Leçon 7

Mesure de probabilité

1 Éléments de langage

Le calcul des probabilités fait usage, sur quelques aspects, d'une terminologie propre qu'il convient toutefois de limiter à son minimum. Ce paragraphe décrit ainsi quelques éléments de langage qui prendront petit à petit leur sens et leur pertinence.

Étant donné un espace mesurable $(E, \mathcal{B})$, une partie B de $\mathcal{B}$ sera souvent appelée un *événement*.

Définition 1 (Mesure de probabilité)**.** *Une mesure P sur un espace mesurable $(E, \mathcal{B})$ de masse totale $P(E) = 1$, est appelée mesure de probabilité, ou parfois plus simplement probabilité. L'espace $(E, \mathcal{B}, P)$ est alors un espace probabilisé.*

Une mesure μ sur $(E, \mathcal{B})$ de masse totale finie $\mu(E) < \infty$ (et non nulle) peut toujours être normalisée en une mesure de probabilité $P = \frac{1}{\mu(E)}\,\mu$. Par exemple, la mesure de Lebesgue λ définit une mesure de probabilité sur tout intervalle (a, b), $a < b$, (ouvert, fermé, semi-ouvert) borné en posant $P = \frac{1}{b-a}\lambda$ (sur les boréliens de (a, b)). Un autre exemple fondamental de mesure de probabilité est une mesure de Dirac $P = \delta_e$, $e \in E$.

Un autre bénéfice d'une masse totale finie est l'existence de l'intégrale $\int_E \phi\, dP$ pour toutes les fonctions $\phi : (E, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurables bornées puisque si $|\phi(x)| \leq c$, $x \in E$, pour une constante $c \geq 0$,

$$\int_E |\phi| dP \;\leq\; \int_E c\, dP \;=\; cP(E) \;=\; c \;<\; \infty.$$

En particulier, il sera souvent commode de décrire une propriété sur les familles de fonctions boréliennes bornées (plutôt qu'intégrables, évitant la référence à une mesure donnée).

Pour une mesure de probabilité P sur $(E, \mathcal{B})$, quelques propriétés de la théorie de la mesure sont simplifiées ; par exemple, si B_n, $n \in \mathbb{N}$, est une suite décroissante d'éléments de $\mathcal{B}$ d'intersection B, il est toujours vrai que $P(B) = \lim_{n\to\infty} P(B_n)$. Dire que $B \in \mathcal{B}$ est un ensemble de mesure nulle, $P(B) = 0$, est équivalent à dire que son complémentaire B^c est de mesure 1 (ou pleine), $P(B^c) = 1$.

Noter également que pour une suite B_n, $n \in \mathbb{N}$, d'éléments de $\mathcal{B}$, $P(B_n) = 1$ pour tout $n \in \mathbb{N}$ équivaut à $P(\bigcap_{n\in\mathbb{N}} B_n) = 1$. En effet, par passage au complémentaire, $0 \leq P\big(\bigcup_{n\in\mathbb{N}} B_n^c\big) \leq \sum_{n\in\mathbb{N}} P(B_n^c) = 0$.

L'usage de systèmes complets d'événements sera constant, notamment à travers la relation de Chasles.

Définition 2 (Système complet d'événements)**.** *Une famille dénombrable B_n, $n \in \mathbb{N}$, d'événements de $\mathcal{B}$ est un système complet d'événements (de l'espace probabilisé $(E, \mathcal{B}, P)$) si les B_n sont disjoints deux à deux de réunion de probabilité* 1*,*

$$\mathbb{P}\bigg(\bigcup_{n\in\mathbb{N}} B_n\bigg) = \sum_{n\in\mathbb{N}} P(B_n) = 1.$$

Comme annoncé, pour un tel système complet d'événements, si $\phi : (E, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable, positive ou bornée, et si $B = \bigcup_{n\in\mathbb{N}} B_n$, d'après la relation de Chasles,

$$\int_E \phi \, dP = \int_B \phi \, dP = \sum_{n\in\mathbb{N}} \int_{B_n} \phi \, dP.$$

La propriété presque partout est habituellement traduite en *presque sûrement*.

2 Exemples de mesures de probabilité

La définition d'une mesure de probabilité P sur un espace mesurable $(E, \mathcal{B})$ est donc celle d'une mesure telle que $P(E) = 1$. Ce paragraphe décrit quelques exemples, à la fois génériques et traditionnels, de mesures de probabilité. Ce petit inventaire est très incomplet, et d'autres exemples apparaîtront naturellement au fil des leçons.

Un premier cadre est celui d'un ensemble $E = \{e_n\,;\, n \in \mathbb{N}\}$ fini ou dénombrable, muni de la σ-algèbre des parties $\mathcal{B} = \mathcal{P}(E)$ (typiquement $E = \mathbb{N}$). Une mesure de probabilité P sur $(E, \mathcal{P}(E))$ est alors simplement de la forme

$$P \,=\, \sum_{n\in\mathbb{N}} p_n \delta_{e_n}$$

où les réels p_n, $n \in \mathbb{N}$, sont positifs ou nuls de somme égale à 1, $\sum_{n\in\mathbb{N}} p_n = 1$. S'il n'y a qu'un nombre fini $p_1, \ldots, p_N$, $N \geq 1$, de p_n non nuls, $P = \sum_{n=1}^N p_n \delta_{e_n}$ (de sorte que le cas dénombrable inclut le cas fini). Le nombre p_n est à entendre comme le *poids* de la probabilité P sur l'entier *(atome)* e_n, $n \in \mathbb{N}$, puisque $P(\{e_n\}) = p_n$. Seuls les poids strictement positifs sont en fait à prendre en compte.

