Espaces probabilisés finis et variables aléatoires

1 Vocabulaire des 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*, généralement noté Ω. Plutôt que de résultats possibles, on parle aussi 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 $\{N,B\}^2$ des 2-listes de $\{N,B\}$, i.e. l'ensemble des mots de 2 lettres sur l'alphabet $\{N,B\}$.

En MPSI, nos univers seront officiellement des ensembles **finis**.

Nous tricherons 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, mais vous vous y frotterez un peu en deuxième année. L'ennui bien sûr, c'est qu'en limitant la taille des univers, on limite drastiquement la nature des expériences aléatoires autorisées. Il nous sera par exemple impossible en MPSI :

— de jouer à pile ou face indéfiniment et d'étudier la probabilité de voir la séquence pile-pile-face revenir régulièrement,

— de jouer aux fléchettes contre un disque et d'étudier la probabilité d'atterrir dans telle ou telle portion du disque.

Les résultats possibles d'une expérience aléatoire peuvent être appréhendés chacun en tant qu'élément de l'univers Ω, mais ce qui nous intéresse, ce sont plutôt des **ensembles** de résultats. Par exemple, 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 $\{2,4,6\}$. De même, dans le tirage de 2 boules 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 $\{BN,BB\}$. De façon générale, toute partie de Ω est appelée un *événement*. On décrit le plus souvent un événement A en énonçant une proposition $\mathscr{P}$ qui caractérise de ses éléments : $A=\{\omega\in\Omega \mid \mathscr{P}(\omega)\}$, mais en tant qu'objet mathématique, A est un ensemble et non pas une proposition.

ω	Résultat possible, issue, réalisation
$\{\omega\}$	Événement élémentaire
Ω	Univers, événement certain
$\varnothing$	Événement impossible
$\mathscr{P}(\Omega)$	Ensemble de tous les événements

Quand on réalise une expérience aléatoire, on est certain que le résultat obtenu appartient à l'univers Ω. Cela justifie qu'on appelle *événement certain* l'ensemble Ω, et pour une raison semblable, *événement impossible* l'ensemble vide $\varnothing$.

$\overline{A}$	non $\mathscr{P}$
$A\cup B$	$\mathscr{P}$ ou $\mathscr{Q}$
$A\cap B$	$\mathscr{P}$ et $\mathscr{Q}$

À présent, dans le tableau de vocabulaire de gauche, les événements A et B sont définis respectivement par des propositions $\mathscr{P}$ et $\mathscr{Q}$. L'événement $\overline{A}$ est appelé l'*événement contraire de* A et on dit de A et B qu'ils sont *incompatibles* quand ils sont disjoints.

Enfin, un lien important unit les concepts d'inclusion et d'implication. Savoir que $\mathscr{P}\Longrightarrow\mathscr{Q}$ revient en effet à savoir, en termes d'événements, que $A\subset B$.

Exemple On lance une pièce n fois successivement. Pour tout $k\in[\![1,n]\!]$, on note F_k l'événement « On obtient face au $k^{\text{ème}}$ lancer ». Les événements $F_1,\dots,F_n$ sont en quelque sorte les événements les plus basiques auxquels 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 $A=\bigcup_{k=1}^{n}\overline{F_k}$,

— l'événement B « On n'obtient jamais pile » peut être écrit $B=\bigcap_{k=1}^{n}F_k$,

— et l'événement C « On obtient deux faces consécutifs au cours des n lancers » peut être écrit $C=\bigcup_{k=1}^{n-1}\big(F_k\cap F_{k+1}\big)$.

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

Définition **(Système complet d'événements)** Soit Ω un univers fini. On appelle *système complet d'événements de* Ω tout ensemble $\{A_1, \ldots, A_n\}$ d'événements deux à deux incompatibles dont la réunion est l'événement certain, i.e. pour lequel $\Omega = \bigsqcup_{i=1}^{n} A_i$ — réunion **DISJOINTE**, attention !

Exemple 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 parle d'alternative plutôt que de système complet d'événements.

Exemple Soient $\Omega = \{\omega_1, \ldots, \omega_n\}$ un univers fini et $A, B \in \mathscr{P}(\Omega)$. La paire $\{A, \overline{A}\}$ est un système complet d'événements et il n'est pas dur de se convaincre sur quelques patates que $\{A \cap B, A \cap \overline{B}, \overline{A} \cap B, \overline{A} \cap \overline{B}\}$ et $\{A, \overline{A} \cap B, \overline{A} \cap \overline{B}\}$ en sont aussi. C'est enfin le cas de l'ensemble $\big\{\{\omega_1\}, \ldots, \{\omega_n\}\big\}$ des événements élémentaires car $\Omega = \{\omega_1\} \sqcup \ldots \sqcup \{\omega_n\}$.

La plupart du temps, les événements qu'on manipule sont définis par des 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 **(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. Si $E = \mathbb{R}$, on précise généralement que X est une *variable aléatoire réelle*.

Événements construits à partir d'une variable aléatoire : Pour toute partie A de E, l'événement $X^{-1}(A)$ est noté $\{X \in A\}$ ou $(X \in A)$: $\quad \{X \in A\} = X^{-1}(A) = \big\{\omega \in \Omega \mid \ X(\omega) \in A\big\}$.

En particulier, si X est une variable aléatoire *réelle*, i.e. à valeurs dans $\mathbb{R}$, alors pour tout $x \in \mathbb{R}$:
- pour $A = \{x\}$: $\quad \{X = x\} = X^{-1}\big(\{x\}\big) = \big\{\omega \in \Omega \mid \ X(\omega) = x\big\}$,
- pour $A =]-\infty, x]$: $\quad \{X \leqslant x\} = X^{-1}\big(]-\infty, x]\big) = \big\{\omega \in \Omega \mid \ X(\omega) \leqslant x\big\}$.

En dépit de son nom, une variable aléatoire sur Ω n'est ni variable, ni aléatoire, c'est une fonction, et en tant que telle, elle envoie tout élément de Ω sur un élément parfaitement fixé de E. L'aléatoire fera irruption quand nous attribuerons à tout événement de Ω une probabilité, i.e. une vraisemblance.

Exemple

- Pour modéliser le lancer d'un dé à 6 faces 2 fois, on peut choisir pour univers Ω l'ensemble $[\![1, 6]\!]^2$ des 2-listes de $[\![1, 6]\!]$. La fonction X_1 (resp. X_2) « valeur obtenue au premier (resp. 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 $\omega = (2, 5) \in \Omega$: $\quad X_1(\omega) = 2, \quad X_2(\omega) = 5 \quad$ et $\quad S(\omega) = 7$.

- Pour modéliser le tirage simultané de 4 entiers entre 1 et 10, on peut choisir pour univers Ω l'ensemble $\mathscr{P}_4\big([\![1, 10]\!]\big)$ 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 entiers, il est impossible que le plus petit entier tiré vaille 8, 9 ou 10.

  Par exemple, pour $\omega = \{2, 5, 6, 8\} \in \Omega$: $\quad X(\omega) = 2$.

Définition-théorème **(Système complet d'événements associé à une variable aléatoire)** Soient Ω un univers fini et X une variable aléatoire sur Ω. 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 $A \in \mathscr{P}\big(X(\Omega)\big)$: $\quad \{X \in A\} = \bigsqcup_{x \in A} \{X = x\}$.

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 $\bigsqcup$.

Démonstration Tout élément ω de Ω est envoyé par X sur un et un seul élément de $X(\Omega)$, à savoir $X(\omega)$, donc appartient à un et un seul événement $\{X = x\}$ — en l'occurrence pour $x = X(\omega)$. Bref : $\quad \Omega = \bigsqcup_{x \in X(\Omega)} \{X = x\}$. ■

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 (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 :

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

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

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

Exemple 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 :

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

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

Définition (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 Soit $\Omega = \{\omega_1, \ldots, \omega_n\}$ un univers fini. Si on note P la probabilité uniforme sur Ω, les événements élémentaires $\{\omega_1\}, \ldots, \{\omega_n\}$ sont équiprobables de probabilité commune $\dfrac{1}{|\Omega|}$.

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

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

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

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

(iv) **Réunion :** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Si $A_1, \ldots, A_n$ sont deux à deux incompatibles, alors : $P\left(\bigsqcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$ (*additivité*),

mais en toute généralité : $P\left(\bigcup_{k=1}^{n} A_k\right) \leqslant \sum_{k=1}^{n} P(A_k)$ (*sous-additivité*).

