Chapitre 23 – Probabilités

Table des matières

1 Expérience aléatoire, événements et variables aléatoires

Le concept d'expérience aléatoire n'est pas mathématique à proprement parler. On appelle ainsi toute expérience — expérience matérielle ou expérience de pensée — susceptible a priori de résultats différents quand on la répète. L'ensemble des résultats possibles d'une telle expérience est appelé son univers et généralement noté Ω. Plutôt que de résultats possibles, on parle aussi parfois d'issues et de réalisations.

– L'expérience d'un lancer de dé à 6 faces peut conduire à 6 résultats selon la face obtenue. Pour modéliser un tel lancer, on peut choisir pour univers Ω l'ensemble $[\![1, 6]\!]$.
– L'expérience du tirage de 2 boules successivement avec remise dans une urne contenant des boules noires et des boules blanches peut conduire à 4 résultats : NN, NB, BN et BB si on choisit d'associer « N » à la couleur noire et « B » à la couleur blanche. Pour modéliser un tel tirage, on peut choisir pour univers Ω l'ensemble $\{N, B\}$ des 2-listes de $\{N, B\}$, i.e. l'ensemble des mots de 2 lettres sur l'alphabet $\{N, B\}$.

En PCSI, nos univers seront officiellement toujours des ensembles FINIS.

Nous tricherons parfois un peu dans certains exercices. Cette restriction du programme aux univers finis est uniquement technique. La théorie des probabilités sur un univers quelconque est techniquement délicate et vous la découvrirez un peu en deuxième année — mais juste un peu. L'ennui bien sûr, c'est qu'en limitant la taille des univers, on limite drastiquement le nombre des expériences aléatoires autorisées. Il nous sera par exemple impossible en PCSI :

– de jouer à pile ou face INDÉFINIMENT,
– de jouer aux fléchettes contre un disque de rayon 2 et de calculer la probabilité d'atterrir dans le disque central de rayon 1, intuitivement égale au quotient d'aires $\dfrac{\pi \times 1^2}{\pi \times 2^2} = \dfrac{1}{4}$ – mais comment le justifier ?

Pour une expérience aléatoire donnée, on peut bien sûr appréhender chaque résultat isolément en tant qu'ÉLÉMENT de Ω, mais ce qui nous intéresse généralement, ce sont des ensembles de résultats définis le plus souvent par une propriété commune et appelés <u>événements</u>. Toute partie de Ω est appelée un <u>événement</u>. L'ensemble des événements de l'expérience aléatoire étudiée est ainsi l'ensemble total $\mathscr{P}(\Omega)$ des parties de Ω.

– Dans le lancer de dé précédent, la proposition « La face obtenue est paire » est satisfaite par 3 résultats, en l'occurrence 2, 4 et 6, et peut être identifiée à l'ensemble $\{2, 4, 6\}$.
– Dans le tirage dans l'urne précédent, la proposition « La première boule est blanche » est satisfaite par 2 résultats, en l'occurrence BN et BB, et peut être identifiée à l'ensemble $\{BN, BB\}$.

Exercice 1. Déterminer l'univers Ω dans chacun des cas suivants :

- on tire à pile ou face avec une pièce de monnaie : l'univers Ω a deux éléments
- on jette trois pièces :
- on tire une carte dans un jeu de 52 cartes : Ω a 52 éléments
- on jette un dé :
- on jette deux dés :
- on prévoit le temps qu'il fera dans 30 jours :
- on prévoit si un photon va passer dans la fente de gauche ou dans celle de droite
- on fait un sondage sur une population de taupins avec une unique question : « la fonction $x \mapsto \dfrac{x^5+\sin x \tan x^3}{x^4 \cos x}$ est-elle prolongeable par continuité en 0 ? »
- Lance une fléchette sur un disque de centre O et de rayon 1.

Définition 2 (Vocabulaire probabiliste usuel sur les événements). Soit Ω un univers fini.
- Les singletons de Ω sont appelés les événements élémentaires de Ω. L'événement Ω est appelé l'événement certain et l'événement $\varnothing$ l'événement impossible.
- Pour tous événements $A, B \in \mathscr{P}(\Omega)$, l'événement $A \cup B$ est appelé l'événement « A ou B » et l'événement $A \cap B$ l'événement « A et B ». L'événement $\overline{A}$ est quant à lui appelé l'événement contraire de A. On dit enfin que les événements A et B sont incompatibles s'ils sont disjoints, i.e. si $A \cap B = \varnothing$.

La théorie des probabilités nous mettra généralement en présence d'événements plus compliqués à décrire que de simples « A ou B » ou « A et B ». Nous aurons par exemple affaire à des réunions d'intersection qui vous paniqueront au départ, mais que vous trouverez vite naturelles.

Exemple 3. On lance une pièce n fois successivement. Pour tout $k \in [\![1,n]\!]$, on note F_k l'événment « On obtient face au k-ième lancer ». Les événements $F_1, \dots, F_n$ sont en quelque sorte les événements les plus basiques aux quels cette expérience aléatoire nous confronte. Il paraît raisonnable que tout événement puisse être écrit en fonction d'eux. Par exemple :

- l'événement A « On obtient pile au moins une fois au cours des n lancers » peut être écrit $A = \displaystyle\bigcup_{k=1}^{n} \overline{F_k}$,
- l'événement B « On n'obtient jamais pile » peut être écrit $B = \displaystyle\bigcap_{k=1}^{n} F_k$,
- l'événement C « On obtient deux faces consécutifs au cours es n lancers » peut être écrit $C = \displaystyle\bigcup_{k=1}^{n-1} (F_k \cap F_{k+1})$.

Remarquez bien que nous n'avons pas eu besoin d'évoquer l'univers Ω pour penser et écrire ces événements. Nous y reviendrons bientôt.

Définition 4 (Système complet d'événements). Soit Ω un univers fini. On appelle système complet d'événements de Ω tout ensemble $\{A_1, \dots, A_n\}$ d'événements deux à deux incompatibles dont la réunion est l'événement certain, i.e. pour lequel : $\Omega = \displaystyle\bigcup_{k=1}^{n} A_i$ et $\forall i, j \in [\![1,n]\!], i \neq j, A_i \cap A_j = \varnothing$.

Exemple 5. Pour l'expérience aléatoire du lancer d'un dé à 6 faces, les événements « La face obtenue est paire » et « La face obtenue est impaire » forment un système complet d'événements car on obtient forcément une face paire ou une face impaire et jamais les deux en même temps. En français, on ne dit pas « système complet d'événements », mais « alternative ».

Plus généralement, soient Ω un univers fini et $A, B \in \mathscr{P}(\Omega)$. La paire d'événements $\{A, \overline{A}\}$ est un système complet d'événements car $A \cup \overline{A} = \Omega$ et $A \cap \overline{A} = \varnothing$.

Exercice 6. À l'aide de deux événements A et B, construire un système complet d'événements

Exemple 7. Pour tout univers fini $\Omega = \{\omega_1, \dots, \omega_n\}$: $\Omega = \sqcup_{k=1}^{n} \{\omega_i\}$, donc les événements élémentaires $\{\omega_1\}, \dots, \{\omega_n\}$ forment un système complet d'événements.

En réalité, la plupart du temps, les événements qu'on manipule sont construits à partir de grandeurs, numériques ou non, qu'on appelle des variables aléatoires. Par exemple, quand on lance une pièce 10 fois successivement, l'événement « On a obtenu exactement 3 piles » peut être écrit « $N = 3$ » si on note N le nombre de piles obtenus. Ce nombre N n'est pas a priori une constante. Il DÉPEND du lancer qu'on effectue et doit donc être vu comme une FONCTION.

Définition 8 (Variable aléatoire). Soient Ω un univers fini et E un ensemble quelconque.

