Probabilités finies discrètes

« On est toujours forcé de donner quelque chose au hasard. »
Bonaparte.

Cette année, les probabilités se restreignent au cas des univers finis. Il ne sera donc pas question de probabilités continues, de loi normale, exponentielle...
Ce chapitre est l'occasion de mettre à profit les compétences acquises dans le chapitre de dénombrement : il trouve ici un champ d'application sur mesure.

Après une longue introduction pour mettre en perspective l'usage des probabilités, ce chapitre commencera par une rapide approche intuitive avant de formaliser rigoureusement l'ensemble des concepts des probabilités finies discrètes.
Les variables aléatoires et les lois usuelles (Bernoulli, binomiale...) sont vues à part dans le chapitre correspondant.

1 En guise d'introduction

Histoire des mathématiques : Les premières traces de raisonnements probabilistes ont été trouvées en Inde vers le IV$^{\text{ème}}$ siècle avant Jésus-Christ. La culture védique étudie les combinaisons possibles des syllabes qui permettent de composer des poèmes, les combinaisons des cinq sens...
Les probabilités modernes sont nées avec Pascal et Fermat au XVII$^{\text{e}}$ siècle. À l'origine, elles se concentrèrent essentiellement sur la théorie des jeux et des paris. Cette nouvelle science du pari fut une révolution dans notre capacité à penser l'avenir. Pascal transposa ses découvertes scientifiques dans le domaine de la métaphysique avec le fameux *pari de Pascal.*
Au début du XVIII$^{\text{e}}$, Bernoulli introduisit l'idée de *la loi des grands nombres* qui stipule que les résultats d'une expérience se rapprochent de la probabilité lorsque cette première est effectuée un très grand nombre de fois.
Au XX$^{\text{e}}$ siècle, Kolmogorov donne un langage formel aux probabilités et construit les bases de la théorie moderne des probabilités. L'intuition cède la place au formalisme mathématique. Ceci permet notamment l'étude des modèles infinis et continus en plus des modèles discrets. La nouvelle théorie tire partie de la récente définition de l'intégration par Lebesgue (un raffinement subtil de la théorie de l'intégration vue en MPSI).
Finalement, les probabilités sont restées fidèles à elles-mêmes car un de leur champ d'application important est actuellement dans les marchés financiers qui sont pour partie un jeu d'argent où l'on parie sur la fluctuation des cours.

De l'expérience physique aux mathématiques : On peut voir les probabilités comme une façon de prévoir l'avenir, ou tout au moins de faire des paris sur celui-ci. Les probabilités seraient donc la science des paris sur l'avenir, et c'est comme telles qu'elles sont utilisées.
Leur objet se rapproche de celui de la physique ou de la chimie qui tentent de prévoir l'évolution d'un système dans une situation donnée (prévoir la trajectoire d'une fusée en fonction de sa conception, la réaction d'un mélange si on lui ajoute un produit...).
Tout ceci s'appuie sur une connaissance *a priori* de lois qui régissent le système[1] et sur leur transcription mathématique. C'est la partie *modélisation* en laquelle interviennent des choix et hypothèses qui peuvent être remis en cause par la suite. Cette modélisation provient de l'observation attentive de la nature (par exemple, si je lâche une pomme, alors elle tombe vers le bas), mais pourra toujours être affinée voire contestée plus tard.
On distingue deux types de modèles :

- le modèle **déterministe** selon lequel une connaissance suffisamment précise des conditions de l'expérience permet de prévoir, *à coup sûr*, le résultat de celle-ci.
- le modèle **probabiliste** pour lequel il est impossible de savoir avec certitude l'issue de l'expérience. On étudie les différents résultats possibles et on estime leur probabilité d'occurrence : s'ils ont plus ou moins de chances de se produire.

1. Ce qui suppose un phénomène observable et que l'on peut reproduire, au moins par la pensée.

La frontière entre ces deux modèles est assez floue. On peut considérer que le modèle déterministe est un modèle probabiliste simplifié en lequel une issue possède une très forte probabilité de réalisation. On néglige alors les autres issues possibles et on fait l'approximation que cette issue *très probable* est *certaine*. A contrario, on peut aussi considérer que le modèle probabiliste décrit un système déterministe mais dont la méconnaissance des conditions précises de réalisation ne nous permet pas de prédire avec certitude son issue.

On voit dès lors que l'étape la plus difficile et la plus risquée est la *modélisation* en laquelle se concentrent tous les choix. Je laisserai donc les professeurs de Physique-Chimie, SVT... s'occuper de cette science difficile et nous nous concentrerons sur l'étape suivante : on suppose que le modèle est établi. Les issues possibles sont donc déjà toutes répertoriées et leur probabilités élémentaires sont estimées. À partir de ce modèle *élémentaire*, nous construirons des modèles plus complexes dont nous estimerons à nouveau les probabilités.

2 Vocabulaire

A Approche intuitive

Vocabulaire	Exemple du lancer de dé
Expérience aléatoire : une expérience dont on ne peut prévoir l'issue.	le lancer de dé
Issue (ou **événement élémentaire**) : résultat possible d'une expérience aléatoire.	1, 2, ... 6, sont les issues
Univers : ensemble des issues possibles d'une expérience aléatoire.	$\Omega = \{1, 2, 3, 4, 5, 6\}$

On peut choisir de regarder par exemple si le résultat du dé est « au moins égal à 4 ». Pour cela on crée l'événement[2] correspondant :

Événement : un ensemble d'issues parmi toutes celles possibles.	$A = \{4, 5, 6\}$ (c'est une partie de Ω).
Un événement A est **réalisé**, si le résultat de l'expérience aléatoire est dans A.	A est réalisé si on obtient 4, 5 ou 6.
Les issues qui composent A, sont appelées issues **favorables** à A (ou qui **réalisent** A).	4, 5 et 6
$\overline{A}$ est **l'événement contraire** de A.	$\overline{A} = \{1; 2; 3\}$

2. Depuis la réforme de l'orthographe de 1990, les deux graphies *événement* et *évènement* sont acceptées bien que l'Académie Française recommande la seconde en accord avec la prononciation. Il semble que les deux graphies aient toujours eu cours à proportions variables selon les époques, et les dictionnaires de l'Académie Française passent de l'une à l'autre selon les versions. Ici, nous utiliserons la graphie « classique » qui était majoritaire du temps de Pascal et qui est celle utilisée au programme officiel.

B Approche mathématique

Définition 2.1

On appelle **univers** un ensemble (fini non vide). On note souvent l'univers Ω.

Les **événements** sont les parties de Ω.

Les **événements élémentaires** sont les événements **singleton**.
On parle aussi d'**issues possibles** et on les note souvent ω.

Remarque : Dans le cadre de ce cours, nous nous limitons aux univers finis. La formalisation est plus complexe pour les univers infinis.

Propriété 2.2 *(Dénombrement)*

Si Card $(\Omega) = n$, alors il y a n événements élémentaires possibles et 2^n événements possibles.

