Chapitre 3

Probabilité et espérance conditionnelles

3.1 Probabilité conditionnelle

Rappelons que nous comprenons la probabilité d'un évenement $A \subseteq \Omega$ comme une mesure de l'espérance que nous avons que c'est l'évènement A qui se réalisera et non son contraire A^c, c'est-à-dire une mesure de l'espérance que nous avons que l'état du monde ω^* où l'on se trouve est tel que $\omega^* \in A$. Imaginons à présent que nous souhaitons déterminer comment réévaluer cette espérance si nous considérons comme acquis que $\omega^* \in B$ (soit qu'on a une information qui nous assure de cela, soit qu'on souhaite simplement traiter séparément les deux cas $\omega^* \in B$ et $\omega^* \notin B$). On appelle cette nouvelle probabilité la *probabilité* de *A sachant B*, et on la note $\mathbb{P}_B(A)$ (ou $\mathbb{P}(A|B)$). Soulignons que $\mathbb{P}_B$ est bien une probabilité sur (tout) Ω, simplement veut-on au-moins que $\mathbb{P}_B(A) = 0$ si $A \subseteq B^c$ (puisque si $\omega^* \in B$, on est certain que $\omega^* \notin A$) ; on veut aussi que $\mathbb{P}_B(B) = 1$ (puisqu'on envisage ici que le cas où $\omega^* \in B$). L'idée est alors de poser $\mathbb{P}_B(A) = c\mathbb{P}(A \cap B)$ ce qui assurera facilement que $\mathbb{P}_B$ est une probabilité, et comme $\mathbb{P}_B(B) = 1$, nous voyons que $c = 1/\mathbb{P}(B)$; bien-entendu ceci impose que $\mathbb{P}(B) \neq 0$. Le fait qu'un évènement B soit de probabilité nulle ou non étant souvent capitale, on dit qu'un évènement est *négligeable* si et seulement si $\mathbb{P}(B) = 0$.

D'où finalement la définition

Définition : Soit B un évenement non-négligeable de $(\Omega, \mathcal{B})$. On appelle *probabilité conditionnelle sachant B* la fonction $\mathbb{P}_B : \mathcal{B} \longrightarrow [0,1]$ définie, pour tout évènement $A \in \mathcal{B}$, par $\boxed{\mathbb{P}_B(A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$.

Proposition 3.1 *Si $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{B})$, et si $B \in \mathcal{B}$ n'est pas négligeable, alors $\mathbb{P}_B$ est également une probabilité sur $(\Omega, \mathcal{B})$.*

Exercice : Montrer la proposition 3.1 et vérifier que $\mathbb{P}_B(A) = \mathbb{E}(\mathbb{I}_A \mathbb{I}_B)/\mathbb{E}(\mathbb{I}_B)$.

On dit qu'une famille d'évènements $(A_i)_{i=1..n}$ forme une système complet d'évènements (s.c.é) si et seulement si [1] $\Omega = \dot{\bigcup}_{i=1..n} A_i$, c'est-à-dire que chaque $\omega \in \Omega$ appartient à un A_i et un seul. Dans ce cas, pour tout évènement $B \in \mathcal{B}$ on a $B = \dot{\bigcup}_{i=1..n} (B \cap A_i)$ et donc $\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B \cap A_i) = \sum_{i=1}^n \frac{\mathbb{P}(B \cap A_i)}{\mathbb{P}(A_i)} \mathbb{P}(A_i)$, d'où la *formule de la probabilité totale*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i). \tag{3.1}$$

Imaginons connues les probabilité $\mathbb{P}(B|A_i)$ et $\mathbb{P}(A_i)$ pour tout $i = 1..n$. La *formule de Bayes*[2] suivante permet d'en déduire les $\mathbb{P}(A_i|B)$.

[1]Certains auteurs se contentent de demander que les A_i soient deux-à-deux disjoints et $\mathbb{P}\left(\bigcup_{i=1..n} A_i\right) = 1$; voyez-vous la différence ? Que peut-on dire de $N := \left(\bigcup_{i=1..n} A_i\right)^c$?

[2]Thomas Bayes (1702-1761), publication (postume) de *An Essay Toward Solving a Problem of Chances*, en 1754.

