

Chapitre 2

Applications mesurables

2.1 Topologie et tribus boréliennes de $\overline{\mathbb{R}}$ et $\overline{\mathbb{R}}_+$

Dans la théorie de l'intégration de Lebesgue, il est très commode de travailler avec des fonctions à valeurs dans la droite achevée $\overline{\mathbb{R}}$, réunion de \mathbb{R} et des points à l'infini $-\infty$ et $+\infty$. Nous ferons aussi un usage intensif de la demi droite achevée $\overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{+\infty\}$.

Pour munir $\overline{\mathbb{R}}$ d'une topologie¹ compatible avec celle de \mathbb{R} , on utilise une application strictement croissante continue f de $] - 1, +1[$ sur \mathbb{R} , par exemple $t \mapsto \tan(\pi t/2)$. L'application f est une bijection *bicontinue* de $] - 1, +1[$ sur \mathbb{R} (un homéomorphisme). On la prolonge en une bijection \tilde{f} de $[-1, +1]$ sur $\overline{\mathbb{R}}$ en posant $\tilde{f}(x) := f(x)$ pour $x \in] - 1, +1[$, $\tilde{f}(-1) := -\infty$ et $\tilde{f}(+1) := +\infty$. En transportant par f la topologie de $[-1, +1]$ sur $\overline{\mathbb{R}}$, on fait de \tilde{f} un homéomorphisme. Ainsi les intervalles $[-1, a[$, ($-1 < a < +1$) forment un système fondamental de voisinages de $-\infty$ dans $\overline{\mathbb{R}}$, les intervalles $[-\infty, b[$, ($b \in \mathbb{R}$). De même un système fondamental de voisinages de $+\infty$ est formé des $]b, +\infty]$. Comme $] - 1, +1[$ est un ouvert de $[-1, +1]$, son image \mathbb{R} par \tilde{f} est un ouvert de $\overline{\mathbb{R}}$. On utilise aussi \tilde{f} pour transporter la relation d'ordre dans $[-1, +1]$ à $\overline{\mathbb{R}}$ de façon à ce que \tilde{f} soit strictement croissante.

Dans $\overline{\mathbb{R}}$, toute suite monotone a une limite. Toute série à termes positifs converge dans $\overline{\mathbb{R}}_+$. Tout sous-ensemble E de $\overline{\mathbb{R}}$ possède une borne inférieure $\inf E$ et une borne supérieure $\sup E$ dans $\overline{\mathbb{R}}$. Cependant on réservera l'appellation *borné* à un ensemble E pour lequel $-\infty < \inf E$ et $\sup E < +\infty$. De même on dira qu'une fonction $g : \Omega \rightarrow \overline{\mathbb{R}}$ est bornée si $E = g(\Omega)$ est une partie bornée de $\overline{\mathbb{R}}$ au sens précédent.

Reprenons la construction de la topologie de $\overline{\mathbb{R}}$ avec une approche métrique. On fixe \tilde{f} ayant les propriétés ci-dessus, disons $f(t) = \tan(\pi t/2)$ pour tout $t \in \mathbb{R}$, $\tilde{f}(-1) = -\infty$ et $\tilde{f}(+1) = +\infty$. On munit alors $\overline{\mathbb{R}}$ de la distance

$$d(x, y) := |\tilde{f}^{-1}(x) - \tilde{f}^{-1}(y)| = \frac{2}{\pi} |\arctan x - \arctan y| = \left| \int_x^y \frac{2 du}{\pi(1+u^2)} \right|. \quad (2.1)$$

¹Pour les lecteurs ne connaissant que les espaces métriques, nous définirons aussi cette topologie à partir d'une distance.

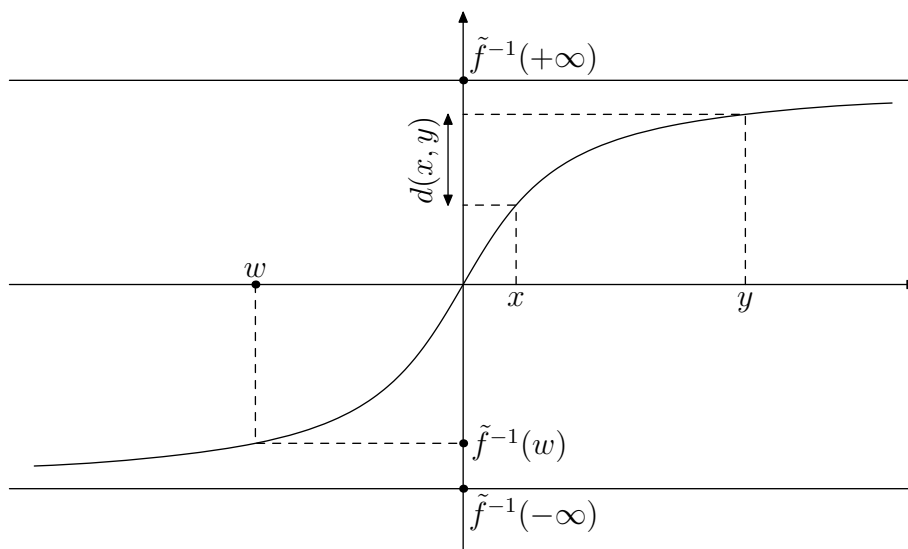


FIG. 2.1 – La distance d sur $\overline{\mathbb{R}}$

Remarquons que pour cette distance, $d(-\infty, +\infty) = 2$. Notons au passage que dans l'espace $(\overline{\mathbb{R}}, d)$, \mathbb{R} est la boule ouverte de centre 0 et de rayon 1, c'est donc un ouvert de $\overline{\mathbb{R}}$. Voyons de plus près la relation entre les boules ouvertes de $[-1, 1]$ et celles de $(\overline{\mathbb{R}}, d)$. Notons $\Delta(t_0, r)$ la boule ouverte de centre t_0 et de rayon r dans l'espace métrique $([-1, 1], \delta)$ où δ est la métrique usuelle $\delta(s, t) := |s - t|$. Il est clair que $\Delta(t_0, r) =]t_0 - r, t_0 + r[\cap [-1, 1]$. Soit $B(c, r)$ la boule de centre $c \in \overline{\mathbb{R}}$ et de rayon r dans l'espace métrique $(\overline{\mathbb{R}}, d)$. Vu la définition de d , cette boule peut s'écrire en posant $t_0 := \tilde{f}^{-1}(c)$,

$$\begin{aligned}
 B(c, r) &= \{x \in \overline{\mathbb{R}}; |\tilde{f}^{-1}(x) - \tilde{f}^{-1}(c)| < r\} \\
 &= \{x \in \overline{\mathbb{R}}; t_0 - r < \tilde{f}^{-1}(x) < t_0 + r\} \\
 &= \{x \in \overline{\mathbb{R}}; \exists t \in [-1, 1], x = \tilde{f}(t), t_0 - r < t < t_0 + r\} \\
 &= \tilde{f}(]t_0 - r, t_0 + r[\cap [-1, 1]) \\
 &= \tilde{f}(\Delta(t_0, r)).
 \end{aligned}$$

Ainsi les boules ouvertes de $(\overline{\mathbb{R}}, d)$ sont exactement les images par \tilde{f} des boules ouvertes de $([-1, 1], \delta)$. En voici la liste exhaustive, classée en trois types, les droites, les demi-droites et les segments ouverts.

- droites : $]-\infty, +\infty[$, $]-\infty, +\infty]$, $[-\infty, +\infty[$, $]-\infty, +\infty[$;
- demi-droites : $]-\infty, a[$, $[-\infty, a[$, $]a, +\infty[$, $]a, +\infty]$, a étant un réel quelconque ;
- segments ouverts : $]a, b[$ (a, b réels quelconques, avec $a < b$). L'intervalle $]a, b[$ est la boule ouverte de centre c tel que $2 \arctan c = \arctan a + \arctan b$ et de rayon $r = \frac{1}{\pi}(\arctan b - \arctan a)$.

Par restriction à \mathbb{R} , on voit ainsi que dans l'espace métrique (\mathbb{R}, d) , la famille des boules ouvertes est constituée des segments ouverts et des intervalles de la forme $]-\infty, a[$, $]a, +\infty[$, $a \in \mathbb{R}$ et de $\mathbb{R} =]-\infty, +\infty[$. Les boules ouvertes de (\mathbb{R}, d) qui ne sont pas des segments ouverts sont clairement réunions de segments ouverts. Les ouverts de (\mathbb{R}, d)

qui sont par définition, les réunions de boules ouvertes sont donc aussi les réunions de segments ouverts. Or la famille des segments ouverts est exactement la famille des boules ouvertes de l'espace métrique (\mathbb{R}, δ) où δ est la métrique usuelle $\delta(x, y) := |x - y|$. Ainsi les deux métriques d et δ génèrent la même topologie sur \mathbb{R} (i.e. ont les mêmes ouverts). Elles sont donc équivalentes.

Remarque 2.1.

- i) Tout ouvert de \mathbb{R} est aussi un ouvert de $\overline{\mathbb{R}}$.
- ii) Si W est un ouvert de $\overline{\mathbb{R}}$, $W \cap \mathbb{R}$ est ouvert de \mathbb{R} .

Justification. Si V est ouvert de \mathbb{R} , il est union de segments ouverts qui sont aussi des boules ouvertes de $\overline{\mathbb{R}}$, donc V est ouvert de $\overline{\mathbb{R}}$. Soit W un ouvert de $\overline{\mathbb{R}}$. Il s'écrit $\cup_{i \in I} B_i$ où les B_i sont des boules ouvertes de $\overline{\mathbb{R}}$ et $W \cap \mathbb{R} = \cup_{i \in I} (B_i \cap \mathbb{R})$. Si certaines des B_i contiennent $-\infty$ ou $+\infty$, l'intersection avec \mathbb{R} les transforme en intervalles ouverts de \mathbb{R} (exemple : $[-\infty, a[\cap \mathbb{R} =]-\infty, a[$) qui sont eux-mêmes unions de *segments* ouverts. Ainsi $W \cap \mathbb{R}$ est union de segments ouverts de \mathbb{R} , donc ouvert de \mathbb{R} . □

Ayant maintenant défini une topologie sur $\overline{\mathbb{R}}$, on peut le munir de la tribu borélienne correspondante. Il est alors naturel de se demander si les boréliens de $\overline{\mathbb{R}}$ diffèrent beaucoup de ceux de \mathbb{R} . Notons auparavant que $\text{Bor}(\mathbb{R})$ n'est pas une tribu sur $\overline{\mathbb{R}}$ (puisque $\overline{\mathbb{R}} \notin \text{Bor}(\mathbb{R})$) et que $\text{Bor}(\overline{\mathbb{R}})$ n'est pas davantage une tribu sur \mathbb{R} puisque $\overline{\mathbb{R}} \in \text{Bor}(\overline{\mathbb{R}})$ et $\overline{\mathbb{R}} \not\subset \mathbb{R}$.

Proposition 2.2.

- a) Si B est un borélien de $\overline{\mathbb{R}}$, $B \cap \mathbb{R}$ est un borélien de \mathbb{R} .
- b) Tout borélien de \mathbb{R} est aussi un borélien de $\overline{\mathbb{R}}$.
- c) B est un borélien de $\overline{\mathbb{R}}$ si et seulement si $B = A \cup C$ où A est un borélien de \mathbb{R} et C est l'un des ensembles $\emptyset, \{-\infty\}, \{+\infty\}, \{-\infty, +\infty\}$.

