Probabilité (discrète)

Florent Capelli

09/11/2026

Beaucoup d'exemples de ces notes sont tirés du livre *Introduction aux probabilités, Modèles et applications*, par Quentin Berger, Francesco Caravenna et Paolo Dai Pra (Dunod).

1 Notion d'espaces probabilistes

On veut décrire un processus aléatoire, c'est-à-dire, un processus qui peut avoir plusieurs issues que nous ne sommes pas en mesure de prévoir exactement et on souhaiterait quantifier le fait que certains événements ont une plus grande chance d'arriver que d'autres. Pour cela, nous avons besoin de **formaliser** ces quantités et les espaces dans lesquels elles ont du sens. On utilise pour cela un espace probabiliste, et, pour aujourd'hui, un espace discret.

Pour expliquer les définitions, on va se placer dans les deux cas suivants :

1. On veut modéliser la valeur de deux dés après deux lancers indépendants (l'un après l'autre).
2. On veut modéliser le résultat d'une élection avec trois candidats.

La définition d'espace probabiliste repose sur les trois concepts suivants :

- Un *univers* ou *espace d'état* Ω. Intuitivement, c'est l'espace des "résultats" possibles de notre processus.
    1. Dans le cas du lancer de deux dés, on pourra choisir $\Omega = [6] \times [6]$ où $[6] = \{1, \dots, 6\}$, c'est-à-dire qu'on ne regardera que la valeur de chaque dé lancé.
    2. Dans le cas du sable, on pourra prendre $\Omega = [M]^3$ où $[M]$ est le nombre d'inscrits sur la liste électorale. Cela dit, on peut aussi prendre $\Omega = \mathbb{N}^3$ sans faire d'hypothèses a priori sur ce nombre.
- La notion d'*événements*. Un événement A est un sous-ensemble de Ω.
    1. Dans le cas des dés, on peut considérer l'évènement : "la somme des dés est paire", c'est-à-dire qu'on définit $A = \{(i, j) \in [6] \times [6] \mid i + j = 0 \mod 2\} \subseteq \Omega$.
    2. Dans le cas des élections, on peut considérer l'événement "le candidat 1 remporte l'élection avec plus de la moitié des voix", qui correspond

à $A = \{(m_1, m_2, m_3) \in \mathbb{N}^3 \mid m_1 > \frac{m_1+m_2+m_3}{2}\}$. On peut aussi considérer l'évènement $B = \{(m_1, m_2, m_3) \in \mathbb{N}^3 \mid m_2 > m_3\}$ qui correspond aux résultats où le candidat 2 remporte plus de voix que le candidat 3. Remarquons que $A \cap B$ est l'évènement où A et B sont tous les deux simultanéments vrais, c'est-à-dire qu'ici, l'intersection des évènements corresponds à la conjunction. De même, l'union correspond à la disjonction et le complément $(\Omega \setminus A)$ à la négation.

- Enfin, on a besoin d'une **mesure de probabilité** $\Pr\colon \mathcal{P}(\Omega) \to [0,1]$ qui mesure la probabilité qu'un événement $\mathcal{A}$ se produise (point de vue fréquentiste).
  1. Dans le cas du lancer de dés, si les dés ne sont pas pipés et que les deux lancers sont indépendants, on peut choisir $\Pr(\{(i,j)\}) = \frac{1}{36}$ pour tout $(i,j) \in [6] \times [6]$ et $\Pr(A) = \sum_{(i,j)\in A} \Pr(\{(i,j)\}) = \frac{|A|}{36}$. Si les dés sont pipés cependant, on choisira une mesure plus adaptées.
  2. Dans le cas des élections, la mesure de probabilité sous-jacente est déjà bien moins clairement définie mais on peut supposer qu'elle existe implicitement, par le comportement individuel de chaque électeur par exemple. L'enjeu ici n'est alors plus de formaliser cette mesure de probabilité, mais de l'estimer à partir d'observations. C'est l'objet d'étude des **statistiques**.

