2.2 Tribu ou σ−algèbre

Définition 2.1 (Tribu ou σ−algèbre) *Soient* E *un ensemble,* T *une famille de parties de* E *(i.e.* $T \subset \mathcal{P}(E)$*). La famille* T *est une tribu (on dit aussi une* σ*−algèbre) sur* E *si* T *vérifie :*

1. $\emptyset \in T, E \in T,$
2. T *est stable par union dénombrable, c'est-à-dire que pour toute famille dénombrable* $(A_n)_{n\in\mathbb{N}}$ *d'éléments de* T*, on a* $\bigcup_{n\in\mathbb{N}} A_n \in T$.
3. T *est stable par intersection dénombrable, c'est-à-dire que pour toute famille dénombrable* $(A_n)_{n\in\mathbb{N}}$ *d'éléments de* T*, on a* $\bigcap_{n\in\mathbb{N}} A_n \in T$.
4. T *est stable par passage au complémentaire, c'est-à-dire que pour tout* $A \in T$*, on a* $A^c \in T$ *(On rappelle que* $A^c = E \setminus A$*).*

Il est clair que, pour montrer qu'une partie T de $\mathcal{P}(E)$ est une tribu, il est inutile de vérifier les propriétés 1-4 de la définition précédente. Il suffit de vérifier par exemple $\emptyset \in T$ (ou $E \in T$), 2 (ou 3) et 4.

Exemples de tribus sur E : $\{\emptyset, E\}$ et $\mathcal{P}(E)$ sont des tribus sur E.

Définition 2.2 (Langage probabiliste) *Soient* E *un ensemble quelconque (parfois appelé l'univers des possibles) et* T *une tribu ; on appelle éventualités les éléments de* E *et événements les éléments de* T*. On appelle événement élémentaire un singleton appartenant à* T*. On dit que deux événements* $A, B \in T$ *sont incompatibles si* $A \cap B = \emptyset$*.*

Proposition 2.3 (Stabilité par intersection des tribus) *Soient* E *et* I *deux ensembles. Pour tout* $i \in I$*, on se donne une tribu* T_i *sur* E*. Alors, la famille (de parties de* E*)*

$$\bigcap_{i\in I} T_i = \{A \subset E;\ A \in T_i,\ \forall i \in I\}$$

est encore une tribu sur E*.*

DÉMONSTRATION – La démonstration de cette proposition fait l'objet de la première question de l'exercice 2.2. ■

Cette proposition nous permet de définir ci-après la notion de tribu engendrée.

Définition 2.4 (Tribu engendrée) *Soient* E *un ensemble et* $\mathcal{C} \subset \mathcal{P}(E)$*. On appelle tribu engendrée par* $\mathcal{C}$ *la plus petite tribu contenant* $\mathcal{C}$*, c'est-à-dire la tribu* $T(\mathcal{C})$ *intersection de toutes les tribus sur* E *contenant* $\mathcal{C}$ *(cette intersection est non vide car* $\mathcal{P}(E)$ *est une tribu contenant* $\mathcal{C}$*).*

2.3 Mesure, probabilité

Définition 2.13 (Mesure) *Soit* (E, T) *un espace mesurable. On appelle mesure une application* $m : T \to \overline{\mathbb{R}}_+$ *(avec* $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup (+\infty)$*) vérifiant :*

1. $m(\emptyset) = 0$,
2. *m est σ-additive, c'est-à-dire que pour toute famille* $(A_n)_{n\in\mathbb{N}}$ *d'éléments de* T *disjoints deux à deux (i.e. tels que* $A_n \cap A_m = \emptyset$, *si* $n \neq m$*) on a :*
$$m\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} m(A_n). \tag{2.1}$$