Preuve du a). Considérons la famille \mathcal{F} de parties de $\overline{\mathbb{R}}$ définie par :

$$\mathcal{F} := \{E \in \mathcal{P}(\overline{\mathbb{R}}); E \cap \mathbb{R} \in \text{Bor}(\mathbb{R})\}.$$

Il est immédiat de vérifier que \mathcal{F} est une tribu sur $\overline{\mathbb{R}}$. Grâce à la remarque 2.1 ii), \mathcal{F} possède tous les ouverts de $\overline{\mathbb{R}}$. Par minimalité, elle contient donc la tribu engendrée par ces ouverts, c'est-à-dire $\text{Bor}(\overline{\mathbb{R}})$. L'inclusion $\text{Bor}(\overline{\mathbb{R}}) \subset \mathcal{F}$ n'est qu'une autre façon d'écrire a) qui est ainsi établie. □

Preuve du b). La famille \mathcal{G} de parties de \mathbb{R} définie par :

$$\mathcal{G} := \{A \in \text{Bor}(\mathbb{R}); A \in \text{Bor}(\overline{\mathbb{R}})\} = \text{Bor}(\mathbb{R}) \cap \text{Bor}(\overline{\mathbb{R}})$$

est une tribu² sur \mathbb{R} . En effet il est clair que \mathcal{G} possède l'ensemble vide et est stable par union dénombrable. Pour la stabilité par complémentaire, on observe que si $A \in \mathcal{G}$,

²Attention, c'est l'intersection de deux tribus sur des ensembles Ω_1 et Ω_2 différents, on ne peut donc pas appliquer ici la proposition 1.5.

alors $\overline{\mathbb{R}} \setminus A \in \text{Bor}(\overline{\mathbb{R}})$, puisque $A \in \text{Bor}(\overline{\mathbb{R}})$; en écrivant $\mathbb{R} \setminus A = (\overline{\mathbb{R}} \setminus A) \cap \mathbb{R}$, on voit que $\mathbb{R} \setminus A$ est l'intersection de deux boréliens de $\overline{\mathbb{R}}$, donc appartient aussi à $\text{Bor}(\overline{\mathbb{R}})$ et $\mathbb{R} \setminus A$ est bien élément de \mathcal{G} . Par construction, \mathcal{G} est une sous tribu de $\text{Bor}(\overline{\mathbb{R}})$. D'autre part elle contient les ouverts de \mathbb{R} (remarque 2.1 i)), donc par minimalité la tribu qu'ils engendrent, c'est à dire $\text{Bor}(\mathbb{R})$. Finalement $\mathcal{G} = \text{Bor}(\mathbb{R})$ et $\text{Bor}(\mathbb{R}) \subset \text{Bor}(\overline{\mathbb{R}})$, ce qui établit b). Attention $\text{Bor}(\mathbb{R})$ n'est pas une sous-tribu de $\text{Bor}(\overline{\mathbb{R}})$ car elles ne sont pas construites sur le même ensemble Ω . \square

Preuve du c). Soit $B \in \text{Bor}(\overline{\mathbb{R}})$, on peut l'écrire $B = (B \cap \mathbb{R}) \cup C$ où C est inclus dans $\{-\infty, +\infty\}$. Par le a), $A := B \cap \mathbb{R}$ est un borélien de \mathbb{R} .

Pour la réciproque, notons d'abord que dans $(\overline{\mathbb{R}}, d)$, les singletons sont des fermés (c'est vrai dans n'importe quel espace métrique) donc des boréliens. Par conséquent, chacun des quatre sous-ensembles de $\{-\infty, +\infty\}$ est un borélien de $\overline{\mathbb{R}}$. Cette remarque et le b) nous montrent que si A est un borélien de \mathbb{R} et C un sous-ensemble de $\{-\infty, +\infty\}$, $A \cup C$ est un borélien de $\overline{\mathbb{R}}$. \square

Passons maintenant à la topologie de $\overline{\mathbb{R}}_+$. On l'obtient naturellement comme la topologie trace de $\overline{\mathbb{R}}$, ou pour rester dans un cadre métrique, comme celle de l'espace métrique $(\overline{\mathbb{R}}_+, d)$, d étant la restriction à $\overline{\mathbb{R}}_+$ de la métrique définie en (2.1). La restriction à $[0, 1]$ de \tilde{f} est alors un homéomorphisme de cet intervalle sur $\overline{\mathbb{R}}_+$. Les boules ouvertes de $(\overline{\mathbb{R}}_+, d)$ sont les images par \tilde{f} de celles de $([0, 1], \delta)$. On peut alors reprendre tout le travail que nous venons de faire pour $\overline{\mathbb{R}}$. Le lecteur pourra établir en exercice les analogues de la remarque 2.1 et de la proposition 2.2 a) et b) obtenus en remplaçant \mathbb{R} par \mathbb{R}_+ et $\overline{\mathbb{R}}$ par $\overline{\mathbb{R}}_+$ et en déduire le résultat suivant.

Proposition 2.3. *B est un borélien de $\overline{\mathbb{R}}_+$ si et seulement si B est un borélien de \mathbb{R}_+ ou s'écrit $B = A \cup \{+\infty\}$, A étant un borélien de \mathbb{R}_+ .*

Proposition 2.4.

a) *La tribu borélienne de $\overline{\mathbb{R}}$ est engendrée par chacune des familles*

$$\mathcal{I}_1 := \{[-\infty, x[; x \in \mathbb{R}\}, \quad \mathcal{I}_2 := \{[-\infty, x]; x \in \mathbb{R}\}.$$

b) *La tribu borélienne de $\overline{\mathbb{R}}_+$ est engendrée par chacune des familles*

$$\mathcal{J}_1 := \{[0, x[; x \in \mathbb{R}_+\}, \quad \mathcal{J}_2 := \{[0, x]; x \in \mathbb{R}_+\}.$$

Démonstration. Montrons l'égalité $\sigma(\mathcal{I}_1) = \text{Bor}(\overline{\mathbb{R}})$ en établissant l'inclusion dans les deux sens. Soit V un ouvert de $\overline{\mathbb{R}}$. On peut l'écrire $V = (V \cap \mathbb{R}) \cup (V \cap \{-\infty, +\infty\})$. En notant que $\{-\infty\} = \bigcap_{n \in \mathbb{N}} [-\infty, -n[$ et $\{+\infty\} = \bigcap_{n \in \mathbb{N}} [n, +\infty[$, on voit que les ensembles $\{-\infty\}$, $\{+\infty\}$ et $] -\infty, x[$ appartiennent à la tribu $\sigma(\mathcal{I}_1)$. Par la remarque 2.1 ii), $V \cap \mathbb{R}$ appartient à $\text{Bor}(\mathbb{R})$, laquelle est engendrée par les $] -\infty, x[$ donc contenue dans $\sigma(\mathcal{I}_1)$. On en déduit l'appartenance de V à $\sigma(\mathcal{I}_1)$, puis par minimalité, l'inclusion $\text{Bor}(\overline{\mathbb{R}}) \subset \sigma(\mathcal{I}_1)$. L'inclusion dans l'autre sens résulte immédiatement de ce que les intervalles $] -\infty, x[$ sont dans $\overline{\mathbb{R}}$ des boules ouvertes de centre $-\infty$, donc des ouverts.

La justification des égalités $\sigma(\mathcal{I}_2) = \text{Bor}(\overline{\mathbb{R}})$, $\sigma(\mathcal{J}_1) = \text{Bor}(\overline{\mathbb{R}}_+)$ et $\sigma(\mathcal{J}_2) = \text{Bor}(\overline{\mathbb{R}}_+)$ est analogue et laissée en exercice au lecteur. \square

2.2 Arithmétique dans $\overline{\mathbb{R}}_+$

La nécessité d'intégrer des fonctions sur des ensembles de mesure infinie ou d'intégrer des fonctions pouvant naturellement prendre une valeur infinie en certains points (par exemple une limite supérieure d'une suite de fonctions ou la somme d'une série à termes positifs) nous amène à *prolonger* à $\overline{\mathbb{R}}_+$ l'addition et la multiplication usuelles en adoptant les conventions suivantes.

$$a + (+\infty) = (+\infty) + a = +\infty, \quad \forall a \in \overline{\mathbb{R}}_+. \quad (2.2)$$

$$a \times (+\infty) = (+\infty) \times a = \begin{cases} +\infty & \text{si } 0 < a \leq +\infty, \\ 0 & \text{si } a = 0. \end{cases} \quad (2.3)$$

L'addition et la multiplication ainsi prolongées restent commutatives et associatives. La multiplication reste distributive par rapport à l'addition. Par contre il convient d'être prudent chaque fois qu'apparaissent une soustraction ou une division. En particulier les règles de simplification habituelles ne s'étendent pas : $a + b = a + c$ n'implique $b = c$ que si a est fini et $ab = ac$ n'implique $b = c$ que si $0 < a < +\infty$.

Au premier abord, la convention $0 \times (+\infty) = 0$ dégage une odeur sulfureuse car elle paraît contradictoire avec la notion de *forme indéterminée du type* « $0 \times (+\infty)$ ». Cette contradiction n'est qu'apparente. La multiplication M , définie sur \mathbb{R}_+^2 par $M(a, b) = ab$, a été prolongée par (2.3) à $\overline{\mathbb{R}}_+^2$ en conservant de bonnes propriétés algébriques, mais on ne prétend pas que ce prolongement de M soit *continu* aux points $(0, +\infty)$ et $(+\infty, 0)$. La notion de *forme indéterminée du type* « $0 \times (+\infty)$ » nous dit précisément qu'il est *impossible de prolonger M en ces points en préservant sa continuité* : si (x_n) et (y_n) sont deux suites dans \mathbb{R}_+ convergentes l'une vers 0 et l'autre vers $+\infty$ (donc (x_n, y_n) converge vers $(0, +\infty)$ dans $\overline{\mathbb{R}}_+^2$), la suite $M(x_n, y_n) = x_n y_n$ peut selon les cas converger vers n'importe quel élément de $\overline{\mathbb{R}}_+$ ou même n'avoir aucune limite.

2.3 Mesurabilité

Définition 2.5. Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces mesurables. L'application $f : \Omega_1 \rightarrow \Omega_2$ est dite \mathcal{F}_1 - \mathcal{F}_2 mesurable si pour tout $B \in \mathcal{F}_2$, $f^{-1}(B) \in \mathcal{F}_1$, autrement dit si $f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1$.

Remarque 2.6. Il est clair d'après la définition de la mesurabilité d'une application, que celle-ci est conservée chaque fois que l'on *diminue* (au sens de l'inclusion) la tribu de l'espace d'arrivée ou que l'on *agrandit* celle de l'espace de départ.

S'il n'y a pas d'ambiguïté sur les tribus concernées, on pourra se contenter de dire « mesurable » au lieu de « \mathcal{F}_1 - \mathcal{F}_2 mesurable ». Les applications mesurables jouent un rôle central dans la théorie de l'intégration abstraite. Elles permettent de « transporter » la mesure d'un espace vers un autre. Quand l'espace d'arrivée est $\Omega_2 = \overline{\mathbb{R}}$ avec $\mathcal{F}_2 = \text{Bor}(\overline{\mathbb{R}})$, ce sont elles qui ont vocation à être intégrées. Dans le langage de la théorie des probabilités, les applications mesurables correspondent à la notion de variable aléatoire.

Définition 2.7. Soit (Ω, \mathcal{F}) un espace probabilisable. On appelle variable aléatoire réelle sur (Ω, \mathcal{F}) toute application $X : \Omega \rightarrow \mathbb{R}$, \mathcal{F} -Bor(\mathbb{R}) mesurable. De même on appelle variable aléatoire complexe sur (Ω, \mathcal{F}) toute application $X : \Omega \rightarrow \mathbb{C}$, \mathcal{F} -Bor(\mathbb{C}) mesurable et vecteur aléatoire, toute application $X : \Omega \rightarrow \mathbb{R}^d$, \mathcal{F} -Bor(\mathbb{R}^d) mesurable ($d > 1$).

Rappelons que du point de vue mathématique, un espace probabilisable n'est rien d'autre qu'un espace mesurable, c'est-à-dire un ensemble Ω muni d'une tribu de parties de Ω . L'emploi de l'adjectif « probabilisable » au lieu de mesurable indique seulement l'intention que l'on a de munir cet espace d'une (ou de plusieurs!) mesure de probabilité \mathbf{P} .

Définition 2.8. Soit (Ω, \mathcal{F}) un espace probabilisable. On appelle variable aléatoire discrète (réelle, resp. complexe) sur (Ω, \mathcal{F}) toute application $X : \Omega \rightarrow \mathbb{K}$ ($\mathbb{K} = \mathbb{R}$, resp. \mathbb{C}), telle que $X(\Omega)$ est au plus dénombrable et X est \mathcal{F} - $\mathcal{P}(\mathbb{K})$ mesurable. De même on appelle vecteur aléatoire discret, toute application $X : \Omega \rightarrow \mathbb{R}^d$, telle que $X(\Omega)$ est au plus dénombrable et X est \mathcal{F} - $\mathcal{P}(\mathbb{R}^d)$ mesurable ($d > 1$).

En vertu de la remarque 2.6, une variable aléatoire *discrète* réelle (resp. complexe) est aussi une variable aléatoire réelle (resp. complexe) et un vecteur aléatoire *discret* est aussi un vecteur aléatoire, au sens de la définition 2.7.

