,Y)}(x,y) := \frac{\partial^2 F_{(X,Y)}(x,y)}{\partial x \partial y}.$$

Ensuite, les var X et Y s'appellent **indépendantes** *si*

$$F_{(X,Y)}(x,y) = F_X(x)F_Y(y) \quad \textit{pour tout } x, y \in \mathbb{R}.$$

En dérivant cette équation p.r. à x et y, on trouve que les densités satisfont

$$\boxed{f_{(X,Y)}(x,y) = f_X(x)f_Y(y) \quad \textit{pour tout } x, y \in \mathbb{R}.}$$

Comme dans En. 4.5, on peut montrer les propriétés suivantes de la fonction de répartition d'un couple de var.

Enoncé 4.12 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $(X,Y) : \Omega \to \mathbb{R}^2$ un couple de var et $F_{(X,Y)}$ sa fonction de répartition. Alors :*

(a) $F_{(X,Y)}(x,y)$ est croissant au sens large p.r. à chacune des deux variables.

(b) On a $F_{(X,Y)}(x,y) \in [0,1]$ pour tout $x, y \in \mathbb{R}$ et

$$\lim_{x\to-\infty} F_{(X,Y)}(x,y) = 0 \quad \textit{pour tout } y \in \mathbb{R},$$
$$\lim_{y\to-\infty} F_{(X,Y)}(x,y) = 0 \quad \textit{pour tout } x \in \mathbb{R},$$
$$\lim_{x,y\to\infty} F_{(X,Y)}(x,y) = 1.$$

(c) Si la loi est absolument continue, on a

$$\boxed{F_{(X,Y)}(x,y) = \int_{-\infty}^{x} \mathrm{d}t \int_{-\infty}^{y} \mathrm{d}s\, f_{(X,Y)}(t,s) \quad \textit{pour tout } x, y \in \mathbb{R}.}$$

(d) Les **lois marginales** *s'obtiennent par*

$$F_X(x) = \lim_{y\to\infty} F_{(X,Y)}(x,y) \quad \textit{pour tout } x \in \mathbb{R},$$
$$F_Y(y) = \lim_{x\to\infty} F_{(X,Y)}(x,y) \quad \textit{pour tout } y \in \mathbb{R}.$$

Pour les densités, on a

$$f_X(t) = \int_{-\infty}^{\infty} \mathrm{d}s\, f_{(X,Y)}(t,s) \quad \textit{pour tout } t \in \mathbb{R},$$
$$f_Y(s) = \int_{-\infty}^{\infty} \mathrm{d}t\, f_{(X,Y)}(t,s) \quad \textit{pour tout } s \in \mathbb{R}.$$

4.4 Espérance et variance

A présent, nous arrivons aux **moments** des var que nous ne définirons que pour les vaac.

Définition 4.13 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une vaac de densité f. L'***espérance (mathématique)** *de X est définie par*

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

et on suppose que cette intégrale converge absolument. La **variance** *de X est définie par*

$$\boxed{V(X) := E([X - E(X)]^2) = \int_{-\infty}^{\infty} \mathrm{d}x\; (x - E(X))^2 f(x)}$$

*et l'***écart type** *par $\sigma(X) := \sqrt{V(X)}$ (comme dans le cas discret, cf. Déf. 3.21).*

Comme dans le cas discret (cf. En. 3.18), on peut montrer que l'espérance a la propriété suivante.

Enoncé 4.14 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $(X, Y) : \Omega \to \mathbb{R}^2$ un couple de vaac ayant la densité $f_{(X,Y)}$ et $h : \mathbb{R}^2 \to \mathbb{R}$ une fonction continue. Alors, l'espérance (si elle existe) de $h(X, Y)$ est donnée par le* **théorème de transfert***,*

$$E(h(X,Y)) = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y\, h(x,y) f_{(X,Y)}(x,y).$$

Notamment, si X est une vaac de densité f_X et $g : \mathbb{R} \to \mathbb{R}$ une fonction continue, on a

$$\boxed{E(g(X)) = \int_{-\infty}^{\infty} \mathrm{d}x\, g(x) f_X(x).}$$

Finalement, on peut montrer l'énoncé suivant.

Enoncé 4.15 *L'espérance et la variance ont les mêmes propriétés que dans le cas discret (notamment, En. 3.19 et En. 3.23 sont également vrais dans le cas continu).*

5 Lois classiques

Nous allons étudier quelques lois typiques qui peuvent être retenues pour une modélisation probabiliste d'une expérience aléatoire. Nous commençons par le cas discret.

