

FIMFA

**Intégration, Probabilités
et Processus Aléatoires**

Jean-François Le Gall

Septembre 2006

Département Mathématiques et Applications
Ecole normale supérieure de Paris

Chapitre 8

Fondements de la théorie des probabilités

Ce chapitre introduit les notions fondamentales de la théorie des probabilités : variables aléatoires, espérance, loi, moments de variables aléatoires, fonctions caractéristiques, etc. Puisque un espace de probabilité n'est rien d'autre qu'un espace mesurable muni d'une mesure de masse totale 1, beaucoup de ces notions correspondent à ce qui a déjà été vu dans le cadre de la théorie de l'intégration. Par exemple une variable aléatoire n'est rien d'autre qu'une fonction mesurable, et la notion d'espérance coïncide avec l'intégrale. Cependant, le point de vue de la théorie des probabilités, qui est expliqué ci-dessous, est bien différent, et une difficulté importante est de comprendre ce point de vue. Ainsi, la notion de loi, qui est un cas particulier de la notion de mesure-image, devient-elle maintenant fondamentale car elle permet d'évaluer la probabilité qu'une variable aléatoire “tombe” dans un ensemble donné.

8.1 Définitions générales

8.1.1 Espaces de probabilité

Soit (Ω, \mathcal{A}) un espace mesurable, et soit P une mesure de probabilité sur (Ω, \mathcal{A}) . On dit alors que (Ω, \mathcal{A}) est un espace de probabilité.

Un espace de probabilité est donc un cas particulier d'espace mesuré, pour lequel la masse totale de la mesure est égale à 1. En fait, le point de vue diffère de la théorie de l'intégration : dans le cadre de la théorie des probabilités, on cherche à fournir un modèle mathématique pour une “expérience aléatoire”.

- Ω représente l'ensemble de toutes les éventualités possibles, toutes les déterminations du hasard dans l'expérience considérée.
- \mathcal{A} est l'ensemble des “événements”, qui sont les parties de Ω dont on peut évaluer la probabilité. Il faut voir un événement $A \in \mathcal{A}$ comme un sous-ensemble de Ω contenant toutes les éventualités ω pour lesquelles une certaine propriété est vérifiée.

- Pour $A \in \mathcal{A}$, $P(A)$ représente la probabilité d'occurrence de l'événement A . Dans les premiers traités de théorie des probabilités, longtemps avant l'introduction de la théorie de la mesure, la probabilité $P(A)$ était définie de la manière suivante : on imagine qu'on répète l'expérience aléatoire un nombre N de fois, et on note N_A le nombre de répétitions pour lesquelles l'événement A est réalisé; alors, la proportion N_A/N converge quand $N \rightarrow \infty$ vers la probabilité $P(A)$. Nous verrons plus loin le lien entre cette définition "historique" et l'approche moderne.

Exemples. (1) On lance un dé deux fois :

$$\Omega = \{1, 2, \dots, 6\}^2, \mathcal{A} = \mathcal{P}(\Omega), P(A) = \frac{\text{Card}(A)}{36}.$$

Le choix de la probabilité correspond à l'idée que tous les résultats possibles pour les deux tirages sont équiprobables.

(2) On lance le dé jusqu'à obtenir un 6. Ici le choix de Ω est déjà moins évident. Comme le nombre de lancers nécessaires n'est a priori pas borné, le bon choix est d'imaginer qu'on fait une infinité de lancers :

$$\Omega = \{1, 2, \dots, 6\}^{\mathbb{N}^*}$$

de sorte qu'un élément de Ω est une suite $\omega = (\omega_1, \omega_2, \dots)$ qui donne les résultats des tirages successifs. La tribu \mathcal{A} sur Ω est la tribu-produit définie comme la plus petite tribu rendant mesurables tous les ensembles de la forme

$$\{\omega : \omega_1 = i_1, \omega_2 = i_2, \dots, \omega_n = i_n\}$$

où $n \geq 1$ et $i_1, \dots, i_n \in \{1, 2, \dots, 6\}$ (\mathcal{A} coïncide aussi avec la tribu borélienne pour la topologie produit sur Ω). Enfin P est l'unique mesure de probabilité sur Ω telle que, pour tout choix de n et de i_1, \dots, i_n ,

$$P(\{\omega : \omega_1 = i_1, \omega_2 = i_2, \dots, \omega_n = i_n\}) = \left(\frac{1}{6}\right)^n.$$

