Probabilités

UE M407

Résumé de cours

Université Lille 1
Sciences et Technologies

Préambule : Comme son titre l'indique, ce document ne constitue pas un polycopié de cours qui se suffit à lui-même mais un **résumé de cours**. Ainsi, il contient les définitions de tous les objets qui ont été introduits dans le cours mais pas tous les exemples qui ont été développés pendant les séances; il contient un ensemble de théorèmes et propositions qui ont été vus (en cours ou en TD) mais pas leur démonstration. Il constitue pour vous un bilan de ce qu'il faut savoir en probabilités à l'issue de ce cours de tronc commun.
Pour le partiel comme pour l'examen, vous aurez le droit de consulter librement votre exemplaire **non annoté** de ce document à l'exclusion de tout autre document personnel. A l'exception des énoncés ou paragraphes précédés de la mention *(Admis)*, je considère que vous devez savoir démontrer l'ensemble des résultats de ce document, même si la démonstration n'a pas été faite en cours parce qu'elle est considérée comme un prérequis (par exemple de théorie de la mesure).
Cela veut par exemple dire que je n'attends pas de vous que vous connaissiez par cœur quelle est l'espérance d'une loi géométrique de paramètre p mais que vous devez être en mesure de montrer qu'elle est égale à $1/p$. Vous devez savoir montrer l'inégalité de Jensen ou le théorème de la classe monotone.

`Mylene.Maida@math.univ-lille1.fr`
2014-2015

Petite bibliographie commentée
(non exhaustive)

Quelques grands classiques, du L3 à l'agrégation, en français :

- **Barbe-Ledoux, Probabilité L3-M1** *[référence principale du cours][BU, @]*
- Foata-Fuchs, Calcul des probabilités *[BU, @]*
- **Ouvrard, Probabilités** *[tome 1 assez élémentaire, tome 2 convergence, LGN, TCL][BU]*
- Jacod-Protter, L'essentiel en théorie des probabilités *[BU]*
- Toulouse, Thèmes de probabilités et statistiques *[bon complément sur certains sujets précis][BU]*

En anglais, pour un point de vue plus pragmatique :

- **Grimmett-Stirzaker, Probability and Random Processes** *[beaucoup d'exemples, présentation plus concrète][BU]*
- Grimmett-Stirzaker, One thousand exercises in probability *[les exercices du précédent]*
- Pitman, Probability *[plus facile (un peu trop ?), également beaucoup d'exemples]*

En anglais, niveau plus avancé :

- Billingsley, Probability and measure *[BU]*
- Dudley, Real analysis and probability *[BU]*
- Durrett Probability : theory and examples*[BFM]*

On donnera les références entre crochets en tête de chaque chapitre ou paragraphe.

Le sigle *BU* signifie que l'ouvrage est présent dans le catalogue de la bibliothèque universitaire de Lille 1, @ qu'il y est présent sous forme électronique et *BFM* qu'on peut le trouver à la bibliothèque de formation des maîtres.

Partie 1. Concepts et outils probabilistes de base

Plan de la première partie :

I. Généralités sur les espaces de probabilités

[Barbe-Ledoux (chap. 3), Ouvrard 1]

1. Triplet probabiliste

L'un de nos objectifs est de **modéliser des expériences aléatoires.**

Ici, modéliser signifie associer à une expérience aléatoire (qui n'est pas un objet mathématique) un **espace de probabilité (ou triplet probabiliste),** c'est-à-dire un triplet $(\Omega, \mathcal{F}, P)$, où

- Ω est un ensemble (qui code l'ensemble des résultats possibles de l'expérience)
- $\mathcal{F}$ est une **tribu** (ou σ-algèbre)
- et P est une **mesure de probabilité.**

On rappelle qu'une **algèbre** sur Ω est un ensemble de parties de Ω qui contient Ω, stable par complémentaire et réunion finie ; une **tribu** (ou σ-algèbre) sur Ω est un ensemble de parties de Ω qui contient Ω, stable par complémentaire et réunion **dénombrable**.

Une mesure (ou mesure positive) μ sur $(\Omega, \mathcal{F})$ est une application de $\mathcal{F}$ dans $\mathbb{R}^+ \cup \{+\infty\}$ telle que $\mu(\emptyset) = 0$ et μ est σ-additive (c'est-à-dire que si $(A_n)_{n\in\mathbb{N}}$ est ue famille dénombrable d'événements disjoints, $P(\cup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n)$).
Une **(mesure de) probabilité** P est une mesure telle que $P(\Omega) = 1$.

On appelle **événement** tout ensemble mesurable, c'est-à-dire tout élément de $\mathcal{F}$.

Proposition 1 ***(Propriétés utiles d'une probabilité)***

Si P est une probabilité sur $(\Omega, \mathcal{F})$, A, B et $(A_n)_{n\in\mathbb{N}}$ des ensembles mesurables,

1. $A \subset B \Rightarrow P(A) \leq P(B)$
2. $P(A^c) = 1 - P(A)$

3. *(Continuité par limite monotone) Si* $(A_n)_{n\in\mathbb{N}}$ *est une suite croissante ou décroissante d'événements,*

$$P(\lim A_n) = \lim P(A_n),$$

avec la convention que $\lim A_n = \cup_{n\in\mathbb{N}} A_n$ *si la suite est croissante et* $\lim A_n = \cap_{n\in\mathbb{N}} A_n$ *si la suite est décroissante.*

4. $P(\cup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} P(A_n)$
5. $P(A\cup B) = P(A) + P(B) - P(A\cup B)$
6. *(Formule du crible ou de Poincaré)*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\leq i_1<\ldots<i_n\leq n} P\left(\bigcap_{j=1}^{k} A_{i_j}\right).$$

Aparté 1 : dénombrabilité

En probabilité, la notion de dénombrabilité est très importante.

On dit qu'un ensemble E est **dénombrable** s'il existe une injection de E dans l'ensemble $\mathbb{N}$ des entiers naturels [1].
Exemples d'ensembles dénombrables : $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$, l'ensemble des suites finies de 0 et de 1, les ensembles finis.
Une union dénombrable d'ensembles dénombrable est encore dénombrable.
L'ensemble $\mathbb{R}$ des nombres réels n'est pas dénombrable. Plus généralement, un produit dénombrable d'ensembles finis n'est pas dénombrable.

En pratique, les cas les plus courants que vous rencontrerez seront de l'un des types suivants :

- Ω dénombrable, $\mathcal{F}$ est l'ensemble des parties de Ω, notée $\mathcal{P}(\Omega)$, P donnée par son germe

Aparté 2 : Germe de probabilité

Un **germe de probabilité** est une famille dénombrable de réels positifs ou nuls $(p_i)_{i\geq 1}$ avec $\sum_{i\geq 1} p_i = 1$.
Si Ω est dénombrable, $(P(\{x_i\}))_{i\geq 1}$ est un germe de probabilité qui détermine P de manière unique.

Cas particulier : Ω fini, P uniforme. Dans ce cas, la probabilité d'un événement se calcule en divisant le "nombre de cas favorables" par le "nombre de cas possibles", c'est-à-dire en divisant la cardinal de l'événement par le cardinal de Ω). Il faut donc savoir **dénombrer.**

Aparté 3 : Dénombrements usuels

Soit E un ensemble de cardinal n et F un ensemble de cardinal p. Soit $q \in \mathbb{N}^*$. Alors
⋆ le nombre de q-listes d'éléments de E est n^q
⋆ le nombre d'arrangements de q éléments de E est $\frac{n!}{(n-q)!}$
⋆ le nombre de combinaisons de q éléments de E est $\binom{n}{q} = \frac{n!}{q!(n-q)!}$
⋆ le nombre de parties de E est 2^n
⋆ le cardinal de $E\times F$ est np
⋆ le nombre d'applications de E dans F est p^n
⋆ le nombre d'injections de E dans F est $\frac{p!}{(p-n)!}$
⋆ le nombre de permutations de E est $n!$

- $\Omega = \mathbb{R}^n$, $\mathcal{F}$ tribu borélienne, P absolument continue par rapport à Lebesgue (cas particulier : probabilité uniforme sur un intervalle)

Rappel : on dit qu'une mesure de probabilité Q est absolument continue par rapport à une autre P ssi $\forall A \in \mathcal{F}, P(A) = 0 \Rightarrow Q(A) = 0$.

