
Panorama des probabilités de Licence¹

Dans ces notes, on rappelle les principales notions probabilistes. Les résultats sont cités ici sans preuve. On renvoie à tout cours de niveau L3 Mathématiques pour une présentation plus détaillée, par exemple [BL], [Ouv] ou [JCB-proba] (web) pour les probabilités et [BP] ou [JCB-mesure] (web) pour des notions de théorie de la mesure.

1 Rappels de théorie de la mesure

Comme son nom l'indique, une mesure mesure des ensembles. Des exemples typiques de mesure sont le cardinal (d'un ensemble discret), la longueur (d'une courbe), l'aire d'une surface, la probabilité (d'un évènement). En général, une mesure ne peut pas mesurer n'importe quel objet mais seulement certains objets adaptés à la nature de la mesure. Mathématiquement, l'ensemble des objets d'un ensemble X que peut mesurer une mesure s'appelle une tribu (ou σ -algèbre). Cette famille (de sous-ensembles de X) doit vérifier certaines propriétés de stabilités :

Définition 1.1 $\mathcal{A} \subset \mathcal{P}(X)$ est une **tribu** (ou une σ -algèbre) lorsque les conditions suivantes sont satisfaites

- $X \in \mathcal{A}$;
- \mathcal{A} est stable par complémentaire : si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$;
- \mathcal{A} est stable par réunion dénombrable : lorsque $A_n \in \mathcal{A}$ pour tout $n \geq 1$ alors $\bigcup_{n \geq 1} A_n \in \mathcal{A}$.

On appelle (ensemble) mesurable tout ensemble A élément d'une tribu \mathcal{A} . Un ensemble muni d'une tribu (X, \mathcal{A}) s'appelle un espace mesurable.

L'ensemble $\mathcal{P}(X)$ de toutes les parties de X est bien entendu une tribu de X (tribu discrète). C'est la plus grosse tribu possible mais justement elle est trop grosse et on peut avoir des difficultés pour définir facilement une mesure sur une famille aussi large. À l'opposé, la famille $\{\emptyset, X\}$ est la plus petite tribu possible (tribu grossière) mais elle trop petite pour qu'il soit vraiment intéressant de considérer une mesure sur cette si petite tribu. On considère souvent des tribus intermédiaires. Par exemple sur un espace topologique (typiquement \mathbb{R}), on considère une tribu qui tient compte de la structure topologique déjà existante, il s'agit de la tribu borélienne.

Définition 1.2 Lorsque X est un espace topologique (c'est à dire muni d'une famille d'ouverts), la plus petite tribu contenant tous les ouverts est appelée **tribu borélienne**. Elle est notée $\mathcal{B}(X)$.

1. Auteur : [Jean-Christophe Breton](#)

Les mesurables $A \in \mathcal{B}(X)$ s'appellent aussi les boréliens. D'après la définition, des boréliens typiques sont les ouverts, les fermés.

En général, sur \mathbb{R} , on travaillera avec la tribu borélienne $\mathcal{B}(\mathbb{R})$ et sur les espaces discrets X (comme \mathbb{N}, \mathbb{Z}), on peut considérer la plus grosse tribu $\mathcal{P}(X)$.

De façon générale la tribu $\sigma(\mathcal{M})$ engendrée par une famille \mathcal{M} de parties de X est la plus petite tribu de X contenant \mathcal{M} . C'est l'intersection de toutes les tribus de X contenant \mathcal{M} .

Sur des espaces munis de tribus, on considère des fonctions qui se comportent bien vis à vis des tribus de ces espaces :

Définition 1.3 Une application $f : (X, \mathcal{A}) \longrightarrow (Y, \mathcal{B})$ est dite **mesurable** si

$$\forall B \in \mathcal{B}, \quad f^{-1}(B) \in \mathcal{A}.$$

L'ensemble des fonctions mesurables est stable par la plupart des opérations usuelles (addition $+$, soustraction $-$, multiplication par un scalaire \cdot , multiplication \times , quotient $/$, composition \circ , passage à la limite \lim).

Exemple 1.4 Lorsqu'on travaille avec les tribus boréliennes, les fonctions $f : X \longrightarrow Y$ continues sont mesurables.

Définition 1.5 Une **mesure** μ sur (X, \mathcal{A}) est une application de $\mathcal{A} \rightarrow [0, +\infty]$ telle que

- $\mu(\emptyset) = 0$;
- si $(A_n)_{n \geq 1}$ est une suite dénombrable d'ensembles de \mathcal{A} deux à deux disjoints alors

$$\mu\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mu(A_n) \quad (\text{propriété de } \sigma\text{-additivité}).$$

Le triplet (X, \mathcal{A}, μ) est appelé un **espace mesuré** (espace mesurable + mesure).

Exemple 1.6 (Mesures classiques)

- Mesure de **Dirac** sur $(X, \mathcal{P}(X))$: soit $a \in X$,

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

- Mesure de **comptage** sur $(X, \mathcal{P}(X))$: c'est la mesure η qui donne le cardinal :

$$\eta(A) = \text{card}(A).$$

- Mesure de **Lebesgue** sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$: c'est la mesure qui généralise la notion de longueur des intervalles aux boréliens :

$$\lambda([a, b]) = b - a, \quad \lambda(A + x) = \lambda(A).$$

- **Mesure image** : Soit $f : (X, \mathcal{A}, \mu) \rightarrow (Y, \mathcal{B})$ une fonction mesurable. On définit sur (Y, \mathcal{B}) la mesure image de f notée μ_f par :

$$\mu_f(B) = \mu(f^{-1}(B)). \quad (1)$$

- Une mesure μ est **σ -finie** sur (X, \mathcal{A}) si

$$X = \bigcup_{n \geq 1} X_n \quad \text{avec } X_n \in \mathcal{A} \text{ et } \mu(X_n) < +\infty. \quad (2)$$

- Mesure de **probabilités** : c'est une mesure μ de poids $\mu(X) = 1$.
- Une mesure ν est dite **absolument continue par** rapport à μ sur (X, \mathcal{A}) si $\mu(A) = 0$ implique $\nu(A) = 0$. Lorsque μ, ν sont σ -finies sur (X, \mathcal{A}) (cf. (2)), le théorème de Radon-Nikodym donne l'existence d'une fonction f \mathcal{A} -mesurable telle que

$$\nu(A) = \int_A f d\nu, \quad \int_A g d\nu = \int_A gf d\mu. \quad (3)$$

La fonction $f = \frac{d\nu}{d\mu}$ s'appelle la dérivée de Radon-Nikodym. Ce résultat est à l'origine de la notion de variable aléatoire à densité, cf. Déf. 2.10.

Intégrales

La théorie de la mesure permet de définir des intégrales par rapport à une mesure μ abstraite (comme en Déf. 1.5). Ainsi si $f : (X, \mathcal{A}, \mu) \rightarrow \mathbb{R}$ est mesurable (plus précisément, il faut f positive ou intégrable), on construit l'intégrale

$$\int_X f d\mu = \int_X f(x)\mu(dx).$$

On utilise indifféremment l'une ou l'autre notation, la deuxième insistant sur l'identité de la variable (muette) d'intégration (ici : x). Ainsi

- si $X = \mathbb{N}$ et $\mu = \eta$ la mesure de comptage, l'intégrale devient une somme : $\int_X f d\eta = \sum_{n \in \mathbb{N}} f(n)$;
- si $X = \mathbb{R}$ et $\mu = \lambda$, l'intégrale "devient" une intégrale de Riemann habituelle $\int_X f d\mu = \int_{\mathbb{R}} f(x)dx$ (au moins pour les fonctions usuelles) ;
- si $X = \Omega$ et $\mu = \mathbb{P}$, l'intégrale devient une espérance probabiliste $\int_{\Omega} f d\mathbb{P} = \mathbb{E}[f]$.

