Probabilité - Chapitre 1

Probabilités sur un ensemble fini

I Espaces probabilisés finis

1) Modélisation ensembliste

a) Expérience et univers

On appelle "expérience aléatoire" toute action dont le résultat ne peut être prévu avec certitude.

Exemples :

- lancer d'une pièce : à chaque lancer, même en jetant la pièce exactement de la même façon (a priori...) , le côté visible est différent.
- la plupart des expériences scientifiques sont en réalité aléatoires : aléas des appareils de mesure, des objets étudiés, etc.

On cherche à relier les expériences aléatoires, donc incertaines, aux mathématiques, a priori science "déterministe". Pour cela, on effectue ce que l'on appelle une *modélisation.*

Définition :

On appelle **univers**, noté Ω, l'ensemble des résultats possibles d'une expérience aléatoire.

Exemples :

- Pour "pile ou face", $\Omega =$
- Lancer d'un dé 6 faces : $\Omega =$
- Lancer de deux dés 6 faces :

Remarques :

- L'univers dépend de l'expérience étudiée, mais aussi de la façon dont on veut la décrire.
- Cette année, nous n'étudierons que le cas où l'univers est de cardinal fini.

b) Evénements aléatoires

Définition :

On appelle **événement aléatoire**, ou **événement** tout sous ensemble de résultats.
Autrement dit, dire qu'un ensemble A est un événement signifie que A est une partie de l'univers Ω, ou encore $A \subset \Omega$.

Exemples :
Prenons le lancer d'un unique dé 6 faces. On choisit $\Omega = \{1, 2, 3, 4, 5, 6\}$.

- l'événement "le lancer a donné un nombre pair" est modélisé par l'ensemble

- $\{4, 5, 6\}$ modélise l'événement

- $\{1, 2, 3, 4, 5, 6\}$ modélise l'événement

- $\emptyset$ modélise l'événement

Remarques :
- Ω et $\emptyset$ sont des parties de Ω : ils modélisent donc bien des événements...
- Si $card(\Omega) = n$, on a   événements possibles.
- On n'est pas obligé de prendre tout $\mathscr{P}(\Omega)$ comme ensemble des événements (car il peut vite être trop grand), mais en probabilité finie, ce sera très souvent le cas.

c) Opération sur les ensembles et traduction en termes probabilistes

Définition :

Soit A et B deux événements, c'est à dire deux sous ensembles de Ω.
- L'ensemble $A \cup B = \{\omega \in \Omega; \omega \in A \text{ ou } \omega \in B\}$ modélise l'événement "A s'est produit ou B s'est produit (ou les deux)"
- L'ensemble $A \cap B = \{\omega \in \Omega; \omega \in A \text{ et } \omega \in B\}$ modélise l'événement "A et B ont eu lieu (simultanément)."
- L'ensemble $A \setminus B = \{\omega \in \Omega; \omega \in A \text{ et } \omega \notin B\}$ modélise l'événement "A a eu lieu mais pas B."
- L'ensemble $\bar{A} = \{\omega \in \Omega; \omega \notin A\}$ modélise l'événement "A n'a pas eu lieu"

Exemple :
"Le lancer de dé est pair et supérieur ou égal à 4" se modélise par $\{2, 4, 6\} \cap \{4, 5, 6\} = \{4, 6\}$

Remarque :
- Si A et B sont tels que $A \cap B = \emptyset$, on dit qu'ils sont **incompatibles**. Par exemple "obtenir un lancer pair" et "obtenir un lancer impair" sont incompatibles.
- Il sera souvent très intéressant pour le calcul de décomposer un événement en union ou intersection d'événements... Cela sera un outil de justification essentiel.

d) Système complet d'événements :

Définition :

On appelle **système complet d'événements** toute famille d'événements A_1, A_2,...,A_n deux à deux incompatibles et tels que la réunion de tous donne l'univers.
Autrement dit :
1. $\displaystyle\bigcup_{i=1}^{n} A_i = \Omega$
2. si $i \neq j$, $A_i \cap A_j = \emptyset$.