5.1 Cas discret

(a) Loi de Dirac

Définition 5.1 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vad avec*

$$X(\Omega) = \{1\}.$$

On dit alors que **X suit une loi de Dirac***, c.-à-d.,*

$$\boxed{P(X = 1) = 1.}$$

L'application $x \mapsto P(X = x)$ *a le graphe suivant :*

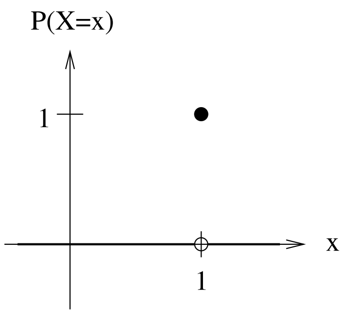

Figure 5.1

Exemple 5.2 *La variable certaine de Ex. 3.7 (b) suit une loi de Dirac. Cette vad est définie par* $X(\omega) = c$ *pour tout* $\omega \in \Omega$ *(et* $c = 1$*). On a donc* $X(\Omega) = \{c\}$*,*

$$P(X = c) = P_X(c) = P_X(X(\Omega)) = P(\Omega) = 1,$$

et, en plus, $P(X = x) = P(X^{-1}(\{x\})) = P(\emptyset) = 0$ *pour tout* $x \neq c$.
La fonction de répartition $F(x) = P(X < x)$ *est déterminée dans Exr. 14 (a). On y trouve qu'elle est donnée par*

$$F(x) = \begin{cases} 0, & x \leq 1, \\ 1, & x > 1, \end{cases}$$

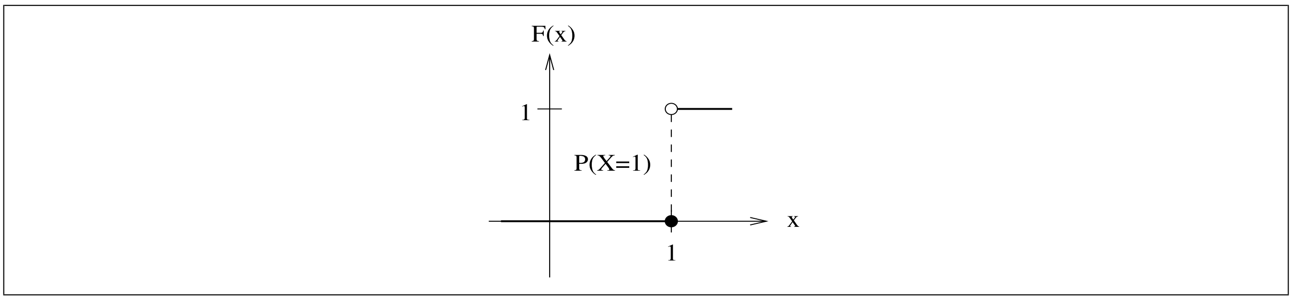

Figure 5.2

ayant donc le graphe donnée dans Fig. 5.2. En plus, pour l'espérance et la variance, on obtient (cf. Exr. 16 (a))

$$\boxed{\begin{aligned} E(X) &= 1, \\ V(X) &= 0. \end{aligned}}$$

(b) Loi uniforme (discrète)

Définition 5.3 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vad avec*

$$X(\Omega) = \{1, 2, \dots, n\}$$

pour un $n \in \mathbb{N}^*$*. On dit que* **X suit une loi uniforme** *si*

$$\boxed{P(X = x) = \frac{1}{n} \quad \text{pour tout } x \in X(\Omega).}$$

L'application $x \mapsto P(X = x)$ *a le graphe suivant :*

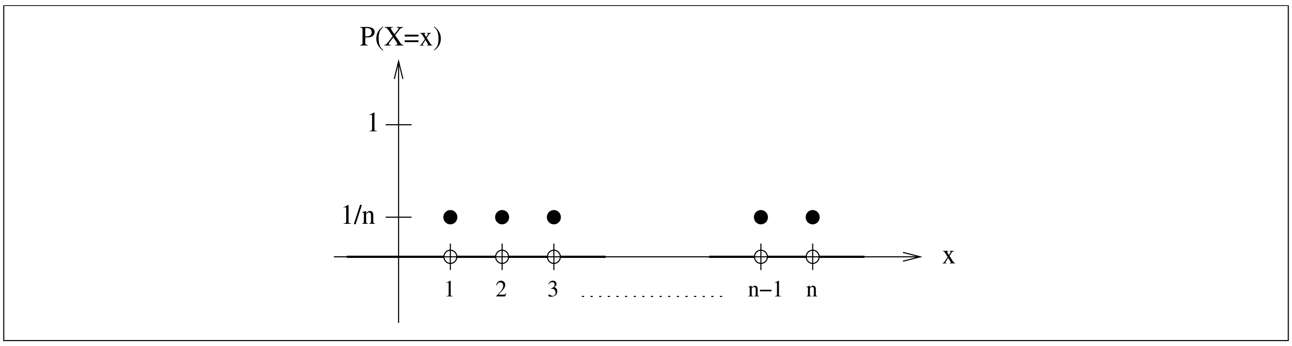

Figure 5.3a

La fonction de répartition a donc la forme donnée dans Fig. 5.3b (cf. En. 3.9).