Théorème 2 *(Radon-Nikodym (Admis))*

> *Si P et Q sont deux mesures de probabilité telles que Q est absolument continue par rapport à P, il existe une fonction $h \in L^1(P)$ telle que $\forall A \in \mathcal{F}, Q(A) = \int_A f dP$. On appelle h la densité de Q par rapport à P.*

> **Aparté 4 : Rappel sur la mesure de Lebesgue (Admis)**
>
> Sur $\mathbb{R}^d$ muni de sa tribu borélienne, on peut définir une mesure λ_d, appelée **mesure de Lebesgue** sur $\mathbb{R}^d$, ayant les propriété suivantes :
> - sur les pavés, elle coïncide avec la mesure du volume
> - elle est invariante par toute isométrie euclidienne
> - elle ne charge pas les points ni aucun sous-espace affine strict de $\mathbb{R}^d$
> - elle est homogène de degré d (c'est-à-dire que pour tout borélien B de $\mathbb{R}^d$ et $a \in \mathbb{R}$, on a $\lambda_d(aB) = |a|^d \lambda_d(B)$.

2. Variable aléatoire réelle

2.1. Définition

On donne d'abord une définition générale d'une **variable aléatoire.**
Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilités et $(E, \mathcal{G})$ un ensemble muni d'une tribu donnés. Une **variable aléatoire (v.a.)** est une fonction mesurable de $(\Omega, \mathcal{F})$ dans $(E, \mathcal{G})$.

Dans le cas particulier où $(E, \mathcal{G}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on parle de **v.a. réelle**.
Si $(E, \mathcal{G}) = (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$, on parle de **vecteur aléatoire**.

Si X est une variable aléatoire de $(\Omega, \mathcal{F})$ dans $(E, \mathcal{G})$, la **loi** de X, notée P_X est la mesure-image de P par X, c'est-à-dire la mesure de probabilité sur $(E, \mathcal{G})$ définie par :

$$\begin{aligned} \forall B \in \mathcal{G},\ P_X(B) &\stackrel{\text{def}}{=} P(X \in B) \\ &= P(\{\omega \in \Omega \mid X(\omega) \in B\}) \end{aligned}$$

Notation : si $P_X = Q$, on notera souvent $X \sim Q$.

En pratique, les lois des v.a. réelles que vous rencontrerez, seront pour la plupart **discrètes** ou **à densité**.

2.2. Lois discrètes

La **masse de Dirac en** x, notée δ_x, est la mesure de probabilité donnée par :

$$\forall A \subset \Omega,\ \delta_x(A) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$$

Une probabilité Q est dite **discrète** si elle s'écrit comme combinaison linéaire dénombrable de masses de Dirac :

$$Q = \sum_{i \in I} p_i \delta_{x_i}, \textit{où } I \textit{ est dénombrable, } p_i \geq 0 \textit{ et } \textstyle\sum_{i \in I} p_i = 1.$$

Une v.a. est dite discrète si sa loi est une mesure de probabilité discrète. Autrement dit, ssi X prend (p.s.) ses valeurs dans un ensemble dénombrable.

2.3. Lois à densité (par rapport à Lebesque)

Une probabilité Q définie sur $\mathbb{R}$ munie de sa tribu borélienne, est dite **à densité** si et seulement si elle est absolument continue par rapport à la mesure de Lebesgue.

De manière équivalente (grâce au théorème 2 (Radon-Nikodym)), une probabilité Q définie sur $\mathbb{R}$ munie de sa tribu borélienne est dite **à densité** s'il existe une fonction mesurable $f : \mathbb{R} \to \mathbb{R}$ positive telle que, pour tout $B \in \mathcal{B}(\mathbb{R})$, on a $Q(B) \stackrel{\text{def}}{=} \int_B f(x)\,dx = \int_\Omega \mathbb{1}_B(x) f(x)\,dx$.

f est alors la **densité** de Q. On note $dQ = f(x)dx$.

NB : on a nécessairement $\int_{\mathbb{R}} f(x)\,dx = 1$.

Si X est une v.a. à densité, alors $P(X = a) = 0$ pour tout $a \in \mathbb{R}$ (X n'a pas d'atome).

2.4. Décomposition de Lebesgue d'une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

Théorème 3 *(Admis).*

> *Toute mesure de probabilité P sur $\mathbb{R}$ se décompose de manière unique comme $P = \alpha P_1 + \beta P_2 + \gamma P_3$, avec P_1 probabilité discrète, P_2 à densité par rapport à la mesure de Lebesque λ et P_3 singulière sans atome.*

2.5. Moments d'une v.a.

Si X est une v.a. réelle P-intégrable ou positive, on définit son **espérance** par :

$$E(X) \stackrel{\text{def}}{=} \int_\Omega X(\omega)\,dP(\omega).$$

L'espérance est un cas particulier d'intégrale par rapport à une mesure positive. Par conséquent :

Proposition 4

> - *(linéarité) $E(aX+bY) = a\,E(X) + b\,E(Y)$*
> - *(convergence monotone) Si $X_n \geq 0$ et $X_n \uparrow X$, alors $E(X_n) \uparrow E(X)$*
> - *(convergence dominée) Si $|X_n| \leq Z$ avec $E(Z) < \infty$ et $X_n \to X$, alors $E(X_n) \to E(X)$.*
> - *(lemme de Fatou) Si $X_n \geq 0$, $E(\liminf X_n) \leq \liminf E(X_n)$.*

De même, si X est dans $L^2(\Omega, \mathcal{F}, P)$, sa **variance** est donnée par :

$$\text{Var}(X) \stackrel{\text{def}}{=} E\left((X - E(X))^2\right) = E(X^2) - E(X)^2.$$

$\sigma(X) = \sqrt{\text{Var}(X)}$ est appelé l'**écart-type** de X.

Propriétés :

> - Si $\text{Var}X = 0$, X est p.s. constante
> - $\text{Var}(aX) = a^2 \text{Var}X$
> - $\text{Var}X = \inf_{a \in \mathbb{R}} E((X-a)^2)$.

De même, si X^n est P-intégrable ou positive, $E(X^n)$ est appelé moment d'ordre n.

On utilise très souvent la **formule de transfert** pour le calcul des moments.

Proposition 5

> *Soit $X : \Omega \to E$ une v.a. et $\varphi : E \to \mathbb{R}^n$ une fonction mesurable bornée. Alors*
>
> $$E(\varphi(X)) \overset{\text{def}}{=} \int_\Omega \varphi(X(\omega))\, dP(\omega) = \int_{\mathbb{R}^n} \varphi(x)\, dP_X(x).$$

Si $P_X = \sum_{i\geq 1} p_i \delta_{x_i}$, $E(\varphi(X)) = \sum_{i\geq 1} p_i \varphi(x_i)$.
Si X est de densité f, $E(\varphi(X)) = \int_{\mathbb{R}} \varphi(x) f(x) dx$.

On peut ensuite étendre la formule à toute fonction φ mesurable P_X-intégrable.
En particulier,

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

2.6. Quelques inégalités très utiles

Proposition 6 ***(Inégalités classiques)***

> - *(Inégalité de Jensen) Si φ est une fonction convexe sur $\mathbb{R}$ et X et $\varphi(X)$ sont dans L^1, alors $\varphi(E(X)) \leq E(\varphi(X))$.*
> - *(Inégalité de Hölder) Si $X \in L^p, Y \in L^q, p, q \geq 1$ et $\frac{1}{p} + \frac{1}{q} = 1$ alors $XY \in L^1$ et*
>
> $$E(|XY|) \leq (E(|X|^p)^{1/p}(E(|X|^q)^{1/q}$$
>
>   *($p = q = 2$: Cauchy-Schwarz)*
> - *$p \mapsto (E(|X|^p)^{1/p}$ croissante*
> - *Pour $p \geq 1$, $(E(|\cdot|^p)^{1/p}$ est une norme [l'inégalité triangulaire donne Minkowski].*

On énonce maintenant quelques inégalités plus spécifiquement probabilistes :

Proposition 7

> - *(Inégalité de Markov) Si X intégrable ou positive, pour tout $t > 0$,*
>
> $$P(X > t) \leq \frac{E(X^+)}{t} \leq \frac{E(|X|)}{t}.$$
>
> - *(Inégalité de Bienaymé-Tchebichev) Si $X \in L^2$, pour tout $t > 0$,*
>
> $$P(|X - E(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$
>
> - *(Inégalité de Tchebichev exponentielle) Si $\exists \lambda > 0$ tel que $E(e^{\lambda X}) < \infty$,*
>
> $$P(X \geq t) \leq e^{-I(t)},$$
>
>   *où $I(t) = \sup_\lambda(\lambda t - \ln E(e^{\lambda X}))$.*

II. Des outils pour déterminer la loi d'une v.a. réelle

L'idée générale est que si on connaît $E(\varphi(X))$ pour un ensemble suffisant $\mathcal{T}$ de fonctions-tests, cela va caractériser la loi de X.

1. Fonction de répartition

[Foata-Fuchs p 48 sq]

Dans ce cas, $\mathcal{T}$ est l'ensemble des fonctions de la forme $\mathbf{1}_{]-\infty,t]}$ pour $t \in \mathbb{R}$.