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

Démonstration

(i) et (ii) $A = (A \setminus B) \sqcup (A \cap B)$, donc $P(A) = P(A \setminus B) + P(A \cap B)$.

Ainsi : $P(\overline{A}) = P(\Omega \setminus A) = P(\Omega) - P(\Omega \cap A) = 1 - P(A)$, donc $P(\varnothing) = P(\overline{\Omega}) = 1 - P(\Omega) = 1 - 1 = 0$.

(iii) Si $A \subset B$, alors $B = A \sqcup (B \setminus A)$, donc $P(B) = P(A) + P(B \setminus A) \geqslant P(A) + 0 = P(A)$.

(iv) Cette fois $A \cup B = (A \setminus B) \sqcup B$, donc $P(A \cup B) = P(A \setminus B) + P(B) \overset{\text{(ii)}}{=} P(A) + P(B) - P(A \cap B) \leqslant P(A) + P(B)$.

On étend ensuite ce premier cas de sous-additivité à davantage d'événements par récurrence. La formule du crible sera quant à elle prouvée au chapitre « Position et dispersion d'une variable aléatoire réelle ».

Au-delà de l'exemple naturel des probabilités uniformes, notre prochain théorème décrit 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 la probabilité P choisie 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 (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 Soit (Ω, P) un espace probabilisé fini. La famille $\big(P(\{\omega\})\big)_{\omega \in \Omega}$ des probabilités des événements élémentaires de Ω est une distribution de probabilités sur Ω car $\displaystyle\sum_{\omega \in \Omega} P(\{\omega\}) = P\left(\bigsqcup_{\omega \in \Omega} \{\omega\}\right) = P(\Omega) = 1$.

Théorème (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 $A \in \mathscr{P}(\Omega)$: $\displaystyle P(A) = \sum_{\omega \in A} p_\omega$.

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

Démonstration

- **Analyse :** Soit P une probabilité sur Ω pour laquelle $P(\{\omega\}) = p_\omega$ pour tout $\omega \in \Omega$. Pour tout $A \in \mathscr{P}(\Omega)$:
$\displaystyle P(A) = P\left(\bigsqcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p_\omega = \sum_{\omega \in \Omega} \mathbb{1}_A(\omega) p_\omega$.
- **Synthèse :** Posons pour tout $A \in \mathscr{P}(\Omega)$: $\displaystyle P(A) = \sum_{\omega \in \Omega} \mathbb{1}_A(\omega) p_\omega$.
  - Pour tout $A \in \mathscr{P}(\Omega)$: $P(A) \in [0,1]$ car $\displaystyle 0 \leqslant P(A) \leqslant \sum_{\omega \in \Omega} p_\omega = 1$.
  - Ensuite $\displaystyle P(\Omega) = \sum_{\omega \in \Omega} \mathbb{1}_\Omega(\omega) p_\omega = \sum_{\omega \in \Omega} p_\omega = 1$.
  - Également, pour tout $\omega \in \Omega$: $\displaystyle P(\{\omega\}) = \sum_{\omega' \in \Omega} \mathbb{1}_{\{\omega\}}(\omega') p_{\omega'} = p_\omega$.
  - Enfin, pour tous $A, B \in \mathscr{P}(\Omega)$ incompatibles : $\mathbb{1}_{A \sqcup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A \cap B} = \mathbb{1}_A + \mathbb{1}_B$, donc :
  $$P(A \sqcup B) = \sum_{\omega \in \Omega} \mathbb{1}_{A \sqcup B}(\omega) p_\omega = \sum_{\omega \in \Omega} \mathbb{1}_A(\omega) p_\omega + \sum_{\omega \in \Omega} \mathbb{1}_B(\omega) p_\omega = P(A) + P(B).$$

Exemple On lance une fois un dé pipé à 6 faces qui fournit le numéro 1 avec probabilité $\frac{1}{4}$ et les autres faces avec une même probabilité p — ainsi $\frac{1}{4} + 5p = 1$, 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 un numéro impair » 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 (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 $A \overset{P_X}{\longmapsto} P(X \in A)$ de $\mathscr{P}\big(X(\Omega)\big)$ dans $[0,1]$.

- L'application P_X est une probabilité sur $X(\Omega)$.
- Cette probabilité est entièrement déterminée par la distribution de probabilités $\big(P(X=x)\big)_{x \in X(\Omega)}$ sur $X(\Omega)$, qu'on appelle justement la *distribution de probabilités de X*.
  En l'occurrence, pour tout $A \in \mathscr{P}\big(X(\Omega)\big)$: $\quad P_X(A) = P(X \in A) = \sum_{x \in A} P(X = x).$

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.

Démonstration Montrons que P_X est une probabilité sur $X(\Omega)$. Or $P_X\big(X(\Omega)\big) = P\big(X \in X(\Omega)\big) = P(\Omega) = 1$, et pour tous $A, B \in \mathscr{P}\big(X(\Omega)\big)$ incompatibles : $\quad P_X(A \sqcup B) = P(X \in A \sqcup B) = P\Big(\big\{X \in A\big\} \sqcup \big\{X \in B\big\}\Big)$

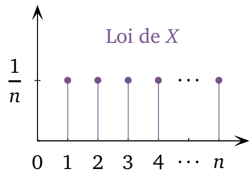
$$= P(X \in A) + P(X \in B) = P_X(A) + P_X(B).$$

Par ailleurs, pour tout $A \in \mathscr{P}\big(X(\Omega)\big)$: $\quad P_X(A) = P(X \in A) = P\left(\bigsqcup_{x \in A} \big\{X = x\big\}\right) = \sum_{x \in A} P(X = x).$ ■

Définition (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 $A \in \mathscr{P}(E)$: $\quad P_X(A) = P(X \in A) = \dfrac{|A|}{|E|}$.

La loi uniforme est la loi pour laquelle les événements $\big\{X = x\big\}$, x décrivant E, sont équiprobables de probabilité $\dfrac{1}{|E|}$.

La formule $P(X \in A) = \dfrac{|A|}{|E|} = \dfrac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$ est essentielle en pratique et ramène tout problème de calcul sur la loi uniforme à un problème de dénombrement.



Exemple

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

Exemple Une urne contient 3 boules blanches et 5 noires. On en tire simultanément 4 boules. Avec quelle probabilité n'a-t-on 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 $\mathscr{P}_4\big([\![1,8]\!]\big)$ des 4-combinaisons de $[\![1,8]\!]$ et la probabilité cherchée vaut :

$$P\Big(X \in \mathscr{P}_4\big([\![4,8]\!]\big)\Big) = \frac{\Big|\mathscr{P}_4\big([\![4,8]\!]\big)\Big|}{\Big|\mathscr{P}_4\big([\![1,8]\!]\big)\Big|} = \frac{\binom{5}{4}}{\binom{8}{4}} = \frac{1}{14}.$$

Définition (Loi de Bernoulli) Soient X une variable aléatoire à valeurs dans $\big\{0,1\big\}$ 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 $A \in \mathscr{P}(\Omega)$: $\quad \mathbb{1}_A \sim \mathscr{B}\big(P(A)\big)$.

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

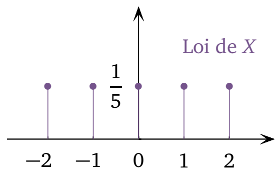
On s'intéresse à présent à la loi de l'image d'une variable aléatoire par une fonction. Commençons par un exemple.

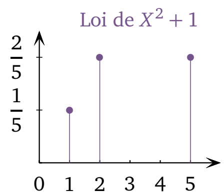
Exemple Soit X une variable uniforme sur $[\![-2,2]\!]$. Quelle est la loi de la variable aléatoire X^2+1 ?

Démonstration Par hypothèse, pour tout $k \in [\![-2,2]\!]$: $P(X=k) = \dfrac{1}{5}$ et la variable X^2+1 a pour image $\{1,2,5\}$. Ensuite : $P\big(X^2+1=1\big) = P\big(X^2=0\big) = P(X=0) = \dfrac{1}{5}$, puis :

$$P\big(X^2+1=2\big) = P\big(X^2=1\big) = P(X=1 \text{ ou } X=-1) = P(X=1)+P(X=-1) = \frac{2}{5},$$

et par un calcul analogue : $P\big(X^2+1=5\big) = \dfrac{2}{5}$.





