Probabilités sur un univers fini

1 Expérience aléatoire et événements

- 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 de ces résultats observables est appelé l'*univers* de l'expérience étudiée et très souvent noté Ω. Plutôt que de résultats, on parle aussi souvent 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 donc 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 donc choisir pour univers Ω l'ensemble $\{\mathrm{N},\mathrm{B}\}^2$ des 2-listes de $\{\mathrm{N},\mathrm{B}\}$, i.e. l'ensemble des mots de 2 lettres sur l'alphabet $\{\mathrm{N},\mathrm{B}\}$.

  > En MPSI, tous nos univers seront des ensembles **finis**. Vous étudierez certains univers infinis en deuxième année.

  Cette restriction du programme aux univers finis est uniquement technique. La théorie générale des probabilités sur un univers quelconque, pour peu qu'on la développe rigoureusement, requiert l'introduction d'objets et de techniques sensiblement plus sophistiqués que ceux que nous aborderons ensemble. 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 MPSI :

  — de jouer à pile ou face **indéfiniment** — nous tricherons un peu parfois...

  — de jouer aux fléchettes contre un disque de rayon $2R$ et de calculer la probabilité d'atterrir dans le disque central de rayon R — cette probabilité vaut intuitivement : $\dfrac{\pi R^2}{\pi(2R)^2}=\dfrac{1}{4}$, mais comment le démontrer ?

- 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 regroupements de résultats définis génériquement par une propriété commune et appelés *événements*.

  — Dans le lancer de dé précédent, la propriété « La face obtenue est paire » est satisfaite par 3 résultats, en l'occurrence 2, 4 et 6, et sera identifiée à l'ensemble $\{2,4,6\}$.

  — Dans le tirage dans l'urne précédent, la propriété « La première boule est blanche » est satisfaite par 2 résultats, en l'occurrence BN et BB, et sera identifiée à l'ensemble $\{\mathrm{BN},\mathrm{BB}\}$.

  Plus généralement, toute partie de Ω est appelée un *événement*. L'ensemble des événements de l'expérience aléatoire étudiée est ainsi l'ensemble total $\mathscr{P}(\Omega)$ des parties de Ω.

Définition **(Vocabulaire usuel sur les événements)** Soit Ω l'univers fini d'une certaine expérience aléatoire.

- 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 : $A\cap B=\varnothing$.
- 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 : $\Omega=\displaystyle\bigsqcup_{i=1}^n A_i$ — 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.

Plus généralement, soient Ω l'univers fini d'une certaine expérience aléatoire et $A, B \in \mathscr{P}(\Omega)$. La paire d'événements $\left\{A, \overline{A}\right\}$ est alors un système complet d'événements car : $A \cup \overline{A} = \Omega$ et $A \cap \overline{A} = \varnothing$, et il n'est pas dur de vérifier que la collection $\left\{A \cap B, A \cap \overline{B}, \overline{A} \cap B, \overline{A} \cap \overline{B}\right\}$ en est un également.

Exemple Pour tout univers fini $\Omega = \left\{\omega_1, \ldots, \omega_n\right\}$, il est clair que : $\Omega = \bigsqcup_{i=1}^{n} \left\{\omega_i\right\}$, donc les événements élémentaires $\left\{\omega_1\right\}, \ldots, \left\{\omega_n\right\}$ forment un système complet d'événements.

2 Probabilités sur un univers fini

Définition (Probabilité sur un univers fini) Soit Ω l'univers fini d'une certaine expérience aléatoire. On appelle *probabilité sur* Ω toute application $P : \mathscr{P}(\Omega) \longrightarrow [0,1]$ telle que :

— $P(\Omega) = 1$,

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

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

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 ensemble fini. L'application $A \stackrel{P}{\longmapsto} \dfrac{|A|}{|\Omega|}$ est une probabilité sur Ω appelé sa *probabilité uniforme*.

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

Exemple On lance un dé à 6 faces une fois. 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 — car à moins que le dé soit pipé, les faces ont toutes autant de chances d'apparaître.
Si on note A l'événement « On obtient un nombre premier » et B l'événement « On obtient un nombre pair », alors :

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

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 $\Omega = \left\{\omega_1, \ldots, \omega_n\right\}$ un univers fini de probabilité uniforme P. Les événements élémentaires $\left\{\omega_1\right\}, \ldots, \left\{\omega_n\right\}$ sont alors 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 :** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. En général : $P\left(\bigcup_{k=1}^{n} A_k\right) \leqslant \sum_{k=1}^{n} P(A_k)$ (*sous-additivité*), mais cette inégalité est une égalité si $A_1, \ldots, A_n$ sont DEUX À DEUX INCOMPATIBLES : $P\left(\bigsqcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$.

