Chapitre 5

Probabilités Conditionnelles

5.1 Introduction historique

À chaque instant, il n'y a qu'un infime espace entre le « trivial »et l'impossible. C'est dans cet espace que sont faites les découvertes mathématiques.

Andreï Kolmogorov - 1903-1987

Issus de la Renaissance, les premiers travaux mathématiques sur les probabilités s'attachaient à des problèmes pratiques, notamment au gain dans les jeux de hasard, et durent attendre un singulier personnage du XVIIIème siècle pour connaître un essor remarquable, en voyant apparaitre son premier théorème fondamental.

5.1.1 Inférence Bayesienne

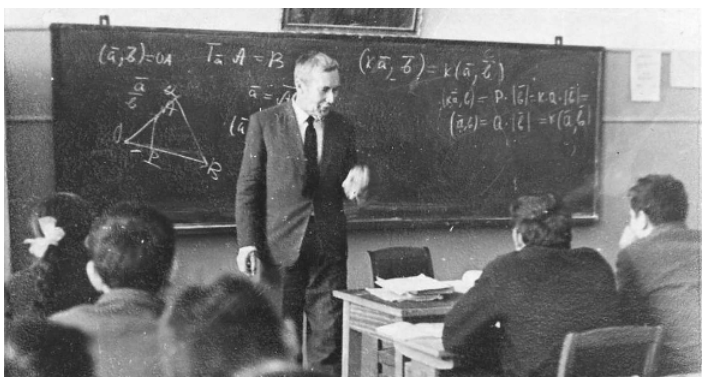
Né vers 1702, **Thomas Bayes** est un pasteur presbytérien britannique qui dévoua sa vie à l'étude de la théologie et de son approche extrêmement rationnelle, au point d'intéresser aux dernières trouvailles mathématiques de son temps, comme celle de **Isaac Newton**. À la fin de sa vie, il se pencha notamment sur une preuve probabiliste de l'existence de Dieu, ou comment prouver qu'avec les innombrables coïncidences qu'on peut observer dans le monde qui nous entoure, l'existence d'un Dieu omnipotent devient nécessaire et très probablement vraie. Son ami **Richard Price** (1723-1791) publie après sa mort un ouvrage qui résume ses idées, et notamment une fameuse formule qui portera désormais le nom de Théorème de Bayes :

$$P_A(B) = \frac{P(B) \times P_B(A)}{P(A)}$$

Peu concerné par les conséquences mathématiques d'une telle trouvaille, Price insiste principalement dans ce livre sur les conséquences théologiques d'une telle formule, et ne développera pas lui-même le formalisme qu'on utilise aujourd'hui.

5.1.2 Kolmogorov et les fondements

Comment rendre parfaitement rigoureux et mathématique une science qui certes compte, mais s'attache essentiellement au monde réel, si loin de l'abstrait mathématique ? L'illustre mathématicien russe **Andreï Nikolaïevitch Kolmogorov** (1903-1987) s'y appliquera au travers de ses fameux axiomes, qui définissent tels qu'on les utilise aujourd'hui les fondements des probabilités modernes. Basé sur des propriétés à remplir pour qu'un ensemble d'objets mathématiques soient « mesurables », il introduit également les notations définitives des probabilités, et la vision moderne qu'on a de la formule de Bayes.



5.2 Définitions

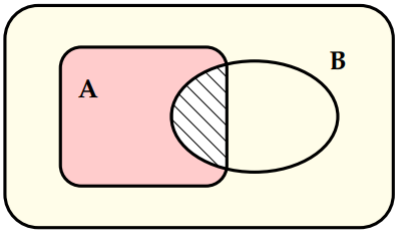
Définition : Soient A et B deux évènements quelconques d'un univers Ω.

• On appelle alors **probabilité conditionnelle** la probabilité que B soit réalisé sachant que l'évènement A est réalisé. Cette probabilité est notée $P_A(B)$, et on pose :

$$P_A(B) = \frac{P(A \cap B)}{P(A)}$$

Remarque :

On peut comprendre intuitivement cette définition à partir du diagramme de Venn ci-contre.