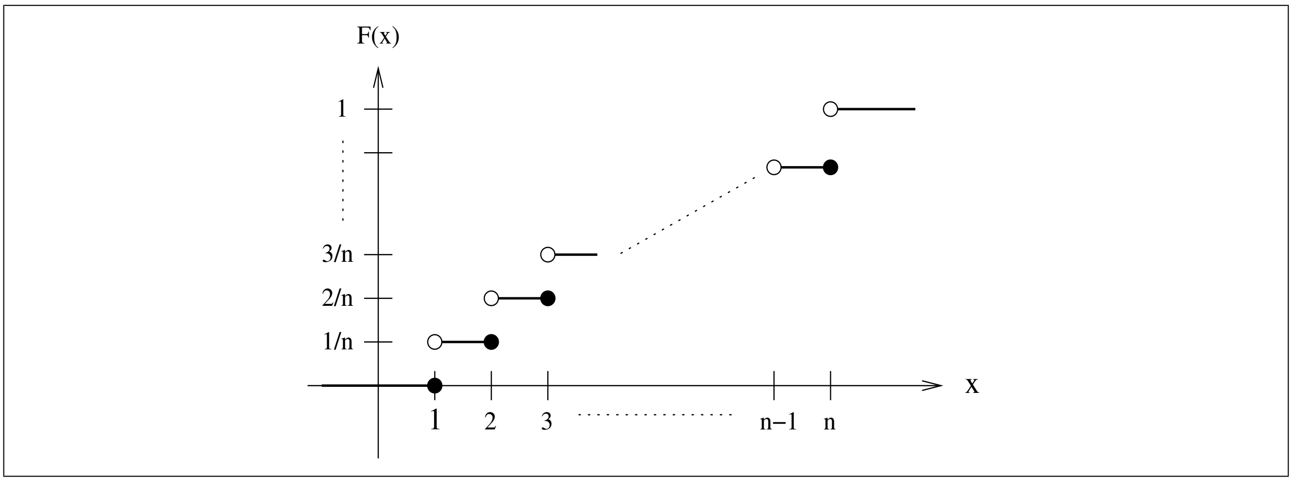

Figure 5.3b

En plus, pour l'espérance et la variance, on a

$$\boxed{\begin{aligned} E(X) &= \frac{n+1}{2}, \\ V(X) &= \frac{n^2-1}{12}. \end{aligned}}$$

Exemple 5.4 *Le lancer d'un dé équilibré à n faces.*

(c) Loi de Bernoulli

Définition 5.5 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une vad avec*

$$X(\Omega) = \{0, 1\}.$$

On dit que **X suit une loi de Bernoulli (de paramètre** $p \in]0,1[$**)**, *noté* $\boxed{X \sim \mathcal{B}(1,p)}$, *si*

$$\boxed{P(X = x) = \begin{cases} 1-p, & x = 0, \\ p, & x = 1, \end{cases} \quad \textit{pour tout } x \in X(\Omega).}$$

L'application $x \mapsto P(X = x)$ a le graphe suivant :

Exemple 5.6 *La variable indicatrice de Ex. 3.7 (c) suit une loi de Bernoulli de paramètre $p := P(A_0)$ (où $A_0 \in \mathcal{A}$ est fixé). Cette vad est définie par $X(\omega) = 1$ si $\omega \in A_0$ et $X(\omega) = 0$ si $\omega \notin A_0$. On a donc $X(\Omega) = \{0, 1\}$ et*

$$\begin{aligned} P(X = 0) &= P_X(0) = P(X^{-1}(0)) = P(A_0^c) = 1 - P(A_0) = 1 - p, \\ P(X = 1) &= P_X(1) = P(X^{-1}(1)) = P(A_0) = p, \end{aligned}$$

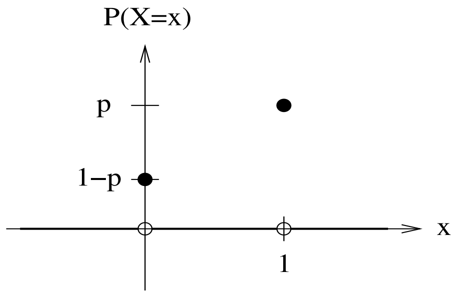

Figure 5.5

et $P(X = x) = 0$ *pour tout* $x \notin \{0, 1\}$*. La fonction de répartition* $F(x) = P(X < x)$ *est déterminée dans Exr. 14 (b). On y trouve qu'elle a la forme*