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

$$\forall t \in \mathbb{R},\ F_X(t) \stackrel{\text{def}}{=} P(X \leq t).$$

Propriétés immédiates.

- $P(a < X \leq b) = F_X(b) - F_X(a)$ si $a < b$.
- $P(X > a) = 1 - F_X(a)$.

Théorème 8

Soient X et Y des v.a. réelles de lois respectives P_X et P_Y. Alors $P_X = P_Y$ si et seulement si $F_X = F_Y$.

Proposition 9

Soit X une v.a. réelle.

- *F_X est une fonction croissante à valeurs dans $[0,1]$.*
- $\lim_{t\to-\infty} F_X(t) = 0$ *et* $\lim_{t\to+\infty} F_X(t) = 1$.
- *$\forall a \in \mathbb{R}$, F_X est continue à droite de a et $F_X(a-) := \lim_{t\uparrow a} F_X(t)$ existe et vaut $P(X < a)$. Elle est donc càdlàg.*

On a donc aussi $P(X = a) = F_X(a) - F_X(a-)$.
Si X est une v.a. continue de densité $f(x)$, alors $F_X(t) = \displaystyle\int_{-\infty}^{t} f(x)\,dx$. La fonction F_X est continue sur $\mathbb{R}$, et si f est continue au point t alors F_X est dérivable au point t avec $F_X'(t) = f(t)$.

Exemple.
Si $P_X = \mathcal{E}(\lambda)$, alors $F_X(t) = 0$ si $t \leq 0$ et $F_X(t) = \displaystyle\int_0^t \lambda e^{-\lambda x}\,dx = 1 - e^{-\lambda t}$ si $t \geq 0$.

Aparté 7 : Lois sans mémoire

Les seules variables aléatoires à valeurs dans $\mathbb{R}_+$ qui ont la propriété d'être **sans mémoire**, c'est-à-dire qui vérifient

$$\forall s,t > 0, P(X > t+s) = P(X > s)P(X > t),$$

sont les v.a. de lois exponentielles (de paramètre $\lambda > 0$).

Les seules variables aléatoires à valeurs dans $\mathbb{N}^*$ qui ont la propriété d'être **sans mémoire**, c'est-à-dire qui vérifient

$$\forall n,m \in \mathbb{N}^*, P(X > n+m) = P(X > m)P(X > n),$$

sont les v.a. de lois géométriques (de paramètre $p \in]0,1[$).

Si X est une v.a. discrète, la fonction de répartition est en escaliers, les marches sont aux valeurs prises par X, la hauteur de la marche en x_i est $P(x_i)$.

La méthode de la transformée inverse a des applications importantes pour la simulation de variables aléatoires.

Proposition 10

> *Soit F une fonction de répartition. On définit sa fonction quantile par*
>
> $$F^-(u) = \inf\{x; F(x) > u\}, \forall u \in]0,1[.$$
>
> *Si U est de loi uniforme sur $[0,1]$, $F^-(U)$ a pour fonction de répartition F.*

2. Théorème de transfert
Dans ce cas, $\mathcal{T}$ est l'ensemble des fonctions continues bornées.

Théorème 11

> *Soient X et Y des v.a. réelles de lois respectives P_X et P_Y.*
> *On a $P_X = P_Y$ ssi pour toute fonction φ continue bornée, $E(\varphi(X)) = E(\varphi(Y))$.*

3. Fonction caractéristique
Dans ce cas, $\mathcal{T}$ est l'ensemble des fonctions de la forme $x \mapsto e^{itx}$ pour $t \in \mathbb{R}$.

Soit X une v.a. réelle. La **fonction caractéristique** de X est la fonction $\varphi_X \colon \mathbb{R} \to \mathbb{C}$ telle que $\forall t \in \mathbb{R}$,

$$\varphi_X(t) \stackrel{\text{def}}{=} E(e^{itX}) = \int_\Omega e^{itX(\omega)}\, dP(\omega) = \int_{\mathbb{R}} e^{itx}\, dP_X(x).$$

Exemple. La fonction caractéristique de la loi $\mathcal{N}(0,1)$ est donnée par $\varphi(t) = e^{-t^2/2}, \forall t \in \mathbb{R}$.

Théorème 12 *(Théorème de Lévy)*

> *Soit X et Y des v.a. réelles de lois respectives P_X et P_Y. Alors $P_X = P_Y$ si et seulement si $\varphi_X = \varphi_Y$.*

On peut retrouver les moments d'une v.a. à partir des dérivées en zéro de sa fonction caractéristique.

Proposition 13

> *Soit X une v.a. réelle. Si $X \in L^n$ alors φ_X est n fois dérivable et*
>
> $$\forall 1 \leq k \leq n,\ \varphi_X^{(k)}(0) = i^k E(X^k).$$
>
> *Réciproquement, si φ_X est k fois dérivable en 0 et $2n \leq k$ alors $X \in L^{2n}$.*

4. Fonction génératrice
[Ouvrard 1 p138, Foata-Fuchs chap 9]
Dans ce cas, $\mathcal{T}$ est l'ensemble des fonctions de la forme $x \mapsto t^x$ pour $t \in [-1,1]$.

NB : un peu moins usitée que les précédentes et restreinte aux v.a. à valeurs dans $\mathbb{N}$.

Soit X une v.a. à valeurs dans $\mathbb{N}$ et $p_n = P(X = n)$. La **fonction génératrice** de X est définie par

$$G_X(x) \stackrel{\text{def}}{=} \sum_{n=0}^{+\infty} p_n x^n = E(t^X).$$

Proposition 14

> *Soit X une v.a. à valeurs dans $\mathbb{N}$. Sa fonction génératrice G_X est bien définie et continue sur $[-1,1]$, $\mathcal{C}^\infty$ sur $]-1,1[$.*
> *De plus, pour tout $r \in \mathbb{N}^*$, $X \in L^r$ si et seulement si $G_X^{(r)}(1^-) := \lim_{s\uparrow 1} G_X^{(r)}(s)$ existe et dans ce cas, on a $E(X(X-1)\dots(X-r+1)) = G_X^{(r)}(1^-)$.*

La fonction génératrice caractérise la loi de la variable aléatoire.

Théorème 15

> *Soit X et Y deux v.a. à valeurs dans $\mathbb{N}$. Alors*
> *$P_X = P_Y \iff G_X = G_Y$ sur un voisinage de 0.*

Exemple.
Soit X une v.a. de loi de Poisson $\mathcal{P}(\lambda)$.
On a $G_X(x) = e^{\lambda(x-1)}$. $G'_X(x) = \lambda e^{\lambda(x-1)}$ et $G'_X(1) = \lambda$. Donc $E(X) = \lambda$.

5. Résumé des méthodes

Théorème 16

> *Soient X et Y deux v.a. réelles de loi respectives P_X et P_Y. Alors si **l'une des conditions** suivantes est vérifiée*
>
> - *(Théorème de transfert) pour toute fonction φ continue bornée,*
>
> $$\int_{\mathbb{R}} \varphi(t)dP_X(t) = \int_{\mathbb{R}} \varphi(t)dP_Y(t)$$
>
> - $F_X = F_Y$
> - *(Théorème de Lévy)* $\varphi_X = \varphi_Y$
> - *X et Y sont à valeurs dans $\mathbb{N}$ et $G_X = G_Y$ sur un voisinage de zéro,*
>
> *on a $P_X = P_Y$, autrement dit X et Y ont même loi.*

6. Compléments sur les vecteurs aléatoires

6.1. Définitions et vocabulaire spécifique aux vecteurs aléatoires

Un **vecteur aléatoire** est une v.a. à valeurs dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. On le note $X = (X_1, \dots, X_n)$. Les projections sur les coordonnées étant mesurables, les $(X_i)_{1\le i\le n}$ sont des v.a. réelles.

La loi P_X de X est une mesure sur $\mathbb{R}^n$, on l'appelle parfois **loi jointe** de $(X_1, \dots, X_n)$. La loi de la v.a. X_i est appelée $i$ème **marginale** de la loi de X.

Si la loi P_X s'écrit comme une somme dénombrable de masses de Dirac, on dit encore que P_X est **discrète.**
S'il existe une fonction $f : \mathbb{R}^n \to \mathbb{R}$ mesurable positive telle que pour tout B borélien de $\mathbb{R}^n$, on a

$$P_X(B) = \int_{\mathbb{R}^n} f(x_1, \dots, x_n)dx_1 \dots dx_n,$$

on dit que X est un vecteur aléatoire de densité f.
Si P_X est de densité f, sa $i$ème marginale est aussi à densité, de densité

$$f_i(x) = \iint_{\mathbb{R}^{n-1}} f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)dx_1 \dots dx_{i-1}dx_{i+1} \dots dx_n.$$

La réciproque est fausse.