Un exemple simple, mais important, d'application mesurable de Ω dans \mathbb{R} (et de variable aléatoire discrète) est l'indicatrice $\mathbf{1}_A$ d'un élément A de la tribu \mathcal{F} :

$$\mathbf{1}_A(\omega) := \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

En effet pour tout $B \subset \mathbb{R}$, on a

$$(\mathbf{1}_A)^{-1}(B) = \begin{cases} \emptyset & \text{si } 0 \notin B \text{ et } 1 \notin B, \\ A & \text{si } 0 \notin B \text{ et } 1 \in B, \\ A^c & \text{si } 0 \in B \text{ et } 1 \notin B, \\ \Omega & \text{si } 0 \in B \text{ et } 1 \in B. \end{cases}$$

On vient de vérifier que $(\mathbf{1}_A)^{-1}(\mathcal{P}(\mathbb{R})) = \sigma(\{A\})$. L'application $\mathbf{1}_A$ est donc $\sigma(\{A\})$ - $\mathcal{P}(\mathbb{R})$ mesurable. Elle est donc aussi \mathcal{F} - \mathcal{B} mesurable pour toute tribu \mathcal{F} possédant A et toute tribu \mathcal{B} incluse dans $\mathcal{P}(\mathbb{R})$ d'après la remarque 2.6. De même, le lecteur vérifiera aisément qu'une fonction constante $\Omega \rightarrow \mathbb{R}$ est mesurable pour toute tribu sur Ω et toute tribu sur \mathbb{R} .

En dehors de ces exemples élémentaires, il est généralement assez difficile, voire parfois impossible, de tester directement l'appartenance de $f^{-1}(B)$ à \mathcal{F}_1 pour *tout* $B \in \mathcal{F}_2$, en raison notamment de l'absence de description exhaustive des éléments de \mathcal{F}_2 quand cette tribu est assez riche (cas de la tribu Bor(\mathbb{R})). La proposition suivante nous montre qu'en fait, il suffit de restreindre ce test aux éléments B d'une famille génératrice de \mathcal{F}_2 .

Proposition 2.9. Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces mesurables et \mathcal{S} une famille de parties de Ω_2 engendrant \mathcal{F}_2 ($\sigma(\mathcal{S}) = \mathcal{F}_2$). L'application $f : \Omega_1 \rightarrow \Omega_2$ est \mathcal{F}_1 - \mathcal{F}_2 mesurable si pour tout $B \in \mathcal{S}$, $f^{-1}(B) \in \mathcal{F}_1$, autrement dit si $f^{-1}(\mathcal{S}) \subset \mathcal{F}_1$.

Démonstration. Comme \mathcal{F}_1 est une tribu sur Ω_1 , l'hypothèse $f^{-1}(\mathcal{S}) \subset \mathcal{F}_1$ implique par minimalité l'inclusion de tribus $\sigma(f^{-1}(\mathcal{S})) \subset \mathcal{F}_1$. Or nous savons par la Proposition 1.11 que $\sigma(f^{-1}(\mathcal{S})) = f^{-1}(\sigma(\mathcal{S}))$. Comme $\sigma(\mathcal{S}) = \mathcal{F}_2$, nous venons ainsi de vérifier l'inclusion $f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1$, autrement dit la mesurabilité de f . \square

Voici une liste (non exhaustive) d'applications courantes de la proposition 2.9.

Corollaire 2.10. *Soient Ω un ensemble muni d'une tribu \mathcal{F} et une application $f : \Omega \rightarrow \mathbb{K}$ où \mathbb{K} désigne l'un des ensembles $\mathbb{R}, \overline{\mathbb{R}}, \mathbb{R}_+, \overline{\mathbb{R}}_+$, ou \mathbb{R}^d . Notons $\mathcal{B} := \text{Bor}(\mathbb{K})$.*

- i) *Pour $\mathbb{K} = \mathbb{R}$, f est \mathcal{F} - \mathcal{B} mesurable si pour tous $a, b \in \mathbb{R}$, $f^{-1}(]a, b[) \in \mathcal{F}$.*
- ii) *Pour $\mathbb{K} = \mathbb{R}$, f est \mathcal{F} - \mathcal{B} mesurable si pour tout $x \in \mathbb{R}$, $f^{-1}(]-\infty, x]) \in \mathcal{F}$.*
- iii) *Pour $\mathbb{K} = \overline{\mathbb{R}}$, f est \mathcal{F} - \mathcal{B} mesurable si pour tout $x \in \mathbb{R}$, $f^{-1}(]-\infty, x]) \in \mathcal{F}$.*
- iv) *Pour $\mathbb{K} = \mathbb{R}_+$ ou $\overline{\mathbb{R}}_+$, f est \mathcal{F} - \mathcal{B} mesurable si pour tout $x \in \mathbb{R}_+$, $f^{-1}([0, x]) \in \mathcal{F}$.*
- v) *Pour $\mathbb{K} = \mathbb{R}^d$, f est \mathcal{F} - \mathcal{B} mesurable si pour tout pavé $C := \prod_{i=1}^d]a_i, b_i[$, $f^{-1}(C) \in \mathcal{F}$.*

Corollaire 2.11. *Soit (Ω, \mathcal{F}) un espace probabilisable et X une application $\Omega \rightarrow \mathbb{K}$ (où $\mathbb{K} = \mathbb{R}, \mathbb{C}$ ou \mathbb{R}^d , $d > 1$) telle que $X(\Omega)$ soit une partie au plus dénombrable de \mathbb{K} . Alors X est une variable aléatoire discrète (resp. un vecteur aléatoire discret) si :*

$$\forall x \in X(\Omega), \quad X^{-1}(\{x\}) \in \mathcal{F}. \quad (2.4)$$

Comme la tribu $\mathcal{P}(\mathbb{K})$ possède les singletons, il est clair que cette condition suffisante est aussi nécessaire. La caractérisation de la \mathcal{F} - $\mathcal{P}(\mathbb{K})$ mesurabilité de X par (2.4) avait été prise comme définition d'une variable aléatoire discrète en DEUG (voir [ICP, Déf. 3.1]).

Démonstration. Posons $\Omega_2 := X(\Omega)$, $\Omega_2^c := \mathbb{K} \setminus \Omega_2$ et notons \mathcal{S} la famille des singletons de Ω_2 . Comme Ω_2 est au plus dénombrable, $\sigma(\mathcal{S}) = \mathcal{P}(\Omega_2)$. Donc par la proposition 2.9, X considérée comme application $\Omega \rightarrow \Omega_2$ est \mathcal{F} - $\mathcal{P}(\Omega_2)$ mesurable, d'où

$$\forall A \in \mathcal{P}(\Omega_2), \quad X^{-1}(A) \in \mathcal{F}. \quad (2.5)$$

Notons qu'il n'y a pas ici d'ambiguïté sur l'écriture $X^{-1}(A)$ qui représente le même ensemble, que l'on prenne Ω_2 ou \mathbb{K} pour ensemble d'arrivée de l'application X . Soit maintenant $B \in \mathcal{P}(\mathbb{K})$ quelconque. En écrivant

$$X^{-1}(B) = X^{-1}(B \cap \Omega_2) \cup X^{-1}(B \cap \Omega_2^c),$$

en remarquant que $X^{-1}(B \cap \Omega_2^c) = \emptyset$ et en appliquant (2.5) à $A := B \cap \Omega_2$, on obtient l'appartenance à \mathcal{F} de $X^{-1}(B)$. L'application X est donc bien \mathcal{F} - $\mathcal{P}(\mathbb{K})$ mesurable. \square

Revenant à la définition 2.5, on observe son analogie formelle avec la définition de la continuité d'une application entre deux espaces topologiques : en remplaçant les tribus \mathcal{F}_1 et \mathcal{F}_2 par des *topologies* \mathcal{T}_1 et \mathcal{T}_2 , la définition deviendrait celle d'une application continue. Un cas important est celui où Ω_1 et Ω_2 sont munis de leurs tribus boréliennes $\sigma(\mathcal{T}_1)$ et $\sigma(\mathcal{T}_2)$.

Définition 2.12. Soient $(\Omega_1, \mathcal{T}_1)$ et $(\Omega_2, \mathcal{T}_2)$ deux espaces topologiques, munis de leurs tribus boréliennes respectives $\mathcal{B}_1 = \sigma(\mathcal{T}_1)$ et $\mathcal{B}_2 = \sigma(\mathcal{T}_2)$. Une application $f : \Omega_1 \rightarrow \Omega_2$ est dite borélienne si elle est \mathcal{B}_1 - \mathcal{B}_2 mesurable.

Proposition 2.13. Soient $(\Omega_1, \mathcal{T}_1)$ et $(\Omega_2, \mathcal{T}_2)$ deux espaces topologiques. Toute application continue $f : \Omega_1 \rightarrow \Omega_2$ est borélienne.

Démonstration. Soit $V \in \mathcal{T}_2$ un ouvert de Ω_2 . Par continuité de f , $f^{-1}(V)$ est un ouvert de Ω_1 , c'est-à-dire un élément de \mathcal{T}_1 . C'est donc aussi un élément de $\mathcal{B}_1 = \sigma(\mathcal{T}_1)$. Ceci étant vrai pour tout V , on a ainsi établi l'inclusion $f^{-1}(\mathcal{T}_2) \subset \mathcal{B}_1$. Par la proposition 2.9, ceci implique la \mathcal{B}_1 - \mathcal{B}_2 mesurabilité de f qui est donc bien borélienne. \square

Nous examinons maintenant l'effet sur la mesurabilité des opérations usuelles sur les fonctions.

Proposition 2.14. Soient $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2, 3$ des espaces mesurables, $f : \Omega_1 \rightarrow \Omega_2$ une application \mathcal{F}_1 - \mathcal{F}_2 mesurable et $g : \Omega_2 \rightarrow \Omega_3$, \mathcal{F}_2 - \mathcal{F}_3 mesurable. Alors $g \circ f$ est \mathcal{F}_1 - \mathcal{F}_3 mesurable.

Démonstration. Soit $C \subset \Omega_3$. En appliquant de façon répétée la définition de l'inverse ensembliste, on obtient :

$$\begin{aligned} (g \circ f)^{-1}(C) &= \{\omega \in \Omega_1; (g \circ f)(\omega) \in C\} \\ &= \{\omega \in \Omega_1; g(f(\omega)) \in C\} \\ &= \{\omega \in \Omega_1; f(\omega) \in g^{-1}(C)\} \\ &= \{\omega \in \Omega_1; \omega \in f^{-1}(g^{-1}(C))\} \\ &= f^{-1}(g^{-1}(C)). \end{aligned}$$

En prenant C quelconque dans \mathcal{F}_3 , on en déduit

$$(g \circ f)^{-1}(\mathcal{F}_3) = f^{-1}(g^{-1}(\mathcal{F}_3)).$$

Les mesurabilités respectives de g et f se traduisant par $g^{-1}(\mathcal{F}_3) \subset \mathcal{F}_2$ et $f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1$, on en déduit l'inclusion $(g \circ f)^{-1}(\mathcal{F}_3) \subset \mathcal{F}_1$ qui exprime la \mathcal{F}_1 - \mathcal{F}_3 mesurabilité de $g \circ f$. \square

Proposition 2.15. Soient (Ω, \mathcal{F}) un espace mesurable, f et g deux applications de Ω dans \mathbb{R} . L'application $h = (f, g) : \Omega \rightarrow \mathbb{R}^2$ est \mathcal{F} -Bor(\mathbb{R}^2) mesurable si et seulement si f et g sont \mathcal{F} -Bor(\mathbb{R}) mesurables.

Démonstration. Vérifions d'abord que la mesurabilité de f et g est nécessaire à celle de h . Supposons donc h mesurable. Les projections $\pi_1 : (x, y) \mapsto x$ et $\pi_2 : (x, y) \mapsto y$ étant continues sont boréliennes et il suffit d'appliquer la proposition 2.14 à $f = \pi_1 \circ h$ et $g = \pi_2 \circ h$ pour obtenir la mesurabilité de f et g .

Réciproquement, supposons f et g \mathcal{F} -Bor(\mathbb{R}) mesurables. Soit \mathcal{C} la classe des pavés $]a, b[\times]c, d[$ dans \mathbb{R}^2 . On a

$$\begin{aligned} h^{-1}(]a, b[\times]c, d[) &= \{\omega \in \Omega; (f(\omega), g(\omega)) \in]a, b[\times]c, d[\} \\ &= \{\omega \in \Omega; f(\omega) \in]a, b[\text{ et } g(\omega) \in]c, d[\} \\ &= \{\omega \in \Omega; \omega \in f^{-1}(]a, b[) \text{ et } \omega \in g^{-1}(]c, d[) \} \\ &= f^{-1}(]a, b[) \cap g^{-1}(]c, d[). \end{aligned}$$

Grâce à la mesurabilité de f et g on en déduit $h^{-1}(\mathcal{C}) \subset \mathcal{F}$. Comme $\sigma(\mathcal{C}) = \text{Bor}(\mathbb{R}^2)$, la mesurabilité de h en découle via la proposition 2.9. \square

Corollaire 2.16. *Soit (Ω, \mathcal{F}) un espace mesurable, f et g deux applications de Ω dans \mathbb{R} .*

- i) *La fonction complexe $f + ig : \Omega \rightarrow \mathbb{C}$ est \mathcal{F} -Bor(\mathbb{C}) mesurable si et seulement si f et g sont \mathcal{F} -Bor(\mathbb{R}) mesurables.*
- ii) *Si f et g sont \mathcal{F} -Bor(\mathbb{R}) mesurables, il en est de même pour les fonctions $f + g$, fg , cf ($c \in \mathbb{R}$ constante) et $|f|$. Ceci se généralise aux fonctions à valeurs complexes en remplaçant Bor(\mathbb{R}) par Bor(\mathbb{C}).*
- iii) *Si f et g sont \mathcal{F} -Bor(\mathbb{R}) mesurables, il en est de même pour les fonctions $\min(f, g)$ et $\max(f, g)$.*
- iv) *f est \mathcal{F} -Bor(\mathbb{R}) mesurable si et seulement si $f^+ = \max(f, 0)$ et $f^- = -\min(f, 0)$ le sont.*

Démonstration. Par souci de concision (on ne rit pas...), nous laissons au lecteur le soin de préciser les tribus concernées par les mesurabilités évoquées ci-dessous.