Théorème (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 $A \in \mathscr{P}\big(f\big(X(\Omega)\big)\big)$: $\displaystyle P_{f(X)}(A) = P\big(f(X)\in A\big) = \sum_{\substack{x\in X(\Omega)\\ f(x)\in A}} P(X=x)$.

En particulier, si deux variables aléatoires X et Y ont la même loi, les variables $f(X)$ et $f(Y)$ ont aussi la même loi pour peu que la composition par f ait un sens.

Démonstration Pour tout $A \in \mathscr{P}\big(f\big(X(\Omega)\big)\big)$, les éléments de l'événement $\big\{f(X)\in A\big\}$ peuvent être regroupés en fonction de la valeur que X leur donne :

$$\big\{f(X)\in A\big\} = \big\{f(X)\in A\big\}\cap\Omega = \big\{f(X)\in A\big\}\cap \bigsqcup_{x\in X(\Omega)} \{X=x\} = \bigsqcup_{x\in X(\Omega)} \big\{f(X)\in A \text{ et } X=x\big\}$$
$$= \bigsqcup_{x\in X(\Omega)} \big\{f(x)\in A \text{ et } X=x\big\} = \bigsqcup_{\substack{x\in X(\Omega)\\ f(x)\in A}} \{X=x\}.$$

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

Démonstration Pour tout $k \in [\![1,2n]\!]$: $P(X=k) = \dfrac{1}{2n}$ et la variable $(-1)^X$ a pour image $\{-1,1\}$. Elle suit précisément la loi de Rademacher, i.e. la loi uniforme sur $\{-1,1\}$ car :

$$P\big((-1)^X=1\big) = \sum_{i=1}^{n} P(X=2i) = \sum_{i=1}^{n}\frac{1}{2n} = \frac{1}{2} \qquad \text{et} \qquad P\big((-1)^X=-1\big) = \sum_{i=0}^{n-1} P(X=2i+1) = \sum_{i=0}^{n-1}\frac{1}{2n} = \frac{1}{2}.$$

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 autrement le même genre de situations 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 (…) ». Mais quel sens cela a-t-il d'introduire une variable aléatoire hors sol sans définir au préablable un espace probabilisé susceptible de la porter ? Qui nous garantit que nos désirs de variables sont des réalités ? 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 $P(X=e) = p_e$ pour tout $e \in E$.

Démonstration On cherche un exemple satisfaisant, rien qu'un, d'espace probabilisé (Ω, P) et de variable aléatoire X. Posons $\Omega = E$ et $X = \mathrm{Id}_E$ et notons P l'unique probabilité sur Ω pour laquelle $P\big(\{\omega\}\big) = p_\omega$ pour tout $\omega \in \Omega$. Comme voulu, X admet E pour image et pour tout $e \in E$: $P(X = e) = P\big(\{e\}\big) = p_e$.

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 les porte. Voué à rester caché, celui-ci ne présente de toute façon aucun caractère d'unicité, de nombreux choix d'espace probabilisé sont possibles pour la description d'une même situation. Par exemple, 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. Changera-t-on d'univers chaque fois qu'on change de question ? Bien sûr que non, car nous sommes au fond toujours confrontés 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 pas 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).

Exemple 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 ?

Démonstration L'égalité $\displaystyle\sum_{k=1}^{n} \frac{2}{3n} + \sum_{k=1}^{n} \frac{1}{3n} = 1$ justifie qu'on puisse se donner une variable aléatoire X, représentant le numéro de la boule tirée, pour laquelle $P(X = 2k) = \dfrac{2}{3n}$ et $P(X = 2k-1) = \dfrac{1}{3n}$ pour tout $k \in [\![1,n]\!]$. Dans ce cadre : $\displaystyle P(X \text{ est pair}) = \sum_{k=1}^{n} P(X = 2k) = \sum_{k=1}^{n} \frac{2}{3n} = \frac{2}{3}$.

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. Par exemple, sous cette hypothèse, B est réalisé avec probabilité 1 et $\overline{B}$ avec probabilité 0 alors que rien ne dit au départ que $P(B) = 1$. En fait, P ne nous offre qu'un point de vue initial sur les événements, sans hypothèse. En guise de nouveau point de vue, on appelle *probabilité conditionnelle sachant B*, notée P_B, la probabilité sur Ω qui mesure la vraisemblance des événements sous l'hypothèse que B est réalisé. C'est pour cette nouvelle probabilité que $P_B(B) = 1$ et $P_B\big(\overline{B}\big) = 0$.

C'est très bien, mais comment calcule-t-on les valeurs de P_B en toute généralité ? Mesurer la vraisemblance d'un événement A sous l'hypothèse que B est réalisé, c'est évidemment s'intéresser à $P(A \cap B)$, mais l'application $A \longmapsto P(A \cap B)$ n'a pas toutes les qualités requises, elle joue trop petit. Au lieu de mesurer la vraisemblance des événements sur une échelle de 0 à 1, elle le fait petitement sur une échelle de 0 à $P(B)$. Qu'à cela ne tienne, normalisons-la en posant $P_B(A) = \dfrac{P(A \cap B)}{P(B)}$ pour tout événement A, sous l'hypothèse technique que $P(B) > 0$. Cette définition fait de P_B une probabilité indifférente à tous les résultats possibles dans Ω qui pourraient contredire B. Et ça tombe bien : $P_B(B) = 1$ et $P_B\big(\overline{B}\big) = 0$.

Définition (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 $P(A|B) = P_B(A) = \dfrac{P(A \cap B)}{P(B)}$ 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 $P(A \cap B) = P(B)\,P_B(A)$ est tenue pour vraie dans ce cas. De fait, si $P(B) = 0$, alors $P(A \cap B) \leqslant P(B)$ par croissance de P, donc $P(A \cap B) = 0$.

Démonstration Montrons que P_B est une probabilité sur Ω. Pour tout $A \in \mathscr{P}(\Omega)$: $0 \leqslant P(A \cap B) \leqslant P(B)$ par croissance de P, donc $P_B(A) \in [0,1]$. Ensuite $P_B(\Omega) = \dfrac{P(\Omega \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$. Enfin, pour tous $A, A' \in \mathscr{P}(\Omega)$ incompatibles, $A \cap B$ et $A' \cap B$ sont également incompatibles, donc :

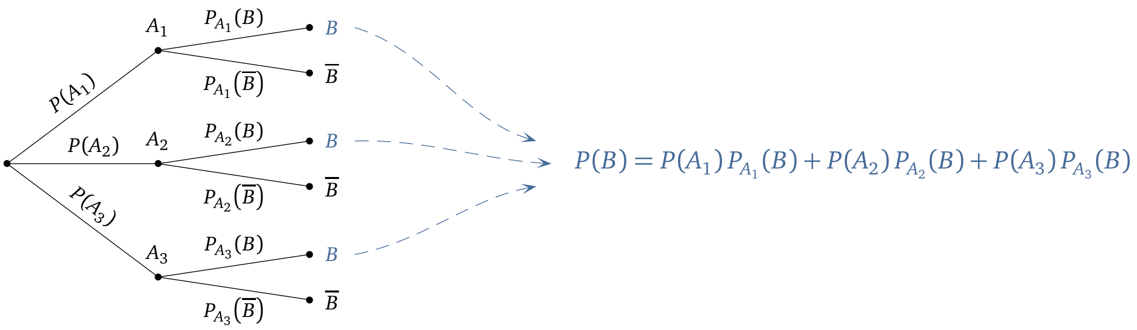
$$P_B(A \sqcup A') = \frac{P\big((A \sqcup A') \cap B\big)}{P(B)} = \frac{P\big((A \cap B) \sqcup (A' \cap B)\big)}{P(B)} = \frac{P(A \cap B) + P(A' \cap B)}{P(B)} = P_B(A) + P_B(A').$$

Théorème (Formule des probabilités totales) Soient (Ω, P) un espace probabilisé fini et $\{A_1, \dots, A_n\}$ un système complet d'événements de Ω. Pour tout $B \in \mathscr{P}(\Omega)$: $\quad P(B) = \sum_{i=1}^{n} P(A_i)\, P_{A_i}(B)$.

Démonstration

$$B = B \cap \Omega = B \cap \bigsqcup_{i=1}^{n} A_i = \bigsqcup_{i=1}^{n} (B \cap A_i) \quad \text{donc } P(B) = P\left(\bigsqcup_{i=1}^{n} (B \cap A_i)\right) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(A_i)\, P_{A_i}(B).$$

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.