Il n'est pas inutile de rappeler l'intégration par rapport à ces mesures : si $\phi : E \to \mathbb{R}$ est à valeurs positives ou bornées,

$$\int_E \phi \, dP \,=\, \sum_{n\in\mathbb{N}} p_n \phi(e_n).$$

Cette formule peut être avantageusement réalisée comme l'application de la relation de Chasles au système complet d'événements $B_n = \{e_n\}$, $n \in \mathbb{N}$, puisque

$$\int_E \phi \, dP \,=\, \sum_{n\in\mathbb{N}} \int_{\{e_n\}} \phi \, dP \,=\, \sum_{n\in\mathbb{N}} \phi(e_n) P\big(\{e_n\}\big) \,=\, \sum_{n\in\mathbb{N}} p_n \phi(e_n).$$

Une mesure de probabilité sur une ensemble fini ou dénombrable est souvent qualifiée de *discrète* ou *atomique*. Il y a autant de mesures discrètes que de séries à termes positifs convergentes, mais bien entendu, comme toujours en mathématique, certaines sommes ou séries sont plus intéressantes que d'autres.

Les premiers exemples suivants font partie des mesures de probabilité discrètes les plus classiques.

(Mesure) Masse de Dirac δ_e. Si $(E, \mathcal{B})$ est un espace mesurable et $e \in E$, $P = \delta_e$ est la masse de Dirac en e.

(Mesure de) Probabilité de Bernoulli[1] $\mathcal{B}(p)$. Si $(E, \mathcal{B})$ est un espace mesurable, $a, b \in E$, distincts, et $p \in]0,1[$,

$$P \;=\; (1-p)\delta_a + p\delta_b$$

est la mesure de probabilité de Bernoulli sur $\{a, b\}$ de paramètre p. En d'autres termes, $P(\{a\}) = 1 - p$, $P(\{b\}) = p$. Traditionnellement $\{a, b\} = \{0, 1\}$ (ou $\{-1, +1\}$), et p est appelé *paramètre de succès* (communément associé à δ_1). Le paramètre p peut être égal à 0 ou à 1, la probabilité de Bernoulli se résumant dans ces cas à une masse de Dirac. Elle est notée $\mathcal{B}(p)$ (si l'espace sous-jacent est clairement exprimé).

(Mesure de) Probabilité uniforme (discrète) $\mathcal{U}(\{1, \ldots, n\})$. Pour $n \geq 1$, la probabilité uniforme P sur $E = \{1, \ldots, n\}$ est la probabilité attribuant le même poids à tous les entiers $k \in \{1, \ldots, n\}$, autrement dit $P = \frac{1}{n} \sum_{k=1}^{n} \delta_k$, ou $P(\{k\}) = \frac{1}{n}$, $k = 1, \ldots, n$. C'est la mesure de comptage normalisée. Elle est éventuellement désignée par $\mathcal{U}(\{1, \ldots, n\})$. Si $n = 1$, c'est la masse de Dirac δ_1, si $n = 2$, elle coïncide avec la mesure de Bernoulli $\mathcal{B}(\frac{1}{2})$ (sur $\{1, 2\}$).

(Mesure de) Probabilité binomiale $\mathcal{B}(n, p)$. Pour $n \geq 1$ et $p \in]0,1[$, la probabilité binomiale P sur $E = \{0, 1, \ldots, n\}$ est définie par

$$P \;=\; \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}\, \delta_k$$

1. Jakob Bernoulli, mathématicien et physicien suisse (1654–1705).

où, pour $k \in \{0,1,\dots,n\}$,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

est le coefficient binomial[2]. Autrement dit $P(\{k\}) = \binom{n}{k} p^k(1-p)^{n-k}$, $k = 0,1,\dots,n$. C'est une probabilité puisque d'après la formule du binôme (d'où le nom de la mesure)

$$\sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} = \big[p + (1-p)\big]^n = 1.$$

Elle est notée $\mathcal{B}(n,p)$, mesure de probabilité binomiale de taille n et de paramètre de succès p. Noter que $\mathcal{B}(1,p)$ est la mesure de Bernoulli $\mathcal{B}(p)$ (sur $\{0,1\}$). Il est possible d'inclure $p=0$ et $p=1$ dans la définition, conduisant à la masse de Dirac aux points 0 et n respectivement (avec la convention $0^0 = 1$).

(Mesure de) Probabilité de Poisson[3] $\mathcal{P}(\theta)$. Cette mesure correspond à la série exponentielle $e^\theta = \sum_{n\in\mathbb{N}\cup\{0\}} \frac{\theta^n}{n!}$, $\theta \in \mathbb{R}$. Après normalisation, la probabilité de Poisson $\mathcal{P}(\theta)$ pour $\theta > 0$ est la mesure sur $(\mathbb{N}\cup\{0\}, \mathcal{P}(\mathbb{N}\cup\{0\}))$ définie par

$$P = \sum_{n\in\mathbb{N}\cup\{0\}} e^{-\theta}\,\frac{\theta^n}{n!}\,\delta_n.$$

Autrement dit $P(\{n\}) = e^{-\theta}\,\frac{\theta^n}{n!}$, $n \in \mathbb{N}\cup\{0\}$.