- On appelle variable aléatoire sur Ω à valeurs dans E toute application X de Ω dans E.
  Dans la plupart des situations $E = \mathbb{R}$ et on appelle X une variable aléatoire réelle.
- Pour tout partie A de E, l'événement $X^{-1}(A)$ est noté $\{X \in A\}$ ou $(X \in A)$:

$$\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}.$$

Si A est un singleton $\{x\}$, on emploie plutôt les notations $\{X = x\}$ ou $(X = x)$:

$$\{X = x\} = X^{-1}(\{x\}) = \{\omega \in \Omega \mid X(\omega) = x\}.$$

Enfin, si $E = \mathbb{R}$ et si A est un intervalle $]-\infty, x]$, on emploie plutôt les notations $\{x \leq x\}$ et $(X \leq x)$ – même chose avec toutes les autres formes d'intervalles :

$$\{X \leq x\} = X^{-1}(]-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\}.$$

En dépit de son nom, une variable aléatoire sur Ω n'est pas une variable, c'est une fonction. De plus, en tant que fonction, une variable aléatoire associe fixement une valeur à tout élément de Ω et ceci n'a rien d'aléatoire. L'aléatoire fera irruption quand nous attribuerons à tout événement de Ω une probabilité, i.e. une certaine mesure de sa vraisemblance.

Exemple 9.

- Pour modéliser le lancer d'un dé à 6 faces 2 fois, on peut choisir pour univers Ω l'ensemble $[\![1, 6]\!]$ des 2-listes de $[\![1, 6]\!]$. La fonction X_1 (resp. X_2) « valeur obtenue au premier (resp. au deuxième) lancer » est une variable aléatoire réelle sur Ω d'image $[\![1, 6]\!]$. La fonction $S = X_1 + X_2$ en est une autre, d'image $[\![2, 12]\!]$.
  Par exemple, pour $\omega = (2, 5) \in \Omega$: $X_1(\omega) = 2$, $X_2(\omega) = 5$ et $S(\omega) = 7$.
- Pour modéliser le tirage simultané de 4 entiers entre 1 et 10, on peut choisir pour univers Ω l'ensemble $\mathscr{P}_4([\![1, 10]\!])$ des 4-combinaisons de $[\![1, 10]\!]$. La fonction X « plus petit entier tiré » est une variable aléatoire réelle sur Ω d'image $[\![1, 7]\!]$ – comme on tire 4 entier, il est impossible que le plus petit entier tiré vaille 8, 9 ou 10.
  Par exemple, pour $\omega = \{2, 5, 6, 8\} \in \Omega$, on a : $\qquad X(\omega) = 2$.

Définition-théorème 10 (Systeme complet d'événements associé à une variable aléatoire). soient Ω un univers fini et X une variable aléatoire. Les événements $\{X = x\}$, x décrivant $X(\omega)$ forment un système complet d'événements de Ω appelé le système complet d'événements de Ω associé à X. En outre, pour tout $A \in \mathscr{P}(X(\Omega))$: $\qquad \{X \in A\} = \bigsqcup_{x \in A} \{X = x\}$.

La dernière relation est facile à lire. Dire que X appartient à A revient simplement à dire que X est égal à x pour un et un seul élément x de A. L'unicité est indiquée par la notation « $\sqcup$ »

2 Probabilités sur un univers fini

Le concept d'expérience aléatoire décrit des situations susceptibles de conduire à plusieurs résultats, mais rien ne nous permet pour le moment de mesurer la VRAISEMBLANCE de ces résultats les uns par rapport aux autres. Nous allons associer dans ce paragraphe à tout événement d'une expérience aléatoire une probabilité, i.e. un réel compris entre 0 et 1 dont la valeur 0 représente le plus bas niveau de vraisemblance et la valeur 1 le niveau le plus élevé.

Définition 11 (Probabilité sur un univers fini). Soit Ω un univers fini. On appelle probabilité sur Ω toute application $P : \mathscr{P}(\Omega) \longrightarrow [0, 1]$ pour laquelle d'une part : $P(\Omega) = 1$ et d'autre part :

$$\forall A, B \in \mathscr{P}(\Omega), \quad A \cap B = \varnothing \quad \Longrightarrow \quad P(A \sqcup B) = P(A) + P(B) \quad \text{(additivité)}.$$

Le couple (Ω, P) est alors appelé un espace probabilisé (fini).

ATTENTION ! Le mot « probabilité » est utilisé de deux manières différentes qu'il convient de bien distinguer :

- La « probabilité de A » est le RÉEL $P(A)$ de $[0, 1]$ qu'on associe à l'événement A pour mesurer sa vraisemblance.
- La « probabilité P » est l'APPLICATION de Ω dans $[0, 1]$ qu'on choisit pour mesurer la vraisemblance de tous les événements de Ω.

Les probabilités uniformes définies ci-dessous constituent l'exemple le plus naturel de probabilité sur un ensemble fini, mais ce ne sont pas les seules probabilités que nous aurons l'occasion d'utiliser — loin de là.

Définition-théorème 12 (Probabilité uniforme). Soit Ω un univers fini. L'application $A \overset{P}{\longmapsto} \dfrac{|A|}{|\Omega|}$ est une probabilité sur Ω appelé sa probabilité uniforme.

Vous connaissez la relation : $P(A) = \dfrac{|A|}{|\Omega|}$ sous la forme suivante : $\dfrac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$.

Exemple 13. On lance une fois un dé équilibre à 6 faces. Quel espace probabilisé pour cette expérience aléatoire ? On peut choisir pour univers Ω l'ensemble $[\![1,6]\!]$ des résultats potentiels et pour probabilité P sur Ω la probabilité uniforme. Avec un dé pipé, on n'aurait bien sûr pas choisi pour P la probabilité uniforme.

Si on note A l'événement « On obtient un nombre premier » et B l'événement « On obtient un nombre pair », alors :

$$P(A) = P(\{2,3,5\}) = \frac{|\{2,3,5\}|}{|\Omega|} = \frac{1}{2} \quad \text{et} \quad P(B) = P(\{2,4,6\}) = \frac{|\{2,4,6\}|}{|\Omega|} = \frac{1}{2}.$$

Dans ce cas particulier où : $P(A) = P(B)$, on dit que les événements A et B ont équiprobables.

Plus généralement :

Définition 14 (Événements équiprobables). Soient (Ω, P) un espace probabilisé fini et $A, B \in \mathscr{P}(\Omega)$. On dit que les événements A et B sont équiprobables si : $P(A) = P(B)$.

Exemple 15. Soit $\Omega = \{\omega_1, \dots, \omega_n\}$ un univers fini. On note P la probabilité uniforme sur Ω.
Les événements élémentaires $\{\omega_1\}, \dots, \{\omega_n\}$ sont équiprobables de probabilité commune $\dfrac{1}{|\Omega|}$.

Proposition 16 (Propriétés des probabilités). Soient (Ω, P) un espace probabilisé fini et $A, B, A_1, \dots, A_n \in \mathscr{P}(\Omega)$.

(i) **Ensemble vide** : $P(\varnothing) = 0$

(ii) **Complémentaire et différence** : $P(\overline{A}) = 1 - P(A)$ et $P(A \backslash B) = P(A) - P(A \cap B)$.

(iii) **Croissance** : Si : $A \subset B$, alors : $P(A) \leq P(B)$.

(iv) **Réunion** : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. En général : $P\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} P(A_k)$ (sous-additivité), et si $A_1, \dots, A_n$ sont DEUX À DEUX INCOMPATIBLES : $P\left(\bigsqcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$.

Formule du crible : $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} P\left(A_{i_1} \cap \ldots \cap A_{i_k}\right)$ (hors programme).

Nous avons présenté plus haut l'exemple naturel des probabilités uniformes. Notre prochain théorème décrit quant à lui l'ensemble de TOUTES les probabilités qu'on peut installer sur un univers fini. Se donner une probabilité P sur un univers Ω revient toujours à faire un choix dans ce vaste ensemble de toutes les probabilités possibles. Fixer P, c'est décider une fois pour toutes que P mesure fidèlement la vraisemblance des événements de Ω et qu'aucune autre probabilité ne le fait aussi bien. On ne travaille pas dans le cas d'un lancer de dé pipé avec la même probabilité que dans le cas d'un dé équilibré, et pourtant on réalise la même expérience. Mais d'abord, une définition.