Exemple *(Exemple suivi...)*

Si l'expérience consiste à lancer successivement deux dés, alors l'univers est l'ensemble des couples de $[\![1;6]\!]^2$.
Il y a donc Card $\left([\![1;6]\!]^2\right) = 36$ issues possibles.
Les événements correspondent aux parties de $[\![1;6]\!]^2$. Il y en a Card $(\mathcal{P}(\Omega)) = 2^{36}$.
On peut par exemple définir les événements

A : « les deux dés donnent chacun un résultat pair »
B : « les deux dés donnent le même résultat »

ainsi, $A = \left\{(a, b) \in [\![1;6]\!]^2,\ \text{tel que } a \text{ et } b \text{ pairs}\right\}$ et Card $(A) = 3 \times 3 = 9$.
et $B = \{(a, a),\ \text{tel que } a \in [\![1;6]\!]\}$, ce qui donne Card $B = 6$.

Définition 2.3 *(Logique et ensembles)*

1. L'événement « A **ou** B » correspond à l'événement « $A \cup B$ ».
2. L'événement « A **et** B » correspond à l'événement « $A \cap B$ ».
3. L'événement **contraire** de « A » correspond à l'événement « $\overline{A}$ ».

Exemple *(suite de l'exemple suivi)*

- « A ou B » est l'événement : « les dés sont tous les deux pairs, ou ils sont égaux ». On ajoute les combinaisons $(1, 1); (3, 3)$ et $(5, 5)$ à celles de A.
- « A et B » est l'événement : « les deux dés sont à la fois égaux et pairs ». $A \cap B = \{(2, 2), (4, 4), (6, 6)\}$.
- Le contraire de A est « au moins un des deux dés donne un résultat impair ».

Définition 2.4

Lorsque $A \cap B = \emptyset$, les deux événements A et B sont dits **incompatibles**.

En particulier, deux événements incompatibles ne peuvent pas être réalisés en même temps.

Exemple

A et $\overline{A}$ sont toujours incompatibles.

Exemple *(suite de l'exemple suivi)*

Si on définit l'événement C par « le premier dé donne un résultat pair et le second donne un résultat impair », alors C est incompatible avec B (car les deux dés ne peuvent pas être de parité contraire tout en donnant le même résultat).

Définition 2.5 *(Système complet d'événements)*

Soit Ω un univers et $\{A_i\}_{i \in I}$ un système d'événements.
Le système d'événements est dit **complet** s'il vérifie les deux conditions :

1. $\bigcup\limits_{i \in I} A_i = \Omega$,
2. $\forall (i,j) \in I^2, \;$ si $i \neq j$ alors $A_i \cap A_j = \emptyset$.

Remarque : Un système complet d'événements est un recouvrement disjoint de l'univers. Si on rajoute la condition, $\forall i \in I, \; A_i \neq \emptyset$, alors ce système forme une **partition** de Ω.

Explications

Un système complet d'événements correspond à une disjonction des cas : on liste l'ensemble des situations (disjointes) possibles.

Exemple

A et $\overline{A}$ forment un système complet d'événements.

Exemple

Avec le double lancer de dés, on peut définir un système complet :

- « Les deux dés donnent chacun un résultat pair »,
- « Les deux dés donnent chacun un résultat impair »,
- « Les deux dés donnent des résultats de parité différente ».

3 Probabilités

Jusqu'ici on n'a fait que décrire les issues possibles. Maintenant, on leur affecte des probabilités qui correspondent aux chances de succès de chaque issue.

Définition 3.1 *(Probabilité sur un univers)*

Soit un univers fini Ω,
une probabilité $\mathbf{P}$ sur Ω, est une application de $\mathcal{P}(\Omega)$ dans $[0,1]$ telle que

1. $\forall A \in \mathcal{P}(\Omega), \; 0 \leqslant \mathbf{P}(A) \leqslant 1$,
2. $\mathbf{P}(\Omega) = 1$,
3. Si $A \cap B = \emptyset$, alors $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.

Explications

Dans le cas fini, ces règles sont suffisantes pour définir proprement ce qu'est une probabilité.
Les trois axiomes formalisent des notions très intuitives :

1. La probabilité est dans $[0,1]$: entre 0% et 100% de chances d'être réalisé.
2. La probabilité d'avoir une issue qui fait partie de l'univers est 1 : il n'est pas possible d'avoir une autre issue que celles de l'univers.
3. Si on a « p chances de réaliser A » et « q chances de réaliser B » et que A et B ne contiennent aucune issue commune ($A \cap B = \emptyset$), alors on a $p+q$ chances de réaliser l'un des deux.
   C'est la traduction probabiliste du cardinal de deux ensembles disjoints.

Définition 3.2

- Si $\mathbf{P}(A) = 0$, alors l'événement A est négligeable ou impossible (cas fini).
- Si $\mathbf{P}(A) = 1$, alors l'événement A est presque sûr.

Avec un univers infini, lorsque $\mathbf{P}(A) = 0$, on dit que l'événement est négligeable, pour $\mathbf{P}(A) = 1$, qu'il est presque certain.

Définition 3.3 *(Espace probabilisé fini)*

Un **espace probabilisé fini** est un couple $(\Omega, \mathbf{P})$, avec Ω un univers fini, et $\mathbf{P}$ une probabilité sur Ω.

Exemple

Cas d'une épreuve de Bernoulli : tirage à pile ou face.
On code l'événement pile avec 1 et l'événement face avec 0.
Si la pièce est équilibrée, on a la même probabilité d'avoir pile ou face : $\frac{1}{2}$.
Ainsi, on définit l'espace probabilité $(\Omega, \mathbf{P})$, avec :

$$\Omega = \{0, 1\}$$

$$\mathbf{P} : \begin{cases} \emptyset & \mapsto 0 \\ \{0\} & \mapsto \frac{1}{2} \\ \{1\} & \mapsto \frac{1}{2} \\ \{0,1\} & \mapsto 1 \end{cases}$$

Exemple

Soit une urne contenant n boules dont p sont blanches et $n-p$ sont noires.
On tire une boule au hasard.
Quelle est la probabilité qu'elle soit blanche ? noire ?
Solution :
Si on note les événements B : « la boule est blanche », et N :« la boule est noire », alors

$$\mathbf{P}(B) = \frac{p}{n} \qquad \text{et} \qquad \mathbf{P}(N) = \mathbf{P}\left(\overline{B}\right) = \frac{n-p}{p}.$$

Ici, en accord avec l'énoncé, on a supposé que chaque boule de l'urne avait la même probabilité d'être tirée.

Les règles qui suivent sont capitales mais il n'y a aucun effort de mémoire à faire : il suffit de les comprendre pour être capable de les retrouver dès que nécessaire.

Corollaire 3.4 *(Règles sur les probabilités)*

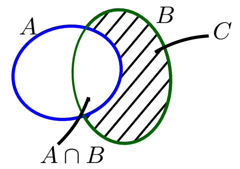
1. $\mathbf{P}(\Omega) = 1$.
2. $\mathbf{P}(\emptyset) = 0$: l'événement « vide » est impossible.
3. $\mathbf{P}\left(\overline{A}\right) = 1 - \mathbf{P}(A)$.
4. Si $A \subset B$, alors $\mathbf{P}(A) \leqslant \mathbf{P}(B)$ (croissance).
5. Si $A \cap B = \emptyset$ alors $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.
6. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.
7. $\mathbf{P}(A \cup B) \leqslant \mathbf{P}(A) + \mathbf{P}(B)$.
8. Si $(A_i)_{1 \leqslant i \leqslant n}$ sont incompatibles deux à deux, alors $\mathbf{P}\left(\bigcup\limits_{1 \leqslant i \leqslant n} A_i\right) = \sum\limits_{i=1}^{n} \mathbf{P}(A_i)$.
9. $\mathbf{P}\left(\bigcup\limits_{1 \leqslant i \leqslant n} A_i\right) \leqslant \sum\limits_{i=1}^{n} \mathbf{P}(A_i)$.