L'unicité de P est une conséquence simple du lemme de classe monotone. L'existence est un cas particulier de la construction de mesures sur des produits infinis. On peut aussi construire P facilement partir de la mesure de Lebesgue sur $[0, 1]$: si à tout réel $x \in [0, 1]$ on associe la suite $(\varepsilon_k)_{k \in \mathbb{N}^*} \in \Omega$ telle que $x = \sum_{k=1}^{\infty} (\varepsilon_k - 1) 6^{-k}$ (cette suite est unique pour presque tout x), la probabilité P est obtenue comme mesure-image de la mesure de Lebesgue sur $[0, 1]$ par l'application $x \rightarrow (\varepsilon_k)_{k \in \mathbb{N}^*}$.

(3) On s'intéresse au déplacement dans l'espace d'une particule ponctuelle soumise à des perturbations aléatoires. Si on se limite à l'intervalle de temps $[0, 1]$, l'espace de probabilité naturel est $C([0, 1], \mathbb{R}^3)$: un élément de Ω , une trajectoire possible, est une fonction continue $\omega : [0, 1] \rightarrow \mathbb{R}^3$. La tribu sur Ω est alors la plus petite tribu qui rende mesurables toutes les applications coordonnées $\omega \rightarrow \omega(t)$ pour $t \in \mathbb{R}_+$. Cette tribu coïncide avec la tribu borélienne pour la topologie de la convergence uniforme sur Ω . Il resterait à construire la probabilité P , pour laquelle de multiples choix sont possibles. L'exemple le plus important,

à la fois du point de vue théorique et pour les applications, est la mesure de Wiener, qui est la loi du mouvement brownien.

Remarque importante. Très souvent dans la suite, on ne spécifiera pas le choix de l'espace de probabilité. Les données importantes seront les propriétés des fonctions définies sur cet espace, les variables aléatoires.

8.1.2 Variables aléatoires

Définition 8.1.1 Soit (E, \mathcal{E}) un espace mesurable. Une application mesurable $X : \Omega \longrightarrow E$ est appelée variable aléatoire (v.a. en abrégé) à valeurs dans E .

Exemples. En reprenant les trois exemples ci-dessus :

(1) $X((i, j)) = i + j$ définit une variable aléatoire à valeurs dans $\{1, 2, \dots, 12\}$.

(2) $X(\omega) = \inf\{j : \omega_j = 6\}$, avec la convention $\inf \emptyset = \infty$, définit une v.a. à valeurs dans $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. Pour vérifier la mesurabilité, on observe que, pour tout $k \geq 1$,

$$X^{-1}(\{k\}) = \{\omega \in \Omega : \omega_1 \neq 6, \omega_2 \neq 6, \dots, \omega_{k-1} \neq 6, \omega_k = 6\}.$$

(3) Pour $t \in [0, 1]$ fixé, $X(\omega) = \omega(t)$ est une v.a. à valeurs dans \mathbb{R}^3 . (Remarquons que nous n'avons pas construit P dans cet exemple, mais cela n'intervient pas pour les questions de mesurabilité.)

Définition 8.1.2 La loi de la variable aléatoire X est la mesure-image de P par X . C'est donc la mesure de probabilité sur (E, \mathcal{E}) , notée P_X , définie par

$$P_X(B) = P(X^{-1}(B)) , \quad \forall B \in \mathcal{E}.$$

En pratique on écrit plutôt :

$$P_X(B) = P(X \in B) (= P(\{\omega \in \Omega : X(\omega) \in B\})).$$

La loi P_X permet de calculer la probabilité des événements qui "dépendent" de la v.a. X . Il faut comprendre qu'à chaque $\omega \in \Omega$ on a associé un "point aléatoire" $X(\omega)$ dans E , et que $P_X(B)$ est la probabilité que ce point aléatoire tombe dans B .

Remarque. Si μ est une mesure de probabilité sur \mathbb{R}^d , ou sur un espace plus général, il existe une manière canonique de construire une variable aléatoire dont la loi est μ . Il suffit de prendre $\Omega = \mathbb{R}^d$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^d)$, $P = \mu$, puis de poser $X(\omega) = \omega$. La loi de X est μ , de manière évidente.