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - p, & 0 < x \leq 1, \\ 1, & x > 1, \end{cases}$$

ayant donc le graphe suivant :

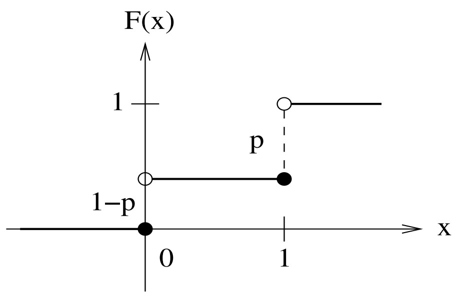

Figure 5.6

En plus, pour l'espérance et la variance, on a (cf. Exr. 16 (b))

$$\boxed{\begin{array}{l} E(X) = p, \\ V(X) = p(1 - p). \end{array}}$$

Exemple 5.7

(a) Le jet d'un dé équilibré où on code 1 *pour les résultats impairs et* 0 *pour les résultats pairs (cf. Ex. 3.5).*

(b) Le lancer d'une pièce biaisée où on code 1 *pour pile et* 0 *pour face.*

(d) Loi binomiale

Définition 5.8 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vad avec*

$$X(\Omega) = \{0, 1, \ldots, n\}.$$

On dit que **X suit une loi binomiale (de paramètres** $n \in \mathbb{N}$ **et** $p \in]0, 1[$**)**, *noté* $\boxed{X \sim \mathcal{B}(n, p)}$, *si*

$$\boxed{P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \textit{pour tout } x \in X(\Omega).}$$

L'application $x \mapsto P(X = x)$ *a le graphe donné dans Fig. 5.8.*

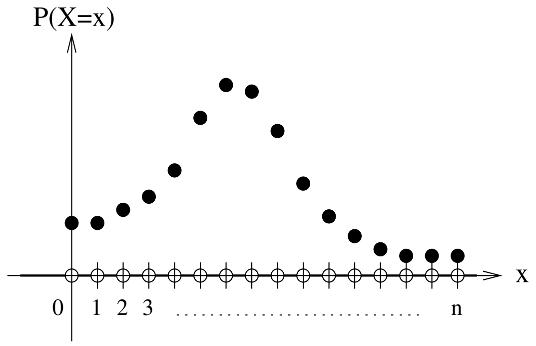

Figure 5.8

On peut montrer les propriétés suivantes.

Enoncé 5.9 *Si* $X \sim \mathcal{B}(n, p)$, *alors*

$$\boxed{\begin{aligned} E(X) &= np, \\ V(X) &= np(1-p). \end{aligned}}$$

En plus, si $X \sim \mathcal{B}(n, p)$ *et* $Y \sim \mathcal{B}(m, p)$ *et si* X *et* Y *sont indépendants, alors* $X + Y \sim \mathcal{B}(n + m, p)$.

Exemple 5.10

(a) *Une urne est constituée de boules présentant une caractéristique dite* A *dans la proportion* p. *On effectue* n *tirages avec remise et on compte le nombre de boules de type* A *obtenues.*

Pour décrire cette expérience aléatoire, on peut retenir l'espace probabilisé $(\Omega, \mathcal{A}, P)$ *specifié par les ingrédients*

$$\begin{aligned} \Omega &:= \{\mathsf{A}, \overline{\mathsf{A}}\}^n, \\ \mathcal{A} &:= \mathcal{P}(\Omega), \\ P(\{\omega\}) &:= p^{N_\omega}(1-p)^{n-N_\omega} \quad \textit{pour tout } \omega \in \Omega, \end{aligned}$$

où N_ω *est le nombre de* A *dans* ω*. La vad* $X : \Omega \to \mathbb{R}$ *qui compte le nombre de boules de type* A *obtenues est donc définie par*

$$X(\omega) := N_\omega \quad \textit{pour tout } \omega \in \Omega.$$

Pour $x \in X(\Omega) = \{0, 1, \ldots, n\}$ *(noter que le nombre de boules de type* A *obtenues est entre* 0 *et* n*), on peut calculer*

$$\begin{aligned} P(X = x) &= P_X(x) = P(X^{-1}(x)) = P\big(\textstyle\bigcup_{\omega : N_\omega = x}\{\omega\}\big) = \displaystyle\sum_{\omega :\, N_\omega = x} P(\{\omega\}) \\ &= \sum_{\omega :\, N_\omega = x} p^{N_\omega}(1-p)^{n-N_\omega} = \sum_{\omega :\, N_\omega = x} p^x(1-p)^{n-x} = \big(\textstyle\sum_{\omega :\, N_\omega = x} 1\big)\, p^x(1-p)^{n-x} \\ &= \binom{n}{x} p^x (1-p)^{n-x}, \end{aligned}$$

et $P(X = x) = 0$ *pour tout* $x \notin \{0, \ldots, n\}$*. D'après Déf. 1.10 et En. 1.11, le coéfficient binomial* $\binom{n}{x}$ *décrit le nombre d'expériences où* x *boules ont la caratéristique* A *parmi les* n *boules tirées (le nombre de combinaisons).*