Exemples :
1. Pour le lancer de dé, "obtenir un résultat pair" et "obtenir un résultat impair" forment un système complet d'événements.
2. Pour le lancer de deux dés, "obtenir deux dés différents" (ce qu'on avait noté Ω_1) et "obtenir un doublon" (ce qu'on a noté Ω_2) forment un système complet d'événements.

Remarque :
Si aucun des A_i est vide, on parle de **partition de** Ω.
En général, ce n'est pas un problème d'avoir des événements vides, donc on ne cherchera pas forcément à avoir une partition.

2) Probabilité

a) Définition

Définition :

Soit Ω un ensemble fini. On appelle **probabilité** sur Ω toute application P définie sur $\mathscr{P}(\Omega)$ à valeur dans $\mathbb{R}$ et vérifiant les propriétés suivantes :

1. $\forall A \in \mathscr{P}(\Omega)$, $P(A) \in [0,1]$
2. $P(\Omega) = 1$
3. si A et B sont deux événements incompatibles, alors $P(A \cup B) = P(A) + P(B)$

Remarques :

- Par une récurrence immédiate, on peut montrer à partir de la propriété 3 que si $A_1, A_2, \ldots, A_n$ sont n événements deux à deux incompatibles, alors

$$P(A_1 \cup A_2 \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) = \sum_{k=1}^{n} P(A_k)$$

- Le couple $(\underbrace{\Omega}_{\text{"l'univers"}}, \underbrace{P}_{\text{"la proba"}})$ est appelé "espace probabilisé".

  Modéliser une expérience aléatoire, c'est donner un espace probabilisé qui rend compte de cette expérience aléatoire.

b) Propriétés

Propriété 1 :

Soient A et B deux événements,

1. $P(\bar{A}) = 1 - P(A)$
2. $P(\emptyset) = 0$
3. Si $A \subset B$, alors
   - $P(A) \leq P(B)$ (on dit que P est croissante)
   - $P(B \setminus A) = P(B) - P(A)$.
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

▷ *Preuve* :

◁

? Le saviez-vous ? **ANDREÏ KOLMOGOROV**

Issu d'une famille de la noblesse russe, et élevé par ses tantes (sa mère meurt en couche) sous la protection de son grand père, Andreï Kolmogorov fait preuve très tôt de qualités mathématiques absolument hors normes, associées à une curiosité débordante sur tous les domaines.

Malgré les perturbations de la révolution soviétique en 1918 (il doit travailler, entre autres, à la construction de voies ferrées pour le "bien de la patrie"), Kolmogorov parvient à poursuivre ses études à Moscou et se forme aux idées mathématiques novatrices venue de France (théorie de l'intégrale de Lebesgue en particulier).

A partir de 1925, il s'engage dans une refondation des probabilités pour leur donner un ancrage plus moderne : malgré son jeune âge, ses résultats sont reconnus unanimement et c'est son axiomatique qu'on utilise aujourd'hui.

Kolmogorov et trois de ses étudiants en 1965

source : Oberwolfach Photo Collection

Ses recherches touchent de nombreux domaines, principalement en probabilités et statistiques (avec des applications déterminantes en cybernétique, dans l'armement, dans la finance, etc.) mais aussi en poésie et en linguistique.

Sa volonté de faire de l'URSS un pays à la pointe des sciences et des mathématiques en particulier le pousse à fonder en 1960 un internat/école à Moscou, réservé aux jeunes prodiges repérés partout sur le territoire, y compris dans les villages reculés. Il s'y investit énormément en tant que professeur, et est reconnu comme excellent pédagogue. Plusieurs de ses élèves sont aujourd'hui encore au coeur de la recherche.

Il meurt en 1987, entrant dans le panthéon non seulement des plus grands mathématiciens de l'histoire, mais aussi des plus grands professeurs.

c) Distribution de probabilités

Définition :

Soit E un ensemble fini. On appelle **distribution de probabilités** toute famille d'éléments de $\mathbb{R}^+$ indexée par E et de somme 1.

