

Laurent BERNIS

# COURS DE MATHÉMATIQUES MP\*

## Chapitre I

# LES PROBABILITÉS, RÉVISIONS DE MPSI

Le présent chapitre couvre le programme de première année en probabilité. Le cours complet, est émaillé d'exercices qui en assurent l'assimilation.

## INTRODUCTION

Les probabilités se distinguent du reste des mathématiques en ce qu'elles consistent non seulement en une théorie mathématique, mais encore en une activité de modélisation, activité que les autres domaines des mathématiques une fois fondés et développés ont rejetée en dehors de leurs préoccupations, se coupant ainsi des activités scientifiques qui les ont parfois fait naître, et qu'ils abandonnent volontiers aux autres sciences. Ainsi l'étude des équations différentielles mène-t-elle une existence autonome, indépendamment des problèmes physiques qui l'alimentent. Cette situation s'explique aisément, puisque c'est précisément l'impuissance de la science à apporter une réponse prédictive à un phénomène qui fait que l'on recourt aux probabilités. Les lois de la mécanique ne sauraient prédire le résultat d'un lancé de dé, puisqu'il dépend de façon sûrement non continue de conditions initiales nombreuses et inconnues. Bref, les probabilités apparaissent là où la science renonce à exercer son pouvoir, et le travail de modélisation, peu conséquent et très variable d'une situation à l'autre, a atterri naturellement dans les mains des mathématiciens.

De cette double essence, les probabilités héritent d'une terminologie propre qui se réfère au contexte concret qu'elles modélisent qui se substitue au langage mathématique, ainsi là où un mathématicien parlerait de partie d'un ensemble, de complémentaire, le probabiliste use des termes *événement*, *événement contraire*. Cette contamination langagière constitue la principale difficulté dans l'abord de cette discipline.

L'objet des probabilités est d'attribuer à un événement qui n'est jugé probable d'aucune science un nombre compris entre 0 et 1 et qui est sensé traduire la fréquence d'apparition de l'événement si l'on répétait un grand nombre de fois les circonstances qui l'ont produit. Dans les sciences on part d'une circonstance connue avec précision (conditions initiales) et on en déduit avec certitude la situation qui en résultera. La position et la vitesse d'un satellite en un instant donné fournissent sa position future. Les probabilités partent d'un événement précisément connu et lui accorde une plausibilité plus ou moins grande, sa probabilité ; ce sont cette fois les conditions initiales qui échappent à notre connaissances, ou qui sont tellement complexes qu'elles en deviennent inexploitables : lorsque on jette un dé on sait seulement que le dé roulera s'arrêtera sur une face. La probabilité d'un événement se veut proportionnelle à la fréquence de réalisation du dit événement si l'on reproduisait un grand nombre de fois les circonstances qui l'ont créé, circonstances dont les détails échappent à notre savoir et conduisent à chaque expérience à une issue *a priori* différente. Ce faisant on suppose implicitement qu'il y a une certaine régularité dans l'apparition de l'événement.

Les probabilités sont nées au XVII<sup>e</sup> siècle, à la suite des travaux de Blaise Pascal, Pierre de Fermat et Christian Huygens motivés par les jeux de hasard. Le terme *probabilité* prend peu à peu son sens actuel avec les développements du traitement mathématique du sujet par Jakob Bernoulli. Il faut attendre le début du XX<sup>e</sup> siècle et les travaux de Kolmogorov, pour voir se développer une théorie axiomatisée des probabilités, théorie qui se fonde dans la théorie de l'intégration de Lebesgue.

Pour ce qui nous concerne, obéissant au programme nous traiterons des probabilités finies c'est-à-dire pour lesquels une expérience conduit à un nombre fini de résultats possibles. Dans ce cadre la part de la modélisation est importante, celle du traitement mathématique reste fort élémentaire.

On signalera enfin les deux jolis ouvrages [?] et [?] dont le début est abordable.

# PROBABILITÉ, ESPACE PROBABILISÉ

Nous allons commencer par décrire clairement le cadre des probabilités et en fixer la terminologie. Nous ne séparerons pas la présentation des définitions premières du rôle qu'elles jouent dans la modélisation d'une situation concrète.

## 2.1 Expérience aléatoire et univers

Notre but, comme nous l'avons dit dans l'introduction, est d'étudier des phénomènes concrets : jet de dé, tirage du loto, déplacement d'une microbille en suspension dans un fluide sous l'action des collisions avec les molécules du fluide, entre deux instants, prix à venir d'une action etc., phénomènes que les sciences sont impuissantes à prévoir et qui cependant présentent une certaine régularité dans la répétition. On parle pour un tel phénomène d'*expérience aléatoire* ou *épreuve aléatoire*. Pour modéliser l'ensemble des résultats d'une expérience aléatoire on recourt à un ensemble, appelé *univers* que la coutume désigne par  $\Omega$ . Cet ensemble est de natures diverses, donnons des exemples :

**Exemple 2.1.1.** —

1. Dans le cas d'un lancé d'un dé (à six face), l'univers est tout naturellement  $\{1, 2, 3, 6\}$  le résultat d'un lancé est représenté par le nombre marqué sur la face tournée vers le ciel.
2. Dans le cas d'un lancé de deux dés, l'univers est  $\{1, 2, 3, 6\}^2$  le résultat d'un lancé est représenté par le couple formé par le nombre indiqué par le premier dé, puis celui indiqué par le second.
3. Dans le cas de la distribution dans un jeu de 52 cartes d'une main de 5 cartes,  $\Omega$  est l'ensemble des parties à 5 éléments de l'ensemble  $\{1, \dots, 52\}$ , on a numéroté les cartes de 1 à 52.
4. On dispose de deux urnes contenant chacune  $N$  boules blanches ou noires, (le nombre de boules d'un couleur pouvant varier d'une urne à l'autre), on choisit une urne puis on tire une boule de l'urne choisie, ici  $\Omega$  est  $\{0, 1\} \times \{1, \dots, N\}$ , la première composante d'un élément de  $\Omega$  désigne le numéro de l'urne, la seconde celui attribué aux  $N$  boules de chaque urne.
5. On regarde la suite des lancés successifs d'une pièce,  $\Omega = \{P, F\}^{\mathbb{N}^*}$  le  $n^{\text{e}}$  terme d'une suite élément de  $\Omega$  est  $P$  si au  $n^{\text{e}}$  lancé on obtient pile,  $F$  sinon.
6. Dans le cas du déplacement entre deux instants d'une microbille livrée aux chocs des molécules d'un fluide,  $\Omega = \mathbb{R}^3$ .

**Tout univers est par convention non vide et dans la suite de ce chapitre on ne s'intéressera qu'à des univers  $\Omega$  finis.**

L'univers étant fixé, on peut étudier différents événements qui peuvent advenir suite à l'expérience aléatoire. la réalisation d'un événement peut se modéliser par une partie de  $\Omega$ . Par exemple dans le cadre du premier exemple 2.1.1., l'événement « le résultat du lancé est pair » est modélisé par la partie  $\{0, 2, 4, 6\}$ . Dans l'exemple 5, l'événement « les 3 premier lancés donnent le même résultat » se modélise par la partie  $\{(P, P, P, x_4, x_5, \dots) | (x_4, x_5, \dots) \in \{P, F\}^{\mathbb{N}}\} \cup \{(F, F, F, x_4, x_5, \dots) | (x_4, x_5, \dots) \in \{P, F\}^{\mathbb{N}}\}$ . Si la réalisation d'un premier événement se modélise par une partie  $A$ , de  $\Omega$  la réalisation d'un second par une partie  $B$ , alors la partie  $A \cup B$  modélise la réalisation du premier événement ou du second, la partie  $A \cap B$  modélise la réalisation du premier événement et du second, enfin le complémentaire de  $A$  modélise la réalisation de l'événement contraire au premier.

Pratiquement on a tendance à confondre la situation réelle modélisée et le modèle mathématique, si bien que le langage courant lui-même en vient à contaminer le langage mathématique, conduisant à la terminologie suivante :

**Terminologie 2.1.1.** —

- Les éléments de l'univers  $\Omega$  sont appelés *issues* ou *résultats possibles* ou *réalisations* ;
- Les parties de  $\Omega$  sont appelées *événements* ;
- On appelle *événement élémentaire* toute partie de  $\Omega$  qui est un singleton ;
- Si  $A$  et  $B$  sont des événements on appelle *événement  $A$  ou  $B$*  l'événement  $A \cup B$ , *événement  $A$  et  $B$*  l'événement  $A \cap B$  et l'*événement contraire* de  $A$ , le *complémentaire* (dans  $\Omega$ ) de  $A$ , noté  $\bar{A}$  ;
- Des éléments  $A$  et  $B$  sont dit *incompatibles* si ils sont *disjoints* ( $A \cap B = \emptyset$ ) ;
- L'élément  $\Omega$  est dit *événement certain*, l'événement  $\emptyset$  est dit *événement impossible*.
- Un *système complet d'événements* est une famille de parties de  $\Omega$ ,  $(A_1, \dots, A_n)$  telle que  $\{A_1, \dots, A_n\}$  soit une partition de  $\Omega$ <sup>1</sup>.

La porosité entre le réel et le modèle fait qu'on pourra par exemple écrire à propos de l'exemple 2.1.1.-2 : « soit  $A$  l'événement "le résultat du lancé de dé est pair" ». Dans cette phrase  $A = \{2, 4, 6\}$  événement au sens mathématique est entièrement confondu avec l'événement réel qu'il décrit, à savoir le fait que le lancé de dé donne un

1. Rappelons que  $\{A_1, \dots, A_n\}$  est une partition de  $\Omega$  si par définition aucun des  $A_i$  n'est vide, les  $A_i$  sont deux à deux disjoints, et leur réunion est  $\Omega$ .

nombre pair. Cette négligence coupable est coutumière en probabilité et cette hérésie ne semble guère émouvoir ceux-là mêmes qui distinguent scrupuleusement le modèle mathématique du réel qu'il représente, dès lors qu'il s'agit d'autres sciences.

## 2.2 Probabilité

Nous en arrivons à l'objet essentiel de ce cours : définir la probabilité d'un événement. Soit un événement  $A$  quelconque. Il sera réalisé si l'issue de l'expérience est un élément de  $A$ . Par exemple dans le cadre de l'exemple 1 de 2.1.1 l'événement  $\{2, 4, 6\}$  qui modélise un score pair, est réalisé dès que l'issue de l'expérience est 2, 4, ou 6. On cherche à attribuer à  $A$  un nombre réel compris entre 0 et 1 qui mesure le degré de plausibilité de la réalisation de  $A$ , ce nombre étant d'autant plus proche de 1 que la réalisation de  $A$  est vraisemblable. Pour avoir une idée intuitive de ce que doit être ce nombre et de ce que seront ses propriétés on peut imaginer que l'on répète  $n$  fois la même expérience et que l'on observe la fréquence  $f_n(A)$  de la réalisation de  $A$  (nombre de fois où l'issue de l'expérience est élément de  $A$  divisé par  $n$ ). L'acte de fondation des probabilités est que  $f_n(A)$  tend en un certain sens vers une limite lorsque  $n$  tend vers  $+\infty$ , c'est cette limite que nous appellerons probabilité de l'événement  $A$  et noterons  $\mathbf{P}(\{A\})$ .

Les propriétés des fréquences d'un événement imposent alors immédiatement à l'application  $\mathbf{P}$  que nous cherchons à définir, les propriétés suivantes :

- i.  $0 \leq P(A) \leq 1$ .
- ii.  $\mathbf{P}(\Omega) = 1$ .
- iii.  $\mathbf{P}(\emptyset) = 0$ .
- iv. Si  $B$  est un élément tel que  $A \cap B = \emptyset$ , alors  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ .
- v.  $\mathbf{P}(A) = \sum_{a \in A} \mathbf{P}(\{a\})$ .

Ces propriétés ne sont pas indépendantes les unes des autres. On pourrait pour une définition rigoureuse de la fonction  $\mathbf{P}$ , se donner la probabilité des événements élémentaires et supposer i., ii. et v., Cette démarche serait évidemment la plus naturelle et redonnerait iii. et iv. Mais une telle définition n'est pas généralisable au cas d'un univers  $\Omega$  de cardinal infini. Aussi allons-nous adopter la définition d'une probabilité sur  $\Omega$  suivante :

### Définition 2.2.1. — PROBABILITÉ —

On appelle probabilité sur un univers fini  $\Omega$ , toute application  $\mathbf{P}$  de  $\mathcal{P}(\Omega)$  dans  $[0, 1]$  qui jouit des propriétés suivantes

1.  $\mathbf{P}(\Omega) = 1$ .
2. Pour tout  $A$  et tout  $B$  parties disjointes de  $\Omega$  (événements incompatibles),

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

On appelle espace probabilisé fini tout couple  $(\mathbf{P}, \Omega)$  où  $\Omega$  est un univers fini et  $\mathbf{P}$  une probabilité sur  $\Omega$ .

Examinons les propriétés d'une probabilité

**Proposition 2.2.2.** — Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $A_1, A_2, \dots, A_n$  des événements deux à deux incompatibles (disjoints si l'on préfère), alors

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

**Corollaire 2.2.3.** — Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $A$  un événement (une partie si l'on veut) de  $\Omega$ . Alors

$$\mathbf{P}(A) = \sum_{x \in A} \mathbf{P}(\{x\}).$$

*Preuve de la proposition 2.2.2.* — Notons  $\mathbf{H}_n$  la propriété à prouver

- $\mathbf{H}_2$  est vraie par définition d'une probabilité.
- Soit un entier  $m \geq 2$ . On suppose  $\mathbf{H}_2$  vraie. Soient  $A_1, A_2, \dots, A_{m+1}$  des événements deux à deux incompatibles.

$$\left(\bigcup_{i=1}^n A_i\right) \cap A_{m+1} = \bigcup_{i=1}^n (A_i \cap A_{m+1}) = \bigcup_{i=1}^n \emptyset = \emptyset.$$

Donc par  $\mathbf{H}_2$ ,

$$\mathbf{P}\left(\bigcup_{i=1}^{m+1} A_i\right) = \mathbf{P}\left(\left(\bigcup_{i=1}^m A_i\right) \cup A_{m+1}\right) = \mathbf{P}\left(\bigcup_{i=1}^m A_i\right) + \mathbf{P}(A_{m+1}),$$

et donc d'après  $\mathbf{H}_m$ ,

$$\mathbf{P}\left(\bigcup_{i=1}^{m+1} A_i\right) = \left(\sum_{i=1}^m \mathbf{P}(A_i)\right) + A_{m+1} = \sum_{i=1}^{m+1} \mathbf{P}(A_i).$$

Voici  $\mathbf{H}_{m+1}$  prouvée.

Par récurrence nous venons de montrer la propriété  $\mathbf{H}_n$  pour tout entier  $n \geq 2$ . □

*Preuve du corollaire 2.2.3.* — L'univers  $\Omega$  étant fini il s'écrit  $\{x_1, x_2, \dots, x_n\}$  (où les  $x_i$  sont deux à deux distincts). Si  $n = 1$  alors il n'y a rien à prouver sinon, les événements élémentaires  $\{x_1\}, \{x_2\}, \dots, \{x_n\}$  étant incompatibles et leur réunion  $\Omega$ , 2.2.2. donne le résultat. □

Venons en aux propriétés d'une probabilité, dont certaines ont été intuitées dans l'approche heuristique par les fréquence.

Le corollaire 2.2.3 montre que la probabilité  $\mathbf{P}$  est définie par sa valeur sur les événements élémentaires. On a un peu mieux comme le montre l'exercice suivant.

**Exercice 2.2.4.** — Soit  $\Omega$  un ensemble fini. Soit  $(p_x)_{x \in \Omega}$  une famille de réels.

1. Montrer qu'il existe au plus une probabilité  $\mathbf{P}$  sur l'univers  $\Omega$  telle que, pour tout élément  $x$  de  $\Omega$ ,

$$\mathbf{P}(\{x\}) = p_x.$$

2. Montrer qu'il existe une probabilité  $\mathbf{P}$  sur  $\Omega$  telle que, pour tout élément  $x$  de  $\Omega$ ,

$$\mathbf{P}(\{x\}) = p_x$$

si et seulement si, pour tout  $x \in \Omega$ ,  $p_x \geq 0$  et  $\sum_{x \in \Omega} p_x = 1$ .

**Proposition 2.2.5.** — Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $A$  et  $B$  des événements de  $\Omega$ .

1.  $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$ .
2.  $\mathbf{P}(\emptyset) = 0$ .
3.  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$ .
4. Si  $A \subset B$  alors  $\mathbf{P}(A) \leq \mathbf{P}(B)$ .