Définition 17 (Distribution de probabilités). Soit E un ensemble fini. On appelle distribution de probabilités sur E toute famille d'éléments de $[0,1]$ indexée par E dont la somme est égale à 1.

Exemple 18. Soit (Ω, P) un espace probabilisé fini. La famille $(P(\{\omega\}))_{\omega \in \Omega}$ des probabilités des événements élémentaires de Ω est une distribution de probabilités sur Ω car : $\sum_{\omega \in \Omega} P(\{\omega\}) = P\left(\bigsqcup_{\omega \in \Omega} \{\omega\}\right) = P(\Omega) = 1$.

Proposition 19 (Détermination d'une probabilité sur les événements élémentaires). Soient Ω un univers fini et $(p_\omega)_{\omega\in\Omega}$ une distribution de probabilités sur Ω. Il existe une et une seule probabilité P sur Ω pour laquelle : $P(\{\omega\}) = p_\omega$ pour tout $\omega \in \Omega$. En l'occurrence, pour tout $A \in \mathscr{P}(\Omega)$: $P(A) = \sum_{\omega\in A} p_\omega$.

En résumé, on connaît tout d'une probabilité quand on connaît la distribution de probabilités des événements élémentaires qui lui est associée.

Exemple 20. On lance une fois un dé pipé à 6 faces qui donne la face « 1 » avec probabilité $\frac{1}{4}$ et les autres faces avec une même probabilité p. Dans ces conditions : $\frac{1}{4} + 5p = 1$, donc : $p = \frac{3}{20}$. Pour modéliser ce lancer, il paraît naturel qu'on choisisse pour univers Ω l'ensemble $[\![1, 6]\!]$ des résultats possibles de l'expérience et pour probabilité P la probabilité définie par la distribution de probabilités $\left(\frac{1}{4}, \frac{3}{20}, \frac{3}{20}, \frac{3}{20}, \frac{3}{20}, \frac{3}{20}\right)$ sur Ω.

L'événement A « On obtient une face impaire » a pour probabilité : $P(A) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{1}{4} + 2 \times \frac{3}{20} = \frac{11}{20}$.

3 Loi d'une variable aléatoire

Sur un espace probabilisé fini (Ω, P) fixé, P mesure la vraisemblance de tous les événements concevables, donc en particulier de tous les événements qu'on peut construire à partir d'une variable aléatoire X sur Ω. Les valeurs de X possèdent ainsi chacune leur propre vraisemblance, ce que la définition suivante formalise proprement.

Définition-théorème 21 (Loi d'une variable aléatoire). Soient (Ω, P) un espace probabilisé fini et X une variable aléatoire sur Ω. On appelle loi de X (pour la probabilité P) l'application $A \xrightarrow{P_X} P(X \in A)$ de $\mathscr{P}(X(\Omega))$ dans $[0, 1]$.

- L'application P_X est une probabilité sur $X(\Omega)$.
- Elle est entièrement déterminée par la distribution de probabilités $(P(X = x))_{x\in X(\Omega)}$ sur $X(\Omega)$, qu'on appelle justement la distribution de probabilités de X.

  En l'occurrence, pour tout $A \in \mathscr{P}(X(\Omega))$: $P_X(A) = P(X \in A) = \sum_{x\in A} P(X = x)$.

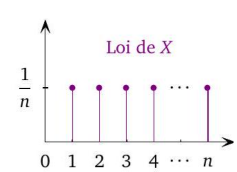
À partir d'une probabilité P sur Ω, on définit ici une nouvelle probabilité P_X sur l'ensemble $X(\Omega)$ des valeurs de X.

Définition 22 (Loi uniforme). Soient E un ensemble fini non vide et X une variable aléatoire à valeurs dans E. On dit que X suit la loi uniforme sur E ou que X est une variable uniforme sur E, ce qu'on note : $X \sim \mathscr{U}(E)$, si P_X est la probabilité uniforme sur E, i.e. si pour tout $A \in \mathscr{P}(E)$: $P_X(A) = P(X \in A) = \frac{|A|}{|E|}$.

La loi uniforme est la loi pour laquelle les événements $\{X = x\}, x$ décrivant E, sont équiprobables de probabilité $\frac{1}{|E|}$. La formule : $P(X \in A) = \frac{|A|}{|E|} = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$ est essentielle en pratique et ramène tout problème de calcul sur la loi uniforme à un problème de dénombrement.

Exemple 23.

- Lorsqu'on choisit au hasard un entier X entre 1 et n, X suit la loi uniforme sur $[\![1, n]\!]$.
- Lorsqu'on lance 2 fois successivement un dé équilibré à 6 faces, le couple (X_1, X_2) des valeurs obtenues suit la loi uniforme sur $[\![1, 6]\!]^2$.



- La loi uniforme sur $\{-1, 1\}$ est souvent appelée la loi de Rademacher. On dit donc d'une variable aléatoire Y pour laquelle : $P(Y = -1) = P(Y = 1) = \frac{1}{2}$ qu'elle suit la loi de Rademacher ou que c'est une variable de Rademacher.

Exemple 24. Une urne contient 3 boules blanches et 5 noires. On en tire simultanément 4 boules. Avec quelle probabilité n'a-ton tiré que des boules noires ? On peut bien sûr supposer sans perte de généralité que les 8 boules de l'urne sont numérotées de 1 à 8 histoire de les distinguer - de 1 à 3 pour les blanches et de 4 à 8 pour les noires. Notons alors X la 4-combinaison de $[\![1,8]\!]$ des boules tirées. Cette variable aléatoire suit la loi uniforme sur l'ensemble $\mathscr{P}_4\left([\![1,8]\!]\right)$ des 4-combinaisons de $[\![1,8]\!]$ et la probabilité cherchée vaut :

$$P\left(X \in \mathscr{P}_4([\![4,8]\!])\right) = \frac{|\mathscr{P}_4([\![4,8]\!])|}{|\mathscr{P}_4([\![1,8]\!])|} = \frac{\begin{pmatrix} 5 \\ 4 \end{pmatrix}}{\begin{pmatrix} 8 \\ 4 \end{pmatrix}} = \frac{1}{14}.$$

Définition 25 (Loi de Bernoulli).
- Soient X une variable aléatoire à valeurs dans $\{0,1\}$ et $p \in [0,1]$. On dit que X suit la loi de Bernoulli de paramètre p ou que X est une variable de Bernoulli de paramètre p, ce qu'on note : $X \sim \mathscr{B}(p)$, si : $P(X=1) = p$ et $P(X=0) = 1-p$
- **Exemple fondamental** : Pour tout espace probabilisé fini (Ω, P) et pour tout $A \in \mathscr{P}(\Omega)$: $\mathbb{1}_A \sim \mathscr{B}(P(A))$.

Les indicatrices jouent un rôle naturel en probabilités que nous exploiterons davantage au chapitre « Position et dispersion d'une variable aléatoire réelle », mais qui mérite d'être compris dès maintenant. Compter les bruns dans une assemblée, c'est donner la valeur « 1 » aux bruns et la valeur « 0 » aux autres, puis additionner toutes ces valeurs. Plus généralement :

Toute variable aléatoire qui représente un cardinal peut être exprimée comme une somme d'indicatrices, i.e. comme une somme de variables aléatoires de loi de Bernoulli.

Exemple 26. On lance n fois un dé à 6 faces. Si on note A_k l'événement « On obtient 6 au k ème lancer » pour tout $k \in [\![1, n]\!]$ et N le nombre de 6 obtenus, alors : $N = \mathbb{1}_{A_1} + \ldots + \mathbb{1}_{A_n}$.

On s'intéresse à présent à la loi de l'image d'une variable aléatoire par une fonction. Commençons par un exemple.