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

Démonstration

(i) et (ii) Comme : $A = \big(A \setminus B\big) \sqcup (A \cap B)$, alors : $P(A) = P\big(A \setminus B\big) + P(A \cap B)$.
Ainsi : $P\big(\overline{A}\big) = P\big(\Omega \setminus A\big) = P(\Omega) - P(\Omega \cap A) = 1 - P(A)$, mais aussi : $P(\varnothing) = 1 - P(\Omega) = 1 - 1 = 0$.

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

(iv) Comme : $A = \big(A \setminus B\big) \sqcup (A \cap B)$, alors : $P(A) = P\big(A \setminus B\big) + P(A \cap B)$. Également : $A \cup B = \big(A \setminus B\big) \sqcup B$, donc : $P(A \cup B) = P\big(A \setminus B\big) + P(B)$. Ensuite, il suffit de mélanger.

Pour la sous-additivité, récurrence ! **Initialisation :** Pour tout $A_1 \in \mathscr{P}(\Omega)$: $P(A_1) \leqslant P(A_1)$.

Hérédité : Soit $n \in \mathbb{N}^*$. On suppose que pour tous $A_1, \ldots, A_n \in \mathscr{P}(\Omega)$: $P\left(\bigcup_{k=1}^{n} A_k\right) \leqslant \sum_{k=1}^{n} P(A_k)$.

Alors pour tous $A_1, \ldots, A_{n+1} \in \mathscr{P}(\Omega)$: $P\left(\bigcup_{k=1}^{n+1} A_k\right) = P\left(\left(\bigcup_{k=1}^{n} A_k\right) \cup A_{n+1}\right)$

$$= P\left(\bigcup_{k=1}^{n} A_k\right) + P(A_{n+1}) - P\left(\left(\bigcup_{k=1}^{n} A_k\right) \cap A_{n+1}\right) \leqslant P\left(\bigcup_{k=1}^{n} A_k\right) + P(A_{n+1}) \overset{\text{HDR}}{\leqslant} \sum_{k=1}^{n} P(A_k) + P(A_{n+1}) = \sum_{k=1}^{n+1} P(A_k).$$

La formule du crible sera prouvée au chapitre « Variables aléatoires sur un espace probabilisé fini ». ■

Nous avons rencontré plus haut l'exemple naturel et fondamental des probabilités uniformes. Le théorème qui suit décrit l'ensemble de TOUTES les probabilités qu'on peut installer sur un ensemble fini.

Théorème (Détermination d'une probabilité sur les événements élémentaires) Soient $\Omega = \big\{\omega_1, \ldots, \omega_n\big\}$ un univers fini et $p_1, \ldots, p_n \in [0, 1]$ des réels pour lesquels : $\sum_{i=1}^{n} p_i = 1$. Il existe une et une seule probabilité P sur Ω telle que pour tout $i \in [\![1, n]\!]$: $P\big(\{\omega_i\}\big) = p_i$. Précisément, pour tout $A \in \mathscr{P}(\Omega)$: $P(A) = \sum_{\substack{1 \leqslant i \leqslant n \\ \omega_i \in A}} p_i$.

Démonstration

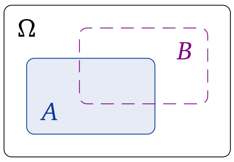
- **Analyse :** Soit P une probabilité sur Ω telle que pour tout $i \in [\![1, n]\!]$: $P\big(\{\omega_i\}\big) = p_i$. Dans ces conditions, pour tout $A \in \mathscr{P}(\Omega)$: $P(A) = P\left(\bigsqcup_{\substack{1 \leqslant i \leqslant n \\ \omega_i \in A}} \{\omega_i\}\right) = \sum_{\substack{1 \leqslant i \leqslant n \\ \omega_i \in A}} P\big(\{\omega_i\}\big) = \sum_{\substack{1 \leqslant i \leqslant n \\ \omega_i \in A}} p_i = \sum_{i=1}^{n} \mathbb{1}_A(\omega_i)\, p_i$.
- **Synthèse :** Pour tout $A \in \mathscr{P}(\Omega)$, posons : $P(A) = \sum_{i=1}^{n} \mathbb{1}_A(\omega_i)\, p_i$.
  - Pour tout $A \in \mathscr{P}(\Omega)$: $P(A) \in [0, 1]$ car : $0 \leqslant P(A) \leqslant \sum_{i=1}^{n} p_i = 1$.
  - En outre : $P(\Omega) = \sum_{i=1}^{n} \mathbb{1}_\Omega(\omega_i)\, p_i = \sum_{i=1}^{n} p_i = 1$.
  - Également, pour tout $j \in [\![1, n]\!]$: $P\big(\{\omega_j\}\big) = \sum_{i=1}^{n} \mathbb{1}_{\{\omega_j\}}(\omega_i)\, p_i = \sum_{i=1}^{n} \delta_{ij}\, p_i = p_j$.
  - 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 :

$$P(A \sqcup B) = \sum_{i=1}^{n} \mathbb{1}_{A \sqcup B}(\omega_i)\, p_i = \sum_{i=1}^{n} \mathbb{1}_A(\omega_i)\, p_i + \sum_{i=1}^{n} \mathbb{1}_B(\omega_i)\, p_i = P(A) + P(B).$$ ■