**Remarque** — le premier point de cette proposition revêt un caractère pratique important. Il est parfois, en effet, plus facile de calculer la probabilité de l'événement contraire que de l'événement lui même (cf. exercice 2.2.7.).

*Preuve de la proposition 2.2.5.* —

1. Les événements  $A$  et  $\bar{A}$  sont incompatibles donc d'après la définition d'une probabilité,  $1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup \bar{A}) = \mathbf{P}(A) + \mathbf{P}(\bar{A})$ , et donc

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

2. Résulte du point 1 appliqué à  $\Omega$ .

3. Les événements  $A$  et  $B \cap \bar{A}$  sont disjoints et leur réunion est  $A \cup B$ , donc :

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \cup (B \cap \bar{A})) = \mathbf{P}(A) + \mathbf{P}(B \cap \bar{A}). \quad (\text{I.1})$$

D'autre part les événements  $B \cap \bar{A}$  et  $B \cap A$  sont disjoints de réunion  $B$ , donc :

$$\mathbf{P}(B) = \mathbf{P}((B \cap \bar{A}) \cup (B \cap A)) = \mathbf{P}(B \cap \bar{A}) + \mathbf{P}(B \cap A).. \quad (\text{I.2})$$

Finalement grâce à (I.1) et (I.2),

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(B \cap \bar{A})$$

4. Si  $A \subset B$ , alors  $A \cup B = B$  et (I.1) s'écrit :

$$\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \cap \bar{A}) \geq \mathbf{P}(A)$$

Ainsi s'achève la preuve de 2.2.4. □

Lors de la modélisation d'une expérience aléatoire réelle, le choix de la probabilité  $\mathbf{P}$  peut être plus ou moins expérimentale, il peut aussi se fonder sur les symétries du problème. Ainsi lorsque l'expérience est le jet d'un dé, la symétrie même du dé fait qu'il n'y a pas plus de raison qu'il fasse un chiffre plutôt qu'un autre, dans un jeu de carte, une main n'a pas plus de raison de sortir qu'une autre. Pour ce genre d'expérience aléatoire on introduit la probabilité uniforme

qui accorde à chaque événement élémentaire la même probabilité. C'est par défaut la probabilité que l'on considère sans indication particulière.

**Proposition-définition 2.2.6.** — **PROBABILITÉ UNIFORME** — *Soit  $\Omega$  un ensemble fini. Il existe une et une seule probabilité  $\mathbf{P}$  sur l'univers  $\Omega$  qui soit constante sur les événements élémentaires.*

Elle est donnée par :

$$\mathbf{P}(A) = \frac{\text{card} A}{\text{card} \Omega},$$

pour toute partie  $A$  de  $\Omega$ .

Elle vérifie :

$$\mathbf{P}(\{x\}) = \frac{1}{\text{card} \Omega},$$

pour tout élément  $x$  de  $\Omega$ .

*Preuve de la proposition 2.2.6.* —

- Supposons que  $\mathbf{P}$  soit une probabilité sur l'univers  $\Omega$ , constante sur les événements élémentaires, de valeur  $p$ , alors comme  $\Omega$  est la réunion disjointe de tous ses événements élémentaires, d'après le corollaire 2.2.3.,

$$1 = \mathbf{P}(\Omega) = \sum_{x \in \Omega} \mathbf{P}(\{x\}) = \sum_{x \in \Omega} p = \text{card}(\Omega) p.$$

Soit  $p = \frac{1}{\text{card}(\Omega)}$ . Toujours d'après 2.2.3. on a aussi, pour toute partie  $A$  de  $\Omega$

$$\mathbf{P}(A) = \sum_{x \in A} \mathbf{P}(\{x\}) = \sum_{x \in A} p = \text{card}(A) p = \frac{\text{card}(A)}{\text{card}(\Omega)}. \quad (\text{I.3})$$

Il existe donc au plus une probabilité constante sur les singletons de  $\Omega$ .

- Réciproquement l'application

$$\mathbf{P} : \mathcal{P}(\Omega) \rightarrow \mathbf{R}; A \mapsto \frac{\text{card}(A)}{\text{card}(\Omega)},$$

est une probabilité. En effet :

- Les propriétés des cardinaux assurent que  $\mathbf{P}$  est à valeurs dans  $[0, 1]$  ;
- $\mathbf{P}(\Omega) = \frac{\text{card}(\Omega)}{\text{card}(\Omega)} = 1$  ;
- Soit enfin  $A$  et  $B$  des événements disjoints,

$$\mathbf{P}(A \cup B) = \frac{\text{card}(A \cup B)}{\text{card}(\Omega)} = \frac{\text{card}(A) + \text{card}(B)}{\text{card}(\Omega)} = \frac{\text{card}(A)}{\text{card}(\Omega)} + \frac{\text{card}(B)}{\text{card}(\Omega)} = \mathbf{P}(A) + \mathbf{P}(B).$$

Il existe donc une et une seule probabilité  $\mathbf{P}$  sur l'univers  $\Omega$  qui soit constante sur les événements élémentaires. On a montré dans la preuve de l'unicité qu'elle vérifiait les propriétés annoncées.  $\square$

**Exercices 2.2.7.** —

1. Soit  $(\Omega, p)$  un espace probabilisé fini. Soient  $A$  et  $B$  des parties de  $\Omega$  telles que  $\mathbf{P}(A) = \frac{3}{4}$  et  $\mathbf{P}(B) = \frac{1}{3}$ . Montrer que  $\frac{1}{12} \leq \mathbf{P}(A \cap B) \leq \frac{1}{3}$ .
2. Soit un entier  $n \geq 2$ .
  - (a) Quelle est la probabilité qu'une famille à  $n$  enfants soit constituée d'enfants des deux sexes ?
  - (b) Quelle est la probabilité pour qu'une famille ait au plus une fille ?
  - (c) On désigne par  $A$  l'événement décrit dans la première question et par  $B$  celui décrit dans la seconde. Comparez  $\mathbf{P}(A \cap B)$  et  $\mathbf{P}(A)\mathbf{P}(B)$ .

Réponse :  $1 - \frac{1}{2^{n-1}}$  ;  $\frac{n+1}{2^n}$  ;  $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$  si et seulement si  $n = 3$ .

3. Soit un entier  $n \leq 365$ . On considère une classe de  $n$  élèves nés une même année (non bisextile). Quelle est la probabilité qu'au moins deux étudiants soit nés le même jour ?

Réponse :  $1 - \frac{A_{365}^n}{365^n}$ .

4. Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $A$ ,  $B$ , et  $C$  des parties de  $\Omega$ . Montrer que

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

Pour une généralisation (Formule de Poincaré) voir exercices.

## 2.3 Conditionnement

On se pose maintenant la question suivante : Dans une expérience aléatoire lorsque l'on sait qu'un événement  $B$  est réalisé comment mesurer les chances qu'un élément  $A$  le soit ? Intuitivement on conçoit bien que la réalisation de  $B$  change les chances de voir  $A$  se réaliser, par exemple dans le cas de deux jets successifs d'un dé l'événement  $A$  : « la somme des deux jets est supérieure ou égale à 10 » est modifié par la réalisation de l'événement  $B$  « le premier jet donne  $k$  ». Si  $k$  est inférieur ou égal à 3, on voit bien qu'il n'y a aucune chance que  $A$  soit réalisé, dans le cas contraire on sent bien intuitivement que plus  $k$  est grand plus augmentent les chances de voir  $A$  se réaliser. Pour définir la plausibilité de  $A$  sachant que  $B$  est réalisé, reprenons l'approche déjà faite pour définir une probabilité en termes de fréquences. Effectuons  $n$  expériences successives, afin de voir la plausibilité de  $A$  sachant  $B$  réalisé, regardons la fréquence d'apparition de  $A$  lorsque  $B$  est réalisé, c'est-à-dire le rapport du nombre d'expériences à l'issue desquelles  $A$  et  $B$  sont réalisés et du nombre d'expériences à l'issue desquelles  $B$  l'est. Avec les notations du 2.2., ce rapport vaut

$$\frac{nf_n(A \cap B)}{nf_n(B)} = \frac{f_n(A \cap B)}{f_n(B)}$$

Par un passage à la limite, acte fondateur des probabilités, cette quantité tend vers :

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

C'est ce rapport qui va définir la probabilité d'avoir  $A$  sachant  $B$  réalisé.

**Définition 2.3.1.** — Soit  $(\Omega, \mathbf{P})$  un espace probabilisé fini,  $B$  une partie de  $\Omega$  telle que  $\mathbf{P}(B) \neq 0$ . Pour toute partie  $A$  de  $\Omega$ , on appelle probabilité conditionnelle de l'événement  $A$  sachant l'événement  $B$  et on note  $\mathbf{P}(A|B)$  le réel :

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

On retiendra

$$\boxed{\mathbf{P}(A \cap B) = \mathbf{P}(A|B)\mathbf{P}(B)} \quad (\text{I.4})$$

**Proposition 2.3.2.** — Soit  $(\Omega, \mathbf{P})$  un espace probabilisé fini,  $B$  une partie de  $\Omega$  telle que  $\mathbf{P}(B) \neq 0$ , la probabilité conditionnelle sachant  $B$ ,

$$\mathbf{P}(\cdot|B) : \mathcal{P}(\Omega) \rightarrow \mathbf{R}; A \mapsto \mathbf{P}(A|B)$$

est une probabilité sur  $\Omega$ . On la note souvent  $\mathbf{P}_B$ .

*Preuve de la proposition 2.3.2.* —

- Comme  $\mathbf{P}$  est à valeurs positives,  $(\mathbf{P}(\cdot|B))$  l'est aussi et comme, pour toute partie  $A$  de  $\Omega$ ,  $(A \cap B) \subset B$ , la propriété 4 de 2.2.5., assure que  $(\mathbf{P}(\cdot|B))$  est à valeurs dans  $[0, 1]$ ;
- $\mathbf{P}(\Omega|B) = \frac{\mathbf{P}(\Omega \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B)}{\mathbf{P}(B)} = 1$ ;
- Enfin soient  $A_1$  et  $A_2$ , des parties de  $\Omega$  disjointes.

$$\begin{aligned} \mathbf{P}(A_1 \cup A_2|B) &= \frac{\mathbf{P}((A_1 \cup A_2) \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}((A_1 \cap B) \cup (A_2 \cap B))}{\mathbf{P}(B)} = \\ &= \frac{\mathbf{P}(A_1 \cap B) + \mathbf{P}(A_2 \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_1 \cap B)}{\mathbf{P}(B)} + \frac{\mathbf{P}(A_2 \cap B)}{\mathbf{P}(B)} = \mathbf{P}(A_1|B) + \mathbf{P}(A_2|B). \end{aligned}$$

Ainsi  $\mathbf{P}(\cdot|B)$  est-elle une probabilité. □

**Exercice d'application** — Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $A$ ,  $B$  et  $C$  des événements de  $\Omega$ . On suppose  $\mathbf{P}(C) \neq 0$ . Montrer que  $\mathbf{P}(A \cup B|C) = \mathbf{P}(A|C) + \mathbf{P}(B|C) - \mathbf{P}(A \cap B|C)$ .

Nous allons maintenant généraliser (I.4).

**Proposition 2.3.3** — THÉORÈME DES PROBABILITÉS COMPOSÉES —

Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $n$  un entier supérieur ou égal à 2. Pour toute famille  $(A_k)_{k=1, \dots, n}$  de parties de  $\Omega$ , telle que  $\mathbf{P}\left(\bigcap_{k=1}^{n-1} A_k\right) \neq 0$ . Alors :

$$\mathbf{P}\left(\bigcap_{k=1}^n A_k\right) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_1 \cap A_2) \dots \mathbf{P}(A_{n-1}|A_1 \cap A_2 \dots \cap A_{n-2}) \mathbf{P}(A_n|A_1 \cap A_2 \dots \cap A_{n-1}).$$

*Preuve de 2.3.3.* — Observons que puisque  $\mathbf{P}\left(\bigcap_{k=1}^{n-1} A_k\right) \neq 0$ , d'après 2.2.5.-4,  $\mathbf{P}\left(\bigcap_{k=1}^k A_k\right) \neq 0$  pour  $k = 1, 2, \dots, n-1$ , ce qui confère un sens à la formule à prouver.

La propriété à prouver notée  $(H_n)$  se démontre sans surprises par récurrence sur  $n$ .

- La propriété  $(H_2)$  n'est autre que (I.4).
- Soit un entier  $m \geq 2$ . Supposons  $(H_m)$ . Considérons alors  $(A_k)_{k=1, \dots, m+1}$  une famille de parties de  $\Omega$ , telle que

$$\mathbf{P}\left(\bigcap_{k=1}^m A_k\right) \neq 0. \text{ Par (I.4);}$$

$$\mathbf{P}\left(\bigcap_{k=1}^{m+1} A_k\right) = \mathbf{P}(A_1 \cap A_2 \cdots \cap A_m) \mathbf{P}(A_{m+1} | A_1 \cap A_2 \cdots \cap A_m).$$

Donc, compte tenu de  $(H_m)$ ,

$$\mathbf{P}\left(\bigcap_{k=1}^n A_k\right) = \mathbf{P}(A_1) \mathbf{P}(A_2 | A_1) \dots \mathbf{P}(A_{m-1} | A_1 \cap A_2 \cdots \cap A_{m-2}) \mathbf{P}(A_m | A_1 \cap A_2 \cdots \cap A_{m-1}) \mathbf{P}(A_{m+1} | A_1 \cap A_2 \cdots \cap A_m).$$

Voici  $(H_m)$  prouvée.

Donc pour tout entier  $n \geq 2$ ,  $(H_n)$  est vraie. □

**Exercice 2.3.4.** — Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini,  $A$  et  $B$  des parties de  $\Omega$ . On suppose  $\mathbf{P}(A) \neq 0$ .

1. On suppose que pour tout  $x \in \Omega$ ,  $\mathbf{P}(\{x\}) \neq 0$ . Montrer que  $\mathbf{P}(B|A) = 1$  si et seulement si  $B \subset A$ . Si l'on ne suppose plus que pour tout  $x \in \Omega$ ,  $\mathbf{P}(\{x\}) \neq 0$ , qu'elle implication reste vraie dans l'implication précédente ?

**Proposition 2.3.5.** — FORMULE DES PROBABILITÉS TOTALES —

Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $(B_1, \dots, B_n)$  un système complet d'événements, tous de probabilité non nulle. Alors, pour toute partie  $A$  de  $\Omega$ ,

$$\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A|B_i) \mathbf{P}(B_i).$$

*Preuve de la proposition 2.3.5.* — Les  $A \cap B_i$ ,  $i = 1, \dots, n$  sont deux à deux distincts de réunion  $A$ . Donc :

$$\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A \cap B_i) = \sum_{i=1}^n \mathbf{P}(A|B_i) \mathbf{P}(B_i).$$

□

**Exemple 2.3.6.** — PROBLÈME DU CHEVALIER DE MÉRÉ —

Donnons à titre d'exemple un problème historique à l'origine du développement des probabilités. Il fut posé par le Chevalier de Méré à Blaise Pascal. Le voici. Deux joueurs jouent à un jeu de hasard (sans parties nulles), le gagnant et le premier à remporter trois parties. Les joueurs doivent mettre fin à leur rencontre avant la fin, à un moment où le premier joueur a gagné deux parties, le second une seule. Comment partager équitablement la mise entre les deux protagonistes ?

Dit en langage probabiliste, la question sous-entend que la part de la mise que doit empocher chaque joueur est proportionnelle à la probabilité qu'il a de gagner. Voyons ça. On peut faire comme si les joueurs jouaient quoiqu'il arrive encore deux parties, en considérant que si le premier joueur emporte la première de ces deux parties, comme il a gagné, la suivante est jouée pour le plaisir. Ceci permet de prendre comme univers  $\Omega = \{0, 1\}^2$ , un élément  $(x, y)$  de  $\Omega$  représente la situation où la première partie a été gagnée par le joueur  $x$  la suivante par le joueur  $y$ . On munira  $\Omega$  de la probabilité uniforme.

Remarquons que pour un tel univers à chaque partie les deux joueurs ont bien la même probabilité de gagner. En effet l'événement  $B_i$  modélisant « le joueur  $i$  emporte, disons la première partie, » est pour  $i = 1, 2$ ,  $\{B_i = \{(i, y), y \in \{1, 2\}\}\}$  ensemble de cardinal 2. Du reste nous aurions pu ne pas expliciter  $\Omega$  et nous contenter de cette propriété.