Explications

$\mathbf{P}(\Omega) = 1$: toutes les issues possibles de l'expérience sont dans l'univers.
$\mathbf{P}(\emptyset) = 0$: l'expérience donne nécessairement un résultat.

Preuve

1. cela fait partie de la définition.
2. $\Omega \cap \emptyset = \emptyset$, donc $\mathbf{P}(\Omega) = \mathbf{P}(\Omega \cup \emptyset) = \mathbf{P}(\Omega) + \mathbf{P}(\emptyset)$. Donc $\mathbf{P}(\emptyset) = 0$.
3. $A \cap \overline{A} = \emptyset$ et $A \cup \overline{A} = \Omega$, donc $1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup \overline{A}) = \mathbf{P}(A) + \mathbf{P}(\overline{A})$.
Donc $\mathbf{P}\left(\overline{A}\right) = 1 - \mathbf{P}(A)$.
4. Si $A \subset B$, alors $B = A \cup (B \setminus A)$ et les deux ensembles sont disjoints.
Donc $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \geqslant \mathbf{P}(A)$ par positivité de la probabilité.
5. fait partie de la définition
6. *Correspond à Grassmann pour les ensembles finis, c'est la même preuve et on peut la visualiser sur un schéma*
Soit $C = B \setminus A$, alors $A \cap C = \emptyset$, et $A \cup C = A \cup B$.
Donc $\mathbf{P}(A \cup B) = \mathbf{P}(A \cup C) = \mathbf{P}(A) + \mathbf{P}(C)$. $(\star)$
Or $C \cup (A \cap B) = B$ et $C \cap (A \cap B) = \emptyset$.
Donc $\mathbf{P}(B) = \mathbf{P}(C \cup (A \cap B)) = \mathbf{P}(C) + \mathbf{P}(A \cap B)$
Donc avec $(\star)$ $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.



7. Découle du 6 et de la positivité de la probabilité.
8. On montre la propriété avec une récurrence sur n.
Le résultat est vrai pour $n = 2$ (déjà vu),
Pour $n \geqslant 3$ fixé, on suppose la propriété vraie au rang $n-1$ et on considère n événements incompatibles $(A_i)_{1 \leqslant i \leqslant n}$.
On vérifie d'abord que A_n est incompatible avec $\bigcup_{1 \leqslant i \leqslant n-1} A_i$.
C'est évident car $A_n \cap \left(\bigcup_{1 \leqslant i \leqslant n-1} A_i\right) = \bigcup_{1 \leqslant i \leqslant n-1} (A_n \cap A_i) = \bigcup_{1 \leqslant i \leqslant n-1} \emptyset = \emptyset$
Donc on peut appliquer le cas $n = 2$
$$\begin{aligned} \mathbf{P}\left(\left(\bigcup_{1 \leqslant i \leqslant n-1} A_i\right) \cup A_n\right) &= \mathbf{P}\left(\bigcup_{1 \leqslant i \leqslant n-1} A_i\right) + \mathbf{P}(A_n) \\ &= \sum_{i=1}^{n} \mathbf{P}(A_i) \qquad \text{(hypothèse de récurrence)} \end{aligned}$$
9. On considère pour chaque i, $B_i = A_i \backslash \left(\bigcup_{j=1}^{i-1} A_j\right)$ de telle sorte que les $(B_i)_{1 \leqslant i \leqslant n}$ soient deux à deux incompatibles et que leur union soit la même que celle des $(A_i)_{1 \leqslant i \leqslant n}$.
On applique ensuite le point précédent aux $(B_i)_{1 \leqslant i \leqslant n}$, et on utilise la croissance de la probabilité : $\forall i \in [\![1, n]\!]$, $\mathbf{P}(B_i) \leqslant \mathbf{P}(A_i)$.

■

Théorème 3.5 *(Description par les événements élémentaires)*

Une probabilité $\mathbf{P}$ sur un univers fini $\Omega = (\omega_1, \omega_2, \cdots, \omega_n)$ est entièrement déterminée par les valeurs des $(p_i = \mathbf{P}(\omega_i))_{i \in [\![1,n]\!]}$.

Dans ce cas, $\mathbf{P}$ est une probabilité si et seulement si $\forall i \in [\![1,n]\!]$, $p_i \geqslant 0$ et $\sum_{i=1}^{n} p_i = 1$.

La probabilité d'un événement A est la somme des probabilités des issues qui le composent :

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

Concrètement avec un univers fini : on associe une probabilité $p_i \in [0,1]$ à chaque issue et la somme des probabilités « élémentaires » est égale à 1.

Preuve

Trivial ■

Définition 3.6 *(Modèle équiprobable)*

Si tous les p_i sont égaux[3], la situation est dite **équiprobable**.

Explications

C'est le modèle de base pour la plupart des expériences. Tirage aux dés, tirage de cartes, pile ou face...

Théorème 3.7 *(Probabilité d'un événement dans un modèle équiprobable)*

Si Ω est un univers fini composé de n issues possibles, et si $\mathbf{P}$ définit un modèle équiprobable sur Ω,

alors la probabilité d'une issue est $p_i = \dfrac{1}{n}$.

Si A est un événement, alors $\mathbf{P}(A) = \dfrac{\text{Card}\,(A)}{\text{Card}\,(\Omega)}$.

Explications

$$\mathbf{P}(A) = \frac{\text{nombre de cas favorables à } A}{\text{nombre de cas au total}}.$$

La détermination d'une probabilité ramène souvent à un problème de dénombrement.

Exemple

Lors du lancer successif de deux dés, la probabilité d'avoir une paire est $\frac{1}{6}$
En effet, le nombre total de possibilités est $\text{Card}\,([\![1;6]\!]^2) = 36$ et le nombre de façon d'avoir une paire est 6 (pour la paire de 1, la paire de 2, ... , la paire de 6).
La probabilité totale est donc bien $\frac{6}{36} = \frac{1}{6}$.

3. Toutes les issues ont la même chance de se produire.

Exemple *(Fondamental)*

Une urne contient n boules : p blanches et $n-p$ noires. On effectue deux tirages. Donner la probabilité d'obtenir deux boules blanches selon le type de tirage.

Solution :

Si les tirages sont **successifs sans remise**, alors il y a $n(n-1)$ tirages possibles, et parmi ces tirages possibles, $p(p-1)$ tirages donnent deux boules blanches.
La probabilité est donc $\mathbf{P}(B,B) = \frac{p(p-1)}{n(n-1)}$.

Si les tirages sont **simultanés**, alors il y a $\binom{n}{2}$ tirages possibles, et parmi ces tirages, $\binom{p}{2}$ donnent deux boules blanches.

La probabilité est donc $\mathbf{P}(B,B) = \frac{\binom{p}{2}}{\binom{n}{2}} = \frac{p(p-1)}{n(n-1)}$.

Si les tirages sont **successifs avec remise**, alors il y a n^2 tirages possibles, et parmi ces tirages, p^2 donnent deux boules blanches.
La probabilité est donc $\mathbf{P}(B,B) = \frac{p^2}{n^2}$.