Intuitivement, du point de vue fréquentiste, la mesure de probabilité mesure la quantité suivante : si on répète l'expérience un grand nombre N de fois et qu'on regarde le nombre de fois $S_A(N)$ où A se produit (c'est-à-dire que le résultat R de l'expérience est dans A), alors $\frac{S_A(N)}{N}$ converge vers $\Pr(A)$. Cette interprétation est purement théorique : rien ne garantit que la répétition de cette expérience soit possible ni que $\frac{S_A(N)}{N}$ ne convergera mais elle nous permettra de comprendre de façon intuitive la plupart des résultats présentés ici.

D'autres interprétations des probabilités ont été proposées. Du point de vue **bayésien**, $\Pr(A)$ représente le degré de croyance en le fait que l'événement A soit vrai. Si la théorie sous-jacente est similaire, l'interprétation des quantité est philosophiquement différente et mène à d'autres considérations statistiques (on reviendra dessus lors du cours sur les statistiques).

Un modèle probabiliste $(\Omega, \Pr)$ est donc simplement un **modèle** d'un phénomène aléatoire à travers la mesure de différents événements. On voit cependant, que sous cette interprétation du modèle, les choses suivantes doivent être vérifier par la mesure de probabilité Pr choisie :

- On a $\Pr(\Omega) = 1$.
- On a $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ pour tout $A, B \in \mathcal{A}$ tels que $A \cap B = \emptyset$. Plus généralement, on veut $\Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$ pour tout ensemble d'événement deux à deux disjoints.

Techniquement, pour que tout fonctionne, on a besoin d'un peu plus dans les espaces non discrets mais nous parlerons de tout cela plus en détail à la prochaine séance.

Exercice 1. Montrez que :

- $\Pr(\Omega \setminus A) = 1 - \Pr(A)$.
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
- $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$.
- Si $B \subseteq A$, $\Pr(A \setminus B) = \Pr(A) - \Pr(B)$.
- $\Pr(A) = \sum_{a \in A} \Pr(\{a\})$.
-

En fait, dans le cas discret, pour définir un espace probabilisé, on peut montrer qu'il est suffisant de définir $\Pr(\{\omega\}) \geq 0$ pour tout $\omega \in \Omega$ avec $\sum_{\omega \in \Omega} \Pr(\{\omega\}) = 1$.

Exemple 1.1.

- Probabilité uniforme sur un ensemble fini Ω : $\Pr(A) = \frac{|A|}{|\Omega|}$. Alternativement, $\Pr(\{\omega\}) = \frac{1}{|\Omega|}$ pour tout $\omega \in \Omega$.
- Probabilité pondérée sur un ensemble fini Ω. Si on a une fonction $w \colon \Omega \to \mathbb{R}_+$, on définit $\Pr(\{A\}) = \frac{\sum_{\omega \in A} w(A)}{\sum_{\omega \in \Omega} w(\omega)}$.
- Probabilité de Poisson pour $\Omega = \mathbb{N}$: $\Pr(\{i\}) = e^{-\lambda} \frac{\lambda^i}{i!}$.

2 Quelques rappels combinatoires

Lorsqu'on fait des probabilités discrètes, on a souvent besoin de raisonner sur le nombre possibles de sous-ensembles produits par certains processus aléatoires. On fait ici quelques rappels qui seront utiles par la suite.

Tirage avec remise. On commence par s'intéresser au cas où on réitère k fois le processus suivant : on tire un objet dans un ensemble O de taille n et on le remet dans l'ensemble. On construit ainsi une suite $o_1, \ldots, o_k \in O^k$ d'objets. Comme chaque objet peut être pioché chaque fois, le nombre de suites $o_1, \ldots, o_k$ que l'on peut former ainsi est n^k.

Tirage sans remise. On s'intéresse désormais au cas où l'objet n'est pas remis dans l'ensemble après avoir été pioché. Dans ce cas, on a n choix au premier tirage, puis $n-1$ au second, puis $n-2$ au troisième etc. On a donc $n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$ tirages possibles.