Exemple 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 probabilité $\frac{3}{20}$. Pour modéliser ce lancer, on peut choisir pour univers Ω l'ensemble $[\![1,6]\!]$ des résultats possibles de l'expérience et pour probabilité P la probabilité définie par : $P\big(\{1\}\big)=\frac{1}{4}$ et pour tout $k\in[\![2,6]\!]$: $P\big(\{k\}\big)=\frac{3}{20}$. Il s'agit bien là d'une probabilité car : $\sum_{k=1}^{6}P\big(\{k\}\big)=\frac{1}{4}+5\times\frac{3}{20}=1$. Attention tout de même — si cette vérification nous garantit que P est une probabilité, elle ne garantit pas l'adéquation de P à la réalité qu'elle est censée décrire.
L'événement A « On obtient une face impaire » a pour probabilité : $P(A)=P\big(\{1\}\big)+P\big(\{3\}\big)+P\big(\{5\}\big)=\frac{1}{4}+2\times\frac{3}{20}=\frac{11}{20}$.

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 $2k-1$. On tire de cette urne une boule au hasard et on regarde son numéro. Pour modéliser ce tirage, on peut choisir pour univers Ω l'ensemble $[\![1,2n]\!]$ de tous les numéros possibles et pour probabilité P la probabilité définie pour tout $k\in[\![1,n]\!]$ par : $P\big(\{2k\}\big)=\frac{2}{3n}$ et $P\big(\{2k-1\}\big)=\frac{1}{3n}$. Il s'agit bien là d'une probabilité car : $\sum_{k=1}^{2n}P\big(\{k\}\big)=\sum_{k=1}^{n}P\big(\{2k\}\big)+\sum_{k=1}^{n}P\big(\{2k-1\}\big)=\sum_{k=1}^{n}\frac{2}{3n}+\sum_{k=1}^{n}\frac{1}{3n}=1$.
L'événement A « On tire une boule de numéro pair » a pour probabilité : $P(A)=\sum_{k=1}^{n}P\big(\{2k\}\big)=\sum_{k=1}^{n}\frac{2}{3n}=\frac{2}{3}$.

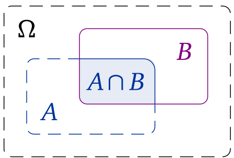
3 Probabilités conditionnelles



Soit Ω un univers fini de probabilité uniforme P. Dans ce contexte, tout événement $A\in\mathscr{P}(\Omega)$ a pour probabilité : $P(A)=\frac{|A|}{|\Omega|}$.

Faisons maintenant l'hypothèse qu'un certain événement $B\in\mathscr{P}(\Omega)$ est réalisé avec : $P(B)>0$. Cette hypothèse modifie a priori la possibilité qu'ont les autres événements de Ω de se réaliser eux-mêmes. Par exemple, sous l'hypothèse que B est réalisé, on a bien envie de dire que B est réalisé « avec probabilité 1 » et que $\overline{B}$ l'est « avec probabilité nulle », mais rien ne garantit pourtant que : $P(B)=1$ et $P\big(\overline{B}\big)=0$. Que faire alors ? De même que nous avons introduit au départ une probabilité P sur Ω pour mesurer la vraisemblance des événements de Ω, nous sommes obligés d'introduire une nouvelle probabilité P_B sur Ω pour mesurer la nouvelle vraisemblance des événements de Ω sous l'hypothèse que B est réalisé. On appelle P_B la *probabilité conditionnelle sur Ω sachant B*. Avec cette probabilité, il devient vrai que : $P_B(B)=1$ et $P_B\big(\overline{B}\big)=0$.