Si chacune des coordonnées X_i est dans $L^2(\Omega, \mathcal{F}, P)$, on peut définir la matrice D, de taille $n \times n$, telle que $D_{ii} = Var(X_i)$ pour tout $1 \leq i \leq n$ et $D_{ij} = cov(X_i, X_j)$, pour tout $i \neq j$. Cette matrice est appelée **matrice de variance-covariance** ou simplement **matrice de covariance** de X. Elle est positive[2] et symétrique.

6.2. Extension des outils déjà connus pour déterminer la loi d'un vecteur aléatoire

On peut définir la **fonction de répartition** de X : pour tout $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$,

$$F_X(t_1, \ldots, t_n) := P(X_1 \leq t_1, \ldots, X_n \leq t_n).$$

Elle caractérise la loi P_X mais est beaucoup moins usitée que dans le cas réel.

On utilise plutôt l'extension du théorème de transfert suivante : soient X et Y deux vecteurs aléatoires de $\mathbb{R}^n$. Si pour toute fonction continue bornée φ de $\mathbb{R}^n$ dans $\mathbb{R}$, $E(\varphi(X)) = E(\varphi(Y))$ alors $P_X = P_Y$.

On peut aussi définir la fonction caractéristique, qui est une fonction à n variables donnée par

$$\forall t = (t_1, \ldots, t_n) \in \mathbb{R}^n, \;\; \varphi_X(t) = E(e^{i\langle t, X\rangle}) = E(e^{i(t_1X_1 + \ldots + t_nX_n)}),$$

qui caractérise encore la loi de X.

6.3. Changement de variables

Théorème 17 *(Admis)*

Soient A et B deux ouverts de $\mathbb{R}^n$ et $h : A \to B$ un $\mathcal{C}^1$-difféomorphisme de A dans B. Alors pour toute f mesurable positive ou intégrable sur B, on a

$$\int_A f(h(x_1, \ldots, x_n))|Jac\ h(x_1, \ldots, x_n)|dx_1 \ldots dx_n = \int_B f(y_1, \ldots, _n)dy_1, \ldots, dy_n,$$

et pour toute g mesurable positive ou telle que $g \circ h$ intégrable sur A, on a

$$\int_A g(h(x_1, \ldots, x_n))dx_1 \ldots dx_n = \int_B g(y_1, \ldots, y_n)|Jac\ h^{-1}(y_1, \ldots, y_n)|dy_1, \ldots, dy_n,$$

avec

$$Jac\ h(x_1, \ldots, x_n) := \det \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1} & \cdots & \frac{\partial h_n}{\partial x_n} \end{pmatrix}.$$

Remarque. On peut vérifier que

$$|\text{Jac}\ h^{-1}(y_1, \ldots, y_n)| = \frac{1}{|\text{Jac}\ h(h^{-1}(y_1, \ldots, y_n))|}.$$

---

2. Une matrice M est dite positive (on dit aussi semi-définie positive) ssi pour tout vecteur x, on a $x^tMx \geq 0$.

Appendice 1 : Algèbre, tribu, classe monotone, un bref rappel

[appendice de Ouvrard 2, Barbe-Ledoux, Foata-Fuchs chap 1 et 2]

Ω donné

On rappelle que : un ensemble de parties de Ω

- $\mathcal{C}$ est une algèbre ssi $\Omega \in \mathcal{C}$ et $\mathcal{C}$ est stable par complémentaire et réunion **finie.**
- $\mathcal{F}$ est une tribu (ou σ-algèbre) ssi $\Omega \in \mathcal{F}$ et $\mathcal{F}$ est stable par complémentaire et par union **dénombrable.**
- $\mathcal{M}$ est une classe monotone ssi $\Omega \in \mathcal{M}$ et $\mathcal{M}$ est stable par différence et réunion **croissante.**

On sait que

- Une intersection quelconque d'algèbres est une algèbre. Une intersection quelconque de tribus est une tribu. Une intersection quelconque de classes monotones est une classe monotone.
  On peut donc définir la notion d'algèbre (respectivement tribu, classe monotone) **engendrée**.
- Une union **croissante** de tribus est une algèbre.
- Une tribu est une classe monotone.
- Une classe monotone stable par intersection finie est une tribu.

Théorème 18 *(dit de la classe monotone)*

Si $\mathcal{E}$ est une partie de $\mathcal{P}(\Omega)$ **stable par intersection finie**, *alors la classe monotone engendrée par $\mathcal{E}$, notée $\mathcal{M}(\mathcal{E})$ coïncide avec $\sigma(\mathcal{E})$ la tribu engendrée par $\mathcal{E}$.*

Remarque. *On utilise souvent ce résultat sous la forme suivante : Si $\mathcal{M}$ est une classe monotone contenant $\mathcal{E}$, elle contient $\sigma(\mathcal{E})$.*

Théorème 19 *(dit de prolongement de Carathéodory)(Admis)*

Si μ est une fonction additive, positive, définie sur une algèbre $\mathcal{C}$ de parties de Ω avec $\mu(\Omega) < \infty$, elle se prolonge de façon unique en une mesure sur $(\Omega, \sigma(\mathcal{C}))$.

Appendice 2 : Vade-mecum sur les lois usuelles pour des variables aléatoires réelles

Lois discrètes usuelles

• Loi uniforme sur un ensemble fini $\{x_1, \ldots, x_N\}$, avec $N \in \mathbb{N}^*$.
X prend ses valeurs (p.s.) dans $\{x_1, \ldots, x_N\}$, avec $\mathbb{P}(X = x_i) = 1/N$ pour tout $i \in \{1, \ldots, N\}$. On a

$$\mathbb{P}_X = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}, \text{ ce que l'on peut noter } X \sim \mathcal{U}(\{x_1, \ldots, x_N\}).$$

Exemple : résultat d'un lancer de dé non pipé.

• Loi de Bernoulli de paramètre p, avec $0 < p < 1$.
X prend ses valeurs (p.s.) dans $\{0, 1\}$, avec $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = 0) = 1 - p$. On a

$$\mathbb{P}_X = (1 - p)\delta_0 + p\delta_1, \text{ ce que l'on peut noter } X \sim \mathcal{B}(1, p).$$

$$\mathrm{E}(X) = p, \quad \mathrm{Var}(X) = p(1 - p), \quad \forall t \in \mathbb{R}, G_X(t) = pt + 1 - p.$$

Exemple : nombre de "face" obtenu lors du lancer d'une pièce.

• Loi binomiale de paramètres n et p, avec $n \in \mathbb{N}^*$ et $0 < p < 1$.
X prend ses valeurs dans $\{0, 1, \ldots, n\}$ et $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ pour $0 \leq k \leq n$. On a

$$\mathbb{P}_X = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} \delta_k, \text{ ce que l'on peut noter } X \sim \mathcal{B}(n, p).$$

$$\mathrm{E}(X) = np, \quad \mathrm{Var}(X) = np(1 - p), \quad \forall t \in \mathbb{R}, G_X(t) = (pt + 1 - p)^n.$$

Exemple : nombre de 6 obtenus au bout de n lancers d'un dé.

• Loi hypergéométrique de paramètres n, N, p, avec $n \in \mathbb{N}^*, N \in \mathbb{N}^*, n \leq N, 0 < p < 1, Np \in \mathbb{N}^*$.
X prend ses valeurs dans $\{0, \ldots, n\}$ et $\mathbb{P}(X = k) = \dfrac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}$ pour $0 \leq k \leq n$. On a

$$\mathbb{P}_X = \sum_{k=0}^{n} \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}} \delta_k, \text{ ce que l'on peut noter } X \sim \mathcal{H}(n, N, p).$$

$$\mathrm{E}(X) = np, \quad \mathrm{Var}(X) = np(1 - p)\frac{N - n}{N - 1}.$$

Exemple : nombre de boules rouges obtenues au bout de 5 tirages sans remise dans une urne contenant 10 boules rouges et 20 boules blanches.

• **Loi géométrique de paramètre p, avec $0 < p < 1$.**
X prend ses valeurs dans $\mathbb{N}^*$ et $\mathbb{P}(X = k) = p(1-p)^{k-1}$ pour $k \geq 1$. On a

$$\mathbb{P}_X = \sum_{k=1}^{\infty} p(1-p)^{k-1}\delta_k, \text{ ce que l'on peut noter } X \sim \mathcal{G}(p).$$

$$\mathrm{E}(X) = 1/p, \quad \mathrm{Var}(X) = (1-p)/p^2, \quad \forall t \in \mathbb{R}, \text{ tel que } |t| < \frac{1}{1-p}, G_X(t) = \frac{pt}{1-(1-p)t}.$$

Exemple : nombre de lancers nécessaires pour obtenir le premier 6.