(b) Le nombre de résultats "pile" apparus au cours de n *jets d'une pièce de monnaie suit une loi* $\mathcal{B}(n, \frac{1}{2})$*. Pour* $x \in \{0, \ldots, n\}$*, on a*

$$\begin{aligned} P(X = x) &= \binom{n}{x} \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{n-x} = \binom{n}{x} \frac{1}{2^n}, \\ E(X) &= \frac{n}{2}, \\ V(X) &= \frac{n}{4}. \end{aligned}$$

Dans le cas des lois **(a) – (d)**, l'univers image comportait un nombre fini d'éléments. Pour les deux lois à venir, l'univers image sera infini (mais dénombrable).

(e) Loi de Poisson

Définition 5.11 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vad avec*

$$X(\Omega) = \mathbb{N}.$$

On dit que **X suit une loi de Poisson (de paramètre $\lambda > 0$)**, *noté* $\boxed{X \sim \mathcal{P}(\lambda)}$, *si*

$$\boxed{P(X = x) = \mathrm{e}^{-\lambda}\,\frac{\lambda^x}{x!} \quad \textit{pour tout } x \in X(\Omega).}$$

L'application $x \mapsto P(X = x)$ a le graphe donné dans Fig. 5.11. On peut montrer les propriétés suivantes.

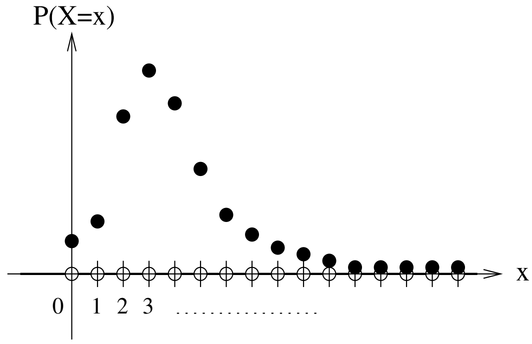

Figure 5.11

Enoncé 5.12 *Si $X \sim \mathcal{P}(\lambda)$, alors*

$$\boxed{E(X) = V(X) = \lambda.}$$

En plus, si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ et si X et Y sont indépendants, alors $X + Y \sim \mathcal{P}(\lambda + \mu)$.

Exemple 5.13

(a) La loi de Poisson est utilisée pour modéliser le nombre d'apparitions d'un événement rare.

(b) Modéliser la distribution d'une espéce végétale sans intervention humaine (contrairement à une plantation de peupliers où on aurait une loi uniforme).

Remarque 5.14 *La loi binomiale peut être approximée par la loi de Poisson dans le sens suivant. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de vad (sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ fixé) t.q. on a $X_n \sim \mathcal{B}(n, p_n)$ pour tout $n \in \mathbb{N}$ et $\lim_{n\to\infty} np_n = \lambda$. Alors, on peut montrer que*

$$\lim_{n\to\infty} P(X_n = x) = \mathrm{e}^{-\lambda}\,\frac{\lambda^x}{x!} \quad \textit{pour tout } x \in X(\Omega).$$

Si n est grand et p assez petit, c.-à-d., si,

$$n \geq 30, \quad p \leq \frac{1}{10}, \quad np \leq 10,$$

on peut remplacer la loi binoimiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$ (à noter que l'espérance reste la même). Cette approximation peut rendre les calculs plus simples.

(f) Loi géométrique

Définition 5.15 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une vad avec*

$$X(\Omega) = \mathbb{N}^*.$$

On dit que **X suit une loi géométrique (de paramètre** $p \in]0, 1[$**)***, si*

$$\boxed{P(X = x) = p(1-p)^{x-1} \quad \text{pour tout } x \in X(\Omega).}$$

L'application $x \mapsto P(X = x)$ a le graphe donné dans Fig. 5.15.

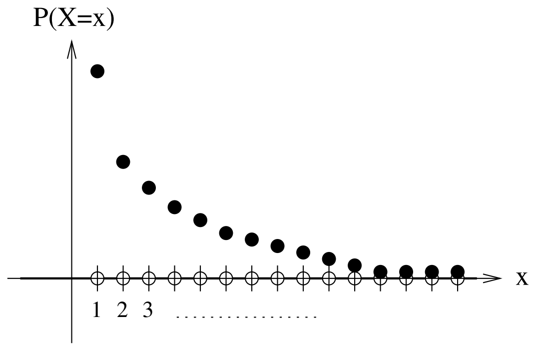

Figure 5.15

On peut montrer les propriétés suivantes.