(Mesure de) Probabilité géométrique $\mathcal{G}(p)$. Si $p \in]0,1[$, la série géométrique de raison p est sommable et $\sum_{n\in\mathbb{N}\cup\{0\}} p^n = \frac{1}{1-p}$. Elle conduit à la forme de la probabilité géométrique de paramètre p sur $\mathbb{N}\cup\{0\}$ exprimée par

$$P = \sum_{n\in\mathbb{N}\cup\{0\}} (1-p)p^n\,\delta_n.$$

2. Ce coefficient est traditionnellement noté C_n^k dans les mathématiques françaises du 20e siècle.
3. Siméon Denis Poisson, mathématicien, géomètre et physicien français (1781-1840).

De nombreuses illustrations probabilistes privilégient la forme sur $\mathbb{N}$, en échangeant en outre p et $1-p$ pour définir la mesure de probabilité géométrique de paramètre $p \in]0,1[$, notée $\mathcal{G}(p)$, comme

$$P = \sum_{n \in \mathbb{N}} p(1-p)^{n-1} \, \delta_n.$$

En termes de poids, $P(\{n\}) = p(1-p)^{n-1}$, $n \in \mathbb{N}$. C'est la forme adoptée dans la suite. La valeur $p = 1$ peut être comprise dans la définition mais conduit à nouveau à une forme dégénérée, la masse de Dirac en 1.

Après les mesures discrètes, la seconde grande famille de mesures de probabilité est celle des mesures ayant une densité, le plus souvent par rapport à la mesure de Lebesgue sur $E = \mathbb{R}$, ou $\mathbb{R}^d$ dans la suite, muni de la tribu $\mathcal{B} = \mathcal{B}(\mathbb{R})$ des boréliens. Précisément, si λ est la mesure de Lebesgue sur $\mathcal{B}(\mathbb{R})$, et si $f : \mathbb{R} \to \mathbb{R}$ est borélienne positive, intégrable par rapport à λ, telle que $\int_{\mathbb{R}} f d\lambda = 1$, la mesure définie par

$$P(B) = \int_B f \, d\lambda = \int_{\mathbb{R}} (\mathbb{1}_B f) d\lambda, \quad B \in \mathcal{B}(\mathbb{R}),$$

est la mesure de probabilité P dite *de densité f par rapport à* λ, notée $dP = f d\lambda$ (densité de Radon-Nikodym, voir Leçon 3). Comme dans le cas discret, la fonction f peut être nulle (ou nulle λ-presque partout) sur des parties (mesurables) de $\mathbb{R}$, restreignant ainsi P aux parties complémentaires (voir les exemples ci-dessous).

Il convient de comparer les mesures discrètes et les mesures à densité comme deux façons différentes de répartir une masse (de sable pour une illustration) totale égale à 1, soit en la concentrant en un nombre fini ou dénombrable de points (atomes), soit en la répartissant de façon diffuse suivant une fonction de densité f, la probabilité d'un intervalle (a,b) correspondant, à travers l'expression $\int_{(a,b)} f d\lambda$, à l'intégrale sous le graphe de la fonction f sur l'intervalle (a,b). Il faut rappeler ici la notation abusive (a,b) pour l'intervalle ouvert, fermé, semi-ouvert, en soulignant que le choix importe peu puisque la mesure de Lebesgue ne charge pas les points. En particulier $\int_{(a,a)} f d\lambda = 0$. Bien entendu, il est facile d'imaginer, et il appartiendra aux leçons suivantes d'en discuter, des combinaisons de ces deux modèles.

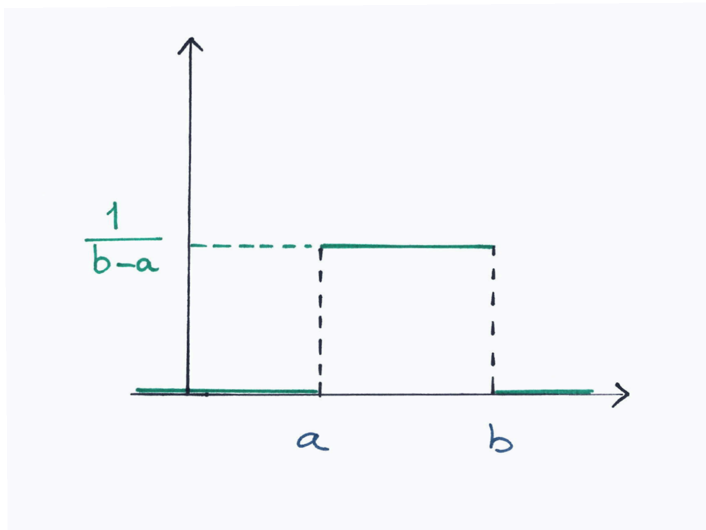
Comme dans le cas discret, il est bon de rappeler l'intégration par rapport aux mesures à densité, résumée dans l'expression $dP = f d\lambda$: si $\phi : \mathbb{R} \to \mathbb{R}$ est borélienne à valeurs positives ou bornées,

$$\int_{\mathbb{R}} \phi \, dP = \int_{\mathbb{R}} \phi f \, d\lambda.$$

Pour paraphraser les exemples discrets, il y a autant de mesures à densité que de fonctions mesurables positives d'intégrale 1 par rapport à la mesure de Lebesgue. Mais certaines sont plus intéressantes, ou communes, que d'autres. Pour tous ces exemples, une représentation graphique de la fonction de densité aidera à la compréhension.