La famille  $(B_1, B_2)$  est un système complet d'événements. Soit l'événement  $A$  le premier joueur l'emporte la rencontre. La formule des probabilités totales dit :

$$\mathbf{P}(A) = \mathbf{P}(B_1) \mathbf{P}(A|B_1) + \mathbf{P}(B_2) \mathbf{P}(A|B_2).$$

Nous avons mentionné que  $\mathbf{P}(B_1) = \mathbf{P}(B_2) = \frac{1}{2}$ . Par ailleurs,  $\mathbf{P}(A|B_1) = 1$  (le duel voit quoi qu'il arrive la victoire du premier joueur), tandis que  $\mathbf{P}(A|B_2) = \frac{1}{2}$ , puisque le premier joueur gagne la rencontre si et seulement si il gagne la seconde partie.



Au total

$$\mathbf{P}(A) = \frac{1}{2} \times 1 + \frac{1}{2} \times \frac{1}{2} = \frac{3}{4}.$$

Le premier joueur doit légitimement empocher les  $\frac{3}{4}$  de la mise.

On retiendra que sur cet exemple on n'a pas besoin réellement d'expliciter  $\Omega$ .

Passons à ce que l'on appelle la probabilité des causes.

**Proposition 2.3.7.** — FORMULE DE BAYES OU DE PROBABILITÉ DES CAUSES —

Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $(A_1, \dots, A_n)$  un système complet d'événements de probabilité non nulle. Alors pour toute partie  $B$  de  $\Omega$  de probabilité non nulle, et tout  $j \in \{1, \dots, n\}$

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

Souvent dans la pratique les événements  $A_i$  précèdent chronologiquement l'événement  $B$  et peuvent même être pensé comme des causes possibles de cette événement. La formule de Bayes permet de déterminer la probabilité de chaque cause possible, d'où son nom. D'un pure point de vue mathématique, il ne s'agit que d'un calcul de probabilité conditionnelle et l'aspect temporel n'a pas de réalité.

*Preuve de la proposition 2.3.7.* — Soient  $B$  une partie de  $\Omega$  de probabilité non nulle et  $j$  un élément de  $\{1, \dots, n\}$ . Par définition des probabilités conditionnelles,

$$\mathbf{P}(A_j|B)\mathbf{P}(B) = \mathbf{P}(A_j \cap B) = \mathbf{P}(B|A_j)\mathbf{P}(A_j),$$

cf. (I.4). En appliquant alors à  $B$  la formule des probabilités totales, on a bien le résultat.  $\square$

**Exemple 2.3.8.** — On reprend l'exemple 2.1.1.-4. On note  $b_i$  le nombre de boules blanches supposé non nul, dans l'urne numéro  $i$ ,  $n_i = N - b_i$  désigne le nombre de boules noires. On suppose avoir tiré une boule blanche. Quelle est *a posteriori* la probabilité que le tirage ait eu lieu dans la première urne? Autrement dit notons  $A_i$  l'événement « on a choisi pour le tirage l'urne numéro  $i$  » pour  $i=1,2$  et  $B$  l'événement « on a tiré une boule blanche » ; déterminons  $\mathbf{P}(A_1|B)$ .

$(A_1, A_2)$  est un système complet d'événements, la formule de Bayes affirme :

$$\mathbf{P}(A_1|B) = \frac{\mathbf{P}(B|A_1)\mathbf{P}(A_1)}{\mathbf{P}(B|A_1)\mathbf{P}(A_1) + \mathbf{P}(B|A_2)\mathbf{P}(A_2)}.$$

On peut munir l'univers  $\Omega$ , défini au 2.1.1., d'une probabilité uniforme, ce n'est pas ici l'essentiel, ce qui est primordial c'est que pour ce choix, la probabilité de tirer une urne ou une autre est la même et que la probabilité de tirer une boule dans chaque urne est uniforme. On aurait pu se limiter à cette hypothèse sans préciser ni  $\Omega$ , ni la probabilité dont on l'équipe, comme dans l'exemple 2.2.6.

Ainsi  $\mathbf{P}(B|A_i) = \frac{b_i}{N}$ , et  $\mathbf{P}(A_i) = \frac{1}{2}$ . Donc :

$$\mathbf{P}(A_1|B) = \frac{\frac{b_1}{N} \times \frac{1}{2}}{\frac{b_1}{N} \times \frac{1}{2} + \frac{b_2}{N} \times \frac{1}{2}} = \frac{b_1}{b_1 + b_2}.$$

**Exercice 2.3.9.** — Un client achète une ampoule dans un magasin. Dans ce magasin 30% des ampoules proviennent d'une usine  $F_1$  le reste d'une usine  $F_2$ . A la sortie de l'usine  $F_1$  2 % des ampoules sont défectueuses, seulement 1% à la sortie de l'usine  $F_2$ . Sachant que l'ampoule achetée par le client est défectueuse, quelle est la probabilité qu'elle ait été fabriquée par la première usine.

Réponse :  $\frac{6}{13}$ .

## 2.4 Événements indépendants

Commençons par une démarche heuristique. Cherchons à modéliser le fait que dans une expérience aléatoire des événements sont indépendants, c'est-à-dire que la réalisation de l'un ne modifie en rien les chances de réalisation de l'autre. Supposons la situation modélisée par un espace probabilisé fini  $(\Omega, \mathbf{P})$  et les deux événements par des parties  $A$  et  $B$  de  $\Omega$ . Admettons momentanément que ces événements sont de probabilités non nulles. Dire que les chances de réalisation du premier événement  $A$  ne sont pas modifiées par la réalisation du second événement se traduit naturellement par le fait que  $\mathbf{P}(A|B) = \mathbf{P}(A)$ . Ce qui s'écrit encore  $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$ . Cette dernière expression

est symétrique en  $A$  et  $B$  et va servir de définition à l'indépendance de deux événements (de probabilité non nulle ou pas).

**Définition 2.4.1.** — Soit  $(\Omega, \mathbf{P})$  un espace probabilisé fini. On dit que des événements (parties)  $A$  et  $B$  de  $\Omega$  sont indépendants si

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

Soient un entier  $m \geq 2$  et  $A_1, A_2, \dots, A_m$  des événements de  $\Omega$ , on dit qu'ils sont deux à deux indépendants si, pour tout couple  $(i, j)$  d'éléments distincts de  $\{1, \dots, m\}$ ,

$$\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j).$$

De la définition d'une probabilité conditionnelle il vient immédiatement :

**Proposition 2.4.1.** — Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $A$  et  $B$  des événements de  $\Omega$ . On suppose de surcroît que  $\mathbf{P}(B) \neq 0$ . Alors  $A$  et  $B$  sont indépendants si et seulement si  $\mathbf{P}(A|B) = \mathbf{P}(A)$ .

**Exemple 2.4.2.** —

- Reprenons l'exemple 2.1.1.-5 sur les  $N$  lancers successifs d'un dé. On regarde l'événement  $A$  « le premier lancé donne face » et l'événement  $B$  « le second lancé donne face ». Montrons qu'ils sont indépendants.

$$A = \{(F, X_2, X_3, \dots, X_N) | (X_2, X_3, \dots, X_N) \in \{P, F\}^{N-1}\},$$

$$B = \{(X_1, F, X_3, \dots, X_N) | (X_1, X_3, \dots, X_N) \in \{P, F\}^{N-1}\}.$$

Le cardinal de  $A$  et  $B$  est  $2^{N-1}$ . Donc  $\mathbf{P}(A) = \mathbf{P}(B) = \frac{2^{N-1}}{\text{card}(\Omega)} = \frac{2^{N-1}}{2^N} = \frac{1}{2}$ .

Par ailleurs  $A \cap B = \{(F, F, X_3, \dots, X_N) | (X_3, \dots, X_N) \in \{P, F\}^{N-2}\}$  et donc :

$$\mathbf{P}(A \cap B) = \frac{2^{N-2}}{2^N} = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbf{P}(A)\mathbf{P}(B).$$

Les événements  $A$  et  $B$  sont bien indépendants, ce qui est conforme à notre intuition le fait de faire face au premier lancé et de faire face au deuxième non aucun rapport l'un avec l'autre.

- Prenons à présent l'exemple 2.3.8 où l'on modélisait le choix d'une urne puis le tirage d'une boule dans l'urne choisie. Considérons l'événement  $B$  « la boule tirée est blanche » et l'événement  $U_1$  « l'urne choisie est la première ». Intuitivement on conçoit fort bien que la probabilité d'obtenir une boule blanche va dépendre du choix de l'urne si les deux urnes n'ont pas le même nombre de boules blanches. Voyons cela.

$$U_1 = \{(1, y), y \in \{1, \dots, N\}\}.$$

Supposons que dans les deux urnes nous ayons numéroté d'abord les boules blanches (dans l'urne numéro  $i$ , les boules blanches sont celles numérotées de 1 à  $n_i$ ). L'événement  $B$  vaut alors

$$B = \{(1, y), y \in \{1, n_1\}\} \cup \{(2, y), y \in \{1, n_2\}\}.$$

Donc,  $B$  étant écrit comme la réunion de deux événements incompatibles

$$\mathbf{P}(B) = \mathbf{P}(\{(1, y), y \in \{1, n_1\}\}) + \mathbf{P}(\{(2, y), y \in \{1, n_2\}\}) = \frac{n_1}{\text{card}\Omega} + \frac{n_2}{\text{card}\Omega} = \frac{n_1 + n_2}{2 \times N}; \quad \mathbf{P}(U_1) = \frac{1}{2}.$$

L'événement  $B \cap U_1 = \{(1, y), y \in \{1, n_1\}\}$ , donc  $\mathbf{P}(B \cap U_1) = \frac{n_1}{2N}$ .

Ainsi a-t-on :  $\mathbf{P}(B \cap U_1) = \mathbf{P}(B)\mathbf{P}(U_1)$  si et seulement si :

$$\frac{n_1}{2N} = \frac{1}{2} \times \frac{n_1 + n_2}{2 \times N},$$

soit si et seulement si  $\frac{n_1}{2} = \frac{n_1 + n_2}{2}$ . Finalement  $B$  et  $U_1$  sont indépendants si et seulement si  $n_1 = n_2$  ce qui est conforme à notre intuition.

**Exercices 2.4.3.** — Soit  $A$  et  $B$  des événements d'un espace probabilisé fini  $(\Omega, \mathbf{P})$ .

1. Montrer que si  $A$ , et  $B$  sont indépendants, alors  $A$  et  $\bar{B}$  le sont, ainsi que  $\bar{A}$  et  $\bar{B}$ .
2. On suppose que  $A \cap B = \emptyset$ . Montrer que  $A$  et  $B$  sont indépendants si et seulement si  $\mathbf{P}(A) = 0$  ou  $\mathbf{P}(B) = 0$ .
3. On suppose  $A$  et  $\bar{A}$  indépendants. Que dire de la probabilité de  $A$  ?

**Définition 2.4.4.** — soit  $(\Omega, \mathbf{P})$  un espace probabilisé fini. Soient un entier  $m \geq 2$  et  $A_1, A_2, \dots, A_m$  des événements de  $\Omega$ , on dit qu'ils sont (mutuellement) indépendants si, pour toute partie finie  $J$  de  $\{1, \dots, m\}$  non vide,

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

Pour vérifier concrètement la mutuelle indépendance de  $m$  éléments il faut vérifier  $2^m - m - 1$  conditions (le nombre de parties de  $\{1, \dots, m\}$  ayant strictement plus d'un élément).

Des événements de  $\Omega$ , mutuellement indépendants, sont deux à deux indépendants, il suffit pour le voir de prendre en particulier dans la définition de la mutuellement indépendance, les parties  $J$  à 2 éléments. En revanche la réciproque est fautive, comme le montre l'exemple suivant. Prenons  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$  et munissons cette ensemble de la probabilité uniforme. Posons  $A = \{\omega_1, \omega_2\}$ ,  $B = \{\omega_1, \omega_3\}$ ,  $C = \{\omega_1, \omega_4\}$ .

$$\mathbf{P}(A \cap B) = \mathbf{P}(A \cap C) = \mathbf{P}(B \cap C) = \mathbf{P}(\{\omega_1\}) = \frac{1}{4}.$$

Les événements  $A$ ,  $B$  et  $C$  sont donc bien deux à deux indépendants. En revanche :  $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(\{\omega_1\}) = \frac{1}{4}$  tandis que  $\mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) = \left(\frac{2}{4}\right)^3 = \frac{1}{8}$ .

Dans le même ordre d'idées, on peut avoir  $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$  sans que  $A, B, C$  soient indépendants deux à deux et donc *a fortiori* mutuellement indépendants, comme le lecteur pourra s'en convaincre en prenant  $A = B$ , avec  $\mathbf{P}(A) \in ]0, 1[$ .

**Exercice 2.4.5.** — Soient  $(\Omega, \mathbf{P})$  un espace probabilisé fini et  $A_1, \dots, A_n$  des événements de  $\Omega$ .

- Supposons les événements  $A_1, \dots, A_n$  (mutuellement) indépendants. Montrer que les événements  $A_1$  et  $\bigcap_{i=2}^n A_i$  sont indépendants.
- Montrer que  $A_1, \dots, A_n$  sont (mutuellement) indépendants si et seulement si, pour toute partie  $J$  non vide de  $\{1, \dots, n\}$  telle que  $\mathbf{P}\left(\bigcap_{j \in J} A_j\right) \neq 0$ , et tout élément  $i$  de  $\{1, \dots, n\}$ , si  $i \notin J$  alors :

$$\mathbf{P}\left(A_i \mid \bigcap_{j \in J} A_j\right) = \mathbf{P}(A_i).$$

Comparer avec 2.4.1.

## VARIABLES ALÉATOIRES

Lorsque l'on a modélisé une expérience aléatoire simple (par exemple par la loi uniforme), il est alors possible d'étudier des événements plus complexes qui sont fonctions des issues de notre expérience... C'est le rôle des variables aléatoires, qui sont, contrairement à ce que leur nom laisse accroire des fonctions définies sur l'univers. Elles vont nous permettre d'associer une probabilité mesurant des événements complexes et difficiles à modéliser.

### 3.1 Premières définitions

**Définition 3.1.1.** — Soit  $(\Omega, \mathbf{P})$  un espace probabilisé. On appelle variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbf{P})$ , toute application  $X$  de  $\Omega$  dans un ensemble quelconque  $E$ .

Si de plus  $E$  est une partie de  $\mathbf{R}$ , on parle de variable aléatoire réelle.

**Exemples 3.1.2.** — On reprend l'espace probabilisé fini défini en 2.1.1-2 qui modélise le lancé de deux dés. On peut définir sur  $(\Omega, \mathcal{A}, \mathbf{P})$  les variables aléatoires suivantes :

- (a)  $X_1 : \{1, \dots, 6\}^2 \rightarrow \mathbf{R}; (x, y) \mapsto x$  (résultat du premier dé),  
 (b)  $X_2 : \{1, \dots, 6\}^2 \rightarrow \mathbf{R}; (x, y) \mapsto \max\{x, y\}$  (résultat maximum des deux dés),  
 (c)  $X_3 : \{1, \dots, 6\}^2 \rightarrow \mathbf{R}; (x, y) \mapsto x + y$  (somme des résultats des deux dés),  
 (d)  $X_4 : \{1, \dots, 6\}^2 \rightarrow \mathbf{R}; (x, y) \mapsto 1$  si  $x \geq y$ , 2 sinon (numéro du dé affichant le plus grand nombre),
- On considère l'espace probabilisé fini  $(S_n, \mathbf{P})$  où  $S_n$  est le groupe symétrique à  $n$  éléments et  $p$  la probabilité uniforme sur  $S_n$ .

- (a) On définit sur cet espace probabilisé la variable aléatoire  $N$  qui à un élément  $\sigma$  de  $S_n$  associe le nombre de ses points fixes.
- (b) On peut aussi considérer la variable aléatoire  $M$  qui à un élément  $\sigma$  de  $S_n$  associe la matrice de permutation  $P_\sigma$ , c'est-à-dire l'élément de  $\mathcal{M}_n(\mathbf{R})$  dont la  $j^{\text{e}}$  colonne est le  $\sigma(j)^{\text{e}}$  vecteur de la base canonique de  $\mathcal{M}_{n,1}(\mathbf{R})$  (un 1 en  $\sigma(j)^{\text{e}}$  ligne, des 0 ailleurs).
3. Soient  $X_1, X_2, \dots, X_m$  des variables aléatoires définies sur un espace probabilisé  $(\Omega, \mathbf{P})$  à valeurs respectivement dans des espaces  $E_1, E_2, \dots, E_m$ , alors on dispose d'une variable aléatoire

$$Z : \Omega \rightarrow E_1 \times \dots \times E_m ; \omega \mapsto (X_1(\omega), X_2(\omega), \dots, X_m(\omega)).$$

On la notera abusivement  $(X_1, \dots, X_m)$ .

Dans ces exemples, les variables aléatoires  $M$  et  $Z$  ne sont pas réelles, les autres le sont.