Produits de mesures

Si (X, \mathcal{A}) et (Y, \mathcal{B}) sont deux espaces, on peut considérer l'espace produit $X \times Y = \{(x, y) : x \in X, y \in Y\}$. On le munit de la tribu produit $\mathcal{A} \otimes \mathcal{B}$ qui est (en gros) constituée des produits d'ensemble $A \times B$ pour $A \in \mathcal{A}$ et $B \in \mathcal{B}$. Si μ est une mesure sur X et ν sur Y , on peut considérer la **mesure produit** $\mu \otimes \nu$ définie sur $\mathcal{A} \otimes \mathcal{B}$ par

$$(\mu \otimes \nu)(A \otimes B) = \mu(A)\nu(B).$$

Par un argument de classes monotones, on montre que $(\mu \otimes \nu)$ ainsi définie est effectivement bien définie sur $\mathcal{A} \otimes \mathcal{B} = \sigma(A \times B : A \in \mathcal{A}, B \in \mathcal{B})$.

Par exemple si $X = Y = \mathbb{R}$ et $\mu = \nu = \lambda$ (la mesure de Lebesgue), on construit sur $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ la mesure $\lambda \otimes \lambda = \lambda_2$, il s'agit de la mesure de Lebesgue en dimension 2. Celle en dimension 1 mesure les longueurs, celle en dimension 2 mesure les surfaces. De même, en dimension 3, la mesure de Lebesgue λ_3 mesure les volumes. On définit de la sorte λ_d sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Les intégrales par rapport aux mesures produits (σ -finies) s'expriment comme des intégrales multiples par les théorèmes de Fubini : si μ, ν sont σ -finies (cf. (2)) alors

$$\int_{X \times Y} f(x, y)(\mu \otimes \nu)(dx, dy) = \int_X \left(\int_Y f(x, y)\nu(dy) \right) \mu(dx) = \int_Y \left(\int_X f(x, y)\mu(dx) \right) \nu(dy) \quad (4)$$

lorsque f est $(\mathcal{A} \otimes \mathcal{B})$ -mesurable et positive (Fubini-Tonelli) ou $(\mu \otimes \nu)$ -intégrable (Fubini).

Si μ et ν sont deux mesures sur le même espace (vectoriel), on peut considérer un autre type de produit : le **produit de convolution**. Il est défini pour tout $A \in \mathcal{A}$ par

$$(\mu * \nu)(A) = \int_X \mu(A - x)\nu(dx).$$

Noter que $\mu * \nu = \nu * \mu$ et $\mu * \delta_0 = \mu$. De plus, si $\mu(dx) = f(x)dx$ et $\nu(dx) = g(x)dx$ alors $\mu * \nu$ a pour densité $f * g$ donnée par $(f * g)(x) = \int f(y)g(x - y)dy$

Limites monotones de probabilités

Soit $(A_n)_{n \geq 1}$ une suite d'évènements croissants $A_n \subset A_{n+1}$ alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n \geq 1} A_n\right).$$

Soit $(B_n)_{n \geq 1}$ une suite d'évènements décroissants $B_n \supset B_{n+1}$ alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcap_{n \geq 1} B_n\right).$$

On rappelle ci-dessous, dans un contexte probabiliste, les résultats clés de convergence des intégrales de Lebesgue : théorème de convergence monotone (Th. 7.4), lemme de Fatou (Lemme 7.5), théorème de convergence dominée (Th. 7.6).

2 Variable aléatoire

La probabilité d'un évènement est une certaine mesure de l'évènement. Les probabilités entrent donc naturellement dans le cadre de la théorie de la mesure.

Définition 2.1 Un espace de probabilité est un espace mesurable (Ω, \mathcal{F}) muni d'une mesure de probabilité \mathbb{P} , c'est à dire une mesure de masse totale 1 : $\mathbb{P}(\Omega) = 1$. Les ensembles mesurables $A \in \mathcal{F}$ sont appelés les évènements².

La tribu \mathcal{F} s'interprète comme l'ensemble de tous les évènements.

Définition 2.2 On appelle **variable aléatoire** (v.a.) toute application mesurable X d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R} muni de la tribu borélienne $\mathcal{B}(\mathbb{R})$. Si l'application est à valeurs dans \mathbb{R}^d , X est un vecteur aléatoire (couple si $d = 2$).

Une variable aléatoire est donc tout simplement une application mesurable sur un espace de probabilité.

Définition 2.3 Soit $X : \Omega \rightarrow \mathbb{R}$, on appelle **loi** de X la mesure \mathbb{P}_X , mesure image sur \mathbb{R} de \mathbb{P} par X , comme en (1) :

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\omega \in \Omega : X(\omega) \in A), \quad A \in \mathcal{B}(\mathbb{R}).$$

Dans la suite, on note \sim pour signifier "a pour loi". Ainsi, $X \sim Y$ signifie X, Y ont même loi, $X \sim \mathcal{P}(\lambda)$ signifie X a pour loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ , cf. ci-dessous.

La loi \mathbb{P}_X d'une variable aléatoire X définit alors une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Il est facile en effet de vérifier que $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = 1$ et que \mathbb{P}_X est σ -additive. Si X est un vecteur aléatoire à valeurs dans \mathbb{R}^d , sa loi est une mesure (de probabilité) sur \mathbb{R}^d .

Si (X, Y) est un couple aléatoire, X est appelée la première **marginale** et Y la seconde. Leurs lois se retrouvent par intégration partielle de la loi $\mathbb{P}_{X,Y}$ du couple :

$$\mathbb{P}_X(A) = \mathbb{P}_{X,Y}(A \times \mathbb{R}), \quad \mathbb{P}_Y(B) = \mathbb{P}_{X,Y}(\mathbb{R} \times B).$$

Attention sans information supplémentaire, on ne retrouve pas la loi $\mathbb{P}_{X,Y}$ à partir des lois \mathbb{P}_X et \mathbb{P}_Y . Un cas particulier important est lorsque les variables aléatoires X et Y sont indépendantes, cf. (10).

Définition 2.4 Soit X une variable aléatoire.

- On appelle **atome** de X (ou de sa loi) tout $a \in \mathbb{R}$ tel que $\mathbb{P}(X = a) \neq 0$.
- On appelle **support** (topologique) d'une variable aléatoire $X : (\Omega, \mathcal{F}) \rightarrow (B, \mathcal{B})$ le plus petit ensemble F fermé de B tel que $\mathbb{P}(X \in F) = 1$. Avec un abus de notation, on notera dans la suite $\mathcal{S}(X)$ ce support.

On peut associer à une variable aléatoire X une sous-tribu $\sigma(X)$ de la tribu de base \mathcal{F} de l'espace de probabilité.

Définition 2.5 La tribu engendrée par X , notée $\sigma(X)$, est la plus petite tribu \mathcal{G} sur Ω qui rend l'application $X : (\Omega, \mathcal{G}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable.

2. Pour l'accent voir ce [lien](#)

Il faut interpréter $\sigma(X)$ comme l'ensemble de tous les évènements qui s'expriment à partir de la variable aléatoire X . Cela représente en quelque sorte toute l'information véhiculée par X .

Lorsqu'on considère plusieurs variables aléatoires X_1, \dots, X_n , on peut considérer la tribu $\sigma(X_1, \dots, X_n)$ qui est la plus petite tribu à rendre en même temps chaque X_i mesurable. On montre que $\sigma(X_1) \cup \dots \cup \sigma(X_n) \subset \sigma(X_1, \dots, X_n)$. Mais il n'y a pas égalité (en général une union de tribus n'est pas une tribu). La tribu $\sigma(X_1, \dots, X_n)$ représente toute l'information véhiculée par X_1, X_2, \dots et X_n .

Définition 2.6 On appelle **fonction de répartition** d'une variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ la fonction F_X définie sur \mathbb{R} par

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X(]-\infty, x]).$$

Proposition 2.7 (Propriétés de la fonction de répartition)

1. F_X est croissante.
2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$.
3. F_X est continue à droite et admet une limite à gauche :

$$\lim_{x \searrow x_0} F_X(x) = F_X(x_0), \quad F_X(x^-) := \lim_{x \nearrow x_0} F_X(x) = \mathbb{P}(X < x_0).$$

On dit que la fonction F_X est càdlàg (acronyme de continue à droite avec une limite à gauche).

4. En fait si x n'est pas un atome de X , alors F_X est continue à gauche (donc continue) en x .

Le dernier point suit de l'observation plus générale : $F_X(x) - F_X(x^-) = \mathbb{P}(X = x)$.

Variable aléatoire discrète

Définition 2.8 Une variable aléatoire X est **discrète** si elle est à valeurs dans un ensemble au plus dénombrable (en bijection avec une partie de \mathbb{N}). Autrement dit : $\mathcal{S}(X)$ est fini ou dénombrable (on peut compter ses éléments).

