

M61B : Probabilités et intégration, *parcours classique*

Clément Erignoux *

*Pour l'écriture de ces notes de cours, je me suis particulièrement inspiré du livre **Mesures, Intégration et Probabilités**, de Thierry Gallouët et Raphaële Herbin, disponible sur HAL, <https://hal.archives-ouvertes.fr/hal-01283567>, dont je remercie les auteurs. Pour plus d'exercices, d'exemples, et d'approfondissements, je vous invite donc à consulter ce dernier. Les résultats et notions clés de ce cours sont indiqués par une ★.*

Introduction : pourquoi une nouvelle intégrale ?

La construction de la théorie des probabilités sur des ensembles continus (ou, en tout cas, non dénombrables) nécessite l'introduction d'une nouvelle notion d'intégrale. Cette intégrale est l'*intégrale de Lebesgue*, qui repose elle même sur la *théorie de la mesure*, dont nous allons voir ensemble quelques éléments. Cette théorie va bien au-delà de ce que nous allons voir en cours, puisque nous ne traiterons que les éléments nécessaires à l'introduction des probabilités continues. Avant tout, toutefois, commençons par nous poser la question : pourquoi une nouvelle intégrale pour compléter la notion, plus élémentaire, d'intégrale de Riemann ?

Commençons par quelques rappels sur l'intégrale de Riemann. Étant donnée une fonction sur un segment, par exemple $[0, 1]$, on définit pour toute suite finie $0 \leq x_1 < x_2 < \dots < x_n \leq 1$ la somme de Riemann correspondante

$$\sum_{i=1}^{n-1} f(x_i)(x_{i+1} - x_i).$$

En pratique, on pourra par exemple considérer la suite $x_i = i/n$, dont la somme de Riemann est donnée par $\frac{1}{n} \sum_{i=1}^n f(x_i)$. Une fonction est Riemann intégrable si, en faisant tendre le *pas* d'une $\max\{x_{i+1} - x_i\}$ vers 0, la somme de Riemann converge, vers une quantité qui est alors l'intégrale (de Riemann) de la fonction sur le segment $[0, 1]$.

Alors, pourquoi y a-t-il besoin d'une autre construction de l'intégrale ? Un premier élément de réponse vient de l'espace des fonctions considéré : L'intégrale

*Pour toute typo/question, me contacter à clement.erignoux@inria.fr. Les notes de cours seront uploadées sur ma page web, <http://chercheurs.lille.inria.fr/cerignou/homepage.html>

de Riemann est très limitée, puisque sa construction n'a de sens que pour les fonctions *continues* (ou plus rigoureusement presque-partout continues) définies sur un segment. La théorie de la mesure substitue aux fonctions continues les fonctions *mesurables*, qui est une notion beaucoup plus générale.

Prenons par exemple l'ensemble des nombres rationnels \mathbb{Q} , et considérons la fonction indicatrice $\mathbf{1}_{\mathbb{Q}}$, définie sur \mathbb{R} par

$$\mathbf{1}_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{si } x \in \mathbb{Q} \\ 0 & \text{si } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

Cette fonction n'est *continue* en aucun point de \mathbb{R} puisque tout réel peut être approché par une suite de nombres rationnels, et n'est donc pas *intégrable* sous la théorie de l'intégrale de Riemann. Toutefois, cette fonction est *mesurable*, et donc *intégrable*, dans le contexte de l'intégrale de Lebesgue.

La notion de *convergence* est une autre raison naturelle pour l'emploi de l'intégrale de Lebesgue. Considérons l'ensemble $E = C([0, 1], \mathbb{R})$ des fonctions continues sur $[0, 1]$ à valeurs dans \mathbb{R} . Étant donnée une suite $(f_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ de fonctions continues, on veut pouvoir formaliser mathématiquement la notion de convergence sous l'intégrale, c'est-à-dire trouver des conditions sur la suite $(f_n)_n$ permettant d'affirmer que

$$\left[f_n \xrightarrow[n \rightarrow \infty]{} f \right] \implies \left[\int_{[0,1]} f_n(x) dx \xrightarrow[n \rightarrow \infty]{} \int_{[0,1]} f(x) dx \right].$$

L'intégrale de Riemann n'offre qu'un résultat faible de convergence, qui repose sur la *convergence uniforme* :

Théorème 1 : Théorème de convergence uniforme

Supposons qu'une suite $(f_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ converge uniformément vers une fonction $f \in E$, c'est-à-dire que $\max_{x \in [0,1]} |f_n(x) - f(x)| \xrightarrow[n \rightarrow \infty]{} 0$. Alors,

$$\int_{[0,1]} f_n(x) dx \xrightarrow[n \rightarrow \infty]{} \int_{[0,1]} f(x) dx.$$

En pratique, ce théorème est très faible car l'hypothèse de convergence uniforme est extrêmement forte, et ne peut donc souvent pas être obtenue. L'intégrale de Lebesgue donne par exemple accès au *théorème de convergence dominée*, beaucoup plus utile, car il repose sur des hypothèses plus faibles :

Théorème 2 : Théorème de convergence dominée

Supposons qu'une suite $(f_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ converge simplement vers $f \in E$, c'est-à-dire que pour tout $x \in [0, 1]$, $f_n(x) \xrightarrow[n \rightarrow \infty]{} f(x)$. Supposons également que les f_n sont uniformément bornées, i.e. qu'il existe une constante C telle que pour

tout $x \in [0, 1]$, $n \in \mathbb{N}$, on ait $|f_n(x)| \leq C$. Alors,

$$\int_{[0,1]} f_n(x) dx \xrightarrow{n \rightarrow \infty} \int_{[0,1]} f(x) dx.$$

Pour illustrer la différence entre ces deux résultats, considérons pour $n \geq 1$ la suite de fonctions f_n , représentées en Figure 1,

$$f_n(x) = \begin{cases} 2nx & \text{pour } x \in [0, 1/2n] \\ 2 - 2nx & \text{pour } x \in [1/2n, 1/n] \\ 0 & \text{pour } x \in [1/n, 1]. \end{cases}$$

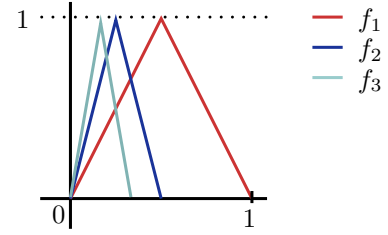


FIGURE 1

On remarque facilement que pour tout $x \in [0, 1]$, $f_n(x) \rightarrow f(x) := 0$ quand $n \rightarrow \infty$. Pour cette suite de fonctions, le *théorème de convergence* uniforme ne s'applique pas, car $\max_{x \in [0,1]} |f_n(x) - f(x)| = 1$ pour tout $n \geq 1$.

En revanche, on peut appliquer le *théorème de convergence dominée* avec $C = 1$, et on obtient alors $\int_{[0,1]} f_n(x) dx \xrightarrow{n \rightarrow \infty} \int_{[0,1]} f(x) dx = 0$, ce qui est vrai puisque $\int_{[0,1]} f_n(x) dx = 1/2n$.

1 Théorie de la mesure

1.1 Tribus

Afin de construire la *mesure de Lebesgue*, et ainsi l'*intégrale de Lebesgue* mentionnée ci-dessus, et plus généralement les probabilités continues, nous allons avoir besoin de donner du sens à la notion d'*ensembles mesurables*. En effet, en prenant par exemple \mathbb{R}^n comme espace de travail, on veut choisir un sous-ensemble $\mathcal{T} \subset \mathcal{P}(\mathbb{R}^n)$ de parties de \mathbb{R}^n dont la *mesure* sera bien définie.

Or, il s'avère que, pour que notre mesure respecte bien les propriétés de σ -additivité que l'on attend (cf. Définition 8 ci-dessous), on ne peut pas choisir $\mathcal{T} = \mathcal{P}(\mathbb{R}^n)$ (c'est-à-dire que toute partie de \mathbb{R}^n aurait une mesure), puisque cet espace est trop gros et fait apparaître des contradictions dans la construction de notre mesure. Les curieux pourront par exemple se renseigner sur le paradoxe de Banach-Tarski : sous ce paradoxe, et en admettant que l'on arrive à construire la mesure de Lebesgue sur toutes les parties de \mathbb{R}^n , on montrerait alors que $1 = 2$. Ce qui est un peu problématique.

Définition 3 : Tribu, espace mesurable



Soit E un ensemble (on peut penser par exemple à $E = \mathbb{R}^n$) et $\mathcal{T} \subset \mathcal{P}(E)$ un ensemble de parties de E . La famille \mathcal{T} est une *tribu*, ou σ -*algèbre*, sur E si \mathcal{T} vérifie

- $\emptyset, E \in \mathcal{T}$,

- \mathcal{T} est stable par union et intersection dénombrable, c'est-à-dire que si $A_n \in \mathcal{T}$ pour tout $n \in \mathbb{N}$, alors $\cap_{n \in \mathbb{N}} A_n \in \mathcal{T}$ et $\cup_{n \in \mathbb{N}} A_n \in \mathcal{T}$,
- \mathcal{T} est stable par passage au complémentaire, c'est-à-dire que si $A \in \mathcal{T}$, alors $A^c := E \setminus A \in \mathcal{T}$.

Un couple (E, \mathcal{T}) , où \mathcal{T} est une tribu sur E , est appelé un *espace mesurable*.

Exemple : La tribu dite *grossière* $\mathcal{T} = \{\emptyset, E\}$ et la tribu *discrète* $\mathcal{T} = \mathcal{P}(E)$ sont toujours des tribus (pas très intéressantes).

Remarque : En pratique, on n'aura pas à vérifier à la fois la stabilité par union et celle par intersection, une seule des deux suffit, puisque l'on peut passer d'une union à une intersection en passant au complémentaire : $\cup_{n \in \mathbb{N}} A_n = (\cap_{n \in \mathbb{N}} A_n^c)^c$.

Proposition 4 : Intersection de tribus

Soient I un ensemble non vide (pas nécessairement dénombrable), et une famille $(\mathcal{T}_i)_{i \in I}$ de tribus. Alors,

$$\bigcap_{i \in I} \mathcal{T}_i = \{A \in \mathcal{P}(E) ; A \in \mathcal{T}_i \forall i \in I\}$$

est aussi une tribu.

Preuve : Il s'agit simplement de vérifier les trois propriétés. Comme \emptyset et E appartiennent à chacun des \mathcal{T}_i , ils appartiennent également à l'intersection. Par ailleurs, si $A_j \in \mathcal{T}_i \forall j \in \mathbb{N}$, $i \in I$, alors, comme les \mathcal{T}_i sont des tribus, $\cup_j A_j \in \mathcal{T}_i$ pour tout i et donc $\cup_j A_j \in \bigcap_{i \in I} \mathcal{T}_i$. Il en va de même pour les intersections dénombrables et pour le passage au complémentaire. \square

Ce résultat permet de définir la tribu engendrée par un ensemble de parties de E .

Définition 5 : Tribu engendrée par un ensemble de parties

Soit $C \subset \mathcal{P}(E)$ un ensemble de parties de E . On appelle *tribu engendrée par C* la plus petite tribu contenant C , notée $\sigma(C)$ définie par

$$\sigma(C) = \bigcap \left\{ \mathcal{T} \subset \mathcal{P}(E) ; \mathcal{T} \text{ tribu contenant } C \right\}.$$

A noter que le membre de droite est bien défini, puisque la tribu discrète $\mathcal{T} = \mathcal{P}(E)$ contient C , et donc l'intersection est non vide.

Pour $E = \mathbb{R}$, on définit une tribu particulière sur laquelle va être construite la mesure de Lebesgue.

Définition 6 : tribu borélienne sur \mathbb{R}



On appelle *tribu borélienne* sur \mathbb{R} la tribu, notée $\mathcal{B}(\mathbb{R})$, engendrée par les in-

tervalles ouverts de \mathbb{R} . Autrement dit, étant donné

$$\mathcal{I} = \{]a, b[, a \in \mathbb{R} \cup \{-\infty\}, b \in \mathbb{R} \cup \{+\infty\}, a \leq b\},$$

on définit

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I}) = \bigcap \left\{ \mathcal{T} \subset \mathcal{P}(\mathbb{R}); \mathcal{T} \text{ tribu contenant } \mathcal{I} \right\}.$$

On appelle *boréliens* les éléments de la tribu borélienne.

Remarque : La tribu borélienne peut en réalité être définie pour tout espace topologique E . Elle est alors définie comme la plus petite tribu contenant tous les ouverts de la topologie de E .

En particulier, étant donné un intervalle $\Gamma \subset \mathbb{R}$, on peut définir la tribu borélienne sur Γ , notée $\mathcal{B}(\Gamma)$, qui est engendrée par les intervalles de Γ . Cette dernière est également donnée par

$$\mathcal{B}(\Gamma) = \{A \cap \Gamma, A \in \mathcal{B}(\mathbb{R})\},$$

ce qui est complètement équivalent.

1.2 Fonction mesurable, mesure

Maintenant munis de la notion de *tribus*, nous allons pouvoir définir une classe de fonctions beaucoup plus générales que les fonctions *continues*, sur laquelle l'intégrale de Lebesgue sera construite.