• **Loi de Poisson de paramètre λ, avec $\lambda > 0$**
X prend ses valeurs dans $\mathbb{N}$ et $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ pour tout entier $k \in \mathbb{N}$. On a

$$\mathbb{P}_X = \sum_{k \in \mathbb{N}} e^{-\lambda}\frac{\lambda^k}{k!}\delta_k, \text{ ce que l'on peut noter } X \sim \mathcal{P}(\lambda).$$

$$\mathrm{E}(X) = \mathrm{Var}(X) = \lambda, \quad \forall t \in \mathbb{R}, G_X(t) = e^{\lambda(t-1)}.$$

Lois à densité usuelles

• **Loi uniforme sur un intervalle $[a, b]$, avec $a < b$.**
La densité de $\mathbb{P}_X$ est $\frac{1}{b-a}\mathbb{1}_{[a,b]}$, ce que l'on peut noter $X \sim \mathcal{U}([a, b])$.

$$\mathrm{E}(X) = (a+b)/2, \quad \mathrm{Var}(X) = (b-a)^2/12, \quad \forall t \in \mathbb{R}, \varphi_X(t) = e^{it\frac{a+b}{2}}\frac{\sin(\frac{(b-a)t}{2})}{\frac{(b-a)t}{2}}.$$

• **Loi exponentielle de paramètre λ, avec $\lambda > 0$.**
La densité de $\mathbb{P}_X$ est $x \mapsto \mathbb{1}_{\mathbb{R}_+}(x)\lambda e^{-\lambda x}$, ce que l'on peut noter $X \sim \mathcal{E}(\lambda)$.

$$\mathrm{E}(X) = 1/\lambda, \quad \mathrm{Var}(X) = 1/\lambda^2, \quad \forall t \in \mathbb{R}, \varphi_X(t) = \frac{\lambda}{it - \lambda}.$$

• **Loi de Gauss, loi normale de pramètres** m **et** σ, **avec** $m \in \mathbb{R}$ **et** $\sigma^2 > 0$.
La densité de la loi $\mathcal{N}(m, \sigma^2)$ est $x \mapsto \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{(x-m)^2}{2\sigma^2}\right)$, ce que l'on peut noter $X \sim \mathcal{N}(m, \sigma^2)$.

$$\mathrm{E}(X) = m, \quad \mathrm{Var}(X) = \sigma^2, \quad \forall t \in \mathbb{R}, \varphi_X(t) = e^{itm}e^{-\frac{1}{2}\sigma^2 t^2}.$$

• **Loi de Cauchy standard.**
La densité de $\mathbb{P}_X$ est $x \mapsto \dfrac{1}{\pi}\dfrac{1}{1+x^2}$, ce que l'on peut noter $X \sim \mathcal{C}(0, 1)$.
Cette loi n'a ni espérance, ni variance.

Partie 2. Indépendance et conditionnement

Plan de la deuxième partie :

III. Indépendance et applications

[Barbe-Ledoux]

1. Indépendance d'événements
Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilités.
Deux événements A, B sont dits **indépendants** si $P(A \cap B) = P(A)P(B)$. On note $A \perp B$.

Remarque. Si A et B sont indépendants, A et B^c le sont aussi.

La famille d'événements $(A_i)_{i \in I}$ est dite **indépendante** (on dit parfois **mutuellement indépendante**) si pour tout sous-ensemble fini $\{i_1, \ldots, i_n\}$ de I,

$$P\left(\bigcap_{k=1}^{n} A_{i_k}\right) = \prod_{k=1}^{n} P(A_{i_k}).$$

Contre-exemple.
Soit $\Omega = \{1, \ldots, 4\}$ et P la probabilité uniforme.
Soit $A = \{1,2\}$, $B = \{1,3\}$, $C = \{1,4\}$.
On a $P(A) = P(B) = P(C) = 1/2$. $P(A \cap B) = P(\{1\}) = 1/4 = P(A)P(B)$ donc $A \perp B$. De même, $A \perp C$ et $B \perp C$, autrement dit A, B, C sont 2 à 2 indépendants. $P(A \cap B \cap C) = P(\{1\}) = 1/4 \neq P(A)P(B)P(C)$ donc A, B, C ne sont pas indépendants.

2. Indépendance de sous-tribus
Une famille de sous-tribus (ou d'algèbres) $(\mathcal{F}_i)_{i \in I}$ (avec $\mathcal{F}_i \subset \mathcal{F}$) est dite **indépendante** si toute famille d'événements $(A_i \in \mathcal{F}_i)_{i \in I}$ est indépendante.

Remarque. Si la famille d'événements $(A_i)_{i \in I}$ est dite indépendante, la famille de sous-tribus $(\sigma(\{A_i\}))_{i \in I}$ l'est aussi.

Proposition 20

Si $\mathcal{C}_1$ et $\mathcal{C}_2$ sont deux sous-algèbres indépendantes, alors $\sigma(\mathcal{C}_1)$ et $\sigma(\mathcal{C}_2)$ sont deux sous-tribus indépendantes.

Proposition 21 *(Regroupement par paquets)*

Soit $(\mathcal{F}_i)_{i \in I}$ une famille indépendante de sous-tribus d'une tribu $\mathcal{F}$. Soit $(J_\ell)_{\ell \in L}$ une partition arbitraire de I. La famille de tribus $(\sigma(\mathcal{F}_i, i \in J_\ell))_{\ell \in L}$ est une famille indépendante.

3. Notion de mesure-produit
3.1. Définition

Théorème 22 *(Admis)*

Soit μ une mesure de probabilité sur $(\Omega, \mathcal{F})$ et ν une mesure de probabilité sur $(\Omega', \mathcal{F}')$. $\mathcal{F} \otimes \mathcal{F}'$ est la tribu de $\Omega \times \Omega'$ engendrée par les pavés, c'est-à-dire les ensembles $A \times B$ avec $A \in \mathcal{F}, B \in \mathcal{F}'$.

Pour tous $A \in \mathcal{F}, B \in \mathcal{F}'$, on définit

$$\pi(A \times B) = \mu(A)\nu(B).$$

Alors π s'étend de façon unique en une mesure de probabilité sur $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$. On note cette mesure $\mu \otimes \nu$ et on l'appelle mesure-produit de μ par ν.

3.2. Fubini

Soient $(\Omega, \mathcal{F}, \mu)$ et $(\Omega', \mathcal{F}', \nu)$ deux espaces de probabilités. On munit $\Omega \times \Omega'$ de la tribu produit $\mathcal{F} \otimes \mathcal{F}'$.

Théorème 23 *(Fubini-Tonelli, Fubini)*

Soit f une fonction à valeurs réelles, définie sur $\Omega \times \Omega'$, $\mathcal{F} \otimes \mathcal{F}'$-mesurable et $\mu \otimes \nu$-intégrable ou positive. Alors on a

$$\int f d(\mu \otimes \nu) = \int_\Omega \left(\int_{\Omega'} f(\omega, \omega') d\nu(\omega') \right) d\mu(\omega) = \int_{\Omega'} \left(\int_\Omega f(\omega, \omega') d\mu(\omega) \right) d\nu(\omega').$$

4. Indépendance de variables aléatoires
4.1. Définition et conséquences

On considère des v.a. $X_i : \Omega \to (E_i, \mathcal{G}_i)$.

La famille de v.a. $(X_i)_{i \in I}$ est **indépendante** si et seulement si la famille de tribus $(X_i^{-1}(\mathcal{G}_i))_{i \in I}$ est indépendante ; autrement dit, si pour tout sous-ensemble fini $\{i_1, \ldots, i_n\}$ de I et $A_j \in \mathcal{G}_{i_j}$,

$$P\left(X_{i_k} \in A_k, 1 \leq k \leq n\right) = \prod_{k=1}^{n} P(X_{i_k} \in A_k).$$

Remarque. Les X_i doivent être définis sur le même espace Ω sinon l'expression de gauche n'a pas de sens.

Si $X_1, X_2, \ldots, X_n$ des v.a. indépendantes et $\varphi_1, \varphi_2, \ldots, \varphi_n$ des fonctions mesurables, alors les v.a. $\varphi_1(X_1), \ldots, \varphi_n(X_n)$ sont indépendantes. *(φ_i doit être définie sur E_i (à valeurs dans un certain F_i))* et les v.a. $Y = (X_1, \ldots, X_k)$ et $Y' = (X_{k+1}, \ldots, X_n)$ sont indépendantes (théorème des coalitions).