Remarque 2.14

1. Dans la définition précédente on a étendu à $\overline{\mathbb{R}}_+$ l'addition dans $\mathbb{R}_+$. On a simplement posé $x + (+\infty) = +\infty$, pour tout $x \in \overline{\mathbb{R}}_+$. Noter également que la somme de la série dans la définition précédente est à prendre dans $\overline{\mathbb{R}}_+$ et que, bien sûr, $a = \sum_{n\in\mathbb{N}} a_n$ signifie simplement que $\sum_{p=0}^{n} a_p \to a$ (dans $\overline{\mathbb{R}}_+$) quand $n \to +\infty$.
2. Soient $x, y, z \in \overline{\mathbb{R}}_+$. Remarquer que $x + y = x + z$ implique $y = z$ si $x \neq +\infty$.
3. Dans la définition précédente, la condition 1. peut être remplacée par la condition : $\exists\ A \in T, m(A) < \infty$. La vérification de cette affirmation est laissée au lecteur attentif.
4. Il est intéressant de remarquer que, pour une série à termes positifs, l'ordre de sommation est sans importance. Plus précisément, si $(a_n)_{n\in\mathbb{N}} \subset \overline{\mathbb{R}}_+$ et si φ est une bijection de $\mathbb{N}$ dans $\mathbb{N}$, on a $\sum_{n\in\mathbb{N}} a_n = \sum_{n\in\mathbb{N}} a_{\varphi(n)}$. C'est l'objet du lemme 2.15.
5. Une conséquence immédiate de la σ-additivité est l'additivité, c'est-à-dire que
$$m\Big(\bigcup_{p=0}^{n} A_p\Big) = \sum_{p=0}^{n} m(A_p)$$
pour toute famille finie $(A_p)_{p=0,\ldots,n}$ d'éléments de T, disjoints deux à deux. L'additivité se démontre avec la σ-additivité en prenant $A_p = \emptyset$ pour $p > n$ dans (2.1).
6. Dans le cas $E = \mathbb{R}$ et $T = \mathcal{P}(\mathbb{R})$, il est facile de construire des mesures sur T, mais il n'existe pas de mesure sur T, notée m, telle que $m(]a, b[) = b - a$ pour tout a, $b \in \mathbb{R}$, $a < b$ (voir les exercices 2.29 et 2.28). Une telle mesure existe si on prend pour T la tribu borélienne de $\mathbb{R}$, c'est l'objet de la section 2.5.

Lemme 2.15 *Soit* $(a_n)_{n\in\mathbb{N}} \subset \overline{\mathbb{R}}_+$ *et soit* $\varphi : \mathbb{N} \to \mathbb{N}$ *bijective ; alors*
$$\sum_{n\in\mathbb{N}} a_n = \sum_{n\in\mathbb{N}} a_{\varphi(n)}.$$

DÉMONSTRATION – On pose
$$A = \sum_{n\in\mathbb{N}} a_n \Big(= \lim_{n\to+\infty} \sum_{p=0}^{n} a_p\Big) \text{ et } B = \sum_{n\in\mathbb{N}} a_{\varphi(n)} \Big(= \lim_{n\to+\infty} \sum_{p=0}^{n} a_{\varphi(p)}\Big).$$
Noter que $A, B \in \overline{\mathbb{R}}_+$. On veut montrer que $A = B$.

On montre d'abord que $B \leq A$. Soit $n \in \mathbb{N}$. On pose $N = \max\{\varphi(0), \ldots, \varphi(n)\}$. Comme $a_q \geq 0$ pour tout $q \in \mathbb{N}$, on a $\sum_{p=0}^{n} a_{\varphi(p)} \leq \sum_{p=0}^{N} a_p \leq A$. On en déduit, faisant tendre n vers ∞ que $B \leq A$.

En raisonnant avec l'inverse de φ on a aussi $A \leq B$ et finalement $A = B$. ∎

Définition 2.16 (Mesure finie et probabilité) *Soit* (E, T) *un espace mesurable.*

1. On appelle mesure finie une mesure m *sur* T *telle que* $m(E) < \infty$.

2. On appelle probabilité une mesure p *sur* T *telle que* $p(E) = 1$.

Définition 2.17 (Espace mesuré, espace probabilisé) *Soient* (E, T) *un espace mesurable, et* m *une mesure (resp. une probabilité) sur* T. *Le triplet* (E, T, m) *est appelé espace mesuré (resp. espace probabilisé).*

Définition 2.18 (Mesure σ-finie) *Soit* (E, T, m) *un espace mesuré, on dit que* m *est σ-finie (ou que* (E, T, m) *est σ-fini) si il existe une suite* $(A_n)_{n\in\mathbb{N}}$ *telle que*

$$A_n \in T, \quad m(A_n) < +\infty, \text{ pour tout } n \in \mathbb{N}, \text{ et } \quad E = \bigcup_{n\in\mathbb{N}} A_n.$$

Remarque 2.19 Soit (E, T, m) un espace mesuré σ-fini. Une conséquence de la définition 2.18 est qu'il existe alors une suite $(E_n)_{n\in\mathbb{N}}$ d'élément de T telle que $m(E_n) < +\infty$ pour tout n, $E = \cup_{n\in\mathbb{N}} E_n$ et $E_n \cap E_m = \emptyset$ si $n \neq m$. En effet, à partir de la suite $(A_n)_{n\in\mathbb{N}}$ donnée par la définition 2.18, on construit une suite $(E_n)_{n\in\mathbb{N}}$ en posant

$$E_0 = A_0 \text{ et, pour } n > 0,\ E_n = A_n \setminus \bigcup_{p=0}^{n-1} E_p.$$

La suite $(E_n)_{n\in\mathbb{N}}$ vérifie bien les propriétés désirées.