Théorème 3.2 (formule de Bayes) *Soit $(A_i)_{i=1..n}$ un s.c.é. de $(\Omega, \mathcal{B})$. Alors, pour tout évènement B non-négligeable, et tout $j = 1..n$, on a*

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\mathbb{P}(B)} \left(= \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)} \right).$$

Preuve : Elle se réduit à l'application de la définition de $\mathbb{P}(A_j|B)$ et de $\mathbb{P}(B|A_j)$:

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(A_j \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \cap A_j)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\mathbb{P}(B)},$$

et on termine par la formule (3.1). □

Exercice : Deux usines se partagent la production mondiale de vistemboirs[3]. Un vistemboir choisi au hasard[4] a la probabilité 0.5 de provenir de l'une comme de l'autre usine. Les vistemboirs de la première usine sont défectueux avec une probabilité de 2% et ceux de la seconde usine sont défectueux avec une probabilité de 6%. On choisit un vistemboir au hasard : qu'elle est la probabilité qu'il provienne de la première usine (Indication : poser $A_i :=$"le vistemboir est produit dans l'usine i", et $B :=$"le vistemboir est défectueux". Cet exemple explique pourquoi cette formule est parfois appelée "formule des probabilités des causes".)

3.2 Indépendance

Cette définition assez anodine de probabilité conditionnelle nous conduit à une notion capitale en probabilité : celle d'indépendance. Au chapitre 1 nous avons vu ce qu'est une v.a. Y qui est X-mesurable : c'est une simplement fonction déterministe de X dans le sens où $Y = g(X)$ pour une fonction déterministe $x \mapsto g(x)$; donc $Y(\omega)$ est entièrement connu dès que $X(\omega)$ est connu. A l'inverse, nous souhaitons une notion d'indépendance telle la connaissance de $X(\omega)$ ne nous apprenne rien sur $Y(\omega)$. En fait nous allons définir ce que sont deux évènements A et B indépendants et demanderons que les évènements $\{X \le x\}$ et $\{Y \le y\}$ soient (tous) indépendants. L'idée que deux évènements A et B sont indépendants est que le fait de supposer, par exemple, que B a lieu ne change pas la probabilité de A, c'est-à-dire que $\mathbb{P}_B(A) = \mathbb{P}(A)$, ce qu'il n'a de sens que si $\mathbb{P}(B) \neq 0$. On veux aussi s'affranchir de l'apparente disymétrie et que $\mathbb{P}_A(B) = \mathbb{P}(B)$ (sous reserve cette fois que $\mathbb{P}(A) \neq 0$). Explicitons ces deux identités ; nous obtenons

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A) \text{ et } \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B).$$

En "chassant les dénominateurs", nous voyons que les deux relations se réduisent à l'unique relation $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ qui a en outre l'avantage d'être définie même si $\mathbb{P}(A) = 0$ ou $\mathbb{P}(B) = 0$. Ceci nous conduit donc tout naturellement à la définition

Définition : Soit $\mathbb{P}$ une probabilité sur $(\Omega, \mathcal{B})$, et soient $A \in \mathcal{B}$ et $B \in \mathcal{B}$ deux évènements. On dit que les évènement A et B sont *indépendants pour la probabilité* $\mathbb{P}$ si et seulement si $\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)}$.

Soient X et Y deux v.a. sur $(\Omega, \mathcal{B})$. On dit que les v.a. X et Y sont *indépendantes pour la probabilité* $\mathbb{P}$ si et seulement si pour tous x et y les évènements $\{X \le x\}$ et $\{Y \le y\}$ sont indépendants.

Notations : Il est commode d'écrire $A \perp\!\!\!\perp B$ pour noter que les évènements A et B sont indépendants, et d'écrire $X \perp\!\!\!\perp Y$ pour noter que les v.a. X et Y sont indépendantes.