Le support d'une variable aléatoire X discrète est l'ensemble de ses atomes $\mathcal{S}(X) = \{a_i : i \in I\}$ (où I est fini ou dénombrable). La loi d'une telle variable aléatoire est la somme des mesures de Dirac en ses atomes a_i pondérées par son importance probabiliste p_i :

$$\mathbb{P}_X = \sum_{i \in I} p_i \delta_{a_i}.$$

Autrement dit, la loi de X est donnée par son support $\mathcal{S}(X) = \{a_i : i \in I\}$ et ses probabilités ponctuelles $p_i = \mathbb{P}(X = a_i)$, $i \in I$. Noter que $\sum_{i \in I} p_i = 1$.

Si (X, Y) est un couple de variables aléatoires discrètes, la loi de X s'obtient à partir de celle du couple en sommant partiellement les probabilités ponctuelles :

$$\mathbb{P}_X(x) = \sum_{y \in \mathcal{S}(Y)} \mathbb{P}_{X,Y}(x, y).$$

Variabes aléatoires discrètes usuelles

- Si $X = c$ est une variable aléatoire **constante**, alors sa loi est $\mathbb{P}_X = \delta_c$ et $\mathcal{S}(X) = \{c\}$.
En effet

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(c \in A) = \delta_c(A).$$

- Soit X une variable aléatoire qui prend les valeurs a_1, \dots, a_n avec les probabilités respectives p_1, \dots, p_n et $p_1 + \dots + p_n = 1$. Alors son support est $\mathcal{S}(X) = \{a_1, \dots, a_n\}$ et sa loi est donnée par :

$$\mathbb{P}_X = p_1 \delta_{a_1} + \dots + p_n \delta_{a_n}.$$

Ici $p_i = \mathbb{P}(X = a_i)$. C'est le cas général d'une loi discrète à support fini.

- Une variable aléatoire X suit la loi **uniforme** ou **équirépartie** sur un ensemble fini $\{x_1, \dots, x_n\}$ si $\mathcal{S}(X) = \{a_1, \dots, a_n\}$ et

$$\mathbb{P}_X = \sum_{i=1}^n \frac{1}{n} \delta_{a_i}.$$

Par exemple, la variable aléatoire X qui indique la face $1, \dots, 6$ obtenue par le lancer d'un dé (à six faces équilibrées) est équirépartie sur $\{1, 2, \dots, 6\}$: $\mathbb{P}_X = \frac{1}{6} \delta_1 + \dots + \frac{1}{6} \delta_6$.

- Soit X une variable aléatoire qui modélise le succès (par un 1) ou l'échec (par un 0) lors d'une expérience aléatoire. Son support est donc $\mathcal{S}(X) = \{0, 1\}$ et sa loi est

$$\mathbb{P}_X = p \delta_1 + (1 - p) \delta_0$$

où p est la probabilité de succès. Il s'agit de la loi de **Bernoulli** $b(p)$.

Exemple : pile ou face avec $p = 1/2$ si la pièce est équilibrée, $p \neq 1/2$ si elle est truquée.

- La variable aléatoire X qui indique le nombre de succès dans une suite de n épreuves indépendantes chacune ayant une probabilité p de succès est discrète $\mathcal{S}(X) = \{0, \dots, n\}$.

$$\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k.$$

Il s'agit de la loi **binomiale** $\mathcal{B}(n, p)$ où p est la probabilité de succès à chaque épreuve. On rappelle que $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ est le coefficient binomial et désigne le nombre de choix pour sélectionner k objets parmi n .

- La variable aléatoire X qui indique le rang du premier succès dans une suite d'expériences indépendantes où chacune a une probabilité p de succès suit la loi **géométrique** $\mathcal{G}(p)$ donnée par $\mathcal{S}(X) = \mathbb{N}^*$ et

$$\mathbb{P}_X = \sum_{k=1}^{+\infty} (1-p)^{k-1} p \delta_k. \quad (5)$$

Une version continue de cette loi est donnée par la loi exponentielle, cf. Exemple 2.11.

- Une variable aléatoire X suit la loi de **Poisson** $\mathcal{P}(\lambda)$ si $\mathcal{S}(X) = \mathbb{N}$ et

$$\mathbb{P}_X = \sum_{k=0}^{+\infty} \frac{e^{-\lambda} \lambda^k}{k!} \delta_k.$$

Cette loi sert à modéliser de petits effectifs. C'est un cas limite de loi binomiale : avec la notion de convergence en loi rappelée en Déf. 7.1, on a $\mathcal{B}(n, \lambda/n) \Rightarrow \mathcal{P}(\lambda)$, $n \rightarrow +\infty$ (théorème de Poisson).

Proposition 2.9 *Soit X une variable aléatoire discrète de support $\mathcal{S}(X) = \{a_i : i \in I\}$ avec les probabilités ponctuelles $p_i, i \in I$. La fonction de répartition F_X est constante par morceaux avec des sauts de taille p_i en a_i :*

$$F_X(t) = \sum_{j \leq i} p_i \quad \text{si } t \in [a_i, a_{i+1}[.$$

Noter que pour des variables aléatoires discrètes, la probabilité de points peut ne pas être nulle (c'est le cas exactement pour les atomes). Conséquemment la nature des inégalités (strictes ou larges) dans les calculs de probabilité est importante dans le cas discret.

Variables aléatoires à densité

Définition 2.10 *Une variable aléatoire X est une variable aléatoire de **densité** f si pour tout borélien A*

$$\mathbb{P}(X \in A) = \int_A f(x) dx.$$

Autrement dit, une telle variable aléatoire X est de loi \mathbb{P}_X absolument continue par rapport à la mesure de Lebesgue ($\mathbb{P}_X \ll \lambda$) et sa densité f est la densité de Radon-Nikodym, cf. (3). On observe facilement que la densité f doit vérifier $f(x) \geq 0$ (car \mathbb{P}_X est une mesure positive) et $\int_{\mathbb{R}} f(x) dx = 1$ (car \mathbb{P}_X est une probabilité). Dans les calculs d'intégration, on a alors l'écriture symbolique $\mathbb{P}_X(dx) = f(x) dx$.

La même notion existe pour les vecteurs aléatoires dans \mathbb{R}^d , dans ce cas les densités sont des fonctions $f(x_1, \dots, x_d)$ à d variables :

$$\mathbb{P}_X(A) = \int_A f(x_1, \dots, x_d) dx_1 \dots dx_d, \quad A \in \mathcal{B}(\mathbb{R}^d).$$

Exemple 2.11 (Lois à densité usuelles)

- Loi **uniforme** sur $[a, b]$ borné : $X \sim \mathcal{U}([a, b])$ est de densité $f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$. Cette loi modélise les quantités uniformément répartie sur un domaine borné.
- Loi **exponentielle** de paramètre $\lambda > 0$: $X \sim \mathcal{E}(\lambda)$ est de densité $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}^+}(x)$. Cette loi modélise des durées de vie, des temps d'attente. Elle est caractérisée par une propriété d'absence de mémoire. C'est une version continue de la loi géométrique (5).
- Loi **normale standard** : $X \sim \mathcal{N}(0, 1)$ est de densité $f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$. C'est une loi centrale en probabilité notamment à cause du TCL (Th. 8.2) qui justifie qu'elle modélise le résultat de petites erreurs indépendantes.
- Loi **normale générale** ($m \in \mathbb{R}, \sigma^2 > 0$) : $X \sim \mathcal{N}(m, \sigma^2)$ est de densité

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

Noter que si $X \sim \mathcal{N}(m, \sigma^2)$ alors

$$\frac{X-m}{\sigma} \sim \mathcal{N}(0, 1) \quad (\text{centrer-réduire}). \quad (6)$$

- Loi de **Cauchy** de paramètre $a > 0$: $X \sim \mathcal{C}(a)$ est de densité $f(x) = \frac{a}{\pi(a^2+x^2)}$.
- Loi **Gamma** de paramètre $n \in \mathbb{N}^*, \lambda > 0$: $X \sim \gamma(n, \lambda)$ est de densité

$$f(x) = \frac{x^{n-1} \lambda^n e^{-\lambda x}}{\Gamma(n)} \mathbf{1}_{\mathbb{R}^+}(x).$$

Dans le cas à densité, la densité f est reliée à la fonction de répartition F_X de la façon suivante.