Exercice 27. Soit X une variable uniforme sur $[\![-2, 2]\!]$. Quelle est la loi de la variable aléatoire $X^2 + 1$?

Proposition 28 (Loi de $f(X)$). Soient (Ω, P) un espace probabilisé fini, F un ensemble, X une variable aléatoire sur Ω et $f : X(\Omega) \longrightarrow F$ une fonction. La loi $P_{f(X)}$ de $f(X) = f \circ X$ est entièrement déterminée par f et la loi de X. En l'occurrence, pour tout $A \in \mathscr{P}(f(X(\Omega)))$: $\quad P_{f(X)}(A) = P(f(X) \in A) = \displaystyle\sum_{\substack{x \in X(\Omega) \\ f(x) \in A}} P(X = x)$.

Ce résultat montre en particulier que si deux variables aléatoires X et Y ont la même loi, les variables $f(X)$ et $f(Y)$ ont aussi la même loi pour toute fonction f pour laquelle la composition a un sens.

Exercice 29. Soit X un variable aléatoire uniforme sur $[\![1, 2n]\!]$. Quelle est la loi de $(-1)^X$?

En début de chapitre, nous avons décrit des situations probabilistes plutôt concrètes — lancers de dés, tirages dans une urne... - en définissant proprement un univers Ω adapté et une probabilité P sur Ω. Plus loin, nous avons décrit le même genre de situations d'une manière différente, en l'occurrence en introduisant une variable aléatoire de loi présumée connue. Quant à nos deux derniers exemples, ils ne relèvent plus du tout de la modélisation et commencent abstraitement ainsi : « Soit X une variable aléatoire de loi $(\ldots)$ ». Hélas, quel sens cela a-t-il d'introduire une variable aléatoire hors sol sans définir au préalable un espace probabilisé susceptible de la porter ? En outre, qui nous garantit que nos désirs sont des réalités ? Suffit-il en mathématiques de dire : « Fiat lux » pour que la lumière soit ? Nous savons bien que non. Le théorème qui suit, hors programme mais très simple, justifie une fois pour toutes la consistance de tous les énoncés dont le point de départ est une variable aléatoire de loi prescrite.

Théorème (Définition implicite d'un espace probabilisé par la donnée d'une distribution de probabilités). Soient E un ensemble fini et $(p_e)_{e \in E}$ une distribution de probabilités sur E. Il existe alors un espace probabilisé fini (Ω, P) et une variable aléatoire X sur Ω d'image E pour lesquels pour tout $e \in E$: $\quad P(X = e) = p_e$.

Dans la plupart des situations que nous étudierons en pratique, le travail commencera par la donnée d'une ou plusieurs variables aléatoires de lois prescrites que nous nous donnerons sans nous préoccuper jamais de l'espace probabilisé (Ω, P) qui leur sert de support. Celui-ci est voué à rester caché et ne présente de toute façon aucun caractère d'unicité, de nombreux choix d'espace probabilisé sont possibles pour la description d'une même situation. En effet, si je lance un dé à 6 faces une fois, j'ai envie de choisir l'ensemble $[\![1, 6]\!]$ pour univers Ω_1, mais si je relance le dé, l'univers $\Omega_2 = [\![1, 6]\!]^2$ paraît plus approprié. Toute probabilité d'un événement qui fait intervenir les deux lancers de dés doit être calculée dans le cadre de l'univers Ω_2, mais Ω_1 suffit si on s'intéresse seulement au premier lancer. Or va-t-on ainsi changer d'univers à chaque fois qu'on change de question ? Ce serait donner l'impression que nous ne sommes pas confrontés toujours à la même réalité avec une conception unique de ce qui est vraisemblable ou non.

La réponse des probabilistes à ce problème des univers multiples consiste à les négliger une bonne fois pour toutes et à ne jamais accorder le moindre regard à l'univers Ω. Sur quel Ω travaillerons-nous désormais ? Peu importe. Nous venons de voir que toute donnée intuitive d'une variable aléatoire peut être formalisée proprement, c'est suffisant. NOUS NE DÉFINIRONS PLUS JAMAIS L'UNIVERS Ω DE NOS TRIBULATIONS PROBABILISTES. Par exemple, si un exercice vous met dans la situation suivante : "On lance un dé à 6 faces une fois et on note X la face obtenue », vous pouvez affirmer que X est une variable aléatoire de loi uniforme sur $[\![1, 6]\!]$ sans évoquer aucun espace probabilisé fini (Ω, P).

Exercice 30. Une urne contient $3n$ boules numérotées de 1 à $2n$ avec, pour tout $k \in [\![1, n]\!]$, exactement 2 boules indiscernables de numéro $2k$ et exactement une boule de numéro $2k - 1$. Avec quelle probabilité une boule tirée au hasard dans cette urne est-elle de numéro pair ?

4 Probabilités conditionnelles

Soit (Ω, P) un espace probabilisé fini. Faisons l'hypothèse qu'un certain événement B est réalisé. Cette hypothèse modifie a priori la vraisemblance de tous les événements du contexte étudié. Il paraît par exemple raisonnable de considérer que, sous cette hypothèse, B est réalisé « avec probabilité 1 » et que $\overline{B}$ l'est « avec probabilité nulle », pourtant il est faux a priori que : $P(B) = 1$ et $P(\overline{B}) = 0$. Comment tenir compte de notre hypothèse sur B ? La probabilité P mesure la vraisemblance de tout événement, mais ceci avant toute hypothèse selon laquelle B est réalisé. Sous cette hypothèse, une autre mesure de vraisemblance doit être introduite, une autre probabilité sur Ω que nous noterons P_B et qu'on appelle la probabilité conditionnelle sur Ω sachant B. C'est pour cette nouvelle probabilité qu'on peut écrire : $P_B(B) = 1$ et $P_B(\overline{B}) = 0$.

À présent, comment calcule-t-on $P_B(A)$ pour un événement A quelconque ? La probabilité $P_B(A)$ mesure la vraisemblance de l'intersection $A \cap B$ au sens de la probabilité initiale P, mais si on veut que l'univers Ω lui-même ait une nouvelle vraisemblance de 1, il convient de poser : $P_B(A) = \dfrac{P(A \cap B)}{P(B)}$ sous l'hypothèse technique que : $P(B) > 0$. Cette définition fait de P_B une sorte de "probabilité d'oubli », une émanation de P aveugle à tout ce qui contredit B. La division par $P(B)$ n'est qu'une manière de normaliser P_B. Elle garantit l'égalité : $P_B(B) = 1$ selon laquelle « B est certain »- un nouvel univers apparent. En outre, comme voulu : $P_B(\overline{B}) = 0$.

Définition 31 (Probabilité conditionnelle). Soient (Ω, P) un espace probabilisé fini et $B \in \mathscr{P}(\Omega)$ pour lequel : $P(B) > 0$. Pour tout $A \in \mathscr{P}(\Omega)$, le réel : $P(A \mid B) = P_B(A) = \dfrac{P(A \cap B)}{P(B)}$ est appelé la probabilité conditionnelle de A sachant B. L'application P_B est alors une probabilité sur Ω appelée sa probabilité conditionnelle sachant B (pour la probabilité P).

On ne sait pas définir $P_B(A)$ lorsque : $P(B) = 0$, mais par convention, l'égalité : $P(A \cap B) = P(B)P_B(A)$ reste tout de même vraie dans ce cas. De fait, si : $P(B) = 0$, alors par croissance de P : $P(A \cap B) \leq P(B)$, donc : $P(A \cap B) = 0$.