(Mesure de) Probabilité uniforme (continue) $\mathcal{U}(a, b)$. Comme la masse de Dirac, c'est à la fois la plus simple des mesures à densité et peut être la moins immédiate à appréhender. Elle correspond au choix d'une fonction de densité constante sur un intervalle (a, b), $a < b$, normalisée donc par la longueur de l'intervalle afin d'obtenir une intégrale égale à 1. Ainsi, la densité de P est

$$f = \frac{1}{b-a} \mathbb{1}_{(a,b)}.$$

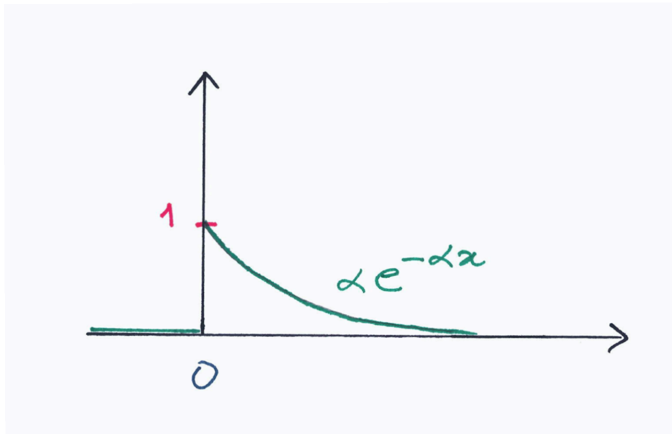


Pour la bien comprendre, si $(u, v) \subset (a, b)$, $P((u, v)$ est la mesure de Lebesgue (longueur) de l'intervalle (u, v) normalisée, soit $P((u, v)) = \frac{v-u}{b-a}$. Cette mesure est donc appelée uniforme sur l'intervalle (a, b), notée $\mathcal{U}(a, b)$. Rappeler que l'intervalle (a, b) peut être ouvert, fermé, semi-ouvert, ne changeant pas la nature de la mesure. Pour quelque souplesse dans certaines illustrations, il sera commode de garder ces options.

(Mesure de) Probabilité exponentielle $\mathcal{E}(\alpha)$. La mesure de probabilité exponentielle $\mathcal{E}(\alpha)$ de paramètre $\alpha > 0$ a pour densité

$$f(x) = \alpha\, e^{-\alpha x}\, \mathbb{1}_{]0,\infty[}(x), \quad x \in \mathbb{R},$$

par rapport à la mesure de Lebesgue sur $\mathbb{R}$.



L'intervalle ouvert $]0, \infty[$ peut être remplacé par l'intervalle ouvert $[0, \infty[$, et f peut être entendue de façon équivalente comme une densité par rapport à la mesure de Lebesgue sur l'un ou l'autre de ces intervalles. Le plus souvent, l'intervalle ouvert $]0, \infty[$ sera employé. La probabilité exponentielle est utilisée pour décrire la durée de vie de modèles atomiques.

(Mesure de) Probabilité gaussienne ou normale $\mathcal{N}(m, \sigma^2)$. Un calcul intégral qui a été expliqué à l'Exercice 4, Leçon 4 démontre que

$$\int_{\mathbb{R}} e^{-\frac{1}{2}x^2} d\lambda = \sqrt{2\pi}\,.$$

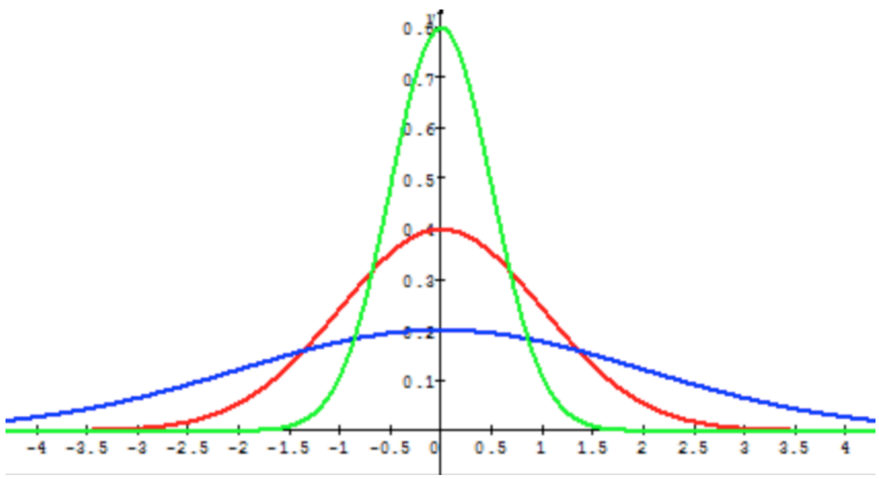
(La normalisation par 2 dans l'exponentielle est une convention probabiliste très avantageuse.) La mesure de probabilité gaussienne, ou normale, $\mathcal{N}(0, 1)$, de paramètres 0 et 1 est ainsi définie comme la mesure P sur les boréliens de $\mathbb{R}$ de densité

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}.$$

Les paramètres 0 et 1 seront expliqués par la suite (Leçon 9). En fait, un changement de variable par translation et homothétie permet de définir de la même façon la mesure $\mathcal{N}(m, \sigma^2)$ de paramètres $m \in \mathbb{R}$ et $\sigma > 0$, de densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-m)^2}, \quad x \in \mathbb{R},$$

par rapport à la mesure de Lebesgue sur $\mathbb{R}$. La représentation graphique de cette fonction est instructive, notamment la symétrie par rapport à la verticale passant par m, et le rôle de σ : plus celui-ci est grand, plus la courbe est aplatie, plus σ est petit, plus elle est pointue (en m). Il faut imaginer ce que $\sigma = 0$ pourrait signifier.