Proposition 2.12 *Si X est une variable aléatoire de densité f , sa fonction de répartition F_X vérifie :*

1. $\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x f(t) dt$.
2. F_X est continue sur \mathbb{R} .
3. Si f est continue au point x_0 , alors F_X est dérivable en x_0 de dérivée $F'_X(x_0) = f(x_0)$.

D'après 2), la fonction de répartition est toujours continue. De là, vient le nom qu'on donne parfois aux variables aléatoires à densité : variables aléatoires continues.

Par ailleurs, noter que, quand X est de densité f , pour tout $a \in \mathbb{R}, \mathbb{P}(X = a) = \int_a^a f(t) dt = 0$ car l'intégrale sur un intervalle réduit à un point est nulle. Pour les lois à densité, les probabilités de points sont nulles (en contraste avec les lois discrètes). Conséquemment, dans les calculs de probabilité, il est égale de considérer les inégalités larges ou strictes dans le cas à densité.

Si (X, Y) est un couple de densité $f(x, y)$, alors X est une variable aléatoire de densité

$$f_X(x, y) = \int_{\mathbb{R}} f(x, y) dy.$$

3 Espérance et variance

Définition 3.1

Moment : Une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ a un moment d'ordre $p \geq 1$ si et seulement si

$$\mathbb{E}[|X|^p] = \int_{\Omega} |X|^p d\mathbb{P} < +\infty.$$

Espérance : Si X a un moment d'ordre 1, l'espérance d'une variable aléatoire X est donnée par

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P}.$$

Si en plus $\mathbb{E}[X] = 0$, la variable aléatoire X est dite centrée.

Variance : Si X a un moment d'ordre 2, la variance de X est donnée par

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Écart-type : L'écart-type d'une variable aléatoire X qui admet une variance est donnée par $\sigma_X = \sqrt{\text{Var}(X)}$.

Remarque 3.2 L'espérance d'une variable aléatoire donne la valeur moyenne (au sens probabiliste) de la variable aléatoire. Sa variance (ou son écart-type) mesure la dispersion des valeurs de la variable aléatoire autour de sa moyenne. Ou encore, l'écart-type donne l'écart moyen de la variable aléatoire par rapport à sa moyenne. Il est équivalent de dire que la variance de X est finie et que X admet un moment d'ordre 2 fini.

Propriétés de l'espérance

Comme l'espérance n'est rien d'autre qu'une intégrale (par rapport à \mathbb{P}), des propriétés de l'intégration, on déduit pour des variables aléatoires intégrables X, Y et des réels a, b :

Proposition 3.3 Soit X, Y des variables aléatoires (intégrables), $a, b \in \mathbb{R}$.

- **Linéarité de \mathbb{E} :** $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
- **Inégalité de Markov :** Si X est une variable aléatoire positive avec un moment d'ordre un fini :

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

- **Transfert :** On peut exprimer l'espérance comme une intégrale sur \mathbb{R} par rapport à la loi de X :

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \mathbb{P}_X(dx).$$

De façon générale si $h(X)$ a un moment d'ordre 1 fini : $\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) \mathbb{P}_X(dx)$.

Cas d'une variable aléatoire discrète

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ avec $\mathcal{S}(X) = \{a_i : i \in I\}$ et des probabilités ponctuelles $p_i, i \in I$. La loi de X est $\mathbb{P}_X = \sum_{i \in I} p_i \delta_{a_i}$. Alors X est intégrable si et seulement si $\mathbb{E}[|X|] = \sum_{i \in I} |a_i| p_i < +\infty$ et dans ce cas

$$\mathbb{E}[X] = \sum_{i \in I} a_i p_i.$$

De même, si $h : \mathbb{R} \rightarrow \mathbb{R}$ est mesurable et $h(X)$ est intégrable si et seulement si $\sum_{i \in I} |h(a_i)| p_i < +\infty$. Son espérance est alors $\mathbb{E}[h(X)] = \sum_{i \in I} h(a_i) p_i$.

Exemple 3.4 (Espérance des variables aléatoires discrètes classiques)

- $X \sim \delta_c$ (ie. une variable aléatoire constante $X = c$) : $\mathbb{E}[X] = c$.
- $X \sim b(p)$: $\mathbb{E}[X] = p$.
- X uniforme sur $\{x_1, \dots, x_n\}$: $\mathbb{E}[X] = \frac{x_1 + \dots + x_n}{n}$.
- $X \sim \mathcal{B}(n, p)$: $\mathbb{E}[X] = np$.
- $X \sim \mathcal{G}(p)$: $\mathbb{E}[X] = \frac{1}{p}$.
- $X \sim \mathcal{P}(\lambda)$: $\mathbb{E}[X] = \lambda$.

Cas d'une variable aléatoire à densité

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ une variable de densité f . La loi de X est donnée par $\mathbb{P}_X(dx) = f(x)dx$. Alors X est intégrable si et seulement si $\mathbb{E}[|X|] = \int_{\mathbb{R}} |x| f(x) dx < +\infty$ et dans ce cas

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx.$$

De même, si $h : \mathbb{R} \rightarrow \mathbb{R}$ est mesurable, alors $h(X)$ est intégrable si et seulement si $\int_{\mathbb{R}} |h(x)| f(x) dx < +\infty$ et son espérance est alors $\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) f(x) dx$.

Exemple 3.5 (Espérances des variables aléatoires à densité classiques)

- $X \sim \mathcal{U}([a, b])$ (avec $-\infty < a < b < +\infty$) : $\mathbb{E}[X] = (b + a)/2$.
- $X \sim \mathcal{E}(\lambda)$: $\mathbb{E}[X] = 1/\lambda$.
- $X \sim \mathcal{C}(a)$: il n'y a pas d'espérance car $\mathbb{E}[|X|] = \int_{\mathbb{R}} \frac{a|x|}{\pi(a^2+x^2)} dx = +\infty$.
- $X \sim \mathcal{N}(0, 1)$: $\mathbb{E}[X] = 0$.
- $X \sim \mathcal{N}(m, \sigma^2)$: $\mathbb{E}[X] = m$.
- $X \sim \gamma(n, \lambda)$: $\mathbb{E}[X] = n/\lambda$.

Propriétés de la variance

La variance est un opérateur quadratique non linéaire.

Proposition 3.6 — $\text{Var}(X) \geq 0$.

- **Formule de Koenig** : $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
- $\text{Var}(aX) = a^2 \text{Var}(X)$.
- $\text{Var}(X + b) = \text{Var}(X)$ pour toute constante $b \in \mathbb{R}$.
- $\text{Var}(X) = 0$ si et seulement si X est constante ps (et vaut alors $\mathbb{E}[X]$).
- **Inégalité de Tchebychev** : Si $\text{Var}(X) < +\infty$, on a

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}.$$

En général, pour calculer la variance, on utilise la formule de Koenig (cf. Prop. 3.6) en calculant $\mathbb{E}[X]$ puis $\mathbb{E}[X^2]$.

Exemples de variances classiques

$$\begin{array}{ll} X \sim \delta_c \text{ (} X \text{ constante)} : \text{Var}(X) = 0; & X \sim \mathcal{U}([a, b]) : \text{Var}(X) = \frac{(a-b)^2}{12}; \\ X \sim b(p) : \text{Var}(X) = p(1-p); & X \sim \mathcal{E}(\lambda) : \text{Var}(X) = 1/\lambda^2; \\ X \sim \mathcal{B}(n, p) : \text{Var}(X) = np(1-p); & X \sim \mathcal{E}(\lambda) : \text{Var}(X) = 1/\lambda^2; \\ X \sim \mathcal{G}(p) : \text{Var}(X) = \frac{1-p}{p^2}; & X \sim \mathcal{C}(a) : \text{Var}(X) \text{ n'est pas définie}; \\ X \sim \mathcal{P}(\lambda) : \text{Var}(X) = 1/\lambda; & X \sim \gamma(n, \lambda) : \text{Var}(X) = n/\lambda^2; \\ X \sim \mathcal{N}(0, 1) : \text{Var}(X) = 1; & X \sim \mathcal{N}(m, \sigma^2) : \text{Var}(X) = \sigma^2. \end{array}$$

Propriétés des moments

On dispose des propriétés suivantes pour les moments :

- **Inégalité de Hölder** : $\|XY\|_1 \leq \|X\|_p \|Y\|_q$ pour p, q exposants conjugués ($1/p + 1/q = 1$).
- **Inégalité de Cauchy-Schwarz** : $\|XY\|_1 \leq \|X\|_2 \|Y\|_2$ ($p = q = 2$).
- **Inégalité de Minkowski** : $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$ ($1 \leq p < +\infty$).
- Si une variable aléatoire est bornée, elle admet des moments de tous les ordres.
- Si X possède un moment d'ordre r , alors X en possède un d'ordre n pour tout $n \leq r$.
- **Théorème de Riesz-Fisher** : Lorsqu'on identifie les variables presque sûrement égales, l'espace $L^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R} \text{ tel que } \mathbb{E}[|X|^p] < +\infty\}$ est un espace vectoriel normé pour la norme $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$. Il est même complet, c'est donc un espace de Banach.