⚠ On voit que les hypothèses de tirage peuvent modifier les résultats et qu'il faut bien lire **toutes les hypothèses**.

Méthodes usuelles pour déterminer une probabilité :

Il existe plusieurs méthodes pour construire des probabilités sur un univers :

- Toutes les issues ont la même probabilité d'occurrence : **modèle équiprobable**.
- La répétition de l'expérience aléatoire donne la fréquence d'occurrence de chaque issue et permet d'établir le modèle.
- La probabilité est calculée à partir d'autres modèles mécaniques, mathématiques...

4 Probabilités conditionnelles

L'explication des probabilités conditionnelles est très simple : en fonction des informations dont on dispose, les probabilités ne sont pas identiques.
Dans les marchés financiers par exemple, si on est bien informé, la probabilité de gains est a priori supérieure : c'est le problème des délits d'initiés.

A Probabilité conditionnelle de B sachant A

Définition 4.1 *(Probabilité conditionnelle de B sachant A)*

A et B sont deux événements, tels que $\mathbf{P}(A) \neq 0$.
Sachant que A est réalisé, la probabilité que B le soit aussi est :

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

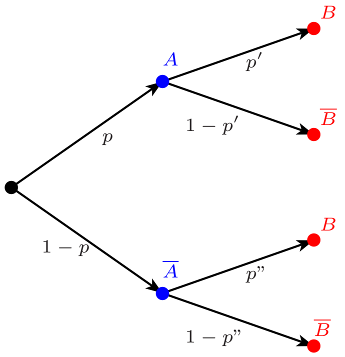
On lit : « la probabilité de B sachant A ».
Les deux notations $\mathbf{P}(B|A)$ ou $\mathbf{P}_A(B)$ sont équivalentes.

Explications

On sait que A est réalisé et on restreint donc l'univers à A.
Dans le cas équiprobable, cela revient au calcul :

$$\mathbf{P}_A(B) = \frac{\text{Card }(A\cap B)}{\text{Card }(A)} = \frac{\text{Card }(A\cap B)\,\text{Card }(\Omega)}{\text{Card }(A)\text{Card }(\Omega)} = \frac{\text{Card }(A\cap B)}{\text{Card }(\Omega)}\frac{\text{Card }(\Omega)}{\text{Card }(A)}$$
$$= \mathbf{P}(A\cap B)\frac{1}{\mathbf{P}(A)}.$$

On peut essayer de comprendre cela avec des arbres de probabilité comme en terminale. Ceci ne constitue pas une *justification*, car les calculs de l'arbre utilisent cette définition et celles qui suivent, mais cela a au moins le mérite de donner une visualisation à cette formule.



Les probabilités conditionnelles se lisent le long des branches.
On peut par exemple, suivre la branche supérieure et on obtient :
$\mathbf{P}\left(A\cap B\right) = pp' = \mathbf{P}(A)\mathbf{P}_A(B)$ avec $\mathbf{P}(A) = p$ et $\mathbf{P}_A(B) = p'$.

Théorème 4.2 *(Formule des probabilités composées)*

Pour deux événements A et B, avec $\mathbf{P}(A) \neq 0$,

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

Pour n événements $A_1, A_2, \cdots, A_n$,

$$\mathbf{P}(A_1\cap A_2\cap A_3\cap\cdots\cap A_n) = \mathbf{P}(A_1)\mathbf{P}_{A_1}(A_2)\mathbf{P}_{A_1\cap A_2}(A_3)\cdots\mathbf{P}_{A_1\cap A_2\cap\ldots\cap A_{n-1}}(A_n).$$

(on suppose que $\mathbf{P}(A_1), \mathbf{P}(A_1\cap A_2), \cdots, \mathbf{P}\left(A_1\cap\cdots\cap A_{n-1}\right)$ sont tous non nuls)

Preuve

La formule pour deux événements s'obtient directement avec la définition.
La formule générale, s'obtient par récurrence.
L'initialisation est immédiate pour $n = 1$.
Soit $n \geqslant 1$. On suppose que le résultat est vrai au rang n et on considère $(n+1)$ événements $(A_i)_{1\leqslant i\leqslant n+1}$ qui vérifient les hypothèses du théorème.
On applique l'hypothèse de récurrence aux n événements $A_1, A_2, \cdots, A_{n-1}$ et $A_n\cap A_{n+1}$.
Ainsi $\mathbf{P}(A_1\cap A_2\cap A_3\cap\cdots\cap A_n\cap A_{n+1})$

$$= \mathbf{P}(A_1)\,\mathbf{P}_{A_1}(A_2)\,\mathbf{P}_{A_1\cap A_2}(A_3)\ \cdots\ \mathbf{P}_{A_1\cap A_2\cap\ldots\cap A_{n-1}}(A_n\cap A_{n+1}).$$

Pour simplifier on note $B = A_1\cap A_2\cap\cdots\cap A_{n-1}$, et on trouve

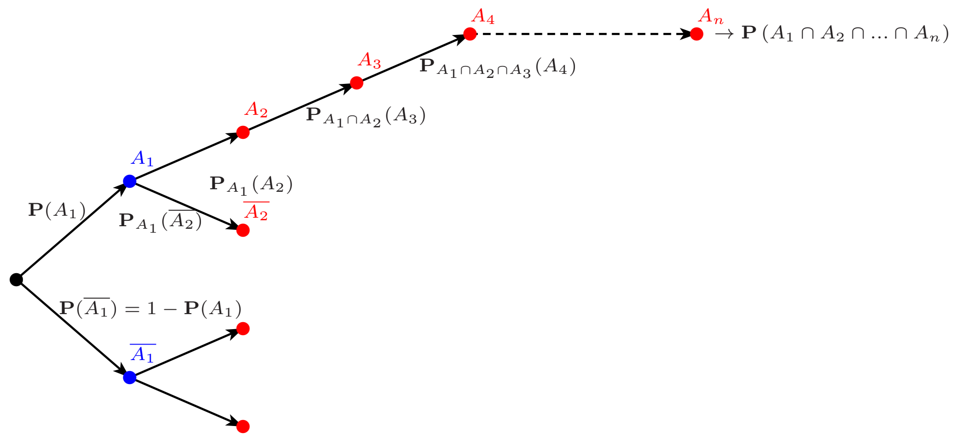
$$\mathbf{P}_B\left(A_n\cap A_{n+1}\right) = \frac{\mathbf{P}\left(B\cap A_n\cap A_{n+1}\right)}{\mathbf{P}(B)} = \frac{\mathbf{P}\left(B\cap A_n\cap A_{n+1}\right)}{\mathbf{P}(B)}\ \frac{\mathbf{P}\left(B\cap A_n\right)}{\mathbf{P}\left(B\cap A_n\right)}$$
$$= \frac{\mathbf{P}\left(B\cap A_n\cap A_{n+1}\right)}{\mathbf{P}\left(B\cap A_n\right)}\ \frac{\mathbf{P}\left(B\cap A_n\right)}{\mathbf{P}\left(B\right)}$$
$$= \mathbf{P}_{B\cap A_n}\left(A_{n+1}\right)\ \mathbf{P}_B\left(A_n\right).$$

D'où le résultat voulu. ∎

Remarque : Il suffit de supposer $\mathbf{P}\left((A_1\cap\cdots\cap A_{n-1}\right) \neq 0$ car alors, par croissance de la probabilité, pour tout $k \leqslant n-1$, $\mathbf{P}\left(A_1\cap\cdots\cap A_k\right) \geqslant \mathbf{P}\left(A_1\cap\cdots\cap A_n\right) > 0$.