Exemple 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 ?

Démonstration Notons B_1 (resp. B_2) l'événement « La première (resp. deuxième) boule tirée est blanche ». Nous cherchons $P(B_2)$. Trois probabilités sont immédiatement connues : $P(B_1) = \frac{b}{n+b}$, $P_{B_1}(B_2) = \frac{b-1}{n+b-1}$ et $P_{\overline{B_1}}(B_2) = \frac{b}{n+b-1}$. Mais $\{B_1, \overline{B_1}\}$ est un système complet d'événements, donc d'après la formule des probabilités totales : $P(B_2) = P(B_1)\, P_{B_1}(B_2) + P\big(\overline{B_1}\big)\, P_{\overline{B_1}}(B_2) = \frac{b}{n+b} \times \frac{b-1}{n+b-1} + \frac{n}{n+b} \times \frac{b}{n+b-1} = \frac{b}{n+b}$. En d'autres termes, les événements B_1 et B_2 sont équiprobables. Est-ce étonnant après coup ?

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

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

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

Démonstration Pour l'assertion (i) : $\quad P_B(A) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\, P_A(B)}{P(B)}$. L'assertion (ii) n'est quant à elle qu'un mélange peu subtil d'assertion (i) et de formule des probabilités totales.

Dans l'assertion (ii), pour tout $j \in [\![1, n]\!]$, $P(A_j)$ est souvent qualifiée de *probabilité a priori* et $P_B(A_j)$ de *probabilité a posteriori*. Avec cette terminologie, la formule de Bayes permet le passage de probabilités a priori à des probabilités a posteriori. Ce qu'il faut ici comprendre, c'est qu'une probabilité n'est pas tant une mesure du hasard dans une situation donnée qu'une mesure de l'information dont on dispose sur cette situation. Si on connaît les probabilités a priori $P(A_1), \dots, P(A_n)$ et si tout à coup on apprend que l'événement B est réalisé, la formule de Bayes nous permet d'actualiser notre perception des événements $A_1,\ldots,A_n$ en nous donnant accès aux probabilités conditionnelles $P_B(A_1),\ldots,P_B(A_n)$ — connaissance plus fine puisque le conditionnement restreint le champ des possibles. La formule de Bayes est une formule de *révision des croyances*.

Exemple 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 non la compagnie des taxis bleus ?

Démonstration Notons B l'événement « Le taxi coupable est bleu » et T_B l'événement « Le témoin croit savoir que le taxi coupable est bleu ». Je déciderai du sort de la compagnie des taxis bleus après avoir calculé $P_{T_B}(B)$ (probabilité a posteriori).

- De quelles informations disposons-nous ? D'une part : $P\left(\overline{B}\right)=\dfrac{90}{100}=\dfrac{9}{10}$ et $P(B)=\dfrac{1}{10}$ (probabilités a priori), et d'autre part : $P_B(T_B)=\dfrac{90}{100}=\dfrac{9}{10}$ et $P_{\overline{B}}\left(\overline{T_B}\right)=\dfrac{80}{100}=\dfrac{8}{10}$. Cette dernière égalité repose sur l'hypothèse que le témoin ne pouvait hésiter qu'entre le vert et le bleu puisque tous les taxis sont verts ou bleus. Il en découle que $P_{\overline{B}}(T_B)=1-P_{\overline{B}}\left(\overline{T_B}\right)=\dfrac{2}{10}$. Finalement, d'après la formule de Bayes utilisée avec le système complet d'événements $\left\{B,\overline{B}\right\}$:
$$P_{T_B}(B)=\frac{P(B)\,P_B(T_B)}{P(B)\,P_B(T_B)+P\left(\overline{B}\right)P_{\overline{B}}(T_B)}=\frac{\dfrac{1}{10}\times\dfrac{9}{10}}{\dfrac{1}{10}\times\dfrac{9}{10}+\dfrac{9}{10}\times\dfrac{2}{10}}=\frac{1}{3}.$$
Conclusion : le témoignage du témoin est invalidé, je peux blanchir la compagnie des taxis bleus.
- Une remarque pour finir. Si le témoin avait été fiable à 95% pour la couleur verte au lieu de 80%, on aurait obtenu $P_{T_B}(B)=\dfrac{2}{3}$ et là, j'aurais été bien embêté pour rendre mon jugement !

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

Démonstration Si $P(A_1\cap\ldots\cap A_{n-1})=0$, alors $P(A_1\cap\ldots\cap A_n)=0$ par croissance de P et le résultat s'écrit $0=0$. Supposons désormais $P(A_1\cap\ldots\cap A_{n-1})>0$. Dans ce cas, par croissance de P : $P(A_1\cap\ldots\cap A_i)>0$ pour tout $i\in[\![1,n-1]\!]$, donc en vertu d'une simplification télescopique :
$$P(A_1)\,P_{A_1}(A_2)\ \ldots\ P_{A_1\cap\ldots\cap A_{n-1}}(A_n)=P(A_1)\prod_{k=2}^{n}P_{A_1\cap\ldots\cap A_{k-1}}(A_k)=P(A_1)\prod_{k=2}^{n}\frac{P(A_1\cap\ldots\cap A_k)}{P(A_1\cap\ldots\cap A_{k-1})}=P(A_1\cap\ldots\cap A_n).$$

Exemple 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) ?

Démonstration Pour tout $i\in[\![1,3]\!]$, notons B_i l'événement « La $i^{\text{ème}}$ boule tirée est blanche ». Nous cherchons $P\left(\overline{B_1}\cap B_2\cap\overline{B_3}\right)$. D'après la formule des probabilités composées :

— pour des tirages avec remise : $P\left(\overline{B_1}\cap B_2\cap\overline{B_3}\right)=P\left(\overline{B_1}\right)P_{\overline{B_1}}(B_2)\,P_{\overline{B_1}\cap B_2}\left(\overline{B_3}\right)=\dfrac{n}{2n}\times\dfrac{n}{2n}\times\dfrac{n}{2n}=\dfrac{1}{8}$,

— pour des tirages sans remise : $P\left(\overline{B_1}\cap B_2\cap\overline{B_3}\right)=\dfrac{n}{2n}\times\dfrac{n}{2n-1}\times\dfrac{n-1}{2n-2}=\dfrac{n}{4(2n-1)}$.

Lorsque n est très grand, ces deux calculs conduisent à peu près au même résultat, ce qui n'est pas étonnant. Dans ce cas, en effet, la non-remise des boules tirées modifie à peine l'équilibre des couleurs au sein de l'urne.

Exemple 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 ?

Démonstration Nous noterons G l'événement « L'un au moins des tickets achetés est gagnant », d'événement contraire $\overline{G}$ « Tous les tickets achetés sont perdants ». Nous allons calculer $P(G)$ grâce à $P\left(\overline{G}\right)$.

- **Première preuve à l'aide d'une variable aléatoire :** On peut numéroter sans perte de généralité les tickets gagnants 1, 2 et 3 et les tickets perdants de 4 à 50. Notons alors X la 6-combinaison de $[\![1,50]\!]$ des tickets achetés. Clairement $X\sim\mathscr{U}\left(\mathscr{P}_6\left([\![1,50]\!]\right)\right)$ et $\overline{G}=\left\{X\in\mathscr{P}_6\left([\![4,50]\!]\right)\right\}$, donc :

$$P\left(\overline{G}\right)=\frac{\left|\mathscr{P}_6\left([\![4,50]\!]\right)\right|}{\left|\mathscr{P}_6\left([\![1,50]\!]\right)\right|}=\frac{\binom{47}{6}}{\binom{50}{6}}=\frac{47\times 46\times 45\times 44\times 43\times 42}{50\times 49\times 48\times 47\times 46\times 45}=\frac{44\times 43\times 42}{50\times 49\times 48}=\frac{473}{700},\qquad \text{donc } P(G)=\frac{227}{700}.$$

- **Deuxième preuve sans variable aléatoire :** On peut considérer sans perte de généralité que les 6 tickets achetés l'ont été dans un certain ordre et pour tout $i\in[\![1,6]\!]$, noter T_i l'événement « Le $i^{\text{ème}}$ ticket acheté est perdant ». Dans ces conditions : $P_{T_1\cap\ldots\cap T_{i-1}}(T_i)=\dfrac{48-i}{51-i}$ pour tout $i\in[\![1,6]\!]$, car sous l'hypothèse que les $i-1$ premiers tickets achetés ont été perdants, il reste au commerçant $51-i$ tickets à vendre dont $48-i$ perdants. Finalement, d'après la formule des probabilités composées : $P(G)=\dfrac{227}{700}$ car :
$$P\left(\overline{G}\right)=P(T_1\cap\ldots\cap T_n)=P(T_1)P_{T_1}(T_2)\ldots P_{T_1\cap\ldots\cap T_{n-1}}(T_n)=\prod_{i=1}^{6}\frac{48-i}{51-i}=\frac{473}{700}.$$

Nos deux preuves ont conduit au même résultat, mais la première offre un cadre formel plus satisfaisant. Nous avons en effet démontré proprement l'existence d'un espace probabilisé fini susceptible d'accueillir une variable aléatoire de loi uniforme sur $\mathscr{P}_6\left([\![1,50]\!]\right)$. Il est au contraire plus délicat de justifier les calculs de probabilités conditionnelles que nous avons effectués dans la deuxième preuve. Comment construire un espace probabilisé fini adapté ? C'est faisable, mais il vaut mieux privilégier le point de vue des variables aléatoires.