Cas particuliers.

• **Variables aléatoires discrètes.** C'est le cas où E est dénombrable (et \mathcal{E} est l'ensemble des parties de E). La loi de X est alors

$$P_X = \sum_{x \in E} p_x \delta_x$$

où $p_x = P(X = x)$ et δ_x désigne la mesure de Dirac en x . En effet,

$$P_X(B) = P(X \in B) = P\left(\bigcup_{x \in B} \{X = x\}\right) = \sum_{x \in B} P(X = x) = \sum_{x \in E} p_x \delta_x(B).$$

En pratique, trouver la loi d'une v.a. discrète, c'est donc calculer toutes les probabilités $P(X = x)$ pour $x \in E$.

Exemple. Revenons à l'exemple (2) ci-dessus, avec $X(\omega) = \inf\{j : \omega_j = 6\}$. Alors, pour tout $k \geq 1$,

$$P(X = k) = P\left(\bigcup_{i_1, \dots, i_{k-1} \neq 6} \{\omega_1 = i_1, \dots, \omega_{k-1} = i_{k-1}, \omega_k = 6\}\right) = 5^{k-1} \left(\frac{1}{6}\right)^k = \frac{1}{6} \left(\frac{5}{6}\right)^{k-1}.$$

Remarquons que $\sum_{k=1}^{\infty} P(X = k) = 1$ et donc $P(X = \infty) = 1 - P(X \in \mathbb{N}) = 0$. Observons que l'ensemble $\{X = \infty\}$ est loin d'être vide puisqu'il contient toutes les suites (i_1, i_2, \dots) qui ne prennent pas la valeur 6.

• **Variables aléatoires à densité.** Une variable aléatoire X à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est dite à densité si P_X est absolument continue par rapport à la mesure de Lebesgue λ .

Dans ce cas, le théorème de Radon-Nikodym montre qu'il existe une fonction borélienne $p : \mathbb{R}^d \rightarrow \mathbb{R}_+$ telle que

$$P_X(B) = \int_B p(x) dx.$$

On a en particulier $\int_{\mathbb{R}^d} p(x) dx = P(X \in \mathbb{R}^d) = 1$. La fonction p , qui est unique à en ensemble de mesure de Lebesgue nulle près, est appelée la densité de (la loi de) X .

Si $d = 1$, on a en particulier, pour tous $\alpha \leq \beta$,

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} p(x) dx.$$

8.1.3 Espérance mathématique

Définition 8.1.3 Soit X une variable aléatoire réelle (i.e. à valeurs dans \mathbb{R}). On note alors

$$E[X] = \int_{\Omega} X(\omega) P(d\omega),$$

qui est bien définie dans les deux cas suivants :

- si $X \geq 0$ (alors $E[X] \in [0, \infty]$),
- si X est de signe quelconque et $E[|X|] = \int |X| dP < \infty$.

On étend cette définition au cas où $X = (X_1, \dots, X_d)$ est une variable aléatoire à valeurs dans \mathbb{R}^d en prenant alors $E[X] = (E[X_1], \dots, E[X_d])$, pourvu bien sûr que chacune des espérances $E[X_i]$ soit bien définie.

Remarque. Si $X = 1_B$, $E[X] = P(B)$. En général, $E[X]$ s'interprète comme la moyenne de la v.a. X . Dans le cas particulier où Ω est fini et P attribue la même valeur à chaque singleton, $E[X]$ est bien la moyenne au sens usuel des valeurs prises par X .

Proposition 8.1.1 Soit X une variable aléatoire à valeurs dans (E, \mathcal{E}) . Pour toute fonction mesurable $f : E \rightarrow [0, \infty]$, on a

$$E[f(X)] = \int_E f(x) P_X(dx).$$

Preuve. C'est évidemment une propriété générale des mesures-images déjà rencontrée dans le cours d'intégration. On remarque que le résultat est vrai par définition pour $f = 1_B$ puis par linéarité pour toute fonction étagée positive. Dans le cas général, on utilise le théorème de convergence monotone et le fait que toute fonction mesurable positive est limite croissante d'une suite de fonctions étagées positives. \square

Si f est de signe quelconque, la formule de la proposition reste vraie à condition que les intégrales soient bien définies, ce qui revient à $E[|f(X)|] < \infty$.