Pour i), supposons f et g mesurables, alors $(f, g) : \Omega \rightarrow \mathbb{R}^2$ l'est aussi (cf. Prop. 2.15). En composant avec l'application continue $\mathbb{R}^2 \rightarrow \mathbb{C}$, $(x, y) \mapsto x + iy$, on obtient la mesurabilité de $f + ig$ (cf. Prop. 2.13). Réciproquement, les projections π_1 et π_2 de \mathbb{C} sur respectivement l'axe réel et l'axe des imaginaires purs étant continues, la mesurabilité de $f + ig$ entraîne celles de $f = \pi_1 \circ (f + ig)$ et $g = \pi_2 \circ (f + ig)$ (cf. Prop. 2.13 et Prop. 2.14).

Pour ii), on pose $h = (f, g)$, $s : \mathbb{R}^2 \rightarrow \mathbb{R}$, $(x, y) \mapsto x + y$, $p : \mathbb{R}^2 \rightarrow \mathbb{R}$, $(x, y) \mapsto xy$, $p_c : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto cx$, $a : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto |x|$. Les applications p , s , p_c et a sont continues donc boréliennes. On utilise alors les propositions 2.14 et 2.15 en notant que $f + g = s \circ h$, $fg = p \circ h$, $cf = p_c \circ f$, $|f| = a \circ f$. Le cas complexe se ramène au cas réel après séparation des parties réelles et imaginaires des fonctions.

Pour iii), on note m et M les applications continues de \mathbb{R}^2 dans \mathbb{R} définies par $m(x, y) = \min(x, y)$ et $M(x, y) = \max(x, y)$ et on obtient les mesurabilités souhaitées par composition puisque $\min(f, g) = m \circ h$ et $\max(f, g) = M \circ h$. Rappelons à cette occasion, que si $\min(x, y)$ est toujours égal à l'un des deux nombres x ou y , la fonction $\min(f, g)$ n'est en général égale à aucune des deux fonctions f et g .

La nécessité de la mesurabilité de f^+ et f^- dans iv) découle de iii) avec $g = 0$ (et aussi de ii) avec $c = -1$). La suffisance provient de ii) en remarquant que $f = f^+ - f^-$. \square

La mesurabilité des applications est un outil commode pour prouver l'appartenance à \mathcal{F} de certains sous ensembles de Ω . Voici un exemple d'usage courant. Soient $f, g : \Omega \rightarrow \mathbb{R}$, \mathcal{F} -Bor(\mathbb{R}) mesurables. Définissons l'ensemble $\{f \leq g\}$ par

$$\{f \leq g\} := \{\omega \in \Omega; f(\omega) \leq g(\omega)\}.$$

Alors $\{f \leq g\}$ appartient à la tribu \mathcal{F} . Il y a au moins deux façons de le voir. On peut noter que $h := g - f$ est \mathcal{F} -Bor(\mathbb{R}) mesurable (corollaire 2.16 ii) et que $\{f \leq g\} = h^{-1}([0, +\infty[)$ est élément de \mathcal{F} comme image réciproque par h d'un fermé (donc

borélien) de \mathbb{R} . On peut aussi utiliser directement la proposition 2.15 en exploitant la \mathcal{F} -Bor(\mathbb{R}^2) mesurabilité de $H := (f, g)$ et en notant que $\{f \leq g\} = H^{-1}(A)$, où $A := \{(x, y) \in \mathbb{R}^2; x \leq y\}$ est fermé dans \mathbb{R}^2 . Cette digression sournoise permet à l'auteur d'amener la question : a-t-on la même conclusion si f et g sont à valeurs dans $\overline{\mathbb{R}}_+$ et sont \mathcal{F} -Bor($\overline{\mathbb{R}}_+$) mesurables ? Il est clair que la première solution basée sur $g - f$ est irrecevable puisque $g - f$ n'est pas forcément défini sur tout Ω . La deuxième solution supposerait une exploration préalable de la topologie de $\overline{\mathbb{R}}_+^2$ et de sa tribu borélienne (a-t-elle une famille génératrice formée de produits cartésiens de boréliens de $\overline{\mathbb{R}}_+$?...) en vue de prouver l'analogie de la proposition 2.15. Nous proposons une troisième voie basée sur une méthode de discrétisation pour établir le résultat. Celui-ci sera réutilisé au chapitre 3.

Proposition 2.17. *Si les applications $f, g : \Omega \rightarrow \overline{\mathbb{R}}_+$ sont \mathcal{F} -Bor($\overline{\mathbb{R}}_+$) mesurables, l'ensemble $\{f \leq g\}$ appartient à \mathcal{F} .*

Démonstration. En remarquant que si $g(\omega) = +\infty$, l'inégalité $f(\omega) \leq g(\omega)$ est automatiquement satisfaite, autrement dit que $\{g = +\infty\} \subset \{f \leq g\}$, on peut écrire

$$\{f \leq g\} = \{f \leq g \text{ et } g = +\infty\} \cup \{f \leq g < +\infty\} = \{g = +\infty\} \cup \{f \leq g < +\infty\}.$$

Comme g est \mathcal{F} -Bor($\overline{\mathbb{R}}_+$) mesurable, $g^{-1}(\{+\infty\}) \in \mathcal{F}$ et on est ramenés à établir l'appartenance à \mathcal{F} de $\{f \leq g < +\infty\}$. Pour cela, commençons par prouver l'équivalence

$$f(\omega) \leq g(\omega) < +\infty \Leftrightarrow \forall n \in \mathbb{N}, \exists k \in \mathbb{N}, \frac{k}{2^n} \leq g(\omega) < \frac{k+1}{2^n} \text{ et } f(\omega) < \frac{k+1}{2^n}. \quad (2.6)$$

L'implication directe est évidente. Pour la réciproque, le cas $n = 0$ donne la finitude de $g(\omega)$. D'autre part $k = k(n, \omega)$ est unique pour ω et n fixés. Il est clair que $u_n := 2^{-n}(k(n, \omega) + 1)$ tend vers $g(\omega)$ quand n tend vers $+\infty$ puisque $0 \leq u_n - g(\omega) \leq 2^{-n}$. En passant à la limite dans l'inégalité $f(\omega) < u_n$, on en déduit $f(\omega) \leq g(\omega)$. La traduction du second membre de l'équivalence (2.6) en opérations ensemblistes nous donne maintenant la représentation

$$\{f \leq g < +\infty\} = \bigcap_{n \in \mathbb{N}} \bigcup_{k \in \mathbb{N}} \left(g^{-1} \left(\left[\frac{k}{2^n}, \frac{k+1}{2^n} \right[\right) \cap f^{-1} \left(\left[0, \frac{k+1}{2^n} \right[\right) \right).$$

L'appartenance à \mathcal{F} de $\{f \leq g < +\infty\}$ résulte alors clairement de la mesurabilité de f et de g . \square

Dans l'étude des suites de fonctions mesurables à valeurs dans \mathbb{R} ou \mathbb{R}_+ , il est souvent commode d'agrandir l'espace d'arrivée en y rajoutant les (ou le) points à l'infini. La proposition suivante nous rassure quant à la préservation de la mesurabilité.

Proposition 2.18. *Soit $f : \Omega \rightarrow \mathbb{R}$, une application \mathcal{F} -Bor(\mathbb{R}) mesurable. Elle est aussi \mathcal{F} -Bor($\overline{\mathbb{R}}$) mesurable lorsqu'on la considère comme application $\Omega \rightarrow \overline{\mathbb{R}}$. De même si $f : \Omega \rightarrow \mathbb{R}_+$ est \mathcal{F} -Bor(\mathbb{R}_+) mesurable, elle est aussi \mathcal{F} -Bor(\mathbb{K}) mesurable comme application $\Omega \rightarrow \mathbb{K}$, pour chacun des ensembles d'arrivée $\mathbb{K} = \overline{\mathbb{R}}_+, \mathbb{R}, \overline{\mathbb{R}}$.*

Démonstration. Pour traiter d'un coup les deux situations, notons \mathbb{K}_0 l'ensemble d'arrivée « initial » ($\mathbb{K}_0 = \mathbb{R}$ ou \mathbb{R}_+), \mathbb{K} l'ensemble d'arrivée « élargi » et $\mathbb{K}_0^c := \mathbb{K} \setminus \mathbb{K}_0$.

Soit $B \in \text{Bor}(\mathbb{K})$, on peut l'écrire $B = (B \cap \mathbb{K}_0) \cup (B \cap \mathbb{K}_0^c)$, d'où

$$f^{-1}(B) = f^{-1}(B \cap \mathbb{K}_0) \cup f^{-1}(B \cap \mathbb{K}_0^c) = f^{-1}(B \cap \mathbb{K}_0),$$

puisque f étant à valeurs dans \mathbb{K}_0 , $f^{-1}(B \cap \mathbb{K}_0^c) = \emptyset$ (l'agrandissement de l'ensemble d'arrivée n'a pas changé $f(\Omega)$). Grâce aux propositions 2.2 et 2.3, on voit que $B \cap \mathbb{K}_0$ est aussi un borélien de \mathbb{K}_0 . Par l'hypothèse de mesurabilité $\mathcal{F}\text{-Bor}(\mathbb{K}_0)$ de f , on en déduit l'appartenance à \mathcal{F} de $f^{-1}(B \cap \mathbb{K}_0)$. Ainsi $f^{-1}(B) \in \mathcal{F}$ pour tout $B \in \text{Bor}(\mathbb{K})$. \square

Proposition 2.19. *Soient (Ω, \mathcal{F}) un espace mesurable et $(f_n)_{n \geq 1}$ une suite d'applications $\mathcal{F}\text{-Bor}(\overline{\mathbb{R}})$ mesurables $\Omega \rightarrow \overline{\mathbb{R}}$. Alors $h := \sup_{n \geq 1} f_n$ est $\mathcal{F}\text{-Bor}(\overline{\mathbb{R}})$ mesurable. De même, si les f_n sont à valeurs dans $\overline{\mathbb{R}}_+$ et $\mathcal{F}\text{-Bor}(\overline{\mathbb{R}}_+)$ mesurables, h est $\mathcal{F}\text{-Bor}(\overline{\mathbb{R}}_+)$ mesurable.*

En raison de la proposition 2.18, la mesurabilité $\mathcal{F}\text{-Bor}(\overline{\mathbb{K}})$ de h reste vraie si les f_n sont à valeurs dans $\mathbb{K} = \mathbb{R}$ ou \mathbb{R}_+ et $\mathcal{F}\text{-Bor}(\mathbb{K})$ mesurables³.

Démonstration. La tribu borélienne de $\overline{\mathbb{R}}$ étant engendrée par les $[-\infty, a]$, il suffit de montrer que pour tout $a \in \mathbb{R}$, $h^{-1}([-\infty, a]) \in \mathcal{F}$ (cf. Cor. 2.10 iii). On remarque d'abord que si $(u_n)_{n \geq 1}$ est une suite dans $\overline{\mathbb{R}}$, on a l'équivalence⁴

$$\sup_{n \in \mathbb{N}^*} u_n \leq a \iff \forall n \in \mathbb{N}^*, u_n \leq a.$$

Cette équivalence nous permet d'écrire :

$$\begin{aligned} h^{-1}([-\infty, a]) &= \{\omega \in \Omega; \sup_{n \geq 1} f_n(\omega) \leq a\} \\ &= \{\omega \in \Omega; \forall n \in \mathbb{N}^*, f_n(\omega) \leq a\} \\ &= \{\omega \in \Omega; \forall n \in \mathbb{N}^*, f_n(\omega) \in [-\infty, a]\} \\ &= \bigcap_{n \in \mathbb{N}^*} f_n^{-1}([-\infty, a]), \end{aligned}$$

ce qui nous donne la conclusion souhaitée grâce à la mesurabilité de chaque f_n et à la stabilité de la tribu \mathcal{F} par intersection *dénombrable*.