Exemples :

1. On suppose que une pièce truquée de sorte que la probabilité de donner Pile soit un certain $p \in]0,1[$. Notons $\Omega = \{P, F\}$.
   Une distribution de probabilités associée à Ω est $(x_P, x_F) = (p, 1-p)$
2. On lance un dé 6 face équilibrée. La distribution de probabilités associée à cette expérience va être

Propriété 2 :
Une probabilité sur un univers Ω est déterminée par la distribution de probabilités $(P(\{\omega\}))_{\omega\in\Omega}$

▷ *Preuve* :

◁

Remarque :
Une distribution de probabilités est en fait l'ensemble ordonné des images par P de Ω.

3) Un cas particulier de probabilité : la probabilité uniforme

a) Construction :

Définition :
Soit Ω un ensemble fini non vide. On appelle **probabilité uniforme** sur Ω l'application P définie par :

$$P : \begin{cases} \mathscr{P}(\Omega) & \to \mathbb{R} \\ A & \mapsto \dfrac{card(A)}{card(\Omega)} \end{cases}$$

Proposition 1 :
La probabilité uniforme est une probabilité.

▷ *Preuve* :

◁

b) Utilisation typique

On emploie la probabilité uniforme pour modéliser des expériences dont aucun résultat n'est privilégié. Dans le langage courant, c'est l'idée lorsqu'on utilise le terme "au hasard".

Un dé non pipé, une pièce non truquée, le tirage d'une carte dans un jeu de tarot sont autant d'expériences qui se modélisent bien par une probabilité uniforme.

Remarque :
On peut retenir la probabilité uniforme sous la forme suivante :

$$P(A) = \frac{\text{nbre de cas favorables}}{\text{nbre de cas possibles}}$$

Le terme "au hasard" ne veut pas toujours dire que la proba est uniforme. On en verra des exemples en TD, mais si on prend par exemple le lancer de deux dés équilibrés :

II Probabilités conditionnelles

1) Probabilité conditionnée par un événement

a) Introduction

On lance un dé 6 faces mais on cache le résultat avant qu'on ait eu le temps de le voir.

Le maître du jeu regarde le nombre obtenu et annonce que le résultat est pair. Quelle est la probabilité d'avoir un 6 ?
Notons A l'événement "avoir un 6".
Sans l'information du maître du jeu, on modélise l'expérience en posant

$$\Omega =$$

muni de la probabilité uniforme. On a donc $P(A) = \dfrac{1}{6}$.

Or, on sait que le résultat est pair. On change alors de Ω en posant

$$\tilde{\Omega} =$$

Alors $P(A) =$
Le problème de ce raisonnement, c'est que l'on modélise à nouveau l'expérience, nous obligeant à changer d'espace probabilisé en remplaçant Ω par l'événement B.

b) Définition

Définition :

Soit (Ω, P) un espace probabilisé. Soit B un événement tel que $P(B) \neq 0$.
On définit l'application P_B (lire "P sachant B")

$$P_B : \left\{ \begin{array}{rl} \mathscr{P}(\Omega) & \to \mathbb{R} \\ A & \mapsto \dfrac{P(A \cap B)}{P(B)} \end{array} \right.$$

.
P_B est une probabilité et est appelée **probabilité conditionnelle de P par B**.

▷ *Preuve* : Montrons que P_B est bien une probabilité :

◁

Exemple :
Dans l'introduction, on cherche en fait $P_B(A)$ avec comme espace probabilisé $\Omega = \{1, 2, 3, 4, 5, 6\}$ muni de la probabilité uniforme.
On a $P(B) = P(\{2, 4, 6\}) = 3/6 = \dfrac{1}{2}$ et $P(A \cap B) = P(\{6\}) = 1/6$, d'où $P_B(A) = \dfrac{1/6}{1/2} = \dfrac{1}{3}$.

c) Commentaires et interprétations

- L'espace probabilisé (Ω, P_B) revient à identifier B et Ω : en effet, $P_B(B) = \dfrac{P(B \cap B)}{P(B)} = 1$.
- On trouve aussi la notation :
$$P(A|B),$$
pour "probabilité de A sachant B". Cette notation date d'avant la formalisation rigoureuse des probabilités et doit être évitée. En effet, $P(A|B)$ signifierait que $(A|B)$ est un événement, ce qui n'est pas le cas.

Danger ! Confusion entre proba conditionnelle et intersection

Attention à ne pas confondre $P_B(A)$ et $P(A \cap B)$! Si on réfléchit trop vite, on peut prendre l'un pour l'autre....