Exemple 2.20 (Mesure de Dirac) *Soient* (E, T) *un espace mesurable et* $a \in E$. *On définit sur* T *la mesure* δ_a *par (pour* $A \in T$*) :*

$$\delta_a(A) = \begin{cases} 0 & \text{si } a \notin A, \\ 1 & \text{si } a \in A. \end{cases} \tag{2.2}$$

On peut remarquer que la mesure de Dirac est une probabilité.

Remarque 2.21 (Comment choisir la probabilité) Soit (E, T) un espace probabilisable, on peut évidemment définir plusieurs probabilités sur T. C'est tout l'art de la modélisation que de choisir une probabilité qui rende compte du phénomène aléatoire que l'on veut observer. On se base pour cela souvent sur la notion de fréquence, qui est une notion expérimentale à l'origine. Soit $A \in T$ un événement, dont on cherche à évaluer la probabilité $p(A)$. On effectue pour cela N fois l'expérience dont l'univers des possibles est E, et on note N_A le nombre de fois où l'événement A est réalisé. A N fixé, on définit alors la *fréquence* $f_N(A)$ de l'événement A par :

$$f_N(A) = \frac{N_A}{N}.$$

Expérimentalement, il s'avère que $f_N(A)$ admet une limite lorsque $N \to +\infty$. C'est ce qu'on appelle la loi empirique des grands nombres. On peut donc définir expérimentalement $p(A) = \lim_{N\to+\infty} f_N(A)$. Cependant, on n'a pas ainsi démontré que p est une probabilité : il ne s'agit pour l'instant que d'une approche intuitive. On donnera plus loin la loi forte des grands nombres (proposition 6.99), qui permettra de justifier mathématiquement la loi empirique. On peut remarquer que $f_N(E) = \frac{N}{N} = 1$.

Exemple 2.22 (Le cas équiprobable) Soit (E, T, p) un espace probabilisé. On suppose que tous les singletons appartiennent à la tribu et que les événements élémentaires sont équiprobables. On a alors : $p(\{x\}) = \frac{1}{\mathrm{card} E}$ pour tout $x \in E$.

Définition 2.23 (Mesure atomique) *Soit (E, T, m) un espace mesuré tel que : $\{x\} \in T$ pour tout x de E. On dit que m est portée par $S \in T$ si $m(S^c) = 0$. Soit $x \in E$, on dit que x est un atome ponctuel de m si $m(\{x\}) \neq 0$. On dit que m est purement atomique si elle est portée par la partie de E formée par l'ensemble de ses atomes ponctuels.*

Définition 2.24 (Mesure diffuse) *Soient (E, T) un espace mesurable et m une mesure sur T. On dit que m est diffuse si $\{x\} \in T$ et $m(\{x\}) = 0$ pour tout $x \in E$. (Cette définition est aussi valable pour une mesure signée sur T, définie dans la section 2.4.)*

Définition 2.25 (Partie négligeable) *Soient (E, T, m) un espace mesuré et $A \subset E$. On dit que A est négligeable s'il existe un ensemble $B \in T$ tel que $A \subset B$ et $m(B) = 0$.*

Définition 2.26 (Mesure complète) *Soit (E, T, m) un espace mesuré, on dit que m est complète (ou que l'espace (E, T, m) est complet) si toutes les parties négligeables sont mesurables, c'est-à-dire appartiennent à T.*

La proposition suivante donne les principales propriétés d'une mesure.