Comment calculerons-nous à présent $P_B(A)$ pour un événement $A\in\mathscr{P}(\Omega)$ quelconque ? Maintenant qu'on sait que B est réalisé, on peut toujours garder Ω comme univers si on veut, mais Ω n'est plus l'ensemble des résultats possibles de l'expérience modélisée, c'est B notre nouvel ensemble de tous les cas possibles. Quant à l'ensemble des cas favorables à la réalisation de A, cet ensemble n'est plus A tout entier, mais sa restriction $A\cap B$. Finalement, sous l'hypothèse que B est réalisé, la probabilité $P_B(A)$ de A vaut : $P_B(A)=\frac{|A\cap B|}{|B|}=\frac{\frac{|A\cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}}=\frac{P(A\cap B)}{P(B)}$.



Par rapport à P, la probabilité conditionnelle $A\overset{P_B}{\longmapsto}\frac{P(A\cap B)}{P(B)}$ se présente comme une probabilité d'oubli, une émanation de P aveugle à tout ce qui dépasse le cadre B. La division par $P(B)$ nous garantit l'égalité : $P_B(B)=1$ selon laquelle « B est tout » — un nouvel univers apparent. En outre : $P\big(\overline{B}\big)=0$. En résumé :

Conditionner, c'est redimensionner l'ensemble des résultats possibles.

Nous venons de travailler avec une probabilité uniforme par simplicité d'exposition, mais les probabilités conditionnelles peuvent être définies dans n'importe quel contexte probabiliste conformément à la définition suivante.

Définition (Probabilité conditionnelle) Soient (Ω, P) un espace probabilisé fini et $B \in \mathscr{P}(\Omega)$ un événement pour lequel : $P(B) > 0$.

- Pour tout $A \in \mathscr{P}(\Omega)$, on appelle *probabilité conditionnelle de A sachant B* le réel : $P(A|B) = P_B(A) = \dfrac{P(A \cap B)}{P(B)}$.
- L'application $P_B : \mathscr{P}(\Omega) \longrightarrow [0,1]$ est une probabilité sur Ω appelée sa *probabilité conditionnelle sachant B* (*issue de P*).

Démonstration Il s'agit de montrer que P_B est une probabilité sur Ω.

- Pour tout $A \in \mathscr{P}(\Omega)$: $P_B(A) \in [0,1]$ car par croissance de P : $0 \leqslant P(A \cap B) \leqslant P(B)$.
- 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 :

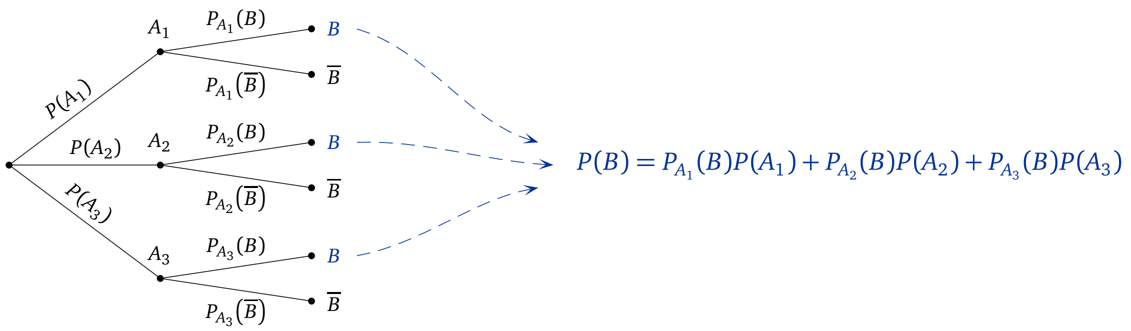
$$P_B(A \sqcup A') = \frac{P\big((A \sqcup A') \cap B\big)}{P(B)} = \frac{P\big((A \cap B) \sqcup (A' \cap B)\big)}{P(B)} = \frac{P(A \cap B) + P(A' \cap B)}{P(B)} = P_B(A) + P_B(A'). \qquad \blacksquare$$

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

Démonstration Comme : $B = B \cap \Omega = B \cap \displaystyle\bigsqcup_{i=1}^{n} A_i = \bigsqcup_{i=1}^{n} (B \cap A_i)$ alors :

$$P(B) = P\left(\bigsqcup_{i=1}^{n} (B \cap A_i)\right) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P_{A_i}(B)\, P(A_i). \qquad \blacksquare$$

✌ **Explication** ✌ Vous avez déjà rencontré la formule des probabilités totales auparavant, mais sans doute à travers des *arbres de probabilité*. Voilà comment vous vous l'êtes sans doute représentée 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 PLUS CONSIDÉRÉ COMME UNE PREUVE BIEN FORMALISÉE. Pourquoi ? Parce qu'il faut un jour apprendre à maîtriser le formalisme de la formule des probabilités totales si on veut pouvoir résoudre des problèmes un peu plus compliqués que ceux auxquels vous avez jusqu'ici été confrontés.