Théorème 32 (Formule des probabilités totales V1). Soient (Ω, P) un espace probabilisé fini et $\{A_1, \ldots, A_n\}$ un système complet d'événements de Ω. Pour tout $B \in \mathscr{P}(\Omega)$:

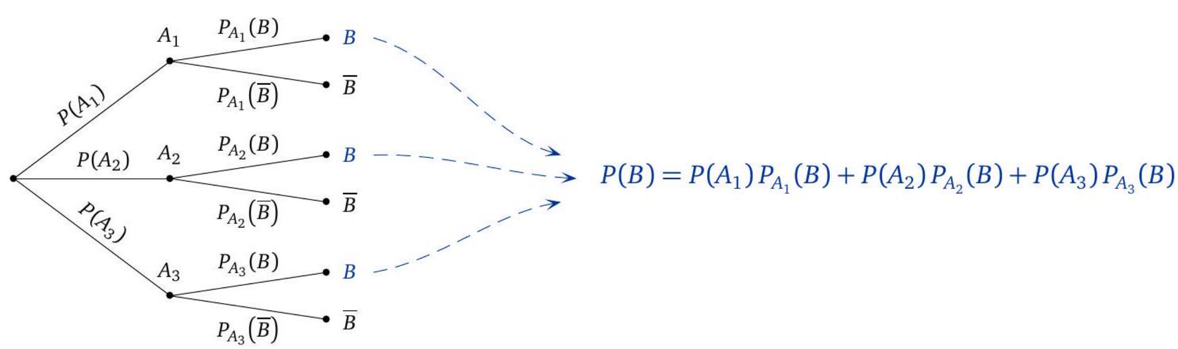
$$P(B) = \sum_{i=1}^{n} P\left(A_i \cap B\right) \qquad \textbf{Formule des probabilités totales V1}$$

Théorème 33 (Formule des probabilités totales V2). Soient (Ω, P) un espace probabilisé fini et $\{A_1, \ldots, A_n\}$ un système complet d'événements de Ω. Si les événements A_i ne sont pas négligeables i.e. : $\forall i \in [\![1, n]\!], P(A_i) \neq 0$, alors pour tout $B \in \mathscr{P}(\Omega)$:

$$P(B) = \sum_{i=1}^{n} P\left(A_i\right) P(B \mid A_i) = \sum_{i=1}^{n} P\left(A_i\right) P_{A_i}(B) \qquad \textbf{Formule des probabilités totales V2}$$

Remarque. Par convention, si $P(B) = 0$, on a : $P(A \mid B)P(B) = 0$, la formule des probabilités totales V2 reste vraie même si certains des événements A_i sont négligeables.

Vous avez déjà rencontré la formule des probabilités totales auparavant, mais sans doute à travers des arbres de probabilité, par exemple sous la forme suivante pour $n = 3$:



À partir d'aujourd'hui, vous gardez le droit de penser en termes d'arbres de probabilité si cela vous aide, mais UN ARBRE DE PROBABILITÉ NE SERA JAMAIS CONSIDÉRÉ COMME UNE PREUVE BIEN FORMALISÉE. Pourquoi ? Parce qu'on ne peut PAS résoudre des problèmes un peu sophistiqués sans maîtriser la formule des probabilités totales en tant que formule.

Exercice 34. Une urne contient n boules noires et b blanches et on en tire 2 boules successivement sans remise. Avec quelle probabilité la deuxième boule tirée est-elle blanche ?

Cet exemple diffère sensiblement de ceux qui l'ont précédé. Alors que nous avons tué plus haut les univers Ω au profit des seules variables aléatoires, nous n'avons cette fois pas introduit de variables aléatoires, mais seulement des événements B_1 et B_2 et certaines probabilités, conditionnelles ou non, qui leur sont associées. De nouveau, il doit bien y avoir un espace probabilisé (Ω, P) quelque part, mais nous ne chercherons pas à formaliser davantage ce genre de situations.

Théorème 35 (Formules de Bayes). Soient (Ω, P) un espace probabilisé fini et $B \in \mathscr{P}(\Omega)$ pour lequel : $P(B) > 0$.

(i) Pour tout $A \in \mathscr{P}(\Omega)$: $\quad P_B(A) = \dfrac{P(A)P_A(B)}{P(B)}$.

(ii) Soit $\{A_1, \ldots, A_n\}$ un système complet d'événements de Ω. Pour tout $j \in [\![1, n]\!]$: $\quad P_B(A_j) = \dfrac{P(B)}{\sum_{i=1}^n P(A_j) P_{A_j}(B) P_{A_i}(B)}$.

Exercice 36. Juge au tribunal, je dois juger de la culpabilité d'une compagnie de taxis bleus. Un soir de brouillard, un taxi a percuté un piéton qui traversait la rue dans son bon droit, puis a pris la fuite. Un témoin affirme que le taxi était bleu et c'est sur la base de ce témoignage que le procès a été instruit. Or dans la ville, deux compagnies de taxis se partagent le marché, la compagnie des taxis bleus et la compagnie des taxis verts, mais tout de même les taxis verts dominent le marché au sens où 90% des taxis dans la ville sont verts.

On demande au témoin d'effectuer des tests de reconnaissance des couleurs pour mesurer la fiabilité de son témoignage. Il s'avère qu'il est fiable dans 90% des cas pour la couleur bleue et 80% des cas pour la couleur verte.

Dois-je condamner ou la compagnie des taxis bleus ?

Théorème 37 (Formule des probabilités composées). Soient (Ω, P) un espace probabilisé fini et $A_1, \ldots, A_n \in \mathscr{P}(\Omega)$.
Alors : $\quad P(A_1 \cap \ldots \cap A_n) = P(A_1) P_{A_1}(A_2) P_{A_1 \cap A_2}(A_3) \ldots P_{A_1 \cap \ldots \cap A_{n-1}}(A_n)$.

Exemple 38. Une urne contient $2n$ boules dont n noires et n blanches, et on en tire 3 boules successivement. Avec quelle probabilité les tire-t-on dans l'ordre noire/blanche/noire si les tirages se font avec remise (resp. sans remise) ?

Exemple 39. Un commerçant met en vente 50 tickets d'un jeu dont exactement 3 sont gagnants. Je lui achète 6 tickets. Avec quelle probabilité en ai-je acheté au moins un gagnant ?

Nous achèverons ce paragraphe par un cas particulier de conditionnement appliqué au concept de variable aléatoire.

Définition 40 (Lois conditionnelles d'une variable aléatoire). Soient (Ω, P) un espace probabilisé fini, X une variable aléatoire sur Ω et $A \in \mathscr{P}(\Omega)$ pour lequel : $\quad P(A) > 0$. On appelle loi conditionnelle de X sachant A (pour la probabilité P) la loi de X pour la probabilité P_A, i.e. l'application $B \longmapsto P_A(X \in B)$ de $\mathscr{P}(X(\Omega))$ dans $[0, 1]$.

La loi conditionnelle de X sachant A est entièrement déterminée par la distribution de probabilités $(P_A(X = x))_{x \in X(\Omega)}$.

Exercice 41. On choisit un entier X au hasard entre 1 et $2n$. Quelle est la loi de X sachant $\{X \leq n\}$?

5 Indépendance

Intuitivement, étant donnés deux événements A et B pour lesquels : $\quad P(B) > 0$, on a envie de dire que A et B sont indépendants si la probabilité $P(A)$ ne dépend pas de la réalisation de B, i.e. si : $P_B(A) = P(A)$, ce qui s'écrit aussi : $P(A \cap B) = P(A)P(B)$. Cette remarque conduit à la définition suivante.

Définition 42 (Événements indépendants). Soit (Ω, P) un espace probabilisé fini.

- Soient $A, B \in \mathscr{P}(\Omega)$ des événements. On dit que A et B sont indépendants si : $P(A \cap B) = P(A)P(B)$.
- Soient $A_1, \ldots, A_n \in \mathscr{P}(\Omega)$ des événements. On dit que $A_1, \ldots, A_n$ sont (mutuellement) indépendants si pour toute partie I de $[\![1, n]\!]$: $\quad P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$.

Remarquez bien que si $A_1, \ldots, A_n$ sont indépendants, alors : $P(A_1 \cap \ldots A_n) = P(A_1) \ldots P(A_n)$, mais la définition de l'indépendance est plus exigeante et exige aussi par exemple qu'on ait : $P(A_1 \cap A_2) = P(A_1)\,P(A_2)$. De fait, par définition, toute sous-famille d'une famille d'événements indépendants est une famille d'événements indépendants.