4.2. Indépendance et loi produit

Proposition 24

Soit $(X_i)_{1 \leq i \leq n}$ des v.a. définies sur Ω. Les v.a. $X_1, \ldots, X_n$ sont indépendantes si et seulement si la loi de la v.a. $Y = (X_1, \ldots, X_n)$ est égale à $P_{X_1} \otimes P_{X_2} \otimes \cdots \otimes P_{X_n}$.

Autrement dit, on peut appliquer le critère suivant :

Proposition 25

Une famille de v.a. réelles $(X_i)_{i \in I}$ sur Ω est indépendante si et seulement si pour toute famille finie $J \subset I$ et $(\varphi_i)_{i \in J}$ des fonctions mesurables telles que $(\varphi_i(X_i))_{i \in J}$ intégrables, on a

$$E\left(\prod_{i \in J} \varphi_i(X_i)\right) = \prod_{i \in J} E\left(\varphi_i(X_i)\right).$$

4.3. Indépendance et corrélation

Deux v.a. réelles X, Y sont dites **non corrélées** si $\mathrm{Cov}(X,Y) := E(XY) - E(X)E(Y) = 0$.

Conséquence : Si X et Y sont deux v.a. réelles non corrélées, alors $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Proposition 26

Soit X, Y des v.a. réelles intégrables indépendantes. Alors XY est intégrable et $E(XY) = E(X)E(Y)$. Autrement dit, deux variables indépendantes sont non-corrélées.

La réciproque est fausse. ***Exemple.*** $X \sim \mathcal{N}(0,1)$ *et* $Y = X^2$.

4.4. Somme de v.a. indépendantes et convolution

Soit μ et ν deux mesures de probabilité sur $\mathbb{R}$. Le **produit de convolution** $\mu * \nu$ est la mesure de probabilité sur $\mathbb{R}$ définie par : $\forall \phi\colon \mathbb{R} \to \mathbb{R}$ fonction borélienne bornée,

$$\int_{\mathbb{R}} \phi \, d(\mu * \nu) \overset{\text{def}}{=} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \phi(x+y) \, d\nu(y) \right) d\mu(x).$$

On vérifie facilement que l'opération $*$ est commutative, associative, distributive par rapport à l'addition.

Proposition 27

Soit X et Y deux v.a. réelles indépendantes. La loi de $X+Y$ est égale à $P_X * P_Y$.

Lois continues. Soit X de densité f et Y de densité g. Si X et Y sont indépendantes, alors $X+Y$ est continue de densité $f * g$, où $f * g(x) = \int_{\mathbb{R}} f(x-y) g(y) \, dy$.

Lois discrètes. Pour le calcul des convolutions discrètes, le point crucial est

$$\delta_a * \delta_b = \delta_{a+b}$$

On utilise ensuite la distributivité.

Les fonctions caractéristiques sont souvent d'une grande aide pour le calcul de produit de convolution, pour la raison suivante :

Proposition 28

Si X et Y sont deux v.a. indépendantes, alors pour tout $t \in \mathbb{R}$,

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

Stabilité par convolution

On a les égalités en loi suivantes :

- $\mathcal{B}(n,p) * \mathcal{B}(m,p) = \mathcal{B}(n+m,p)$
- $\mathcal{N}(m_1,\sigma_1^2) * \mathcal{N}(m_2,\sigma_2^2) = \mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$
- $\mathcal{P}(\lambda_1) * \mathcal{P}(\lambda_2) = \mathcal{P}(\lambda_1+\lambda_2)$
- $\Gamma(\alpha,\nu) * \Gamma(\alpha,\nu') = \Gamma(\alpha,\nu+\nu')$
- Si X et Y sont des $\mathcal{C}(0,1)$ indépendantes, $\frac{X+Y}{2}$ est aussi $\mathcal{C}(0,1)$.

5. Lemme de Borel-Cantelli
Notation. Soit $(A_n)_{n\geq 1}$ une suite d'événements. On note

$$\limsup_{n\to+\infty} A_n \stackrel{\text{def}}{=} \bigcap_{k=1}^{+\infty} \bigcup_{n\geq k} A_n. = \{\omega \in \Omega \mid \omega \in A_n \text{ pour une infinité de } n\}$$

$$\liminf_{n\to+\infty} A_n \stackrel{\text{def}}{=} \bigcup_{k=1}^{+\infty} \bigcap_{n\geq k} A_n = \{\omega \in \Omega \mid \omega \in A_n \text{ à partir d'un certain rang}\}$$

On a $(\limsup A_n)^c = \liminf(A_n^c)$.

Théorème 29 *(lemme de Borel-Cantelli)*

Soit $(A_n)_{n\geq 1}$ une suite d'événements.

- *Si $\sum_{n\geq 1} P(A_n) < +\infty$ alors $P(\limsup_{n\to+\infty} A_n) = 0$.*
- *Si $\sum_{n\geq 1} P(A_n) = +\infty$ et si la famille $(A_n)_{n\geq 1}$ est indépendante alors $P(\limsup_{n\to+\infty} A_n) = 1$.*

Exemple. (le singe dactylographe de Borel) *Soit m une suite de lettres de longueur L. Soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes de loi uniforme sur les caractères $\{c_1, \ldots, c_p\}$. Existe-t-il $n \geq 0$ tel que $X_{n+1} \cdots X_{n+L}$ forment le mot m ? Autrement dit, un singe qui tape au hasard sur une machine à écrire finira-t-il par écrire Hamlet ?*

6. Loi du 0-1
Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilités et $(\mathcal{F}_n)_{n\geq 1}$ des sous-tribus de $\mathcal{F}$.

On note $\sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \ldots)$ la tribu engendrée par les tribus $(\mathcal{F}_k)_{k\geq n}$, et on pose

$$\mathcal{F}^\infty = \bigcap_{n\geq 1} \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \ldots).$$

$\mathcal{F}^\infty$ est appelée **tribu asymptotique**.

Exemple. Soit $(X_n)_{n\geq 1}$ des v.a. réelles sur $(\Omega, \mathcal{F}, P)$, et $\mathcal{F}_n = \sigma(X_n) = \{X^{-1}(B) \in \mathcal{F} \mid B \in \mathcal{B}(\mathbb{R})\}$.
L'idée est que les événements qui ne dépendent pas d'un nombre fini de X_i sont dans la tribu asymptotique. Si $A = \{\omega \in \Omega \mid X_n(\omega) = 0 \text{ pour une infinité de } n\}$, alors $A \in \mathcal{F}^\infty$, car $\omega \in A \Leftrightarrow \forall N,\ \exists n \geq N, X_n(\omega) = 0$ donc

$$A = \underbrace{\bigcap_{n\geq N} \underbrace{\bigcup_{n\geq N} X_n^{-1}(\{0\})}_{\in\sigma(\mathcal{F}_N, \mathcal{F}_{N+1}, \ldots)}}_{\in\mathcal{F}^\infty}$$

$$B = \left\{\omega \in \Omega \mid \lim_{n\to+\infty} \frac{X_1(\omega) + \cdots + X_n(\omega)}{n} \text{ existe}\right\} \in \mathcal{F}^\infty.$$

Théorème 30 *(loi du 0-1 de Kolmogorov)*

Soit $(\mathcal{F}_n)_{n\geq 1}$ une famille de sous-tribus indépendantes. Alors pour tout événement $A \in \mathcal{F}^\infty$, $P(A) = 0$ ou 1.

IV. Conditionnement

Dans tout le chapitre, on se place dans un espace de probabilités $(\Omega, \mathcal{F}, P)$.
Dans tous les énoncés, l'unicité s'entend "à un ensemble de mesure nulle près". Toutes les égalités ou inégalités faisant intervenir des espérances conditionnelles sont à comprendre au sens presque sûr.

1. Définition de l'espérance conditionnelle

1.1. Pour des variables dans $\mathbf{L^1(\Omega, \mathcal{F}, P)}$

Théorème-Définition 31

> *Soit $\mathcal{G}$ une sous tribu de $\mathcal{F}$ et $X \in L^1(\Omega, \mathcal{F}, P)$. Il existe une unique variable aléatoire dans $L^1(\Omega, \mathcal{G}, P)$ (donc $\mathcal{G}$-mesurable), notée $E_{\mathcal{G}}(X)$ et appelée **espérance conditionnelle** de X sachant $\mathcal{G}$, telle que*
>
> $$\forall B \in \mathcal{G}, \quad E(X\mathbf{1}_B) = E(E_{\mathcal{G}}(X)\mathbf{1}_B).$$
>
> *La relation ci-dessus est appelée **propriété caractéristique de l'espérance conditionnelle**.*
> *On a plus généralement que pour toute v.a. Z $\mathcal{G}$-mesurable bornée,*
>
> $$E(XZ) = E(E_{\mathcal{G}}(X)Z).$$
>
> *Si $X \geq 0$, on a aussi $E_{\mathcal{G}}(X) \geq 0$.*

Dans le cas où la sous-tribu $\mathcal{G}$ est la tribu engendrée par une v.a. Y, notée $\sigma(Y)$, on notera l'espérance conditionnelle correspondante $E_{\sigma(Y)}(X)$ ou $E_Y(X)$.