4 Transformées exponentielles

Définition 4.1 La fonction caractéristique d'une variable aléatoire X est la fonction de \mathbb{R} dans \mathbb{C} donnée par

$$\varphi_X(t) = \mathbb{E}[e^{itX}]. \quad (7)$$

Il s'agit de la transformée de Fourier de la loi \mathbb{P}_X de X . Elle est définie pour tout $t \in \mathbb{R}$, elle est continue et elle caractérise la loi de X : $X \sim Y$ si et seulement si $\varphi_X = \varphi_Y$.

Exemple 4.2 (Fonctions caractéristiques usuelles)

$$\begin{aligned} X \sim \mathcal{U}(x_1, \dots, x_n) : \varphi_X(t) &= \sum_{k=1}^n \frac{1}{n} e^{itx_k} & X \sim U([a, b]) : \varphi_X(t) &= \frac{e^{ibt} - e^{iat}}{it} \\ X \sim b(p) : \varphi_X(t) &= 1 - p + pe^{it} & X \sim \mathcal{E}(\lambda) : \varphi_X(t) &= \frac{\lambda}{\lambda - it} \\ X \sim \mathcal{B}(n, p) : \varphi_X(t) &= (1 - p + pe^{it})^n & X \sim \gamma(n, \lambda) : \varphi_X(t) &= \left(\frac{\lambda}{\lambda - it}\right)^n \\ X \sim \mathcal{C}(a) : \varphi_X(t) &= \exp(-a|t|) & X \sim \mathcal{P}(\lambda) : \varphi_X(t) &= \exp(\lambda(e^{it} - 1)) \\ X \sim \mathcal{N}(0, 1) : \varphi_X(t) &= \exp(-t^2/2) & X \sim \mathcal{N}(m, \sigma^2) : \varphi_X(t) &= \exp(imt - \sigma^2 t^2/2). \end{aligned}$$

La fonction caractéristique est (uniformément) continue, bornée par 1, vaut 1 en 0 et est de type positif : $\forall n \geq 1, t_i \in \mathbb{R}, z_i \in \mathbb{C}$ pour $1 \leq i \leq n$, on a $\sum_{k,l=1}^n \varphi_X(t_k - t_l) z_k \bar{z}_l \geq 0$. Réciproquement, si φ est continue en 0 de type positif avec $\varphi(0) = 1$ alors φ est une fonction caractéristique (théorème de Bochner-Herglotz).

Proposition 4.3 (Fonction caractéristique et moments) Si une variable aléatoire X a un moment d'ordre p , alors sa fonction caractéristique φ_X est dérivable p fois et on a

$$\varphi_X^{(p)}(0) = i^p \mathbb{E}[X^p].$$

Une réciproque partielle existe : si φ_X est dérivable p fois en 0 alors X a des moments jusqu'à l'ordre $2[p/2]$ (la réciproque est entière lorsque p est pair, partielle lorsque p est impair).

Si X est un vecteur aléatoire à valeurs dans \mathbb{R}^d , sa fonction caractéristique est une fonction de \mathbb{R}^d dans \mathbb{C} donnée par $\varphi_X(t_1, \dots, t_d) = \mathbb{E}[\exp(i\langle t, X \rangle)]$ où (X_1, \dots, X_d) sont les coordonnées de $X \in \mathbb{R}^d$ et $\langle t, X \rangle = t_1 X_1 + \dots + t_d X_d$.

D'autres fonctions caractérisent la loi, on en introduit deux fréquemment utilisées ci-dessous.

Fonction génératrice. On considère une variable aléatoire X entière (de rayon $R \geq 1$) $\mathcal{S}(X) \subset \mathbb{N}$ et $p_k = \mathbb{P}(X = k)$. La fonction génératrice de X est

$$G_X(t) = \mathbb{E}[t^X] = \sum_{k \geq 0} p_k t^k. \quad (8)$$

La fonction génératrice est C^∞ sur $[0, 1[$ et dérivable à l'ordre p en 1 si et seulement si $\mathbb{E}[X^p] < +\infty$. De plus, les dérivées successives en 0 permettent de retrouver la loi de X :

$$G_X^{(k)}(0) = k! p_k.$$

En particulier $G_X(0) = p_0, G'_X(1) = \mathbb{E}[X]$ (si le moment d'ordre 1 existe).

Exemple 4.4 Des calculs simples donnent :

- $X \sim b(p) : G_X(t) = (1 - p) + pt$;
- $X \sim \mathcal{B}(n, p) : G_X(t) = (1 - p + pt)^n$;
- $X \sim \mathcal{G}(p) : G_X = \frac{pt}{1 - (1-p)t}$;
- $X \sim \mathcal{P}(\lambda) : G_X(t) = \exp(\lambda(t - 1))$.

Pour les vecteurs aléatoires à valeurs dans \mathbb{N}^d , on définit aussi une notion de fonction génératrice : $G_{(X_1, \dots, X_d)}(t_1, \dots, t_d) = \mathbb{E}[t_1^{X_1} \dots t_d^{X_d}]$.

Transformée de Laplace. Pour une variable aléatoire X , on appelle transformée de Laplace de X

$$\phi_X(t) = \mathbb{E}[e^{tX}]. \quad (9)$$

La fonction ϕ_X est définie pour les valeurs de $t \in \mathbb{R}$ pour lesquelles e^{tX} est intégrable, il s'agit d'un intervalle contenant 0. Il peut être ouvert, fermé ou semi-fermé. Lorsque cet intervalle n'est pas réduit à $\{0\}$, ϕ_X est analytique sur un voisinage de 0 et sur ce voisinage on a

$$\phi_X(t) = \sum_{n \geq 0} \frac{t^n}{n!} \mathbb{E}[X^n].$$

En particulier, $\mathbb{E}[X^n] = \phi_X^{(n)}(0)$. Lorsque $X = (X_1, \dots, X_d)$ est un vecteur aléatoire à valeurs dans \mathbb{R}^d , on définit encore sa transformée de Laplace par $\phi_X(t) = \mathbb{E}[\exp(\langle t, X \rangle)]$ pour les $t \in \mathbb{R}^d$ tels que $\exp(\langle t, X \rangle)$ est intégrable.

Formellement, on a les liens suivants entre toutes ces transformées exponentielles :

$$\varphi_X(t) = G_x(e^{it}), \quad \phi_X(t) = G_X(e^t) = \varphi_X(-it).$$

5 Indépendance

Il s'agit d'une notion typiquement probabiliste et à ce titre fondamentale dans les raisonnements en probabilité.

Définition 5.1 (Indépendance)

- Deux évènements $A, B \in \mathcal{F}$ d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sont indépendants (et on note $A \perp B$) si

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

- Deux variables aléatoires X et Y sont indépendantes (et on note encore $X \perp Y$) si et seulement si pour tout $A, B \in \mathcal{B}(\mathbb{R})$, on a

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B).$$

- Deux tribus \mathcal{F} et \mathcal{G} sont indépendantes si pour tout $A \in \mathcal{F}$ et $B \in \mathcal{G}$, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. On note $\mathcal{F} \perp \mathcal{G}$. On a en particulier : $X \perp Y$ si et seulement si pour les tribus engendrées $\sigma(X) \perp \sigma(Y)$.

En particulier, deux évènements A et B incompatibles (disjoints) ne peuvent pas être indépendants à moins que l'un des deux ne soit de probabilité nulle. Sinon $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$, tandis que $\mathbb{P}(A)\mathbb{P}(B) > 0$. Il ne faut donc pas confondre ces deux notions (indépendance et incompatibilité).