Exemple : Pièces de monnaie indépendantes : Considérons l'ensemble des états du monde Ω suivant :

$$\Omega = \{0,1\}^2 = \{0,1\} \times \{0,1\} = \{\omega = (\omega_1, \omega_2), \omega_i = 0..1, i = 1..2\}.$$

Posons $X_1(\omega_1, \omega_2) = \omega_1$ et $X_2(\omega_1, \omega_2) = \omega_2$. Considérons deux pièces de monnaie, et modélisons le fait que la première tombe sur Pile par l'évènement $\{X_1 = 1\}$, l'évènement $\{X_2 = 1\}$ modélisant quant à lui que la seconde pièce tombe sur Pile. Soient $p := \mathbb{P}(\{X_1 = 1\})$ et $q := \mathbb{P}(\{X_2 = 1\})$. Supposons les deux pièces indépendantes. Notons $A := \{X_1 = 1\}$ et $B := \{X_2 = 1\}$; l'indépendance des deux pièces implique celle de A et B, donc $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = pq$, et donc $\mathbb{P}(\{(X_1, X_2) = (1,1)\}) = pq$. En procédant de même pour les autres combinaisons de valeurs de X_1 et X_2, nous voyons que les deux pièces sont indépendantes si et seulement si on a le tableau des probabilités suivant :

[3] "Vistemboir" : nouvelle de Jacques Perret, La Machin, NRF, Gallimard, 1953.

[4] de "azzahr", le jeu de dés, en arabe ; ne me demandez pas comment on fait pour "choisir au hasard"...

$\downarrow X_1 = *, X_2 = * \longrightarrow$	1	0	$\mathbb{P}(\{X_1 = *\}) \downarrow$
1	pq	$p(1-q)$	p
0	$(1-p)q$	$(1-p)(1-q)$	$1-p$
$\mathbb{P}(\{X_2 = *\} \longrightarrow$	q	$1-q$	

Exercice : Pièces indiscernables : On reprend les notation de l'exemple précèdent. Montrer que $X_1 \perp\!\!\!\perp X_2 \quad \Rightarrow \quad \mathbb{P}(X_1 \neq X_2) = p + q - 2pq$. On dit que les deux pièces sont *indiscernables* si $\mathbb{P}(X_1 \neq X_2) = \mathbb{P}(X_1 = X_2 = 1) = \mathbb{P}(X_1 = X_2 = 0) = \frac{1}{3}$; montrer que deux pièces indiscernables ne peuvent être indépendantes.

Définition : Soit $(A_i)_{i\in I}$ une famille d'évènements (l'ensemble I peut être infini). On dit que ces évènements sont indépendants si et seulement si $I_0\mathbb{P}(\bigcap_{i\in I_0} A_i) = \Pi_{i\in I_0}\mathbb{P}(A_i)$ pour tout sous-ensemble *fini*. Il est commode de dénoter par $\coprod_{i\in I} A_i$ la propriété que ces évènements sont indépendants.

On dit que les v.a. d'une famille $(X_i)_{i\in I}$ sont indépendantes si et seulement si pour tous $x_i \in \mathcal{X}_i := X_i(\Omega)$ les évènements $(\{X_i \leq x_i\})_{i\in I}$ sont indépendants. Il est commode de dénoter par $\coprod_{i\in I} X_i$ la propriété que ces v.a. sont indépendantes.

Exercice : Evènements (mutuellement) indépendants : On reprend les v.a. indépendantes de l'exemple, et on choisit $p = \frac{1}{2} = q$; on pose $X = X_1$, $Y = X_2$, et $Z = |X - Y|$. Quelles sont les valeurs possibles pour la v.a. Z ? Les v.a. X et Y sont indépendantes par construction ; montrer que les v.a. X et Z sont indépendantes, et que les v.a. Y et Z sont indépendantes. On pose $A := \{X \leq 0\}$, $B := \{Y \leq 0\}$, et $C := \{Z \leq 0\}$. Calculer $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(C)$, et $\mathbb{P}(A \cap B \cap C)$. Montrer que les v.a. X, Y, et Z *ne* sont *pas* indépendantes. Cet exemple montre que trois v.a. peuvent être indépendantes deux-à-deux sans être indépendantes. C'est pourquoi, pour des v.a. qui sont indépendantes au sens de la définition que nous avons données, on dit parfois qu'elles sont *mutuellement* indépendantes.

Proposition 3.3 *Si les v.a. $(X_i)_{i\in I}$ ne prennent chacune qu'un ensemble fini ou dénombrable de valeurs $\mathcal{X}_i := X_i(\Omega) = \{x^i_j, j = 1, 2, \ldots\}$, les v.a. $(X_i)_{i\in I}$ sont indépendantes si et seulement si pour tous $x^i \in \mathcal{X}_i$ les évènements $(\{X_i = x^i\})_{i\in I}$ sont indépendants.*

Soient $(X_i)_{i\in I}$ une famille de v.a. et $(g_i)_{i\in I}$ une famille de fonctions $g_i : X_i(\Omega) \longrightarrow \mathbb{R}$. Posons $Y_i := g_i(X_i)$. Si les v.a. $(X_i)_{i\in I}$ sont indépendantes, alors les v.a $(Y_i)_{i\in I}$ sont également indépendantes.