Exemple Dans une classe de 40 étudiants — 25 filles et 15 garçons — le professeur principal se propose de désigner brutalement deux délégués provisoires. Il prend une liste de la classe, ferme les yeux et pointe au hasard un premier nom avec la pointe de son stylo, puis de même un deuxième. Avec quelle probabilité le deuxième nom tiré est-il celui d'un garçon ?

Démonstration Notons G_1 (resp. G_2) l'événement « Le premier (resp. deuxième) nom tiré est celui d'un garçon ». Nous cherchons $P(G_2)$. De quelles informations disposons-nous ?

$$P(G_1) = \frac{15}{40} = \frac{3}{8}, \qquad P\big(\overline{G_1}\big) = \frac{25}{40} = \frac{5}{8}, \qquad P_{G_1}(G_2) = \frac{14}{39} \qquad \text{et} \qquad P_{\overline{G_1}}(G_2) = \frac{15}{39}.$$

Or $\{G_1, \overline{G_1}\}$ est un système complet d'événements de probabilités strictement positives, donc d'après la formule des probabilités totales : $P(G_2) = P_{G_1}(G_2)\, P(G_1) + P_{\overline{G_1}}(G_2)\, P\big(\overline{G_1}\big) = \dfrac{14}{39} \times \dfrac{3}{8} + \dfrac{15}{39} \times \dfrac{5}{8} = \dfrac{3}{8}$.

✌ **Explication** ✌ Cet exemple diffère sensiblement de ceux qui l'ont précédé dans ce cours. Dans nos premiers exemples, la résolution du problème posé passait par la définition propre d'un univers Ω et d'une probabilité P sur Ω. Ici au contraire, point d'univers — seulement des événements et certaines probabilités les concernant. Il y a bien un espace probabilisé (Ω, P) quelque part, mais il reste dans l'ombre. Les événements et les probabilités qu'on se donne au départ nous épargnent de l'expliciter. On peut même en faire une règle.

Un exercice de modélisation probabiliste se résout :

— soit en explicitant l'espace probabilisé (Ω, P) de l'expérience aléatoire étudiée et en faisant du dénombrement,

— soit en acceptant comme point de départ un certain jeu d'événements et de probabilités, conditionnelles ou non, qu'utiliseront tous les calculs de probabilités ultérieurs.

La deuxième démarche pose un vrai problème dont il faut avoir conscience, mais sur lequel nous ne nous étendrons pas. Qu'est-ce qui nous garantit, dans l'exemple précédent où nous n'avons pas explicité (Ω, P), que la définition naïve des événements G_1 et G_2 et des probabilités $P(G_1)$, $P\left(\overline{G_1}\right)$, $P_{G_1}(G_2)$ et $P_{\overline{G_1}}(G_2)$ ne sont pas contradictoires ? En toute rigueur, de tels objets ne peuvent être introduits que sur un espace probabilisé **PRÉALABLEMENT** défini — or là, nous les introduisons de force sans la moindre allusion à un (Ω, P). Nous **ADMETTRONS** sans le dire, toutes les fois où nous ferons un tel raisonnement, l'**EXISTENCE** d'un espace probabilisé sous-jacent.

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

(i) Pour tout $A \in \mathscr{P}(\Omega)$, si : $P(A) > 0$, alors : $P_B(A) = \dfrac{P_A(B)\,P(A)}{P(B)}$.

(ii) Soit $\{A_1, \ldots, A_n\}$ un système complet d'événements de Ω de probabilités strictement positives.

Pour tout $j \in [\![1, n]\!]$: $P_B(A_j) = \dfrac{P_{A_j}(B)\,P(A_j)}{\displaystyle\sum_{i=1}^{n} P_{A_i}(B)\,P(A_i)}$.

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

✌ **Explication** ✌ 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 $P(A_1), \ldots, P(A_n)$ 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. En résumé, 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 ?

$$P\left(\overline{B}\right) = \frac{90}{100} = \frac{9}{10} \quad \text{et} \quad P(B) = \frac{1}{10} \quad \text{(probabilités a priori)} \quad \text{et} \quad P_B(T_B) = \frac{90}{100} = \frac{9}{10} \quad \text{et} \quad P_{\overline{B}}\left(\overline{T_B}\right) = \frac{80}{100} = \frac{8}{10}.$$

L'égalité : $P_{\overline{B}}\big(\overline{T_B}\big) = \dfrac{80}{100}$ 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 : $P_{\overline{B}}(T_B) = 1 - P_{\overline{B}}\big(\overline{T_B}\big) = \dfrac{2}{10}$.
Finalement, d'après la formule de Bayes utilisée avec le système complet d'événements $\big\{B, \overline{B}\big\}$:

$$P_{T_B}(B) = \frac{P_B(T_B)\,P(B)}{P_B(T_B)\,P(B) + P_{\overline{B}}(T_B)\,P\big(\overline{B}\big)} = \frac{\dfrac{9}{10} \times \dfrac{1}{10}}{\dfrac{9}{10} \times \dfrac{1}{10} + \dfrac{2}{10} \times \dfrac{9}{10}} = \frac{1}{3}.$$

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 !