Enoncé 5.16 *Si X suit une loi géométrique de paramètre p, on a*

$$\boxed{\begin{aligned} E(X) &= \frac{1}{p}, \\ V(X) &= \frac{1-p}{p^2}. \end{aligned}}$$

Exemple 5.17 *On effectue des expériences successives indépendantes jusqu'à la réalisation d'un événement particulier* A *et on note X le nombre (aléatoire) d'expériences effectuées (le temps d'attente de* A*). P.ex., on lance une pièce biaisée jusqu'à obtenir "pile".*

5.2 Cas continu

Par la suite, nous discuterons quelques lois importants dans le cas absolument continu.

(a) Loi uniforme (continue)

Définition 5.18 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vaac. On dit que* **X suit une loi uniforme (de paramètres** $a, b \in \mathbb{R}$ **avec** $a < b$**)**, *noté* $\boxed{X \sim \mathcal{U}(a,b)}$, *si sa densité a la forme*

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

L'application $x \mapsto f(x)$ *a le graphe donné dans Fig. 5.18. Pour les propriétés d'une vaac* $X \sim \mathcal{U}(a,b)$, *cf. Exr. 25.*

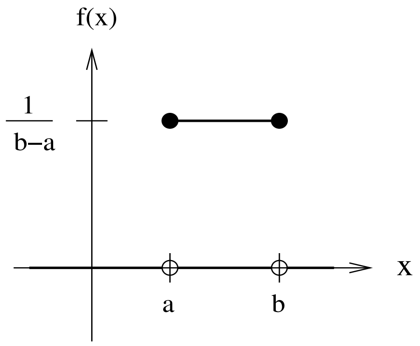

Figure 5.18

Exemple 5.19 *Un modèle élémentaire pour un générateur de nombres pseudo-aléatoires est basé sur la loi* $\mathcal{U}(0,1)$.

(b) Loi exponentielle

Définition 5.20 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}$ *une vaac. On dit que* **X suit une loi exponentielle (de paramètre** $\theta > 0$**)**, *noté* $\boxed{X \sim \mathcal{E}(\theta)}$, *si*

$$\boxed{f(x) = \begin{cases} \theta \mathrm{e}^{-\theta x}, & x \geq 0, \\ 0, & \textit{sinon.} \end{cases}}$$

L'application $x \mapsto f(x)$ *a le graphe suivant :*

Pour les propriétés d'une vaac $X \sim \mathcal{E}(\theta)$, cf. Ex. 4.8 et Exr. 21.

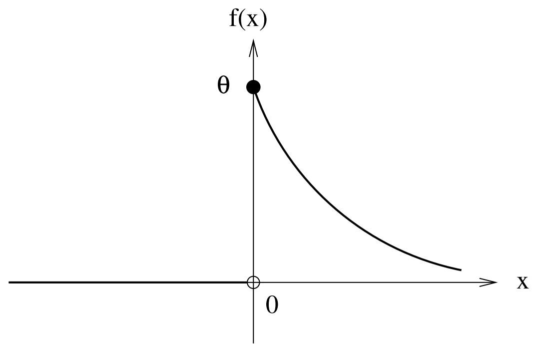

Figure 5.20

Exemple 5.21 *La loi exponentielle peut s'interpréter comme un modèle pour une durée de vie X ayant la propriété de* **non-vieillissement** *(on dira également que sa loi est* **sans mémoire***), c.-à-d., la loi exponentielle satisfait*

$$\boxed{P(X > T + t \,|\, X > T) = P(X > t) \quad \textit{pour tout}\, T, t \geq 0,}$$

où nous avons utilisé la notation $P(X > T + t \,|\, X > T) := P(X^{-1}(]T + t, \infty[) \,|\, X^{-1}(]T, \infty[))$*.*

Solution D'après Déf. 2.11 de la probabilité conditionnelle, on a

$$P(X > T + t \,|\, X > T) = \frac{P(X^{-1}(]T + t, \infty[) \cap X^{-1}(]T, \infty[))}{P(X^{-1}(]T, \infty[))} = \frac{P(X > T + t)}{P(X > T)}.$$

Pour tout $r \geq 0$, on a $\mathbb{R} =]-\infty, r] \cup \{r\} \cup]r, \infty[$ et donc, d'après Déf. 2.6, on peut écrire

$$1 = \underbrace{P_X(]-\infty, r[)}_{=F(r)} + \underbrace{P_X(\{r\})}_{=0} + \underbrace{P_X(]r, \infty[)}_{=P(X>r)}.$$