Les exemples canoniques précédents décrivent des mesures de probabilités sur $\mathbb{N}$, ou une partie de $\mathbb{N}$, et sur $\mathbb{R}$, ou une partie de $\mathbb{R}$. Il sera commode pour l'analyse, en particulier dès le paragraphe suivant, de considérer suivant le cas que toutes ces mesures sont des mesures sur $\mathbb{R}$ muni de la tribu borélienne $\mathcal{B}(\mathbb{R})$, de support fini, dénombrable, ou un intervalle de $\mathbb{R}$ (voir l'extension d'une mesure en Leçon 2). Il est en effet profitable de tirer bénéfice des divers points de vue pour la meilleure compréhension des mesures de probabilité concernées.

3 Fonction de répartition

La fonction de répartition est en fait la fonction croissante qui définit une mesure de Stieltjes. Le point de vue est ici renversé, la mesure (de probabilité) est donnée, lui est associée une fonction croissante qui la détermine complètement.

Définition 3 (Fonction de répartition)**.** *Soit P une mesure de probabilité sur $\mathcal{B}(\mathbb{R})$; la fonction de répartition de P est la fonction*

$$F(t) \,=\, P\big(\,]-\infty, t]\big), \quad t \in \mathbb{R}.$$

Lorsque t varie en croissant de $-\infty$ à $+\infty$, la fonction de répartition « cumule » donc la masse de P (et la terminologie anglo-saxonne de « cumulative distribution function » traduit peut-être mieux sa signification). Il est clair en effet sur sa définition que F est croissante (car $P(\,]-\infty, s]) \leq P(\,]-\infty, t])$ si $]-\infty, s] \subset \,]-\infty, t]$) et à valeurs dans $[0, 1]$. Comme

$$\bigcap_{n\in\mathbb{N}} \,]-\infty, -n] \,=\, \emptyset \quad \text{et} \quad \bigcup_{n\in\mathbb{N}} \,]-\infty, n] \,=\, \mathbb{R},$$

les propriétés de convergence monotone d'une mesure (de probabilité) assurent que

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

En outre, F est continue à droite en tout point, autrement dit, pour tout $t \in \mathbb{R}$,

$$\lim_{s\to t,\, s>t} F(s) \,=\, F(t)$$

(soit $F(t+) = F(t)$). En effet, comme F est croissante, il suffit de vérifier que $\lim_{n\to\infty} F(t + \frac{1}{n}) \,=\, F(t)$ ce qui résulte des propriétés d'une probabilité pour les intersections dénombrables, à savoir

$$\begin{aligned}
\lim_{n\to\infty} F\big(t + \tfrac{1}{n}\big) \,&=\, \lim_{n\to\infty} P\big(\,]-\infty, t + \tfrac{1}{n}]\big) \\
&=\, P\bigg(\bigcap_{n\geq 1} \,]-\infty, t + \tfrac{1}{n}]\bigg) \\
&=\, P\big(\,]-\infty, t]\big) \,=\, F(t).
\end{aligned}$$

Un argument similaire, basé sur la croissance, montre que F admet des limites à gauche, c'est-à-dire que la limite

$$\begin{aligned}\lim_{s\to t,\, s<t} F(s) \,=\, \lim_{n\to\infty} F\big(t-\tfrac{1}{n}\big) \,&=\, \lim_{n\to\infty} P\big(\,]-\infty, t-\tfrac{1}{n}]\big)\\ &=\, P\bigg(\bigcup_{n\geq 1}\,]-\infty, t-\tfrac{1}{n}]\bigg)\\ &=\, P\big(\,]-\infty, t[\big) \,=\, F(t-)\end{aligned}$$

existe en tout point $t\in\mathbb{R}$.

Il résulte de ces propriétés que si la fonction de répartition F de P détermine (par définition) la probabilité P sur les intervalles $]-\infty, t]$, $t\in\mathbb{R}$, elle détermine en fait P sur toutes les formes d'intervalles ; par exemple, si $a<b$,

$$\begin{aligned}P\big([a,b]\big) \,&=\, P\big(]-\infty, b]\setminus\,]-\infty, a[\big)\\ &=\, P\big(]-\infty, b]\big) - P\big(]-\infty, a[\big)\\ &=\, F(b)-F(a-),\end{aligned}$$

ou encore $P(]a,b[) = F(b-) - F(a)$.

L'Exercice 2, Leçon 10, explique comment toute fonction $F:\mathbb{R}\to[0,1]$, croissante, continue à droite et telle que $\lim_{t\to-\infty}F(t)=0$, $\lim_{t\to+\infty}F(t)=1$, est la fonction de répartition d'une mesure de probabilité sur $\mathcal{B}(\mathbb{R})$. Cette mesure est en outre unique en vertu de l'énoncé suivant (qui constitue la partie unicité du théorème de Carathéodory, Théorème 5, Leçon 2).

Proposition 4. *La fonction de répartition détermine de façon unique la probabilité : si deux mesures de probabilité P et Q sur $\mathcal{B}(\mathbb{R})$ ont même fonction de répartition, alors $P=Q$.*