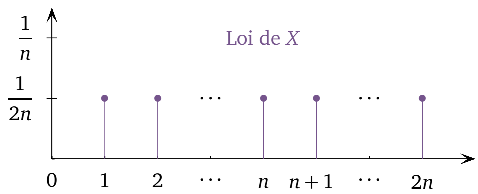
Nous achèverons ce paragraphe par un cas particulier de conditionnement appliqué au concept de variable aléatoire.

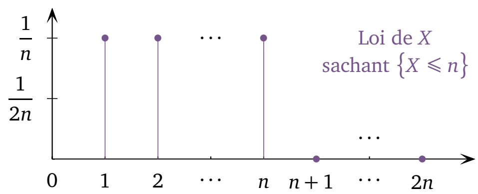
Définition (Lois conditionnelles d'une variable aléatoire) Soient (Ω,P) un espace probabilisé fini, X une variable aléatoire sur Ω et $A\in\mathscr{P}(\Omega)$ pour lequel $P(A)>0$. On appelle *loi conditionnelle de X sachant A (pour la probabilité P)* la loi de X pour la probabilité P_A, i.e. l'application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

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

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

Démonstration Par définition $X\sim\mathscr{U}\big([\![1,2n]\!]\big)$, donc $P(X\leqslant n)=\displaystyle\sum_{k=1}^{n}P(X=k)=\sum_{k=1}^{n}\frac{1}{2n}=\frac{1}{2}$. Ainsi, pour tout $k\in[\![1,n]\!]$: $P_{\{X\leqslant n\}}(X=k)=\dfrac{P(X=k)}{P(X\leqslant n)}=\dfrac{1}{n}$ et pour tout $k\in[\![n+1,2n]\!]$: $P_{\{X\leqslant n\}}(X=k)=0$, donc la loi conditionnelle de X sachant $\{X\leqslant n\}$ est la loi uniforme sur $[\![1,n]\!]$.





5 Indépendance

Intuitivement, étant donnés deux événements A et B pour lesquels $P(B)>0$, 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 (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 $[\![1,n]\!]$: $P\left(\displaystyle\bigcap_{i\in I}A_i\right)=\prod_{i\in I}P(A_i)$.

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 avec son « POUR TOUTE PARTIE I DE $[\![1,n]\!]$ » et demande 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 $\Omega = [\![1,4]\!]$ 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 (Indépendance et complémentaires) Soient (Ω, P) un espace probabilisé fini et $A_1,\dots,A_n \in \mathscr{P}(\Omega)$. Si $A_1,\dots,A_n$ sont indépendants, les événements $A_1^\circ,\dots,A_n^\circ$ le sont aussi pour tous $A_1^\circ \in \{A_1,\overline{A_1}\},\dots,A_n^\circ \in \{A_n,\overline{A_n}\}$.

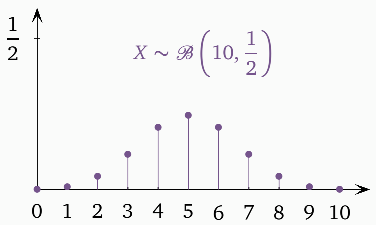
Démonstration Nous nous contenterons du cas de deux événements A et B.

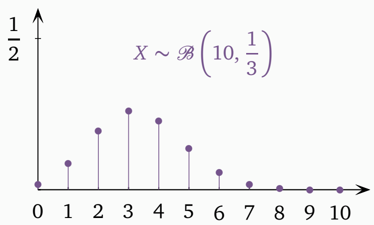
- Si A et B sont indépendants, $\overline{A}$ et B le sont aussi car :
$$P\big(\overline{A}\cap B\big) = P(B\setminus A) = P(B) - P(A\cap B) \overset{\text{Indép.}}{=} P(B) - P(A)\,P(B) = \big(1-P(A)\big)\,P(B) = P\big(\overline{A}\big)\,P(B).$$
- Quitte à permuter A et B, le premier point montre que si A et B sont indépendants, A et $\overline{B}$ le sont aussi, et donc a fortiori $\overline{A}$ et $\overline{B}$ aussi, toujours d'après le premier point.

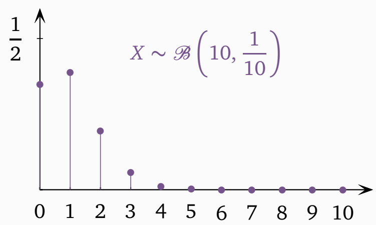
Grâce au concept d'indépendance, nous allons enfin pouvoir définir une loi usuelle très importante en pratique — la *loi binomiale*. Intéressons-nous pour cela à la répétition n fois indépendamment d'une expérience aléatoire à deux issues dites « favorable » et « défavorable » et dont l'issue favorable est de probabilité p. Quelle est la loi du nombre X d'issues favorables ? Pour tout $i \in [\![1,n]\!]$, notons F_i l'événement « La $i^{\text{ème}}$ expérience a une issue favorable ». Les événements $F_1,\dots,F_n$ sont indépendants par hypothèse. La variable aléatoire X est à valeurs dans $[\![0,n]\!]$ et pour tout $k \in [\![0,n]\!]$, 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 $\mathscr{P}_k\big([\![1,n]\!]\big)$ l'ensemble des parties de $[\![1,n]\!]$ de cardinal k : $\displaystyle \{X=k\} = \bigsqcup_{I\in\mathscr{P}_k([\![1,n]\!])} \left(\bigcap_{i\in I} F_i \ \cap \bigcap_{j\in[\![1,n]\!]\setminus I} \overline{F_j}\right)$, donc :
$$P(X=k) = \sum_{I\in\mathscr{P}_k([\![1,n]\!])} P\left(\bigcap_{i\in I} F_i \ \cap \bigcap_{j\in[\![1,n]\!]\setminus I} \overline{F_j}\right) \overset{\text{Indép.}}{=} \sum_{I\in\mathscr{P}_k([\![1,n]\!])} \prod_{i\in I} P(F_i) \prod_{j\in[\![1,n]\!]\setminus I} P\big(\overline{F_j}\big) = \sum_{I\in\mathscr{P}_k([\![1,n]\!])} p^k(1-p)^{n-k} = \binom{n}{k} p^k(1-p)^{n-k}.$$

Prenez le temps d'observer à quel point l'indépendance des événements $F_1,\dots,F_n$ 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 $\left(\dbinom{n}{k} p^k(1-p)^{n-k}\right)_{k\in[\![0,n]\!]}$ est une distribution de probabilités sur $[\![0,n]\!]$.

Définition (Loi binomiale) Soient X une variable aléatoire à valeurs dans $[\![0,n]\!]$ 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 $k \in [\![0,n]\!]$: $\quad P(X=k) = \dbinom{n}{k} p^k(1-p)^{n-k}$.







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 $\mathscr{B}(p)$.

Exemple 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 ?