> **Aparté 10 : tribu engendrée par une v.a.**
>
> Soit Y une variable aléatoire réelle définie sur $(\Omega, \mathcal{F}, P)$. On note $\sigma(Y)$ la plus petite sous-tribu de $\mathcal{F}$ qui rend Y mesurable.
> On a alors
> $$\sigma(Y) = \{Y^{-1}(B); B \in \mathcal{B}(\mathbb{R}^d)\}.$$
> Soit Z est une variable aléatoire réelle définie $(\Omega, \mathcal{F}, P)$. Z est $\sigma(Y)$-mesurable si et seulement si il existe une fonction mesurable $h : \mathbb{R} \to \mathbb{R}$ telle que $Z = h(Y)$ (lemme de Doob).

La démonstration du Théorème-Définition ci-dessus utilise le résultat important de théorie de la mesure suivant :

Théorème 32 *(Radon-Nikodym)(Admis)*

> *Soient μ et ν deux mesures de masse finie sur $(\Omega, \mathcal{F}, P)$. On suppose que ν est absolument continue par rapport à μ. Alors il existe une unique fonction h positive dans $L^1(\Omega, \mathcal{F}, P)$ telle que*
>
> $$\forall A \in \mathcal{F}, \quad \nu(A) = \int_A h d\mu.$$
>
> *La fonction h est appelée dérivée de Radon-Nikodym (ou encore densité) de ν par rapport à μ.*

1.2. Pour des variables positives

Théorème 33

> *Soit X une v.a. à valeurs dans $[0,\infty]$.*
> *La formule $E_{\mathcal{G}}(X) := \lim_{n\to\infty} E_{\mathcal{G}}(\inf(X,n))$ (où la limite est croissante) définit une v.a. à valeurs dans $[0,\infty]$ qui est caractérisée par la propriété suivante :*
>
> $$\forall Z \ \mathcal{G}\text{-mesurable positive, } E(XZ) = E(E_{\mathcal{G}}(X)Z).$$

2. Propriétés de l'espérance conditionnelle

Proposition 34

> *On suppose que X et X' sont positives ou dans $L^1(\Omega,\mathcal{F},P)$.*
>
> 1. *(linéarité) Pour tout $a,b\in\mathbb{R}$, $E_{\mathcal{G}}(aX+bX') = aE_{\mathcal{G}}(X) + bE_{\mathcal{G}}(X')$.*
> 2. *Si X est $\mathcal{G}$-mesurable, $E_{\mathcal{G}}(X) = X$.*
> 3. *(convergence monotone) Si $(X_n)_{n\in\mathbb{N}}$ est une suite croissante de v.a. positives et X la limite croissante des X_n, alors*
> $$E_{\mathcal{G}}(X) = \lim_{n\to\infty} E_{\mathcal{G}}(X_n),$$
> *où la limite est croissante.*
> 4. *(Fatou) Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a. positives, alors*
> $$E_{\mathcal{G}}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} E_{\mathcal{G}}(X_n).$$
> 5. *(convergence dominée) Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. intégrables qui converge p.s. vers X. Suppposons qu'il existe une v.a. Z telle que $\forall n$, $|X_n| \leq Z$ p.s. et $E(Z) < \infty$. Alors*
> $$E_{\mathcal{G}}(X) = \lim_{n\to\infty} E_{\mathcal{G}}(X_n), \quad \text{p.s. et dans } L^1.$$
> 6. *Si $X\in L^1(\Omega,\mathcal{F},P)$, $E(E_{\mathcal{G}}(X)) = E(X)$.*
> 7. *Si $X\in L^1(\Omega,\mathcal{F},P)$, $|E_{\mathcal{G}}(X)| \leq E_{\mathcal{G}}(|X|)$ et donc $E(|E_{\mathcal{G}}(X)|) \leq E(|X|)$.*
> 8. *(Jensen) Si f est une fonction convexe positive ou telle que $f(X)\in L^1$, alors*
> $$E_{\mathcal{G}}(f(X)) \geq f(E_{\mathcal{G}}(X)).$$
> 9. *Si Y est $\mathcal{G}$-mesurable, avec X et Y positives ou bien X et XY dans L^1, alors*
> $$E_{\mathcal{G}}(XY) = YE_{\mathcal{G}}(X).$$
> 10. *Si $\mathcal{G}_1$ et $\mathcal{G}_2$ sont deux sous-tribus de $\mathcal{F}$ telles que $\mathcal{G}_1\subset\mathcal{G}_2$, $E_{\mathcal{G}_1}(E_{\mathcal{G}_2}(X)) = E_{\mathcal{G}_1}(X) = E_{\mathcal{G}_2}(E_{\mathcal{G}_1}(X))$.*

3. Espérance conditionnelle dans quelques cas particuliers

3.1. Cas particulier des variables de carré intégrable

On appelle $L^2(\Omega,\mathcal{G},P)$ le sous-espace fermé des éléments de $L^2(\Omega,\mathcal{F},P)$ dont au moins un représentant est $\mathcal{G}$-mesurable.

Théorème 35

> *Si $X\in L^2(\Omega,\mathcal{F},P)$, alors $E_{\mathcal{G}}(X)$ est la projection orthogonale de X sur $L^2(\Omega,\mathcal{G},P)$, pour le produit scalaire $(X,Y)\mapsto \mathrm{E}(XY)$.*

3.2. Retour sur le conditionnement discret

Soit B un événement tel que $P(B) > 0$. On peut alors définir une nouvelle probabilité P_B telle que

$$\forall A \in \mathcal{F}, P_B(A) = \frac{P(A \cap B)}{P(B)}.$$

Pour toute variable aléatoire X positive ou dans $L^1(\Omega, \mathcal{F}, P)$, on peut donc définir son espérance sous P_B, donnée par

$$E_B(X) := \int_\Omega X(\omega) dP_B(\omega).$$

On peut vérifier que

$$E_B(X) = \frac{E(X\mathbf{1}_B)}{P(B)}.$$

La plus petite tribu qui contient B est $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$. On peut vérifier que $E_B(X) = E_\mathcal{G}(X)$.

De même, si Y est une v.a. discrète à valeurs dans E et $E' = \{y \in E; P(Y = y) > 0\}$, alors on peut vérifier que $E_Y(X) = \varphi(Y)$ où la fonction $\varphi : E \mapsto \mathbb{R}$ vaut 0 sur $E \setminus E'$ et $\varphi(y) = \frac{E(X\mathbf{1}_{Y=y})}{P(Y=y)}$ si $y \in E'$.

3.3. Conditionnement gaussien

Aparté 11 : Vecteurs gaussiens

Soit C une matrice de taille $d \times d$ symétrique positive, à coefficients dans $\mathbb{R}$. Un **vecteur gaussien centré de matrice de covariance C** est un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$, dans L^2, dont la fonction caractéristique est donnée par

$$\varphi_X(t_1, \ldots, t_d) = E(e^{i\langle t, X \rangle}) = \exp\left(-\frac{1}{2}\sum_{j,k=1}^{d} C_{jk} t_j t_k\right), \quad \forall t = (t_1, \ldots, t_d) \in \mathbb{R}^d.$$

On dit alors que X suit une loi $\mathcal{N}(0, C)$. Cette notation est justifiée par le fait que son espérance est nulle et sa matrice de covariance est C.

Pour toute matrice réelle symétrique, positive, il existe un vecteur gaussien centré de matrice de covariance C, obtenu comme AY, où $Y = (Y^1, \ldots, Y^d)$ un vecteur aléatoire dont les coordonnées sont des variables aléatoires réelles indépendantes, toutes de loi $\mathcal{N}(0, 1)$ et $A = \sqrt{C}$.

Un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$ est un vecteur gaussien centré si et seulement si toute combinaison linéaire de ses coordonnées est une variable aléatoire (réelle) gaussienne centrée.

Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$, de loi $\mathcal{N}(0, C)$. Si C n'est pas inversible, la loi de X n'a pas de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Si C est inversible, la densité de la loi de X est donnée par, $\forall x \in \mathbb{R}^d$

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{det(C)}} exp\left(-\frac{1}{2}\langle x, C^{-1}x\rangle\right).$$

Proposition 36

Soit $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ un vecteur gaussien centré. Alors les vecteurs $(X_1, \ldots, X_m)$ et $(Y_1, \ldots, Y_n)$ sont indépendants si et seulement si, pour tout couple (i, j) avec $i \in \{1, \ldots, m\}$ et $j \in \{1, \ldots, n\}$, $cov(X_i, Y_j) = 0$.