Définition 7 : Fonction mesurable



Soient deux espaces mesurables (E, \mathcal{T}) et (E', \mathcal{T}') . Une fonction $f : E \rightarrow E'$ est dite *mesurable* (ou, plus précisément, $(\mathcal{T}, \mathcal{T}')$ -mesurable) si pour tout $A \in \mathcal{T}'$, on a

$$f^{-1}(A) := \{x \in E \mid f(x) \in A\} \in \mathcal{T}.$$

Si la tribu d'arrivée \mathcal{T}' est la tribu borélienne, on dira que la fonction est *borel-mesurable*, ou *borélienne*.

Remarque : La mesurabilité, tant pour les fonctions que pour les ensembles, est une propriété extrêmement stable : en pratique, il est extrêmement difficile de produire des fonctions réelles non-mesurables, parce qu'il est très difficile de produire des *ensembles* non-mesurables. Par ailleurs, comme vous le verrez en exercices, la plupart des opérations auxquelles on pourrait penser sur les fonctions mesurables, somme, produit, composition, limites, liminf, limsup, etc. préservent toutes la mesurabilité.

Exercice 1 : Mesurabilité des fonctions continues

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue. On munit \mathbb{R} , à l'arrivée comme au départ, de la tribu borélienne.

- 1) Soit $\mathcal{G} = \{B \in \mathcal{B}(\mathbb{R}), f^{-1}(B) \in \mathcal{B}(\mathbb{R})\}$, montrer que \mathcal{G} est une tribu.
- 2) Montrer que tout ensemble ouvert O est un borélien, i.e. $O \in \mathcal{B}(\mathbb{R})$. Indice : on pourra considérer les $I_{x,r} =]x-r, x+r[$ pour $x \in \mathbb{Q}$ et $r \in \mathbb{Q}_+$.
- 3) Montrer que \mathcal{G} contient les intervalles.
- 4) En déduire que $\mathcal{G} = \mathcal{B}(\mathbb{R})$, et conclure qu'une fonction continue est mesurable.

SOLUTION :

- 1) On commence par remarquer que $f^{-1}(\emptyset) = \emptyset$ et $f^{-1}(\mathbb{R}) = \mathbb{R}$, donc $\emptyset, \mathbb{R} \in \mathcal{G}$. Par ailleurs, si $A_n \in \mathcal{G} \forall n \in \mathbb{N}$, alors $f^{-1}(\cup A_n) = \cup f^{-1}(A_n) \in \mathcal{B}(\mathbb{R})$. Enfin, $f^{-1}(A)^c = f^{-1}(A^c)$ donc si $A \in \mathcal{G}$, on a également $A^c \in \mathcal{G}$. \mathcal{G} est donc une tribu.
- 2) Soit O un ouvert de \mathbb{R} , pour tout $x \in \mathbb{Q}$ on définit $\mathbb{Q}_x = \{r \in \mathbb{Q}_+,]x-r, x+r[\subset O\}$. On peut alors écrire

$$O = \bigcup_{x \in \mathbb{Q} \cap O} \bigcup_{r \in \mathbb{Q}_x}]x-r, x+r[,$$

qui est une union dénombrable d'intervalles, et est donc un borélien.

- 3) Soit I un intervalle (ouvert), f étant continue, $f^{-1}(I)$ est un ouvert de \mathbb{R} , et est donc borélien d'après la question précédente. On en déduit que $I \in \mathcal{G}$.
- 4) \mathcal{G} étant une tribu qui contient les intervalles, par définition elle contient la tribu Borélienne, qui est engendrée par les intervalles. Or par construction, $\mathcal{G} \subset \mathcal{B}(\mathbb{R})$, donc $\mathcal{G} = \mathcal{B}(\mathbb{R})$. Cela signifie que pour tout $B \in \mathcal{B}(\mathbb{R})$, $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$, donc f est une fonction mesurable.

□

Définition 8 : Mesure (positive)



Soit (E, \mathcal{T}) un espace mesurable, une application $\mu : \mathcal{T} \rightarrow \mathbb{R}_+ \cup \{\infty\}$ est une *mesure* si les deux conditions suivantes sont satisfaites :

- l'ensemble vide a mesure nulle, $\mu(\emptyset) = 0$,
- μ est σ -additive, c'est-à-dire que pour toute famille dénombrable $(A_n)_{n \in \mathbb{N}}$ d'ensembles disjoints deux-à-deux (i.e. tels que $A_n \cap A_m = \emptyset$ si $n \neq m$), on a

$$\mu(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

Le triplet (E, \mathcal{T}, μ) est alors appelé *espace mesuré*.

Notation : On notera dans toute la suite du cours par " \sqcup " les unions disjointes, c'est-à-dire que

$$[A = B \sqcup C] \iff [A = B \cup C \text{ et } B \cap C = \emptyset].$$



ATTENTION : il ne faut pas confondre fonction mesurable et mesure ! Les deux sont des fonctions, mais une fonction mesurable prend pour variable un élément $e \in E$, alors qu'une mesure prend pour variable un événement $A \in \mathcal{T}$.

Définition 9 : Mesures finies, σ -finies, de probabilité

Soit (E, \mathcal{T}, μ) un espace mesuré. On dit que la mesure μ est

- *finie* si $\mu(E) < \infty$.
- σ -*finie* si il existe une *partition* dénombrable $E = \sqcup_{i \geq 1} E_i$, telle que pour tout $i \in \mathbb{N}$, $E_i \in \mathcal{T}$ et $\mu(E_i) < \infty$.
- une *mesure de probabilités* si $\mu(E) = 1$.

Exemple : la *mesure de Lebesgue* sur \mathbb{R} , notée λ est l'unique mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout $a \leq b$, on ait

$$\lambda([a, b]) = b - a.$$

en d'autres termes, la mesure de Lebesgue d'un intervalle est la *longueur* de cet intervalle. Cette mesure est un objet fondamental à la fois en théorie de la mesure et en théorie des probabilités : comme on le verra plus tard dans le cours, elle permet d'étendre la notion d'intégrale au sens de Riemann à toute fonction mesurable (borélienne).

La mesure de Lebesgue sur \mathbb{R} est σ -*finie*. La mesure de Lebesgue sur $[0, 1]$, définie par $\lambda_{[0,1]}(A) = \lambda(A \cap [0, 1])$, est une mesure de probabilité.

Remarque : L'existence et l'unicité de la mesure de Lebesgue telle que définie ci-dessus n'est *a priori* pas garantie. La démonstration de l'existence et l'unicité de λ n'est pas à notre programme, et donc nous admettrons donc la "définition" ci-dessus comme la définition de la mesure de Lebesgue (cf paragraphe 3.3).

Exemple : Soit (E, \mathcal{T}) un ensemble mesurable, la *mesure de Dirac* en x , notée δ_x , est la mesure définie par

$$\delta_x(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon.} \end{cases}$$

1.3 Suites d'événements

On finit cette section par quelques notions et résultats utiles sur les mesures et sur les suites d'événements.

Proposition 10 : Propriétés générales des mesures.

Soit (E, \mathcal{T}, μ) un espace mesuré.

1. (Monotonie \star) Si A et B sont mesurables et $A \subset B$, alors $\mu(A) \leq \mu(B)$.

2. (Sous-additivité ★) Pour toute suite d'événements $A_n \in \mathcal{F}^{\mathbb{N}}$, on a

$$\mu(\cup_{n \in \mathbb{N}} A_n) \leq \sum_{n \in \mathbb{N}} \mu(A_n).$$

3. (Continuité par rapport aux unions croissantes) Soit (A_n) une suite croissante (pour l'inclusion) d'ensembles mesurables, i.e. $A_n \subset A_{n+1} \forall n$. Alors, on a

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu(\cup_{n \in \mathbb{N}} A_n).$$

4. (Continuité par rapport aux intersections décroissantes) Par ailleurs, soit (A_n) une suite décroissante (pour l'inclusion) d'ensembles mesurables, i.e. $A_{n+1} \subset A_n \forall n$. Alors, **en supposant qu'il existe $n_0 \in \mathbb{N}$ tel que $\mu(A_{n_0}) < \infty$** , on a

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu(\cap_{n \in \mathbb{N}} A_n).$$

Preuve :

1. On écrit $\mu(B) = \mu(A) + \mu(B \setminus A)$, le dernier terme étant positif, l'inégalité suit.
2. On définit $B_n = A_n \setminus \cup_{k < n} A_k$. Les B_n sont disjoints, et $\sqcup_{n \in \mathbb{N}} B_n = \cup_{n \in \mathbb{N}} A_n$. Par ailleurs, pour tout n , $B_n \subset A_n$, et donc par monotonie $\mu(B_n) \leq \mu(A_n)$. On obtient

$$\mu(\cup_{n \in \mathbb{N}} A_n) = \mu(\cup_{n \in \mathbb{N}} B_n) = \sum_{n \in \mathbb{N}} \mu(B_n) \leq \sum_{n \in \mathbb{N}} \mu(A_n).$$

3. Soit (A_n) une suite croissante, on définit la suite d'événements incompatibles $B_0 = A_0$ et $B_n = A_n \setminus A_{n-1}$ pour tout $n \geq 1$. On a alors, par σ -additivité de la mesure μ

$$\mu(\cup_{n \in \mathbb{N}} A_n) = \mu(\cup_{n \in \mathbb{N}} B_n) = \sum_{n \in \mathbb{N}} \mu(B_n) = \lim_{N \rightarrow \infty} \sum_{n \leq N} \mu(B_n) = \lim_{N \rightarrow \infty} \mu(\cup_{n \leq N} B_n) = \lim_{N \rightarrow \infty} \mu(A_N).$$

4. Soit (A_n) une suite décroissante, on définit la suite d'événements croissante $B_n = A_{n_0} \setminus A_{n_0+n}$, à laquelle on peut appliquer l'identité précédente.

$$\begin{aligned} \mu(A_{n_0}) - \lim_{n \rightarrow \infty} \mu(A_n) &= \lim_{n \rightarrow \infty} \mu(B_n) \\ &= \mu(\cup_{n \in \mathbb{N}} B_n) = \mu(A_{n_0} \setminus \cap_{n \geq n_0} A_n) = \mu(A_{n_0}) - \mu(\cap_{n \geq n_0} A_n) = \mu(A_{n_0}) - \mu(\cap_{n \geq 0} A_n). \end{aligned}$$

□

Exercice 2

Trouver un contre exemple à la seconde affirmation si l'on ne suppose plus qu'un des A_n est de mesure finie.

SOLUTION : On peut considérer $A_n =]n, +\infty[$: chacun des A_n a une mesure infinie, pourtant l'intersection des A_n est vide, et a donc une mesure nulle. □

Définition 11 : Limsup et liminf d'ensembles

Soit (E, \mathcal{T}) un espace mesurable. Soit $(A_n)_{n \in \mathbb{N}} \in \mathcal{T}^{\mathbb{N}}$ une suite d'ensembles mesurables. On définit les ensembles

$$\limsup A_n = \bigcap_{n \geq 0} \bigcup_{p \geq n} A_p \quad \text{et} \quad \liminf A_n = \bigcup_{n \geq 0} \bigcap_{p \geq n} A_p.$$

Remarque : Comme vous le montrerez en TD, la $\limsup A_n$ est l'ensemble des éléments de Ω qui appartiennent à une infinité de A_n . La $\liminf A_n$ est l'ensemble des éléments de Ω qui appartient à tous les A_n à partir d'un certain rang.

Proposition 12 : Lemme de Borel-Cantelli

Soit (E, \mathcal{T}, μ) un espace mesuré, alors pour toute suite d'événements $A_n \in \mathcal{T}^{\mathbb{N}}$ telle que $\sum_{n=1}^{\infty} \mu(A_n) < \infty$, on a

$$\mu(\limsup A_n) = 0.$$

En d'autres termes, si une suite d'événements est de somme de probabilité finie, avec probabilité 1, un nombre fini de ces événements se réalisent.

Preuve : On définit $B_n = \bigcup_{p \geq n} A_p$. C'est une suite décroissante d'événements, et par hypothèse ils sont tous de mesure finie puisque $\mu(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n) < \infty < \infty$. On peut donc appliquer la Propriété 10, on obtient par sous-additivité

$$\mu(\limsup A_n) = \mu(\bigcap_{n \in \mathbb{N}} B_n) = \lim_{n \rightarrow \infty} \mu(B_n) \leq \lim_{n \rightarrow \infty} \sum_{p \geq n} \mu(A_p) = 0,$$

puisque la série $\sum \mu(A_p)$ est absolument convergente. \square

2 Variable aléatoire

2.1 Loi d'une variable aléatoire

Avec ces éléments de théorie de la mesure, nous pouvons maintenant introduire plus généralement la notion de *variable aléatoire (v.a.)*.

Définition 13 : Espace de probabilités, variables aléatoires

★

Un *espace de probabilités* est un espace mesuré $(\Omega, \mathcal{F}, \mathbb{P})$, formé d'un ensemble Ω appelé *univers*, une tribu \mathcal{F} sur Ω et \mathbb{P} une mesure de probabilité sur (Ω, \mathcal{F}) . Une *variable aléatoire* est une fonction *mesurable* $X : \Omega \rightarrow E$, où (E, \mathcal{T}) est un *espace mesurable donné*. Dans le contexte de la théorie des probabilités, les éléments de \mathcal{F} sont appelés *événements*. L'espace E est alors appelé *espace d'états* de la variable aléatoire X .

La fonction X étant mesurable, on peut alors définir sa *distribution*, ou *loi*, \mathbb{P}_X , comme la mesure sur (E, \mathcal{T}) définie pour tout ensemble mesurable $A \in \mathcal{T}$ par

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)).$$

Exercice 3

Montrer que, étant donnée une variable aléatoire X , l'application $\mathbb{P}_X : A \in \mathcal{T} \mapsto \mathbb{P}(X^{-1}(A))$ est bien une mesure, et qui plus est une mesure de probabilité, sur (E, \mathcal{T}) .