La donnée de P_X permet donc de calculer la valeur moyenne de variables aléatoires de la forme $f(X)$. Inversement, on utilise souvent la proposition pour calculer la loi d'une v.a. X : si on arrive à écrire

$$E[f(X)] = \int f d\nu$$

pour toute fonction f "suffisamment" générale, alors on peut identifier ν à la loi de X . Donnons un exemple simple de ce principe.

Proposition 8.1.2 Soit $X = (X_1, \dots, X_d)$ une v.a. à valeurs dans \mathbb{R}^d . Supposons que la loi de X a une densité $p(x_1, \dots, x_d)$. Alors, pour tout $j \in \{1, \dots, d\}$, la loi de X_j a une densité donnée par

$$p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_d) dx_1 \dots dx_{j-1} dx_{j+1} \dots dx_d$$

(par exemple, si $d = 2$,

$$p_1(x) = \int_{\mathbb{R}} p(x, y) dy, \quad p_2(y) = \int_{\mathbb{R}} p(x, y) dx).$$

Preuve. Soit π_j la projection $\pi_j(x_1, \dots, x_d) = x_j$. En utilisant le théorème de Fubini, on écrit, pour toute fonction borélienne $f : \mathbb{R} \rightarrow \mathbb{R}_+$,

$$\begin{aligned} E[f(X_j)] = E[f(\pi_j(X))] &= \int_{\mathbb{R}^d} f(x_j) p(x_1, \dots, x_d) dx_1 \dots dx_d \\ &= \int_{\mathbb{R}^d} f(x_j) \left(\int_{\mathbb{R}^{d-1}} p(x_1, \dots, x_d) dx_1 \dots dx_{j-1} dx_{j+1} \dots dx_d \right) dx_j \\ &= \int_{\mathbb{R}} f(x_j) p_j(x_j) dx_j, \end{aligned}$$

ce qui donne le résultat voulu. \square

Remarque. Si $X = (X_1, \dots, X_d)$ est une v.a. à valeurs dans \mathbb{R}^d , les lois P_{X_j} , qu'on appelle souvent les lois marginales de X , sont déterminées par la loi de X , simplement parce que $P_{X_j} = \pi_j(P_X)$, avec la notation ci-dessous. Il est important d'observer que :

la réciproque est fautive !

Pour un exemple, considérons une densité de probabilité q sur \mathbb{R} , et observons que la fonction $p(x_1, x_2) = q(x_1)q(x_2)$ est alors aussi une densité de probabilité sur \mathbb{R}^2 . D'après une remarque ci-dessus on peut construire une v.a. $X = (X_1, X_2)$ à valeurs dans \mathbb{R}^2 dont la loi est la mesure de densité p par rapport à la mesure de Lebesgue. Mais alors les deux v.a. X et $X' = (X_1, X_1)$ ont mêmes lois marginales (la proposition ci-dessus montre que $P_{X_1}(dx) = P_{X_2}(dx) = q(x)dx$) alors que les lois P_X et $P_{X'}$ sont très différentes, simplement parce que $P_{X'}$ est portée par la diagonale de \mathbb{R}^2 , qui est de mesure de Lebesgue nulle.

8.1.4 Exemple : le paradoxe de Bertrand

Pour illustrer les notions introduites dans les paragraphes précédents, considérons le problème suivant. On s'intéresse à la probabilité qu'une corde choisie au hasard sur un cercle ait une longueur plus grande que le côté du triangle équilatéral inscrit. Sans perte de généralité on peut supposer que le cercle est le cercle unité. Bertrand proposait deux méthodes de calcul :

- (a) On choisit les deux extrémités de la corde au hasard sur le cercle. La première étant choisie, la longueur de la corde sera plus grande que le côté du triangle équilatéral inscrit si et seulement si la seconde extrémité est dans un secteur angulaire d'ouverture $2\pi/3$. La probabilité est donc $\frac{2\pi/3}{2\pi} = \frac{1}{3}$.
- (b) On choisit le centre de la corde au hasard sur le disque unité. La probabilité désirée est la probabilité que le centre tombe dans le disque de rayon $1/2$ centré à l'origine. Comme l'aire de ce disque est un quart de l'aire du disque unité, on trouve comme probabilité $\frac{1}{4}$.