Proposition 2.27 (Propriétés des mesures) *Soit (E, T, m) un espace mesuré. La mesure m vérifie les quatre propriétés suivantes :*

1. Monotonie : Soit $A, B \in T$, $A \subset B$, alors

$$m(A) \leq m(B). \tag{2.3}$$

2. σ-sous-additivité : Soit $(A_n)_{n\in\mathbb{N}} \subset T$, alors

$$m\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \leq \sum_{n\in\mathbb{N}} m(A_n). \tag{2.4}$$

3. Continuité croissante : Soit $(A_n)_{n\in\mathbb{N}} \subset T$, telle que $A_n \subset A_{n+1}$, pour tout $n \in \mathbb{N}$, alors

$$m\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty}(m(A_n)) = \sup_{n\in\mathbb{N}}(m(A_n)). \tag{2.5}$$

4. Continuité décroissante : Soit $(A_n)_{n\in\mathbb{N}} \subset T$, telle que $A_{n+1} \subset A_n$, pour tout $n \in \mathbb{N}$, et telle que il existe $n_0 \in \mathbb{N}$, $m(A_{n_0}) < \infty$, alors

$$m\Big(\bigcap_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty}(m(A_n)) = \inf_{n\in\mathbb{N}}(m(A_n)). \tag{2.6}$$

2.6 Indépendance et probabilité conditionnelle

2.6.1 Probabilité conditionnelle

Commençons par expliquer la notion de probabilité conditionnelle sur l'exemple du lancer de dé. On se place dans le modèle équiprobable : soient $E = \{1, 2, 3, 4, 5, 6\}$, $T = \mathcal{P}(E)$ et p la probabilité définie par $p(\{x\}) = \frac{1}{6}$, $\forall x \in E$. La probabilité de l'événement A "obtenir 6" est $\frac{1}{6}$. Supposons maintenant que l'on veuille évaluer la chance d'obtenir un 6, alors que l'on sait déjà que le résultat est pair (événement $B = \{2, 4, 6\}$). Intuitivement, on a envie de dire que la "chance" d'obtenir un 6 est alors $\frac{1}{\text{card} B} = \frac{1}{3}$.

Définition 2.51 (Probabilité conditionnelle) *Soient* (E, T, p) *un espace probabilisé et* $A, B \in T$.

Si $p(B) \neq 0$ *la probabilité conditionnelle de* A *par rapport à* B *(on dit aussi probabilité de* A *par rapport à* B*), notée* $p(A|B)$*, est définie par* $p(A|B) = \frac{p(A \cap B)}{p(B)}$.

Si $p(B) = 0$ *la probabilité conditionnelle de* A *par rapport à* B*, notée* $p(A|B)$*, n'est pas définie. C'est un nombre arbitraire entre* 0 *et* 1.

De cette définition on déduit la formule de Bayes : soient (E, T, p) un espace probabilisé et $A, B \in T$, alors :

$$p(B)p(A|B) = p(A \cap B) \tag{2.16}$$

Remarque 2.52 Soient (E, T, p) un espace probabilisé et A un événement tel que $p(A) \neq 0$. Alors l'application $p_A : T \to [0, 1]$ définie par :

$$p_A(B) = p(B|A) = \frac{p(A \cap B)}{p(A)}, \forall B \in T$$

est une probabilité sur T. On dit que "la masse de p_A est concentrée en A" : on a en effet : $p_A(B) = 0$, pour tout $B \in T$ tel que $A \cap B = \emptyset$. On a aussi $p_A(A) = 1$.

Remarque 2.53 *Voici un corollaire immédiat de la relation 2.16. Soit* (E, T, p) *est un espace probabilisé et* $(C_n)_{n\in\mathbb{N}} \subset T$ *une partition de* E *telle que* $p(C_n) \neq 0$. *On a alors, pour tout* $A \in T$,

$$p(A) = \sum_{n\in\mathbb{N}} p(C_n)p(A|C_n).$$

2.6.2 Evénements indépendants, tribus indépendantes

Définition 2.54 (Indépendance de deux événements) *Soient* (E, T, p) *un espace probabilisé. On dit que deux événements* A *et* B *sont indépendants si* $p(A)p(B) = p(A \cap B)$.

Remarque 2.55 Lors de la modélisation d'un phénomène aléatoire, il y a des événements qui semblent *a priori* indépendants, c'est-à-dire que la réalisation de l'un semble n'avoir aucune influence sur la réalisation de l'autre. On choisira alors, pour le modèle probabiliste, une probabilité qui respecte cette indépendance. Attention toutefois, pour une probabilité p donnée, deux événements peuvent être indépendants alors qu'ils ne paraissent pas intuitivement indépendants, voir à ce sujet l'exercice 9.14 page 249 sur les variables aléatoires indépendantes.