SOLUTION : On commence par remarquer $X^{-1}(\emptyset) = \emptyset$ et donc $\mathbb{P}_X(\emptyset) = \mathbb{P}(\emptyset) = 0$ car \mathbb{P} est une mesure. Par ailleurs, on a pour toute famille disjointe d'événements $X^{-1}(\sqcup_k A_k) = \sqcup_k X^{-1}(A_k)$, et donc la σ -additivité de \mathbb{P}_X découle de celle de \mathbb{P} :

$$\mathbb{P}_X(\sqcup_k A_k) = \mathbb{P}(\sqcup_k X^{-1}(A_k)) = \sum_k \mathbb{P}_X(A_k).$$

L'application \mathbb{P}_X est donc une mesure. C'est une mesure de probabilités car \mathbb{P} en est une, et $X^{-1}(E) = \Omega$.

□

Notation : On notera $\{X \in A\}$ l'événement $\{X \in A\} := X^{-1}(A) \in \mathcal{F}$, et par conséquent

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

Dans le contexte de deux variables aléatoires X et Y , on notera également

$$\{X \in A, Y \in B\} := \{X \in A\} \cap \{Y \in B\}.$$

En pratique, l'espace Ω n'est pas accessible, car il est gigantesque : on n'y a accès que par des observations, qui sont les *variables aléatoires*. Prenons l'exemple d'un lancer de dé : chaque élément ω de l'univers Ω contient l'ensemble des paramètres du lancer, toutes les positions et vitesses de la main qui lance le dé, la position et orientation du dé, toutes les caractéristiques de la surface sur lequel le dé est lancé, etc. En somme, tous les paramètres physiques de l'expérience qui peuvent avoir une influence sur son résultat. Comme on n'a pas accès à toutes ces données, on va donc *encoder* le résultat de l'expérience, à travers la *variable aléatoire* X ="résultat du dé", qui prend ses valeurs dans $\{1, 2, \dots, 6\}$. Si le dé est bien équilibré, on estime que la loi de cette variable aléatoire est uniforme sur $\{1, 2, \dots, 6\}$.

2.2 Rappels : variables aléatoires discrètes

Pour comprendre une variable aléatoire X , il nous faut maintenant comprendre sa distribution \mathbb{P}_X . Commençons par quelques rappels sur les lois discrètes : dans ce contexte, on se donne un espace E fini ou dénombrable, identifié, quitte à numéroter ses éléments, à $E = \{1, \dots, N\}$ (cas fini) ou $E = \mathbb{N}$ (cas infini dénombrable). On munit E de la *tribu discrète* $\mathcal{T} = \mathcal{P}(E)$.

Pour caractériser la distribution \mathbb{P}_X d'une v.a. discrète X , il suffit alors d'estimer $\mathbb{P}_X(\{e\}) = \mathbb{P}(X = e)$ pour chaque élément $e \in E$. Cela permet alors de caractériser \mathbb{P}_X sur toute la tribu discrète \mathcal{T} . Pour cette raison, la notion de tribu n'est pas nécessaire pour les variables *discrètes*. On verra toutefois qu'elle est fondamentale pour les variables *continues*.

Définition 14 : Variables aléatoires discrètes



Une *variable aléatoire discrète* est une fonction mesurable $X : \Omega \rightarrow E$ d'un *espace de probabilités* $(\Omega, \mathcal{F}, \mathbb{P})$ dans un espace mesurable (E, \mathcal{T}) fini ou dénombrable. La tribu \mathcal{T} étant discrète, la *loi*, ou *distribution*, de X est entièrement caractérisée par la famille $(\mu_k)_{k \in E}$, où $\mu_k := \mathbb{P}(X = k)$, que l'on représentera par un vecteur ligne $\mu = (\mu_1, \mu_2, \dots)$.

Définition 15 : Espérance et variance d'une v.a. discrète



Rappelons nous que E est vu comme un sous-ensemble de \mathbb{N} . L'*espérance*, ou *moyenne* d'une v.a. discrète $X : \Omega \rightarrow E$ de distribution μ est donnée par

$$\mathbb{E}(X) = \sum_{k \in E} k \mu_k.$$

Sa variance est donnée par

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \sum_{k \in E} k^2 \mu_k - \mathbb{E}(X)^2.$$

Ces deux quantités sont bien définies si E est un ensemble fini, et sont définies sous réserve d'absolue convergence de la série si $E = \mathbb{N}$.

Définition 16 : Probabilités conditionnelles et indépendance



Étant donnés deux événements A et B , en supposant que $\mathbb{P}(B) \neq 0$, on définit la *probabilité conditionnelle de A sachant B* , notée $\mathbb{P}(A | B)$ ou parfois $\mathbb{P}_B(A)$, par

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

Les événements A et B sont *indépendants* si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Exercice 4

Montrer que pour tout $B \in \mathcal{F}$, $\mathbb{P}_B : A \mapsto \mathbb{P}_B(A)$ est une mesure de probabilités.

SOLUTION : On remarque que $\emptyset \cap B = \emptyset$, et donc $\mathbb{P}_B(\emptyset) = 0$. Par ailleurs, si les (A_k) sont une famille d'événements disjoints, les $(B \cap A_k)_k$ sont aussi des événements disjoints, et donc \mathbb{P}_B est σ -additive car \mathbb{P} l'est. Finalement, $\mathbb{P}_B(\Omega) = \mathbb{P}(B)/\mathbb{P}(B) = 1$, \mathbb{P}_B est donc une mesure de probabilités.

□

Proposition 17 : Formule des probabilités totales

Étant donnés des événements $A \in \mathcal{F}$ et une partition $\Omega = \sqcup_{k=1}^n B_k$, où $B_1, \dots, B_n \in \mathcal{F}$ (on dira que $(B_k)_{k \leq n}$ est un *système complet d'événements*), on a

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A \mid B_k) \mathbb{P}(B_k).$$

Preuve : Les événements $A \cap B_k, A \cap B_{k'}$ étant disjoints deux à deux, par σ -additivité de la mesure \mathbb{P} , on peut écrire

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap (\sqcup_{k=1}^n B_k)) = \sum_{k=1}^n \mathbb{P}(A \cap B_k) \\ &= \sum_{k=1}^n \mathbb{P}(A \mid B_k) \mathbb{P}(B_k). \end{aligned}$$

□

Proposition 18 : Théorème de transfert pour les v.a. discrètes

Soit une fonction $\varphi : E \rightarrow \mathbb{N}$, alors pour toute v.a. discrète $X : \Omega \rightarrow E$, la fonction $\varphi(X) : \Omega \rightarrow \mathbb{N}$ est également une variable aléatoire discrète, et son espérance s'exprime en fonction de la distribution μ de X par

$$\mathbb{E}(\varphi(X)) = \sum_{k \in E} \varphi(k) \mu_k.$$

Preuve : On commence par montrer que $\varphi(X) : \omega \mapsto \varphi(X(\omega))$ est bien une variable aléatoire, c'est-à-dire qu'elle est mesurable. Pour cela, il faut montrer que $\varphi(X)^{-1}(k) \in \mathcal{F}$ pour tout $k \in \mathbb{N}$. Or

$$\varphi(X)^{-1}(k) = \{\omega \in \Omega \mid \varphi(X(\omega)) = k\} = \bigcup_{\substack{k' \in E \\ \varphi(k') = k}} X^{-1}(k'),$$

qui est une union dénombrable d'ensembles mesurables, et est donc mesurable.

La formule de transfert en découle immédiatement : par définition de $\mathbb{E}(\varphi(X))$, et en utilisant que la famille d'événements $(\{X = k'\})_{k' \in E}$ est un système complet d'événements,

$$\begin{aligned} \mathbb{E}(\varphi(X)) &= \sum_{k \in \mathbb{N}} k \mathbb{P}(\varphi(X) = k) = \sum_{\substack{k \in \mathbb{N}, \\ k' \in E}} k \mathbb{P}(\varphi(X) = k, X = k') \\ &= \sum_{\substack{k \in \mathbb{N}, \\ k' \in E}} \varphi(k') \mathbb{P}(\varphi(X) = k, X = k') = \sum_{k' \in E} \varphi(k') \mathbb{P}(X = k'). \end{aligned}$$

□

Définition 19 : Indépendance deux à deux de v.a.

Deux variables aléatoires X et Y sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, et à valeurs dans deux espaces discrets (au plus dénombrables) E, F sont indépendantes si pour tout $e \in E, f \in F$,

$$\mathbb{P}(X = e, Y = f) = \mathbb{P}(X = e)\mathbb{P}(Y = f).$$

2.3 Rappel : quelques lois discrètes classiques

Voici pour rappel les lois discrètes les plus importantes :

Définition 20 : Loi de Bernoulli

La loi de Bernoulli(p) modélise une expérience avec deux résultats possibles (succès $X = 1$, échec, $X = 0$), avec probabilité p de succès.

- Paramètre(s) $p \in [0, 1]$.
- Espace d'états $E = \{0, 1\}$.
- Distribution $\mathbb{P}(X = 1) = p, \mathbb{P}(X = 0) = 1 - p$.
- Espérance $\mathbb{E}(X) = p$, variance $\text{Var}(X) = p(1 - p)$.

Définition 21 : Loi binomiale



La loi Binomiale(n, p) modélise le nombre de succès parmi n tentatives, pour une expérience qui a une probabilité p de succès à chaque tentative. C'est la somme de n variables de Bernoulli(p) indépendantes.

- Paramètres $n \in \mathbb{N}, p \in [0, 1]$.
- Espace d'états $E = \{0, 1, \dots, n\}$.
- Distribution $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.
- Espérance $\mathbb{E}(X) = np$, variance $\text{Var}(X) = np(1 - p)$.

Définition 22 : Loi uniforme

La loi uniforme(n) modélise le choix au hasard d'un objet parmi n objets identiques. (exemple, lancer de dé)

- Paramètres $n \in \mathbb{N} \setminus \{0\}$.
- Espace d'états $E = \{1, \dots, n\}$.
- Distribution $\mathbb{P}(X = k) = 1/n$.
- Espérance $\mathbb{E}(X) = (n + 1)/2$, variance $\text{Var}(X) = (n^2 - 1)/12$.

Définition 23 : Loi géométrique



La loi Géométrique(p) modélise le nombre de tentatives pour avoir son premier succès, pour une expérience qui a une probabilité p de succès à chaque tentative.

- Paramètre $p \in [0, 1]$.
- Espace d'états $E = \mathbb{N} \setminus \{0\}$.
- Distribution $\mathbb{P}(X = k) = p(1 - p)^{k-1}$.
- Espérance $\mathbb{E}(X) = 1/p$, variance $\text{Var}(X) = (1 - p)/p^2$.

Définition 24 : Loi de Poisson

La loi de Poisson(λ) modélise le nombre d'événements arrivant dans un intervalle de temps donné, si ces dernières arrivent avec une fréquence moyenne, $1/\lambda$, connue.

- Paramètre $\lambda \in \mathbb{R}_+^*$.
- Espace d'états $E = \mathbb{N}$.
- Distribution $\mathbb{P}(X = k) = \lambda^k e^{-\lambda} / k!$.
- Espérance $\mathbb{E}(X) = \lambda$, variance $\text{Var}(X) = \lambda$.

2.4 Variables aléatoires réelles à densité

Après ces rappels sur les lois discrètes, intéressons nous maintenant aux lois continues. Dans tout ce qui suit, et sauf mention contraire, **nous nous placerons sur la droite réelle $E = \mathbb{R}$, munie de la tribu borélienne $\mathcal{B}(\mathbb{R})$** . Commençons par définir les lois continues dont la densité est C^1 par morceaux, que l'on peut introduire via l'intégrale de Riemann. L'intégrale de Riemann d'une fonction n'est définie *a priori* que sur un segment. Pour introduire des lois classiques sur \mathbb{R} , toutefois, on a besoin de pouvoir définir l'intégrale sur \mathbb{R} , que nous allons définir pour les fonctions C^1 par morceaux comme l'extension de l'intégrale de Riemann sur un segment.

Rappelons qu'une fonction est C^1 par morceaux sur \mathbb{R} si pour tous $a < b \in \mathbb{R}$, il existe $x_0 = a < x_1 < \dots < x_{n-1} < x_n = b$ tels que f est C^1 sur chaque intervalle $]x_{k-1}, x_k[$, et prolongeable en une fonction continue sur $[x_{k-1}, x_k]$.

Définition 25 : Intégrale sur \mathbb{R}

Soit une fonction f positive et C^1 par morceaux sur \mathbb{R} , Riemann-intégrable sur tout segment $[-A, A]$, telle que la limite

$$\lim_{A \rightarrow \infty} \int_{-A}^A f(x) dx$$

existe et soit finie. On dit alors que la fonction f est *intégrable* sur \mathbb{R} , et on

définit

$$\int_{\mathbb{R}} f(x)dx = \lim_{A \rightarrow \infty} \int_{-A}^A f(x)dx.$$

Une fonction f de classe C^1 sur \mathbb{R} est *intégrable* si la fonction *positive* $|f|$ (qui est C^1 par morceaux) est *intégrable*. On admettra alors que l'intégrale $\int_{\mathbb{R}} f(x)dx = \lim_{A \rightarrow \infty} \int_{-A}^A f(x)dx$ existe et est finie.