**Notations** — Les probabilistes recourent sans vergogne pour les variables aléatoires à des notations abusives proscrites partout ailleurs. Voyons cela :

Soient  $X$  une variable aléatoire définie sur un espace probabilisé fini  $(\Omega, \mathbf{P})$ , à valeurs dans un ensemble  $E$ ,  $A$  et  $B$  des parties de  $E$  et  $a$  un élément de  $E$ .

NOTATIONS RIGOUREUSES	NOTATIONS PROBABILISTES
$X^{-1}(A)$ ou $\{\omega \in \Omega   X(\omega) \in A\}$	$\{X \in A\}$ ou $(X \in A)$
$X^{-1}(\{a\})$	$\{X = a\}$ ou $(X = a)$
Pour $E = \mathbf{R}$ , $X^{-1}([a, +\infty[)$	$\{X \geq a\}$ ou $(X \geq a)$
$\mathbf{P}(X^{-1}(A))$	$\mathbf{P}(X \in A)$
$\mathbf{P}(X^{-1}(\{a\}))$	$\mathbf{P}(X = a)$
Pour $E = \mathbf{R}$ , $\mathbf{P}(X^{-1}([a, +\infty[))$	$\mathbf{P}(X \geq a)$
$\vdots$	$\vdots$

**Définition 3.1.3.** — Soient  $X$  une variable aléatoire sur un espace probabilisé  $(\Omega, \mathbf{P})$  à valeurs dans  $E$  et  $f$  une application de  $E$  dans un ensemble  $F$ . L'application  $f \circ X$  définit alors une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbf{P})$  à valeurs dans  $F$ , on l'appelle image de  $X$  par  $f$  et on la note abusivement  $f(X)$ .

## 3.2 Loi d'une variable aléatoire

Nous allons maintenant associer à une variable aléatoire une probabilité définie sur son image.

**Proposition-définition 3.2.1.** — LOI D'UNE VARIABLE ALÉATOIRE —

Soit  $X$  une variable aléatoire définie sur un espace probabilisé fini  $(\Omega, \mathbf{P})$  à valeurs dans un ensemble  $E$ . Alors l'application

$$\mathbf{P}_X : \mathcal{P}(X(\Omega)) \rightarrow [0, 1] ; A \mapsto \mathbf{P}(X \in A)$$

est une probabilité sur  $X(\Omega)$ .

On appelle cette probabilité, loi de la variable aléatoire  $X$ .

*Preuve de la proposition 3.2.1.* —

Remarquons pour commencer que  $X(\Omega)$  est fini puisque  $\Omega$  l'est, ce, que  $E$  le soit ou non. Par ailleurs  $\mathbf{P}_X$  est bien à valeurs dans  $[0, 1]$  puisque  $\mathbf{P}$  l'est.

- $\{X \in X(\Omega)\} = \Omega$ , donc  $\mathbf{P}_X(X(\Omega)) = \mathbf{P}(\Omega) = 1$ .
- Soient  $A$  et  $B$  des parties disjointes de  $X(\Omega)$ .

D'une part, par définition de  $\mathbf{P}_X$ ,

$$\mathbf{P}_X(A \cup B) = \mathbf{P}(X \in A \cup B) = \mathbf{P}(X^{-1}(A \cup B)) = \mathbf{P}(X^{-1}(A) \cup X^{-1}(B)) = \mathbf{P}(\{X \in A\} \cup \{X \in B\}).$$

D'autre part,

$$\{X \in A\} \cap \{X \in B\} = X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B) = X^{-1}(\emptyset) = \emptyset,$$

donc  $\{X \in A\}$  et  $\{X \in B\}$  sont deux événements de  $\Omega$  disjoints. Donc

$$\mathbf{P}_X(A \cup B) = \mathbf{P}(\{X \in A\}) + \mathbf{P}(\{X \in B\}) = \mathbf{P}_X(A) + \mathbf{P}_X(B).$$

De ces deux points il vient que  $\mathbf{P}_X$  est une probabilité. □

**Remarque :** On pourrait songer à prolonger  $\mathbf{P}_X$  en une probabilité définie sur  $E$ , par  $(\mathbf{P}_X(A) = \mathbf{P}(X = A)$  pour tout  $A$  dans  $E$ ) mais comme  $E$  n'est pas nécessairement fini, on sort du cadre du programme de première année et l'on verra cette année que la chose nécessite quelques précautions.

On écrira

Soient  $X$  une variable aléatoire définie sur un espace probabilisé fini  $(\Omega, \mathbf{P})$  à valeurs dans un ensemble  $E$  et  $f$  une application de  $E$  dans un ensemble  $F$ . Soit  $B$  une partie de  $f(X(\Omega))$ , alors :

$$\mathbf{P}_{f(X)}(B) = \mathbf{P}(f(X) \in B) = \mathbf{P}(X \in f^{-1}(B)) = \mathbf{P}_X(f^{-1}(B)).$$

Le corollaire 2.2.3. nous a appris qu'une probabilité sur un univers fini est entièrement définie par ses valeurs sur les singletons, il en résulte le résultat suivant.

**Proposition 3.2.2.** — *Soit  $X$  une variable aléatoire définie sur un espace probabilisé fini  $(\Omega, \mathbf{P})$  à valeurs dans un ensemble  $E$ . Alors la loi de  $X$  est entièrement déterminée par les valeurs  $\mathbf{P}(X = x)$ , où  $x$  décrit  $X(\Omega)$ .*

Plus précisément pour toute partie  $A$  de  $X(\Omega)$ ,

$$\mathbf{P}_X(A) = \sum_{x \in A} \mathbf{P}(X = x).$$

Nous allons donner maintenant des lois de variables aléatoires qui modélisent de nombreuses situations.

Nous supposons qu'est donné un espace probabilisé fini  $(\Omega, \mathbf{P})$ . En vertu de 3.2.2., nous définirons les lois des variables aléatoires  $X$  par les valeurs  $\mathbf{P}(X = x)$ .

LOI UNIFORME —

**Définition 3.2.3.** — *On dit qu'une variable aléatoire  $X$  à valeurs dans  $E$  ensemble fini suit la loi uniforme si, pour tout  $x \in E$ ,*

$$\mathbf{P}(X = x) = \frac{1}{|E|}$$

On note :

$$X \sim \mathcal{U}(E).$$

C'est la situation dégénérée où  $\mathbf{P}$  est la loi uniforme et  $X$  l'identité, mais pas que. Prenons pour  $\Omega$ , l'ensemble  $\{0, \dots, 99\}$  que l'on munit de la probabilité uniforme, et  $X$  la variable qui à un élément  $\omega$  de  $\Omega$  associe le dernier chiffre dans l'écriture décimale de  $\omega$ , ici  $E = \{0, \dots, 9\}$  et  $X \sim \mathcal{U}(E)$ , puisque pour  $i = 0, \dots, 9$ ,

$$\{X = i\} = \{i, \overline{1i}, \overline{2i}, \dots, \overline{9i}\},$$

et donc

$$\mathbf{P}_X(\{i\}) = \mathbf{P}(X = i) = \frac{|\{X = i\}|}{|E|} = \frac{10}{100} = \frac{1}{10} = \frac{1}{|E|}.$$

LOI DE BERNOULLI —

Commençons par un exemple. Le lancer d'une pièce non équilibrée conduit à deux issues possibles *pile* ou *face*, l'univers est donc  $\{P, F\}$ , la probabilité de  $\{P\}$  est prise égale à  $p$  élément de  $]0, 1[$ , donc celle de  $\{F\}$  à  $1 - p$ . La variable aléatoire  $X$  qui associe 1 à  $P$  et 0 à  $F$  suit la loi :

$$\mathbf{P}_X(\{1\}) = p, \mathbf{P}_X(\{0\}) = 1 - p.$$

De façon générale :

**Définition 3.2.4.** — *Soit  $p$  un élément de  $[0, 1]$ . On dit qu'une variable aléatoire  $X$  à valeurs dans  $\{0, 1\}$  suit la loi de Bernoulli de paramètre  $p$  si,  $\mathbf{P}(X = 1) = p$  (donc  $\mathbf{P}(X = 0) = 1 - p$ ) On note :*

$$X \sim \mathcal{B}(p).$$

Notons que  $X \sim \mathcal{B}(\frac{1}{2})$  ne signifie rien d'autre que  $X \sim \mathcal{U}(\{0, 1\})$ .

La loi de Bernoulli est très utile : dans toute expérience à deux issues, (pile ou face, victoire ou défaite...) la variable aléatoire qui renvoie 1 pour une issue, 0 pour l'autre, suit une loi de Bernoulli. Par abus on dit encore que la variable aléatoire qui renvoie le résultat de l'expérience (qui n'est autre que l'identité) suit une loi de Bernoulli, ce faisant on identifie implicitement les deux issues à 0 et 1.

Une variable aléatoire qui suit une loi de Bernoulli de paramètre  $p$  est dite simplement *variable de Bernoulli de paramètre  $p$* .

**Proposition 3.2.5** — Soit  $A$  un événement de  $(\Omega, \mathcal{A}, \mathbf{P})$  de probabilité  $p$ . La fonction indicatrice (caractéristique) de  $A$ ,  $\mathbf{1}_A$  est une variable aléatoire de Bernoulli de paramètre  $p$ .

*Preuve de 2.2.5.* — On a  $\mathbf{P}(\mathbf{1}_A = 1) = \mathbf{P}(A) = p$  et  $\mathbf{P}(\mathbf{1}_A = 0) = \mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A) = 1 - p$ .

## LOI BINOMIALE

Commençons là encore par un exemple.

On considère une urne contenant  $N$  boules blanches et  $M$  boules noires. On tire  $n$  boules avec remise. L'univers est bien sûr, après numérotation des boules,  $\{1, \dots, N + M\}^n$  et il est naturellement muni de la probabilité uniforme. On note  $X$  la variable aléatoire qui à un élément de  $\Omega$ , associe le nombre de boules blanches obtenues, mathématiquement le nombre de ses composantes qui sont le numéro d'une boule blanche. Soit  $k \in \{0, \dots, n\}$ . Etant donné une partie  $J$  à  $k$  éléments de  $\{1, \dots, n\}$  le nombre d'éléments  $(x_1, \dots, x_n)$  de  $\Omega$ , tels que pour tout  $i \in \{1, \dots, n\}$ ,  $x_i$  soit le numéro d'une boule blanche si  $j \in J$  et d'une boule noire sinon, vaut  $N^k M^{n-k}$ . Or il existe  $\binom{n}{k}$  parties de  $\{1, \dots, n\}$  à  $k$  éléments donc le cardinal de  $\{X = k\}$  est  $\binom{n}{k} N^k M^{n-k}$ . Comme le cardinal de  $\Omega$  est  $(N + M)^n$  on a :

$$\mathbf{P}(X = k) = \binom{n}{k} \left( \frac{N}{N + M} \right)^k \left( \frac{M}{N + M} \right)^{n-k}.$$

Posons  $p = \frac{N}{N + M}$ , alors :

$$\mathbf{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Plus généralement :

**Définition 3.2.6.** — Soit un entier  $n \geq 1$  et  $p$  un élément de  $[0, 1]$ . On dit qu'une variable aléatoire  $X$  à valeurs dans  $\{0, \dots, n\}$  suit la loi binomiale de taille  $n$  et de paramètre  $p$  (ou plus négligemment, de paramètres  $n$  et  $p$ ) si,

$$\mathbf{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

On note :

$$X \sim \mathcal{B}(n, p).$$

Une variable aléatoire qui suit une loi binomiale de paramètre  $n$  et  $p$  est dite *variable binomiale de paramètres  $n$  et  $p$* .

## Remarques —

1. On retrouve que  $\mathbf{P}_X(\{1, \dots, n\}) = 1$  grâce au binôme de Newton, qui donne son nom à la loi :

$$\mathbf{P}_X(\{1, \dots, n\}) = \sum_{k=0}^n \mathbf{P}(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} = (p + (1 - p))^n = 1^n = 1.$$

2. La loi binomiale de paramètre 1 et  $p$  est la loi de Bernoulli de paramètre  $p$ .

## Exercice —

Soit  $X$  une variable aléatoire.

1. On suppose que  $X \sim \mathcal{U}(\{1, \dots, n\})$  identifier la loi de  $n - X$ .
2. On suppose que  $X \sim \mathcal{B}(n, p)$  identifier la loi de  $n - X$ .

**Exercice 3.2.7.** — Une urne contient  $N$  boules blanches et  $M$  boules noires. On tire  $n$  boules sans remise, donc  $n \geq N + M$ . On note  $X$  le nombre de boules blanches obtenues.

1. Décrire l'espace probabilisé  $(\Omega, \mathbf{P})$  qui modélise cette expérience aléatoire.
2. Soit  $k \in \{0, n\}$ . Déterminer  $\mathbf{P}(X = k)$ .
3. On suppose que  $N$  et  $M$  tendent vers  $+\infty$  de telle sorte que  $\frac{N}{N+M}$  tende vers une limite finie  $p$ . Montrer que  $\mathbf{P}(X = k)$  tend vers  $\binom{n}{k} p^k (1-p)^{n-k}$ . Interpréter ce résultat.

*Solution de l'exercice 3.2.7.*—

1. Comme on tire les boules sans remise et que l'on ne se soucie pas de l'ordre dans lequel elles ont été prélevées, il est loisible de considérer que l'on tire les  $n$  boules d'un coup et de prendre pour  $\Omega$  l'ensemble des parties à  $n$  éléments de  $\{1, \dots, N+M\}$ , son cardinal est  $\binom{N+M}{n}$ . On munit naturellement  $\Omega$  de la probabilité uniforme.
2. On dispose de  $X$  qui est une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbf{P})$ . Si  $k > N$  ou  $n-k > M$  alors l'ensemble  $\{X = x\}$  est vide, sinon  $|\{X = x\}| = \binom{N}{k} \binom{M}{n-k}$ . Donc :

$$\mathbf{P}(X = k) = \begin{cases} \frac{\binom{N}{k} \binom{M}{n-k}}{\binom{N+M}{n}}, & \text{si } k \leq N \text{ et } n-k \leq M, \\ 0, & \text{sinon.} \end{cases}$$

3. Calcul élémentaire. Lorsque le nombre de boules blanches et noires devient très grand le fait de ne pas remettre ou de remettre les boules ne change pas grand chose, ce qui explique le résultat.

### 3.3 Couples et $n$ uplets de variables aléatoires

Dans tout ce paragraphe nous est donné un espace probabilisé fini  $(\Omega, \mathbf{P})$ . Toutes les variables aléatoires sont définies sur cet espace.

On a vu que si l'on dispose de deux variables aléatoires  $X_1$  et  $X_2$  à valeurs respectivement dans  $E_1$  et  $E_2$ , d'après 3.1.3.-3  $(X_1, X_2)$  est une variable aléatoire à valeurs dans  $E_1 \times E_2$ , la loi de  $(X_1, X_2)$  s'appellera loi conjointe de  $X_1$  et  $X_2$ . Réciproquement si  $X$  est une variable aléatoire à valeurs dans un produit cartésien  $E_1 \times E_2$ , la première et la seconde composante de  $X$  sont des variables aléatoires, leurs lois sont appelées respectivement première et seconde loi marginale de  $X$ . Nous allons donner une définition plus générale dans le cas de  $p$  variables aléatoires.

Jusqu'à la fin du paragraphe  $p$  désigne un entier supérieur ou égal à 2, et  $E_1, E_2, \dots, E_p$  des ensembles.

**Définition 3.3.1.** — Soient  $X_1, X_2, \dots, X_p$  des variables aléatoires à valeurs respectivement dans les ensembles  $E_1, E_2, \dots, E_p$ . On appelle loi conjointe (ou jointe) de ces variables (prises dans cet ordre) la loi de la variable aléatoire  $(X_1, X_2, \dots, X_p)$ . Autrement dit, en notant  $X = (X_1, \dots, X_p)$ , pour  $(\omega_1, \omega_2, \dots, \omega_p)$  un élément de  $X(\Omega)$ ,

$$\mathbf{P}_X(\{(\omega_1, \omega_2, \dots, \omega_k)\}) = \mathbf{P}(X_1 = \omega_1, \dots, X_p = \omega_p).$$

Si  $X$  est une variable aléatoire à valeurs dans  $E_1 \times E_2 \times \dots \times E_p$ , pour  $k = 1, 2, \dots, p$ , on appelle  $k^{\text{e}}$  loi marginale de  $X$  la loi de  $X_k$ ,  $k^{\text{e}}$  composante de  $X$ .