Théorème 5.2 (Loi du 0/1) Soit $(\mathcal{F}_n)_{n \geq 1}$ une suite de tribus indépendantes alors la tribu asymptotique $\mathcal{F}^\infty = \bigcap_{n \geq 1} \sigma\left(\bigcup_{k \geq n} \mathcal{F}_k\right)$ est triviale : si $A \in \mathcal{F}^\infty$ alors $\mathbb{P}(A) = 0$ ou 1.

Proposition 5.3 Deux variables aléatoires X, Y sont indépendantes si et seulement si la loi $\mathbb{P}_{X,Y}$ du couple (X, Y) est la loi produit des lois marginales \mathbb{P}_X et \mathbb{P}_Y :

$$\mathbb{P}_{X,Y} = \mathbb{P}_X \otimes \mathbb{P}_Y. \quad (10)$$

c'est à dire $\mathbb{P}_X(A \times B) = \mathbb{P}_X(A) \mathbb{P}_Y(B)$ pour tout $A \in \mathcal{A}, B \in \mathcal{B}$.

Conséquences. Si $X \perp Y$ alors (quand les espérances sont définies) $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ et plus généralement pour toutes fonctions mesurables f et g alors lorsque les espérances sont bien définies :

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

Pour des variables aléatoires discrètes, le critère d'indépendance s'énonce : $X \perp Y$ si et seulement si $\mathbb{P}(X = a_i, Y = b_j) = \mathbb{P}(X = a_i)\mathbb{P}(Y = b_j)$ pour tout $a_i \in \mathcal{S}(X)$ et $b_j \in \mathcal{S}(Y)$. Pour des variables aléatoires à densité, le critère d'indépendance s'énonce : soit (X, Y) un couple de densité $f_{(X,Y)}$ et de densité marginale $f_X(x)$ pour X , $f_Y(y)$ pour Y . Les variables aléatoires X et Y sont indépendantes si et seulement si $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$ pour tout $x, y \in \mathbb{R}$.

Une conséquence importante : si on connaît les lois de X et de Y , des variables supposées **indépendantes**, on peut reconstruire la loi du couple (X, Y) à partir des marginales par (10). Ce n'est pas vrai en général lorsque X et Y ne sont pas indépendantes.

L'indépendance se caractérise en termes de fonction caractéristique (7) :

Théorème 5.4 Deux variables aléatoires X et Y sont indépendantes si et seulement si

$$\varphi_{(X,Y)}(s, t) = \varphi_X(s) \varphi_Y(t) \quad \text{pour tout } s, t \in \mathbb{R}.$$

Des résultats analogues existent pour les autres fonctions qui caractérisent la loi. Par exemple, pour les fonctions génératrices (8) et de Laplace (9) :

- Si X, Y sont des variables aléatoires entières alors $X \perp Y$ si et seulement si $G_{X,Y}(s, t) = G_X(s) G_Y(t)$.
- Si X, Y sont des variables aléatoires alors $X \perp Y$ si et seulement si $\phi_{(X,Y)}(s, t) = \phi_X(s) \phi_Y(t)$ pour tout $s, t \in \mathbb{R}$.

Sommes de variables aléatoires indépendantes

Dans le cas indépendant, la loi de la somme de deux variables indépendantes est connue.

Proposition 5.5 *Si X et Y sont deux variables aléatoires indépendantes alors la loi de $X + Y$ est donnée par $\mathbb{P}_X * \mathbb{P}_Y$.*

En particulier, si X, Y ont pour densités f et g alors $X + Y$ admet pour densité

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(y)g(x - y) dy.$$

Pour les fonctions caractéristiques, on a :

Proposition 5.6 *Si deux variables aléatoires X et Y sont indépendantes alors pour tout $t \in \mathbb{R}$:*

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t).$$

Des résultats analogues existent pour la fonction génératrice (8) et la transformée de Laplace (9) :

$$G_{X+Y}(t) = G_X(t) G_Y(t), \quad \phi_{X+Y}(t) = \phi_X(t) \phi_Y(t).$$

Avec l'indépendance, la somme d'une variance vérifie une propriété remarquable.

Corollaire 5.7 *Si X et Y sont deux variables aléatoires indépendantes avec des moments d'ordre 2 finis alors*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Attention, sans indépendance, cette propriété est fautive (considérer $X = Y$!). On a alors

$$\text{Var}(X, Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$$

où $\text{Cov}(X, Y)$ est la **covariance** de X, Y donnée par

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

La coraviance est bilinéaire et vérifie $\text{Cov}(X, Y) \leq \sqrt{\text{Var}(X) \text{Var}(Y)}$ (Cauchy-Schwarz).

En sommant des variables aléatoires indépendantes de loi classiques, on retrouve d'autres lois classiques :

- la somme de variables aléatoires normales indépendantes suit encore une loi normale : $\mathcal{N}(m_1, \sigma_1^2) * \mathcal{N}(m_2, \sigma_2^2) = \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$;
- la somme de n variables aléatoires de Bernoulli $b(p)$ indépendantes suit une loi binomiale $\mathcal{B}(n, p)$: $b(p)^{*n} = \mathcal{B}(n, p)$;
- la somme de n variables aléatoires de exponentielle $\mathcal{E}(\lambda)$ indépendantes suit une loi gamma $\gamma(n, \lambda)$: $\mathcal{E}(\lambda)^{*n} = \gamma(n, \lambda)$.

6 Conditionnement

6.1 Conditionnement par un évènement

Définition 6.1 Soit B un évènement de probabilité non nulle $\mathbb{P}(B) \neq 0$. Pour tout évènement A , on définit la probabilité conditionnelle de A sachant B :

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

L'intérêt de cette notion vient du fait que souvent, compte tenu des informations disponibles dans un problème de modélisation, il peut être plus facile d'attribuer une valeur à la probabilité conditionnelle $\mathbb{P}(A|B)$ que de calculer $\mathbb{P}(A \cap B)$ ou $\mathbb{P}(A)$. Si $A \perp B$, évidemment, on a $\mathbb{P}(A|B) = \mathbb{P}(A)$, ie. le conditionnement par B est sans effet.

En fait, la fonction d'ensemble $\mathbb{P}(\cdot|B) : A \in \mathcal{F} \mapsto \mathbb{P}(A|B)$ est une nouvelle probabilité sur (Ω, \mathcal{F}) . De ce fait, on dispose pour les probabilités conditionnelles de toutes les propriétés habituelles d'une probabilité. On a en plus :

Proposition 6.2 (Règle des conditionnements successifs) Si A_1, \dots, A_n sont n évènements tels que $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) \neq 0$, alors

$$\begin{aligned} & \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \times \dots \times \mathbb{P}(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}). \end{aligned} \quad (11)$$

Définition 6.3 On appelle **système complet** d'évènements toute suite $(B_i)_{i \in I}$ d'évènements deux à deux disjoints et dont la somme des probabilités vaut 1 :

$$\sum_{i \in I} \mathbb{P}(B_i) = 1.$$

Le système est dit fini si I est fini, infini si I est infini.