Exemple **(Paradoxe des prisonniers version Schtroumpfs)** Gargamel a réussi à emprisonner la Schtroumpfette, le Grand Schtroumpf et le Schtroumpf à lunettes. Magnanime, il a décidé d'en libérer un et de faire bouillir les deux autres, mais son chat-geôlier Azraël n'est pas autorisé à annoncer le sort de chacun avant l'heure du bouillon.
Nous noterons S l'événement « La Schtroumpfette est épargnée », G l'événement « Le Grand Schtroumpf est épargné » et L l'événement « Le Schtroumpf à lunettes est épargné ».

- Avec quelle probabilité la Schtroumpfette sera-t-elle épargnée a priori ? Avec probabilité : $P(S) = \dfrac{1}{3}$, bien sûr !
- Seulement voilà, la Schtroumpfette est curieuse : « Dis-moi, Azraël, je sais bien que tu n'as pas le droit de me dire si je serai cuite ou non, mais comme nous serons deux à l'être, tu peux au moins me dire qui le sera du Grand Schtroumpf ou du Schtroumpf à lunettes. Je ne te demande qu'un nom, un seul ! » Azraël réfléchit un instant et se dit : « Après tout… », puis à la Schtroumpfette : « Nous ferons bouillir le Schtroumpf à lunettes. »
A posteriori, avec quelle probabilité la Schtroumpfette sera-t-elle donc épargnée ? Elle se dit que puisqu'il ne reste qu'elle et le Grand Schtroumpf en lice pour la marmite finale, elle a maintenant une probabilité : $P_{\overline{L}}(S) = \dfrac{1}{2}$ d'être épargnée — et c'est mieux, se réjouit-elle ! C'est là qu'elle se trompe.
- La Schtroumpfette a l'impression qu'Azraël lui a donné l'information $\overline{L}$, mais ce n'est là qu'une apparence car en réalité Azraël s'interdit de la désigner. L'information qu'Azraël fournit dépend d'un choix qu'il a dû faire ou non — si la Schtroumpfette est cuite, Azraël a donné le seul nom qu'il pouvait donner, mais sinon il a dû faire un choix entre le Grand Schtroumpf et le Schtroumpf à lunettes. Le problème, c'est que la définition initiale des événements S, G et L ne tient pas compte de cette cuisine psychologique d'Azraël. En résumé, la Schtroumpfette a bien raison d'affirmer que : $P_{\overline{L}}(S) = \dfrac{1}{2}$, mais elle a tort de prendre l'événement « AZRAËL DIT QUE le Schtroumpf à lunettes sera bouilli » pour l'événement $\overline{L}$.
La réponse d'Azraël requiert ainsi une modélisation plus fine et nous noterons A_L l'événement « Azraël a désigné le Schtroumpf à lunettes ». D'après la formule de Bayes : $P_{A_L}(S) \overset{\spadesuit}{=} \dfrac{P_S(A_L)\,P(S)}{P_S(A_L)\,P(S) + P_{\overline{S}}(A_L)\,P\big(\overline{S}\big)}$, et d'après la formule des probabilités totales : $P_{\overline{S}}(A_L) \overset{\clubsuit}{=} P_{\overline{S}\cap\overline{G}}(A_L)\,P_{\overline{S}}\big(\overline{G}\big) + P_{\overline{S}\cap\overline{L}}(A_L)\,P_{\overline{S}}\big(\overline{L}\big)$. Quelles probabilités connaissons-nous indépendamment de la cuisine psychologique d'Azraël ?

$$P(S) = \frac{1}{3}, \quad P\big(\overline{S}\big) = \frac{2}{3}, \quad P_S(A_L) = \frac{1}{2}, \quad P_{\overline{S}\cap\overline{G}}(A_L) = 0, \quad P_{\overline{S}}\big(\overline{G}\big) = \frac{1}{2} \quad \text{et} \quad P_{\overline{S}}\big(\overline{L}\big) = \frac{1}{2}.$$

On compare à présent ci-dessous deux modélisations possibles du problème posé.