On peut déduire facilement une loi marginale de la loi d'une variable :

**Proposition 3.3.2** — Soit  $X = (X_1, \dots, X_p)$  une variable aléatoire à valeurs dans  $E_1 \times \dots \times E_p$ . Pour  $k$  élément de  $\{1, \dots, p\}$ , la  $k^{\text{e}}$  loi marginale de  $X_k$ ,  $P_{X_k}$  est donnée par :

$$\mathbf{P}_{X_k}(A) = \mathbf{P}\left(X \in (E_1 \times \dots \times E_{k-1} \times A \times E_{k+1} \times \dots \times E_p)\right),$$

pour tout  $A \subset X_k(\Omega)$ .

**Preuve de la proposition 3.3.2.** —

Soit  $A \subset X_k(\Omega)$ . L'ensemble  $\{X_k \in A\} = \{X \in E_1 \times \dots \times E_{k-1} \times A \times E_{k+1} \times \dots \times E_p\}$ , le résultat en résulte instantanément.  $\square$

**Illustration lorsque  $p = 2$ ,** — Soient  $X$  et  $Y$  des variables aléatoires indépendantes et qui suivent la même loi binomiale d'ordre  $n$  et de paramètre  $p$ . On note  $U = \min\{X, Y\}$  et  $V = \max\{X, Y\}$ .

Déterminons pour commencer la loi du couple  $(U, V)$ . En premier lieu

$$(U, V)(\Omega) = \{(i, j) \in \{0, \dots, n\}^2 | i \leq j\}.$$

Soit alors  $(i, j) \in (U, V)(\Omega)$ . Deux cas.

— Premier cas  $i = j$ .

$$\mathbf{P}(U = i, V = j) = \mathbf{P}(X = i, Y = i) = \mathbf{P}(X = i)\mathbf{P}(Y = i) = \left(\binom{n}{i} p^i (1-p)^{n-i}\right)^2,$$

car  $X$  et  $Y$  sont indépendantes.

— second cas  $i < j$

$$\mathbf{P}(U = i, V = j) = \mathbf{P}\left(\left((X = i) \cap (Y = j)\right) \cup \left((X = j) \cap (Y = i)\right)\right) = \mathbf{P}(X = i, Y = j) + \mathbf{P}(X = j, Y = i),$$

Car les événements  $((X = i) \cap (Y = j))$  et  $((X = j) \cap (Y = i))$  sont incompatibles, or  $X$  et  $Y$  sont indépendantes, donc,

$$\mathbf{P}(U = i, V = j) = \mathbf{P}(X = i)\mathbf{P}(Y = j) + \mathbf{P}(X = j)\mathbf{P}(Y = i) = 2 \binom{n}{i} p^i (1-p)^{n-i} \binom{n}{j} p^j (1-p)^{n-j}.$$

Au total,

$$\mathbf{P}(U = i, V = j) = \begin{cases} \binom{n}{i}^2 p^{2i} (1-p)^{2n-2i}, & \text{si } i = j, \\ 2 \binom{n}{i} \binom{n}{j} p^{i+j} (1-p)^{2n-i-j}, & \text{si } i < j. \end{cases}$$

Etudions par exemple la première loi marginale de  $(U, V)$ .

Soit  $i \in \{0, \dots, n\}$ .

$$\mathbf{P}_U(\{i\}) = \mathbf{P}((U, V) \in \{i\} \times \{1, \dots, n\}) = \sum_{j=0}^n \mathbf{P}((U, V) = (i, j)) = \sum_{j=i}^n \mathbf{P}_{(U,V)}(i, j),$$

en effet les événements  $\{(U, V) = (i, j)\}$ ,  $j = 0, \dots, n$  sont deux à deux disjoints de réunion  $\{(U, V) = (i, j)\}$ .  
Donc

$$\mathbf{P}_U(\{i\}) = \binom{n}{i}^2 p^{2i} (1-p)^{2n-2i} + \sum_{j=i+1}^n 2 \binom{n}{i} \binom{n}{j} p^{i+j} (1-p)^{2n-i-j},$$

En convenant qu'une somme vide est nulle.

On peut présenter synthétiquement ces résultats dans un tableau.

	0	1	...	$j$	...	$n$	
0	$\mathbf{P}_{(U,V)}(\{0,0\})$	$\mathbf{P}_{(U,V)}(\{0,1\})$	...	$\mathbf{P}_{(U,V)}(\{0,j\})$	...	$\mathbf{P}_{(U,V)}(\{0,n\})$	$\rightarrow \mathbf{P}_U(\{0\})$
1	$\mathbf{P}_{(U,V)}(\{1,0\})$	$\mathbf{P}_{(U,V)}(\{1,1\})$	...	$\mathbf{P}_{(U,V)}(\{1,j\})$	...	$\mathbf{P}_{(U,V)}(\{1,n\})$	$\rightarrow \mathbf{P}_U(\{1\})$
$\vdots$							$\vdots$
$i$	$\mathbf{P}_{(U,V)}(\{i,0\})$	$\mathbf{P}_{(U,V)}(\{i,1\})$	...	$\mathbf{P}_{(U,V)}(\{i,j\})$	...	$\mathbf{P}_{(U,V)}(\{i,n\})$	$\rightarrow \mathbf{P}_U(\{i\})$
$\vdots$							$\vdots$
$n$	$\mathbf{P}_{(U,V)}(\{n,0\})$	$\mathbf{P}_{(U,V)}(\{n,1\})$	...	$\mathbf{P}_{(U,V)}(\{n,j\})$	...	$\mathbf{P}_{(U,V)}(\{n,n\})$	$\rightarrow \mathbf{P}_U(\{n\})$
	$\downarrow$	$\downarrow$		$\downarrow$		$\downarrow$	
	$\mathbf{P}_V(\{0\})$	$\mathbf{P}_V(\{1\})$	...	$\mathbf{P}_V(\{j\})$	...	$\mathbf{P}_V(\{n\})$	

Les sommes en lignes donnent  $\mathbf{P}_U$ , celles en colonne  $\mathbf{P}_V$ .

On a noté abusivement  $\mathbf{P}((U, V) = (i, j))$  plus simplement  $\mathbf{P}_{(U,V)}(\{i, j\})$  même si  $(i, j)$  n'est pas élément de  $(U, V)(\Omega)$ .

En revanche, avec les notations de 3.3.1., la donnée pour une variable  $X = (X_1, \dots, X_p)$  de ses  $p$  lois marginales ne permet pas en général de connaître la loi de  $X$ . Donnons un contre-exemple.

**Contre-exemple 3.3.3.** — Une urne contient 4 boules numérotées de 1 à 4. Les deux premières sont blanches les deux dernières noires. Dans une première expérience on tire deux boules avec remise. L'univers  $\Omega_1$  est donc  $\{1, \dots, 4\}^2$  muni, faute de plus d'information, de la probabilité uniforme. On considère la variable aléatoire  $X = (X_1, X_2)$  à valeurs dans  $\{B, N\}^2$ , qui à un élément  $(x_1, x_2)$  de  $\Omega_1$  associe le couple  $(C_1, C_2)$  où  $C_i$  est la couleur de  $x_i$  pour  $i = 1, 2$ .



Ainsi  $X(1, 4) = (B, N)$ ,  $X(3, 3) = (N, N)$ . Les deux variables  $X_1$  et  $X_2$  suivent naturellement une loi de Bernoulli de paramètre  $\frac{1}{2}$ . En effet on a par exemple :

$$\mathbf{P}(X_1 = B) = \frac{|\{(B, B), (B, N)\}|}{|\Omega|} = \frac{2}{4} = \frac{1}{2}.$$

Dans une seconde expérience on prélève sans remise deux boules dans notre urne, l'univers  $\Omega_2$  est alors l'ensemble des couples de 2 éléments distincts de  $\{1, \dots, 4\}^1$ , toujours muni de la probabilité uniforme. Le cardinal de  $\Omega_2$  est le nombre d'arrangements à 2 éléments de  $\{1, \dots, 4\}$ , soit  $4 \times 3 = 12$ . On considère la variable aléatoire  $Y = (Y_1, Y_2)$  à valeurs dans  $\{B, N\}^2$ , qui à un élément  $(y_1, y_2)$  de  $\Omega_2$  associe comme dans la première expérience le couple des couleurs correspondantes.

La variable  $Y_1$  suit clairement une loi de Bernoulli de paramètre  $\frac{1}{2}$ . L'étude de  $Y_2$  est plus délicate. D'après 2.3.5,

$$\mathbf{P}(Y_2 = B) = \mathbf{P}(Y_2 = B | Y_1 = B) \mathbf{P}(Y_1 = B) + \mathbf{P}(Y_2 = B | Y_1 = N) \mathbf{P}(Y_1 = N) = \frac{1}{3} \times \frac{1}{2} + \frac{2}{3} \times \frac{1}{2} = \frac{1}{2}.$$

Ainsi,  $Y_2$  suit-elle aussi une loi de Bernoulli de paramètre  $\frac{1}{2}$ .

Les variables  $X$  et  $Y$  ont mêmes lois marginales. Mais

$$\mathbf{P}(X = (B, B)) = \frac{|\{(1, 1); (1, 2); (2, 1); (2, 2)\}|}{4^2} = \frac{4}{4^2} = \frac{1}{4},$$

tandis que  $\mathbf{P}(Y = (B, B)) = \frac{|\{(1, 2), (2, 1)\}|}{12} = \frac{1}{6}$ . Ainsi donc  $X$  et  $Y$  n'ont-elles pas la même loi.

Remarque : on a noté de la même façon par  $\mathbf{P}$  les probabilités sur  $\Omega_1$  et sur  $\Omega_2$  afin de ne pas alourdir les notations

La loi conjointe permet également de découvrir les lois conditionnelles de variables aléatoires.

**Proposition 3.3.4** — Soient  $X = (X_1, \dots, X_p)$  une variable aléatoire à valeurs dans  $E_1 \times \dots \times E_p$  et  $k$  élément de  $\{1, \dots, p\}$  et  $(A_1, \dots, A_p)$  un élément de  $(X_1(\Omega), \dots, X_p(\Omega))$  tel que :

$$\mathbf{P}(X_1 \in A_1, \dots, X_{k-1} \in A_{k-1}, X_{k+1} \in A_{k+1}, \dots, X_p \in A_p) \neq 0.$$

Alors

$$\mathbf{P}(X_k \in A_k | X_1 \in A_1, \dots, X_{k-1} \in A_{k-1}, X_{k+1} \in A_{k+1} \dots X_p \in A_p) = \frac{\mathbf{P}_X(A_1 \times \dots \times A_p)}{\mathbf{P}_X(A_1 \times \dots \times A_{k-1} \times E_k \times A_{k+1} \times \dots \times A_p)}.$$

pour tout  $A \in E_1$ .

*Preuve de la proposition 3.3.4.* —

Par définition  $\mathbf{P}(X_k \in A_k | X_1 \in A_1, \dots, X_{k-1} \in A_{k-1}, X_{k+1} \in A_{k+1} \dots X_p \in A_p)$  vaut :

$$\frac{\mathbf{P}\left((X_k \in A_k) \cap (X_1 \in A_1, \dots, X_{k-1} \in A_{k-1}, X_{k+1} \in A_{k+1} \dots X_p \in A_p)\right)}{\mathbf{P}(X_1 \in A_1, \dots, X_{k-1} \in A_{k-1}, X_{k+1} \in A_{k+1}, \dots, X_p \in A_p)}$$

D'une part,  $\mathbf{P}((X_k \in A_k) \cap (X_1 \in A_1, \dots, X_{k-1} \in A_{k-1}, X_{k+1} \in A_{k+1} \dots X_p \in A_p))$  vaut :

$$\mathbf{P}(X_1 \in A_1, \dots, X_{k-1} \in A_{k-1}, X_k \in A_k, X_{k+1} \in A_{k+1} \dots X_p \in A_p) = P_X(A_1, \dots, A_p).$$

D'autre part,  $\mathbf{P}(X_1 \in A_1, \dots, X_{k-1} \in A_{k-1}, X_{k+1} \in A_{k+1}, \dots, X_p \in A_p)$  vaut :

$$\mathbf{P}(X \in (A_1 \times \dots \times A_{k-1} \times E_k \times A_{k+1} \times \dots \times A_p)) = \mathbf{P}_X(A_1 \times \dots \times A_{k-1} \times E_k \times A_{k+1} \times \dots \times A_p).$$

D'où le résultat. □

### 3.4 Variables aléatoires indépendantes

Dans tout ce paragraphe on dispose d'un espace probabilisé  $(\Omega, \mathbf{P})$ . Toutes les variables aléatoires sont définies sur cet espace.

1. Contrairement à la situation de l'exercice 3.2.7., l'ordre de tirage à une importance pour la suite et l'on ne se contente pas de prendre pour issue de notre expérience une partie à 2 éléments

**Définition 3.4.1.** — Des variables aléatoires  $X$  et  $Y$  à valeurs respectivement dans des ensembles  $E_1$  et  $E_2$  sont dites indépendantes si, pour toute partie  $A$  de  $E_1$  et toute partie  $B$  de  $E_2$ ,

$$\mathbf{P}(X \in A, Y \in B) = \mathbf{P}(X \in A)\mathbf{P}(Y \in B).$$

autrement dit si les événements  $\{X \in A\}$  et  $\{Y \in B\}$  sont indépendants.

Plus généralement des variables aléatoires  $X_1, \dots, X_p$  à valeurs respectivement dans des ensembles  $E_1, \dots, E_p$  sont dites mutuellement indépendantes si, pour tout  $(A_1, \dots, A_p)$  tel que  $A_1 \subset E_1, \dots, A_p \subset E_p$ ,

$$\mathbf{P}(X_1 \in A_1, \dots, X_p \in A_p) = \mathbf{P}(X_1 \in A_1) \times \dots \times \mathbf{P}(X_p \in A_p).$$

autrement dit si les événements  $\{X_1 \in A_1\}, \dots, \{X_p \in A_p\}$  sont mutuellement indépendants.

Des variables aléatoires  $X_1, \dots, X_p$  à valeurs respectivement dans des ensembles  $E_1, \dots, E_p$  sont dites deux à deux indépendantes si, pour tout couple  $(i, j)$  d'éléments distincts de  $\{1, \dots, p\}$ ,  $X_i$  et  $X_j$  sont indépendantes.

Si avec les notations de la précédente définition,  $X_1, \dots, X_p$  sont mutuellement indépendantes alors elles sont deux à deux indépendantes, en effet considérons sans restreindre la généralité les variables  $X_1$  et  $X_2$ . Pour toute partie  $A_1$  de  $E_1$  et  $A_2$  de  $E_2$ ,

$$\mathbf{P}_{(X_1, X_2)}(A_1 \times A_2) = \mathbf{P}(X_1 \in A_1, X_2 \in A_2) = \mathbf{P}(X_1 \in A_1, X_2 \in A_2, X_3 \in E_3, \dots, X_p \in E_p)$$

Donc par mutuelle indépendance,

$$\mathbf{P}_{(X_1, X_2)}(A_1 \times A_2) = \mathbf{P}(X_1 \in A_1)\mathbf{P}(X_2 \in A_2)\mathbf{P}(X_3 \in E_3) \dots \mathbf{P}(X_p \in E_p) = \mathbf{P}(X_1 \in A_1)\mathbf{P}(X_2 \in A_2) \times 1^{p-2}$$

et finalement

$$\mathbf{P}_{(X_1, X_2)}(A_1 \times A_2) = \mathbf{P}(X_1 \in A_1)\mathbf{P}(X_2 \in A_2).$$

En revanche, on peut avoir  $X_1, \dots, X_p$ , pour  $p \geq 3$  deux à deux indépendantes sans être mutuellement indépendantes.

**Remarque** : Soient des variables aléatoires  $X_1, \dots, X_p$  à valeurs respectivement dans des ensembles  $E_1, \dots, E_p$ . On a vu que les lois des variables  $(X_1, \dots, X_p)$  ne permettent pas en général d'obtenir leur loi conjointe, cf. contre-exemple 3.3.3. Par contre si les variables sont indépendantes, la chose est possible, puisqu'en notant  $X = (X_1, \dots, X_p)$ ,

$$\mathbf{P}_X(A_1 \times \dots \times A_p) = \mathbf{P}(X_1 \in A_1, \dots, X_p \in A_p) = \mathbf{P}(X_1 \in A_1) \times \dots \times \mathbf{P}(X_p \in A_p) = \mathbf{P}_{X_1}(A_1) \times \dots \times \mathbf{P}_{X_p}(A_p).$$

C'est précisément la situation décrite dans la première expérience du contre exemple 3.3.3. Les variables  $X_1$  et  $X_2$  sont indépendantes, par exemple :

$$\mathbf{P}(X_1 \in \{B\}, X_2 \in \{N\}) = \frac{|\{(1, 3); (1, 4); (2, 3); (2, 4)\}|}{4^2} = \frac{1}{4},$$

tandis que

$$\mathbf{P}(X_1 \in \{B\}) \times \mathbf{P}(X_2 \in \{N\}) = \frac{|\{(1, i); (2, i)|i \in \{1, 4\}\}|}{4^2} \times \frac{|\{(i, 3); (i, 4)|i \in \{1, 4\}\}|}{4^2} = \frac{8}{16} \times \frac{8}{16} = \frac{1}{4}.$$

On étudie de même tous les autres cas. Ceci est intuitivement rassurant le second tirage n'est pas influencé par le résultat du premier. On aurait du reste pu modéliser l'expérience par un couple de deux variables aléatoires  $X'_1$  et  $X'_2$  indépendantes qui suivent une loi de Bernoulli (à valeurs dans  $\{B, N\}$ ).