Le cas où les f_n sont à valeurs dans $\overline{\mathbb{R}}_+$ se traite de la même façon, en appliquant le point iv) au lieu du iii) dans le corollaire 2.10. \square

Corollaire 2.20. *Soient (Ω, \mathcal{F}) un espace mesurable et $(f_n)_{n \geq 1}$ une suite d'applications $\mathcal{F}\text{-Bor}(\overline{\mathbb{R}})$ mesurables $\Omega \rightarrow \overline{\mathbb{R}}$.*

i) Les fonctions $\sup_{n \geq 1} f_n$ et $\inf_{n \geq 1} f_n$ sont $\mathcal{F}\text{-Bor}(\overline{\mathbb{R}})$ mesurables.

³On est obligé de garder $\overline{\mathbb{K}}$ comme ensemble d'arrivée pour h car le supremum d'une suite de fonctions à valeurs finies peut très bien prendre la valeur $+\infty$.

⁴Attention, cette équivalence ne subsiste pas si l'on remplace les inégalités larges par des inégalités strictes. Dans ce cas on aurait toujours « \Rightarrow » mais plus « \Leftarrow ».

- ii) Les fonctions $\limsup_{n \rightarrow +\infty} f_n$ et $\liminf_{n \rightarrow +\infty} f_n$ sont $\mathcal{F}\text{-Bor}(\overline{\mathbb{R}})$ mesurables.
 iii) Si f est limite simple sur Ω de f_n (c'est-à-dire si $\forall \omega \in \Omega, f_n(\omega) \rightarrow f(\omega) \in \overline{\mathbb{R}}$), f est $\mathcal{F}\text{-Bor}(\overline{\mathbb{R}})$ mesurable.

Cet énoncé reste valable en remplaçant partout $\overline{\mathbb{R}}$ par $\overline{\mathbb{R}}_+$.

On peut faire ici aussi la même remarque qu'après la proposition 2.19, pour le cas où les f_n sont à valeurs dans \mathbb{R} ou \mathbb{R}_+ .

Démonstration. Pour i), il suffit de remarquer que $\inf_{n \geq 1} f_n = -\sup_{n \geq 1} (-f_n)$. Pour ii), on se ramène à i) en écrivant

$$\liminf_{n \rightarrow +\infty} f_n = \sup_{n \geq 1} \inf_{k \geq n} f_k, \quad \limsup_{n \rightarrow +\infty} f_n = \inf_{n \geq 1} \sup_{k \geq n} f_k.$$

Enfin, si f est limite simple de f_n , $f = \limsup_{n \rightarrow +\infty} f_n$, donc iii) découle de ii).

Dans le cas où les f_n sont à valeurs dans $\overline{\mathbb{R}}_+$, la seule chose à changer est la justification de la mesurabilité de $g := \inf_{n \geq 1} f_n$. La tribu borélienne de $\overline{\mathbb{R}}_+$ est engendrée par la famille $\mathcal{J}'_1 := \{[a, +\infty]; a \in \mathbb{R}_+\}$ et l'équivalence

$$\inf_{n \in \mathbb{N}^*} u_n \geq a \iff \forall n \in \mathbb{N}^*, u_n \geq a,$$

nous permet d'écrire $g^{-1}([a, +\infty]) = \bigcap_{n \in \mathbb{N}^*} f_n^{-1}([a, +\infty])$, justifiant ainsi l'appartenance à \mathcal{F} de $g^{-1}([a, +\infty])$ et la mesurabilité de g . \square

2.4 Fonctions étagées

Les fonctions étagées sont les fonctions les plus simples à partir desquelles on pourra construire l'intégrale abstraite par approximation.

Définition 2.21. On dit qu'une application $f : \Omega \rightarrow \mathbb{R}$ est une fonction étagée si l'ensemble $f(\Omega)$ de ses valeurs est fini. En notant $f(\Omega) = \{y_1, y_2, \dots, y_n\}$ (les y_i étant tous distincts) et $A_i = f^{-1}(\{y_i\})$, on a la décomposition canonique

$$f = \sum_{i=1}^n y_i \mathbf{1}_{A_i} \tag{2.7}$$

et $\{A_i; 1 \leq i \leq n\}$ est une partition finie de Ω .

Si $0 \in f(\Omega)$, il faut l'écrire dans la décomposition (2.7) pour avoir vraiment la décomposition canonique. Dans le même ordre d'idées, si une fonction g s'écrit $g = \sum_{j=1}^m c_j \mathbf{1}_{B_j}$, les c_j étant distincts et tous non nuls, les B_j étant deux à deux disjoints mais $B := \bigcup_{j=1}^m B_j \neq \Omega$, alors g est étagée et sa décomposition canonique est $g = \sum_{j=1}^{m+1} c_j \mathbf{1}_{B_j}$, où l'on a posé $c_{m+1} = 0$ et $B_{m+1} = B^c$. Enfin, si $g = \sum_{j=1}^m c_j \mathbf{1}_{B_j}$, les c_j étant distincts et tous non nuls, les B_j étant quelconques, g est encore étagée. Pour avoir sa décomposition

canonique, il faut considérer la partition de Ω engendrée par la famille B_1, \dots, B_m, B_{m+1} (cette partition est formée de celles des intersections $D = C_1 \cap C_2 \cap \dots \cap C_{m+1}$ qui sont non vides, C_j désignant l'un des deux ensembles B_j ou B_j^c) et calculer la valeur constante de g sur chacun des ensembles D et éventuellement réunir les D donnant la même valeur pour g .

Proposition 2.22. *Soient (Ω, \mathcal{F}) un espace mesurable et $f : \Omega \rightarrow \mathbb{R}$ une fonction étagée, de décomposition canonique $f = \sum_{i=1}^n y_i \mathbf{1}_{A_i}$. Alors f est \mathcal{F} -Bor(\mathbb{R}) mesurable si et seulement si tous les A_i sont dans \mathcal{F} .*

Démonstration. Si tous les A_i sont dans \mathcal{F} , f est mesurable comme combinaison linéaire d'indicatrices d'éléments de \mathcal{F} (cf. p. 50 et Cor. 2.16 ii)).

Réciproquement, si la fonction étagée f est mesurable, comme la tribu borélienne de \mathbb{R} contient les singletons, chaque $A_i = f^{-1}(\{y_i\})$ est dans \mathcal{F} . \square

Si $f = \sum_{i=1}^n y_i \mathbf{1}_{A_i}$, les A_i étant tous dans \mathcal{F} et les y_i des réels positifs, f est mesurable \mathcal{F} -Bor(\mathbb{R}_+). Par la proposition 2.18, elle est aussi mesurable \mathcal{F} -Bor(\mathbb{K}) pour $\mathbb{K} = \mathbb{R}_+, \mathbb{R}, \overline{\mathbb{R}}$. Nous appellerons dans la suite « fonction étagée mesurable positive » une telle fonction f sans préciser l'ensemble d'arrivée ni la tribu borélienne considérés.

Théorème 2.23. *Soit (Ω, \mathcal{F}) un espace mesurable. Toute fonction $f : \Omega \rightarrow \overline{\mathbb{R}}_+$ et mesurable \mathcal{F} -Bor($\overline{\mathbb{R}}_+$) est limite simple sur Ω d'une suite croissante $(f_n)_{n \geq 1}$ de fonctions étagées mesurables positives.*

Démonstration. L'idée est d'utiliser pour construire f_n , les valeurs approchées dyadiques par défaut de f au niveau de résolution n . On peut procéder comme suit en définissant pour $n \in \mathbb{N}$, les ensembles

$$\begin{aligned} A_{n,k} &:= f^{-1}([k2^{-n}, (k+1)2^{-n}[), \quad 0 \leq k \leq n2^n - 1; \\ A_{n,n2^n} &:= f^{-1}([n, +\infty[). \end{aligned}$$

On prend alors pour f_n la fonction étagée positive de décomposition canonique

$$f_n = \sum_{k=0}^{n2^n} \frac{k}{2^n} \mathbf{1}_{A_{n,k}}.$$

Comme f est mesurable, les $A_{n,k}$ sont dans \mathcal{F} , ce qui entraîne la mesurabilité de f_n (Prop. 2.22).

Il reste à vérifier que pour tout $\omega \in \Omega$, la suite de réels $(f_n(\omega))_{n \geq 1}$ est croissante et converge dans $\overline{\mathbb{R}}$ vers $f(\omega)$.

Réglons d'abord le cas où $f(\omega) = +\infty$. Alors pour tout n , $f_n(\omega) = n$, ce qui nous donne bien une suite croissante convergente vers $+\infty = f(\omega)$.

Si $f(\omega) < +\infty$, $f_n(\omega) = n$ pour $n \leq [f(\omega)]$ (on note $[x]$ la partie entière de x , unique entier m tel que $m \leq x < m+1$) et pour $n > [f(\omega)]$,

$$f_n(\omega) = \frac{k(n, \omega)}{2^n} = \max \left\{ \frac{l}{2^n}; \frac{l}{2^n} \leq f(\omega), l \in \mathbb{N} \right\}. \quad (2.8)$$

La suite finie $(f_n(\omega))_{n \leq [f(\omega)]}$ est clairement croissante. Regardons la suite $(f_n(\omega))_{n > [f(\omega)]}$. D'après (2.8), on a

$$f_n(\omega) = \frac{k(n, \omega)}{2^n} = \frac{2k(n, \omega)}{2^{n+1}} \leq f(\omega) \quad \Rightarrow \quad \frac{2k(n, \omega)}{2^{n+1}} \leq \frac{k(n+1, \omega)}{2^{n+1}} = f_{n+1}(\omega),$$

d'où la croissance de la suite $(f_n(\omega))_{n > [f(\omega)]}$. Pour établir définitivement la croissance de toute la suite $(f_n(\omega))_{n \geq 1}$, il ne reste plus qu'à examiner le point de raccord des deux sous-suites, donc à comparer $f_n(\omega)$ et $f_{n+1}(\omega)$ pour $n = [f(\omega)]$. Il suffit de remarquer que $f_n(\omega) = n = (n2^{n+1})2^{-n-1} \leq f(\omega)$ et comme $f_{n+1}(\omega)$ est donné par (2.8), on a $(n2^{n+1})2^{-n-1} \leq k(n+1, \omega)2^{-n-1} = f_{n+1}(\omega)$.

La convergence est immédiate, puisque pour $n > [f(\omega)]$, on a d'après (2.8)

$$f_n(\omega) \leq f(\omega) < f_n(\omega) + \frac{1}{2^n},$$

d'où $0 \leq f(\omega) - f_n(\omega) < 2^{-n}$. □

Remarque : La convergence est *uniforme* sur Ω si f est bornée (i.e. $M := \sup_{\Omega} f < +\infty$) car pour $n > M$, on a pour tout $\omega \in \Omega$, $0 \leq f(\omega) - f_n(\omega) < 2^{-n}$.

Pour illustrer la convergence de f_n vers f , on a choisi ci-dessous $\Omega = [0, 4]$, $f(x) = (2 + \sin(x^2))\sqrt{|x-1|}$ et représenté les fonctions étagées f_1 , f_2 et f_3 (figures 2.2 à 2.4).

Corollaire 2.24. *Soit (Ω, \mathcal{F}) un espace mesurable. Une fonction $g : \Omega \rightarrow \overline{\mathbb{R}}$ est \mathcal{F} -Bor($\overline{\mathbb{R}}$) mesurable si et seulement si elle est limite simple sur Ω d'une suite de fonctions étagées \mathcal{F} -Bor(\mathbb{R}) mesurables.*

Démonstration. On sait déjà qu'une limite simple de fonctions mesurables est mesurable (Cor. 2.20 iii). Dans l'autre sens, on écrit $g = g^+ - g^-$ et on utilise le théorème 2.23 en remarquant que pour tout ω au plus une seule des deux valeurs $g^+(\omega)$ et $g^-(\omega)$ est non nulle et que si $f(\omega) = 0$ dans le Th. 2.23, alors tous les $f_n(\omega)$ sont nuls. Ceci permet de découper Ω en trois parties disjointes $\{g = 0\}$, $\{g > 0\}$ et $\{g < 0\}$ sur chacune desquelles on applique le théorème 2.23. □

2.5 Mesures images, lois

Les applications mesurables permettent de « transporter » la mesure d'un espace à un autre. Dans le cas d'une variable (ou d'un vecteur) aléatoire, la « mesure image » ainsi obtenue s'appelle la loi de la variable (du vecteur) aléatoire.