Explications

En reprenant un arbre de probabilité, on suit une branche qui réalise les événements A_i.



Remarque : Avec cette formule, on a exigé que $\mathbf{P}(A) \neq 0$ pour que la probabilité conditionnelle $\mathbf{P}_A(B)$ ait un sens.
Néanmoins, en cas de nullité de $\mathbf{P}(A)$, on peut encore donner un sens à cette écriture.
En effet, si $\mathbf{P}(A) = 0$, alors par croissance des probabilités $\mathbf{P}(A\cap B) = 0$. De plus, toute probabilité étant bornée, le membre de droite s'écrit $0\times$ "quantité bornée" $= 0$.
On peut donc en conclure que $\mathbf{P}(A)\mathbf{P}_A(B) = 0$, et ce, même si la deuxième quantité est prise arbitrairement dans $[0, 1]$.
Ainsi, l'égalité reste vérifiée dans le cas $\mathbf{P}(A) = 0$.
Il en est de même pour la formule des probabilités composées.

Exemple

Soit une urne contenant n boules blanches et n boules noires.
On tire n boules, quelle est la probabilité de n'avoir que des blanches ?
Indication : distinguer les situations selon les types de tirage.
Solution :

- si les tirages sont simultanés, $p = \frac{(n!)^2}{(2n)!}$.
- si les tirages sont successifs avec remise. Le tirage complet peut être représenté par un n-uplet, avec 2^n cas possibles et un cas favorable. $p = \frac{1}{2^n}$.
- si les tirages sont successifs sans remise, on note A_i l'événement : « on obtient une boule blanche au i-ème tirage ».

$$\begin{aligned}\mathbf{P} &= \mathbf{P}(A_1 \cap A_2 \cap \cdots \cap A_n)\\ &= \mathbf{P}(A_1)\,\mathbf{P}_{A_1}(A_2)\,\mathbf{P}_{A_1\cap A_2}(A_3)\cdots \mathbf{P}_{A_1\cap A_2\cap\ldots\cap A_{n-1}}(A_n) \qquad \text{(probabilités composées)}\end{aligned}$$

Or $\mathbf{P}_{A_1\cap A_2\cap\ldots\cap A_i}(A_{i+1}) = \frac{n-i}{2n-i}$ car si A_1, A_2, $\cdots$ A_i sont réalisés, c'est que l'on a tiré i boules blanches dans l'urne. Il reste donc $n-i$ boules blanches sur un total de $(2n-i)$ boules pour le tirage suivant. Donc

$$\begin{aligned}p &= \frac{n}{2n}\,\frac{n-1}{2n-1}\,\frac{n-2}{2n-2}\cdots\frac{n-(n-1)}{2n-(n-1)}\\ &= \frac{n(n-1)(n-2)\cdots 1}{(2n)(2n-1)(2n-2)\cdots(n+1)} = \frac{(n!)^2}{(2n)!}.\end{aligned}$$

On retrouve la probabilité du tirage simultané. C'est naturel car on s'intéresse au tirage d'une seule couleur et que l'ordre n'a donc pas d'importance.

Théorème 4.3

Soit Ω un univers fini et $\mathbf{P}$ une probabilité sur Ω.
Soit $A \subset \Omega$ tel que $\mathbf{P}(A) > 0$.

$$\mathbf{P}_A \text{ est une probabilité sur } \Omega.$$

Preuve

1. $\mathbf{P}_A$ est bien définie sur Ω à valeurs dans $\mathbf{R}$.
2. Soit $B \subset \Omega$, alors $B \cap A \subset \Omega$ donc $\mathbf{P}(B \cap A)$ est bien défini et positif.
   De plus, par croissance de la probabilité, $\mathbf{P}(B \cap A) \leqslant \mathbf{P}(A)$ donc $\mathbf{P}_A(B) \in [0, 1]$.
3. Si $B \cap C = \emptyset$, alors $(A \cap B) \cap (A \cap C) = A \cap B \cap C = \emptyset$
   ce qui donne donc
   $$\mathbf{P}(A \cap (B \cup C)) = \mathbf{P}((A \cap B) \cup (A \cap C)) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap C).$$
   Donc
   $$\begin{aligned}\mathbf{P}_A(B \cup C) &= \frac{\mathbf{P}(A \cap (B \cup C))}{\mathbf{P}(A)}\\ &= \frac{\mathbf{P}(A \cap B) + \mathbf{P}(A \cap C)}{\mathbf{P}(A)}\\ &= \mathbf{P}_A(B) + \mathbf{P}_A(C).\end{aligned}$$

∎

Explications

Donner un conditionnement, c'est changer de probabilité.
D'un point de vue plus concret, on peut interpréter toute probabilité comme une probabilité conditionnelle, relativement à « tout ce que l'on sait », ou plus simplement, « sachant l'énoncé de l'exercice ».
Par exemple, on voit bien que répondre que la probabilité que la couleur de la boule soit blanche $\mathbf{P}(B) = p$ dépend de l'expérience décrite dans l'énoncé.
Dans certains exercices, l'astuce sera donc de trouver le bon conditionnement pour travailler avec une probabilité avec laquelle on saura faire les calculs.

B Indépendance

Définition 4.4

Deux événements A et B sont **indépendants**
si, et seulement si, $\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$.

La propriété suivante, sert à donner une interprétation plus concrète de l'indépendance, mais elle n'est valable que lorsque $\mathbf{P}(A) \neq 0$.

Propriété 4.5

Si $\mathbf{P}(A) \neq 0$, A et B sont indépendants si et seulement si $\mathbf{P}_A(B) = \mathbf{P}(B)$.

Explications

Cela signifie que le fait que l'on sache que A est réalisé (ou non) n'influence pas notre connaissance sur l'éventuelle réalisation de B.
Comme nous allons le voir dans l'exemple suivant, la notion d'indépendance ne traduit pas nécessairement un lien de cause à conséquence ou une influence directe entre deux événements, mais il s'agit davantage d'une question ayant trait à l'*information reçue*.
Si A influence B, le fait d'avoir une information sur A, nous en donnera également sur B, et les événements ne seront pas indépendants. Par contre, A peut très bien n'avoir aucune influence sur B, et pourtant ne pas être indépendant de B.
Pour tenter d'être plus clair, les deux événements peuvent se réaliser sans influence mutuelle, et pourtant, si je sais que l'un est réalisé, cela peut me donner des informations sur un *contexte général*, qui favoriserait ou défavoriserait la réalisation de B.

Cela invite à comprendre que la notion de probabilité ne désigne donc pas tant les chances qu'un événement se réalise, que ma capacité à prévoir son éventuelle réalisation (indépendamment de ses chances intrinsèques d'arriver).
On restera néanmoins très prudent face à son intuition : les cas d'indépendance peuvent se révéler hautement contre-intuitifs. C'est d'abord l'affaire d'un calcul.

Propriété 4.6

A et B sont indépendants, si, et seulement si A et $\overline{B}$ sont indépendants,
si, et seulement si $\overline{A}$ et $\overline{B}$ sont indépendants.