Plus généralement lorsqu'une expérience est en fait la répétition de  $N$  expériences toutes identiques et sans influence les unes sur les autres, il est possible de la modéliser par un  $N$ -uplet de variables aléatoires mutuellement indépendantes suivant la même loi. Prenons un exemple.

On jette une pièce non nécessairement équilibrée  $N$  fois ( $N$  désigne un entier supérieur ou égal à 2). On pourrait définir pour univers  $\{P, F\}^N$  et définir une probabilité sur cette univers, dans le cas où la pièce est équilibrée, ce serait la probabilité uniforme, dans les autres cas c'est un peu plus ardu et il est plus simple de considérer  $N$  variables aléatoires  $(X_1, \dots, X_N)$  mutuellement indépendantes et qui suivent toute la même loi de Bernoulli (à valeurs dans  $\{P, F\}$ ) et dont le paramètre  $p$  dépend de la pièce. La  $i^e$  variable donne le résultat du  $i^e$  lancer. Ainsi

$$\mathbf{P}(X_1 = P, X_2 = P, \dots, X_N = P) = \mathbf{P}(X_1 = P) \times \dots \times \mathbf{P}(X_N = P)$$

donne la probabilité pour qu'à chaque coup la pièce donne pile.

En général on ne précise pas l'espace probabilisé sur lequel sont définies les variables  $X_1, \dots, X_N$  et l'on ne s'inquiète même pas de son existence. Toutefois il est possible de justifier cette négligence. Les plus curieux trouveront dans l'exercice suivant une présentation rigoureuse de cette façon de procéder.

**Exercice 3.4.2.** — Soit  $(E_1, \mathbf{P}_1), \dots, (E_n, \mathbf{P}_n)$  des espaces probabilisés. On note  $E = E_1 \times \dots \times E_n$  et pour  $i = 1, \dots, n$ ,  $X_i$  la  $i^e$  projection de  $E$ ,

$$X_i : E \rightarrow E_i; (x_1, \dots, x_n) \mapsto x_i.$$

1. Montrer qu'il existe une probabilité  $\mathbf{P}$  et une seule sur  $E$  telle que pour tout élément  $(x_1, \dots, x_N)$  de  $\mathbf{E}$ ,

$$\mathbf{P}(x_1, \dots, x_N) = \mathbf{P}_1(x_1)\mathbf{P}_2(x_2) \times \mathbf{P}_N(x_N).$$

2. Montrer que les variables  $X_i$ ,  $i = 1, \dots, N$  sont mutuellement indépendantes et que pour  $i = 1, \dots, N$  et tout élément  $A_i$  de  $E_i$ ,  $\mathbf{P}(X_i \in A_i) = \mathbf{P}_i(A_i)$ . pour  $i = 1, \dots, N$
3. Application montrer qu'il existe  $N$  variables de Bernoulli toutes de paramètre  $p$ , définies sur un même espace probabilisé et mutuellement indépendantes.

**Proposition 3.4.3.** — Soient  $X_1$  et  $X_2$  des variables aléatoires indépendantes à valeurs respectivement dans des ensembles  $E_1, E_2$  Soient par ailleurs, pour  $i = 1, 2$   $f_i$  une application de  $E_i$  dans une ensemble  $F_i$ . Alors les variables aléatoires  $f_1(X_1)$  et  $f_2(X_2)$  sont indépendantes.

*Preuve de la proposition 3.4.3* — Soient  $A_1 \in F_1$  et  $A_2 \in F_2$ , alors :

$$\mathbf{P}(f_1(X_1) \in A_1, f_2(X_2) \in A_2) = \mathbf{P}(X_1 \in f_1^{-1}(A_1), X_2 \in f_2^{-1}(A_2)),$$

donc par indépendance de  $X_1$  et  $X_2$ ,

$$\mathbf{P}(f_1(X_1) \in A_1, f_2(X_2) \in A_2) = \mathbf{P}(X_1 \in f_1^{-1}(A_1))\mathbf{P}(X_2 \in f_2^{-1}(A_2)) = \mathbf{P}(f_1(X_1) \in A_1)\mathbf{P}(f_2(X_2) \in A_2).$$

Donc les variables aléatoires  $f_1(X_1)$  et  $f_2(X_2)$  sont indépendantes. □

Plus généralement on peut prouver le résultat suivant :

*Proposition 3.4.4.* — Soient un entier  $N \geq 2$ ,  $X_1, X_2, \dots, X_N$  des variables aléatoires mutuellement indépendantes à valeurs respectivement dans des ensembles  $E_1, \dots, E_N$  et  $k$  un entier tel que  $1 \leq k \leq N-1$ . Si  $f$  est une application définie sur  $E_1 \times E_2 \cdots \times E_k$  et  $g$  une application définie sur  $E_{k+1} \times E_{k+2} \cdots \times E_N$  alors les variables  $f(X_1, \dots, X_k)$  et  $g(X_{k+1}, \dots, X_N)$  sont indépendantes.

Attention de ne pas généraliser trop hâtivement la proposition 3.4.4., comme nous le montre l'exercice suivant :

**Exercice 3.4.5.** — Soient deux variables aléatoires à valeurs dans  $\{-1, 1\}$ ,  $U$  et  $V$ , définies sur un même espace probabilisé  $(\Omega, \mathbf{P})$  indépendantes et de même loi :

$$\mathbf{P}_U(-1) = \mathbf{P}_V(-1) = \frac{1}{3}; \quad \mathbf{P}_U(1) = \mathbf{P}_V(1) = \frac{2}{3}.$$

On définit sur  $(\Omega, \mathcal{A}, \mathbf{P})$ , les variables aléatoires sur  $X$  et  $Y$  par :

$$X = U, Y = \text{sign}(U)V.$$

1. Quelle est la loi de la variable aléatoire  $(X, Y)$  ?
2. Les variables aléatoires  $X$  et  $Y$  sont elles indépendantes ?
3. Les variables  $X^2$  et  $Y^2$  sont elles indépendantes ?

*Réponses*

1. La loi de  $(X, Y)$  est entièrement définie par les valeurs de  $P(X = x, Y = y)$ , pour tout couple  $(x, y)$  d'éléments de  $\{-1, 1\}$ , données ci dessous :

	Y	-1	1
X	· · ·		
-1		$\frac{2}{9}$	$\frac{1}{9}$
1		$\frac{2}{9}$	$\frac{4}{9}$

2. Non, par exemple  $\mathbf{P}(X = 1)\mathbf{P}(Y = 1) = \frac{2}{3} \times \left(\frac{1}{3} \times \frac{1}{3} + \frac{2}{3} \times \frac{2}{3}\right) = \frac{10}{27} \neq \mathbf{P}(X = 1, Y = 1)$ .
3. Oui d'après 3.3.4, puisque  $X^2 = U^2$  et  $Y^2 = V^2$ .

Etudions à présent la somme de  $N$  variables aléatoires de Bernoulli indépendantes. Cette étude est d'une grande utilité pratique. Considérons par exemple un joueur qui lance  $N$  fois une pièce non nécessairement équilibrée. Nous avons vu qu'une telle expérience se modélise par  $N$  variables aléatoires indépendantes  $X_1, \dots, X_N$  qui suivent toutes une même loi de Bernoulli, en convenant par exemple que pile est représenté par 1 et face par 0. Alors en notant  $S_N = X_1 + X_2 + \dots + X_N$ ,  $\mathbf{P}(S_n) = k$  représente pour tout entier  $k$  la probabilité que le joueur fasse au cours de ses  $N$  lancers,  $k$  fois pile. La loi de  $S_n$  est remarquable :

**Proposition 3.4.6.** — Soient un entier  $N \geq 1$  et  $N$  variables aléatoires indépendantes sur  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $X_1, \dots, X_N$  qui suivent toutes la loi de Bernoulli de paramètre  $p$ , pour  $i = 1, \dots, N$ ,

$$X_i \sim \mathcal{B}(p).$$

Alors la variable aléatoire somme  $S_N$ , définie par  $S_N = X_1 + X_2 + \dots + X_N$  suit une loi binomiale de paramètres  $N$  et  $p$  :

$$S_N \sim \mathcal{B}(N, p).$$

*Preuve de la proposition 3.4.6.* —

Commençons par une preuve par récurrence sur  $N$ , c'est sans doute la preuve la plus simple, mais elle masque le sens du résultat.

Notons  $(P_N)$  la propriété à prouver.

- $(P_1)$  est vraie puisque, comme nous l'avons déjà dit la loi binomiale  $\mathcal{B}(1, p)$  n'est rien d'autre que la loi de Bernoulli de paramètre  $p$ .
- Soit un entier  $M \geq 1$  tel que  $(P_M)$  soit vraie. Prenons  $X_1, \dots, X_{M+1}$  des variables aléatoires indépendantes qui suivent toutes la loi de Bernoulli de paramètre  $p$  et posons :

$$S_{M+1} = X_1 + \dots + X_{M+1}, \quad S_M = X_1 + \dots + X_M.$$

Notons que  $S_{M+1}$  est à valeurs dans  $\{1, \dots, M+1\}$ , puisque les  $X_i, i = 1, \dots, M+1$  sont à valeurs dans  $\{0, 1\}$ . Soit alors  $k \in \{0, \dots, M+1\}$ .

— *Premier cas* :  $k \geq 1$ . Puisque  $(\{X_{M+1} = 0\}; \{X_{M+1} = 1\})$  est un système complet d'événements,

$$\mathbf{P}(S_{M+1} = k) = \mathbf{P}(S_{M+1} = k | X_{M+1} = 1) \mathbf{P}(X_{M+1} = 1) + \mathbf{P}(S_{M+1} = k | X_{M+1} = 0) \mathbf{P}(X_{M+1} = 0).$$

Soit

$$\mathbf{P}(S_{M+1} = k) = \mathbf{P}(S_M = k-1)p + \mathbf{P}(S_M = k)(1-p).$$

Donc d'après  $(P_M)$ ,

$$\mathbf{P}(S_{M+1} = k) = \binom{M}{k-1} p^{k-1} (1-p)^{M-k+1} p + \binom{M}{k} p^k (1-p)^{M-k} (1-p) = \left( \binom{M}{k-1} + \binom{M}{k} \right) p^k (1-p)^{M+1-k}$$

et finalement par la formule de Pascal,

$$\mathbf{P}(S_{M+1} = k) = \binom{M+1}{k} p^k (1-p)^{M+1-k}$$

— *Second cas* :  $k = 0$

$$\mathbf{P}(S_{M+1} = 0) = \mathbf{P}(X_1 = 0, X_2 = 0, \dots, X_{M+1} = 0).$$

L'indépendance mutuelle des variables  $X_1, \dots, X_{M+1}$  assure donc :

$$\mathbf{P}(S_{M+1} = 0) = (1-p)^{M+1} = \binom{M+1}{0} p^0 (1-p)^{M+1-0}.$$

De ces deux cas il vient que  $(P_{M+1})$  est vraie.

Ainsi a-t-on prouvé par récurrence la propriété  $(P_{M+1})$ .

Donnons à présent une preuve ayant du sens.

Soit  $k$  un élément de  $\{1, \dots, N\}$ . Désignons par  $\mathcal{P}_k$  l'ensemble des parties de  $\{1, \dots, N\}$  à  $k$  éléments.

$$\mathbf{P}(S_N = k) = \mathbf{P} \left( \bigcup_{J \in \mathcal{P}_k} \left( \bigcap_{j \in J} \{X_j = 1\} \cap \bigcap_{j \in \bar{J}} \{X_j = 0\} \right) \right).$$

Les ensembles  $A_J = \left( \bigcap_{j \in J} \{X_j = 1\} \cap \bigcap_{j \in \bar{J}} \{X_j = 0\} \right)$ ,  $J \in \mathcal{P}_k$  sont deux à deux disjoints, en effet  $\omega \in A_J$  si et seulement si  $X_j(\omega) = 1$ , pour  $j \in J$  et  $X_j(\omega) = 0$  sinon. Donc :

$$\mathbf{P}(S_N = k) = \sum_{J \in \mathcal{P}_k} \mathbf{P} \left( \bigcap_{j \in J} \{X_j = 1\} \cap \bigcap_{j \in \bar{J}} \{X_j = 0\} \right).$$

l'indépendance mutuelle des  $X_j$ ,  $j = 1, \dots, n$  assure alors que :

$$\mathbf{P}(S_N = k) = \sum_{J \in \mathcal{P}_k} \prod_{j \in J} \mathbf{P}(X_j = 1) \times \prod_{j \in \bar{J}} \mathbf{P}(X_j = 0) = \sum_{J \in \mathcal{P}_k} p^k (1-p)^{N-k}.$$

finalement :

$$\mathbf{P}(X = k) = \binom{N}{k} p^k (1-p)^{N-k}.$$

Pour ceux qui déploreraient la sècheresse de ce calcul donnons une approche moins rigoureuse mais plus intuitive. Pour une partie  $J$  à  $k$  éléments donnée la probabilité de  $\{X_j = 1\}$  pour tous les éléments de  $J$  et  $\{X_j = 0\}$  pour les autres est  $p^k (1-p)^{N-k}$ . Il y a  $\binom{N}{k}$  parties à  $k$  éléments donc :  $\mathbf{P}(S_N = k) = \binom{N}{k} p^k (1-p)^{N-k}$ .  $\square$

### 3.5 Moments

Dans toute ce paragraphe on considère un espace probabilisé fini  $(\Omega, \mathbf{P})$  sur lequel seront définies, sauf mention contraire, toutes les variables aléatoires.

#### ESPÉRANCE

Commençons par une approche heuristique. On effectue une expérience quelconque et l'on considère une grandeur réelle  $X$  qui dépend de l'issue de l'expérience, autrement dit  $X$  est une fonction définie sur l'ensemble des issues possibles, à valeurs dans une partie  $E$  de  $\mathbf{R}$ . On imagine que l'on effectue un grand nombre  $n$  de fois l'expérience et on fait la moyenne des valeurs prises par  $X$ , on trouve

$$\frac{1}{n} \sum_{\omega \in \Omega} \text{Nb}(\{\omega\}) X(\omega),$$

où  $\text{Nb}(\{\omega\})$  représente le nombre de fois où l'événement élémentaire  $\omega$  se produit. Notons qu'avec les notations de l'introduction,  $\frac{1}{n} \text{Nb}(\{\omega\})$  vaut  $f_n(\{\omega\})$ , fréquence de l'événement  $\{\omega\}$  si bien que la moyenne des valeurs prises par  $X$ , vaut :

$$\sum_{\omega \in \Omega} f_n(\{\omega\}) X(\omega).$$

D'après l'introduction de la probabilité en terme de fréquences, cette quantité tend avec  $n$  en un certain sens vers :

$$\sum_{\omega \in \Omega} \mathbf{P}(\{\omega\}) X(\omega).$$

Cette quantité qui mesure la valeur moyenne de  $X$  s'appellera espérance de  $X$ .

**Définition 3.5.1** — Soit  $X$  une variable aléatoire réelle définie sur  $(\Omega, \mathcal{A}, \mathbf{P})$ . On appelle espérance de  $X$  et l'on note  $E(X)$ , la quantité

$$E(X) = \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\}) X(\omega).$$

L'espérance est donc la moyenne des images par  $X$  des éléments  $\omega$  de  $\Omega$ , pondérée par la probabilité de l'événement  $\{\omega\}$ . On peut donner une autre forme de l'espérance.

**Proposition 3.5.2.** — Soit  $X$  une variable aléatoire définie sur  $(\Omega, \mathcal{A}, \mathbf{P})$ , à valeur dans une partie  $E$  de  $\mathbf{R}$ . Alors

$$E(X) = \sum_{x \in E} \mathbf{P}(X = x) x = \sum_{x \in X(\Omega)} \mathbf{P}(X = x) x.$$

L'espérance de  $X$  apparaît donc comme la moyenne des valeurs  $x$  prises par  $X$  pondérées par la probabilité que  $X$  prenne la valeur  $x$ , en effet  $\sum_{x \in X(\Omega)} \mathbf{P}(X = x) = \sum_{x \in X(\Omega)} \mathbf{P}_X(\{x\}) = 1$ .