Théorème 3.4 *Si les v.a. X et Y sont indépendantes, on a $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*

Preuve : Voir cours ; c'est facile. □

Remarque : nous avions déjà vu que $\mathbb{E}$ est linéaire, et qu'on a donc $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$. Nous voyons ici que si les v.a. X et Y sont indépendantes, on a de plus $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(X)$. Ceci est évidemment très commode pour les calculs d'espérance, mais soulignons bien que si la linéarité est toujours assurée, la relation $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(X)$ est généralement fausse lorsque l'hypothèse d'indépendance n'est pas assurée.

A noter que la réciproque "$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(X) \Rightarrow X \perp\!\!\!\perp Y$" est fausse également : soit X une v.a. (élémentaire) symétrique, c'est-à-dire $\mathbb{P}(\{X \geq +x\} = \mathbb{P}(\{X \leq -x\}))$ pour tout x, alors $Z := X^3$ est également symétrique, et donc $\mathbb{E}(X) = 0 = \mathbb{E}(Z)$. Posons à présent $Y := X^2$; Y n'est bien-entendu pas indépendant de X (sauf si $X = 0$), et pourtant $\mathbb{E}(XY) = \mathbb{E}(X \cdot X^2) = \mathbb{E}(Z) = 0 = 0 \cdot \mathbb{E}(X^2) = \mathbb{E}(X) \cdot \mathbb{E}(X^2) = \mathbb{E}(Y)$.

Exercice : Démontrer ce qui vient d'être affirmé ci-dessus : donner un exemple de v.a. élémentaire symétrique X (choisir $X(\Omega) = \{-1, +1\}$. Montrer que $\mathbb{P}(X = +1) = \mathbb{P}(X = -1) = \frac{1}{2}$; en déduire que $\mathbb{E}(X) = 0 = \mathbb{E}(X^3)$. Montrer que X et $Y := X^2$ ne sont pas indépendantes.

3.3 Espérance conditionnelle

Soit tout d'abord Y une v.a. sur $(\Omega, \mathcal{B})$, et $A \in \mathcal{B}$ un évènement non-négligeable. Nous pouvons donc considérer la probabilité conditionnelle $\mathbb{P}_A$ sur $(\Omega, \mathcal{B})$. L'*espérance* de Y sachant A, notée $\mathbb{E}(Y|A)$, est l'espérance de Y pour cette probabilité $\mathbb{P}_A$, c'est-à-dire le *nombre*

$$\mathbb{E}(Y|A) := \mathbb{E}^{\mathbb{P}_A}(Y) := \frac{1}{\mathbb{P}(A)} \sum_{y\in Y(\Omega)} y\mathbb{P}(\{Y = y\} \cap A\}.$$

C'est donc la moyenne des valeurs de Y pondérée par la probabilité qu'ont ces valeurs lorsque l'évènement A est supposé assuré.

Si l'on considère le cas où l'évènement A est la survenue d'une valeur (non négligeable) d'une v.a. X, l'espérance conditionnelle permet d'associer à la v.a. Y une nouvelle v.a. $\overline{Y}_X$ ne dépendant que des valeurs de X. Voici comment :

Nous supposons que la v.a. X est élémentaire et notons $\mathcal{X} := X(\Omega) = \{x_1, \ldots, x_n\}$ l'ensemble de ses valeurs prises (avec probabilité non nulle, par définition d'une v.a. élémentaire). Ce qui suit n'est pas restreint à ce cas particulier mais l'exposition du cas général implique des constructions masquant trop facilement le but poursuivi.