▶ Lorsque l'on calcule "$P(A \cap B)$", on veut dire que les deux événements A et B sont aléatoires. On ne sait pas si A ou B a eu lieu, et on cherche les "chances" que les deux aient lieu simultanément.

▶ En revanche, quand on calcule $P_B(A)$, on suppose que B a eu lieu : B n'est plus aléatoire, c'est un événement certain (d'ailleurs $P_B(B) = 1$)....

Il s'agira donc de bien réfléchir sur le contexte pour ne pas se tromper de calcul : avez-vous supposé qu'un événement a eu lieu ?

2) Justification des arbres...

a) Formule des probas composées

Proposition 2 : Formule des probabilités composées

(i) Soit A et B deux événements avec $P(A) \neq 0$. Alors
$$P(A \cap B) = P(A)P_A(B)$$

(ii) Soit $n \geq 2$ et soient $A_1, A_2, \ldots, A_n$ n événements tels que
$$P(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \neq 0$$
Alors
$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P_{A_1}(A_2)P_{A_1 \cap A_2}(A_3) \times \ldots \times P_{A_1 \cap A_2 \cap \ldots \cap A_{n-1}}(A_n)$$

▷ *Preuve* :

◁

b) Exemple typique d'utilisation :

Un chapeau magique contient 3 lapins jaunes et 4 lapins bleus. On tire au hasard, successivement et sans remise 3 lapins de ce chapeau.
On s'intéresse à la probabilité de l'événement A="obtenir un lapin jaune, puis bleu, puis jaune".
L'intérêt de la formule des probabilités composées, c'est qu'on va pouvoir se passer de la modélisation complète et raisonner étape par étape, sans effectuer de dénombrement.

Méthode : TRAITER UN PROBLÈME AVEC SUCCESSION D'ÉVÉNEMENTS

Quand un problème traite de la répétition d'une action, il est essentiel de :

$\boxed{\textbf{toujours numéroter les événements.}}$

Dans cet exemple, les événements sont nommés J_1, J_2, etc... et pas seulement J, qui serait imprécis sur le numero de tirage.

c) Représentation par un arbre

Une façon de traduire les probabilités composées est la représentation par arbre pondéré. Chaque branche est pondérée de la probabilité conditionnelle d'être obtenue sachant les noeuds précédents :

Un chemin sur l'arbre correspond à une intersection d'événéments. La formule des probabilités en cascade justifie donc le fait que l'on calcule le produit des probas sur les branches pour avoir l'événement qui nous intéresse.

3) Utilisation des systèmes complets d'événements

a) Première utilisation : formule des probabilités totales

Exemple :
On dispose de deux chapeaux : un haut-de-forme et un chapeau melon. Le premier chapeau contient 3 lapins jaunes et 2 bleus, tandis que le second contient 1 seul jaune pour 3 bleus.
On lance une pièce équilibrée. Si pile tombe, on extrait un lapin du chapeau haut de forme. Sinon, on prend dans le chapeau melon.
Quelle est la probabilité d'avoir un lapin jaune ?

Au lieu de détailler comme fait ici, on pourra directement utiliser la formule dite des "probabilités totales".

Proposition 3 : Formule des probabilités totales

Soit $\{A_1, A_2, \ldots, A_n\}$ un système complet d'événements. Soit B un événement. Avec la convention que si $P(A_i) = 0$, alors $P(A_i)P_{A_i}(B) = 0$, on a :

$$P(B) = \sum_{i=1}^{n} P(A_i)P_{A_i}(B)$$

▷ *Preuve* :

◁

b) Deuxième utilisation : formule de Bayes

Exemple :
On reprend l'exemple précédent. L'expérience a eu lieu, et on a obtenu un lapin jaune. Quelle est la probabilité de l'avoir extrait du chapeau melon ?