La probabilité de B sachant A correspond à la **proportion** de B dans A, c'est à dire la part hachurée dans l'ensemble A. Ainsi :

$$P_A(B) = \frac{\text{Nombres d'éléments communs à } A \text{ et } B}{\text{Nombre d'éléments de } A} = \frac{P(A \cap B)}{P(A)}$$

ex :

On donne ci-contre la répartition des spectateurs sur une journée dans une salle de cinéma selon les séances et le tarif. On choisit un des spectateurs au hasard et on considère les événements :

- *M : « La personne a assisté à la séance du matin. »*
- *D : « La personne a payé demi-tarif. »*

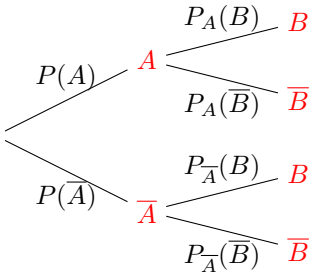
	Plein Tarif	*Demi-tarif*	*Total*
Matin	103	91	194
Soir	280	26	306
Total	383	117	500

La probabilité que la personne ait assisté à la séance du matin sachant qu'elle a payé le demi-tarif est $P_D(M) = \frac{91}{117}$ *car parmi les* 117 *personnes qui sont ont payé le demi-tarif,* 91 *sont venues le matin.*

De même, la probabilité qu'une personne ait payé le demi-tarif sachant qu'elle a assisté à la séance du matin est $P_M(D) = \frac{91}{194}$.

5.3 Arbre pondéré

Soient deux événements A et B. On peut représenter par un arbre pondéré les probabilités suivantes lorsque l'on connaît les probabilités de B ou $\overline{B}$ lorsque A est réalisé.



Pour remplir et utiliser un arbre, on doit respecter les règles suivantes :

- Sur chaque branche de l'arbre, on écrit les probabilités correspondantes (de préférence sous forme fractionnelle).
- La somme des probabilités inscrites sur les branches issues d'un même nœud est égale à 1 (loi des nœuds).
- Le produit des probabilités inscrites sur chaque branche d'un chemin donne la probabilité de l'intersection des événements placés sur ce chemin. Par exemple :

$$P(A) \times P_A(B) = P(A \cap B)$$

Proposition 1 : Probabilités totales

Soit A_1, A_2, …, A_n une partition de l'univers Ω (ensembles deux à deux incompatibles et dont l'union forme Ω), alors, pour tout événement B, on a :

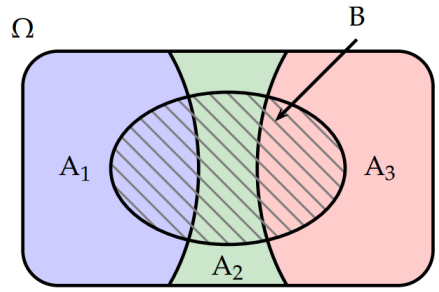
$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

∘ Démonstration : Les évènements A_1, A_2,…, A_n sont mutuellement incompatibles, donc $A_1 \cap B$, $A_2 \cap B$, …, $A_n \cap B$ sont également mutuellement incompatibles. Ainsi :

$$\begin{aligned} P(B) &= P((A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)) \\ &= P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B) \end{aligned}$$

Par exemple dans le cas d'une partition en trois événements :

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

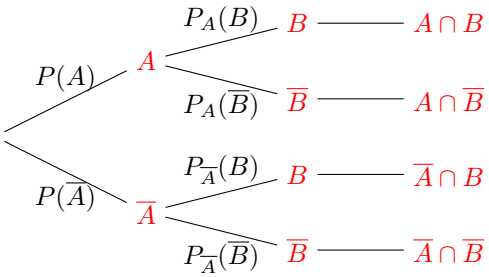


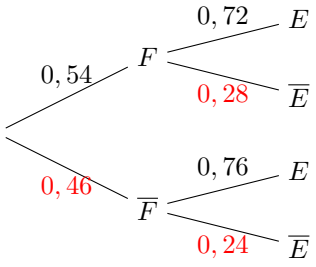
□

Conséquence directe de cette proposition, dans le cas de l'arbre considéré précédemment :

- La probabilité d'un événement est la somme des probabilités des chemins qui aboutissent à cet événement.

$$\begin{aligned} P(B) &= P(A \cap B) + P(\overline{A} \cap B) \\ &= P(A) \times P_A(B) + P(\overline{A}) \times P_{\overline{A}}(B) \end{aligned}$$





ex :