Exemple 2.56 Prenons comme exemple le lancer simultané de deux dés : *a priori*, il parait raisonnable de supposer que les résultats obtenus pour chacun des deux dés n'influent pas l'un sur l'autre, et on va donc chercher une probabilité qui respecte cette indépendance. L'univers des possibles est ici

$$E = \{(i, j), 1 \leq i \leq 6, 1 \leq j \leq 6\}.$$

Les résultats de chaque lancer simultané des deux dés étant équiprobables, on a donc envie de définir, pour $A \in \mathcal{P}(E)$, $p(A) = \frac{\mathrm{card} A}{36}$. Voyons maintenant si deux événements *a priori* indépendants sont indépendants pour cette probabilité. Considérons par exemple l'événement A : "obtenir un double 6" ; on peut écrire : $A = B \cap C$, où B est l'événement "obtenir un 6 sur le premier dé" et C l'événement "obtenir un 6 sur le deuxième dé". On doit donc vérifier que : $p(A) = p(B)p(C)$. Or $B = \{(6, j), 1 \leq j \leq 6\}$ et $C = \{(i, 6), 1 \leq i \leq 6\}$. On a donc $p(B) = p(C) = \frac{1}{6}$, et on a bien $p(A) = p(B)p(C)(= \frac{1}{36})$.

On généralise la notion d'indépendance de deux événements en introduisant la notion d'indépendance de tribus.

Définition 2.57 (Indépendance des tribus) *Soit* (E, T, p) *un espace probabilisé et* $(T_k)_{k\in\mathbb{N}^*}$ *une suite de tribus incluses dans* T.

1. *Soit* $N > 1$. *On dit que les* N *tribus* T_k, $k = 1, \ldots, N$, *sont indépendantes (on dit aussi que la suite* $T_1, \ldots, T_N$ *est indépendante) si pour toute famille* $(A_1, \ldots, A_N)$ *d'événements tels que* $A_k \in T_k$ *pour* $k = 1, \ldots, N$ *on a :* $p(\bigcap_{k=1}^{N} A_k) = p(A_1)p(A_2) \ldots p(A_N)$.
2. *On dit que la suite* $(T_k)_{k\in\mathbb{N}^*}$ *est indépendante (ou que les tribus* $T_1, \ldots, T_n, \ldots$ *sont indépendantes) si pour tout* $N \geq 1$, *les* N *tribus* T_k, $k = 1, \ldots, N$, *sont indépendantes.*

On peut facilement remarquer que si A et B sont deux événements d'un espace probabilisé (E, T, p), ils sont indépendants (au sens de la définition 2.54) si et seulement si les tribus $T_A = \{\emptyset, E, A, A^c\}$ et $T_B = \{\emptyset, E, B, B^c\}$ sont indépendantes (voir l'exercice 3.19). On en déduit la généralisation de la définition d'indépendance à plusieurs événements :

Définition 2.58 (Evénements indépendants) *Soient* (E, T, p) *un espace probabilisé et* $(A_k)_{k=1,\ldots,N}$ *des événements, on dit que les* N *événements* $(A_k)_{k=1,\ldots,N}$ *sont indépendants si les* N *tribus engendrées par les événements* A_k, $k = 1 \ldots, N$ *(c'est-à-dire les* N *tribus définies par* $T_k = \{A_k, A_k^c, E, \emptyset\}$ *pour* $k = 1 \ldots, N$*) sont indépendantes.*

Sous les hypothèses de la définition précédente, on peut remarquer que les événements $A_1, \ldots, A_N$ sont indépendants, c'est-à-dire que les tribus engendrées par $A_1, \ldots, A_N$ sont indépendantes) si et seulement si

$$P(\bigcap_{i\in I} A_i) = \prod_{i\in I} P(A_i) \text{ pour tout } I \subset \{1, \ldots, N\},$$

voir l'exercice 3.19. Nous terminons ce paragraphe par une proposition sur les tribus indépendantes :

Proposition 2.59 *Soit* (E, T, p) *un espace probabilisé.*

1. *Soit* $N > 1$ *et* $(T_k)_{k\in\{0,\ldots N\}}$ *une suite indépendante de tribus incluses dans* T. *La tribu* T_0 *est alors indépendante de la tribu engendrée par les tribus* $T_1, \ldots, T_N$.
2. *(Généralisation) Soit* $N > 1$, $q > 1$, $n_0, \ldots, n_q$ *tel que* $n_0 = 0$, $n_i \leq n_{i+1}$ *(pour* $i = 0, \ldots, q-1$*),* $n_q = N$ *et* $(T_k)_{k\in\{0,\ldots N\}}$ *une suite indépendante de tribus incluses dans* T. *Pour* $i = 1, \ldots, q$, *on note* τ_i *la tribu engendrée par les tribus* T_n *pour* $n = n_{i-1}, \ldots, n_i$. *Alors, les tribus* $\tau_1, \ldots, \tau_q$ *sont indépendantes.*