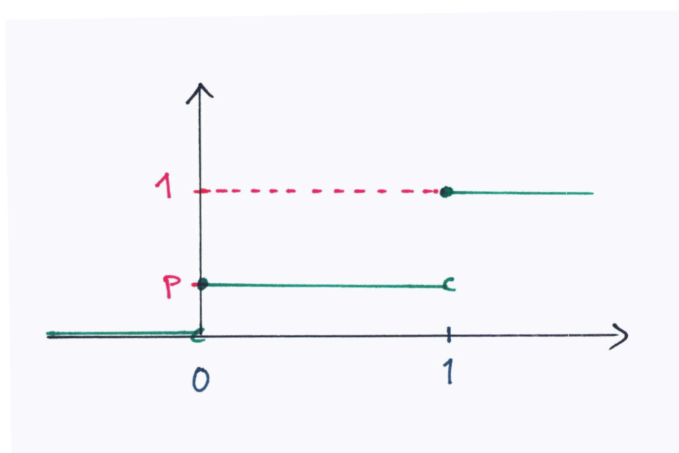
Démonstration. Ce résultat découle d'un argument de classe monotone (Leçon 1). En effet, si P et Q ont même fonction de répartition, elles coïncident sur tous les intervalles de $\mathbb{R}$. Soit alors

$$\mathcal{M} \,=\, \big\{B\in\mathcal{B}(\mathbb{R})\,;\, P(B)=Q(B)\big\}.$$

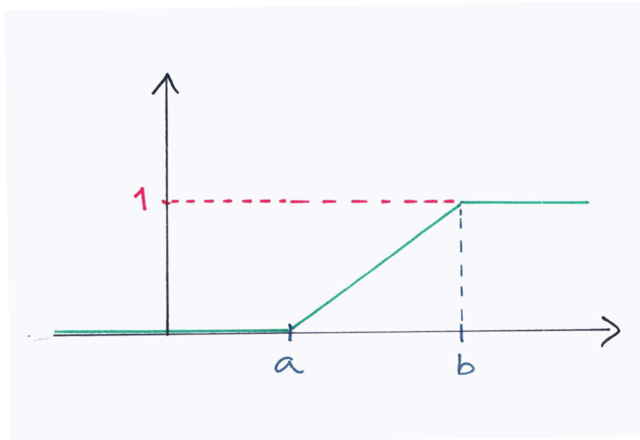
À l'aide des propriétés des mesures, il est aisé de vérifier que $\mathcal{M}$ est une classe monotone ; d'après ce qui précède, elle contient les intervalles, et donc contient la classe monotone engendrée par les intervalles. La famille des intervalles étant stable par intersection finie, cette dernière coïncide, d'après le théorème des classes monotones (Théorème 12, Leçon 1), avec la tribu engendrée par ces mêmes intervalles, qui n'est autre que $\mathcal{B}(\mathbb{R})$. Ainsi P et Q coïncident sur $\mathcal{B}(\mathbb{R})$ comme annoncé. □

La représentation graphique des fonctions de répartition des exemples de mesures de probabilité du Paragraphe 2 est instructive. Prendre soin en particulier de la différence entre fonction de densité et fonction de répartition. Les fonctions de répartition des mesures discrètes sont en escalier, celles des mesures à densité sont continues. Il est à noter en particulier que si $P = \delta_a$, $a \in \mathbb{R}$, alors $F = \mathbb{1}_{[a,\infty[}$.

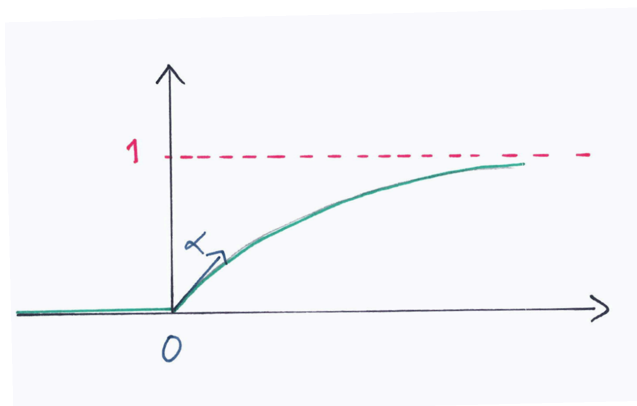
Voici quelques représentations de fonctions de répartition :



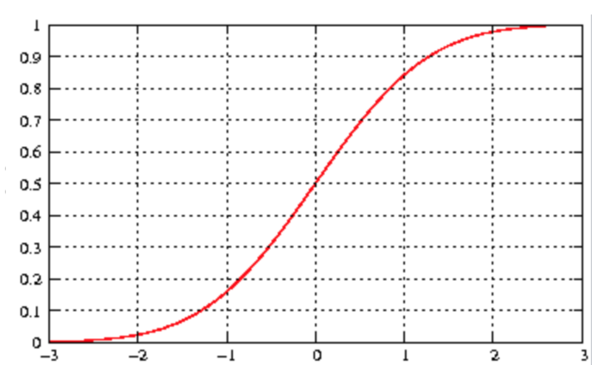
Probabilité de Bernoulli $\mathcal{B}(p)$



Probabilité uniforme $\mathcal{U}(a, b)$



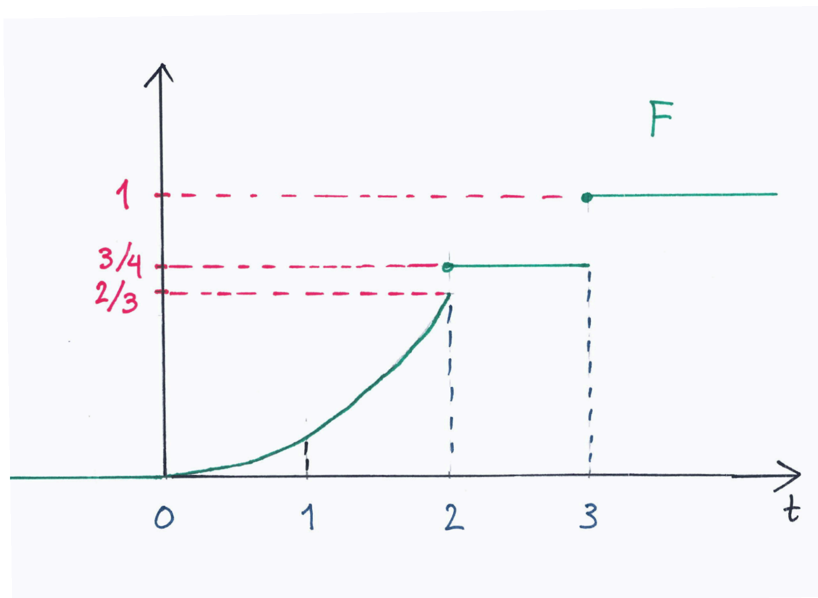
Probabilité exponentielle $\mathcal{E}(\alpha)$