Tirage de k éléments. On s'intéresse enfin au cas où l'on souhaite tirer k objets simultanément dans O. C'est-à-dire que, contrairement au cas précédent, l'ordre du tirage n'importe plus. On veut juste prendre un ensemble $O' \subseteq O$ tel que $|O'| = k$. Le nombre de sous-ensemble de taille k d'un ensemble de taille n se note $\binom{n}{k}$ et se dit "k parmi n". Pour calculer sa valeur, on part de la formule précédente pour le tirage sans remise. Chaque suite $s = (o_1, \ldots, o_k)$ du tirage sans remise induit un unique ensemble $O_s = \{o_1, \ldots, o_k\}$ de taille k. Maintenant, si on fixe O' de taille k, on peut se demander combien de suite s vérifie $O_s = O'$. Pour cela, il suffit de remarquer que s doit être une permutation de O'. On a donc $k!$ suites qui induisent O'. Donc $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Exemple 2.1. Le paradoxe des anniversaires est un faux paradoxe qui dit essentiellement la chose suivante : dans une classe (disons de 30 étudiants), il y a très souvent deux personnes qui ont leur anniversaire en même temps. Cela semble contre-intuitif car il y a 365 dates d'anniversaire possible (on simplifie légèrement en ignorant le cas des années bissextiles). On se dit alors qu'il y a peu de chances que deux personnes aient leur anniversaire en même temps. Cela vient du fait qu'on a tendance à confondre cette probabilité avec la probabilité que deux personnes aient leur anniversaire à une date fixée (par exemple, le 1er janvier), qui, elle, serait petite.

On veut calculer p_n la probabilité que deux personnes aient la même date d'anniversaire dans un groupe de n personnes. On suppose que les dates d'anniversaires uniformément distribuées dans la population, c'est-à-dire qu'une personne est né le $i$ème jour de l'année avec probabilité $\frac{1}{365}$. On se place donc dans l'espace $\Omega = [365]^n$ avec $\Pr((d_1, \dots, d_n)) = \frac{1}{365^n}$ et p_n est la probabilité de l'événement : $B_n = \{(d_1, \dots, d_n) \mid \exists i < j . d_i = d_j\}$. Comme on est dans un espace fini avec une distribution uniforme sur chaque objet, on a $p_n = \frac{|B_n|}{|\Omega|} = \frac{|B_n|}{365^n}$. Notre but est donc d'évaluer $|B_n|$. Il est plus simple d'évaluer $|\Omega \setminus B_n|$. C'est le nombre de suite $(d_1, \dots, d_n)$ sans répétition. C'est-à-dire $365 \times \cdots \times (365 - n + 1)$. On a donc $|B_n| = |\Omega| - \frac{365!}{(365-n)!} = 365^n - \prod_{i=0}^{n-1}(365 - i)$. En d'autre mot, $p_n = \frac{|B_n|}{365^n} = 1 - (\prod_{i=0}^{n-1} 1 - \frac{i}{365})$.

Regardons quelques valeurs de p_n :

```
p = [1-np.prod([1-i/365 for i in range(j)]) for j in range(50)]
```

On voit par exemple que $p_{23} \simeq 0.5$, $p_{30} \simeq 0.7$ et $p_{50} \simeq 0.96$, $p_{57} \simeq 0.99$!

2.1 Probabilités conditionnelles et indépendance

Étant donné deux évènements A, B, on note $P(A|B)$ (la probabilité de A sachant B) la valeur $P(A|B) = \frac{P(A \cap B)}{P(B)}$. Intuitivement, on essaie d'évaluer la probabilité de A parmi l'ensemble des états où B est vrai. Si on regarde dans le cas fini par exemple, cette valeur est le nombre d'ensemble où A et B sont vrais ($|A \cap B|$) divisé par le nombre d'ensembles où B est vrai. Or $\frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{P(A \cap B)}{P(B)}$ ce qui justifie la définition.

Exemple 2.2. Dans le cas du lancé de deux dés. On note A l'événement "Il y a au moins un six" et B l'événement "La somme des dés vaut 9". On a $P(A \cap B) = \frac{2}{36} = \frac{1}{18}$ puisque seul $(3, 6)$ et $(6, 3)$ vérifie les deux événements. Par contre, comme $P(B) = \frac{1}{9}$ (si on fixe la valeur du premier dé à $i \geq 3$, il n'y a qu'une seule valeur $9 - i$ possible pour le second). Donc $P(A|B) = \frac{9}{18} = \frac{1}{2}$. Autrement dit, si on nous donne l'information que la somme des dés fait 9, on sait désormais que la probabilité que l'un des deux dés soit 6 est $\frac{1}{2}$. On peut le vérifier. Les cas où B est vrai sont $(3, 6), (4, 5), (5, 4), (6; 3)$. On voit bien que dans la moitié des cas, un des dés vaut 6.