DÉMONSTRATION – On montre tout d'abord le premier item de la proposition. On note S la tribu engendrée par les tribus $T_1, \ldots, T_N$. Comme S est la plus petite tribu contenant les tribus T_k ($k = 1, \ldots, N$), elle est incluse dans T. On veut montrer que T_0 et S sont indépendantes, c'est-à-dire que $p(A \cap B) = p(A)p(B)$ pour tout $A \in T_0$ et tout $B \in S$. Pour le montrer, on va utiliser la proposition 2.31 (donnant l'unicité d'une mesure). Soit $A \in T_0$, on définit les mesures m et μ sur T en posant :

$$m(B) = p(A \cap B),\ \mu(B) = p(A)p(B),\ \text{pour } B \in T,$$

et on pose :

$$\mathcal{C} = \{\bigcap_{k=1}^{N} A_k,\ A_k \in T_k \text{ pour } k = 1, \ldots, N\}.$$

Pour $B \in \mathcal{C}$, on a $B = \bigcap_{k=1}^{N} A_k$ avec $A_k \in T_k$ avec $k = 1, \ldots, N$. On a donc, en utilisant l'indépendance des tribus T_0, $T_1, \ldots, T_N$,

$$m(B) = p(A \cap B) = p(A)p(A_1)p(A_2) \ldots p(A_N) = p(A)p(B) = \mu(B).$$

On a donc $m = \mu$ sur $\mathcal{C}$. Comme $\mathcal{C}$ est stable par intersection et que $E \in \mathcal{C}$, la proposition 2.31 nous donne $m = \mu$ sur la tribu engendrée par $\mathcal{C}$. Comme cette tribu contient toutes les tribus T_k ($k = 1, \ldots, N$), elle contient aussi S (en fait, elle est égale à S). On a donc bien montré que $p(A \cap B) = p(A)p(B)$ pour tout $B \in S$ et pour tout $A \in T_0$.

Pour montrer le deuxième item (qui est une généralisation du premier), il suffit de faire une récurrence finie de q étapes et d'utiliser la technique précédente. Par exemple, pour $q = 2$ la technique précédente donne :

$$p((\bigcap_{k=0}^{n_1} A_k) \cap B_2) = p(\bigcap_{k=0}^{n_1} A_k)p(B_2),$$

pour $A_k \in T_k$, $k = 0, \ldots, n_1$ et $B_2 \in \tau_2$. Puis en reprenant la technique précédente, on montre $p(B_1 \cap B_2) = p(B_1)p(B_2)$ pour $B_1 \in \tau_1$ et $B_2 \in \tau_2$, ce qui donne bien l'indépendance de τ_1 et τ_2. ■

2.6.3 Probabilités sur les boréliens de $\mathbb{R}$

Une probabilité est définie sur un espace probabilisable. Très souvent, on ne connaît du problème aléatoire que l'on cherche à modéliser ni l'ensemble E ("univers des possibles") ni la tribu T (ensemble des événements) ni la probabilité p. Par contre, on connaît une "image" de la probabilité p par une application (dite mesurable, voir chapitre suivant) X de E dans $\mathbb{R}$. On travaille alors avec l'espace beaucoup plus sympathique (car mieux défini...) $(\mathbb{R}, \mathcal{B}(\mathbb{R}), p_X)$, ou p_X est une probabilité sur $\mathcal{B}(\mathbb{R})$, que les probabilistes appellent "loi de probabilité" (elle dépend de p et de l'application X).
Nous donnons maintenant quelques notions propres aux lois de probabilités (ou probabilités définies sur les boréliens de $\mathbb{R}$), ainsi que quelques exemples concrets utilisés dans la représentation de phénomènes aléatoires.

Théorème 2.60 (Fonction de répartition) *Soit p une probabilité sur les boréliens de $\mathbb{R}$. On appelle fonction de répartition de la probabilité p la fonction* F, *définie de $\mathbb{R}$ dans $[0,1]$ par :* $F(t) = p(]-\infty, t])$.
La fonction F *est croissante et continue à droite. De plus, on a* $\lim_{t \to -\infty} F(t) = 0$ *et* $\lim_{t \to +\infty} F(t) = 1$.