2.5.1 Généralités

Proposition 2.25. *Soient $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$ des espaces mesurables, μ une mesure sur $(\Omega_1, \mathcal{F}_1)$ et $f : \Omega_1 \rightarrow \Omega_2$ une application \mathcal{F}_1 - \mathcal{F}_2 mesurable. La fonction d'ensembles $\mu \circ f^{-1}$ définie sur \mathcal{F}_2 par*

$$\forall B \in \mathcal{F}_2, \quad (\mu \circ f^{-1})(B) := \mu(f^{-1}(B))$$

est une mesure sur $(\Omega_2, \mathcal{F}_2)$. On l'appelle mesure image de μ par f .

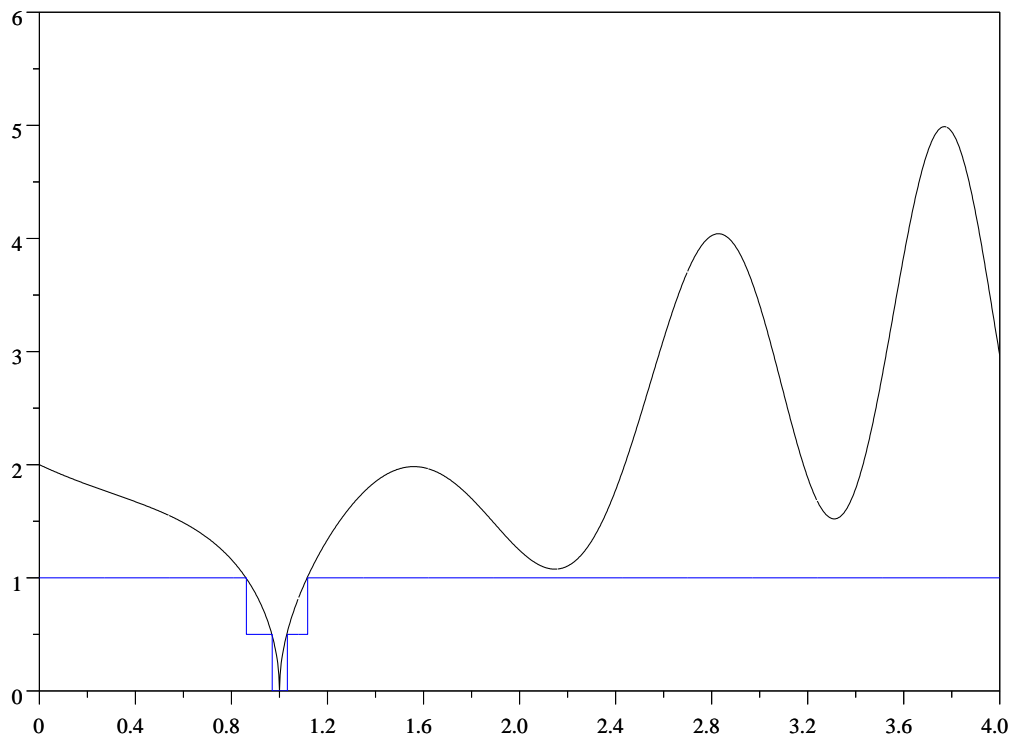


FIG. 2.2 – f et f_1

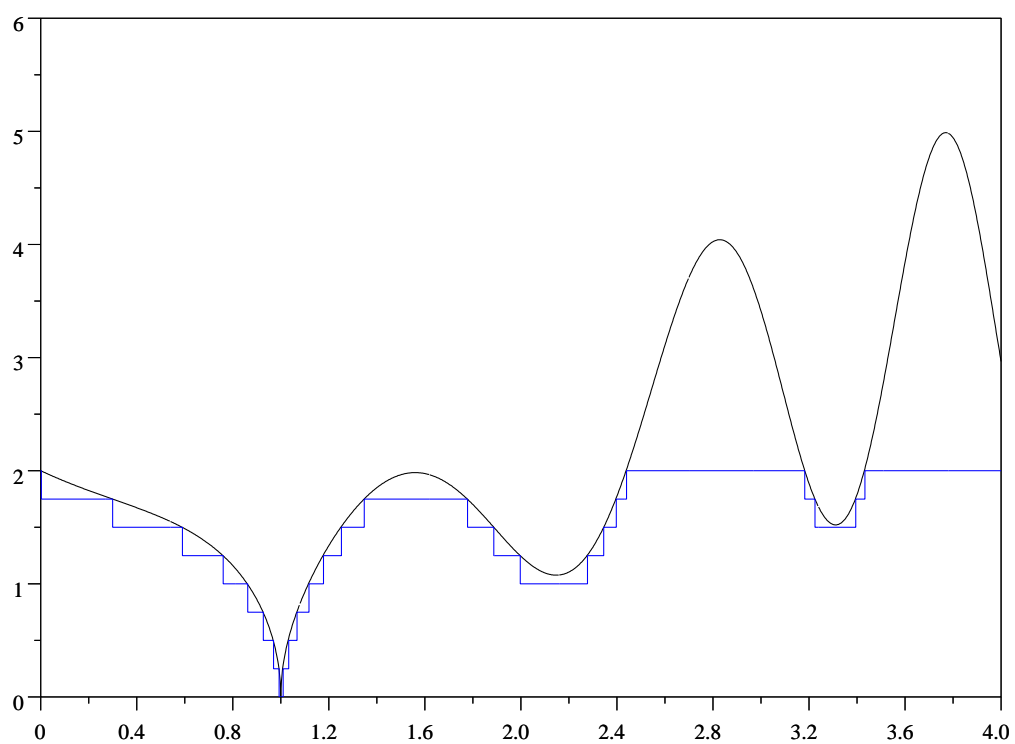


FIG. 2.3 – f et f_2

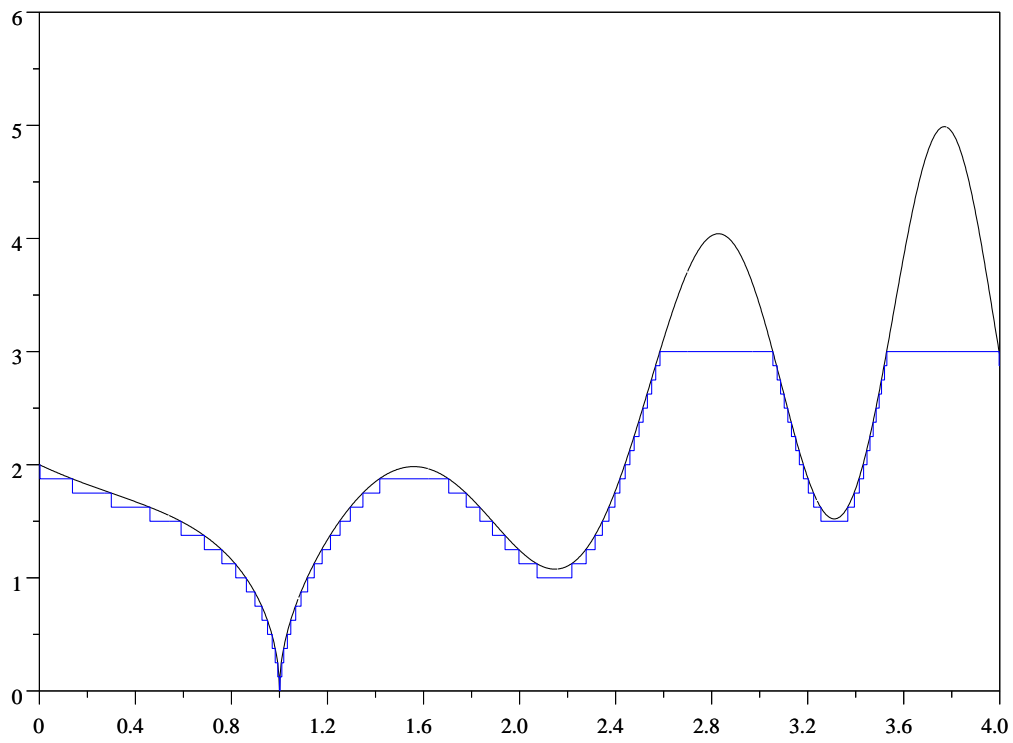


FIG. 2.4 – f et f_3

Démonstration. La fonction d'ensembles $\nu = \mu \circ f^{-1}$ est bien définie sur \mathcal{F}_2 puisque pour $B \in \mathcal{F}_2$, $f^{-1}(B) \in \mathcal{F}_1$ par mesurabilité de f et μ est définie sur \mathcal{F}_1 . Comme $f^{-1}(\emptyset) = \emptyset$, on a immédiatement $\nu(\emptyset) = \mu(\emptyset) = 0$. Pour la σ -additivité de ν , on note que l'image inverse ensembliste f^{-1} commute avec les intersections et les réunions. Soit $(B_k)_{k \in \mathbb{N}}$ une suite d'ensembles deux à deux disjoints, membres de \mathcal{F}_2 . Pour $j \neq k$, $f^{-1}(B_j) \cap f^{-1}(B_k) = f^{-1}(B_j \cap B_k) = f^{-1}(\emptyset) = \emptyset$, donc les $f^{-1}(B_k)$ sont deux à deux disjoints et dans \mathcal{F}_1 . La σ -additivité de ν résulte alors de celle de μ en écrivant

$$\nu\left(\bigcup_{k \in \mathbb{N}} B_k\right) = \mu\left(f^{-1}\left(\bigcup_{k \in \mathbb{N}} B_k\right)\right) = \mu\left(\bigcup_{k \in \mathbb{N}} f^{-1}(B_k)\right) = \sum_{k \in \mathbb{N}} \mu(f^{-1}(B_k)) = \sum_{k \in \mathbb{N}} \nu(B_k).$$

On a donc vérifié que ν est une mesure sur $(\Omega_2, \mathcal{F}_2)$. Remarquons au passage que $\nu(\Omega_2) = \mu(\Omega_1)$ car $f^{-1}(\Omega_2) = \Omega_1$. \square

La notion de mesure image est particulièrement utile en théorie des probabilités puisqu'elle permet de transporter la probabilité d'un espace probabilisé abstrait sur un espace plus familier, comme \mathbb{R}^d ou l'une de ses parties.

Définition 2.26. Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. La mesure image $P_X := \mathbf{P} \circ X^{-1}$ qui est une probabilité sur $\text{Bor}(\mathbb{R}^d)$, est appelée loi du vecteur aléatoire X sous \mathbf{P} (de la variable aléatoire X si $d = 1$).

La loi de X sous \mathbf{P} est donc la mesure de probabilité définie par

$$\forall B \in \text{Bor}(\mathbb{R}^d), \quad P_X(B) = \mathbf{P}(X^{-1}(B)) = \mathbf{P}(\{\omega \in \Omega; X(\omega) \in B\}).$$

Il est courant de simplifier cette écriture en

$$P_X(B) = \mathbf{P}(X \in B).$$

L'usage est aussi d'abrégé « loi de X sous \mathbf{P} » en « loi de X » quand il n'y a pas d'ambiguïté, c'est-à-dire quand on n'envisage qu'une seule mesure de probabilité sur (Ω, \mathcal{F}) . Remarquons que la définition de X ne présuppose pas l'existence de \mathbf{P} . Remarquons aussi que deux variables aléatoires définies sur le même espace peuvent avoir même loi sans être égales, voire même avoir même loi sans être définies sur le même espace probabilisé...

Notons enfin que les situations où l'on munit (Ω, \mathcal{F}) de plusieurs mesures de probabilité (et où il peut être utile de préciser sous quelle mesure on considère la loi de X) n'ont rien d'exceptionnel. On peut en mentionner deux d'importance fondamentale : le conditionnement et le modèle statistique (voir les sous-sections 2.5.4 et 2.5.5 ci-après).

Proposition 2.27 (Loi d'une variable aléatoire discrète). Soient $(\Omega, \mathcal{F}, \mathbf{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{K}$ une variable aléatoire discrète ou un vecteur aléatoire discret (au sens de la définition 2.8, avec $\mathbb{K} = \mathbb{R}, \mathbb{C}$ ou \mathbb{R}^d , $d > 1$). La loi de X sous \mathbf{P} est la mesure P_X sur $\mathcal{P}(\mathbb{K})$, donc aussi par restriction sur $\text{Bor}(\mathbb{K})$, donnée par :

$$P_X = \sum_{x \in X(\Omega)} \mathbf{P}(X = x) \delta_x, \tag{2.9}$$

où δ_x désigne la mesure de Dirac au point x . Rappelons qu'ici $X(\Omega)$ est au plus dénombrable, le second membre de (2.9) est donc une somme finie ou une série de mesures finies.