Définition 26 : Lois à densité C^1 par morceaux



Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est une *densité de probabilité* si elle est

- positive, i.e. $f(x) \in \mathbb{R}_+ \quad \forall x \in \mathbb{R}$.
- borélienne et intégrable sur \mathbb{R} , et satisfait $\int_{\mathbb{R}} f(x)dx = 1$.

Dans le cadre de ce cours, on considérera des densités de probabilités C^1 par morceaux.

Une *variable aléatoire réelle* $X : \Omega \rightarrow \mathbb{R}$ est dite à *densité* f_X si pour tout $a \leq b \in \mathbb{R}$,

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

Définition 27 : Fonction de répartition



Soit X une variable à densité f_X , C^1 par morceaux, on appelle *fonction de répartition* de X la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(y)dy.$$

Remarque : La première partie de cette identité, $F_X(x) = \mathbb{P}(X \leq x)$ est valide pour toutes variables aléatoires à valeurs dans \mathbb{R} , y compris les variables discrètes à valeurs dans \mathbb{N} . Elle définit la fonction de répartition pour toutes les variables aléatoires réelles. Toutefois, on utilise rarement la fonction de répartition pour les variables discrètes, pour lesquelles elle n'apporte pas grand chose.

Proposition 28 : Propriétés de la fonction de répartition, admis

La fonction de répartition F_X d'une v.a. réelle est croissante, continue à droite, et satisfait

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

Par ailleurs, elle caractérise entièrement la loi de la variable X , c'est-à-dire que si deux variables aléatoires X et Y ont même fonction de répartition, alors leurs distributions \mathbb{P}_X et \mathbb{P}_Y sont identiques.

Remarque : Nous reverrons plus tard ces propriétés de la fonction de répartition (cf. Propriété 52 et Exercice 11). En pratique, il est important de savoir reconnaître, éventuellement graphiquement, les fonctions de répartitions classiques, qui peuvent permettre d'un coup d'œil d'identifier les distributions correspondantes.

Définition 29 : Espérance et Variance de lois à densité

★

Soit X une variable à densité f_X , C^1 par morceaux. On suppose que les fonctions $|x|f_X(x)$ et $x^2f_X(x)$ sont intégrables sur \mathbb{R} . On définit l'espérance $\mathbb{E}(X)$ de X par

$$\mathbb{E}(X) = \int_{\mathbb{R}} xf_X(x)dx.$$

On définit la variance $\text{Var}(X)$ de X par

$$\text{Var}(X) = \int_{\mathbb{R}} x^2 f_X(x)dx - \mathbb{E}(X)^2.$$

Exercice 5

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction positive, on admettra le théorème de Fubini, qui garantit que pour tous intervalles $I, I' \subset \mathbb{R}$, l'ordre d'intégration sur $I \times I'$ n'a pas d'importance, c'est-à-dire que

$$\int_I \int_{I'} f(x, y) dx dy = \int_{I'} \int_I f(x, y) dy dx$$

Montrer que si X est une v.a. réelle, positive, à densité f_X dont l'espérance est bien définie, alors

$$\mathbb{E}(X) = \int_0^\infty (1 - F_X(y)) dy.$$

Indice : on pourra commencer par remarquer que $x = \int_0^x dy$, et utiliser la fonction

$$\mathbf{1}_{\{y \leq x\}} = \begin{cases} 1 & \text{si } y \leq x \\ 0 & \text{sinon.} \end{cases}.$$

2.5 Quelques lois classiques à densité C^1 par morceaux

Définition 30 : Loi uniforme sur un segment

La loi uniforme $U([a, b])$ modélise le tirage d'un nombre réel entre a et b .

- Paramètres $a < b \in \mathbb{R}$.
- Densité

$$f_{a,b}(x) = \begin{cases} 1/(b-a) & \text{si } x \in [a, b] \\ 0 & \text{sinon.} \end{cases}$$

— Espérance $\mathbb{E}(X) = \frac{a+b}{2}$, variance $\text{Var}(X) = \frac{(b-a)^2}{12}$.

Exercice 6

Soit X une variable aléatoire uniforme sur $[0, 1]$. Montrer que $Y = 1 - X : \omega \mapsto 1 - X(\omega)$ est une variable uniforme sur $[0, 1]$.

SOLUTION : Étant donné que X est une variable aléatoire uniforme, on calcule/connait sa fonction de répartition :

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x \geq 1. \end{cases}$$

On calcule maintenant la fonction de répartition de Y ,

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(1 - X \leq x) = \mathbb{P}(X \geq 1 - x) = 1 - \mathbb{P}(X < 1 - x) = 1 - F_X(1 - x).$$

car x est une variable à densité et donc $\mathbb{P}(x = \alpha) = 0$ pour tout réel α . un rapide calcul nous donne alors

$$F_Y(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } 0 \leq x \leq 1 = F_X(x). \\ 1 & \text{for } x \geq 1 \end{cases}$$

Les deux variables X et Y ont même fonction de répartition, elles ont donc également la même loi. \square

Définition 31 : Loi gaussienne/normale



La loi normale $\mathcal{N}(\mu, \sigma^2)$ est une loi fondamentale en théorie des probabilités parce qu'elle modélise, entre autres, les fluctuations de la moyenne empirique (cf. équation (3) ci-dessous) d'un grand nombre de tirages indépendants autour de leur espérance. C'est le *théorème central limite*, que nous verrons en fin de cours (cf. Théorème 66).

— Paramètres $\mu \in \mathbb{R}$ et $\sigma^2 > 0$.

— Densité

$$f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

— Espérance $\mathbb{E}(X) = \mu$, variance $\text{Var}(X) = \sigma^2$.

La *fonction de répartition* de la loi normale n'est pas une fonction explicite. Lorsque $\mu = 0$ et $\sigma = 1$, on parlera de loi *normale centrée réduite*, notée $\mathcal{N}(0, 1)$.

Définition 32 : Loi exponentielle



La loi exponentielle représente le temps d'attente pour des événements indépendants intervenant à fréquence moyenne donnée. Elle est très utile pour modéliser des files d'attente.

- Paramètre $\lambda > 0$.
- Densité $f_\lambda(x) = \lambda e^{-\lambda x}$, fonction de répartition $F_\lambda(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$
- Espérance $\mathbb{E}(X) = 1/\lambda$, variance $\text{Var}(X) = 1/\lambda^2$.

Exercice 7 : absence de mémoire de la loi exponentielle

Barnabé attend le bus. Ayant étudié en détail les transports publics, il sait que le temps, exprimé en heures, qui sépare le passage de deux bus successifs suit une loi exponentielle de paramètre $\lambda = 1$.

- 1) *Ilevia* informe Barnabé qu'un bus est passé à 16h. Quelle est la probabilité que le suivant arrive avant 16h30?
- 2) Un bus étant passé à 16h, Barnabé arrive à l'arrêt à 16h30, et remarque que le bus suivant n'est pas encore passé. Quelle est la probabilité que Barnabé attende moins de 30min le bus suivant?
- 3) En déduire la propriété, dite *d'absence de mémoire*, de la loi exponentielle : soit X une v.a. de loi exponentielle, pour tous $a, b \geq 0$

$$\mathbb{P}(X \geq a + b \mid X \geq a) = \mathbb{P}(X \geq b).$$

SOLUTION :

1) On note X_k l'heure à laquelle passe le k -ème bus, et $Y_k = X_k - X_{k-1}$ le temps d'attente du k -ème bus. On suppose arbitrairement que le bus 0 passe à 16h, i.e. $X_0 = 16$. Comme une loi exponentielle ne peut pas être négative, on cherche

$$\mathbb{P}(X_1 \leq 16.5) = \mathbb{P}(Y_1 \leq 0.5) = \mathbb{P}(\text{Exp}(1) \leq 0.5) = F(0.5),$$

où $F(x) = 1 - e^{-x}$ représente la fonction de répartition d'une variable exponentielle de paramètre $\lambda = 1$. Par conséquent, $\mathbb{P}(X_1 \leq 16.5) = 1 - e^{-0.5}$.

2) On pose $X_0 = 0$, on cherche

$$\begin{aligned} \mathbb{P}(Y_1 \leq 1 \mid X_1 \geq 16.5) &= \frac{\mathbb{P}(Y_1 \leq 1, X_1 \geq 16.5)}{\mathbb{P}(X_1 \geq 16.5)} = \frac{\mathbb{P}(0.5 \leq Y_1 \leq 1)}{\mathbb{P}(Y_1 \geq 0.5)} \\ &= \frac{F(1) - F(0.5)}{1 - F(0.5)} = \frac{e^{-0.5} - e^{-1}}{e^{-0.5}} = F(0.5). \end{aligned}$$

3) Pour une loi exponentielle de paramètre λ , on a

$$\mathbb{P}(X \geq a + b \mid X \geq a) = \frac{\mathbb{P}(X \geq a + b)}{\mathbb{P}(X \geq a)} = \frac{1 - F(a + b)}{1 - F(a)} = e^{-\lambda(a+b)} / e^{-\lambda a} = e^{-\lambda b} = \mathbb{P}(X \geq b).$$

Autrement dit, étant donné une loi exponentielle X de paramètre λ , conditionnellement à $X \geq a$, la variable aléatoire $Y - a$ est suit également une loi exponentielle de paramètre λ . \square

Exercice 8 : lois ni discrètes ni à densité

Barnabé arrive à un guichet de poste. S'il n'y a pas de queue, ce qui arrive avec probabilité $p = 1/2$, on s'occupe de lui immédiatement et il n'attend pas. Sinon, il doit attendre un temps exponentiel (en heures) de paramètre $\lambda = 1$ avant que l'on s'occupe de lui.

4) On note X le temps d'attente (en heures) de Barnabé. La loi de X est-elle discrète? À densité?

5) Quelle est la probabilité que Barnabé attende une heure ou moins?

SOLUTION :

1) On note V l'événement "Le guichet est vide". La loi de X n'est pas à densité, puisque

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 0 | V)\mathbb{P}(V) + \mathbb{P}(X = 0 | V^c)\mathbb{P}(V^c) = 1 * 1/2 + 1/2\mathbb{P}(\text{Exp}(1) = 0) = 1/2.$$

Si X était à densité, on aurait $\mathbb{P}(X = 0) = \int_0^0 f_X(x)dx = 0$.

On montre maintenant que X n'est pas discrète. Soit $(a_k)_{k \in \mathbb{N}}$ une suite de réels non nuls. On calcule

$$\mathbb{P}(X \in \{a_k, k \in \mathbb{N}\}) = \sum_k \mathbb{P}(X = a_k | V)\mathbb{P}(V) = \frac{1}{2} \sum_k \mathbb{P}(\text{Exp}(1) = a_k) = 0,$$

En particulier, pour tout ensemble dénombrable Γ , $\mathbb{P}(X \in \Gamma) = \frac{1}{2} \mathbf{1}_{\{0 \in \Gamma\}} < 1$, la variable X n'est donc pas discrète.

2) On écrit

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X \leq 1 | V)\mathbb{P}(V) + \mathbb{P}(X \leq 1 | V^c)\mathbb{P}(V^c) = 1 * \frac{1}{2} + \frac{1}{2}F(1) = 1 - \frac{1}{2}e^{-1},$$

où F est la fonction de répartition d'une loi exponentielle de paramètre 1. \square

3 Intégrale par rapport à une mesure σ -finie.

3.1 Définition de l'intégrale d'une fonction mesurable



Les variables aléatoires discrètes et les variables aléatoires à densité ne sont que des exemples de variables aléatoires réelles. Comme le montre l'exercice précédent, on peut facilement construire des variables aléatoires qui ne sont ni l'un, ni l'autre. Nous allons maintenant construire l'intégrale d'une manière beaucoup plus générale que celle de Riemann, ce qui nous donnera accès à de nombreux résultats fort utiles. On se place pour l'instant dans un cadre aussi général que possible, en fixant un espace mesuré (E, \mathcal{F}, μ) , muni d'une mesure σ -finie μ (cf. Définition 9).

Lorsque $(E, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, où λ est la *mesure de Lebesgue* déjà évoquée à la fin de la Section 1, cette nouvelle intégrale sera l'*intégrale de Lebesgue*, beaucoup plus puissante que l'intégrale de Riemann. La construction de cette nouvelle intégrale se fait par étapes, commençons par définir des fonctions essentielles, les fonctions *indicatrices* d'ensembles mesurables.

Définition 33 : Fonction indicatrice d'un ensemble mesurable

Soit $A \in \mathcal{T}$ un ensemble mesurable, on définit la fonction $\mathbf{1}_A : E \rightarrow \mathbb{R}$, appelée *fonction indicatrice de A*, par

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \in A^c. \end{cases}$$

À noter que $\mathbf{1}_A(x)\mathbf{1}_B(x) = \mathbf{1}_{A \cap B}(x)$.

Les fonctions indicatrices forment la base de la construction de l'intégrale au sens de Lebesgue. Définissons maintenant l'intégrale d'une combinaison linéaire de fonctions indicatrices.

Définition 34 : Fonctions étagées

Une fonction $f : E \rightarrow \mathbb{R}$ est *étagée* s'il existe $n \in \mathbb{N}$, une famille d'ensembles mesurables $A_1, \dots, A_n \in \mathcal{T}$, et $a_1, \dots, a_n \in \mathbb{R}$ tels que

$$f = \sum_{i=1}^n a_i \mathbf{1}_{A_i}.$$

On note \mathcal{E} l'ensemble des fonctions étagées sur E . On note \mathcal{E}_+ l'ensemble des fonctions étagées positives, c'est-à-dire telles que $a_i \in \mathbb{R}_+ \forall i \leq n$.