On obtient donc un résultat différent dans les deux cas. L'explication tient dans le fait que les deux méthodes correspondent à des expériences aléatoires différentes, représentées par des choix différents de l'espace de probabilité. Il n'y a donc aucune raison pour que la loi de la variable aléatoire que l'on considère (la longueur de la corde) soit la même dans les deux cas. Pour nous en convaincre, explicitons les choix des espaces de probabilité.

(a) Dans ce cas,

$$\Omega = [0, 2\pi]^2, \quad \mathcal{A} = \mathcal{B}([0, 2\pi]^2), \quad P(d\omega) = \frac{1}{4\pi^2} d\theta d\theta',$$

où on note $\omega = (\theta, \theta')$ pour $\omega \in \Omega$. La longueur de la corde est

$$X(\omega) = 2 \left| \sin\left(\frac{\theta - \theta'}{2}\right) \right|.$$

On calcule facilement la loi de X :

$$\begin{aligned}
 E[f(X)] &= \int_{\Omega} f(X(\omega)) P(d\omega) \\
 &= \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} f(2|\sin(\frac{\theta - \theta'}{2})|) d\theta d\theta' \\
 &= \frac{1}{\pi} \int_0^{\pi} f(2\sin(\frac{u}{2})) du \\
 &= \frac{1}{\pi} \int_0^2 f(x) \frac{1}{\sqrt{1 - \frac{x^2}{4}}} dx.
 \end{aligned}$$

Donc X est une v.a. réelle à densité : $P_X(dx) = p(x)dx$, avec

$$p(x) = \frac{1}{\pi} \frac{1}{\sqrt{1 - \frac{x^2}{4}}} 1_{[0,2]}(x).$$

En particulier, la probabilité recherchée est

$$P(X \geq \sqrt{3}) = \int_{\sqrt{3}}^2 p(x) dx = \frac{1}{3}.$$

(b) Maintenant,

$$\Omega = \{\omega = (y, z) \in \mathbb{R}^2 : y^2 + z^2 < 1\}, \quad \mathcal{A} = \mathcal{B}(\Omega), \quad P(d\omega) = \frac{1}{\pi} 1_{\Omega}(y, z) dy dz.$$

La longueur de la corde est

$$X(\omega) = 2\sqrt{1 - y^2 - z^2}$$

et pour calculer sa loi on écrit

$$\begin{aligned}
 E[f(X)] &= \frac{1}{\pi} \int_{\mathbb{R}^2} f(2\sqrt{1 - y^2 - z^2}) 1_{\{y^2+z^2 < 1\}} dy dz \\
 &= 2 \int_0^1 f(2\sqrt{1 - r^2}) r dr \\
 &= \frac{1}{2} \int_0^2 f(x) x dx.
 \end{aligned}$$

Donc $P_X(dx) = p(x)dx$, avec

$$p(x) = \frac{1}{2} 1_{[0,2]}(x) x dx.$$

On peut remarquer que la densité obtenue est très différente de celle du cas (a). En particulier,

$$P(X \geq \sqrt{3}) = \int_{\sqrt{3}}^2 p(x) dx = \frac{1}{4}.$$

Exercice. Traiter le cas de la troisième méthode proposée par Bertrand : on choisit au hasard la direction du rayon orthogonal à la corde, puis le centre de la corde uniformément sur ce rayon.

8.1.5 Lois classiques

On donne dans ce paragraphe quelques exemples importants de lois.

Lois discrètes.

- (a) *Loi uniforme.* Si E est un ensemble fini, $\text{Card}(E) = n$, une v.a. X est de loi uniforme sur E si

$$P(X = x) = \frac{1}{n}, \quad \forall x \in E.$$

- (b) *Loi de Bernoulli de paramètre $p \in [0, 1]$.* C'est la loi d'une v.a. X à valeurs dans $\{0, 1\}$ telle que

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

On interprète X comme le résultat du lancer d'une pièce truquée qui tombe sur pile avec probabilité p .