Finalement, en utilisant Ex. 4.8, on obtient $P(X > r) = 1 - (1 - \mathrm{e}^{-\theta r}) = \mathrm{e}^{-\theta r}$ et donc,

$$P(X > T + t \,|\, X > T) = \frac{\mathrm{e}^{-\theta(T+t)}}{\mathrm{e}^{-\theta T}} = \mathrm{e}^{-\theta t} = P(X > t).$$

□

(c) Loi normale

Définition 5.22 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une vaac. On dit que* **X suit une loi normale (de paramètres** $m \in \mathbb{R}$ **et** $\sigma > 0$**)***, noté* $\boxed{X \sim \mathcal{N}(m, \sigma)}$*, si*

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}.}$$

L'application gaussienne $x \mapsto f(x)$ a le graphe suivant :

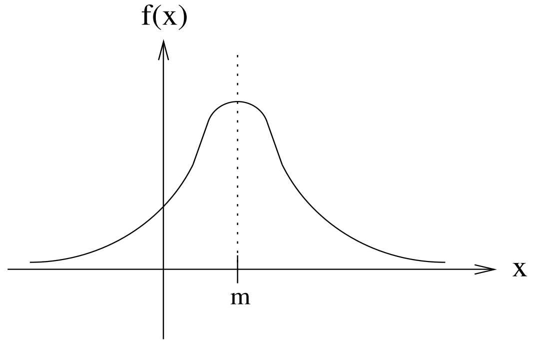

Figure 5.22

On peut montrer les propriétés suivantes.

Enoncé 5.23 *Si $X \sim \mathcal{N}(m, \sigma)$, on a*

$$\boxed{\begin{aligned} E(X) &= m, \\ V(X) &= \sigma^2. \end{aligned}}$$

En plus, si $X \sim \mathcal{N}(m, \sigma)$ et $Y \sim \mathcal{N}(n, \tau)$ et si X et Y sont indépendants, alors $X + Y \sim \mathcal{N}(m + n, \sqrt{\sigma^2 + \tau^2})$.

Exemple 5.24 *La loi normale est l'une des principales distributions de probabilité avec une multitude d'applications comme, p.ex., dans la description de*

- *erreurs de mesure,*
- *déviation de la cote nominale dans la production,*
- *mouvement brownien dans des domaines divers (biologie, physique, finance,. . .),*
- *modèles d'assurance, . . .*

Remarque 5.25

(a) *En faisant le changement de variable*

$$U := \frac{1}{\sigma}(X - m),$$

c.-à-d., en centrant et en réduisant, on obtient de $X \sim \mathcal{N}(m, \sigma)$ que

$$U \sim \mathcal{N}(0, 1).$$

La loi $\mathcal{N}(0, 1)$ s'appelle la **loi normale standard** *ou la* **loi normale centrée réduite**.

(b) *La loi binomiale (discrète) peut être approximée par la loi normale (continue) dans le sens où, si $X \sim \mathcal{B}(n, p)$, n est assez grand et p pas trop proche de 0 et 1, alors X suit à peu près la loi normale $\mathcal{N}(np, \sqrt{np(1-p)})$ (de même espérance et écart type). Dans la pratique, on utilise souvent cette approximation lorsque*

$$n \geq 30, \quad np \geq 5, \quad n(1-p) \geq 5.$$

(c) La fonction de répartition d'une vaac qui suit une loi normale standard $\mathcal{N}(0,1)$ *est habituellement notée*

$$\boxed{\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{d}t\ \mathrm{e}^{-\frac{t^2}{2}} \quad \textit{pour tout } x \in \mathbb{R}.}$$

Comme $\Phi(x) = P(U < x)$ *ne peut s'exprimer par des fonctions élémentaires, ses valeurs sont fournies dans des "tables" (support papier ou électronique).*

Noter les propriétés suivantes.

Enoncé 5.26

(a) $\Phi(-x) = 1 - \Phi(x)$ *pour tout* $x \in \mathbb{R}$

(b) $P(|U| \leq x) \overset{\textit{En. 4.10 (c)}}{=} \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$ *pour tout* $x \geq 0$

(c) $F(x) = P(X < x) = P(U < \frac{x-m}{\sigma}) = \Phi(\frac{x-m}{\sigma})$ *pour tout* $x \in \mathbb{R}$

6 Lois asymptotiques

Dans cette section, nous allons brièvement discuter deux théorèmes fondamentaux de la statistique asymptotique, à savoir, la loi des grands nombres et le théorème central limite qui précisent le comportement asymptotique de la moyenne empirique d'un échantillon de variables aléatoires (notées **va** ; nous ne ferons plus la distinction [pédagogique] entre le cas discret et le cas continu).