Preuve

$\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}\left(A \cap \overline{B}\right)$, donc $\mathbf{P}\left(A \cap \overline{B}\right) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$.
Ainsi, pour A et B indépendants, $\mathbf{P}\left(A \cap \overline{B}\right) = \mathbf{P}(A) - \mathbf{P}(A)\mathbf{P}(B)$

$$= \mathbf{P}(A)(1 - \mathbf{P}(B)) = \mathbf{P}(A)\mathbf{P}\left(\overline{B}\right).$$

Donc A et $\overline{B}$ indépendants.

La réciproque s'obtient en appliquant le même résultat à A et $\overline{B}$, car $B : \overline{\overline{B}}$.
La seconde équivalence en découle en appliquant le complémentaire successivement à A et à B. ■

Exemple

Supposons que nous disposions d'une urne avec n boules de couleur blanche ou noire.

Je tire n boules successivement sans vous les montrer et vous devez deviner la couleur de la dernière tirée. Le but est de savoir si l'annonce des couleurs des $(n-1)$ premières boules vous aide à deviner la couleur de la dernière.

Au moment où vous cherchez à deviner, le tirage a déjà été réalisé. Il ne s'agit donc pas d'évaluer la probabilité que la dernière boule soit d'une couleur ou d'une autre : moi qui la vois, je sais avec certitude. Par contre, il est question évaluer vos chances sur chaque hypothèse afin de choisir la plus probable en fonction de l'information dont vous disposez.

On note A l'événement correspondant aux couleurs obtenues aux $n-1$ premiers tirages (connus) et B_n l'événement selon lequel le dernier tirage donne une boule blanche. On cherche donc à comparer les probabilités $\mathbf{P}(B_n)$ et $\mathbf{P}_A(B_n)$ pour voir si ces événements sont indépendants[4].
Si on suppose que l'urne contient p boules blanches et $n-p$ boules noires.
Si vous n'avez aucune information sur les $n-1$ premières boules tirées, alors, la probabilité que la dernière boule soit blanche est : $\mathbf{P}(B_n) = \frac{p}{n}$.

Voyons si cela est changé lorsque l'on vous donne l'information des $n-1$ premières boules tirées :

1. Si le tirage est sans remise, cela veut dire que vous pouvez savoir avec certitude la couleur de la n-ième boule tirée en fonction des $(n-1)$ précédentes.
   En effet, il ne reste plus qu'une seule boule à tirer et on connait la composition initiale de l'urne ainsi que les couleurs des $(n-1)$ autres boules tirées.
   En fonction des couleurs annoncées précédemment, la probabilité sera donc
   - soit $\mathbf{P}_A(B_n) = 1$ si on n'a tiré que $(p-1)$ boules blanches parmi les $(n-1)$ premières,
   - soit $\mathbf{P}_A(B_n) = 0$ si on a déjà annoncé avoir tiré toutes les boules blanches parmi les $(n-1)$ premières.

   Pour ce type de tirage **les événements ne sont pas indépendants** puisque la connaissance des $(n-1)$ premiers tirages me donne une information sur le dernier : la probabilité n'est plus la même.
2. Si le tirage est réalisé avec remise.
   Dans ce cas, à chaque tirage, la probabilité d'avoir une boule blanche est $\frac{p}{n}$.
   La connaissance des $n-1$ premiers tirages ne me donne absolument aucune information supplémentaire sur ce qu'a été ou sera le dernier tirage. **Les événements sont donc indépendants.**

4. Ici, puisque l'on sait que A est réalisé, il est nécessairement de probabilité non nulle.

> La notion d'indépendance n'est pas une relation d'influence, mais de connaissance.
> L'indépendance entre événements dépend du choix de la loi de probabilité.

Des événements indépendants pour une loi de probabilité, peuvent ne pas l'être pour une autre, même si l'expérience aléatoire est la même.

Exemple

On tire une carte d'un jeu de 52 cartes.
On définit deux événements A : « la carte est un as », B : « la carte est un cœur ».
Ces deux événements sont-ils indépendants ?
Solution :
Oui : que la carte soit un as ou non, la probabilité que ce soit un cœur est toujours $1/4$.

Exemple *(D'après Jacques Rouxel)*

> À gauche du ciel, il y avait la planète Shadok. Elle n'avait pas de forme spéciale... ou plutôt... elle changeait de forme.
> Comme la planète Shadok changeait de forme, il y avait des Shadoks qui tombaient. C'était très gênant... surtout pour les Shadoks.
> Les Shadoks en eurent donc assez, au bout d'un certain temps, de vivre sur une planète qui ne marchait pas bien... alors ils décidèrent d'aller sur la terre... qui avait l'air de mieux marcher...
> Les Shadoks décidèrent donc de construire une fusée interplanétaire. Mais ces malheureuses bêtes n'avait pas de connaissances spéciales en astronautique. D'ailleurs, elles n'avaient pas tellement... de connaissances spéciales. On peut même dire qu'elles n'avaient pas de connaissances du tout...
> À chaque essai de la fusée, la fusée retombait et se cassait... la situation était satisfaisante : les essais rataient très bien ! Car c'était un des principes de base de la logique shadok : « ce n'est qu'en essayant continuellement que l'on finit par réussir. » Ou, en d'autres termes :
>
> **Plus ça rate, plus on a de chances que ça marche.**
>
> Leur fusée n'était pas très très au point mais ils avaient calculé qu'elle avait quand même une chance sur un million de marcher...
> Et ils se dépêchaient de bien rater les 999 999 premiers essais pour être sûrs que le millionième marche.

Commenter la logique shadok.

Définition 4.7 *(Famille d'événements mutuellement indépendants)*

Soit $(A_i)_{1\leqslant i\leqslant n}$ une famille d'événements pour $(\Omega, \mathbf{P})$,
On dit que les $(A_i)_{1\leqslant i\leqslant n}$ sont **mutuellement indépendants**,
si *pour toute partie* I de $[\![1;n]\!]$

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

Exercice
Combien d'égalités faut-il vérifier ?
Solution :
Le nombre de parties de I, sauf celle vide et celles à un seul élément : $2^n - (n+1)$.

Exemple
Avec 3 événements A, B et C, il faut vérifier les égalités :

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

⚠ Des événements peuvent être indépendants deux à deux sans être mutuellement indépendants. Cela doit être vrai **pour toute partie**.

Exemple
On lance deux dés et on définit les 3 événements :

- A :« Le résultat du premier dé est 1 »,
- B :« Le résultat du second dé est 1 »,
- C :« Les deux dés donnent le même résultat ».

Étudier l'indépendance deux à deux et l'indépendance mutuelle.
Solution :
$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{6}$ et $\mathbf{P}(A\cap B) = \mathbf{P}(A\cap C) = \mathbf{P}(B\cap C) = \frac{1}{36}$.
Donc les événements sont deux à deux indépendants.
Pourtant $\mathbf{P}(A\cap B\cap C) = \frac{1}{36} \neq \mathbf{P}(A)\,\mathbf{P}(B)\,\mathbf{P}(C)$.
Donc les événements ne sont pas mutuellement indépendants.

⚠ $\mathbf{P}(A\cap B\cap C) = \mathbf{P}(A)\,\mathbf{P}(B)\,\mathbf{P}(C) \quad \not\Rightarrow \quad A,\ B,\ C$ mutuellement indépendants.