- (c) *Loi binômiale $\mathcal{B}(n, p)$ ($n \in \mathbb{N}^*$, $p \in [0, 1]$).* C'est la loi d'une v.a. X à valeurs dans $\{1, \dots, n\}$ telle que

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}.$$

On interprète X comme le nombre de piles obtenus en n lancers avec la pièce précédente.

- (d) *Loi géométrique de paramètre $p \in]0, 1[$.* C'est la loi d'une v.a. X à valeurs dans \mathbb{N} , telle que

$$P(X = k) = (1 - p) p^k.$$

X est le nombre de piles obtenus avant le premier face.

- (e) *Loi de Poisson de paramètre $\lambda > 0$.* X est une v.a. à valeurs dans \mathbb{N} , et

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

On calcule facilement $E[X] = \lambda$. La loi de Poisson est très importante aussi bien du point de vue théorique que dans les applications. Intuitivement, elle correspond au nombre d'événements rares qui se sont produits durant une période longue. La traduction mathématique de cette intuition est l'*approximation binômiale de la loi de Poisson* : si pour tout $n \geq 1$, X_n suit une loi binômiale $\mathcal{B}(n, p_n)$ et si $np_n \rightarrow \lambda$ quand $n \rightarrow \infty$, alors pour tout entier $k \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Lois continues. Dans les trois exemples qui suivent, X est une v.a. à valeurs dans \mathbb{R} , à densité $p(x)$.

- (a) *Loi uniforme sur $[a, b]$ ($a < b$).*

$$p(x) = \frac{1}{b - a} 1_{[a, b]}(x).$$

(b) *Loi exponentielle de paramètre $\lambda > 0$.*

$$p(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Les lois exponentielles possèdent la propriété caractéristique suivante : si $a, b > 0$,

$$P(X > a + b) = P(X > a) P(X > b),$$

ce qu'on interprète en disant que la probabilité que $X - a > b$ sachant que $X > a$ coïncide avec la probabilité que $X > b$. C'est la propriété d'absence de mémoire de la loi exponentielle, qui explique qu'elle soit utilisée par exemple pour modéliser les temps de vie de machine sans usure.

(c) *Loi gaussienne, ou normale, $\mathcal{N}(m, \sigma^2)$ ($m \in \mathbb{R}, \sigma > 0$).*

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Avec la loi de Poisson, c'est la loi la plus importante en théorie des probabilités. Sa densité est la fameuse courbe en cloche. Les paramètres m et σ s'interprètent comme

$$m = E[X], \quad \sigma^2 = E[(X - m)^2].$$

On remarque aussi que $X - m$ suit la loi $\mathcal{N}(0, \sigma^2)$. La loi gaussienne jouera un rôle important dans le Chapitre 10.

Par convention on dira qu'une v.a. constante égale à m suit la loi gaussienne $\mathcal{N}(m, 0)$. Si X suit la loi $\mathcal{N}(m, \sigma^2)$, pour tous $\lambda, \mu \in \mathbb{R}$, $\lambda X + \mu$ suit la loi $\mathcal{N}(\lambda m + \mu, \lambda^2 \sigma^2)$.

8.1.6 Fonction de répartition d'une variable aléatoire réelle

Si X est une v.a. réelle, la *fonction de répartition* de X est la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

$$F_X(t) = P(X \leq t) = P_X([\!-\infty, t]), \quad \forall t \in \mathbb{R}.$$

La fonction F_X est croissante, continue à droite et a pour limite 0 en $-\infty$ et 1 en $+\infty$.

Inversement, si on se donne une fonction F ayant ces propriétés, on a vu dans le cours d'intégration qu'il existe une (unique) mesure de probabilité μ telle que $\mu([\!-\infty, t]) = F(t)$ pour tout $t \in \mathbb{R}$. Cela montre qu'on peut interpréter F comme la fonction de répartition d'une v.a. réelle.

Il découle des résultats du cours d'intégration que F_X caractérise la loi P_X de X . On a en particulier

$$\begin{aligned} P(a \leq X \leq b) &= F_X(b) - F_X(a-) && \text{si } a \leq b, \\ P(a < X < b) &= F_X(b-) - F_X(a) && \text{si } a < b, \end{aligned}$$

et les sauts de F_X correspondent aux atomes de P_X .