*Preuve de la proposition 3.5.2* — Que  $\sum_{x \in E} \mathbf{P}(X = x) x = \sum_{x \in X(\Omega)} \mathbf{P}(X = x) x$  résulte de ce que  $\mathbf{P}(X = x) = 0$ , pour tout élément  $x$  de  $E$  qui n'est pas élément de  $X(\Omega)$ . On notera au passage que  $\Omega$  étant fini, les sommes de la formule sont bien définies, la première comme somme d'un nombre fini de termes non nuls, l'autre comme somme finie. L'ensemble  $\{X = x\}_{x \in X(\Omega)}$  est une partition de  $\Omega$ . Donc :

$$E(X) = \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\}) X(\omega) = \sum_{x \in X(\Omega)} \sum_{\omega \in \{X=x\}} \mathbf{P}(\{\omega\}) X(\omega),$$

Soit

$$E(X) = \sum_{x \in X(\Omega)} \sum_{\omega \in \{X=x\}} \mathbf{P}(\{\omega\})x = \sum_{x \in X(\Omega)} \left( \mathbf{P} \left( \bigcup_{\omega \in \{X=x\}} \{\omega\} \right) \right) x,$$

puisque  $(\{\omega\})_{\omega \in \{X=x\}}$  est une famille d'éléments deux à deux incompatibles, leur réunion étant de plus  $\{X=x\}$ ,

$$E(X) = \sum_{x \in X(\Omega)} \mathbf{P}(X=x) x.$$

Voici la formule prouvée. □

**Exemple 3.5.3.** — Un jeu consiste à miser un franc sur un des numéros 1, 2, 3, 4, 5 et 6. On lance alors un dé et si le numéro sur lequel on a misé sort, le joueur empoche 5 francs, sinon il ne gagne rien, dans tous les cas il laisse sa mise. Modélisons le jeu et le gain que peut espérer le joueur. L'univers est  $\{1, \dots, 6\}$  que l'on munira de la probabilité uniforme. Le gain est représenté par une variable aléatoire  $G$  défini comme suit. On note  $k$  le numéro sur lequel le joueur a misé, alors pour tout  $x \in \Omega$ ,  $G(x) = -1$  si  $x \neq k$ , et  $G(k) = 5 - 1 = 4$ . L'espérance de  $G$  vaut donc :

$$E(G) = \mathbf{P}(G = -1)(-1) + \mathbf{P}(G = 4)4 = \mathbf{P}(\{1, \dots, 6\} \setminus \{k\})(-1) + \mathbf{P}(\{k\})4 = \frac{5}{6}(-1) + \frac{1}{6}4 = -\frac{1}{6}.$$

En conclusion, puisque l'espérance est négative, si l'on a foi en les probabilités, on peut s'attendre à ce qu'un joueur qui jouerait un grand nombre de parties, perde de l'argent (en gros  $\frac{1}{6}$  de franc).

**Remarque** — L'intérêt de la formule de la proposition 3.4.2. est qu'il est nul besoin de connaître l'espace probabilisé  $(\Omega, \mathbf{P})$  pour calculer l'espérance de  $X$ , seule la loi de  $X$  est nécessaire. En fait l'espérance de  $X$  ne dépend que de sa loi.

Donnons donc l'espérance d'une variables aléatoire réelle  $X$  lorsqu'elle obéit à des lois connues.

- **Variable constante** : la variable  $X$  prend une valeur constante  $c$ , donc  $\mathbf{P}(X=x) = \delta_{x,c}$ , pour tout réel  $x$ .

$$E(x) = 1 \cdot c = c.$$

- **Loi uniforme** : on suppose que  $X$  suit la loi uniforme sur  $\{1, \dots, n\}$ .

$$E(X) = \sum_{k=1}^n \mathbf{P}(X=k)k = \sum_{k=1}^n \frac{1}{n}k,$$

donc

si  $X \sim \mathcal{U}(\{1, \dots, n\})$  alors  $E = \frac{n+1}{2}$

- **Loi de Bernoulli** : on suppose que  $X$  suit la loi de Bernoulli de paramètre  $p$ .

$$E(X) = \mathbf{P}(X=1)1 + \mathbf{P}(X=0)0 = p \times 1 + (1-p) \times 0,$$

Donc

$$\text{si } X \sim \mathcal{B}(p) \text{ alors } E = p$$

- **Loi binomiale** : on suppose que  $X$  suit la loi de Binomiale de paramètre  $(n, p)$ . Le calcul direct est assez délicat, nous allons donner dans la suite une façon rapide de trouver l'espérance, pour lors, nous développons néanmoins une méthode de calcul

$$E(X) = \sum_{k=0}^n \mathbf{P}(X=k)k = \sum_{k=1}^n \binom{n}{k} p^k (1-p)^{n-k} k.$$

Par ailleurs on a l'égalité dans  $\mathbf{R}[T]$  :

$$\sum_{k=0}^n \binom{n}{k} T^k (1-p)^{n-k} = (T + (1-p))^n.$$

Par dérivation formelle :

$$\sum_{k=1}^n \binom{n}{k} k T^{k-1} (1-p)^{n-k} = n(T + (1-p))^{n-1}.$$

Donc en substituant à l'indéterminé  $T$  dans l'égalité précédente le réel  $p$  et en multipliant par  $p$ , on a :

$$\sum_{k=1}^n \binom{n}{k} k p^k (1-p)^{n-k} = p n (p + (1-p))^{n-1},$$

et donc

si  $X \sim \mathcal{B}(n, p)$ . alors  $E = pn$

Passons aux propriétés de l'espérance.

**Proposition 3.5.4.** — Soient  $X$  et  $Y$  des variables aléatoires réelles définies sur  $(\Omega, \mathcal{A}, \mathbf{P})$ , et  $(a, b)$  un couple de réels. L'espérance jouit des propriétés suivantes :

1. Linéarité :  $E(aX + bY) = aE(X) + bE(Y)$ , autrement dit  $E$  est une application linéaire de l'espace vectoriel des variables aléatoires réelles définies sur  $(\Omega, \mathcal{A}, \mathbf{P})$ , dans  $\mathbf{R}$ .
2. Positivité : si  $X$  est à valeurs positives, alors  $E(X) \geq 0$ .
3. Croissance : si  $X \geq Y$ , alors  $E(X) \geq E(Y)$ .

*Preuve de la proposition 3.5.4* —

1.

$$E(aX + bY) = \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\})(aX + bY)(\omega) = \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\})(aX(\omega) + bY(\omega))$$

Donc

$$E(aX + bY) = a \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\})X(\omega) + b \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\})Y(\omega) = aE(X) + bE(Y).$$

2. Supposons  $X \geq 0$ .

$$E(X) = \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\})X(\omega).$$

Donc  $E(X) \geq 0$ , puisque tous les termes de la somme précédente sont positifs.

3. Comme  $X - Y \geq 0$ , le deuxième point dit que  $E(X - Y) \geq 0$ , ce qui compte tenu du premier s'écrit :  $E(X) - E(Y) \geq 0$ .

□

**Application** — La proposition précédente redonne l'espérance d'une variable qui suit une loi binomiale. En effet si  $X_1, \dots, X_n$  sont des variables aléatoires sur  $(\Omega, \mathcal{A}, \mathbf{P})$  qui suivent toutes la loi de Bernoulli de paramètre  $p$ , alors on a vu que  $X_1 = X_2 + \dots + X_n$  suit la loi binomiale de paramètre  $(n, p)$ , cf. 3.3.6. Mais  $E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) = np$  d'après le calcul déjà fait de l'espérance d'une variable de Bernoulli.

**Exercice 3.5.5.** — Généralisons le troisième point de 3.5.4. Soient  $X$  et  $Y$  des variables aléatoires réelles définies sur  $(\Omega, \mathcal{A}, \mathbf{P})$ . On suppose que  $\mathbf{P}(X \geq Y) = 1$ . Montrer que  $E(X) \geq E(Y)$ .

*Solution de l'exercice 3.5.5.* — Notons  $\Omega' = \{X \geq Y\}$ . Par hypothèse  $\mathbf{P}(\Omega') = 1$ , donc  $\mathbf{P}(\bar{\Omega}') = 0$  et donc, pour tout  $\omega \in \bar{\Omega}'$ ,  $\mathbf{P}(\{\omega\}) = 0$ . Par ailleurs

$$E(X) = \sum_{\omega \in \Omega'} X(\omega)\mathbf{P}(\{\omega\}) + \sum_{\omega \in \bar{\Omega}'} X(\omega)\mathbf{P}(\{\omega\}) = \sum_{\omega \in \Omega'} X(\omega)\mathbf{P}(\{\omega\})$$

Donc

$$E(X) \geq \sum_{\omega \in \Omega'} Y(\omega)\mathbf{P}(\{\omega\}) = \sum_{\omega \in \Omega'} Y(\omega)\mathbf{P}(\{\omega\}) + \sum_{\omega \in \bar{\Omega}'} Y(\omega)\mathbf{P}(\{\omega\}) = E(Y).$$

**Définition-proposition 3.5.6.** — Une variable aléatoire réelle  $X_0$  est dite centrée si par définition :  $E(X_0) = 0$ .

Si  $X$  est une variable aléatoire réelle, alors la variable aléatoire  $X - E(X)$  est centrée.

**Preuve de 3.5.6.** — Résulte immédiatement de la linéarité de l'espérance et du calcul de l'espérance d'une variable constante. □

On étudie maintenant l'espérance d'un produit de composition.

**Proposition 3.5.7.** — FORMULE DE TRANSFERT —

Soit  $X$  une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbf{P})$  à valeurs dans un espace  $E$  et  $f$  une application de  $X(\Omega)$  à valeurs réelles. Alors

$$E(f(X)) = \sum_{x \in X(\Omega)} f(x)\mathbf{P}(X = x) = \sum_{x \in X(\Omega)} f(x)\mathbf{P}_X(\{x\})$$

Autrement dit,  $E(f(X))$  est l'espérance de  $f$  sur l'espace probabilisé fini  $(X(\Omega), \mathbf{P}_X)$ .

**Remarque :** on peut écrire encore avec quelques abus,  $E(f(X)) = \sum_{x \in E} f(x)\mathbf{P}(X = x)$ , puisque pour tout  $x \in E$  si  $x \notin X(\omega)$ , alors  $\mathbf{P}(X = x) = 0$ , ( $f$  est alors prolongée à  $E$  de quelconque manière).

*Preuve de la proposition 3.5.7.* —

L'ensemble  $\{X = x\}_{x \in X(\Omega)}$  est une partition de  $\Omega$ . Donc :

$$E(f(X)) = \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\}) f(X(\omega)) = \sum_{x \in X(\Omega)} \sum_{\omega \in \{X=x\}} \mathbf{P}(\{\omega\}) f(X(\omega)),$$

Soit

$$E(f(X)) = \sum_{x \in X(\Omega)} \sum_{\omega \in \{X=x\}} \mathbf{P}(\{\omega\}) f(x) = \sum_{x \in X(\Omega)} \left( \mathbf{P} \left( \bigcup_{\omega \in \{X=x\}} \{\omega\} \right) \right) f(x),$$

puisque  $(\{\omega\})_{\omega \in \{X=x\}}$  est une famille d'éléments deux à deux incompatibles, leur réunion étant de plus  $\{X = x\}$ ,

$$E(f(X)) = \sum_{x \in X(\Omega)} \mathbf{P}(X = x) f(x) = \sum_{x \in X(\Omega)} \mathbf{P}_X(\{x\}) f(x).$$

Voici la formule prouvée. □

Etudions à présent le comportement de l'espérance vis à vis du produit.

**Proposition 3.5.8.** — *soient  $X_1, X_2, \dots, X_p$  des variables aléatoires réelles mutuellement indépendantes définies sur  $(\Omega, \mathcal{A}, \mathbf{P})$ . Alors :*

$$E(X_1 \times X_2 \times \dots \times X_p) = E(X_1) \times E(X_2) \times \dots \times E(X_p).$$

*Preuve de la proposition 3.5.8.* — Posons  $X = (X_1, X_2, \dots, X_p)$  et  $Z = X_1 X_2 \dots X_p$ , notons que  $Z = f(X)$ , où  $f : \mathbf{R}^p \rightarrow \mathbf{R}; (x_1, \dots, x_p) \mapsto x_1 \times \dots \times x_p$ . Par la remarque qui suit 3.5.7.,

$$E(Z) = \sum_{(x_1, x_2, \dots, x_p) \in \mathbf{R}^p} \mathbf{P}(X = (x_1, x_2, \dots, x_p)) f(x_1, x_2, \dots, x_p).$$

Donc compte tenu de la mutuelle indépendance de  $X_1, \dots, X_p$ ,

$$E(Z) = \sum_{(x_1, x_2, \dots, x_p) \in \mathbf{R}^p} \mathbf{P}(X = x_1) \mathbf{P}(X = x_2) \dots \mathbf{P}(X = x_p) x_1 x_2 \dots x_p.$$

Donc

$$E(Z) = \sum_{x_1 \in \mathbf{R}} \mathbf{P}(X = x_1) x_1 \sum_{x_2 \in \mathbf{R}} \mathbf{P}(X = x_2) x_2 \dots \sum_{x_p \in \mathbf{R}} \mathbf{P}(X = x_p) x_p.$$

Finalement :  $E(Z) = E(X_1) E(X_2) \dots E(X_p)$ . □

**Attention !**

- le résultat 3.5.8. est faux sans l'hypothèse de mutuelle indépendance,
- que l'espérance du produit de variables aléatoires soit le produit de leurs espérances ne prouve pas la mutuelle indépendance de ces variables.

Lorsque une variable aléatoire  $X$  prend des valeurs entières il est souvent plus aisé de déterminer  $\mathbf{P}(X \geq x)$  que  $\mathbf{P}(X = x)$ , l'expression de  $\mathbf{P}(X = x)$  qui s'écrit  $\mathbf{P}(X = x) = \mathbf{P}(X \geq x) - \mathbf{P}(X \geq x + 1)$  prend une forme souvent compliquée. Qu'importe l'exercice qui suit donne l'expression de la variance en fonction de  $\mathbf{P}(X \geq x)$ .

**Exercice** — Soit  $X$  une variable aléatoire définie sur  $\Omega$  à valeurs dans  $\{0, n\}$ .

Montrer que :  $E(X) = \sum_{x=1}^n \mathbf{P}(X \geq x)$ .

*Solution* — Pour tout élément  $x$  de  $\{0, \dots, n\}$ ,  $\{X \geq x\}$  est la réunion disjointe de  $\{X \geq x + 1\}$  et de  $\{X = x\}$ , si bien que :

$$\mathbf{P}(X = x) = \mathbf{P}(X \geq x) - \mathbf{P}(X \geq x + 1)$$

Donc

$$E(X) = \sum_{x=0}^n x \mathbf{P}(X = x) = \sum_{x=0}^n x (\mathbf{P}(X \geq x) - \mathbf{P}(X \geq x + 1)) = \sum_{x=0}^n x \mathbf{P}(X \geq x) - \sum_{x=0}^n x \mathbf{P}(X \geq x + 1)$$

Grâce au changement d'indice de sommation «  $y = x + 1$  » dans l'ultime somme, on obtient donc :

$$E(X) = \sum_{x=0}^n x \mathbf{P}(X \geq x) - \sum_{y=1}^{n+1} (y - 1) \mathbf{P}(X \geq y) = .$$



Mais  $X$  étant à valeur dans  $\{1, \dots, n\}$ ,  $\mathbf{P}(X \geq n+1) = 0$ , donc

$$\mathbf{E}(X) = \sum_{x=1}^n x \mathbf{P}(X \geq x) - \sum_{x=1}^n (x-1) \mathbf{P}(X \geq x) = \sum_{x=1}^n \mathbf{P}(X \geq x).$$

### ÉCART TYPE

L'espérance d'une variable aléatoire, donne on l'a vu une idée de la valeur moyenne des valeurs prises par  $X$  sur un grand nombre d'expériences aléatoires. Par contre elle ne donne aucune information sur les fluctuations de ces valeurs autour de cette moyenne. Pour mesurer l'écart moyenne entre les valeurs prises par  $X$  et  $\mathbf{E}(X)$ , il est d'abord naturel de considérer  $\mathbf{E}(|X - \mathbf{E}(X)|)$ . Mais cette quantité est peu propice aux calculs et l'on préfère introduire une quantité, appelée variance de  $X$ .