DÉMONSTRATION – La croissance de F est une conséquence de la monotonie de p (proposition 2.27). En effet, soit $a, b \in \mathbb{R}$, $a < b$. On a $]-\infty, a[\subset]-\infty, b[$ et donc, par monotonie de p, $F(a) = p(]-\infty, a[) \leq p(]-\infty, b[) = F(b)$, ce qui montre bien la croissance de F.

Pour montrer que F est continue à droite, on utilise la continuité décroissante de p (proposition 2.27). Soit $a \in \mathbb{R}$ et $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ telle que $a_n \downarrow a$ (c'est-à-dire $a_{n+1} \leq a_n$ pour tout $n \in \mathbb{N}$ et $\lim_{n \to +\infty} a_n = a$). On remarque que

$$]-\infty, a] = \bigcap_{n \in \mathbb{N}}]-\infty, a_n],\ ]-\infty, a_{n+1}] \subset]-\infty, a_n] \text{ et } p(]-\infty, a_n]) < +\infty$$

pour tout $n \in \mathbb{N}$. La continuité décroissante de p donne alors

$$\mathrm{F}(a_n) = p(]-\infty, a_n]) \to p(]-\infty, a]) = \mathrm{F}(a) \text{ lorsque } n \to +\infty.$$

Ceci montre la continuité à droite de F.

Pour montrer que $\lim_{a\to+\infty} \mathrm{F}(a) = 1$, on utilise la continuité croissante de p. Soit $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ telle que $a_n \uparrow +\infty$ (c'est-à-dire $a_{n+1} \geq a_n$ pour tout $n \in \mathbb{N}$ et $a_n \to +\infty$). On pose $\mathrm{A}_n =]-\infty, a_n]$. On a $\mathrm{A}_n \subset \mathrm{A}_{n+1}$ pour tout $n \in \mathbb{N}$ et $\bigcup_{n\in\mathbb{N}} \mathrm{A}_n = \mathbb{R}$. Par continuité croissante de p (Proposition 2.27), on a donc

$$\mathrm{F}(a_n) = p(\mathrm{A}_n) \to p(\mathbb{R}) = 1 \text{ lorsque } n \to +\infty.$$

Ceci prouve que $\lim_{a\to+\infty} \mathrm{F}(a) = 1$.

Pour montrer que $\lim_{a\to-\infty} \mathrm{F}(a) = 0$, on utilise la continuité décroissante de p. Soit $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ telle que $a_n \downarrow -\infty$ (c'est-à-dire $a_{n+1} \leq a_n$ pour tout $n \in \mathbb{N}$ et $\lim_{n\to+\infty} a_n = -\infty$). On pose $\mathrm{B}_n =]-\infty, a_n]$. On a $\mathrm{B}_{n+1} \subset \mathrm{B}_n$ pour tout $n \in \mathbb{N}$, $p(\mathrm{B}_n) < +\infty$ pour tout $n \in \mathbb{N}$ et $\bigcap_{n\in\mathbb{N}} \mathrm{B}_n = \emptyset$. Par continuité décroissante de p (Proposition 2.27), on a donc

$$\mathrm{F}(a_n) = p(\mathrm{B}_n) \to p(\emptyset) \text{ lorsque } n \to +\infty.$$

Ceci prouve que $\lim_{a\to-\infty} \mathrm{F}(a) = 0$. ■

Le théorème 2.60 a une réciproque que nous énonçons dans le théorème 2.61.

Théorème 2.61 (Fonction de répartition et probabilité) *Soit* F *une fonction de* $\mathbb{R}$ *dans* $\mathbb{R}$*, croissante, continue à droite et telle que*

$$\lim_{t\to-\infty}, \mathrm{F}(t) = 0 \text{ et } \lim_{t\to+\infty} \mathrm{F}(t) = 1.$$

Alors, il existe une unique probabilité p sur $\mathcal{B}(\mathbb{R})$ telle que F *soit la fonction de répartition de p.*

La démonstration du théorème 2.61 n'est pas faite ici car ce théorème est essentiellement contenu dans le théorème 2.62 que nous donnons maintenant.