Proposition 4 : formule de Bayes

1. **Version simple :** Soient A et B deux événements de probabilité non nulles, alors
$$P_B(A) = \frac{P(A)P_A(B)}{P(B)}$$
2. **Version avec système complet d'événements :**
Soit $\{A_1, A_2, \ldots, A_n\}$ un système complet d'événements, tous de probabilités non nulles. Soit B un événement tel que $P(B) \neq 0$. Alors pour tout $i \in [\![1, n]\!]$, on a :
$$P_B(A_i) = \frac{P(A_i)P_{A_i}(B)}{\sum_{k=1}^{n} P(A_k)P_{A_k}(B)}$$

▷ *Preuve* : On écrit la définition $P_B(A_i) = \dfrac{P(A_i \cap B)}{P(B)}$, puis on applique les probas composées au numerateur, les probabilités totales au dénominateur. ◁

Remarque :
Il n'est pas vraiment utile de retenir la formule par coeur : c'est la preuve qu'il faut connaître !

Le saviez-vous ? **Qui est Bayes ?**

Pasteur anglais du XVIII^e et mathématicien à ses heures perdues, Thomas Bayes est considéré comme le précurseur des probabilités conditionnelles. Il cherche à trouver les causes les plus probables à des phénomènes aléatoires observés.
La "formule de Bayes" est son résultat le plus célèbre. Il lui a été attribué à titre posthume par un de ses amis à qui il avait envoyé le manuscrit avant de mourir.

"maths Pantheon" par Paul Epps

Le résultat sera ensuite généralisé par Pierre Simon de Laplace. La formule de Bayes est fréquemment utilisée dans divers domaines, allant de la reconstitution de signal détérioré par des interférences jusqu'à l'intelligence artificielle et les réseaux neuronaux.

4) Indépendance

a) Définition :

Définition :

Soit (Ω, P) un espace probabilisé et soit A et B deux événements.
On dit que A et B sont **indépendants** si et seulement si $P(A \cap B) = P(A)P(B)$.

Remarques :
si A est de probabilité non nulle, A et B indépendants équivaut à $P_A(B) = P(B)$.

Exemple :
On lance un dé 6 faces non pipé, et on s'intéresse aux événements suivants :

- A : le résultat est paire
- B : le résultat est $\geq$ à 3
- C : le résultat est $\leq$ à 3

Propriété 3 :

(i) Pour tout $A \in \mathscr{P}(\Omega)$, A et $\emptyset$ sont indépendants, A et Ω sont indépendants.

(ii) Si A et B sont deux événements indépendants, A et $\bar{B}$ sont indépendants, $\bar{A}$ et B sont indépendants et $\bar{A}$ et $\bar{B}$ sont indépendants.

▷ *Preuve* :

◁

b) Généralisation :

Définition :

Soit (Ω, P) un espace probabilisé et soient n événements $A_1, A_2, \dots, A_n$.
On dit que ces événements sont **mutuellement indépendants** (ou indépendants) si et seulement si

$$\forall k \in [\![1,n]\!], \forall i_1, \dots, i_k \in [\![1,n]\!], P(A_{i_1} \cap A_{i_2} \cap \dots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$$

Autrement dit : n événements sont indépendants, si et seulement si la relation d'indépendance est vérifiées pour toutes les intersections possibles.
Par exemple, A_1, A_2 et A_3 indépendants signifie :

Remarques :

- En général, l'indépendance mutuelle est une hypothèse de l'énoncé. Il est rarissime qu'on demande à le démontrer.
- La propriété 3 (ii) peut-également être montrée pour n événements. Ainsi, si $A_1, A_2, \dots, A_n$ sont indépendants, alors il en est de même si on remplace des A_i par $\overline{A_i}$.

c) Indépendance 2 à 2

Définition :

Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace probabilisé et soient n événements $A_1, A_2, \dots, A_n$.
On dit que ces événements sont **indépendants** 2 **à** 2 si et seulement si $\forall i, k \in [\![1,n]\!]$ avec $i \neq k$, A_i est indépendant de A_k.

L'indépendance 2 à 2 est un hypothèse moins forte que l'indépendance mutuelle.

Exemple :
On lance deux dé 6 faces distinguables. On considère les événements :

— A : la somme des deux dés est 7.

— B : le premier dé donne 4.

— C : le deuxième dé donne 3.

Etudiez l'independance 2 à 2 des événements A, B et C et l'indépendance mutuelle.