Démonstration La face noire a une probabilité $\frac{4}{6}=\frac{2}{3}$ d'apparaître quand on lance le dé une fois. Par indépendance des lancers, le nombre N de faces noires obtenues à l'issue des 5 lancers suit donc la loi binomiale $\mathscr{B}\left(5,\frac{2}{3}\right)$, et ainsi : $P(N=3)=\binom{5}{3}\left(\frac{2}{3}\right)^3\left(1-\frac{2}{3}\right)^{5-3}=\frac{80}{243}\approx 0,33.$

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

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

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

$$P(X_1=x_1,\ \dots,\ X_n=x_n)=P(X_1=x_1)\dots P(X_n=x_n),$$

où la notation $\{X_1=x_1,\ \dots,\ X_n=x_n\}$ désigne l'événement $\{X_1=x_1 \text{ et } \dots \text{ et } X_n=x_n\}=\{X_1=x_1\}\cap\dots\cap\{X_n=x_n\}$.

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 !**

Démonstration Faisons l'hypothèse que $P(X_1=x_1,\dots,X_n=x_n)=P(X_1=x_1)\dots P(X_n=x_n)$ pour tous $x_1\in X_1(\Omega),\dots,x_n\in X_n(\Omega)$ et montrons que $X_1,\dots,X_n$ sont indépendantes.

- Montrons d'abord que $P(X_1\in B_1,\ \dots,\ X_n\in B_n)=P(X_1\in B_1)\dots P(X_n\in B_n)$ pour toutes parties B_1 de $X_1(\Omega)$, …, B_n de $X_n(\Omega)$. En notant $\bigstar$ la proposition $(x_1,\dots,x_n)\in B_1\times\dots\times B_n$:

$$\begin{aligned}P(X_1\in B_1)\dots P(X_n\in B_n)&=\left(\sum_{x_1\in B_1}P(X_1=x_1)\right)\dots\left(\sum_{x_n\in B_n}P(X_n=x_n)\right)=\sum_{\bigstar}P(X_1=x_1)\dots P(X_n=x_n)\\&=\sum_{\bigstar}P(X_1=x_1,\ \dots,\ X_n=x_n)=P\left(\bigsqcup_{\bigstar}\{X_1=x_1,\ \dots,\ X_n=x_n\}\right)\\&=P\Big((X_1,\dots,X_n)\in B_1\times\dots\times B_n\Big)=P(X_1\in B_1,\ \dots,\ X_n\in B_n).\end{aligned}$$

- Fixons à présent des parties A_1 de $X_1(\Omega)$… et A_n de $X_n(\Omega)$ ainsi qu'une partie I de $[\![1,n]\!]$. Nous devons montrer que $P\left(\bigcap_{i\in I}\{X_i\in A_i\}\right)=\prod_{i\in I}P(X_i\in A_i)$. Posons pour tout $i\in[\![1,n]\!]$: $B_i=\begin{cases}A_i & \text{si } i\in I\\ X_i(\Omega) & \text{sinon.}\end{cases}$

  D'après le point précédent : $P\left(\bigcap_{i\in I}\{X_i\in A_i\}\right)=P\left(\bigcap_{i=1}^{n}\{X_i\in B_i\}\right)=\prod_{i=1}^{n}P(X_i\in B_i)=\prod_{i\in I}P(X_i\in A_i).$ ■

Exemple 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 ? Les variables aléatoires X_1 et X_2 sont bien sûr indépendantes et uniformes sur $[\![1,6]\!]$, donc :

$$P\big(X_1 \text{ et } X_2 \text{ sont impairs}\big)\overset{\text{Indép.}}{=}P(X_1\text{ est impair})\,P(X_2\text{ est impair})=\frac{\big|\{1,3,5\}\big|}{6}\times\frac{\big|\{1,3,5\}\big|}{6}=\frac{1}{4}.$$

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

Démonstration La variable aléatoire $X_1\dots X_n$ est à valeurs dans $\{0,1\}$ et :

$$P(X_1\dots X_n=1)=P(X_1=1,\ \dots,\ X_n=1)\overset{\text{Indép.}}{=}P(X_1=1)\dots P(X_n=1)=p_1\dots p_n$$

et $P(X_1\dots X_n=0)=1-P(X_1\dots X_n=1)=1-p_1\dots p_n$.

Nous avons déjà justifié tous les énoncés du genre : « Soit X une variable aléatoire de loi (…) ». Cela dit, en réalité, comme on vient de le voir, on a souvent besoin de se donner non pas **UNE** variable aléatoire de loi donnée, mais une **FAMILLE** de variables aléatoires **INDÉPENDANTES** de lois données. Le théorème qui suit nous garantit que c'est toujours possible.

Théorème (Existence d'une famille finie de variables aléatoires indépendantes de lois prescrites) Soient E et F deux ensembles finis, $(p_e)_{e\in E}$ une distribution de probabilités sur E et $(q_f)_{f\in F}$ une distribution de probabilités sur F. Il existe alors un espace probabilié fini (Ω, P) et des variables aléatoires indépendantes X et Y sur Ω, d'images respectives E et F, pour lesquels $P(X=e)=p_e$ pour tout $e\in E$ et $P(Y=f)=q_f$ pour tout $f\in F$.

Cet énoncé se généralise sans difficulté au cas d'un nombre fini de variables aléatoires.

Démonstration Comme $\displaystyle\sum_{e\in E, f\in F} p_e q_f = \left(\sum_{e\in E} p_e\right)\left(\sum_{f\in F} q_f\right) = 1$, la famille $(p_e q_f)_{(e,f)\in E\times F}$ est une distribution de probabilités sur $E\times F$. Il existe donc un espace probabilisé fini (Ω, P) et une variable aléatoire (X,Y) sur Ω d'image $E\times F$ pour laquelle pour tout $(e,f)\in E\times F$: $P(X=e,\ Y=f)=p_e q_f$.

— Pour tout $e\in E$: $\displaystyle P(X=e)=\sum_{f\in F} P(X=e,\ Y=f)=\sum_{f\in F} p_e q_f = p_e$ et pour tout $f\in F$: $P(Y=f)=q_f$.

— En retour, X et Y sont indépendantes car $P(X=e,\ Y=f)=p_e q_f = P(X=e)P(Y=f)$ pour tous $e\in E$ et $f\in F$.

Exemple 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 ?

Démonstration On peut numéroter sans perte de généralité les 2 as de pique 1 et 2 et les autres cartes de 3 à 32. Pour tout $k\in [\![1,n]\!]$, notons X_k la 4-combinaison des cartes tirées au $k^{\text{ème}}$ tirage. Ces variables suivent toutes la loi uniforme sur $\mathscr{P}_4\big([\![1,32]\!]\big)$ et sont de plus indépendantes car les tirages se font avec remise.

- Calculons d'abord pour tout $k\in[\![1,n]\!]$ la probabilité de déceler la supercherie au $k^{\text{ème}}$ tirage, i.e. de l'événement $A_k=\big\{1\in X_k \text{ et } 2\in X_k\big\}$. Or il y a autant de tirages qui réalisent A_k que de 2-combinaisons de l'ensemble $[\![3,32]\!]$ des cartes qui ne sont pas les as de pique, donc : $P(A_k)=\dfrac{\binom{30}{2}}{\binom{32}{4}}=\dfrac{3}{248}$.
- Notons à présent B l'événement « La supercherie finit par être décelée » d'événement contraire « La supercherie n'est jamais décelée » : $\overline{B}=\overline{A_1}\cap\ldots\cap\overline{A_n}$. Parce que les variables aléatoires $X_1,\ldots,X_n$ le sont, les événements $A_1,\ldots,A_n$ sont indépendants, donc les événements $\overline{A_1},\ldots,\overline{A_n}$ aussi, et du coup :
$$P(B)=1-P\big(\overline{B}\big)=1-P\big(\overline{A_1}\cap\ldots\cap\overline{A_n}\big)=1-P\big(\overline{A_1}\big)\ \ldots\ P\big(\overline{A_n}\big)=1-\left(1-\frac{3}{248}\right)^n=1-\left(\frac{245}{248}\right)^n.$$
Finalement : $$P(B)\geqslant 0{,}9 \iff \left(\frac{245}{248}\right)^n\leqslant\frac{1}{10} \iff n\geqslant\frac{\ln 10}{\ln\dfrac{248}{245}} \iff n\geqslant 190.$$

Théorème (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.