Définition 35 : Intégrale d'une fonction étagée positive ★

Soit $f = \sum_{i=1}^n a_i \mathbf{1}_{A_i} \in \mathcal{E}_+$ une fonction étagée positive, on définit l'intégrale de f (contre μ) par

$$\int f d\mu = \sum_{i=1}^n a_i \mu(A_i).$$

Cette définition reste vraie si f n'est pas supposée positive, mais que la mesure μ est finie.

Exercice 9

Montrer qu'une fonction est étagée si et seulement si il existe une famille $B_1, \dots, B_k \in \mathcal{T}$ d'ensembles mesurables *disjoints*, et $b_1, \dots, b_k \in \mathbb{R}$ tels que

$$f = \sum_{i=1}^k b_i \mathbf{1}_{B_i}.$$

SOLUTION : Soit $f = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ une fonction étagée. Pour $\sigma \subset \{1, \dots, n\}$, on définit $\sigma^c = \{1, \dots, n\} \setminus \sigma$ et l'ensemble mesurable

$$B_\sigma = \left(\bigcap_{k \in \sigma} A_k \right) \cap \left(\bigcap_{\ell \in \sigma^c} A_\ell^c \right).$$

On vérifie facilement que pour $\sigma \neq \sigma'$, B_σ et $B_{\sigma'}$ sont disjoints. On définit également $b_\sigma = \sum_{k \in \sigma} a_k$, et on peut alors vérifier (par récurrence sur n par exemple) que

$$f = \sum_{\sigma \subset \{1, \dots, n\}} b_\sigma \mathbf{1}_{B_\sigma}$$

□

Définition 36 : Intégrale d'une fonction positive ★

Soit $f : E \rightarrow \mathbb{R}_+$ une fonction *mesurable positive*. On définit l'intégrale de f , notée $\int f d\mu \in \mathbb{R}_+ \cup \{+\infty\}$, par

$$\int f d\mu = \sup \left\{ \int g d\mu; g \in \mathcal{E}_+, g \leq f \right\}.$$

La fonction f est dite *intégrable* si $\int f d\mu < \infty$.

Définition 37 : Intégrale d'une fonction mesurable, espace \mathcal{L}^1 ★

Soit f une fonction mesurable, on définit les fonctions $f_+(x) = \max\{f(x), 0\}$ et $f_-(x) = \max\{-f(x), 0\}$. Ces fonctions sont positives et mesurables, et on a les identités

$$f = f_+ - f_- \quad \text{et} \quad |f| = f_+ + f_-.$$

On dit que la fonction f est *intégrable* si f_+ et f_- sont toutes les deux intégrables. L'intégrale de f est alors définie par

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu.$$

On note $\mathcal{L}^1(E, \mathcal{T}, \mu)$, ou plus simplement \mathcal{L}^1 ou $\mathcal{L}^1(\mu)$, l'ensemble des fonctions intégrables sur E .

Notation : Étant donné une fonction f et une mesure μ , on notera indifféremment

$$\int f d\mu, \quad \int f(x) d\mu(x) \quad \text{ou} \quad \int f(x) \mu(dx)$$

pour l'intégrale de f contre μ .

Définition 38 : Intégrale sur un ensemble mesurable

Soit f une fonction mesurable, et $A \in \mathcal{T}$ un ensemble mesurable, on définit, si elle existe,

$$\int_A f d\mu = \int_E f \mathbf{1}_A d\mu.$$

3.2 Convergence d'intégrales

Grâce à cette nouvelle intégrale, nous allons pouvoir obtenir des résultats de convergence généraux, et notamment des résultats forts d'interversion de limites et d'intégrale.

Définition 39 : Convergence simple et dans \mathcal{L}^1 ★

Une suite $(f_n)_{n \in \mathbb{N}}$ de fonctions de E dans \mathbb{R} converge simplement vers une fonction f si pour tout $x \in E$, $f_n(x) \rightarrow f(x)$.

On dira qu'une suite $(f_n)_{n \in \mathbb{N}}$ de fonctions mesurables converge dans \mathcal{L}^1 vers une fonction f si les f_n et f sont intégrables (i.e. $f_n, f \in \mathcal{L}^1$), et si

$$\lim_{n \rightarrow \infty} \int |f_n - f| d\mu$$

existe et vaut 0.



ATTENTION : aucune de ces deux convergences n'implique l'autre ! En fait, dans le contexte de convergence en intégrale, il est généralement préférable de parler de convergence presque-partout, puisque une fonction peut être modifiée arbitrairement sur un ensemble de mesure nulle sans changer son intégrale !

Exercice 10 : Densité des fonctions étagées

Soit $f : E \rightarrow \bar{\mathbb{R}}_+$ une fonction mesurable positive réelle.

- 1) Montrer que f est la limite simple d'une suite de fonctions étagées positives.
- 2) On ne suppose plus f positive, montrer que f est limite simple d'une suite de fonctions étagées.
- 3) Soit f une fonction mesurable bornée. Montrer qu'il existe une suite de fonctions étagées qui converge uniformément vers f .

SOLUTION :

- 1) On définit pour tout $n \geq 0$ et pour tout $1 \leq k \leq n^2$ l'intervalle $I_{n,k} = [(k-1)/n, k/n[$. Posons $A_{n,k} = f^{-1}(I_{n,k})$ et $A_{n,\star} = f^{-1}([n, +\infty])$. On construit alors

$$\hat{f}_n = n \mathbf{1}_{A_{n,\star}} + \sum_{k=1}^{n^2} \frac{k-1}{n} \mathbf{1}_{A_{n,k}}.$$

C'est une suite de fonctions étagées, dont la limite simple est f .

- 2) La partie positive et la partie négative de f , f_+ et f_- sont mesurables positives, on peut donc leur appliquer le 1).
- 3) Notons avec \pm quantités et ensembles relatives à f_{\pm} . Si f est bornée, il existe n tel que $A_{n,\star}^{\pm} = \emptyset$. Or pour tout $k \leq n^2$, pour tout $x \in A_{n,k}$,

$$|\hat{f}_{\pm,n}(x) - f_{\pm}(x)| \leq 1/n.$$

La convergence de \hat{f}_n vers f est donc uniforme. □

Définition 40 : Ensembles négligeables

On dit qu'un ensemble $A \subset E$ est *négligeable* s'il existe un ensemble mesurable $N \supset A$ satisfaisant $\mu(N) = 0$. On dira qu'une propriété est vraie *presque-partout* si elle est vraie partout sauf sur un ensemble négligeable.

Remarque : La limite dans \mathcal{L}^1 d'une suite de fonctions (f_n) est toujours définie à un ensemble négligeable près : en effet, si $f = g$ presque-partout, et si (f_n) converge vers f dans \mathcal{L}^1 , alors (f_n) converge aussi vers g dans \mathcal{L}^1 .

Proposition 41 : Propriétés de l'intégrale, admis



Linéarité : Pour deux fonctions mesurables f et g , et réels α et β , on admettra que la fonction $\alpha f + \beta g$ est mesurable. l'intégrale est linéaire, c'est-à-dire qu'on a

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

Monotonie : l'intégrale est monotone, c'est-à-dire que si $f \leq g$ sont deux fonctions mesurables, on a

$$\int f d\mu \leq \int g d\mu.$$

Définition 42 : lim inf et lim sup d'une suite de réels

Soit $(a_n)_{n \in \mathbb{N}}$ une suite réelle. Remarquons que les suites de termes généraux $v_n = \inf_{p \geq n} a_p$ et $w_n = \sup_{p \geq n} a_p$ sont respectivement croissante et décroissante, et par conséquent admettent une limite. On définit alors

$$\liminf_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} \inf_{p \geq n} a_p \quad \text{et} \quad \limsup_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} \sup_{p \geq n} a_p.$$

Ces deux quantités sont respectivement la plus petite et la plus grande *valeur d'adhérence* de la suite a_n .

Étant donnée une suite de fonctions réelles $(f_n)_{n \in \mathbb{N}}$, $f_n : E \rightarrow \mathbb{R}$, on définit de manière analogue les fonctions $\liminf f_n$ et $\limsup f_n$ en posant pour tout $x \in E$

$$\liminf f_n : x \mapsto \liminf_{n \rightarrow \infty} f_n(x) \quad \text{et} \quad \limsup f_n : x \mapsto \limsup_{n \rightarrow \infty} f_n(x).$$

Grâce à cette définition, on donne un premier résultat très pratique de convergence d'intégrales, le Lemme de Fatou.

Proposition 43 : Lemme de Fatou, admis

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions *mesurables positives* $E \rightarrow \mathbb{R}$. La fonction

$f = \liminf f_n$ est mesurable et

$$\int_E \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu.$$

Donnons deux exemples où l'inégalité est stricte. Définissons les fonctions sur \mathbb{R}

$$f_n(x) = n \mathbf{1}_{]0, 1/n[}(x) \quad \text{et} \quad g_n(x) = (1/n) \mathbf{1}_{[0, n]}(x),$$

les $(f_n)_{n \in \mathbb{N}}$ et les $(g_n)_{n \in \mathbb{N}}$ sont toutes d'intégrale 1, mais $\liminf f_n = \liminf g_n \equiv 0$, et donc $\int_E \liminf_{n \rightarrow \infty} f_n d\mu = 0$.

On énonce maintenant deux résultats fondamentaux, les théorèmes de convergence monotone et dominée.

Théorème 44 : Théorème de convergence monotone, admis



Soit $(f_n)_{n \in \mathbb{N}}$ une suite *croissante* de fonctions *mesurables positives* de E dans \mathbb{R} , c'est-à-dire que $\forall x \in E, \forall n \in \mathbb{N}, f_{n+1}(x) \geq f_n(x)$. Alors, il existe une fonction *mesurable* f qui est limite simple de $(f_n)_{n \in \mathbb{N}}$, et on a

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

À noter que l'on ne fait aucune hypothèse d'intégrabilité, les deux quantités ci-dessus peuvent être infinies. De la même manière, f est à valeurs dans $\mathbb{R} \cup \{+\infty\}$.

Remarque : en appliquant ce théorème à la suite $(-f_n)_{n \in \mathbb{N}}$, on voit immédiatement que le théorème de convergence monotone s'applique également à une suite *décroissante* de fonctions *néglatives*.

Remarque : Nous ne verrons pas la preuve de ce théorème, toutefois grâce au Lemme de Fatou, la convergence est très facile à obtenir. Notons $I_n = \int f_n d\mu$ et $I = \int f d\mu$. La suite $(f_n)_{n \in \mathbb{N}}$ étant croissante, on a $f_n \leq f$, et donc par monotonie $I_n \leq I$. L'inégalité $I \leq I_n$ est une conséquence directe du Lemme de Fatou, puisque pour toute suite réelle $(u_n)_{n \rightarrow \infty}$ admettant une limite, on a

$$\liminf_{n \rightarrow \infty} u_n = \limsup_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} u_n.$$

On énonce maintenant un des résultats les plus important de convergence sous le signe intégral. Ce résultat, très utile, énonce que si une suite de fonction et sa limite simple peuvent toutes majorées par une fonction intégrable, alors la convergence a également lieu dans \mathcal{L}^1 .

Théorème 45 : Théorème de convergence dominée, admis



Soit une suite $(f_n)_{n \in \mathbb{N}}$ de fonctions de E dans \mathbb{R} convergeant simplement vers une fonction f . On suppose qu'il existe une fonction intégrable $g \in \mathcal{L}^1$ telle que

$$|f_n(x)| \leq g(x) \quad \forall n \in \mathbb{N}, \forall x \in E.$$

Alors, f_n converge vers f dans \mathcal{L}^1 . En particulier,

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

Remarque : Reprenons les exemples d'inégalité stricte du Lemma de Fatou,

$$f_n(x) = n\mathbf{1}_{]0,1/n[}(x) \quad \text{et} \quad g_n(x) = (1/n)\mathbf{1}_{[0,n]}(x),$$

qui convergent simplement vers $f \equiv g \equiv 0$ sur $]0, +\infty[$. Comme l'inégalité est stricte pour le Lemme de Fatou, en particulier le théorème de convergence dominée de doit pas pouvoir s'appliquer; en réalité, une domination naturelle de ces deux suites de fonctions serait par la fonction $h : x \mapsto 1/x$ sur l'intervalle $]0, +\infty[$. On peut rapidement vérifier que $|f_n|, |g_n| \leq h$, mais la fonction h n'est *pas* intégrale, que ce soit en 0 (suite $(f_n)_n$) ou en $+\infty$ (suite $(g_n)_n$).

3.3 Intégrale de Lebesgue, intégrale de Riemann

Une mesure fondamentale qui nous intéressera dans le cadre des probabilités réelles est la *mesure de Lebesgue*, sur laquelle nous revenons maintenant.

Théorème 46 : Théorème de Carathéodory, admis



Il existe une unique mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout $a < b \in \mathbb{R} \cup \{-\infty, +\infty\}$,

$$\lambda([a, b]) = b - a$$

Cette mesure est appelée *mesure de Lebesgue*.