Soit $x \in \mathcal{X}$; l'évènement $B_x := \{X = x\}$ est de probabilité non nulle et on peut considérer la probabilité $\mathbb{P}_{B_x}$ conditionnellement à cet évènement. Soit à présent Y une v.a. ; notons $\mathbb{E}_x(Y)$ l'espérance de Y pour la probabilité $\mathbb{P}_{B_x}$. Ce nombre ne prend donc en compte que les valeurs $Y(\omega)$ que prend Y lorsque $X(\omega) = x$; en effet

$$\mathbb{E}_x(Y) = \sum_{y \in Y(\Omega)} y \mathbb{P}_{B_x}(\{Y = y\}) = \frac{1}{\mathbb{P}(\{X = x\})} \sum_{y \in Y(\Omega)} y \mathbb{P}(\{Y = y\} \cap \{X = x\}),$$

et $\mathbb{P}(\{Y = y\} \cap \{X = x\}) = 0$ si $y \notin Y(\{X = x\})$, puisque $\{Y = y\} \cap \{X = x\} = \emptyset$ dans ce cas. Cette formule montre que le nombre $\mathbb{E}_x(Y)$ est la moyenne des valeurs de Y pondérée par la probabilité de $\{Y = y\} \cap \{X = x\}$. Si l'on effectue cette opération pour chaque $x \in X(\Omega)$ on définit une fonction $x \mapsto g(x) := \mathbb{E}_x(Y)$; nous pouvons alors considérer la nouvelle v.a. $\overline{Y}_X := g(X)$; par construction elle est constante sur les parties de la partition de Ω en les $\{X = x\}$, $x \in \mathcal{X}$, et cette valeur constante est une moyenne de Y sur $\{X = x\}$; nous pouvons donc comprendre $\overline{Y}_X$ comme la v.a. X-mesurable "ressemblant le plus à Y" ; en particulier, si Y est (déjà) X-mesurable, on a $\overline{Y}_X = Y$. En revanche, si Y et X sont indépendantes, ce traitement dégrade complétement la v.a. Y en une constante (la constante qui ressemble le plus à Y, à savoir $\mathbb{E}(X)$). Encore un mot : $\overline{Y}_X$ se note $\mathbb{E}(Y|X)$, notation commode dans les calculs, mais peut-être pas très évocatrice.

A noter que la v.a. X n'a été utilisée dans cette construction que pour définir la partition de Ω en les atomes de $\alpha(X)$ que sont les évènement $\{X = x\}$, pour $x \in \mathcal{X}$. Soit $\mathcal{A}$ est une algèbre quelconque et supposons que ses atomes A_i soient de probabilité non nulle. La même construction conduit à la v.a. $\mathbb{E}(Y|\mathcal{A})$, appelée "espérance conditionnelle de Y relativement à l'algèbre $\mathcal{A}$", qui est constante sur les atomes A_i de $\mathcal{A}$, définie par

$$\mathbb{E}(Y|\mathcal{A})(\omega) = \mathbb{E}(Y|A_i) \text{ pour tout } \omega \in A_i.$$

Exercice : Soit X une v.a. élémentaire. Montrer que $\mathbb{E}(Y|X) = \mathbb{E}(Y|\alpha(X)\}$; en déduire que $\mathbb{E}(Y|X) = \mathbb{E}(Y|g(X))$ pour toute application injective g. Ceci montre le rôle réduit des valeurs de la v.a. X dans $\mathbb{E}(Y|X)$: seule l'*information révélée par* X (c'est-à-dire $\alpha(X)$) est essentielle.

Théorème 3.5 *Soient $\mathcal{A}^+ \supseteq \mathcal{A}^-$ deux algèbres, alors* $\mathbb{E}(\mathbb{E}(X|\mathcal{A}^+)|\mathcal{A}^-) = \mathbb{E}(X|\mathcal{A}^-)$.

Si Z est $\mathcal{A}$-mesurable, alors $\mathbb{E}(ZX|\mathcal{A}) = Z\mathbb{E}(X|\mathcal{A})$.

Soit Y une v.a. $\mathcal{A}$-mesurable. Alors $Y = \mathbb{E}(X|\mathcal{A})$ *si et seulement si* $\mathbb{E}(Y\mathbb{I}_A) = \mathbb{E}(X\mathbb{I}_A)$ *pour tout* $A \in \mathcal{A}$

Exercice : Montrer ce théorème 3.5 dans le cas où les atomes de $\mathcal{A}$ et $\mathcal{A}^+$ (et donc ceux de $\mathcal{A}^-$) ne sont pas négligeables et Z est une v.a. élémentaire.

Thomas Bayes (1702-1761) :