Proposition 6.4 (Formule des probabilités totales) Si $(B_i)_{i \in I}$ forme un système complet dénombrable de Ω en évènements avec $\mathbb{P}(B_i) > 0$ pour tout $i \in I$, alors pour tout $A \in \mathcal{F}$:

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i) \mathbb{P}(B_i).$$

Lorsque l'on sait calculer les probabilités conditionnelles $\mathbb{P}(A|B_i)$ pour tous les membres B_i d'un système complet $(B_i)_{i \in I}$, on peut chercher les probabilités conditionnelles avec les conditionnements inverses $\mathbb{P}(B_i|A)$. Elles sont données par :

Proposition 6.5 (Formule de Bayes) Soit A un évènement de probabilité non nulle et $(B_i)_{i \in I}$ un système complet de Ω en évènements de probabilités non nulles. On a

$$\forall j \in I, \quad \mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j) \mathbb{P}(B_j)}{\sum_{i \in I} \mathbb{P}(A|B_i) \mathbb{P}(B_i)}.$$

6.2 Lois conditionnelles

Étant données deux variables aléatoires X, Y , de façon générale, la loi conditionnelle de Y sachant X est définie par le théorème de désintégration des mesures qui affirme que pour tout $A, B \in \mathcal{B}(\mathbb{R})$, on a :

$$\mathbb{P}(Y \in A, X \in B) = \int_B \mathbb{P}(Y \in A | X = x) \mathbb{P}_X(dx) = \mathbb{E}[\mathbf{1}_B(X) \mathbb{P}(Y \in A | X)]. \quad (12)$$

La famille $(\mathbb{P}(Y \in \cdot | X = x))_{x \in \mathbb{R}}$ s'appelle la famille des probabilités conditionnelles de Y sachant X . La loi conditionnelle de Y sachant X est $\mathbb{P}(Y \in \cdot | X)$. L'existence est assurée par le théorème de Radon-Nikodym (3), la régularité de la famille (pour tout $x \in \mathbb{R}$, $\mathbb{P}(Y \in \cdot | X = x)$ est une probabilité) est assurée par le théorème de Jirina (admis). On observe facilement que lorsque $X \perp Y$, $\mathbb{P}(X \in A | X) = \mathbb{P}(Y \in A)$, ie. le conditionnement est sans effet lorsqu'il y a indépendance.

Probabilités ponctuelles conditionnelles

Définition 6.6 Soit (X, Y) un couple de variables aléatoires discrètes sur $(\Omega, \mathcal{F}, \mathbb{P})$. Pour x tel que $\mathbb{P}(X = x) \neq 0$, on appelle loi conditionnelle de Y sachant $X = x$, l'ensemble des probabilités ponctuelles

$$\mathbb{P}(Y = y | X = x) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)} \text{ si } y \in \mathcal{S}(Y), \quad = 0 \text{ sinon.}$$

On vérifie facilement que cette définition vérifie la définition générale (12). Si X, Y sont des variables aléatoires indépendantes alors la loi conditionnelle de X sachant $Y = y$ est la même que celle de X : $\mathbb{P}_Y(\cdot | X = x) = \mathbb{P}_Y$. Autrement dit, à nouveau on observe que le conditionnement par une variable aléatoire indépendante est sans effet.

Lorsque $\mathcal{S}(X) = \{x_i : i \in I\}$, la loi conditionnelle est alors donnée par

$$\mathbb{P}(Y \in A | X) = \sum_{i \in I} \mathbb{P}(Y \in A | X = x_i) \mathbf{1}_{\{X = x_i\}}, \quad A \in \mathcal{A}.$$

Densité conditionnelle

Définition 6.7 Soit (X, Y) un couple de variables aléatoires réelles de densité $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. On définit la densité conditionnelle de Y sachant $X = x$ par

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)} \text{ si } f_X(x) > 0, \quad = 0 \text{ sinon}$$

où $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$ est la densité (marginale) de X .

La loi conditionnelle de Y sachant $X = x$ est alors définie par cette densité $f_{Y|X=x}$:

$$\forall A \in \mathcal{B}(\mathbb{R}), \mathbb{P}(Y \in A|X = x) = \int_A f_{Y|X=x}(y) dy.$$

Cette définition vérifie encore facilement la définition générale (12). La densité conditionnelle $f_{Y|X=x}$ est une fonction de la seule variable y ; x n'y intervient seulement qu'en paramètre de la fonction. Comme dans le cas discret, si les variables aléatoires X et Y sont indépendantes de densités f_X et f_Y alors les densités conditionnelles sont les densités marginales : $f_{Y|X=x}(y) = f_Y(y)$ pour tout $x \in \mathbb{R}$: à nouveau le conditionnement par une variable aléatoire indépendante est sans effet.

6.3 Espérance conditionnelle

L'espérance conditionnelle de Y , variable aléatoire supposée intégrable, sachant X est définie comme l'espérance de la loi conditionnelle $\mathbb{P}(Y \in \cdot | X)$:

$$\mathbb{E}[Y|X] = \int y \mathbb{P}(Y \in dy|X).$$

On montre aussi qu'il s'agit de la (presque sûrement) unique variable aléatoire Z $\sigma(X)$ -mesurable telle que

$$\mathbb{E}[Y\mathbf{1}_{\{X \in B\}}] = \mathbb{E}[Z\mathbf{1}_{\{X \in B\}}] \quad \forall B \in \mathcal{B}(\mathbb{R}). \quad (13)$$

Son existence est assurée par le théorème de Radon-Nikodym (3) appliquée à la mesure $\mathbb{Q}(B) = \mathbb{E}[Y\mathbf{1}_{\{X \in B\}}]$ absolument continue par rapport à \mathbb{P}_X : on a alors $Z = f(X)$ où f est alors la dérivée de Radon-Nikodym : $f = \frac{d\mathbb{Q}}{d\mathbb{P}_X}$.

Noter les propriétés de l'espérance conditionnelle :

- Si Y est $\sigma(X)$ -mesurable alors $\mathbb{E}[Y|X] = Y$ ps.
- Si $Y \perp X$ alors $\mathbb{E}[Y|X] = \mathbb{E}[Y]$ (conditionnement sans effet).
- Conditionnements successifs : $\mathbb{E}[\mathbb{E}[Y|X_1, X_2] | X_1] = \mathbb{E}[Y|X_1]$. Cela rappelle (11).
- En particulier $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$.
- Si Z est $\sigma(X)$ -mesurable : $\mathbb{E}[YZ|X] = \mathbb{E}[Y|X]Z$.
- Lorsque $Y \in L^2(\Omega)$, $\mathbb{E}[Y|\mathcal{G}] = \text{proj}_{L^2(\Omega, \sigma(X))}(Y)$ est la projection de $Y \in L^2(\Omega, \mathcal{F})$ sur $L^2(\Omega, \sigma(X))$.
- Lorsque (X, Y) est un vecteur gaussien, alors $\mathbb{E}[Y|X] \in L^2(\sigma(X))$ s'écrit

$$\mathbb{E}[Y|X] = \frac{\text{Cov}(X, Y)}{\sigma_X^2} X + \left(m_Y - \frac{\text{Cov}(X, Y)}{\sigma_X^2} m_X \right) \sim \mathcal{N}\left(m_Y, \frac{\text{Cov}(X, Y)^2}{\sigma_X^2} \right).$$

- On définit de la même façon d'autres quantités conditionnelles telles que la variance conditionnelle $\text{Var}(Y|X)$ qui est la variance de Y par rapport à la loi conditionnelle $\mathbb{P}(\cdot|X)$. Par Pythagore, on a la décomposition suivante de la variance en :

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X)] + \text{Var}(\mathbb{E}[Y|X]).$$

En pratique pour calculer l'espérance conditionnelle $\mathbb{E}[Y|X]$ sachant X , on calcule l'espérance de Y en supposant (pour ce calcul) que X n'est pas aléatoire. Fixer X a des conséquences sur la loi de Y , si bien qu'on fait le calcul avec la loi conditionnelle de Y sachant X :

$$\mathbb{E}[Y|X] = \int_{\mathbb{R}} y \mathbb{P}_{Y|X}(dy).$$

Une fois ce calcul fait, on obtient une valeur qui dépend de X : $\mathbb{E}[Y|X] = h(X)$.

On peut reobtenir l'espérance sachant $X = x$ par $\mathbb{E}[Y|X = x] = h(x)$.

Quand on prend l'espérance conditionnelle de $Y = \mathbf{1}_A$, on récupère la probabilité conditionnelle de X :

$$\mathbb{E}[\mathbf{1}_A|X] = \mathbb{P}(A|X).$$

7 Convergences probabilistes

On rappelle brièvement les différentes convergences probabilistes.

Définition 7.1 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires et X une variable aléatoire.

Convergence presque sûre. On dit que $(X_n)_{n \geq 1}$ converge presque sûrement (ps) vers X , et on note $X_n \xrightarrow{ps} X$, si la convergence est vraie avec une probabilité 1

$$\mathbb{P}(\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)) = 1.$$

Convergence en norme p . On dit que $(X_n)_{n \geq 1}$ converge en norme p vers X , et on note $X_n \xrightarrow{L^p} X$, si $\|X_n - X\|_p \rightarrow 0$ quand $n \rightarrow +\infty$, c'est à dire

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

Convergence en probabilités. On dit que $(X_n)_{n \geq 1}$ converge en probabilité vers X , et on note $X_n \xrightarrow{\mathbb{P}} X$ si

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|X - X_n| \geq \varepsilon) = 0.$$

Convergence en loi. On dit que $(X_n)_{n \geq 1}$ converge en loi vers X si

$$\mathbb{P}(X_n \in A) \rightarrow \mathbb{P}(X \in A), \quad n \rightarrow +\infty,$$

pour tout $A \in \mathcal{B}(\mathbb{R})$ tel que $\mathbb{P}(X \in \partial A) = 0$.