Azraël NE s'autorise PAS à désigner la Schtroumpfette	Azraël s'autorise à désigner la Schtroumpfette
Dans ce cas : $P_{\overline{S}\cap\overline{L}}(A_L) = 1$, donc : $P_{\overline{S}}(A_L) \overset{\clubsuit}{=} 0 \times \dfrac{1}{2} + 1 \times \dfrac{1}{2} = \dfrac{1}{2}$, donc : $P_{A_L}(S) \overset{\spadesuit}{=} \dfrac{\dfrac{1}{2} \times \dfrac{1}{3}}{\dfrac{1}{2} \times \dfrac{1}{3} + \dfrac{1}{2} \times \dfrac{2}{3}} = \dfrac{1}{3} = P(S)$.	Dans ce cas : $P_{\overline{S}\cap\overline{L}}(A_L) = \dfrac{1}{2}$, donc : $P_{\overline{S}}(A_L) \overset{\clubsuit}{=} 0 \times \dfrac{1}{2} + \dfrac{1}{2} \times \dfrac{1}{2} = \dfrac{1}{4}$, donc : $P_{A_L}(S) \overset{\spadesuit}{=} \dfrac{\dfrac{1}{2} \times \dfrac{1}{3}}{\dfrac{1}{2} \times \dfrac{1}{3} + \dfrac{1}{4} \times \dfrac{2}{3}} = \dfrac{1}{2} > \dfrac{1}{3} = P(S)$.

Dans la première colonne, l'égalité : $P_{A_L}(S) = P(S)$ montre que la confession d'Azraël n'améliore pas les perspectives de survie de la Schtroumpfette. La deuxième colonne met quant à elle en évidence son erreur de raisonnement.

- Faut-il conclure de tout cela que l'information d'Azraël est sans contenu véritable ? Oh que non ! C'est en fait au Grand Schtroumpf qu'elle profite. Le fait qu'Azraël s'interdise de désigner la Schtroumpfette se traduit ainsi : $P_G(A_L)=1$.

En retour : $P_{A_L}(G)=\dfrac{P_G(A_L)\,P(G)}{P_S(A_L)\,P(S)+P_{\overline{S}}(A_L)\,P\big(\overline{S}\big)}=\dfrac{1\times\frac{1}{3}}{\frac{1}{2}\times\frac{1}{3}+\frac{1}{2}\times\frac{2}{3}}=\dfrac{2}{3}>\dfrac{1}{3}=P(G)$. Sacré Grand Schtroumpf.

Théorème (Formule des probabilités composées) Soient (Ω,P) un espace probabilisé fini et $A_1,\dots,A_n\in\mathscr{P}(\Omega)$. On suppose : $P\big(A_1\cap\ldots\cap A_{n-1}\big)>0$. 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 Comme : $\big(A_1\cap\ldots\cap A_{n-1}\big)\subset\big(A_1\cap\ldots\cap A_i\big)$ pour tout $i\in[\![1,n-1]\!]$ et : $P\big(A_1\cap\ldots\cap A_{n-1}\big)>0$, alors par croissance de P : $P\big(A_1\cap\ldots\cap A_i\big)>0$. Finalement, grâce à une simplification télescopique :

$$P(A_1)\,P_{A_1}(A_2)\,\ldots\,P_{A_1\cap\ldots\cap A_{n-1}}(A_n)=P(A_1)\prod_{k=2}^{n}P_{A_1\cap\ldots\cap A_{k-1}}(A_k)=P(A_1)\prod_{k=2}^{n}\frac{P(A_1\cap\ldots\cap A_k)}{P(A_1\cap\ldots\cap A_{k-1})}=P(A_1\cap\ldots\cap A_n).\qquad\blacksquare$$

Exemple J'ai 4 pièces de 1€ et 3 de 2€ en poche et je veux me faire un petit plaisir aujourd'hui. Pour ne pas tout dépenser, je tire 3 pièces au hasard de ce pactole, successivement. Avec quelle probabilité sortent-elles dans l'ordre : 2€ / 1€ / 2€ ?

Démonstration Pour tout $i\in[\![1,3]\!]$, notons D_i l'événement « La $i^{\text{ème}}$ pièce tirée est une pièce de 2€ ». Nous cherchons $P\big(D_1\cap\overline{D_2}\cap D_3\big)$. D'après la formule des probabilités composées :

$$P\big(D_1\cap\overline{D_2}\cap D_3\big)=P(D_1)\,P_{D_1}\big(\overline{D_2}\big)\,P_{D_1\cap\overline{D_2}}(D_3)=\frac{3}{7}\times\frac{4}{6}\times\frac{2}{5}=\frac{4}{35}\approx\frac{5}{35}=\frac{1}{7}.$$

Exemple Un commerçant met en vente 50 tickets d'un certain 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 $P\big(\overline{G}\big)$.