ATTENTION ! (Mutuellement) indépendants $\Longrightarrow$ DEUX À DEUX indépendants mais LA RÉCIPROQUE EST FAUSSE !

Posons par exemple : $\Omega = \llbracket 1, 4 \rrbracket$ et notons P la probabilité uniforme sur Ω. Posons en outre : $A = \{1, 2\}$, $B = \{1, 3\}$ et $C = \{2, 3\}$. Alors : $P(A) = P(B) = P(C) = \dfrac{1}{2}$ et $P(A \cap B) = P(A \cap C) = P(B \cap C) = \dfrac{1}{4}$, donc A, B et C sont deux à deux indépendants. Pourtant : $P(A \cap B \cap C) = P(\varnothing) = 0 \neq \dfrac{1}{8} = P(A)P(B)P(C)$, donc A, B et C ne sont pas (mutuellement) indépendants.

Théorème 43 (Indépendance et complémentaires). Soient (Ω, P) un espace probabilisé fini et $A_1, \ldots, A_n \in \mathscr{P}(\Omega)$ Si $A_1, \ldots, A_n$ sont indépendants, les événements $A_1^\circ, \ldots, A_n^\circ$ le sont aussi pour tous $A_1^\circ \in \left\{A_1, \overline{A_1}\right\}, \ldots, A_n^\circ \in \left\{A_n, \overline{A_n}\right\}$.

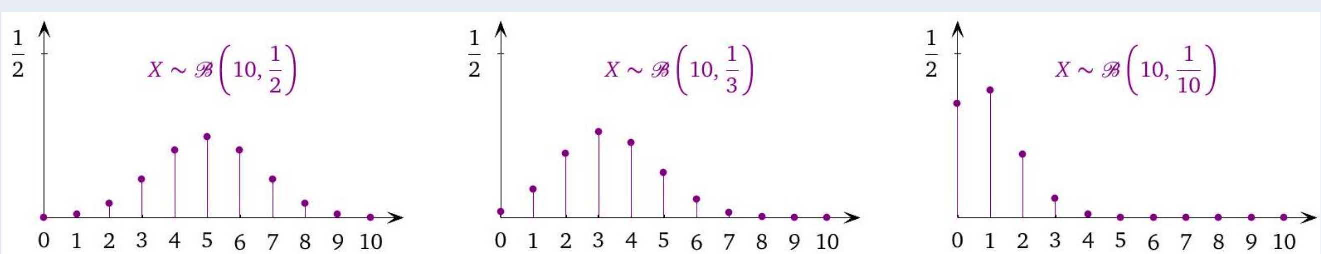
Grâce au concept d'indépendance, nous allons enfin pouvoir définir une loi usuelle très importante en pratique — la loi binomiale. Intéressons-nous pour cela à la répétition n fois indépendamment d'une expérience aléatoire à deux issues, disons « favorable » et « défavorable », dont l'issue favorable est de probabilité p. Quelle est la loi du nombre X d'issues favorables ? Pour tout $i \in \llbracket 1, n \rrbracket$, notons F_i l'événement « La i ème expérience a une issue favorable ». Les événements $F_1, \ldots, F_n$ sont indépendants par hypothèse. La variable aléatoire X est à valeurs dans «0, $n\rrbracket$ et pour tout $k \in \llbracket 0, n \rrbracket$, l'événement $\{X = k\}$ est réalisé si et seulement si on a obtenu k issues favorables et $n - k$ issues défavorables. Ainsi, si on note $\mathscr{P}_k(\llbracket 1, n \rrbracket)$ l'ensemble des parties de $\llbracket 1, n \rrbracket$ de cardinal k :

$$\{X = k\} = \bigsqcup_{I \in \mathscr{P}_k(\llbracket 1, n \rrbracket)} \left(\bigcap_{i \in I} F_i \cap \bigcap_{j \in \llbracket 1, n \rrbracket \setminus I} \overline{F_j} \right), \quad \text{donc :}$$

$$P(X = k) = \sum_{I \in \mathscr{P}_k(\llbracket 1, n \rrbracket)} P\left(\bigcap_{i \in I} F_i \cap \bigcap_{j \in \llbracket 1, n \rrbracket \setminus I} \overline{F_j} \right) \overset{\text{Indép.}}{=} \sum_{I \in \mathscr{P}_k(\llbracket 1, n \rrbracket)} \prod_{i \in I} P(F_i) \prod_{j \in \llbracket 1, n \rrbracket \setminus I} P\left(\overline{F_j}\right) = \sum_{I \in \mathscr{P}_k(\llbracket 1, n \rrbracket)} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Prenez le temps d'observer à quel point l'indépendance des événements $F_1, \ldots, F_n$ a été cruciale dans ce calcul. Sans indépendance, on n'aurait pas su avancer. Observez aussi qu'a posteriori, c'est la formule du binôme qui justifie que la famille $\left(\binom{n}{k} p^k (1-p)^{n-k} \right)_{k \in \llbracket 0, n \rrbracket}$ est une distribution de probabilités sur $\llbracket 0, n \rrbracket$.

Définition 44 (Loi binomiale). Soient X une variable aléatoire à valeurs dans $\llbracket 0, n \rrbracket$ et $p \in [0, 1]$. On dit que X suit la loi binomiale de paramètre (n, p) ou que X est une variable binomiale de paramètre (n, p), ce qu'on note : $X \sim \mathscr{B}(n, p)$, si pour tout $k \in \llbracket 0, n \rrbracket$: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.



Lorsqu'on répète n fois indépendamment une expérience aléatoire à deux issues avec probabilité p pour l'issue favorable, le NOMBRE D'ISSUES FAVORABLES suit la loi binomiale $\mathscr{B}(n, p)$.

La loi binomiale est aussi appelée la loi des tirages avec remise, expression qui trouve son explication dans l'exemple suivant. Lorsqu'on tire AVEC REMISE - donc indépendamment - n boules dans une urne contenant des boules blanches en proportion p et des boules noires en proportion $1 - p$, la variable aléatoire «nombre de boules blanches tirées » suit la loi binomiale $\mathscr{B}(n, p)$.

Pour finir, il est clair que la loi binomiale $\mathscr{B}(1, p)$ n'est autre que la loi de Bernoulli $\mathscr{B}(p)$!

Exercice 45. On lance 5 fois un dé équilibré à 6 faces dont 2 blanches et 4 noires. Avec quelle probabilité obtient-on exactement 3 fois une face noire ?

On adapte à présent aux variables aléatoires le concept d'indépendance.

Définition 46 (Variables aléatoires indépendantes). Soient (Ω, P) un espace probabilisé fini et $X_1, \ldots, X_n$ des variables aléatoires sur Ω. On dit que $X_1, \ldots, X_n$ sont (mutuellement) indépendantes si pour toutes parties A_1 de $X_1(\Omega)$, $\ldots$, A_n de $X_n(\Omega)$, les événements $\{X_1 \in A_1\}, \ldots, \{X_n \in A_n\}$ sont indépendants.

Il est équivalent d'exiger que pour tous $x_1 \in X_1(\Omega), \ldots, x_n \in X_n(\Omega)$:

$$P\left(X_1 = x_1 \text{ et } \ldots \text{ et } X_n = x_n\right) = P\left(X_1 = x_1\right) \ldots P\left(X_n = x_n\right).$$

L'indépendance des variables aléatoires étant définie comme une indépendance d'événements, toute sous-famille d'une famille de variables aléatoires indépendantes est une famille de variables aléatoires indépendantes. ATTENTION ! (Mutuellement) indépendantes $\Longrightarrow$ DEUX À DEUX indépendantes mais LA RÉCIPROQUE EST FAUSSE !

Exercice 47. On lance un dé équilibré à 6 faces 2 fois et on note X_1 (resp. X_2) la face obtenue au premier (resp. second) lancer. Avec quelle probabilité obtient-on les deux fois une face impaire ?