L'intégrale contre la mesure de Lebesgue coïncide avec l'intégrale de Riemann pour toute fonction telle que les deux intégrales sont bien définies (typiquement, les fonctions continues par morceaux sur un segment). On pourra donc écrire pour la mesure de Lebesgue $dx = d\lambda(x) = \lambda(dx)$ pour l'élément infinitésimal d'intégration. En l'absence d'ambiguïté, lorsque l'on intègre contre la mesure de Lebesgue, on utilisera la notation la plus simple, c'est-à-dire la première, i.e.

$$\int f d\lambda = \int f(x) dx.$$

En dehors de la construction différentes des deux intégrales, comme on l'a vu, l'intérêt fondamental de la mesure de Lebesgue est que l'intégrale par rapport à λ est définie pour la classe des fonctions *mesurables*, qui est une classe bien plus générale que celle des fonctions continues par morceau sur un segment!

Nous n'allons pas démontrer le théorème de Carathéodory. Toutefois, il est important de savoir qualifier les mesures, en particulier les mesures sur la tribu borélienne. Les boréliens généraux étant extrêmement difficiles à caractériser, on va avoir besoin d'outils pour montrer des égalités entre mesures, pour pouvoir dire que si deux mesures coïncident sur une classe d'événements C , elles sont égales.

Définition 47 : π -système

Une classe C de parties de E est un π -système si elle est stable par intersection finie :

$$\{A \in C \text{ et } B \in C\} \Rightarrow \{A \cap B \in C\}.$$

Exemple : L'ensemble des intervalles ouverts est un π -système, de même que l'ensemble des singletons.

Définition 48 : Classe monotone

Une classe \mathcal{M} de parties de E est un λ -système, ou *classe monotone* si elle contient E , et est stable différence et par union croissante :

$$\{A \in \mathcal{M} \text{ et } B \in \mathcal{M} \text{ et } B \subset A\} \Rightarrow \{A \setminus B \in \mathcal{M}\}.$$

$$\{\forall n \geq 0, A_n \in \mathcal{M} \text{ et } A_n \subset A_{n+1}\} \Rightarrow \{\cup_{n \geq 0} A_n \in \mathcal{M}\}.$$

Exemple : Étant données deux mesures de probabilité \mathbb{P} et \mathbb{Q} sur (Ω, \mathcal{F}) , la classe $\mathcal{M} = \{A \in \mathcal{F} \mid \mathbb{P}(A) = \mathbb{Q}(A)\}$ est une classe monotone.

Lemme 49 : Lemme des classes monotones (admis)

La plus petite classe monotone contenant le π -système C est la tribu $\sigma(C)$ engendrée par C .

Corollaire 50 : Unicité des mesures de probabilités



Deux mesures de probabilité \mathbb{P} et \mathbb{Q} sur (Ω, \mathcal{F}) qui coïncident sur un π -système $C \subset \mathcal{F}$ coïncident également sur la tribu $\sigma(C)$ engendrée par C .

Preuve : Considérons deux mesures de probabilité \mathbb{P} et \mathbb{Q} qui coïncident sur un π -système $C \subset \mathcal{F}$. La classe $\mathcal{M} = \{A \in \mathcal{F} \mid \mathbb{P}(A) = \mathbb{Q}(A)\}$ est une classe monotone contenant C , d'après le Lemme des classes monotones, $\mathcal{M} = \sigma(C)$. \square

En pratique, ce corollaire est très utile! Il signifie que pour caractériser une mesure (par exemple la loi d'une variable aléatoire), il n'y a pas besoin de la connaître sur la tribu tout entière, mais il suffit de la caractériser sur un π -système engendrant cette tribu. La tribu borélienne, par exemple, est engendrée par le π -système

$$C = \{] - \infty, x], x \in \mathbb{R}\},$$

on en déduira plus loin dans le cours que la fonction de répartition caractérise entièrement la distribution d'une variable aléatoire!

4 Variables aléatoires réelles

4.1 Variable aléatoire réelles et fonction de répartition

On a maintenant tous les outils nécessaires pour introduire les probabilités continues. Dans tout ce qui suit, on fixe un *espace de probabilités* $(\Omega, \mathcal{F}, \mathbb{P})$.

On dira qu'une propriété \mathcal{P} est vérifiée *presque-sûrement* s'il existe $F \in \mathcal{F}$ tel que $\mathbb{P}(F) = 1$ et tel que \mathcal{P} est vérifiée pour tout $\omega \in F$.

Définition 51 : variable aléatoire réelle, fonction de répartition ★

Comme nous l'avons déjà vu dans la Déf. 13, une variable aléatoire réelle est une fonction mesurable $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On note alors $\mathbb{P}_X(\cdot) = \mathbb{P}(X \in \cdot)$ sa distribution, définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Les *variables à densité* introduites dans le paragraphe 2.4 sont un cas particulier de variables aléatoires réelles X pour lesquelles $\mathbb{P}_X(A) = \int_A f_X(x)dx$, pour une fonction f_X borélienne. La fonction f_X est alors appelée la *densité par rapport à la mesure de Lebesgue* de la loi \mathbb{P}_X , et on écrit formellement $\mathbb{P}_X(dx) = f_X(x)dx$.

La *fonction de répartition* de la variable aléatoire X est la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$, définie par

$$F_X(x) = \mathbb{P}(X \leq x).$$

Proposition 52 : Propriétés de la fonction de répartition ★

La fonction de répartition F_X d'une v.a. réelle est croissante, continue à droite, et satisfait

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

Enfin, la fonction de répartition caractérise entièrement la loi d'une v.a. réelle, c'est-à-dire que si X et Y ont même fonction de répartition, $F_X = F_Y$, alors $\mathbb{P}_X = \mathbb{P}_Y$.

Preuve : La croissance est immédiate, et cette dernière garantit l'existence des deux limites, ainsi que l'identité

$$\lim_{x \in \mathbb{R} \rightarrow -\infty} F_X(x) = \lim_{n \in \mathbb{N} \rightarrow -\infty} F_X(n).$$

La continuité à droite est une conséquence directe de la continuité des mesures par rapport aux intersections décroissantes : pour toute suite positive $\varepsilon_n \rightarrow 0$, pour tout $x \in \mathbb{R}$,

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\cap_{n \in \mathbb{N}} \{X \leq x + \varepsilon_n\}) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq x + \varepsilon_n) = \lim_{n \rightarrow \infty} F(x + \varepsilon_n).$$

On applique ensuite la Proposition 10, puisque \mathbb{P}_X est une mesure de probabilités, et tout événement a donc probabilité finie,

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow -\infty} \mathbb{P}_X([-\infty, n]) = \lim_{n \rightarrow \infty} \mathbb{P}_X\left(\bigcap_{\substack{m \in \mathbb{N}, \\ m \leq n}}]-\infty, m[\right) = \mathbb{P}_X(\emptyset) = 0.$$

La seconde limite se montre de manière analogue.

Enfin, la dernière affirmation est une conséquence du Corollaire 50 du Lemme des classes monotones : le π -système

$$C = \{]-\infty, x], x \in \mathbb{R}\}$$

engendre la tribu des boréliens (c'est-à-dire $\sigma(C) = \mathcal{B}(\mathbb{R})$), et si $F_X = F_Y$, on a par définition que $\mathbb{P}_X = \mathbb{P}_Y$ sur C . Le Corollaire 50 garantit alors que $\mathbb{P}_X = \mathbb{P}_Y$ sur $\mathcal{B}(\mathbb{R})$ tout entier. \square

4.2 Moments d'une variable aléatoire réelle

La définition de l'intégrale de Lebesgue dans la section précédente nous permet désormais de définir l'espérance pour des variables aléatoires générales. Nous nous concentrons sur des variables aléatoires réelles, mais la notion d'espérance est beaucoup plus générale.

Définition 53 : Espérance d'une va. réelle ★

Rappelons (cf. Définition 13) que la loi \mathbb{P}_X d'une variable aléatoire $X : \Omega \rightarrow E$ est définie comme la mesure $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$, pour $A \in \mathcal{T}$. On dit qu'une variable aléatoire *réelle* $X : \Omega \rightarrow \mathbb{R}$ est *intégrable*, ou *dans \mathcal{L}^1* , si $\int_{\mathbb{R}} |x| \mathbb{P}_X(dx) < \infty$. On définit alors l'*espérance*, ou *moyenne*, de X par

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \mathbb{P}_X(dx). \quad (1)$$

On dira que la variable X est *centrée* si $\mathbb{E}(X) = 0$. On admettra que l'espérance a les mêmes propriétés que l'intégrale :

$$\text{monotonie : } \forall X, Y \text{ v.a.r. } X \leq Y \text{ p.s.} \implies \mathbb{E}(X) \leq \mathbb{E}(Y),$$

$$\text{linéarité : } \forall a, b \in \mathbb{R} \text{ et } X, Y \text{ v.a.r. } \mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Exercice 11

Dans cet exercice, on suppose à nouveau le théorème de Fubini (cf. Exercice 5) : deux intégrales successives d'une fonction positive peuvent être interverties.

1) Montrer que si X est intégrable, $\mathbb{E}(X) = \int_0^{+\infty} (1 - F_X(x))dx - \int_{-\infty}^0 F_X(x)dx$.

2) En commençant par une variable aléatoire X positive, montrer la formule

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega). \quad (2)$$

3) En déduire que l'espérance a les mêmes propriétés que l'intégrale : elle est *linéaire*

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y),$$

et *monotone*

$$X \leq Y \implies \mathbb{E}(X) \leq \mathbb{E}(Y).$$

SOLUTION : Pour traiter cet exercice, il faut sans cesse se ramener à des indicatrices d'événements, dont on connaît explicitement les intégrales.

1) On commence par montrer le résultat pour $X : \Omega \rightarrow \mathbb{R}^+$ positive. On remarque que $x = \int_0^x dy = \int_0^{+\infty} \mathbf{1}_{[0,x]}(y) dy$, et on peut alors écrire grâce au théorème de Fubini

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{+\infty} x \mathbb{P}_X(dx) = \int_0^{+\infty} \left[\int_0^{+\infty} \mathbf{1}_{[0,x]}(y) dy \right] \mathbb{P}_X(dx) \\ &= \int_0^{+\infty} \left[\int_0^{+\infty} \mathbf{1}_{[0,x]}(y) \mathbb{P}_X(dx) \right] dy. \end{aligned}$$

Or pour une variable positive, on peut écrire

$$\int_0^{+\infty} \mathbf{1}_{[0,x]}(y) \mathbb{P}_X(dx) = \int_0^{+\infty} \mathbf{1}_{\{y < x\}} \mathbb{P}_X(dx) = \mathbb{P}(X > y) = 1 - F_X(y).$$

Pour une variable X négative, on obtient de manière analogue

$$\mathbb{E}(X) = - \int_{-\infty}^0 F_X(y) dy$$

On sépare alors X en partie positive et partie négative, $X_+ = X\mathbf{1}_{X \geq 0}$, $X_- = X\mathbf{1}_{X \leq 0}$, et on vérifie que pour tout $t \geq 0$, $F_{X_+}(t) = F_X(t)$, et pour $t \leq 0$, $F_{X_-}(t) = F_X(t)$, ce qui conclut la preuve.

2) Pour une variable X positive, on écrit

$$\int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \int_{\Omega} \left[\int_0^{+\infty} \mathbf{1}_{[0,X(\omega)]}(y) dy \right] \mathbb{P}(d\omega),$$

et on inverse encore les deux intégrales par le théorème de Fubini, pour obtenir

$$\int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \int_0^{+\infty} \left[\int_{\Omega} \mathbf{1}_{[0,X(\omega)]}(y) \mathbb{P}(d\omega) \right] dy = \int_0^{+\infty} \mathbb{P}(X > y) dy = \int_0^{+\infty} (1 - F_X(y)) dy,$$

on retrouve donc le résultat de la question 1. Le cas où X est négatif se montre de manière analogue, ce qui prouve la formule alternative

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega)$$

3) Avec cette nouvelle formule, la linéarité et la monotonie découlent immédiatement de la linéarité et la monotonie de l'intégrale. \square

On peut maintenant énoncer le théorème de transfert pour des variables aléatoires générales, déjà énoncé pour les variables discrètes.