Exemple
On lance un dé et on définit 3 événements :

- A :« Le résultat du dé est 1, 2 ou 3 »,
- B :« Le résultat du dé est 1, 3 ou 5 »,
- C :« Le résultat du dé est 1, 2, 4 ou 5 ».

Étudier l'indépendance mutuelle.
Solution :

$$\mathbf{P}(A) = \frac{1}{2}, \qquad \mathbf{P}(B) = \frac{1}{2}, \qquad \mathbf{P}(C) = \frac{2}{3}.$$

Or $A\cap B\cap C$ est l'événement : « le résultat du dé est 1 », donc $\mathbf{P}(A\cap B\cap C) = \frac{1}{6} = \frac{1}{2}\frac{1}{2}\frac{2}{3}$.

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

Mais $A\cap B$ est l'événement : « le résultat du dé est 1 ou 3 ».
Donc $\mathbf{P}(A\cap B) = \frac{1}{3} \neq \frac{1}{4} = \mathbf{P}(A)\mathbf{P}(B)$.
Les événements ne sont donc pas mutuellement indépendants.
On voit que $\mathbf{P}(A\cap B\cap C) = \mathbf{P}(A)\,\mathbf{P}(B)\,\mathbf{P}(C) \not\Rightarrow A,\ B,\ C$ mutuellement indépendants.

Propriété 4.8

L'indépendance passe aux sous familles :
Si $(A_i)_{i\in I}$ est une famille d'événements mutuellement indépendants, alors pour toute sous famille $(A_j)_{i\in J}$, les événements sont aussi mutuellement indépendants.

Preuve
Une sous-famille de $(A_j)_{j\in J}$ est une sous-famille de $(A_i)_{i\in I}$. ■

Remarque : La réciproque est aussi vraie en prenant la sous-famille pour laquelle $J = I$.
Une propriété similaire existe pour les événements deux à deux indépendants.

C Formule de Bayes

Les formules qui suivent ne doivent pas solliciter la mémoire, mais l'intelligence. Si elles sont bien comprises, il n'y a rien à apprendre.

L'idée de la formule de Bayes, est d'utiliser la symétrie du produit dans la définition des probabilités conditionnelles.

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

On peut donc « inverser » A et B dans les relations conditionnelles :

Propriété 4.9

Soit A et B deux événements de $(\Omega, \mathbf{P})$ tels que $\mathbf{P}(A) > 0$,

$$\mathbf{P}_A(B) = \frac{\mathbf{P}_B(A)\mathbf{P}(B)}{\mathbf{P}(A)}.$$

La formule de Bayes va généraliser cette relation.

Propriété 4.10 *(Formule des probabilités totales)*

Soit A un événement de $(\Omega, \mathbf{P})$, et $(B_i)_{1 \leqslant i \leqslant n}$ un système complet d'événements tous de probabilités non nulles

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

Preuve

D'un point de vue ensembliste, les B_i forment une partition de Ω.

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{n} B_i \right) = \bigcup_{i=1}^{n} (A \cap B_i).$$

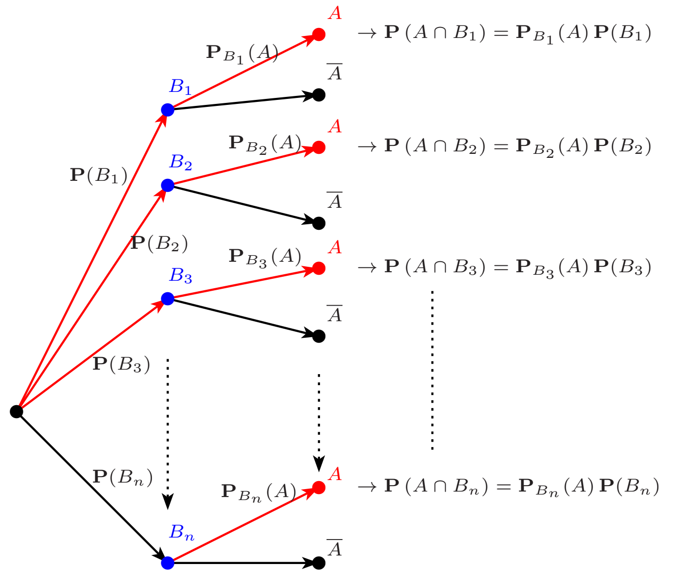
Comme les B_i forment une partition, l'union est disjointe, donc

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

■

Explications

Sur un arbre de probabilités, cela revient à sommer tous les chemins qui aboutissent à A en passant par chacun des B_k.



Remarque : En toute rigueur, on exige du système complet d'événements qu'il n'admette aucun événement de probabilité nulle. Néanmoins, la remarque au sujet des probabilités conditionnelles en cas de nullité reste vraie ici.

Exemple *(Système quasi-complet)*

Un système d'événements $\{A_i\}_{i \in [\![1,\, n]\!]}$ est dit « quasi-complet » si :

- $\forall (i,\, j) \in [\![1,\, n]\!],\, i \neq j \Rightarrow A_i \cap A_j = \emptyset$,
- $\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1$.

Montrer que la formule des probabilités totales s'applique encore avec un tel système.

On remarque deux différences avec la formule du cours : l'union n'est pas nécessairement égale à Ω, et les événements ne sont plus supposés de probabilité non nulle.

Solution :

On note pour tout $i \in [\![1,\, n]\!]$, $B_i = A_i$ et $B_{n+1} = \Omega \setminus \bigcup_{i=1}^{n} A_i = \overline{\bigcup_{i=1}^{n} A_i}$.
Alors, il est clair que $\{B_i\}_{i \in [\![1,\, n+1]\!]}$ forme un système complet d'événements.

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{n+1} B_i \right) = \bigcup_{i=1}^{n+1} (A \cap B_i).$$

L'union est disjointe, donc

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

Or $A \cap B_{n+1} \subset B_{n+1}$ et $\mathbf{P}(B_{n+1}) = 1 - \mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = 0$, donc par croissance de la probabilité, $0 \leqslant \mathbf{P}(A \cap B_{n+1}) \leqslant 0$, donc $\mathbf{P}(A \cap B_{n+1}) = 0$.
On peut donc écrire :

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

Si l'un des B_i est de probabilité nulle, alors, toujours par croissance des probabilités, on obtient également $\mathbf{P}(A \cap B_i) = 0$ et on peut le retirer de la somme avant de passer au probabilités conditionnelles.
Néanmoins, avec la remarque vue plus haut, et en prenant comme convention que « quantité arbitraire finie » $\times 0 = 0$, on peut écrire (abusivement) $\mathbf{P}_{B_i}(A)\,\mathbf{P}(B_i) = 0 = \mathbf{P}(A \cap B_i)$ et on trouve alors

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

Théorème 4.11 *(Formule de Bayes)*

Soit A un événement de $(\Omega, \mathbf{P})$, et $(B_i)_{1 \leqslant i \leqslant n}$ un système complet d'événements, Si $\mathbf{P}(A) > 0$, alors

$$\mathbf{P}_A(B_k) = \frac{\mathbf{P}_{B_k}(A)\mathbf{P}(B_k)}{\mathbf{P}(A)} = \frac{\mathbf{P}_{B_k}(A)\mathbf{P}(B_k)}{\sum\limits_{i=1}^{n} \mathbf{P}_{B_i}(A)\mathbf{P}(B_i)}.$$

Exemple *(Exemple classique à travailler)*

Une population est atteinte d'un virus. On sait que la proportion de personnes atteintes est 10^{-4}.
Un test de dépistage a été mis au point. Les expérimentations ont permis de savoir que les probabilités que l'individu soit détecté positif s'il est atteint ou s'il ne l'est pas sont respectivement égales à $0,99$ et à $0,001$. Sachant que le test donne un résultat positif, quelle est la probabilité que l'individu soit effectivement atteint ?