Dans un lycée 54 % des élèves sont des filles dont 72 % sont externes. De plus, 76 % des garçons sont externes. On choisit un élève au hasard. On pose :

- *F : « l'élève choisi est une fille »*
- *E : « l'élève choisi est externe »*

On traduit alors :

$$P(F) = 0,54 \qquad P_F(E) = 0,72 \qquad P_{\overline{F}}(E) = 0,76$$

Ainsi, si on cherche la probabilité d'obtenir un élève externe, on effectue le calcul :

$$\begin{aligned} P(E) &= P(F \cap E) + P(\overline{F} \cap E) \\ &= P(F) \times P_F(E) + P(\overline{F}) \times P_{\overline{F}}(E) \\ &= 0,54 \times 0,72 + 0,46 \times 0,76 \\ &= 0,7384 \end{aligned}$$

5.4 Indépendance

Définition : Soient A et B deux évènements quelconques d'un univers Ω.

- On dit que A et B sont **indépendants** si et seulement si :

$$P(A \cap B) = P(A) \times P(B) \qquad \text{ou lorsque } P(A) \neq 0 \text{ et } P_A(B) = P(B)$$

Autrement dit, le fait que A soit réalisé n'a aucune influence sur la probabilité que B se réalise, et inversement.

ex :

On donne la répartition des licenciés dans un club. On tire au sort une personne au hasard pour une tombola et on note les événements :

- *A : « La personne est adulte. »*
- *B : « La personne pratique le basket-ball. »*

On constate alors que :

$$P(A) = \frac{132}{528} = 0,25 \qquad et \qquad P_B(A) = \frac{45}{180} = 0,25$$

	Adulte	*Enfant*	*Total*
Handball	73	174	247
Basket-ball	45	135	180
Gymnastique	14	87	101
Total	132	396	528

Ainsi $P(A) = P_B(A)$, or $P(B) \neq 0$, donc les événements A et B sont indépendants.

Proposition 2 :

Soit A et B deux événements d'un univers Ω.

- Alors si A et B sont indépendants, $\overline{A}$ et B sont également indépendants.

◦ Démonstration : On applique la formule des probabilités totales :

$$P(B) = P(A \cap B) + P(\overline{A} \cap B)$$
$$\text{donc : } P(\overline{A} \cap B) = P(B) - P(A \cap B)$$

Or A et B sont indépendants donc $P(A \cap B) = P(A) \times P(B)$, ainsi :

$$\begin{aligned} P(\overline{A} \cap B) &= P(B) - P(A) \times P(B) \\ &= (1 - P(A))P(B) \\ &= P(\overline{A}) \times P(B) \end{aligned}$$

Donc par définition $\overline{A}$ et B sont indépendants.

□

5.5 Variable aléatoire

Définition :

Soit une expérience aléatoire d'univers Ω, composé de $n \in \mathbb{N}^*$ issues.
On appelle **variable aléatoire** l'objet mathématique qui à tout issue de Ω fait correspondre un nombre réel x_i.
On appelle alors **loi de probabilité** la fonction P qui à toute valeur x_i fait correspondre un nombre $p_i \in [0;1]$, tel que :

$$p_1 + p_2 + \cdots + p_n = 1$$

On notera alors :

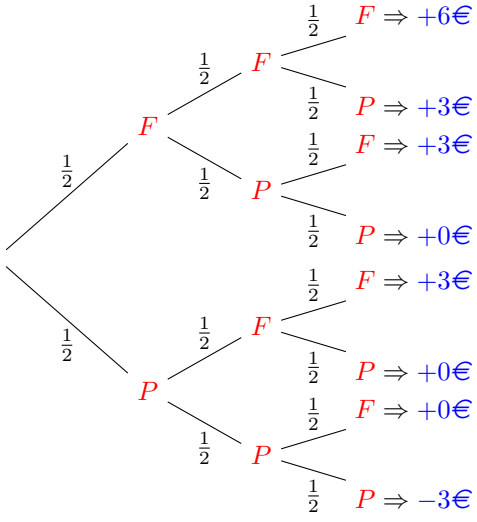
$$\forall i \in [\![1;n]\!],\ P(X = x_i) = p_i$$

ex : Considérons une expérience aléatoire où on lance successivement trois fois une pièce équilibrée afin d'obtenir Pile ou Face. Cette expérience va offrir 8 issues différentes, et on fixe alors que si on obtient Face, le joueur gagne 2€, et si il obtient Pile il perd 1€.