Probabilité gaussienne $\mathcal{N}(0, 1)$

Afin d'avancer dans l'analyse des fonctions de répartition, l'exemple suivant pourra être utile. Soit $F : \mathbb{R} \to [0,1]$ la fonction définie par

$$F(t) = \begin{cases} 0 & \text{si} \quad t < 0, \\ \frac{t^2}{6} & \text{si} \quad 0 \leq t < 2, \\ \frac{3}{4} & \text{si} \quad 2 \leq t < 3, \\ 1 & \text{si} \quad 3 \leq t. \end{cases}$$



La fonction F vérifie les propriétés d'une fonction de répartition, elle détermine donc une mesure de probabilité P sur $\mathcal{B}(\mathbb{R})$. La question est d'essayer de décrire P à partir de F, avec toujours cette idée que F cumule, en croissant, la masse de P depuis $-\infty$. À cet effet, il peut être observé que F admet des points de discontinuité en $t = 2$ et $t = 3$, d'amplitudes respectives $\frac{1}{12}$ et $\frac{1}{4}$. La probabilité P charge ainsi ces points comme la mesure discrète

$$\mu_d = \frac{1}{12}\,\delta_2 + \frac{1}{4}\,\delta_3.$$

En dehors de ces points, F est continue et même dérivable. Autrement dit, la fonction F_c obtenue à partir de F en réduisant à 0 ces amplitudes est continue et dérivable en dehors de ces points, de dérivée (en changeant intentionnellement la lettre t en x)

$$f(x) \;=\; \frac{x}{3}\,\mathbb{1}_{]0,2[}(x)$$

(la dérivée est nulle sur $]2,3[$ puisque F y est constante, traduisant une absence d'accumulation de masse). La probabilité P, en dehors des points 2 et 3, se comporte donc comme la mesure μ_c de densité f par rapport à la mesure de Lebesgue λ.

En conséquence de ces observations, $P = \mu_d + \mu_c$. Par abus de notation, il sera souvent écrit

$$P \;=\; \frac{1}{12}\,\delta_2 + \frac{1}{4}\,\delta_3 + f\,d\lambda$$

dans l'écriture de Radon-Nikodym. Ni μ_d ni μ_c ne sont de probabilité, et suivant le cas, elles peuvent être normalisées par leur masse totale respective, à savoir $\frac{1}{3}$ et $\frac{2}{3}$, pour en faire des probabilités $P_d = 3\mu_d$ et $P_c = \frac{3}{2}\,\mu_d$, conduisant à la décomposition de P sous la forme de la combinaison convexe d'une probabilité discrète et d'une probabilité à densité

$$P \;=\; \frac{1}{3}\,P_d + \frac{2}{3}\,P_c.$$

Cet exemple illustre une procédure générale, basée pour commencer sur l'observation suivante. Celle-ci s'applique plus généralement à toute fonction croissante.

Proposition 5. *Une fonction de répartition $F : \mathbb{R} \to [0,1]$ n'admet qu'un nombre au plus dénombrable de points de discontinuité.*

Démonstration. Un point de discontinuité t de F est caractérisé par $F(t) - F(t-) > 0$, puisque F est continue à droite. Poser alors, pour tout $n \geq 1$,

$$D_n \;=\; \big\{t \in \mathbb{R}\,;\; F(t) - F(t-) \geq \tfrac{1}{n}\big\}.$$

L'ensemble D_n est nécessairement fini car, par croissance de F, il ne peut contenir plus de $n+1$ points. La conclusion s'ensuit en considérant la réunion des D_n, $n \geq 1$. $\square$

Comme annoncé, cette démonstration s'étend à toute fonction croissante $G : \mathbb{R} \to \mathbb{R}$. Une réunion dénombrable d'ensembles au plus dénombrables étant dénombrable, il suffit en effet de vérifier l'assertion sur tout intervalle $]-k, k[$, $k \geq 1$, dont l'image par G est contenue dans l'intervalle $[G(-k), G(k)]$. Il suffit alors de reprendre l'argument précédent en remplaçant $[0, 1]$ par cet intervalle.

Soit alors F la fonction de répartition d'une mesure de probabilité P sur les boréliens de $\mathbb{R}$, et soit t_n, $n \in \mathbb{N}$, la suite de ses points de discontinuité, d'amplitudes respectives $a_n = F(t_n) - F(t_n-)$, $n \in \mathbb{N}$. La mesure discrète $\mu_d = \sum_{n \in \mathbb{N}} a_n \delta_{t_n}$ correspond aux masses atomiques de P ; elle a pour masse totale $\alpha = \mu_d(\mathbb{R}) \leq 1$ (éventuellement nulle). La mesure $\mu_c = P - \mu_d$ n'a plus de masses atomiques. Ainsi P se décompose sous la forme

$$P \;=\; \alpha P_d + (1-\alpha) P_c$$

où $\alpha P_d = \mu_d$ et $(1-\alpha)P_c = \mu_c$, les probabilités P_d et P_c n'étant effectives que si respectivement $\alpha > 0$ et $\alpha < 1$ (autrement dit $P = P_c$ si $\alpha = 0$ et $P = P_d$ si $\alpha = 1$). La probabilité P_d est la partie discrète de P, P_c sa partie continue.