Solution :
Il est important de bien transcrire l'énoncé en langage mathématique.

- On note
  - A :« l'individu est atteint du virus »,
  - B :« le test est positif ».
- On sait que
  - $\mathbf{P}(A) = 10^{-4}$,
  - $\mathbf{P}_A(B) = 0,99$ et $\mathbf{P}_{\overline{A}}(B) = 0,001$,

On cherche $\mathbf{P}_B(A)$.
$(A, \overline{A})$ forme un système complet d'événements et on peut donc appliquer la formule de Bayes :

$$\mathbf{P}_B(A) = \frac{\mathbf{P}_A(B)\mathbf{P}(A)}{\mathbf{P}(B)} = \frac{\mathbf{P}_A(B)\mathbf{P}(A)}{\mathbf{P}_A(B)\mathbf{P}(A) + \mathbf{P}_{\overline{A}}(B)\mathbf{P}(\overline{A})} \approx 9\%$$

(en utilisant $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$).

Interprétation : La probabilité d'être atteint du virus quand le test est positif n'est donc que de 9%. Cela vient du fait que le test a beau être très fiable, la très faible[5] proportion de personnes atteintes dans la population réduit la probabilité.
On comprend mieux ce résultat en étudiant ce tableau :

	positif	négatif	total
malade	99	1	100
sain	1 000	998 900	999 900

On a supposé que l'on effectuait le test sur une population de 1 000 000 de personnes.
Parmi celles-ci, une proportion de 10^{-4} sont atteintes, donc 100 personnes.
Sachant que la personne est malade, il y a 99 chances sur 100, qu'elle soit détectée : ainsi 99 des 100 personnes malades sont détectées par le test, et la dernière ne l'est pas.

5. Admirez la paronymie !

Explications

Ceci illustre une grande difficulté pour détecter un événement rare avec un test ; même si le test est extrêmement fiable. C'est une erreur très néfaste, mais malheureusement très courante qui consiste à confondre deux types de questions :

- Quelle est la probabilité qu'un individu malade ne soit pas détecté ?
- Quelle est la probabilité qu'un individu détecté soit effectivement malade ?

Nous venons de voir que les réponses à ces deux questions sont très différentes.

D De l'importance de la modélisation

Mise en situation : Au début de l'année, j'ai reçu une lettre anonyme qui m'annonçait que l'action `molin-mathématiques` allait monter au cours de la semaine suivante.
La semaine suivante, par simple curiosité, je regarde le cours de bourse de l'action en question, et je vois qu'elle réalise en effet une belle progression.

Le vendredi midi, je reçois une autre lettre anonyme m'indiquant que l'action `prochain DS` baisserait la semaine prochaine.
Intrigué, je regarde la semaine suivante et je constate qu'en effet, cette action baisse.

Le vendredi, je reçois à nouveau une lettre et sa prédication s'avère aussi juste que les précédentes. Et ainsi de suite pendant 10 semaines !

À la fin de la 10^{e} semaine, je reçois une lettre identique, m'invitant à acheter des actions `truands et compagnie` qui m'apporteront à coup sûr une belle plus-value. Cette fois ci, elle est signée :

Le Courtier de Baltimore

Ce dernier me propose donc d'en faire l'achat moyennant une confortable commission à son égard.

Raisonnons un peu avant d'y mettre toutes nos économies.
Supposons que ce fameux courtier ne soit qu'un charlatan et qu'il ne connaisse rien à la bourse. Quelle est alors la probabilité qu'il réussisse, dix semaines de suite, à me fournir la bonne prédiction ?
On suppose donc que le courtier n'y connait rien et donne à chaque fois une action au hasard. Ainsi, chaque prédiction est indépendante des précédentes, et on peut supposer en première approximation qu'à chaque fois, il a une chance sur deux que sa prédiction s'avère juste.
La probabilité d'avoir raison 10 semaines de suite si c'est un charlatan est donc :

$$\mathbf{P}(B) = \mathbf{P}\left(B_1 \cap B_2 \cap \cdots \cap B_{10}\right) = \mathbf{P}(B_1)\mathbf{P}(B_2)\cdots\mathbf{P}(B_{10}) = \frac{1}{2} \times \frac{1}{2} \times \cdots \times \frac{1}{2}$$
$$= \frac{1}{2^{10}} = \frac{1}{1024}.$$

c'est-à-dire, moins de une chance sur 1000. Je peux donc être sûr de mon coup, et j'ai pense même avoir affaire à un excellent courtier puisqu'il ne s'est pas trompé une seule fois.

Le vice caché : En réalité, ce fameux courtier n'est qu'un truand. Comment a-t-il donc fait pour déjouer ainsi les probabilités ?
La première semaine, il a simplement envoyé une lettre à 1024 personnes en promettant à la moitié l'entre eux une hausse de l'action et à l'autre une baisse.
Il observe le cours de l'action, et à ceux auxquels la prédiction s'est révélée juste, il envoie une nouvelle prédiction. Il envoie donc 512 lettres où il annonce, pour moitié une hausse, et pour les autres, une baisse.
Il réitère ainsi sa méthode 10 semaines de suite, en n'envoyant la nouvelle lettre qu'à ceux auxquels les prédictions précédentes se sont révélées justes.
À la fin de la dixième semaine, il ne restait plus qu'une personne, c'était moi. Et il est désormais temps pour lui d'empocher le jackpot. Il n'y a aucun hasard là dedans, mais comme je ne vois qu'un aspect partiel du problème, je me fais duper.

Morale de cette histoire : il n'y en a malheureusement pas. Imaginons que des compagnies financières lancent couramment de nombreux fonds d'investissement qu'elles laissent évoluer pendant quelques années. La plupart d'entre eux, comme les lettres du courtier, s'avèrent très décevants, et sont oubliés. Mais parmi tous ces fonds, certains obtiennent de très bons résultats, et la compagnie peut donc éditer une jolie plaquette sur papier glacé faisant miroiter les taux mirobolants obtenus par ce fond lors des 5 dernières années... avec la mention qui s'avère plus pertinente que nous ne voulions le croire : « les performances passées ne préjugent pas des performances futures » . C'est simplement la dernière lettre du courtier !
Il en est de même, en recherche ou en TIPE expérimentaux : si on essaie de très nombreuses expériences avec des paramètres tous plus farfelus les uns que les autres, on risque à la fin, d'obtenir une expérience qui marche par un malheureux concours du hasard. Il suffit ensuite de cacher toutes les autres expériences, et de proposer votre dernière *lettre*, car

> Aussi improbable que soit un événement, il est hautement probable qu'il se réalise un jour, si on y met la persévérance suffisante.

De là, à faire décoller la fusée Shadok ? Ils y sont arrivés !
Ainsi, quand on donne une information, il faut également s'interroger sur toutes les informations que l'on ne nous donne pas... Les calculs les plus précis et convaincants, peuvent vite perdre de leur superbe.