Démonstration Dans l'énoncé du théorème, $(X_1,\ldots,X_m)$ est la fonction $\omega\longmapsto\big(X_1(\omega),\ldots,X_m(\omega)\big)$ sur Ω, qui est tout autant que $X_1,\ldots,X_n$ une variable aléatoire, mais à valeurs dans un ensemble produit. Son image $(X_1,\ldots,X_n)(\Omega)$ est l'ensemble des valeurs atteintes par $(X_1,\ldots,X_n)$: $(X_1,\ldots,X_n)(\Omega)\subset X_1(\Omega)\times\ldots\times X_n(\Omega)$, sans égalité en général. Même chose avec $(X_{m+1},\ldots,X_n)$.

Posons $X=(X_1,\ldots,X_m)$ et $Y=(X_{m+1},\ldots,X_n)$.

- Montrons que les variables aléatoires X et Y sont indépendantes. Pour tous $x=(x_1,\ldots,x_m)\in(X_1,\ldots,X_m)(\Omega)$ et $y=(x_{m+1},\ldots,x_n)\in(X_{m+1},\ldots,X_n)(\Omega)$:
$$P(X=x,\ Y=y)=P(X_1=x_1,\ \ldots,\ X_n=x_n)\overset{\text{Indép.}}{=}P(X=x_1)\ldots P(X_n=x_n)$$
$$\overset{\text{Indép.}}{=}P(X_1=x_1,\ \ldots,\ X_m=x_m)P(X_{m+1}=x_{m+1},\ \ldots,\ X_n=x_n)=P(X=x)P(Y=y).$$

- Montrons que $f(X)$ et $g(Y)$ sont indépendantes. Pour tous $a \in f(X)(\Omega)$ et $b \in g(Y)(\Omega)$:

$$P\big(f(X)=a,\ g(Y)=b\big) = P\Big(X \in f^{-1}\big(\{a\}\big),\ Y \in g^{-1}\big(\{b\}\big)\Big) \overset{\text{Indép.}}{=} P\Big(X \in f^{-1}\big(\{a\}\big)\Big) P\Big(Y \in g^{-1}\big(\{b\}\big)\Big) = P\big(f(X)=a\big)P\big(g(Y)=b\big).$$

6 Loi d'une famille de variables aléatoires

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

Définition-théorème (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 $\omega \longmapsto \big(X(\omega), Y(\omega)\big)$ 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 $\big(P(X=x,\ Y=y)\big)_{(x,y)\in(X,Y)(\Omega)}$ sur $(X,Y)(\Omega)$.
- 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.

Exemple 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) ?

Démonstration Le couple (C_1,C_2) est à valeurs dans $\{\mathrm{N},\mathrm{B}\}^2$, mais son image est $\big\{(\mathrm{N},\mathrm{B}),(\mathrm{B},\mathrm{N}),(\mathrm{B},\mathrm{B})\big\}$ seulement car l'urne ne contient qu'une seule boule noire. Ainsi $P(C_1=\mathrm{N},\ C_2=\mathrm{N})=0$. Ensuite :

$$P(C_1=\mathrm{B},\ C_2=\mathrm{B}) = P(C_1=\mathrm{B})P_{\{C_1=\mathrm{B}\}}(C_2=\mathrm{B}) = \frac{3}{4}\times\frac{2}{3} = \frac{1}{2}$$

et de même : $P(C_1=\mathrm{B},\ C_2=\mathrm{N}) = \dfrac{3}{4}\times\dfrac{1}{3}=\dfrac{1}{4}$ et $P(C_1=\mathrm{N},\ C_2=\mathrm{B}) = \dfrac{1}{4}\times\dfrac{3}{3}=\dfrac{1}{4}$.

$C_1 \backslash C_2$	B	N
B	$\frac{1}{2}$	$\frac{1}{4}$
N	$\frac{1}{4}$	0

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 $P(X=x,\ Y=y)$ 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 : $P(X=x,\ Y=y)=P(X=x)P(Y=y)$ 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 $X_1,\ldots,X_n$ sont indépendantes
si et seulement si la distribution de probabilités de $(X_1,\ldots,X_n)$
se calcule par simple produit à partir des distributions de probabilités de $X_1,\ldots,X_n$.

Le petit résultat suivant est hors programme mais éclairant et mérite qu'on s'y attarde. Comment modéliser le lancer d'un dé équilibré à 6 faces 2 fois ? On peut, avec des notations évidentes, se donner deux variables aléatoires indépendantes X_1 et X_2 de loi uniforme sur $[\![1,6]\!]$. On peut aussi se donner directement un couple (X_1,X_2) de loi uniforme sur $[\![1,6]\!]^2$. Ces deux modélisations sont-elles cependant si différentes ? Eh bien non, c'est le sens du théorème que voici.

Théorème (Loi uniforme et produit cartésien) Soient E et F deux ensembles finis non vides et X et Y deux variables aléatoires définies sur un même espace probabilisé fini. Les assertions suivantes sont équivalentes :

(i) $(X,Y)\sim\mathscr{U}(E\times F)$. (ii) X et Y sont indépendantes et $X\sim\mathscr{U}(E)$ et $Y\sim\mathscr{U}(F)$.

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

Démonstration

(i) $\Longrightarrow$ (ii) Pour tout $x \in E$: $P(X=x) = \sum_{y\in F} P(X=x,\ Y=y) = \sum_{y \in F} \frac{1}{|E \times F|} = \frac{1}{|E|}$, donc $X \sim \mathscr{U}(E)$, et de même $Y \sim \mathscr{U}(F)$. A fortiori, X et Y sont indépendantes car pour tous $x \in E$ et $y \in F$:
$$P(X=x,\ Y=y) = \frac{1}{|E \times F|} = \frac{1}{|E|} \times \frac{1}{|F|} = P(X=x)P(Y=y).$$

(ii) $\Longrightarrow$ (i) Le couple (X,Y) est à valeurs dans $E \times F$ et pour tout $(x,y) \in E \times F$:
$$P(X=x,\ Y=y) \overset{\text{Indép.}}{=} P(X=x)P(Y=y) = \frac{1}{|E|} \times \frac{1}{|F|} = \frac{1}{|E \times F|}, \qquad \text{donc } (X,Y) \sim \mathscr{U}(E \times F).$$

À défaut de pouvoir calculer la loi conjointe à partir des lois marginales, on peut **TOUJOURS** faire le calcul inverse.

Théorème (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 $x \in X(\Omega)$:
$$P(X=x) = \sum_{y \in Y(\Omega)} P(X=x,\ Y=y) \qquad \text{et pour tout } y \in Y(\Omega) : \quad P(Y=y) = \sum_{x \in X(\Omega)} P(X=x,\ Y=y).$$
Ce résultat se généralise sans difficulté aux familles d'un nombre fini quelconque de variables aléatoires.

Démonstration Pour tout $x \in X(\Omega)$:
$$\sum_{y \in Y(\Omega)} P(X=x,\ Y=y) = P\left(\bigsqcup_{y \in Y(\Omega)} \{X=x,\ Y=y\} \right) = P\left(\{X=x\} \cap \overbrace{\bigsqcup_{y\in Y(\Omega)} \{Y=y\}}^{=\Omega} \right) = P(X=x).$$

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 :
$$X(\Omega) = \{x_1, \ldots, x_m\}, \qquad Y(\Omega) = \{y_1, \ldots, y_n\}$$
et pour tout $(i,j) \in [\![1,m]\!] \times [\![1,n]\!]$:
$$p_{ij} = P(X=x_i,\ Y=y_j).$$

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

Première loi marginale P_X

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

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 (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 $z \in f(X,Y)(\Omega)$: $\quad P\big(f(X,Y)=z\big) = \sum_{\substack{(x,y)\in(X,Y)(\Omega) \\ f(x,y)=z}} P(X=x,\ Y=y).$

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

Cas particulier le plus courant : $$P(X+Y=z) = \sum_{\substack{(x,y)\in(X,Y)(\Omega) \\ x+y=z}} P(X=x,\ Y=y).$$

Théorème (Sommes de variables aléatoires indépendantes de lois binomiales) Soient $p \in [0,1]$, $m,n \in \mathbb{N}^*$ et $X, Y, X_1, \ldots, X_n$ 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 $X_i \sim \mathscr{B}(p)$ pour tout $i \in [\![1,n]\!]$ et si $X_1, \ldots, X_n$ 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.