On la note $X_n \implies X$ ou $X_n \xrightarrow{loi} X$ ou $X_n \xrightarrow{\mathcal{L}} X$. De façon équivalente, la convergence en loi s'exprime en termes des fonctions de répartition (facile à voir) :

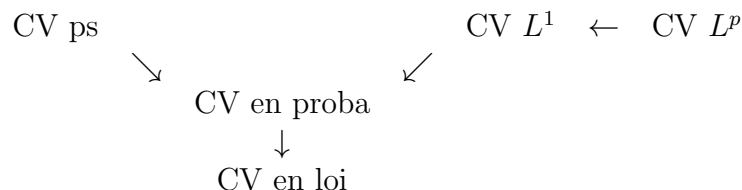
$$\mathbb{P}(X_n \leq x) \rightarrow \mathbb{P}(X \leq x), \quad n \rightarrow +\infty$$

en tout $x \in \mathbb{R}$ tel que $\mathbb{P}(X = x) = 0$, et même en termes des fonctions caractéristiques (plus difficile à voir) :

Théorème 7.2 (Paul Lévy) *La convergence en loi des variables aléatoires est équivalente à la convergence simple des fonctions caractéristiques :*

$$X_n \Longrightarrow X \quad \text{si et seulement si} \quad \lim_{n \rightarrow +\infty} \varphi_{X_n}(t) = \varphi_X(t) \quad \text{pour tout } t \in \mathbb{R}.$$

Les liens entre les différentes convergences sont résumés par le diagramme suivant :



Aucune réciproque n'est (entièrement) vraie. Certaines le sont cependant dans des cas particuliers. Par exemple si $X_n \Rightarrow X$ avec $X = c$ ps (c constante) alors il y a convergence en probabilité : $X_n \xrightarrow{\mathbb{P}} X$.

Un outil important pour prouver des convergences presque sûres est fourni par le lemme de Borel-Cantelli :

Lemme 7.3 (Borel-Cantelli) *Soit $(A_n)_{n \geq 1}$ une suite d'évènements.*

1. *Si $\sum_{n \geq 1} \mathbb{P}(A_n) < +\infty$ alors $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0$, ie. presque sûrement, seuls un nombre fini d'évènements A_n est réalisé.*
2. *Si $\sum_{n \geq 1} \mathbb{P}(A_n) = +\infty$ et les $A_n, n \geq 1$, sont indépendants alors $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 1$, ie. presque sûrement, une infinité d'évènements $A_n, n \geq 1$, est réalisée.*

On rappelle dans un contexte probabiliste les théorèmes fondamentaux de passage à la limite dans les espérances : convergence monotone (Th. 7.4), lemme de Fatou (Lemme 7.5) et théorème de convergence dominée (Th. 7.6).

Théorème 7.4 (Convergence monotone, Beppo Levi) *Soit $(X_n)_{n \geq 1}$ une suite croissante de variables aléatoires positives ($0 \leq X_n \leq X_{n+1}$). Soit $X = \lim_{n \rightarrow +\infty} X_n$ la limite ps de X_n dans $[0, +\infty]$. Alors $\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.*

Lemme 7.5 (Fatou) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires positives. Alors*

$$\mathbb{E} \left[\liminf_{n \rightarrow +\infty} X_n \right] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n].$$

Théorème 7.6 (Convergence dominée) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires telle que $X_n \xrightarrow{\text{ps}} X$ quand $n \rightarrow +\infty$. S'il existe une variable aléatoire Y intégrable ($\mathbb{E}[|Y|] < +\infty$) telle que pour tout n , $|X_n| \leq Y$ ps alors $\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.*

En pratique : quand la convergence est dominée, on peut intervertir limite et espérance.

8 Théorèmes *limite* classiques

Dans la suite, i.i.d. signifie indépendant(e)s et identiquement distribué(e)s, c'est à dire de même loi. On notera aussi souvent *va iid* pour variables aléatoires indépendantes et identiquement distribuées.

Loi des grands nombres (LGN)

Théorème 8.1 (LGN) *Soit $(X_n)_{n \geq 1}$ une suite de va iid avec un moment d'ordre un (ie. $\mathbb{E}[|X_1|] < +\infty$, si bien que l'espérance est bien définie). Alors*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{*} \mathbb{E}[X_1], \quad n \rightarrow +\infty. \quad (14)$$

- On parle de LGN **faible** lorsque $*$ est une convergence en probabilité : $\xrightarrow{\mathbb{P}}$.
- On parle de LGN **forte** lorsque $*$ est une convergence ps \xrightarrow{ps} ou convergence L^p : $\xrightarrow{L^p}$.
- En fait, si pour des variables aléatoires iid X_n , $n \geq 1$, la convergence (14) a lieu ps pour une limite $c \in \mathbb{R}$ donnée alors nécessairement $X_1 \in L^1(\Omega)$ et $c = \mathbb{E}[X_1]$.

La LGN montre que la **moyenne arithmétique** $\frac{1}{n} \sum_{i=1}^n X_i$ converge vers la moyenne probabiliste $\mathbb{E}[X_1]$. C'est grâce à ce résultat qu'on peut estimer une proportion dans une population par une proportion dans un échantillon (représentatif). Plus généralement, c'est l'outil de base pour l'estimation ponctuelle en statistique.

Théorème central limite (TCL)

On déduit par exemple du théorème de Paul Lévy (Th. 7.2) un résultat fondamental en probabilité et statistiques : le théorème central limite (TCL).

Théorème 8.2 (TCL) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires iid, d'espérance m et de variance finie σ^2 . Soit $S_n = X_1 + \dots + X_n$ la somme partielle. Alors quand $n \rightarrow +\infty$*

$$\frac{S_n - nm}{\sqrt{n\sigma^2}} \Longrightarrow \mathcal{N}(0, 1).$$

Remarque 8.3 — Le TCL complète la loi des grands nombres en donnant la vitesse de convergence dans la LGN.

- Le TCL justifie le rôle central de la loi $\mathcal{N}(0, 1)$ qui apparaît au cœur des variations de n'importe quelle loi (avec un moment d'ordre 2) par rapport à sa moyenne.
- Le TCL permet de compléter une estimation ponctuelle par un intervalle de confiance. Après la LGN, c'est le deuxième outil de base en statistique.

- Le TCL justifie que lorsque n est grand, on approxime la loi d'une somme de variables aléatoires iid de $L^2(\Omega)$ par une loi normale : la loi de S_n est approximée par celle de

$$nm + \sigma\sqrt{n}\mathcal{N}(0, 1) \sim \mathcal{N}(nm, n\sigma^2).$$

Application (théorème de Moivre-Laplace). Comme une variable aléatoire X_n de loi binomiale $\mathcal{B}(n, p)$ peut se voir comme la somme de n variables aléatoires ϵ_i indépendantes de loi de Bernoulli $b(p)$, $X_n = \epsilon_1 + \dots + \epsilon_n$, la remarque précédente montre qu'on peut approcher la loi $\mathcal{B}(n, p)$ par la loi normale $\mathcal{N}(np, np(1 - p))$.

Références

- [BL] Philippe Barbe, Michel Ledoux. *Probabilité*. EDP science, 2007.
- [JCB-mesure] Jean-Christophe Breton. *Intégrale de Lebesgue*. [Notes de cours](#) de L3 Mathématiques, Université de Rennes 1, 2014.
- [JCB-proba] Jean-Christophe Breton. *Probabilités*. [Notes de cours](#) de L3 Mathématiques, Université de Rennes 1, 2014.
- [BP] Marc Briane, Gilles Pagès. *Théorie de l'intégration*, 5ème édition. Coll. Vuibert Supérieur, Ed. Vuibert, 2012.
- [Ouv] Jean-Yves Oувrard. *Probabilités*. Tomes 1 et 2. Cassini, 2008.