Définition 6.1 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $X_i : \Omega \to \mathbb{R}$ des va pour $i \in \{1, \ldots, n\}$. La famille $X_1, \ldots, X_n$ est appelé un* **échantillon d'une loi** *(de taille n) si les va sont (mutuellement) indépendantes et si elles ont la même loi.*
La **moyenne empirique** *d'une famille $X_1, \ldots, X_n$ de va sur $(\Omega, \mathcal{A}, P)$ (pas nécessairement un échantillon d'une loi) est définie par*

$$\boxed{\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i.}$$

6.1 Loi des grand nombres

Si $X_1, \ldots, X_n$ est un échantillon d'une loi, on a

$$\begin{aligned} E(\bar{X}_n) &\overset{\text{En. 4.15}}{=} \frac{1}{n} \sum_{i=1}^{n} \underbrace{E(X_i)}_{=: m} = m, \\ V(\bar{X}_n) &\overset{\text{En. 4.15}}{=} \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{V(X_i)}_{=: \sigma^2} = \frac{\sigma^2}{n}, \end{aligned}$$

d'où $\lim_{n\to\infty} V(\bar{X}_n) = 0$ ce qui permet de montrer la loi des grands nombres (en utilisant l'inégalité de Tchebychev que nous n'allons pas discuter ici). Pour énoncer cette loi, nous avons besoin de la notion de convergence suivante.

Définition 6.2 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $(X_n)_{n\in\mathbb{N}}$ une suite de va et X une va. On dit que $(X_n)_{n\in\mathbb{N}}$* **converge en probabilité** *vers X, noté $X_n \underset{P}{\longrightarrow} X$, si, pour tout $\varepsilon > 0$,*

$$\boxed{\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1.}$$

On arrive alors à la **loi (faible) des grand nombres**.

Enoncé 6.3 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $(X_n)_{n\in\mathbb{N}}$ une suite de va (mutuellement) indépendantes t.q., pour tout $n \in \mathbb{N}$,*

$$E(X_n) = m, \quad V(X_n) = \sigma^2.$$

Alors, nous avons que

$$\boxed{\bar{X}_n \underset{P}{\longrightarrow} m.}$$

Remarque 6.4 *Cet énoncé est obtenu sous la condition que les va X_n aient les mêmes moments du 1^{er} et du $2^{ième}$ ordre (c.-à-d., $E(X_n)$ et $V(X_n)$) mais pas nécessairement la même loi.*

Exemple 6.5 *On effectue n expériences successives indépendantes où on s'intéresse à chaque fois à la réalisation d'un certain événement* A *(une variable de Bernoulli).*

A présent, nous allons passer au deuxième théorème fondamental de la statistique asymptotique.

6.2 Théorème central limite

La loi des grands nombres énonce la convergence de la moyenne empirique vers une variable certaine ayant la loi de Dirac, loi dégénérée dont toute la probabilité est concentrée en un seul point.

Si on équilibre la suite des va centrées et réduites

$$\frac{\bar{X}_n - m}{\sigma},$$

qui converge vers zéro, par $\sqrt{n}$ qui tend vers l'infini, la suite a une limite non dégénérée, la loi normale. Voilà le contenu du **théorème central limite** suivant.

Enoncé 6.6 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $X_1, X_2, \ldots$ un échantillon d'une loi (de taille infinie) admettant des moments du 1^{er} et du $2^{ième}$ ordre, notés $m := E(X_n)$ et $\sigma^2 := V(X_n)$, alors, on a*

$$\boxed{\sqrt{n}\, \frac{\bar{X}_n - m}{\sigma} \underset{Loi}{\longrightarrow} \mathcal{N}(0,1),}$$

où la suite **converge en loi***, c.-à-d., les fonctions de répartition convergent en tout point (de continuité).*

Remarque 6.7 *Grâce à ce théorême, la loi normale est la loi de probabilité la plus utilisée en statistique. Bien que beaucoup de phénomènes individuels ne soient pas symétriques et ne peuvent donc pas être modélisés par la loi normale, la résultante d'un grand nombre de composantes indépendantes est bel et bien décrite par la loi normale.*

(Da Capo al) Fine.

Voici quelques références qui ont été utilisées pour la rédaction de ce cours, la référence principale étant la première partie de la monographie [5].

Références

[1] Chung K L 1979 *Elementary probability theory with stochastic processes* (Springer)

[2] Delmas J-F 2010 *Introduction au calcul des probabilités et à la statistique* (ENSTA)

[3] Fredon D 2008 *Statistique et probabilités en 30 fiches* (Dunod)

[4] Gaultier M 1998 *Statistique* (Vuibert)

[5] Lecoutre J-P 2009 *Statistique et probabilités* (Dunod)