Démonstration Il nous suffit bien sûr d'établir l'assertion (i). La variable aléatoire $X+Y$ est à valeurs dans $[\![0,m+n]\!]$ et pour tout $k \in [\![0,m+n]\!]$:

$$P(X+Y=k) = \sum_{i=0}^{k} P(X=i,\ Y=k-i) \overset{\text{Indép.}}{=} \sum_{i=0}^{k} P(X=i)P(Y=k-i)$$
$$= \sum_{i=0}^{k} \binom{m}{i} p^i(1-p)^{m-i} \times \binom{n}{k-i} p^{k-i}(1-p)^{n-k+i} = p^k(1-p)^{m+n-k} \sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}.$$

Or dans l'anneau des polynômes $\mathbb{R}[T]$: $(T+1)^m(T+1)^n = (T+1)^{m+n}$, donc après identification des coefficients de degré k : $\displaystyle\sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i} = \binom{m+n}{k}$ — ce qui généralise la formule de Vandermonde.

Conclusion : $\displaystyle P(X+Y=k) = \binom{m+n}{k} p^k(1-p)^{m+n-k}$, donc en effet $X+Y \sim \mathscr{B}(m+n,p)$.

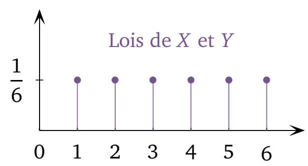
Exemple On lance deux fois de suite un dé équilibré à 6 faces et on note (X,Y) le couple de valeurs obtenues. Ce couple suit la loi uniforme sur $[\![1,6]\!]^2$ et nous allons tâcher de déterminer la loi de la somme $S=X+Y$, la loi conditionnelle de X sachant $\{S=4\}$ et la loi de l'écart $|X-Y|$.

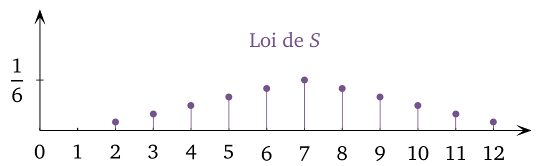
Démonstration

- **Loi de la somme $S=X+Y$:** La variable S est à valeurs dans $[\![2,12]\!]$ et pour tout $s \in [\![2,12]\!]$:

$$P(S=s) = P(X+Y=s) = \sum_{\substack{1\leqslant i,j\leqslant 6\\ i+j=s}} P(X=i,\ Y=j) = \sum_{\substack{1\leqslant i,j\leqslant 6\\ i+j=s}} \frac{1}{36}.$$

Or combien existe-t-il de couples $(i,j) \in [\![1,6]\!]^2$ pour lesquels $i+j=s$? Réponse : autant que de couples $(i,s-i)$ pour lesquels $1 \leqslant i \leqslant 6$ et $1 \leqslant s-i \leqslant 6$, ou encore $1 \leqslant i \leqslant 6$ et $s-6 \leqslant i \leqslant s-1$. Ainsi, pour tout $s \in [\![2,7]\!]$: $\displaystyle P(S=s) = \sum_{i=1}^{s-1} \frac{1}{36} = \frac{s-1}{36}$, et pour tout $s \in [\![8,12]\!]$: $\displaystyle P(S=s) = \sum_{i=s-6}^{6} \frac{1}{36} = \frac{13-s}{36}$.



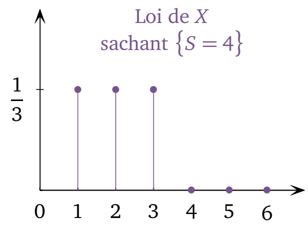


- **Loi conditionnelle de X sachant $\{S=4\}$:** Pour tout $k \in [\![1,6]\!]$:

$$P(X=k,\ S=4) = P(X=k,\ Y=4-k) = \begin{cases} \dfrac{1}{36} & \text{si } k \in [\![1,3]\!] \\ 0 & \text{sinon,} \end{cases}$$

et comme $P(S=4) = \dfrac{4-1}{36} = \dfrac{1}{12}$ d'après le point précédent :

$$P_{\{S=4\}}(X=k) = \frac{P(X=k,\ S=4)}{P(S=4)} = \begin{cases} \dfrac{1}{3} & \text{si } k \in [\![1,3]\!] \\ 0 & \text{sinon.} \end{cases}$$

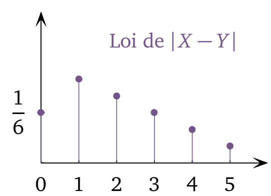


- **Loi de l'écart $|X-Y|$:** La variable aléatoire $|X-Y|$ est à valeurs dans $[\![0,5]\!]$ et pour tout $d \in [\![0,5]\!]$:

$$P\big(|X-Y|=d\big) = P\big(X-Y=d \text{ ou } Y-X=d\big) = \begin{cases} P(X=Y) & \text{si } d=0 \\ P(X-Y=d)+P(Y-X=d) & \text{sinon.} \end{cases}$$

Or $\displaystyle P(X=Y) = \sum_{k=1}^{6} P(X=k,\ Y=k) = \sum_{k=1}^{6} \frac{1}{36} = \frac{1}{6}$ et pour tout $d \in [\![1,5]\!]$:

$$P\big(|X-Y|=d\big) = P(X-Y=d)+P(Y-X=d)$$
$$= \sum_{i=1}^{6-d} P(X=i,\ Y=d+i) + \sum_{j=1}^{6-d} P(X=d+j,\ Y=j)$$
$$= \sum_{i=1}^{6-d} \frac{1}{36} + \sum_{j=1}^{6-d} \frac{1}{36} = 2 \times \frac{6-d}{36} = \frac{6-d}{18}.$$



Exemple 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 ?

Démonstration

- Pour tout $k \in [\![0,n]\!]$, sous l'hypothèse que $N_1=k$, les $n-k$ appels de la deuxième série d'appels satisfont les mêmes hypothèses que ceux de la première série, donc la loi conditionnelle de N_2 sachant $\{N_1=k\}$ est la loi binomiale $\mathscr{B}(n-k,p)$.
- La variable aléatoire N est à valeurs dans $[\![0,n]\!]$ et pour tout $s \in [\![0,n]\!]$:

$$P(N=s) = P(N_1+N_2=s) = \sum_{\substack{0\leqslant i,j\leqslant n\\ i+j=s}} P(N_1=i,\ N_2=j) = \sum_{i=0}^{s} P(N_1=i,\ N_2=s-i)$$
$$= \sum_{i=0}^{s} P(N_1=i)\,P_{\{N_1=i\}}(N_2=s-i) = \sum_{i=0}^{s} \binom{n}{i} p^i(1-p)^{n-i}\binom{n-i}{s-i}p^{s-i}(1-p)^{(n-i)-(s-i)}.$$

Or $\displaystyle\binom{n}{i}\binom{n-i}{s-i} = \frac{n!}{i!(n-i)!}\times\frac{(n-i)!}{(s-i)!(n-s)!} = \frac{n!}{(n-s)!(s-i)!i!} = \frac{n!}{s!(n-s)!}\times\frac{s!}{i!(s-i)!} = \binom{n}{s}\binom{s}{i}$, donc :

$$P(N=s) = \sum_{i=0}^{s}\binom{n}{s}\binom{s}{i}p^s(1-p)^{2n-s-i} = \binom{n}{s}p^s(1-p)^{2n-s}\sum_{i=0}^{s}\binom{s}{i}(1-p)^{-i}$$
$$\overset{\text{Binôme}}{=} \binom{n}{s}p^s(1-p)^{2n-s}\left(1+\frac{1}{1-p}\right)^s = \binom{n}{s}p^s(2-p)^s(1-p)^{2n-2s} = \binom{n}{s}p^s(2-p)^s\Big((1-p)^2\Big)^{n-s}$$
$$= \binom{n}{s}\big(2p-p^2\big)^s\Big(1-\big(2p-p^2\big)\Big)^{n-s}.$$

Finalement $N \sim \mathscr{B}\big(n, 2p-p^2\big)$.

- Le calcul qui précède est classique et il est bon de savoir le refaire, mais on aurait pu atteindre le résultat plus directement. Comment ? En envisageant la situation autrement. Si l'on y réfléchit bien, l'employé réalise n expériences aléatoires indépendantes consistant à joindre ses correspondants en deux temps et la variable aléatoire N compte le nombre de correspondants qu'il a obtenus, donc $N \sim \mathscr{B}(n,q)$ où q est la probabilité d'obtenir un correspondant en un ou deux appels. Or que vaut q ? La probabilité de ne pas obtenir un correspondant après deux appels vaut $(1-p)^2$, donc $q = 1-(1-p)^2 = 2p-p^2$.