Théorème 54 : Théorème de transfert (admis)

Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mesurable. Alors, la variable aléatoire $\varphi(X)$ est intégrable (par rapport à \mathbb{P}) si et seulement si φ est intégrable par rapport à \mathbb{P}_X , et on a alors

$$\mathbb{E}(\varphi(X)) := \int_{\Omega} \varphi(X(\omega)) \mathbb{P}(d\omega) = \int_{\mathbb{R}} \varphi(x) \mathbb{P}_X(dx).$$

Exemple : soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle, Pour tout ensemble borélien B , on peut définir la variable aléatoire

$$\mathbf{1}_{\{X \in B\}} : \omega \mapsto \mathbf{1}_B(X(\omega)).$$

Par le théorème de transfert, son espérance est donnée par

$$\mathbb{E}(\mathbf{1}_{\{X \in B\}}) = \int_{\mathbb{R}} \mathbf{1}_B(x) \mathbb{P}_X(dx) = \int_B \mathbb{P}_X(dx) = \mathbb{P}(X \in B).$$

On vient de donner la définition, grâce à l'intégrale de Lebesgue, de l'espérance d'une variable aléatoire réelle. Voyons maintenant comment cette définition rejoint celle que nous avons déjà vu pour les variables à densité, et les variables discrètes. Commençons par une variable à densité X , dont la mesure \mathbb{P}_X est définie pour tout borélien B par

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \int_B \mathbb{P}_X(dx) = \int_B f_X(x) dx.$$

Cela nous permet d'écrire formellement $\mathbb{P}_X(dx) = f_X(x) dx$. En transposant cette identité dans notre définition générale de l'espérance, (1), on retrouve bien la Définition 29 de l'espérance d'une variable à densité :

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx.$$

Considérons maintenant les variables discrètes. Bien qu'on puisse les définir sur la tribu borélienne, cette dernière n'est pas un choix naturel pour le cas discret, comme on l'a déjà vu. On considère donc une variable aléatoire discrète $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Notez que l'on a muni \mathbb{N} de la tribu *discrète* $\mathcal{P}(\mathbb{N})$, ce qui pour des variables discrètes n'est pas un souci. On suppose dans un premier temps que X ne prend qu'un nombre fini de valeurs, $X \in \{0, \dots, n\}$. Alors, la variable aléatoire X est une *fonction étagée* sur $(\Omega, \mathcal{F}, \mathbb{P})$, puisqu'elle peut s'écrire comme

$$X(\omega) = \sum_{k=0}^n k \mathbf{1}_{\{X=k\}}(\omega),$$

et les événements $\{X = k\} = X^{-1}(\{k\})$ sont mesurables puisque X est une variable aléatoire. Or on connaît l'intégrale d'une fonction étagée contre une mesure finie \mathbb{P} (voir Définition 35), et on peut donc écrire

$$\mathbb{E}(X) = \int_{\{1, \dots, n\}} x \mathbb{P}_X(dx) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \sum_{k=0}^n k \mathbb{P}(X = k).$$

Si maintenant X est intégrable mais n'est pas bornée, on approche X par les fonctions étagées $X^{(n)} = X \mathbf{1}_{X \leq n}$. Alors, grâce au théorème de convergence dominée, qui s'applique car $|X|$ est intégrable, et on a à la fois $|X| \leq |X|$ et $|X^{(n)}| \leq |X|$, on peut écrire

$$\mathbb{E}(X) = \lim_{n \rightarrow \infty} \mathbb{E}(X^{(n)}) = \lim_{n \rightarrow \infty} \sum_{k=0}^n k \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k \mathbb{P}(X = k).$$

On voit bien dans ces deux exemples que l'espérance d'une variable aléatoire peut être vue comme une intégrale sur l'univers Ω (ce qui nous a permis de retrouver l'expression de l'espérance d'une variable aléatoire discrète), ou comme une intégrale sur l'espace d'arrivée de X , ce qui nous permet de retrouver l'expression de l'espérance des variables à densité.

Définition 55 : Moments d'une v.a. réelle ★

Étant donné $k \in \mathbb{N}$, le k -ème moment d'une variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ est donné, s'il existe, par

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

On dira que X admet un k -ème moment fini si X^k est intégrable. En particulier, une variable a son premier moment fini si et seulement si elle est *intégrable*. Pour une variable admettant un second moment fini, on dira aussi qu'elle est *de carré intégrable*.

Définition 56 : Variance d'une v.a. réelle ★

Soit X une variable aléatoire réelle de second moment fini, sa variance est définie par

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{\mathbb{R}} x^2 \mathbb{P}_X(dx) - \mathbb{E}(X)^2.$$

Son écart type est défini comme $\sigma(X) = \sqrt{\text{Var}(X)}$. On dira parfois que la variable X est *réduite* si $\sigma(X) = 1$.

4.3 Propriétés et inégalités classiques

On commence par une inégalité de convexité. Rapellons qu'une fonction φ sur \mathbb{R} est *convexe* si elle est au-dessus de toutes ses tangentes, c'est à dire si

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda \varphi(x) + (1 - \lambda)\varphi(y) \quad \forall x, y \in \mathbb{R}, \lambda \in [0, 1].$$

Une fonction φ de classe C^2 est convexe si et seulement si $\varphi'' \geq 0$.

Théorème 57 : Inégalité de Jensen, admis

Soit X une variable aléatoire, et φ une fonction convexe. Alors,

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

Cette inégalité reste valable si X n'est pas intégrable.

Remarque : On voit naturellement que l'inégalité est inversée si φ est concave, puisque

$$\varphi \text{ convexe} \iff -\varphi \text{ concave.}$$

Exercice 12

Utiliser l'inégalité de Jensen pour montrer que si X admet un k -ème moment, alors elle admet un j -ème moment pour tout $j \leq k$.

SOLUTION : Il suffit de remarquer que pour tout entiers positifs $1 \leq j \leq k$, la fonction $\varphi : x \mapsto x^{j/k}$ est concave sur \mathbb{R}_+ car $j/k \leq 1$. On en déduit que

$$\mathbb{E}(|X|^j) = \mathbb{E}\left[(|X|^k)^{j/k}\right] \leq \mathbb{E}\left[|X|^k\right]^{j/k} < \infty,$$

car X admet un k -ème moment par hypothèse. □

Proposition 58 : Inégalité de Markov ★

Soit $X : \Omega \rightarrow \mathbb{R}_+$ une variable aléatoire *positive*, pour tout $a > 0$, on a

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Remarque : cette inégalité reste valide lorsque X n'est pas intégrable, mais elle devient triviale, puisque le membre de droite est alors infini!

Preuve : Fixons $a > 0$, comme X est une v.a. *positive*, on commence par écrire l'inégalité

$$a\mathbf{1}_{\{X \geq a\}} \leq X,$$

valide pour tout ω . L'espérance étant monotone, on peut prendre l'espérance de part et d'autre :

$$\mathbb{E}(a\mathbf{1}_{\{X \geq a\}}) = a\mathbb{P}(X \geq a) \leq \mathbb{E}(X).$$

On divise ensuite de part et d'autre par a pour conclure la preuve. □

Exercice 13

Trouver un contre-exemple à l'inégalité de Markov si X n'est pas positive.

SOLUTION : Il suffit de prendre une variable centrée, comme une loi normale centrée réduite, pour laquelle $\mathbb{E}(X) = 0$ et pourtant $\mathbb{P}(X \geq a) > 0$ pour tout $a > 0$. □

L'inégalité de Markov est extrêmement utile, c'est l'outil standard pour contrôler la probabilité qu'une variable aléatoire intégrable soit grande. Une de ses conséquences est l'inégalité de Bienaymé-Chebychev, qui permet de contrôler la probabilité qu'une variable aléatoire soit loin de sa moyenne.

Proposition 59 : Inégalité de Bienaymé-Chebychev

Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire de second moment fini, et $a > 0$, on a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}$$

Remarque : Là encore, l'inégalité reste valide, mais inintéressante si X n'admet pas de second moment. Par ailleurs, l'inégalité de Bienaymé-Chebychev peut se réécrire de la manière suivante : pour tout $k > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq k\sigma) \leq \frac{1}{k^2},$$

où $\sigma < \infty$ est l'écart type de X .

Preuve : On commence par remarquer que

$$|X - \mathbb{E}(X)| \geq a \iff (X - \mathbb{E}(X))^2 \geq a^2.$$

On peut donc appliquer l'inégalité de Markov à la v.a. positive $Y := (X - \mathbb{E}(X))^2$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) = \mathbb{P}(Y \geq a^2) \underset{\text{inég. Markov}}{\leq} \frac{\mathbb{E}(Y)}{a^2} = \frac{\text{Var}(X)}{a^2}.$$

□

5 Loi des grands nombres et TCL

5.1 Convergence des variables aléatoires réelles

On définit pour les variables aléatoires des notions de convergence analogues aux convergences d'intégrales, avec un vocabulaire légèrement différent de celui des intégrales introduit dans la Définition 39.

Définition 60 : Convergence p.s., \mathcal{L}^1 , en loi

★

Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles *converge presque-sûrement*, ou *p.s.*, vers X , noté $X_n \xrightarrow{p.s.} X$, si il existe $A \in \mathcal{F}$ tel que $\mathbb{P}(A) = 1$ et $X_n(\omega)$ converge vers $X(\omega)$ pour tout $\omega \in A$. autrement dit

$$X_n \xrightarrow{p.s.} X \iff \mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1.$$

Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires *converge dans \mathcal{L}^1* vers une variable aléatoire X , noté $X_n \xrightarrow{\mathcal{L}^1} X$ si les X_n et X sont intégrables et si

$$\mathbb{E}(|X_n - X|) \xrightarrow{n \rightarrow \infty} 0.$$

Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires *converge en probabilités* vers une variable aléatoire X , noté $X_n \xrightarrow{\mathbb{P}} X$ si pour tout $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0.$$

Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires *converge en loi, ou en distribution*, vers une variable aléatoire X , noté $X_n \xrightarrow{(d)} X$ si pour toute fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée,

$$\mathbb{E}(\varphi(X_n)) \xrightarrow{n \rightarrow \infty} \mathbb{E}(\varphi(X)).$$

Remarque : La convergence en loi est différente des convergences en probas, *p.s.* et \mathcal{L}^1 , car elle porte sur les mesures \mathbb{P}_{X_i} et \mathbb{P}_X et non pas directement sur les variables aléatoires X_i et X . En particulier, la convergence en loi ne requiert pas que les v.a. X_i et X soient définies sur le même espace de probabilité.

Notation : Dans le cadre d'une convergence en loi, on identifiera parfois une variable aléatoire et sa distribution. Dans le contexte du *théorème central limite*, par exemple, introduit ci-dessous, on écrira " $Y_n \xrightarrow{(d)} \mathcal{N}(0, 1)$ " plutôt que " $Y_n \xrightarrow{(d)} N$ ", où N est une variable aléatoire de loi normale centrée réduite $\mathcal{N}(0, 1)$ ".

Remarque : Ces convergences ne sont pas toutes aussi fortes les unes que les autres! En fait, on peut montrer que

$$\left[X_n \xrightarrow{L^1 \text{ ou p.s.}} X \right] \implies \left[X_n \xrightarrow{(d)} X \right] \implies \left[X_n \xrightarrow{\mathbb{P}} X \right].$$

Attention toutefois, on n'a pas $X_n \xrightarrow{p.s.} X \implies X_n \xrightarrow{L^1} X$, ni l'implication inverse! Pour la première, il faut le théorème de convergence dominée (Théorème 45). Pour la seconde, il n'y a pas de théorème général, mais le contre exemple suivant montre que l'implication n'est pas vraie : on considère des bernouillis indépendantes X_n de paramètre $1/n$. On calcule facilement que $\mathbb{E}(|X_n|) = 1/n \rightarrow 0$, et donc $X_n \xrightarrow{L^1} 0$. Pourtant, par le second lemme de Borel-Cantelli (hors-programme), une infinité de X_n valent 1 avec probabilité 1 puisque $\sum_n 1/n = \infty$ et que les événements $\{X_n = 1\}$ sont indépendants. Il en est de même pour les événements $\{X_n = 0\}$, on a donc en particulier, presque-surement, que la suite $(X_n(\omega))_{n \in \mathbb{N}}$ ne converge pas!

Proposition 61 : Caractérisation de la convergence en loi

Soit $F_n = F_{X_n}$ la fonction de répartition de la v.a. X_n et F celle de X . Alors,

$$X_n \xrightarrow{(d)} X \iff (F_n(x) \rightarrow F(x), \forall x \text{ où } F \text{ est continue}).$$

En d'autres termes, pour des variables aléatoires réelles, la convergence en loi est équivalente à la convergence simple de la fonction de répartition.

5.2 Indépendance

Définition 62 : Variables aléatoires indépendantes



Deux variables aléatoires réelles X et Y sont *indépendantes* si pour tous boréliens $A, B \in \mathcal{B}(\mathbb{R})$,

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

Une famille $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires est dite *indépendante* si pour tout $k \in \mathbb{N}$, pour tous $n_1 \neq n_2 \neq \dots \neq n_k \in \mathbb{N}$, et pour tous boréliens $A_1, \dots, A_k \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(X_{n_i} \in A_i \forall i = 1 \dots, k) = \prod_{i=1}^k \mathbb{P}(X_{n_i} \in A_i).$$

Si la famille $(X_n)_{n \in \mathbb{N}}$ est *indépendante* et de plus les X_n ont tous la même loi \mathbb{P}_X , on dira que la famille $(X_n)_{n \in \mathbb{N}}$ est *indépendante et identiquement distribuée*, ou *i.i.d.*.

Notation : On notera $X \perp Y$ pour " X et Y sont indépendantes".



ATTENTION : Pour que la famille $(X_n)_{n \in \mathbb{N}}$ soit indépendante, il ne suffit pas que les $(X_n)_{n \in \mathbb{N}}$ soient indépendantes deux à deux, i.e. $X_n \perp X_m \forall n \neq m$. Considérons deux variables indépendantes X et Y de même loi, où $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$. Alors, en posant $Z = XY$, la famille (X, Y, Z) est indépendante deux à deux, mais pas indépendante.