Démonstration. Nous savons déjà (cf. preuve du corollaire 2.11) que pour tout $B \in \mathcal{P}(\mathbb{K})$, $X^{-1}(B) = X^{-1}(B \cap X(\Omega))$. Par conséquent, comme $P_X = \mathbf{P} \circ X^{-1}$,

$$\forall B \in \mathcal{P}(\mathbb{K}), \quad P_X(B) = P_X(B \cap X(\Omega)). \quad (2.10)$$

L'ensemble $B \cap X(\Omega)$ est au plus dénombrable, donc union finie ou dénombrable de ses singletons. Par additivité ou σ -additivité de la mesure P_X , on en déduit :

$$\begin{aligned} P_X(B \cap X(\Omega)) &= \sum_{x \in B \cap X(\Omega)} P_X(\{x\}) \\ &= \sum_{x \in B \cap X(\Omega)} \mathbf{P}(X = x) \\ &= \sum_{x \in X(\Omega)} \mathbf{P}(X = x) \mathbf{1}_B(x) \\ &= \sum_{x \in X(\Omega)} \mathbf{P}(X = x) \delta_x(B). \end{aligned}$$

Compte tenu de (2.10), nous venons ainsi de vérifier que :

$$\forall B \in \mathcal{P}(\mathbb{K}), \quad P_X(B) = \sum_{x \in X(\Omega)} \mathbf{P}(X = x) \delta_x(B),$$

ce qui est précisément la traduction de (2.9). \square

2.5.2 Lois discrètes classiques

Nous revoyons ici les principales lois discrètes classiques, déjà étudiées en DEUG. Pour les détails, voir [ICP]. La notion d'indépendance des événements ou des variables aléatoires utilisée dans cette présentation a été définie en DEUG. Elle sera généralisée et étudiée systématiquement au chapitre sur les mesures produits. Lorsque la loi de X est une mesure connue μ , ce qui est le cas pour tous les exemples qui suivent, il est d'usage de dire que X « suit » la loi μ ou encore que X « obéit » à la loi μ .

Lois de Bernoulli

Définition 2.28. *La variable aléatoire X suit la loi de Bernoulli de paramètre p ($p \in [0, 1]$) si $X(\Omega) = \{0, 1\}$ et*

$$\mathbf{P}(X = 1) = p, \quad \mathbf{P}(X = 0) = 1 - p = q.$$

On notera $X \sim \text{Bern}(p)$.

Si A est un événement de probabilité p , son indicatrice $\mathbf{1}_A$ est une variable aléatoire suivant la loi de Bernoulli de paramètre p . Réciproquement, si X est une v.a. de Bernoulli, on peut toujours écrire $X = \mathbf{1}_A$ en définissant $A := \{\omega \in \Omega, X(\omega) = 1\}$.

Loi uniforme sur un ensemble fini de réels

Définition 2.29. La variable aléatoire X suit la loi uniforme sur l'ensemble fini de réels $A = \{x_1, \dots, x_n\}$ si $X(\Omega) = A$ et si P_X est l'équiprobabilité sur cet ensemble :

$$\forall k = 1, \dots, n, \quad \mathbf{P}(X = x_k) = \frac{1}{n}.$$

Par exemple le nombre de points obtenus lors du jet d'un dé équilibré suit la loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$. Dire que P_X est la loi uniforme sur $A := \{x_1, \dots, x_n\}$ équivaut à

$$\forall B \in \mathcal{P}(\mathbb{R}), \quad \mathbf{P}(X \in B) = \frac{\text{card}(A \cap B)}{\text{card}(A)}. \quad (2.11)$$

La loi P_X apparaît ainsi comme le conditionnement⁵ par A de la mesure de comptage ν sur \mathbb{R} définie par : $\nu(B) = \text{card } B$ si B est une partie finie de \mathbb{R} et $\nu(B) = +\infty$ si B est une partie infinie de \mathbb{R} .

Lois binomiales

Définition 2.30. La variable aléatoire X suit la loi binomiale de paramètres n et p ($n \in \mathbb{N}^*$ et $p \in [0, 1]$) si l'ensemble des valeurs possibles est $X(\Omega) = \{0, 1, \dots, n\}$ et

$$\forall k = 0, 1, \dots, n, \quad \mathbf{P}(X = k) = C_n^k p^k (1-p)^{n-k}.$$

Notation : $X \sim \text{Bin}(n, p)$.

La loi binomiale $\text{Bin}(n, p)$ est la loi du nombre de succès obtenus en une suite de n épreuves répétées indépendantes avec pour chaque épreuve une probabilité de succès p .

Plus généralement, soit A_1, \dots, A_n une famille de n événements mutuellement indépendants ayant tous même probabilité p et notons $X_i = \mathbf{1}_{A_i}$. Alors la variable aléatoire $S_n = \sum_{i=1}^n X_i$ suit la loi binomiale $\text{Bin}(n, p)$.

Lois hypergéométriques

Alors que la loi binomiale intervient dans les tirages avec remise, la loi hypergéométrique correspond aux tirages sans remise.

Exemple 2.1. Dans une production totale de N objets dont M sont défectueux, on prélève au hasard un échantillon de n objets (tirage sans remise). Soit X le nombre aléatoire d'objets défectueux dans l'échantillon. Quelle est sa loi ?

En considérant tous les échantillons possibles comme équiprobables, un peu de dénombrement mène à la formule suivante :

$$\mathbf{P}(X = k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} \quad \text{si} \quad \begin{cases} 0 \leq k \leq M, \\ 0 \leq n - k \leq N - M. \end{cases} \quad (2.12)$$

⁵Au sens de l'exemple 1.4 du chapitre précédent.

Définition 2.31. La loi définie par (2.12) s'appelle loi hypergéométrique de paramètres N , M et n . Notation : $X \sim \text{Hypg}(N, M, n)$. Le paramètre N est l'effectif de la population totale, M celui de la sous-population à laquelle on s'intéresse et n la taille de l'échantillon observé.

Pour une taille d'échantillon n fixée, plus N et M sont grands, moins les tirages sans remise diffèrent des tirages avec remise. Plus précisément, la loi hypergéométrique converge vers la loi binomiale au sens suivant.

Théorème 2.32. On suppose que quand N tend vers $+\infty$, $M = M(N)$ tend vers $+\infty$ en vérifiant la condition :

$$\lim_{N \rightarrow +\infty} \frac{M}{N} = p \quad \text{avec} \quad 0 < p < 1. \quad (2.13)$$

Alors, n restant fixé, la loi hypergéométrique $\text{Hypg}(N, M, n)$ converge vers la loi binomiale $\text{Bin}(n, p)$, ce qui signifie ici que si $(X_N)_{N \geq 1}$ est une suite de v.a. avec $X_N \sim \text{Hypg}(N, M, n)$ et Y est une v.a. de loi $\text{Bin}(n, p)$, alors :

$$\forall k = 0, 1, \dots, n, \quad \lim_{N \rightarrow +\infty} \mathbf{P}(X_N = k) = \mathbf{P}(Y = k), \quad (2.14)$$

autrement dit :

$$\forall k = 0, 1, \dots, n, \quad \lim_{N \rightarrow +\infty} \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}. \quad (2.15)$$

Pour la preuve, voir [ICP].

Lois multinomiales

Définition 2.33. Le vecteur aléatoire $X : \Omega \rightarrow \mathbb{R}^d$ suit la loi multinomiale de paramètres n et (p_1, \dots, p_d) où $n \in \mathbb{N}^*$ et les p_i sont strictement positifs et de somme 1 si $X(\Omega)$ est l'ensemble des d -uplets (j_1, j_2, \dots, j_d) d'entiers tels que $j_1 + j_2 + \dots + j_d = n$ et si

$$\forall (j_1, j_2, \dots, j_d) \in X(\Omega), \quad \mathbf{P}\{X = (j_1, j_2, \dots, j_d)\} = \frac{n!}{j_1! j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d}.$$

Notation : $X \sim \text{Mult}(n; p_1, \dots, p_d)$

Rappelons que la loi multinomiale sert à modéliser le total des résultats de chaque type observés dans une suite d'épreuves répétées indépendantes ayant chacune d types de résultats possibles. Par exemple si on lance 200 fois un dé, on obtient un vecteur de dimension 6 dont la i -ème composante est le nombre total d'apparitions de la face numéro i au cours des 200 lancers. Ce vecteur suit la loi multinomiale de paramètres 200 et $(p_1, p_2, p_3, p_4, p_5, p_6)$, où p_i est la probabilité d'apparition de la face n° i lors d'un lancer.

Lois géométriques

Définition 2.34. Une variable aléatoire X suit la loi géométrique de paramètre $p \in]0, 1[$, si $X(\Omega) = \mathbb{N}^*$ et :

$$\forall k \in \mathbb{N}^*, \quad \mathbf{P}(X = k) = (1 - p)^{k-1}p.$$

Notation : $X \sim \text{Geom}(p)$.

La situation typique où intervient la loi géométrique est celle du « temps d'attente du premier succès » dans une suite infinie d'épreuves répétées indépendantes avec même probabilité de succès $p \in]0, 1[$. Si X désigne le numéro (aléatoire) de la première épreuve où l'on obtient un succès, on vérifie facilement que $\mathbf{P}(X = k) = (1 - p)^{k-1}p$ pour tout $k \in \mathbb{N}^*$. En toute rigueur X est à valeurs dans $\overline{\mathbb{N}^*}$ en attribuant à X la valeur $+\infty$ lorsqu'aucune épreuve de la suite ne donne un succès. On voit facilement que $\mathbf{P}(X = +\infty) = 0$, ce qui permet de considérer X comme une variable à valeurs dans \mathbb{N}^* (en modifiant Ω et \mathcal{F} , voir à ce sujet le corrigé du Problème de l'examen de septembre 2003).

Lorsque X suit une loi géométrique, les probabilités $P(X > n)$ ont une expression particulièrement simple en fonction de $q = 1 - p$.

$$\mathbf{P}(X > n) = q^n.$$

Cette formule permet de vérifier facilement la propriété d'« absence de mémoire en temps discret » :

Proposition 2.35. Si X suit la loi géométrique de paramètre p ,

$$\forall n, k \in \mathbb{N}^*, \quad \mathbf{P}(X > n + k \mid X > n) = \mathbf{P}(X > k). \quad (2.16)$$

La preuve est laissée en exercice, de même que la réciproque : si une variable aléatoire X à valeurs dans \mathbb{N}^* vérifie (2.16), elle suit une loi géométrique.

Lois de Poisson

Définition 2.36. On dit que la variable aléatoire discrète X suit la loi de Poisson de paramètre $\alpha > 0$ si $X(\Omega) = \mathbb{N}$ et

$$\forall k \in \mathbb{N}, \quad \mathbf{P}(X = k) = \frac{e^{-\alpha} \alpha^k}{k!}.$$

Notation : $X \sim \text{Pois}(\alpha)$.

Une des raisons de l'importance de cette loi est le théorème de convergence de la loi binomiale vers la loi de Poisson.

Théorème 2.37. Si $(p_n)_{n \geq 1}$ est une suite de réels de $[0, 1]$ vérifiant

$$np_n \rightarrow \alpha \in]0, +\infty[, \quad \text{quand } n \rightarrow +\infty, \quad (2.17)$$

alors :

$$\forall k \in \mathbb{N}, \quad C_n^k p_n^k (1 - p_n)^{n-k} \longrightarrow \frac{e^{-\alpha} \alpha^k}{k!}, \quad \text{quand } n \rightarrow +\infty.$$

2.5.3 Lois uniformes

Sur un ensemble infini A , il n'existe pas d'équiprobabilité P . Sinon on pourrait trouver une suite infinie $(x_k)_{k \in \mathbb{N}}$ d'éléments de A tous distincts tels que pour tout $k \in \mathbb{N}$, $P(\{x_k\}) = p > 0$ avec p indépendant de k , d'où pour tout $n \geq 1$, $P(\{x_1, x_2, \dots, x_n\}) = np \leq P(A)$ et en faisant tendre n vers $+\infty$, on trouverait $P(A) = +\infty$, ce qui est absurde. On peut néanmoins définir la loi uniforme sur une partie (borélienne) infinie de \mathbb{R} ou \mathbb{R}^d de façon analogue à (2.11), en remplaçant la mesure de comptage par la mesure de Lebesgue.

Définition 2.38. Soit A un borélien de \mathbb{R}^d tel que $0 < \lambda_d(A) < +\infty$, λ_d désignant la mesure de Lebesgue sur \mathbb{R}^d . Le vecteur aléatoire $X : \Omega \rightarrow \mathbb{R}^d$ suit la loi uniforme sur A si

$$\forall B \in \text{Bor}(\mathbb{R}^d), \quad \mathbf{P}(X \in B) = \frac{\lambda_d(A \cap B)}{\lambda_d(A)}. \quad (2.18)$$

Notation : $X \sim \text{Unif}(A)$.