Théorème 2.62 (Lebesgue-Stieltjes)

1. *Soit m une mesure sur $\mathcal{B}(\mathbb{R})$, finie sur les compacts (on dit "localement finie"). Soit $a \in \mathbb{R}$, on définit la fonction* F *de* $\mathbb{R}$ *dans* $\mathbb{R}$ *par :* $\mathrm{F}(t) = m(]a, t])$ *si $t \geq a$ et* $\mathrm{F}(t) = -m(]t, a])$ *si $t \leq a$. Alors, la fonction* F *est continue à droite et croissante.*
2. *Réciproquement, soit* F *une fonction de* $\mathbb{R}$ *dans* $\mathbb{R}$*, croissante et continue à droite. Alors,il existe une unique mesure m sur $\mathcal{B}(\mathbb{R})$ telle que pour tout $a, b \in \mathbb{R}$ avec $a \leq b$, on ait $m(]a, b]) = \mathrm{F}(b) - \mathrm{F}(a)$. Cette mesure s'appelle la mesure de Lebesgue-Stieltjes associée à* F.

DÉMONSTRATION – La démonstration du premier item est essentiellement la même que celle de la proposition 2.60. Elle n'est pas détaillée ici.

Pour démontrer le deuxième item, on introduit l, application définie de l'ensemble des intervalles de $\mathbb{R}$ de la forme $]a, b]$ dans $\mathbb{R}$ ($a < b$) par : $l(]a, b]) = \mathrm{F}(b) - \mathrm{F}(a)$. La démonstration du fait qu'il existe un prolongement unique de cette application en une mesure sur $\mathcal{B}(\mathbb{R})$ est très voisine à celle du théorème de Carathéodory (théorème 2.35). Elle n'est pas détaillée ici. ■

Donnons, pour clore ce chapitre, quelques exemples de lois de probabilités, c'est-à-dire de probabilités sur les boréliens de $\mathbb{R}$, et leurs fonctions de répartition associées.

Définition 2.63 (Loi de probabilité discrète) *Soit p une loi de probabilité. On dit que p est discrète si elle est purement atomique. L'ensemble de ses atomes $\mathcal{A}$ est nécessairement dénombrable (voir l'exercice 2.23). La probabilité p s'écrit alors*

$$p = \sum_{a\in\mathcal{A}} p(\{a\})\delta_a,$$

où δ_a désigne la mesure de Dirac en a., définie par (2.2) *La fonction de répartition de la probabilité p est définie par :*

$$\mathrm{F}(t) = \sum_{a\in\mathcal{A}, a\leq t} p(\{a\}).$$

Exemple 2.64 (Exemples de lois discrètes) *Donnons quelques exemples de probabilités discrètes, p, sur $\mathcal{B}(\mathbb{R})$ et de $\mathcal{A}$ l'ensemble (dénombrable) de leurs atomes.*

- *La loi uniforme discrète :* $\mathrm{N} \in \mathbb{N}^*$, $\mathcal{A} = \{a_1, \ldots, a_\mathrm{N}\}$, $p(\{a_i\}) = \dfrac{1}{\mathrm{N}}$
- *La loi binomiale :* $\mathrm{N} \in \mathbb{N}^*$, $\mathcal{A} = \{1, \ldots, \mathrm{N}\}$, $\mathrm{P} \in]0,1[$, $p(\{k\}) = \mathrm{C}_\mathrm{N}^k \mathrm{P}^k (1-\mathrm{P})^{\mathrm{N}-k}$
- *La loi de Pascal :* $\mathcal{A} = \mathbb{N}^*$, $\mathrm{P} \in]0,1[$, $p(\{k\}) = \mathrm{P}(1-\mathrm{P})^{k-1}$
- *La loi de Poisson à paramètre λ :* $\mathcal{A} = \mathbb{N}$, $\lambda > 0$, $p(\{k\}) = e^{-\lambda}\dfrac{\lambda^k}{k!}$

Définition 2.65 (Loi continue) *Soit p une probabilité sur les boréliens de $\mathbb{R}$. On dit que p est continue si sa fonction de répartition est continue.*

Exemple 2.66 (Exemple de loi continue) La plupart des exemples de probabilités continues provient de ce qu'on appelle les mesures de densité par rapport à la mesure de Lebesgue, pour lesquelles on a besoin de la notion d'intégrale de Lebesgue qu'on n'a pas encore introduite. On peut toutefois déjà citer l'exemple de la loi uniforme sur un intervalle $[a,b]$ de $\mathbb{R}$: Soient $-\infty < a < b < +\infty$; pour $\mathrm{A} \in \mathcal{B}(\mathbb{R})$, on pose

$$p(\mathrm{A}) = \frac{\lambda(\mathrm{A} \cap [a,b])}{b-a}.$$

On vérifie facilement que p est une probabilité appelée probabilité uniforme sur $[a,b]$.