On note alors :

- *P : « Obtenir Pile »*
- *F : « Obtenir Face »*

Et on peut représenter cette expérience par l'arbre suivant :



On constate alors 4 évènements différents, on peut définir la variable aléatoire X qui à chacun de ces évènements fait correspondre la probabilité de l'obtenir, ainsi :

$$P(X = 6) = \frac{1}{8} \quad \textit{une issue}$$
$$P(X = 3) = \frac{3}{8} \quad \textit{trois issues}$$
$$P(X = 0) = \frac{3}{8} \quad \textit{trois issues}$$
$$P(X = -3) = \frac{1}{8} \quad \textit{une issue}$$

On peut donner plus rapidement la loi de probabilité par un tableau :

x_i	-3	0	3	6
$P(X = x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

5.6 Espérance

Définition :
Soit une expérience aléatoire d'univers Ω ayant n issues, et X une variable aléatoire sur Ω. On appelle alors **espérance mathématique de X**, noté $E[X]$ la valeur de :

$$E[X] = \sum_{k=1}^{n} p_i x_i = p_1 x_1 + p_2 x_2 + \cdots p_n x_n$$

L'espérance de X correspond à la moyenne des valeurs prises par X, pondérées par les probabilités correspondantes. C'est la valeur moyenne qu'on peut espérer obtenir en répétant indéfiniment l'expérience aléatoire.

ex : Si on reprend le tableau de l'exemple précédent, calculer l'espérance revient à multiplier entre elle les valeurs d'une colonne puis additionner le tout :

$$\begin{aligned} E[X] &= P(X=-3)\times(-3) + P(X=0)\times 0 + P(X=3)\times 3 + P(X=6)\times 6 \\ &= \frac{1}{8}\times(-3) + \frac{3}{8}\times 3 + \frac{1}{8}\times 6 = \frac{-3+9+6}{8} = \frac{12}{8} = 1,5 \end{aligned}$$

Ceci signifie qu'en répétant l'expérience indéfiniment, le joueur peut espérer gagner en moyenne $1,5$€ par expérience.

Proposition 3 : Linéarité de l'espérance
Soit X une variable aléatoire sur un univers Ω, et soient a et b deux nombres réels. On note $aX+b$ la variable aléatoire qui à chaque issue associe le nombre $ax_i + b$. On a alors :

$$E[aX+b] = a\,E[X] + b$$

∘ Démonstration : La variable aléatoire $aX+b$ prend les valeurs ax_1+b, ax_2+b, …, ax_n+b, avec exactement les mêmes probabilités p_1, p_2, …, p_n que X, puisque les issues n'ont pas changé. Ainsi :

$$\begin{aligned} E[aX+b] &= p_1(ax_1+b) + p_2(ax_2+b) + \cdots + p_n(ax_n+b) \\ &= a\,p_1x_1 + p_1b + a\,p_2x_2 + p_2b + \cdots + a\,p_nx_n + p_nb \\ &= a\underbrace{(p_1x_1 + p_2x_2 + \cdots + p_nx_n)}_{=E[X]} + b\underbrace{(p_1 + p_2 + \cdots + p_n)}_{=1} \\ &= a\,E[X] + b \end{aligned}$$

la dernière somme valant 1 puisqu'il s'agit de la somme de toutes les probabilités de la loi.

□

*rq : Cette propriété est très commode : elle évite de recalculer entièrement une loi de probabilité lorsque les gains sont modifiés de façon affine. Attention toutefois, elle ne vaut que pour les transformations de la forme $aX+b$: en général $E[X^2]$ n'est **pas** égal à $(E[X])^2$, comme nous le verrons plus bas.*

ex : Reprenons le jeu précédent, dont l'espérance vaut $E[X] = 1,5$. Supposons que l'organisateur double tous les gains, puis prélève un droit d'entrée de 2€. Le nouveau gain est alors la variable aléatoire $Y = 2X-2$, et il est inutile de reconstruire son tableau de loi :

$$E[Y] = 2\times E[X] - 2 = 2\times 1,5 - 2 = 1$$

Le joueur peut donc espérer gagner en moyenne 1€ par partie avec cette nouvelle règle.