Le cas $d = 1$ et $A = [0, 1]$ revêt une importance particulière. En effet on peut démontrer que si Y est une variable aléatoire de fonction de répartition F et si $X \sim \text{Unif}([0, 1])$, alors $F^{-1}(X)$ a même loi que Y . Ici F^{-1} désigne l'inverse généralisé de F défini par

$$\forall u \in]0, 1[, \quad F^{-1}(u) = \inf\{t \in \mathbb{R}; F(t) \geq u\}.$$

Ceci permet de *simuler* n'importe quelle loi sur \mathbb{R} dès que l'on dispose d'un générateur de nombres aléatoires suivant la loi uniforme sur $[0, 1]$. Cette propriété sera vue en T.D.

2.5.4 Lois conditionnelles

Rappelons que si $(\Omega, \mathcal{F}, \mathbf{P})$ est un espace probabilisé et $H \in \mathcal{F}$ un évènement tel que $\mathbf{P}(H) > 0$, on peut définir sur \mathcal{F} une nouvelle *mesure* de probabilité $\mathbf{P}_H = \mathbf{P}(\cdot | H)$ par

$$\forall B \in \mathcal{F}, \quad \mathbf{P}_H(B) := \mathbf{P}(B | H) = \frac{\mathbf{P}(B \cap H)}{\mathbf{P}(H)}.$$

Définition 2.39. Soient $(\Omega, \mathcal{F}, \mathbf{P})$ un espace probabilisé, $H \in \mathcal{F}$ tel que $\mathbf{P}(H) > 0$, X un vecteur aléatoire à valeurs dans \mathbb{R}^d et défini sur (Ω, \mathcal{F}) . On appelle loi conditionnelle de X sachant H , la loi de X sous \mathbf{P}_H . En la notant $P_{X|H}$, on a donc

$$\forall B \in \text{Bor}(\mathbb{R}^d), \quad P_{X|H}(B) = \mathbf{P}_H(X^{-1}(B)) = \mathbf{P}(X \in B | H).$$

Il importe de ne pas se laisser induire en erreur par la notation $P_{X|H}$, elle ne concerne pas une nouvelle variable aléatoire « $X | H$ » mais bien toujours *la même* variable aléatoire X . Ce qui a changé, c'est la mesure dont on munit (Ω, \mathcal{F}) et sous laquelle on considère la loi de X .

Voici un exemple de calcul de loi conditionnelle particulièrement simple. Soit A un borélien de \mathbb{R}^d tel que $0 < \lambda_d(A) < +\infty$ et X un vecteur aléatoire de loi uniforme sur A .

Soit C un borélien inclus dans A et tel que $\lambda_d(C) > 0$. Alors la loi de X sachant $X \in C$ est la loi uniforme sur C . En effet en rappelant (2.18) et en notant que $A \cap C = C$, on peut écrire pour tout $B \in \text{Bor}(\mathbb{R}^d)$:

$$\begin{aligned} P_{X|X \in C}(B) &= \mathbf{P}(X \in B \mid X \in C) = \frac{\mathbf{P}(X \in B \text{ et } X \in C)}{\mathbf{P}(X \in C)} \\ &= \frac{\mathbf{P}(X \in B \cap C)}{\mathbf{P}(X \in C)} \\ &= \frac{\lambda_d(A \cap B \cap C)/\lambda_d(A)}{\lambda_d(A \cap C)/\lambda_d(A)} \\ &= \frac{\lambda_d(B \cap C)}{\lambda_d(C)}. \end{aligned}$$

2.5.5 Un premier exemple de modèle statistique

Terminons avec un exemple simple de modèle statistique où coexistent naturellement plusieurs lois pour *une même* variable aléatoire. On dispose d'une urne contenant des boules rouges et des boules vertes. La proportion θ de boules vertes est inconnue. On se propose de l'estimer en effectuant n tirages avec remise d'une boule et en notant sa couleur. C'est un problème de *sondage*.

On peut utiliser la modélisation suivante. Prenons $\Omega_n = \{r, v\}^n$. Un événement élémentaire ω est donc ici une suite finie de n caractères r ou v codant les résultats des n tirages. Comme Ω_n est fini, on le munit de la tribu $\mathcal{F}_n = \mathcal{P}(\Omega_n)$. Avec cette tribu, n'importe quelle application de Ω_n dans \mathbb{R} est mesurable, donc est une variable aléatoire. Définissons maintenant les n variables aléatoires X_i ($i = 1, \dots, n$) par

$$X_i(\omega) = \begin{cases} 0 & \text{si la } i\text{-ème composante de } \omega \text{ est } r, \\ 1 & \text{si la } i\text{-ème composante de } \omega \text{ est } v. \end{cases}$$

Définissons aussi

$$S_n := \sum_{i=1}^n X_i, \quad M_n := \frac{1}{n} S_n.$$

L'introduction de ces deux variables est bien naturelle. En effet $S_n(\omega)$ est le nombre de caractères v dans l'événement élémentaire ω , autrement dit le nombre de boules vertes observées dans la suite de n tirages codée par ω . De même $M_n(\omega)$ est la proportion de boules vertes dans cette même suite de tirages.

Pour l'instant nous avons des variables aléatoires discrètes définies sur $(\Omega_n, \mathcal{F}_n)$, mais pas encore de mesure de probabilité sur cet espace. On ne peut donc pas encore parler de loi pour ces variables. Comme nous avons choisi un Ω_n ne dépendant pas de θ , il est clair que la probabilité dont on va munir $(\Omega_n, \mathcal{F}_n)$ doit dépendre de θ . Malheureusement nous ignorons la valeur de θ , il faut donc se résigner à payer cette ignorance en munissant $(\Omega_n, \mathcal{F}_n)$ non pas d'une mesure de probabilité \mathbf{P}_θ , mais de *toute une famille* $(\mathbf{P}_\theta)_{\theta \in \Theta}$. La famille d'espaces probabilisés $(\Omega_n, \mathcal{F}_n, (\mathbf{P}_\theta)_{\theta \in \Theta})$ s'appelle un *modèle*

statistique. L'ensemble Θ est l'espace des paramètres du modèle. Dans la situation qui nous intéresse, θ est une proportion, donc un nombre rationnel. On pourra donc prendre $\Theta = \mathbb{Q} \cap [0, 1]$ si on ignore le nombre total N de boules dans l'urne ou se restreindre à $\Theta = \{0, 1/N, 2/N, \dots, 1\}$ si on connaît la valeur de N .

Comment définir maintenant \mathbf{P}_θ ? Notons d'abord que Ω_n étant fini et muni de $\mathcal{F}_n = \mathcal{P}(\Omega_n)$, \mathbf{P}_θ sera caractérisée par les $\mathbf{P}_\theta(\{\omega\})$, pour ω décrivant Ω . Arrivés à ce point, nous allons voir qu'il n'y a plus qu'un seul choix admissible pour \mathbf{P}_θ . Rappelons en effet que les tirages sont avec remise, donc la composition de l'urne est la même avant chaque nouveau tirage. La probabilité de sortir une boule verte lors du i -ème tirage doit donc être θ . Autrement dit la loi de X_i sous \mathbf{P}_θ doit être la loi de Bernoulli de paramètre θ . D'autre part les résultats des tirages passés n'influencent pas le tirage à venir puisque l'urne est toujours dans la même composition avant chaque tirage. On modélise ceci par l'indépendance des tirages, autrement dit sur l'espace $(\Omega_n, \mathcal{F}_n, \mathbf{P}_\theta)$, les X_i doivent être des variables de Bernoulli de même loi et indépendantes. Ceci nous conduit à poser :

$$\forall \omega \in \Omega_n, \quad \mathbf{P}_\theta(\{\omega\}) := \theta^{S_n(\omega)}(1 - \theta)^{n - S_n(\omega)}.$$

Il est alors facile de voir que sous \mathbf{P}_θ , S_n suit la loi binomiale $\text{Bin}(n, \theta)$. Quant à M_n , elle suit une loi qui a les mêmes masses que $\text{Bin}(n, \theta)$, mais localisées sur les rationnels $0, 1/n, 2/n, \dots, 1$, au lieu des entiers $0, 1, 2, \dots, n$. En notant cette loi P_{θ, M_n} , on a

$$P_{\theta, M_n} = \sum_{k=0}^n C_n^k \theta^k (1 - \theta)^{n-k} \delta_{k/n}.$$

Quand n est grand, cette loi est bien concentrée dans un petit voisinage de θ , voir les diagrammes en bâtons⁶ de la figure 2.5 pour une illustration. Ceci permet de proposer une estimation du paramètre inconnu θ par un *intervalle de confiance* construit à partir de la valeur observée $M_n(\omega)$ (voir [ICP] chapitre 6 pour les détails).

On peut aussi proposer une *estimation ponctuelle* de θ en utilisant la convergence presque sûre de M_n vers θ (loi forte des grands nombres, voir [ICP] chapitre 6). On estime alors θ par la valeur observée $M_n(\omega)$. Cette approche suppose que l'on remplace l'espace mesurable $(\Omega_n, \mathcal{F}_n)$ par un espace (Ω, \mathcal{F}) ne dépendant pas de n et assez « riche » pour supporter une suite infinie $(X_i)_{i \geq 1}$ de variables aléatoires de Bernoulli qui soient sous chaque \mathbf{P}_θ , indépendantes et de même loi $\text{Bern}(\theta)$. Ce problème est celui de la modélisation du jeu de pile ou face infini⁷. La difficulté est exactement la même que celle que nous avons rencontrée pour construire la mesure de Lebesgue sur \mathbb{R} .

⁶Pour représenter graphiquement la loi discrète $\mu = \sum p_k \delta_{x_k}$, on trace à partir de chaque point $(x_k, 0)$ un segment vertical de hauteur proportionnelle à p_k . La figure 2.5 affiche théoriquement 101 bâtons pour chacune des deux lois de M_{100} , sous $\mathbf{P}_{0,15}$ et sous $\mathbf{P}_{0,7}$. En pratique seuls sont visibles ceux dont la hauteur est supérieure à l'épaisseur d'un trait d'imprimante.

⁷Voir les Annales d'IFP 2002-2003, D.M. n° 2.

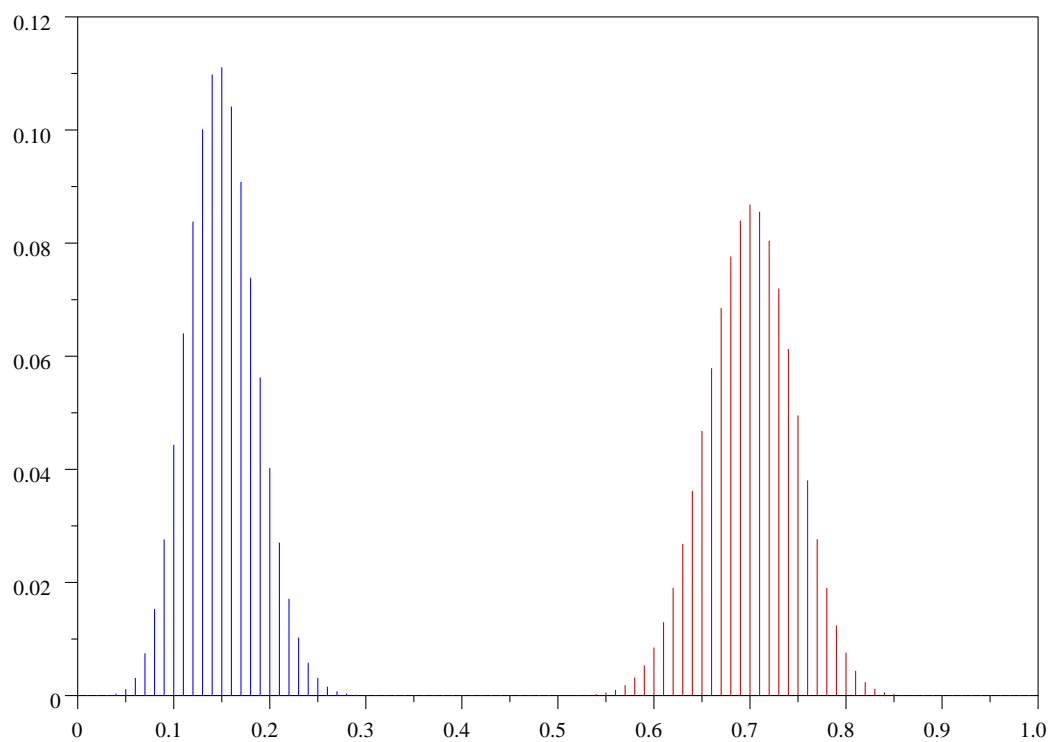


FIG. 2.5 – Loi de M_{100} sous $\mathbf{P}_{0,15}$ et sous $\mathbf{P}_{0,7}$