- **Première preuve :** À base de dénombrement après définition de l'espace probabilisé (Ω,P). On numérote artificiellement les tickets gagnants 1, 2 et 3 et les tickets perdants de 4 à 50. On peut alors choisir pour univers Ω l'ensemble $\mathscr{P}_6\big([\![1,50]\!]\big)$ des 6-combinaisons de $[\![1,50]\!]$ — achat simultané des 6 tickets — et pour probabilité P sur Ω la probabilité uniforme. Dans ce modèle : $\overline{G}=\mathscr{P}_6\big([\![4,50]\!]\big)$, donc :

$$P\big(\overline{G}\big)=\frac{\big|\overline{G}\big|}{|\Omega|}=\frac{\binom{47}{6}}{\binom{50}{6}}=\frac{47\times46\times45\times44\times43\times42}{50\times49\times48\times47\times46\times45}=\frac{44\times43\times42}{50\times49\times48}=\frac{473}{700},\qquad\text{et enfin : }\quad P(G)=\frac{227}{700}.$$

- **Deuxième preuve :** À base de probabilités conditionnelles sans aucune définition de l'espace (Ω,P). On peut considérer 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 : $P\big(\overline{G}\big)=P(T_1\cap\ldots\cap T_n)=P(T_1)\,P_{T_1}(T_2)\,\ldots\,P_{T_1\cap\ldots\cap T_{n-1}}(T_n)=\displaystyle\prod_{i=1}^{6}\frac{48-i}{51-i}$

$$=\frac{47\times46\times45\times44\times43\times42}{50\times49\times48\times47\times46\times45}=\frac{473}{700},\qquad\text{donc : }\quad P(G)=\frac{227}{700}.$$

4 Événements indépendants

Définition (Paire d'événements indépendants) Soient (Ω,P) un espace probabilisé fini et $A,B\in\mathscr{P}(\Omega)$. On dit que les événements A et B sont *indépendants* si : $P(A\cap B)=P(A)\,P(B)$.

Explication Intuitivement — sous l'hypothèse simplificatrice que : $P(B) > 0$ — les événements A et B sont indépendants si $P(A)$ ne dépend pas de la réalisation de B, i.e. si : $P_B(A) = P(A)$, ou encore : $P(A \cap B) = P(A)\, P(B)$.

Définition (Ensemble fini d'événements mutuellement indépendants) Soient (Ω, P) un espace probabilisé fini et $A_1, \ldots, A_n \in \mathscr{P}(\Omega)$. On dit que les événements $A_1, \ldots, A_n$ sont (*mutuellement*) *indépendants* si pour toute partie I de $[\![1, n]\!]$:

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

ATTENTION ! $\boxed{\text{(Mutuellement) indépendants} \implies \text{DEUX À DEUX indépendants}}$ 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 $A_1, \ldots, A_n \in \mathscr{P}(\Omega)$. Si $A_1, \ldots, A_n$ sont indépendants, les événements $A_1^\circ, \ldots, A_n^\circ$ le sont aussi pour tous $A_1^\circ \in \{A_1, \overline{A_1}\}, \ldots, A_n^\circ \in \{A_n, \overline{A_n}\}$.

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. ■

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 alors 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

- Commençons par calculer la probabilité de déceler la supercherie quand on tire une seule fois 4 cartes simultanément. À cette fin, choisissons pour univers Ω l'ensemble des 4-combinaisons des 32 cartes du jeu et pour probabilité P sur Ω la probabilité uniforme. L'événement A « La supercherie est décelée » est l'ensemble des tirages qui contiennent les deux as de pique. Quand on veut constuire un tel tirage, seules les 2 autres cartes restent à choisir sous la forme d'une 2-combinaison de l'ensemble des 30 cartes qui ne sont pas les as de pique ($\binom{30}{2}$ possibilités) — d'où finalement : $P(A) = \dfrac{\binom{30}{2}}{\binom{32}{4}} = \dfrac{3}{248}$.

- À présent, pour tout $k \in [\![1, n]\!]$, notons A_k l'événement « La supercherie est décelée pendant le $k^{\text{ème}}$ tirage » et B l'événement « La supercherie finit par être décelée » d'événement contraire « La supercherie n'est jamais décelée » : $\overline{B} = \overline{A_1} \cap \ldots \cap \overline{A_n}$. Le fait que les n expériences réalisées le soient avec remise signifie que les événements $A_1, \ldots, A_n$ sont indépendants, 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 : $P(B) \geqslant 0,9 \iff \left(\dfrac{245}{248}\right)^n \leqslant \dfrac{1}{10} \iff n \geqslant \dfrac{\ln 10}{\ln \dfrac{248}{245}} \iff n \geqslant 190.$