Dans l'exemple précédent, on voit apparaître une autre interprétation des probabilités conditionnelles. On cherche parfois à établir la probabilité d'un fait en sachant qu'un autre s'est produit. Dans une vision bayésienne, on veut réviser notre degré de croyance en A après avoir observé B. On peut formaliser cette intuition ainsi :

Théorème 2.1. *Pour un évènement B, $(\Omega, \Pr(\cdot|B))$ est un espace probabilisé (où $\Pr(\cdot|B)$ est la fonction qui associe A à $\Pr(A|B)$).*

La formule de Bayes relie la probabilité de A sachant B à celle de B sachant A comme suit :

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

C'est une évidence à prouver mais elle permet facilement de prouver des résultats qui semblent parfois contre-intuitifs.

Exemple 2.3. On considère deux urnes U_1 et U_2. La première contient 3 boules rouges et 1 boules vertes. La seconde contient 1 boule rouge et une boule verte. On choisit une urne avec probabilité 1/2 et on pioche dedans. Il est facile de voir que la probabilité de piocher une boule rouge est $\frac{1}{2} \times (\frac{3}{4} + \frac{1}{2}) = \frac{5}{8}$. La probabilité de piocher une boule rouge sachant qu'on a choisi la première urne est $\frac{3}{4}$. Maintenant, quelle est la probabilité que la première urne ait été choisie sachant qu'on a tiré une boule rouge ? On voit ici que le problème est plus complexe à résoudre car il ne suit pas la temporalité des événements. On choisit d'abord l'urne avant de tirer la boule. On a donc l'impression qu'il est difficile d'évaluer la probabilité d'avoir choisi dans la première urne sachant qu'on a pioché une boule rouge.

Cependant, on peut facilement utiliser la formule de Bayes pour cela. Si A est l'événement "l'urne 1 a été choisie" et B l'événement "une boule rouge a été choisie", on voit facilement que :

- $P(A) = \frac{1}{2}$
- $P(B) = \frac{5}{8}$
- $P(B|A) = \frac{3}{4}$

Donc $P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{1}{2}\frac{3}{4}\frac{8}{5} = \frac{3}{5}$.

Événements indépendants. On dit que deux événements A, B sont **indépendants** si et seulement si $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$. Le nom est justifié parce que, dans le cas d'événements indépendants, $P(A|B) = P(A)$ et $P(B|A) = P(B)$. En d'autre terme, la connaissance de A n'influe pas la probabilité sur B (et vice-versa). Attention, cette notion est purement calculatoire et ne veut pas dire qu'il n'y a aucune relation entre A et B. En effet, si on tire une carte dans un jeu de 52 cartes, la probabilité de "la carte est un as" (événement A) est $\frac{4}{52} = \frac{1}{13}$. La probabilité de "la carte est du cœur" (événement B) est $\frac{1}{4}$. La probabilité de $A \cap B$ est $\Pr(A \cap B) = \frac{1}{52} = \frac{1}{13} \times \frac{1}{4} = \frac{1}{52}$.

Cela dit, lorsqu'on décrit un processus aléatoire sans forcément décrire l'ensemble des probabilités, il est courant que l'on considère deux événements comme étant indépendants car ne partageant aucune information. Cela est une *hypothèse* sur la façon dont le processus s'est déroulé permettant de calculer des probabilités.

Une famille $A_1, \ldots, A_k$ d'événements est indépendante si pour tout $J \subseteq [k]$, on a $\Pr(\bigcap_{j \in J} A_j) = \prod_{j \in J} \Pr(A_j)$. C'est-à-dire qu'on a une indépendance quelque soit le sous-ensemble d'événements qu'on considère.