Exemple 48. Soient $p_1, \ldots, p_n \in [0,1]$ et $X_1, \ldots, X_n$ des variables de Bernoulli indépendantes de paramètres respectifs $p_1, \ldots, p_n$. Alors : $X_1 \ldots X_n \sim \mathscr{B}(p_1 \ldots p_n)$.

Exercice 49. Un jeu de 32 cartes a été malicieusement truqué, on y a remplacé une carte autre que l'as de pique par un deuxième as de pique. On répète n fois avec remise l'expérience consistant à tirer simultanément 4 cartes. À partir de quelle valeur de n la probabilité de déceler la supercherie est-elle supérieure ou égale à 0,9 ?

Théorème 50 (Lemme des coalitions). Soient (Ω, P) un espace probabilisé fini, E et F deux ensembles, $X_1, \ldots, X_n$ des variables aléatoires sur Ω et $f : (X_1, \ldots, X_m)(\Omega) \longrightarrow E$ et $g : (X_{m+1}, \ldots, X_n)(\Omega) \longrightarrow F$ deux fonctions. Si $X_1, \ldots, X_n$ sont indépendantes, alors les coalitions $f(X_1, \ldots, X_m)$ et $g(X_{m+1}, \ldots, X_n)$ le sont aussi.

Cet énoncé se généralise sans difficulté au cas d'un nombre fini quelconque de coalitions.

L'énoncé paraît compliqué, mais il dit simplement, par exemple, que si trois variables aléatoires X, Y et Z définies sur un même espace probabilisé fini sont indépendantes, les variables aléatoires $X + Y$ et Z le sont aussi.

6 Loi d'une famille de variable aléatoires

Comme nous l'avons remarqué dans le lemme des coalitions, toute famille de variables aléatoires est elle-même une variable aléatoires car nous avons autorisé depuis le début nos variables aléatoires à prendre leurs valeurs dans un ensemble quelconque.

Définition-théorème 51 (Loi conjointe et lois marginales d'une famille de variables aléatoires). Soient (Ω, P) un espace probabilisé fini et X et Y deux variables aléatoires sur Ω.

- L'application $\omega \longmapsto (X(\omega), Y(\omega))$ est une variable aléatoire sur Ω notée (X, Y) et sa loi $P_{(X,Y)}$ est souvent appelée la loi conjointe de X et Y. Son image $(X,Y)(\Omega)$ est incluse dans le produit $X(\Omega) \times Y(\Omega)$, sans égalité en général.
  Elle est entièrement déterminée par la distribution de probabilités $(P(X = x \text{ et } Y = y))_{(x,y) \in (X,Y)(\Omega)}$ sur$(X,Y)(\Omega)$. Le réel $P(X = x \text{ et } Y = y)$ est souvent noté $P(X = x, Y = y)$.
- La loi P_X de X est appelée la première loi marginale de (X, Y) et la loi P_Y de Y sa deuxième loi marginale.

Ces définitions se généralisent sans difficulté aux familles d'un nombre fini quelconque de variables aléatoires.

Exercice 52. Une urne contient 3 boules blanches et 1 noire et on en tire successivement 2 boules sans remise. Le couple de couleurs ainsi tirées est noté (C_1, C_2) - avec « B » pour les blanches et « N » pour l'unique noire. Quelle est la loi de (C_1, C_2) ?

ATTENTION ! Nous savons maintenant associer trois notions de lois P_X, P_Y et $P_{(X,Y)}$ à un couple (X, Y) de variables aléatoires, mais comment ces notions sont-elles liées ? La donnée de P_X et P_Y détermine-t-elle $P_{(X,Y)}$? Et dans l'autre sens ?

- En général, le réel $P(X = x \text{ et } Y = y)$ ne dépend pas tant des nombres $P(X = x)$ et $P(Y = y)$, i.e. des variables X et Y isolément, que du lien éventuellement étroit qui unit ces variables. En d'autres termes, on ne connaît pas la loi conjointe du couple (X, Y) quand on connaît séparément ses lois marginales.

- En revanche, lorsque X et Y sont indépendantes : $\quad P(X = x \text{ et } Y = y) = P(X = x)P(Y = y) \quad$ pour tous $x \in X(\Omega)$ et $y \in Y(\Omega)$, donc la donnée des lois marginales P_X et P_Y détermine entièrement la loi conjointe $P_{(X,Y)}$. Plus précisément :

> Les variables $X_1, \dots, X_n$ sont indépendantes
> si et seulement si la distribution de probabilités de $(X_1, \dots, X_n)$
> se calcule par simple produit à partir des distributions de probabilités de $X_1, \dots, X_n$.

Théorème 53 (Calcul des lois marginales à partir de la loi conjointe). Soient (Ω, P) un espace probabilisé fini et X et Y deux variables aléatoires sur Ω. La loi (conjointe) $P_{(X,Y)}$ du couple (X, Y) détermine entièrement ses lois marginales P_X et P_Y. Précisément, pour tout $x \in X(\Omega)$:

$$P(X = x) = \sum_{y \in Y(\Omega)} P(X = x \text{ et } Y = y) \quad \text{et pour tout } y \in Y(\Omega) : \quad P(Y = y) = \sum_{x \in X(\Omega)} P(X = x \text{ et } Y = y).$$

Ce résultat se généralise sans difficulté aux familles d'un nombre fini quelconque de variables aléatoires.

On représente parfois la loi de (X, Y) comme un tableau à deux entrées donnant $P(X = x \text{ et } Y = y)$ en fonction des valeurs possibles de $x \in X(\Omega)$ en lignes et $y \in Y(\Omega)$ en colonnes. La figure ci-contre illustre la manière dont on peut calculer P_X et P_Y à partir de $P_{(X,Y)}$. On a noté :

$$X(\Omega) = \{x_1, \dots, x_m\}, \quad Y(\Omega) = \{y_1, \dots, y_n\}$$

et pour tout $(i, j) \in [\![1, m]\!] \times [\![1, n]\!]$:

$$p_{ij} = P\left(X = x_i \text{ et } Y = y_j\right)$$

$X \backslash Y$	y_1	y_2	$\cdots$	y_n	
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$\rightarrow \sum_{j=1}^{n} p_{1j} = P(X = x_1)$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$\rightarrow \sum_{j=1}^{n} p_{2j} = P(X = x_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\cdots$	p_{mn}	$\rightarrow \sum_{j=1}^{n} p_{mj} = P(X = x_m)$

Première loi marginale P_X

$= y_1) = \sum_{i=1}^{m} p_{i1} \quad \cdots \quad \sum_{i=1}^{m} p_{in} = P(Y = y_n)$

Deuxième loi marginale P_Y

Nous avons déjà appris à calculer la loi d'une variable aléatoire de la forme $f(Z)$. Dans le cas d'un couple $Z = (X, Y)$, on obtient l'énoncé suivant.

Théorème 54 (Loi de $f(X, Y)$). Soient (Ω, P) un espace probabilisé fini, F un ensemble, X et Y deux variables aléatoires sur Ω et $f : (X, Y)(\Omega) \longrightarrow F$ une fonction. La loi $P_{f(X,Y)}$ de $f(X, Y)$ est entièrement déterminée par f et la loi conjointe de (X, Y). En l'occurrence, pour tout $z \in f(X, Y)(\Omega)$: $\quad P(f(X, Y) = z) = \sum_{\substack{(x,y) \in (X,Y)(\Omega) \\ f(x,y)=z}} P(X = x \text{ et } Y = y).$

Ce résultat se généralise sans difficulté aux familles d'un nombre fini quelconque de variables aléatoires.

Cas particulier le plus courant :

$$\boxed{P(X + Y = z) = \sum_{\substack{(x,y) \in (X,Y)(\Omega) \\ x+y=z}} P(X = x \text{ et } Y = y).}$$

Théorème 55 (Sommes de variables aléatoires indépendantes de lois binomiales). Soient $p \in [0, 1]$, $m, n \in \mathbb{N}^*$ et $X, Y, X_1, \dots, X_n$ des variables aléatoires définies sur un même espace probabilisé fini.