Remarque : En pratique, pour vérifier l'indépendance, il suffit de considérer des boréliens de type $A =]a, +\infty[$. Par exemple,

$$X \perp Y \iff \mathbb{P}(X > a, Y > b) = \mathbb{P}(X > a)\mathbb{P}(Y > b) \quad \forall a, b \in \mathbb{R}.$$

Proposition 63 : Espérance et indépendance, admis



Soient X et Y deux variables aléatoires indépendantes à valeurs dans \mathbb{R} , et $f, g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions boréliennes. Alors, $f(X)$ et $g(Y)$ sont également indépendantes. Par ailleurs,

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

De manière analogue, si la famille $(X_n)_{n \in \mathbb{N}}$ est indépendante, alors $\forall k \in \mathbb{N}$, $\forall n_1 \neq n_2 \neq \dots \neq n_k \in \mathbb{N}$, et pour toutes fonctions boréliennes f_1, \dots, f_n , on a

$$\mathbb{E}\left(\prod_{i=1}^k f_i(X_{n_i})\right) = \prod_{i=1}^k \mathbb{E}(f_i(X_{n_i})).$$

Corollaire 64 : Variance et somme de variables indépendantes



Soient X et Y deux variables aléatoires indépendantes de carré intégrable, on a

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

Preuve : Par linéarité de l'espérance,

$$\begin{aligned}\text{Var}(X + Y) &= \mathbb{E}((X + Y)^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\ &= \mathbb{E}(X^2) + \mathbb{E}(Y^2) - 2\mathbb{E}(XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 + 2\mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

Par indépendance, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, ce qui montre l'identité. \square

5.3 Loi des grands nombres

Nous avons maintenant tous les outils nécessaires pour énoncer les deux principaux résultats de ce cours, la loi des grands nombres, et le théorème central limite. Pour cela, étant donné une suite i.i.d. de v.a. $(X_i)_{i \in \mathbb{N}}$, on définit sa *moyenne empirique*

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}. \quad (3)$$

Théorème 65 : Loi forte des grands nombres (LGN)



Soit $(X_i)_{i \in \mathbb{N}}$ une famille i.i.d. de variables aléatoires, on suppose X_1 intégrable. Alors, \bar{X}_n converge p.s. quand $n \rightarrow \infty$, et

$$\lim_{n \rightarrow \infty} \bar{X}_n = \mathbb{E}(X_1).$$



ATTENTION : il est fondamental que la suite soit indépendante pour pouvoir appliquer la loi des grands nombres! (de même que le théorème central limite ci-dessous) Prenons par exemple $X_1 \sim \text{Bernoulli}(1/2)$, et $X_i = X_1$ pour tout $i \in \mathbb{N}$ (c'est à dire que pour tout ω , $X_i(\omega) = X_1(\omega)$). Alors, la suite (X_i) est bien identiquement distribuée, puisque tous les X_i ont également pour loi une $\text{Bernoulli}(1/2)$, mais elle n'est pas indépendante. On remarque facilement que $\bar{X}_n = X_1$ p.s., et reste donc aléatoire, et donc \bar{X}_n ne converge pas p.s. vers $\mathbb{E}(X_1) = 1/2$.

Malheureusement, la loi forte des grands nombres ne dit rien sur la vitesse à laquelle \bar{X}_n converge vers X . Et pour cause, la vitesse de cette convergence dépend des propriétés d'intégrabilité de X_1 ! Par exemple, on peut montrer le résultat suivant.

Exercice 14 : Vitesse de convergence de la LFGN

On suppose que $(X_i)_{i \in \mathbb{N}}$ une famille i.i.d. de variables aléatoires, on suppose que X_1 admet un moment d'ordre deux. Montrer que

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| \geq \varepsilon) \leq \frac{\text{Var}(X_1)}{\varepsilon^2 n}.$$

SOLUTION : On applique l'inégalité de Bienaymé-Chebychev :

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2},$$

il suffit donc maintenant de calculer la variance de \bar{X}_n . Comme l'espérance est linéaire, $\text{Var}(aX) = a^2 \text{Var}(X)$. Grâce à cette affirmation, et en utilisant le Corollaire 64, on obtient donc que

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) \stackrel{(X_k) \text{ i.i.d.}}{=} \frac{n}{n^2} \text{Var}(X_1) = \frac{\text{Var}(X_1)}{n}.$$

□

Remarque : En réalité, si l'on suppose de plus fortes conditions d'intégrabilité sur X_1 , on peut montrer des convergences beaucoup plus rapides vers la moyenne. Par exemple, si $\mathbb{E}(e^{\lambda X_1}) < +\infty$ pour tout $\lambda \in \mathbb{R}$, le *théorème de Cramér* (complètement hors programme!) affirme qu'il existe une fonction $\Lambda(\varepsilon)$ telle que

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| \geq \varepsilon) \leq e^{-\Lambda(\varepsilon)n},$$

qui est une estimation beaucoup plus forte que celle que nous avons obtenue dans l'exercice précédent!

5.4 Théorème central limite et statistiques descriptives

Théorème 66 : Théorème central limite



On suppose que X_1 admet un moment d'ordre deux (c'est-à-dire que sa variance est finie). On note $\mu = \mathbb{E}(X)$ et $\sigma^2 = \text{Var}(X)$. Alors,

$$Y_n := \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{(d)} \mathcal{N}(0, 1),$$

où $\mathcal{N}(0, 1)$ est une loi normale centrée réduite introduite dans la Définition 31.

Le théorème central limite que nous venons de voir est le pilier des statistiques descriptives, et a de nombreuses applications. Nous allons illustrer le principe général des intervalles de confiance statistiques avec l'exercice suivant.

On cherche à déterminer la proportion de personnes en France qui sont actuellement infectées par le Covid. Pour cela, on va tester un certain nombre de personnes, que l'on suppose choisies uniformément dans la population. Évidemment, comme on n'a accès qu'à une partie de la population, le taux d'infections estimé ne sera pas exactement égal au taux d'infections général de la population. On veut donc estimer la probabilité que le taux d'infection en France soit dans un intervalle donné, dit *intervalle de confiance*, à l'aide de tests de coronavirus dans un échantillon de taille n .

Exercice 15

On associe à chaque individu $1 \leq i \leq n$ de l'échantillon une variable de Bernoulli B_i de paramètre p , valant 1 si l'individu i est infecté, et 0 sinon. Construire à l'aide de l'échantillon observé un intervalle I_n tel que

$$\mathbb{P}(p \in I_n) \geq 0.95.$$

Attention, contrairement aux exemples que l'on a vu précédemment, la quantité p est un *paramètre déterministe*, alors que l'intervalle $I_n = [a_n, b_n]$ est *aléatoire*!

SOLUTION : On suppose que la population française est assez grande pour considérer qu'elle est infinie (si, si!). Une proportion p de la population est infectée par le coronavirus. Le paramètre p est inconnu, notre objectif de l'estimer. On va donc tester un échantillon de la population, composé de n individus. On définit la proportion *observée* d'infections dans l'échantillon $p_n = \frac{1}{n} \sum_{i=1}^n B_i$.

Par le théorème central limite,

$$\mathbb{P}\left(\frac{\sqrt{n}|p_n - p|}{\sigma_p} \leq \varepsilon\right) \underset{n \rightarrow \infty}{\simeq} \mathbb{P}(|\mathcal{N}(0, 1)| \leq \varepsilon),$$

où $\sigma_p = \sqrt{p(1-p)}$ est l'écart-type de la loi de Bernoulli. Typiquement, on cherche à construire un intervalle de confiance à 95%, c'est à dire un intervalle dans lequel le paramètre p a une probabilité de 0.95 de se trouver. On va donc choisir ε aussi petit que possible tel que $\mathbb{P}(|\mathcal{N}(0, 1)| \leq \varepsilon) \geq 0.95$. En consultant des tables de lois normales, on obtient

$$\varepsilon = 1.96.$$

On déduit donc du théorème central limite

$$\mathbb{P}\left(p_n - \frac{1.96 \sigma_p}{\sqrt{n}} \leq p \leq p_n + \frac{1.96 \sigma_p}{\sqrt{n}}\right) \geq 0.95. \quad (4)$$

Malheureusement, dans les bornes ci-dessus, l'écart-type σ_p dépend encore de p . On va donc également devoir l'estimer, en écrivant à l'aide de la loi des grands nombres

$$s_n = \frac{1}{n} \sum_{i=1}^n B_i^2 - \left(\frac{1}{n} \sum_{i=1}^n B_i\right)^2 = p_n(1 - p_n) \xrightarrow{n \rightarrow \infty} p(1 - p).$$

On va donc remplacer dans (4) l'écart type σ_p par l'estimateur de l'écart type $\sqrt{s_n}$. On définit donc

$$I_n = \left[p_n - \frac{1.96 \sqrt{s_n}}{\sqrt{n}}, p_n + \frac{1.96 \sqrt{s_n}}{\sqrt{n}} \right].$$

À noter que s_n est construit uniquement à partir des observations (B_k) , et non pas des paramètres théoriques du problème. Par ailleurs, on a comme on le voulait

$$\mathbb{P}(p \in I_n) \geq 0.95.$$

Quelques remarques sur ce raisonnement.

- Dans la dernière identité, on a triché en remplaçant σ_p par son estimateur $\sqrt{s_n}$. Ce faisant, on fait, d'après le théorème central limite appliqué à la suite $(B_k^2)_{k \in \mathbb{N}}$, une erreur de l'ordre de $1/\sqrt{n}$ sur σ_p . Ce n'est pas un problème en réalité, puisque cela entraîne une correction d'ordre $1/\sqrt{n^2} = 1/n$ dans l'intervalle I_n , qui est de taille $O(1/\sqrt{n})$.
- Cet intervalle de confiance permet donc de déterminer la taille d'échantillon nécessaire pour avoir un niveau de précision donné. Imaginons par exemple que le gouvernement souhaite pouvoir estimer, avec un niveau de confiance de 95%, et avec une précision donnée de δ , le taux d'infection dans la population générale. Il suffit pour cela de choisir n assez grand pour que $1.96 \sqrt{s_n}/\sqrt{n} \leq \delta$.
- Tout ce raisonnement repose, comme la loi des grands nombres et le théorème central limite, sur l'indépendance des échantillons B_k . Pour cette raison, il est important, pour que les résultats puissent être pertinents, de tester des personnes qui viennent d'environnements aussi divers et distants que possibles! Imaginez par exemple que tous les tests viennent de patients présentant des symptômes du coronavirus. On surestimerait alors énormément la prévalence p du coronavirus dans la population!

□

On cherche maintenant à estimer si, parmi les patients hospitalisés du covid, la létalité du coronavirus dépend de l'âge du patient.

Exercice 16

On a accès à deux échantillons de patients ayant été hospitalisés, chaque échantillon contenant 100 patients supposés indépendants. Dans le premier groupe, constitué de patients de plus de 50 ans, 50 patients sont décédés. Dans le second groupe, constitué de patients de 15 à 49 ans, 25 personnes sont décédées du covid.

On pose deux hypothèses de travail. Sous

H_0 : le taux de décès du covid est indépendant de l'âge,

le taux théorique de décès suite à l'hospitalisation d'un patient est donné par un paramètre p , commun à tous les patients indépendamment de leur âge.

Sous la seconde hypothèse,

H_1 : le taux de décès du covid dépend de l'âge,

les deux groupes de patients ont deux taux théoriques de décès p_1 et p_2 , respectivement p_1 pour les patients de plus de 50 ans et p_2 pour les 15 – 49 ans. Avec quel degré de confiance peut-on justifier l'hypothèse H_0 ?

On admettra que la somme de deux gaussiennes indépendantes suit également une distribution gaussienne.

SOLUTION : Posons $n = 100$, on va appliquer le même type de raisonnement que dans l'exercice précédent. Supposons que H_0 est vraie. Alors, en définissant p_n la proportion observée de décès dans le premier échantillon et q_n celle dans le second échantillon, on a

$$p_n = 0.5 \quad \text{et} \quad q_n = 0.25.$$

D'autre part, par le théorème central limite,

$$X := \frac{\sqrt{n}(p_n - p)}{\sqrt{p(1-p)}} \simeq \mathcal{N}(0, 1) \quad \text{et} \quad Y := \frac{\sqrt{n}(q_n - p)}{\sqrt{p(1-p)}} \simeq \mathcal{N}(0, 1).$$

On va maintenant supposer que ces deux approximations sont en fait des égalités, et que $X, Y \sim \mathcal{N}(0, 1)$. Par symétrie de la distribution gaussienne, $-Y \sim \mathcal{N}(0, 1)$. Par ailleurs, comme X et Y reposent sur des individus tous indépendants entre eux, on a $X \perp -Y$, et donc $X - Y \sim \mathcal{N}(m, \sigma^2)$, puisque une somme de gaussiennes est encore une gaussienne. Or $m = \mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = 0$. et pour des variables indépendantes

$$\sigma^2 = \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(-Y) = \text{Var}(X) + \text{Var}(Y) = 2,$$

et donc $(X - Y)/\sqrt{2} \sim \mathcal{N}(0, 1)$. On reprend maintenant les résultats numériques de l'expérience, et on obtient

$$\frac{X - Y}{\sqrt{2}} = \frac{\sqrt{n}(p_n - q_n)}{\sqrt{p(1-p)}}.$$

Comme on ne connaît pas p , on va à nouveau estimer l'écart type $\sqrt{p(1-p)}$ grâce à l'écart type empirique dans l'échantillon total,

$$s_n = \sqrt{\frac{p_n + q_n}{2} \left(1 - \frac{p_n + q_n}{2}\right)} = \sqrt{15}/8,$$

et on obtient la valeur numérique $(X - Y)/\sqrt{2} \simeq 40/\sqrt{15} \simeq 10,3$.

Sous l'hypothèse H_0 , notre variable observée $(X - Y)/\sqrt{2} \sim \mathcal{N}(0, 1)$ prend une valeur supérieure à 10. En réalité, cette probabilité est infinitésimale, puisque on a déjà que

$$\mathbb{P}(\mathcal{N}(0, 1) > 5) \simeq 3 \times 10^{-7}.$$

On peut donc rejeter H_0 , et conclure que le taux de décès dû au covid dépend bien de l'âge du patient. \square