Si la fonction de répartition F_c de P_c, qui est obtenue à partir de F en éliminant les points de sauts et en renormalisant, est dérivable (sauf éventuellement en un nombre fini ou dénombrable de points, presque partout en fait), P_c admet une densité f par rapport à la mesure de Lebesgue donnée par la dérivée F_c' presque partout. C'est la situation qui a été décrite dans l'exemple précédent. Il sera alors écrit, dans un abus de notation,

$$P \;=\; \alpha \sum_{n \in \mathbb{N}} a_n \delta_{t_n} + (1-\alpha) f d\lambda.$$

4 Complément : Décomposition d'une mesure

Il ne faudrait pas que la décomposition précédente d'une mesure (de probabilité) sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ laisse à penser qu'il n'y a que des mesures discrètes et des mesures à densité. En fait, le théorème de Radon-Nikodym (Leçon 3) fournit des mesures dites *étrangères* à la mesure de Lebesgue. Le théorème suivant est énoncé pour des probabilités, mais s'étend aux mesures quelconques.

Théorème 6 (Décomposition d'une mesure)**.** *Soit P une mesure de probabilité sur $\mathcal{B}(\mathbb{R})$; alors*

$$P = \alpha P_d + \beta P_c + \gamma P_\perp$$

où $\alpha, \beta, \gamma \geq 0$ et $\alpha + \beta + \gamma = 1$, P_d est une probabilité discrète, P_c admet une densité par rapport à la mesure de Lebesgue λ, et $P_\perp$ est sans atome et étrangère à λ au sens où $P_\perp(B) = 1$ sur un borélien B tel que $\lambda(B) = 0$.

Bien entendu P_d est aussi étrangère à λ, mais $P_\perp$ n'admet pas d'atomes. Un exemple de mesure étrangère sans atome est fourni par la mesure de Cantor (Exercice 16, Leçon 3).

Exercices

(*Une étoile * désignera une question de difficulté supérieure.*)

Exercice 1 (*Formule du crible de Poincaré*). Soit P une mesure de probabilité sur un espace mesurable $(X, \mathcal{A})$; démontrer que pour tous $A_1, \ldots, A_n \in \mathcal{A}$,

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 \cap A_2), \\ P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3), \end{aligned}$$

puis que

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})$$

(*Indication* : effectuer une récurrence sur n, ou développer puis intégrer par rapport à P l'identité

$$1 - \mathbb{1}_A = \mathbb{1}_{A^c} = \prod_{k=1}^{n} \mathbb{1}_{A_k^c} = \prod_{k=1}^{n} (1 - \mathbb{1}_{A_k})$$

où $A = \bigcup_{k=1}^{n} A_k$.)

Exercice 2. Soit $F : \mathbb{R} \to \mathbb{R}$ la fonction définie par

$$F(t) = \begin{cases} 0 & \text{si} \quad t < 0, \\ \frac{1}{4} & \text{si} \quad 0 \leq t < 1, \\ \frac{t}{2} & \text{si} \quad 1 \leq t < 2, \\ 1 & \text{si} \quad 2 \leq t. \end{cases}$$

Représenter F.

a) Montrer que F est la fonction de répartition d'une unique mesure de probabilité P sur les boréliens de $\mathbb{R}$.

b) Calculer $P(\{\frac{1}{2}\}), P(\{1\}), P(]\frac{1}{2}, \frac{3}{2}[)$.

c) Décomposer P en la somme d'une mesure atomique et d'une mesure à densité.

Exercice 3. Soit, pour tout réel $\beta \in [0,1]$, la fonction $F_\beta : \mathbb{R} \to \mathbb{R}$ définie par

$$F(t) \;=\; \begin{cases} 0 & \text{si} \quad t < 0, \\ \beta + (1-\beta)t & \text{si} \quad 0 \leq t < 1, \\ 1 & \text{si} \quad 1 \leq t. \end{cases}$$

Vérifier que F_β est une fonction de répartition. Décrire la probabilité P associée (décomposer P sous la forme $P = \beta P_1 + (1-\beta)P_2$ où P_1 est une probabilité discrète et P_2 une probabilité à densité que l'on précisera). Discuter les cas $\beta = 0$ et $\beta = 1$.

Exercice 4 (*(Mesure de) Probabilité logistique*). Démontrer que la fonction (*sigmoïde*)

$$F(t) \;=\; \frac{1}{1+e^{-t}}\,, \quad t \in \mathbb{R},$$

est la fonction de répartition d'une mesure de probabilité P sur les boréliens de $\mathbb{R}$ admettant une densité par rapport à la mesure de Lebesgue.

Exercice 5 (*(Mesure de) Probabilité de Gumbel*[4]). Démontrer que la fonction (double exponentielle)

$$F(t) \;=\; e^{-e^{-t}}, \quad t \in \mathbb{R},$$

est une fonction de répartition (c'est la fonction de répartition de la loi dite de Gumbel). Si P désigne la mesure de probabilité de fonction de répartition F, calculer $P(A)$ pour $A = [1, \infty[$, $]-\infty, 0[$, $\{0\}$, $[0, \frac{1}{2}[$.

4. Emil Julius Gumbel, mathématicien allemand (1891–1966).