**Définition 3.5.9.** — VARIANCE — Soit  $X$  une variable aléatoire réelle. On appelle variance de  $X$  et l'on note  $V(X)$ , la quantité :

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

La variance est donc donnée par les formules :

$$V(X) = \sum_{\omega \in \Omega} (X(\omega) - \mathbf{E}(X))^2 \mathbf{P}(\{\omega\}); \quad V(x) = \sum_{x \in X(\omega) \text{ (où } E)} (x - \mathbf{E}(X))^2 \mathbf{P}(X = x)$$

$\mathbf{E}((X - \mathbf{E}(X))^2)$  est en quelque sorte la moyenne des écarts au carré entre  $X(\omega)$  et  $\mathbf{E}(X)$ , pondérée par la probabilité de  $\{\omega\}$ . Ce faisant, on diminue par rapport  $\mathbf{E}(|X - \mathbf{E}(X)|)$  l'importance des petits écarts (inférieurs à 1) et l'on augmente celle des plus importants, mais ceci n'a que peu d'importance face au gain en matière de calcul et à la richesse mathématique de cette notion.

Donnons quelques propriétés de la variance utiles à son calcul pratique.

**Proposition 3.5.10.** — Soient  $X$  une variable aléatoire réelle,  $a$  et  $b$  des réels, alors :

$$V(aX + b) = a^2 V(X).$$

*Preuve de la proposition 3.5.10* — Par linéarité de l'espérance et le fait que l'espérance d'une variable constante est précisément cette constante,

$$V(aX + b) = \mathbf{E}((aX + b - \mathbf{E}(aX + b))^2) = \mathbf{E}((aX + b - (a\mathbf{E}(X) + b))^2) = \mathbf{E}(a^2(X - \mathbf{E}(X))^2) = a^2 \mathbf{E}((X - \mathbf{E}(X))^2) = a^2 V(X).$$

□

La variance est donc *homogène de degré 2* il en résulte que si  $X$  représente une grandeur physique, des mètres par exemple  $V(X)$  ne représente pas la même grandeur, mais son carré. C'est pourquoi, on définit un nouveau indicateur de dispersion, l'*écart type*, qui est la racine carrée de la variance.

**Définition 3.5.11** — Soit  $X$  une variable aléatoire réelle. On appelle écart-type  $X$  et l'on note  $\sigma(X)$ , la quantité :

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

**Définition, proposition 3.5.12.** — Soit  $X_0$  une variable aléatoire réelle. On dit que  $X_0$  est réduite si  $V(X_0) = 1$ . Si  $X$  est une variable aléatoire réelle d'écart-type non nul, alors  $\frac{X - \mathbf{E}(X)}{\sigma(X)}$  est une variable centrée et réduite.

*Preuve de la proposition 3.5.12* — Résulte directement de la linéarité de l'espérance et de 3.4.10. □

La formule qui suit est comme nous le verrons utile pour calculer la variance dans de nombreux cas.

**Proposition 3.5.13.** — FORMULE DE KÖNIG-HUYGHENS — Soit  $X$  une variable aléatoire réelle. Alors :

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

*Preuve de 3.5.13.* —

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

Donc par linéarité de l'espérance, et le fait que l'espérance d'une variable constante est cette constante,

$$V(X) = E(X^2) - 2E(x)E(X) + E(X)^2 = E(X^2) - E(X)^2.$$

□

On peut utiliser ce résultat dans le calcul la variance d'une variables aléatoire réelle  $X$  lorsqu'elle obéit à des lois connues.

- **Variable constante** : la variable  $X$  prend une valeur constante  $c$ ,

$$V(X) = E(X^2) - (E(X))^2 = c^2 - c^2 = 0$$

- **Loi de Bernoulli** :

$$V(X) = E(X^2) - (E(X))^2 = E(X) - E(X)^2 = p - p^2$$

Donc

$$\text{si } X \sim \mathcal{B}(p). \text{ alors } V(X) = p(1 - p)$$

La formule de König-Huyghens assure en particulier que  $E(X^2) \geq E(X)^2$ . L'exercice suivant généralise ce résultat.

**Exercice 3.5.14.** — Soient  $X$  et  $Y$  des variables aléatoires. Montrer que :

$$E(XY)^2 \leq E(X)^2 E(Y)^2$$

On retrouve  $E(X^2) \geq E(X)^2$  en choisissant  $Y$  constante égale à 1

*Indication* : utiliser l'inégalité de Cauchy-Schwarz.

On peut aussi pour les calculs de variance utiliser le résultat suivant.

**Proposition 3.5.15.** — Soient  $X$  et  $Y$  des variables aléatoires réelles. Si elle sont indépendantes, alors :

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

*Preuve de la proposition 3.5.15.* — Là encore, 3.5.13. et la linéarité de l'espérance nous sauve :

$$\begin{aligned} V(X+Y) &= E((X+Y)^2) - E(X+Y)^2 = E(X^2 + Y^2 + 2XY) - (E(X) + E(Y))^2 = \\ &= E(X^2) + E(Y^2) + 2E(XY) - E(X)^2 - E(Y)^2 - 2E(X)E(Y). \end{aligned}$$

Donc, toujours 3.5.13,

$$V(X+Y) = V(X) + V(Y) + 2(E(XY) - E(X)E(Y)) \quad (\text{I.5})$$

Mais si  $X$  et  $Y$  sont indépendantes, alors, d'après 3.4.7. :

$$V(X+Y) = V(X) + V(Y) + 2(E(X)E(Y) - E(X)E(Y)) = V(X) + V(Y).$$

□

On déduit de ce résultat la variance d'une loi binomiale.

**Loi binomiale** — La somme de  $n$  variables de Bernoulli indépendantes de paramètre  $p$  suit une loi binomiale de paramètre  $(n, p)$  donc sa variance est par applications itératives de 3.5.15.,  $n(p(1-p))$ .

$$\boxed{\text{si } X \sim \mathcal{B}(n, p), \text{ alors } V(X) = np(1-p)}$$

#### INÉGALITÉS DE MARKOV ET BIENAYMÉ-TCHEBYCHEV

Nous allons étudier la probabilité qu'une variable s'écarte de son espérance.

On appelle inégalité de Markov diverses inégalités qui reposent sur un raisonnement simple à retenir.

Prenons  $X$  une variable aléatoire réelle et  $h$  une application de  $\mathbf{R}$  dans  $\mathbf{R}$  à valeurs positives. Evaluons l'espérance de  $h(X)$ . Soient un réel  $a > 0$  et  $\Omega' := \{h(X) \geq a\}$ .

$$E(h(X)) = \sum_{\omega \in \Omega} h(X(\omega))\mathbf{P}(\{\omega\}) = \sum_{\omega \in \Omega'} h(X(\omega))\mathbf{P}(\{\omega\}) + \sum_{\omega \in \bar{\Omega}'} h(X(\omega))\mathbf{P}(\{\omega\}).$$

Comme  $h$  est à valeur positive, on a donc

$$E(h(X)) \geq \sum_{\omega \in \Omega'} h(X(\omega))\mathbf{P}(\{\omega\}) \geq \sum_{\omega \in \Omega'} a\mathbf{P}(\{\omega\}) = a\mathbf{P}(\Omega')$$

finalement

$$E(h(X)) \geq a\mathbf{P}(h(X) \geq a) \quad (\text{I.6})$$

On en déduit immédiatement pour  $h = |\cdot|$ , l'inégalité de Markov classique :

$$\boxed{\mathbf{P}(|X| \geq a) \leq \frac{E(|X|)}{a}} \quad (\text{inégalité de Markov})$$

Cette égalité a surtout une importance théorique et intervient dans la preuve de la loi faible des grands nombres.

Par le même genre de techniques on traitera l'exercice suivant.

**Exercice 3.5.16.** — Soit  $X$  une variable aléatoire réelle

1. Soit  $g$  une application de  $\mathbf{R}_+$  dans  $\mathbf{R}_+$ , strictement croissante. Montrer que pour tout réel  $a > 0$ ,

$$\mathbf{P}(|X| \geq a) \leq \frac{E(g(|X|))}{g(a)}$$

2. Soient un réel  $\alpha > 0$  et une application  $h : \mathbf{R} \rightarrow [0, \alpha]$ . Montrer que pour tout réel  $a$  tel que  $0 \leq a < \alpha$ , on a

$$\mathbf{P}(h(X) \geq a) \geq \frac{E(h(X)) - a}{\alpha - a}.$$

L'inégalité de Bienaymé-Tchebychev montre bien comment la variance contrôle l'écart d'une variable à son espérance. Soit un réel  $a > 0$ . Appliquons (I.6) à la variable aléatoire  $X - E(X)$  avec  $h : x \mapsto x^2$  :

$$\mathbf{P}((X - E(X))^2 \geq a^2) \leq \frac{E((X - E(X))^2)}{a^2}.$$

Mais comme  $\{|X - E(X)| \geq a\} = \{(X - E(X))^2 \geq a^2\}$ , on obtient la formule suivante :

$$\boxed{\mathbf{P}(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}} \quad (\text{inégalité de Bienaymé-Tchebychev})$$

## COVARIANCE, CORRÉLATION

On cherche à présent à mesurer la « ressemblance » entre deux variables aléatoires  $X$  et  $Y$ . Plus exactement étant donné  $X$  et  $Y$  variables aléatoires on cherche à comparer les variables centrées réduites  $X_0 = \frac{X - \mathbf{E}(X)}{\sigma(X)}$  et  $Y_0 = \frac{Y - \mathbf{E}(Y)}{\sigma(Y)}$ , ce afin de comparer la dispersion de  $X$  et  $Y$  autour de leurs espérances respectives en les ramenant à des quantités adimensionnées.

Pour ce faire remarquons que, dans le cas où aucun événement élémentaire n'est de probabilité nulle,

$$\mathbf{R}^\Omega \times \mathbf{R}^\Omega : (X_1, X_2) \mapsto E(X_1 X_2) = \sum_{\omega \in \Omega} X_1(\omega) X_2(\omega) \mathbf{P}\{\omega\}$$

est un produit scalaire, noté  $\langle \cdot | \cdot \rangle$ , sur l'espace vectoriel des variables aléatoires réelles sur  $(\Omega, \mathcal{A}, \mathbf{P})$ , la variance d'une variable aléatoire  $Z$  n'est alors que la norme au carré de  $Z - \mathbf{E}(Z)$ . Si des événements élémentaires  $\{\omega\}$  sont de probabilité nulle ce n'est plus un produit scalaire mais une forme bilinéaire symétrique positive.

L'inégalité de Cauchy-Schwarz (valable dans tous les cas) donne alors

$$|E(X_0 Y_0)| \leq \sqrt{V(X_0) V(Y_0)} = 1$$

Plus précisément en nous plaçant dans le cas où aucun événement élémentaire n'est de probabilité nulle,  $X_0$  et  $Y_0$  sont normés,  $-1 \leq E(X_0 Y_0) \leq 1$  et d'après la condition d'égalité dans l'inégalité de Cauchy-Schwarz  $E(X_0 Y_0) = 1$  si et seulement si  $X_0 = Y_0$ ,  $E(X_0 Y_0) = -1$  si et seulement si  $X_0 = -Y_0$ ,  $E(X_0 Y_0) = 0$  si et seulement si  $X_0$  est orthogonal à  $Y_0$ .  $E(X_0 Y_0)$  mesure donc la ressemblance entre  $X_0$  et  $Y_0$  si cette quantité est proche de 1 alors  $X_0$  et  $Y_0$  se ressemblent, si elle est proche de  $-1$  alors  $X_0$  est proche de  $-Y_0$  et enfin, si cette quantité est voisine de 0, alors  $X_0$  et  $Y_0$  sont sans rapport.

Notons que  $E(X_0 Y_0) = \frac{E((X - \mathbf{E}(X))(Y - \mathbf{E}(Y)))}{\sigma(X)\sigma(Y)}$  et adoptons les définitions suivantes :

**Définition, proposition 3.5.17** — Soient  $X$  et  $Y$  des variables aléatoires réelles sur  $(\Omega, \mathcal{A}, \mathbf{P})$ . On appelle covariance de  $X$  et  $Y$  et l'on note  $\text{Cov}(X, Y)$  la quantité

$$\text{Cov}(X, Y) = E((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))),$$

lorsque  $\sigma(X)\sigma(Y) \neq 0$ , on appelle coefficient de corrélation de  $X$  et  $Y$  et l'on note  $\rho(X, Y)$  la quantité

$$\rho(X, Y) = \frac{E((X - \mathbf{E}(X))(Y - \mathbf{E}(Y)))}{\sigma(X)\sigma(Y)}$$

Le coefficient de corrélation de  $X$  et  $Y$  est élément de  $[-1, 1]$ .

On dit que  $X$  et  $Y$  sont non corrélées si  $\text{Cov}(X, Y) = 0$ , ou si l'on préfère, lorsque  $\sigma(X)\sigma(Y) \neq 0$ , si  $\rho(X, Y) = 0$ .

L'interprétation en terme de forme bilinéaire symétrique de la covariance, assure les propriétés suivantes :

$$\boxed{\text{Cov}(X, X) = V(X)}$$

$$\boxed{V(X + Y) = V(X) + 2\text{Cov}(X, Y) + V(Y); \text{Cov}(X, Y) = \frac{1}{2}(V(X + Y) - V(X) - V(Y))}$$

On déduit de ces formules l'expression suivante de la covariance qui généralise 3.5.13.

**Proposition 3.5.18.** — Soient  $X$  et  $Y$  des variables aléatoires réelles sur  $(\Omega, \mathcal{A}, \mathbf{P})$ . Alors

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

*Preuve de 3.5.18.* —

$$\text{Cov}(X, Y) = \frac{1}{2}(V(X + Y) - V(X) - V(Y)),$$

donc compte tenu de 3.5.13,

$$\frac{1}{2}(E((X + Y)^2) - E(X + Y)^2 - E(X^2) + E(X)^2 - E(Y^2) + E(Y)^2) = E(XY) - E(X)E(Y).$$



Passons au lien entre non corrélation et indépendance.

**Proposition 3.5.19.** — Soient  $X$  et  $Y$  des variables aléatoires réelles sur  $(\Omega, \mathcal{A}, \mathbf{P})$ . Si  $X$  et  $Y$  sont indépendantes, alors  $X$  et  $Y$  sont non corrélées.

*Preuve de la proposition 3.5.19.* — Résulte directement de l'expression de la covariance de 3.5.18. et de 3.5.8.  $\square$

La réciproque est fautive comme le montre le contre-exemple suivant.

**Contre-exemple 3.5.20.** — On prend pour espace probabilisé  $\{-1, 0, 1\}$  muni de la probabilité uniforme. Soient les variables aléatoires  $X$  qui est l'identité, et  $Y$  définie par  $Y = 1$ , si  $X = 0$ , 0 sinon. On a  $X \sim \mathcal{U}(\{-1, 0, 1\})$ ,  $XY$  nulle et  $\mathbf{E}(X) = 0$  donc :

$$\text{Cov}(XY) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = 0 - 0 = 0.$$

Les variables  $X$  et  $Y$  sont non corrélées.

Mais  $\mathbf{P}(X = 1, Y = 1) = 0$  puisque  $XY = 0$ , et  $\mathbf{P}(X = 1) = \frac{1}{3}$  et  $\mathbf{P}(Y = 1) = \frac{1}{3}$ , ainsi :

$$\mathbf{P}(X = 1, Y = 1) \neq \mathbf{P}(X = 1)\mathbf{P}(Y = 1),$$

et donc  $X$  et  $Y$  ne sont pas indépendantes.

#### MOMENTS

L'intérêt des moments apparaîtra dans le cours de spé. lorsque  $\Omega$  ne sera plus fini.

Brutalement, la définition :

**Définition 3.5.21.** — Soient  $X$  une variable aléatoire réelle sur  $(\Omega, \mathcal{A}, \mathbf{P})$  et un entier  $k \geq 0$ . On appelle moment d'ordre  $k$  de  $X$  la quantité  $\mathbf{E}(X^k)$ .

Ainsi, l'espérance de  $X$  est son moment d'ordre 0, sa variance le moment d'ordre 2 de  $X - \mathbf{E}(X)$ .

En particulier le moment d'ordre 2 de  $X$  vaut :

$$\sum_{\omega \in \Omega} X(\omega)^2 \mathbf{P}(\omega)$$

On notera l'analogie avec le moment<sup>1</sup> d'inertie  $\mathcal{M}$  par rapport à un point  $O$  de  $n$  points matériels  $\omega_1, \dots, \omega_n$  de masses respectives  $\mathbf{M}(x_1), \dots, \mathbf{M}(x_n)$ , distants respectivement de  $X(\omega_1), \dots, X(\omega_n)$  de 0 :

$$\mathcal{M} = \sum_{i=1}^n X(\omega_i)^2 \mathbf{M}(\omega_i)$$

La variance de  $X$  est à rapprocher du moment d'inertie par rapport au centre de masse.

1. Moment, vient du latin *momentum*-i, impulsion, poids...