(i) Si : $X \sim \mathscr{B}(m, p)$ et $Y \sim \mathscr{B}(n, p)$ et si X et Y sont INDÉPENDANTES, alors : $X + Y \sim \mathscr{B}(m + n, p)$.

(ii) Si : $\quad X_i \sim \mathscr{B}(p) \quad$ pour tout $i \in [\![1, n]\!]$ et si $X_1, \dots, X_n$ sont INDÉPENDANTES, alors : $\sum_{i=1}^{n} X_i \sim \mathscr{B}(n, p)$. Bref, quand on répète une expérience aléatoire à deux issues plusieurs fois indépendamment, si X compte le nombre de succès obtenus après m essais et Y le nombre de succès obtenus après n nouveaux essais, $X + Y$ compte le nombre de succès obtenus après $m + n$ essais !

Exemple 56. On lance simultanément 2 dés équilibrés à 6 faces, l'un blanc, l'autre noir. La valeur obtenue avec le dé blanc (resp. noir) est notée X (resp. Y). Le couple (X, Y) suit la loi uniforme sur $[\![1, 6]\!]^2$ et nous allons tâcher de déterminer la loi de la somme $S = X + Y$, la loi conditionnelle de X sachant $\{S = 4\}$ et la loi de l'écart $|X - Y|$.

Exemple 57. Dans un centre d'appels, un employé effectue n appels téléphoniques vers n correspondants distincts dont chacun décroche avec probabilité p.

- On note N_1 le nombre de correspondants qui ont décroché. SANS CALCUL : $N_1 \sim \mathscr{B}(n,p)$, car les appels sont indépendants et la probabilité d'obtenir un correspondant ne dépend pas du correspondant choisi.
- L'employé rappelle un peu plus tard les $n - N_1$ correspondants qui n'ont pas décroché lors de sa première série d'appels. On note N_2 le nombre de ces correspondants qui décrochent cette fois et $N = N_1 + N_2$ le nombre total des correspondants qui ont décroché. Quelle est la loi de N ?

7 Dispersion d'une variable aléatoire

Dans tout ce qui suit, (Ω, P) est un univers probabilisé fini.

7.1 Espérance

Ce sont les notions de moyenne (pondérée) et d'intégrale (discrète) qui sont transcrites dans la théorie des probabilités.
Dans la suite, les variables aléatoires sont à valeurs réelles ou complexes.

Définition 58. Soit X une variable aléatoire sur Ω telle que $X(\Omega) = \{x_1, \dots, x_n\}$, on définit $E(X)$, **l'espérance de** X, par

$$E(X) = \sum_{k=1}^{n} x_k P(X = x_k) = \sum_{x \in X(\Omega)} x P(X = x)$$

Si $E(X) = 0$, on dit que X est **centrée**.

Comme la somme des probabilités des x_k est 1, on voit que $E(X)$ est la moyenne des x_k pondérée par les probabilités respectives.
On peut voir aussi $E(X)$ comme un centre de gravité du système des points x_k disposés sur la droite réelle pondérés par les masses $P(X = x_k)$.

C'est aussi l'intégrale sur $[0,1]$ de la fonction en escalier qui vaut x_k sur l'intervalle $\left[\sum_{i=1}^{k-1} P(X = x_i), \sum_{i=1}^{k} P(X = x_i)\right]$.

Exercice 59. Un dé à 6 faces numérotées de 1 à 6 a été truqué de telle sorte que les faces 1, 2 et 3 tombent avec probabilité $\dfrac{1}{6}$, les faces 4 et 5 avec probabilité $\dfrac{1}{12}$ et la face 6 avec probabilité $\dfrac{1}{3}$. Quel numéro obtient-on en moyenne ?

L'espérance est un indicateur de position de la variable aléatoire.

Proposition 60 (Espérance des lois usuelles). Soit X une variable aléatoire sur Ω.

(i) **Variables aléatoires constantes.** Si X est constante de valeur c, alors $E(X) = c$.

(ii) **Loi uniforme** Soit $E = \{x_1, \dots, x_n\}$ une partie de $\mathbb{C}$. Si : $X \sim \mathscr{U}(E)$, $E(X)$ est la moyenne « naturelle » des valeurs $x_1, \dots, x_n$ de X : $\quad E(X) = \dfrac{1}{n} \sum_{k=1}^{n} x_k$.

(iii) **Loi de Bernoulli.** Si X suit une loi de Bernoulli $\mathcal{B}(p)$, alors $E(X) = p$.
Exemple fondamental. Pour tout événement $A \in \mathscr{P}(\Omega)$: $E(\mathbb{1}_A) = P(A)$.

(iv) **Loi de binomiale.** Si X suit une loi binomiale $\mathcal{B}(n,p)$, alors $E(X) = np$.

L'espérance a les mêmes propriétés que l'intégrale :

Proposition 61 (Propriétés de l'espérance). Soient X et Y deux variables aléatoires (réelle ou complexe) sur Ω.

(i) **Autre expression.** $\quad E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$

(ii) **Linéarité.** $\forall \lambda, \mu \in \mathbb{C}, E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y)$

(iii) **Inégalité triangulaire.** $|E(X)| \leq E(|X|)$

(iv) **Positivité.** $X \geq 0 \Longrightarrow E(X) \geq 0$

(v) **Croissance.** $X \leq Y \Longrightarrow E(X) \leq E(Y)$

L'espérance se calcule bien par composition par une application :

Théorème 62 (de transfert). Soit $X : \Omega \to \mathbb{R}$ et f définie sur $X(\Omega)$ alors

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

ainsi l'espérance de $f(X)$ est déterminée non pas par X mais seulement par la loi de X.

Proposition 63 (Inégalité de Markov). Soit X une variable aléatoire réelle positive, et $a \in \mathbb{R}^+$, alors

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

Proposition 64. Si X et Y sont indépendantes, alors l'espérance du produit est le produit des espérances :

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

La réciproque est fausse en général.

Contre-exemple pour la réciproque : prenons X de loi uniforme à valeurs dans $\{-1, 0, 1\}$: on a $E(X) = 0$. Soit Y telle que $Y = 1$ si $X \neq 0$ et $Y = 0$ sinon.
On a toujours $XY = 0$, donc $E(XY) = 0 = E(X)E(Y)$. Or X et Y ne sont pas indépendantes : c'est clair par la définition de Y et on peut s'en rendre compte explicitement : on a $P(X = 0, Y = 0) = 0$, alors que $P(X = 0)P(Y = 0) = \dfrac{1}{3} \times \dfrac{2}{3} \neq 0$.

7.2 Variance et écart type

On s'intéresse maintenant à la dispersion des valeurs d'une variable aléatoire autour de leur moyenne.

Définition 65. Soit X une variable aléatoire réelle, on appelle **variance de (X)**, le réel

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

On appelle **écart type de** X le réel

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

La variance peut être interprétée comme le moment d'inertie du système de points des valeurs de X autour de son centre de gravité (son espérance).

L'écart type peut être interprété comme une mesure de la dispersion des valeurs autour de la moyenne (l'espérance) : c'est une distance moyenne à la moyenne.

Proposition 66 (Formule de König-Huygens). Soit X une variable aléatoire réelle, on a

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

Proposition 67. Soient X une variable aléatoire réelle, a et b des réels, alors

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

Proposition 68 (Variance des lois usuelles).

(i) Si X suit la loi uniforme sur $[\![1, n]\!]$, alors $V(X) = \dfrac{n^2-1}{12}$.

(ii) Si X est de loi de Bernoulli $\mathcal{B}(p)$, alors $V(X) = p(1-p)$.

(iii) Si X est de loi binomiale $\mathcal{B}(n,p)$, alors $V(X) = np(1-p)$.

Théorème 69 (inégalité de Bienaymé-Tchebytchev). Soient X une variable aléatoire réelle et $a \in \mathbb{R}$, $a > 0$, alors

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