Définition :
Soit une expérience aléatoire d'univers Ω ayant n issues, et X une variable aléatoire sur Ω.
On appelle alors **variance mathématique de X**, noté $V(X)$ la valeur de :

$$V[X] = \sum_{k=1}^{n} p_i(x_i - E[X])^2 = p_1(x_1 - E[X])^2 + p_2(x_2 - E[X])^2 + \cdots p_n(x_n - E[X])^2$$

La variance de X correspond à la moyenne de l'écart entre les valeurs de X et l'espérance, mis au carré et pondérés par les probabilités.
On en déduit ce qu'on appelle l'**écart-type** de X, noté $\sigma(X)$ et qui vaut :

$$\sigma(X) = \sqrt{V(X)}$$

La variance ne représentant pas grand chose par la présence des carrés (nécessaire pour rendre les écarts positifs), on applique une racine afin d'avoir une sorte de moyenne des écarts entre les valeurs et l'espérance de X.

ex : Reprenons une dernière fois l'exemple initial pour calculer la variance :

$$\begin{aligned} V(X) &= P(X=-3)\times(-3-1,5)^2 + P(X=0)\times(0-1,5)^2 + P(X=3)\times(3-1,5)^2 \\ &\quad + P(X=6)\times(6-1,5)^2 \\ &= \frac{1}{8}\times(-4,5)^2 + \frac{3}{8}\times(-1,5)^2 + \frac{3}{8}\times 1,5^2 + \frac{1}{8}\times 4,5^2 \\ &= \frac{20,25+6,75+6,75+20,25}{8} \\ &= \frac{54}{8} = 6,75 \end{aligned}$$

On en déduit l'écart-type :

$$\begin{aligned} \sigma(X) &= \sqrt{V(X)} \\ &= \sqrt{6,75} \\ &\approx 2,6 \end{aligned}$$

L'écart moyen entre les gains et le gain moyen sera donc de $2,6$€.

Proposition 4 : Formule de König-Huygens
Soit X une variable aléatoire sur un univers Ω. On note $E[X^2]$ l'espérance de la variable aléatoire X^2, c'est-à-dire le nombre $p_1x_1^2 + p_2x_2^2 + \cdots + p_nx_n^2$. On a alors :

$$V(X) = E[X^2] - \big(E[X]\big)^2$$

∘ Démonstration : Notons $m = E[X]$ pour alléger l'écriture, et développons le carré présent dans la définition de la variance :

$$\begin{aligned} V(X) &= p_1(x_1-m)^2 + p_2(x_2-m)^2 + \cdots + p_n(x_n-m)^2 \\ &= p_1(x_1^2 - 2mx_1 + m^2) + \cdots + p_n(x_n^2 - 2mx_n + m^2) \\ &= \underbrace{(p_1x_1^2 + \cdots + p_nx_n^2)}_{=E[X^2]} - 2m\underbrace{(p_1x_1 + \cdots + p_nx_n)}_{=E[X]=m} + m^2\underbrace{(p_1 + \cdots + p_n)}_{=1} \\ &= E[X^2] - 2m^2 + m^2 \\ &= E[X^2] - m^2 \end{aligned}$$

□

*rq : Cette formule est presque toujours plus rapide que la définition, car elle évite de soustraire l'espérance à chacune des valeurs : il suffit d'ajouter une ligne au tableau de la loi de probabilité, celle des x_i^2. C'est donc elle que l'on utilisera en pratique. Elle montre au passage que $E[X^2]$ et $(E[X])^2$ ne sont **pas** égaux : leur différence est précisément la variance, qui n'est nulle que si X prend une seule valeur.*

ex : Vérifions la formule sur l'exemple précédent. Calculons d'abord l'espérance des carrés :

$$\begin{aligned} E[X^2] &= \frac{1}{8} \times (-3)^2 + \frac{3}{8} \times 0^2 + \frac{3}{8} \times 3^2 + \frac{1}{8} \times 6^2 \\ &= \frac{9 + 0 + 27 + 36}{8} = \frac{72}{8} = 9 \end{aligned}$$

On en déduit la variance, sans avoir eu besoin des écarts à la moyenne :

$$V(X) = E[X^2] - \big(E[X]\big)^2 = 9 - 1,5^2 = 9 - 2,25 = 6,75$$

On retrouve bien la valeur obtenue plus haut, pour un calcul nettement plus court.