Pour les vecteurs gaussiens, les calculs d'espérance conditionnelle se ramènent à une projection orthogonale, pour le produit scalaire $(X, Y) \mapsto \mathrm{E}(XY)$

Proposition 37

Soit $(Y_1, \ldots, Y_n, X)$ *un vecteur gaussien centré. Alors l'espérance conditionnelle de* X *sachant* $(Y_1, \ldots, Y_n)$ *est donnée par*

$$E[X|Y_1, \ldots, Y_n] = \hat{X},$$

où $\hat{X}$ *est la projection orthogonale de* X *sur l'espace vectoriel* $\{\sum_{k=1}^{n} \alpha_k Y_k; \alpha_k \in \mathbb{R}\}$.

4. Espérance conditionnelle et indépendance

Théorème 38

Deux sous-tribus $\mathcal{G}_1$ *et* $\mathcal{G}_2$ *sont indépendantes si et seulement si pour toute v.a.* $\mathcal{G}_2$*-mesurable positive ou dans* $L^1(\Omega, \mathcal{G}_2, P)$, $E_{\mathcal{G}_1}(X) = E(X)$.

En particulier, si X et Y sont deux v.a. réelles, X et Y sont indépendantes si et seulement si pour toute fonction h borélienne telle que $E(|h(X)|) < \infty$, $E_Y(h(X)) = E(h(X))$.

Théorème 39

Soient X *et* Y *sont deux v.a. réelles. Supposons que* X *est indépendante de* $\mathcal{B}$ *et* Y $\mathcal{B}$*-mesurable. Alors, pour tout* $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$*, mesurable,*

$$E_{\mathcal{B}}(g(X, Y)) = \int g(x, Y) P_X(dx).$$

Partie 2. Les principaux théorèmes limites

Plan de la deuxième partie :

V. Les différentes notions de convergence pour des v.a. réelles

[Barbe-Ledoux, Foata-Fuchs chap 16, Ouvrard 2]

Dans tout ce qui suit, $(X_n)_{n\geq 1}$ et X sont des v.a. définies sur un même espace de probabilités $(\Omega, \mathcal{F}, P)$ et à valeurs dans $\mathbb{R}$.

1. Convergence presque sûre
On dit que la suite de v.a. $(X_n)_{n\geq 1}$ **converge presque sûrement** (p.s.) vers la v.a. X si

$$P\left(\lim_{n\to+\infty} X_n = X\right) = 1.$$

Autrement dit, il existe un sous-ensemble Ω' tel que $P(\Omega') = 1$ et $\forall \omega \in \Omega'$, $\lim\limits_{n\to+\infty} X_n(\omega)$ existe et vaut $X(\omega)$.
On note $X_n \xrightarrow{p.s.} X$.

Proposition 40 *(un critère pour la convergence p.s.)*

1. *Si* $\forall \epsilon > 0, \sum_{n\geq 1} P(|X_n - X| > \epsilon) < \infty$, *alors* $X_n \xrightarrow{p.s.} X$.
2. *On suppose que les* $(X_n)_{n\geq 1}$ *sont indépendants. Alors* $X_n \xrightarrow{p.s.} 0$ *ssi* $\forall \epsilon > 0, \sum_{n\geq 1} P(|X_n| > \epsilon) < \infty$.

Proposition 41 *(une autre caractérisation de la convergence p.s.)*

La suite $(X_n)_{n\geq 1}$ *converge p.s. vers* X *ssi*

$$\forall \epsilon > 0, \lim_{n\to\infty} P\left(\sup_{k\geq n} |X_k - X| \geq \epsilon\right) = 0.$$

2. Convergence en probabilité
On dit que la suite $(X_n)_{n\geq 1}$ **converge en probabilité** vers la v.a. X si

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

On note $X_n \xrightarrow{P} X$.

Proposition 42 *(lien entre convergence p.s. et convergence en probabilité)*

Si $X_n \xrightarrow{p.s.} X$ alors $X_n \xrightarrow{P} X$.

La réciproque est fausse en général.

Contre-exemple : convergence en proba mais pas p.s.

On définit $(A_n)_{n\geq 1}$ des sous-intervalles de $[0,1]$ de la façon suivante : pour $k \geq 0$ et $0 \leq j \leq 2^k - 1$,

$$A_{2^k+j} = \left[\frac{j}{2^k}, \frac{j+1}{2^k}\right].$$

Chacun des intervalles $(A_i)_{2^k \leq i < 2^{k+1}}$ est de longueur $1/2^k$, et leur union vaut $[0,1]$.
On prend P la mesure de Lebesgue sur $\Omega = [0,1]$ et on pose $X_n = \mathbb{1}_{A_n}$.
On peut vérifier que $X_n \xrightarrow{P} 0$ mais $(X_n)_{n\geq 1}$ ne converge pas p.s. vers 0.

On a tout de même la réciproque partielle suivante :

Proposition 43

Si $X_n \xrightarrow{P} X$ alors il existe une sous-suite extraite $(X_{n_k})_{k\geq 1}$ qui converge p.s. vers X.

On a même un résultat plus fort :

Proposition 44

$X_n \xrightarrow{p.s.} X$ si et seulement si de toute sous-suite extraite $(X_{n_k})_{k\geq 1}$ on peut extraire une sous-suite qui converge p.s. vers X.

On en déduit que la convergence en probabilité est stable par les opérations usuelles :

Proposition 45

On suppose $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$ et φ continue de $\mathbb{R}^2$ dans $\mathbb{R}$ alors $\varphi(X_n, Y_n) \xrightarrow{P} \varphi(X, Y)$.

En particulier, si $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, alors pour touts réels α et β, $\alpha X_n + \beta Y_n \xrightarrow{P} \alpha X + \beta Y$, ou encore $X_n Y_n \xrightarrow{P} XY$, etc.

Soit $L^0(\Omega, \mathcal{F}, P)$ l'ensemble des v.a. réelles sur $(\Omega, \mathcal{F}, P)$, quotienté par la relation d'équivalence $X \sim Y$ ssi $X = Y$ p.s. Pour tous $X, Y \in L^0(\Omega, \mathcal{F}, P)$, on définit $d(X,Y) = E(min(|X-Y|, 1))$.

Proposition 46 *(métrisabilité de la convergence en probabilité)*

d est une distance sur $L^0(\Omega, \mathcal{F}, P)$ et $(X_n)_{n\geq 1}$ converge en probabilité vers X ssi $d(X_n, X)$ converge vers 0. On dit que d métrise la convergence en probabilité.

On peut même montrer que $L^0(\Omega, \mathcal{F}, P)$ muni de la distance d est un espace métrique complet :

Proposition 47 *(complétude de $L^0(\Omega, \mathcal{F}, P)$)*

Supposons que $(X_n)_{n\geq 1}$ vérifie le critère de Cauchy pour la distance d, i.e. $\forall \epsilon > 0, \exists n_0, \forall n \geq n_0$, $d(X_n, X_{n_0}) \leq \epsilon$. Alors $(X_n)_{n\geq 1}$ converge en probabilité.

3. Convergences $L^p, p > 0$
La suite de v.a. $(X_n)_{n\geq 1}$ **converge dans** L^p vers la v.a. X si $\lim_{n\to+\infty} \|X_n - X\|_p = 0$, autrement dit

$$\lim_{n\to+\infty} E(|X_n - X|^p) = 0.$$

On note $X_n \xrightarrow{L^p} X$.

Les plus usitées sont les convergences L^1 ou L^2.

Il est facile de vérifier par Jensen que s'il y a convergence L^p, il y a convergence $L^{p'}$ pour tout $0 < p' < p$.

Proposition 48 *(lien avec la convergence en probabilité)*

> *Si $X_n \xrightarrow{L^p} X$ alors $X_n \xrightarrow{P} X$.*
>
> *La réciproque est fausse en général.*

Pour établir une réciproque partielle, on rappelle la notion d'uniforme intégrabilité :

Une famille $(X_i)_{i\in I}$ de v.a. réelles, intégrables, est dite équiintégrable ou uniformément intégrable si

$$\lim_{c\to\infty} \sup_{i\in I} \int_{\{|X_i|>c\}} |X_i| dP = 0.$$

On rappelle quelques critères d'uniforme intégrabilité :
- une famille finie de v.a. intégrables est uniformément intégrable,
- si il existe une v.a. Y intégrable telle que, p.s. $\forall i \in I, |X_i| \leq Y$, alors la famille $(X_i)_{i\in I}$ est uniformément intégrable,
- la famille $(X_i)_{i\in I}$ est uniformément intégrable ssi ($\sup_{i\in I} E(|X_i|) < \infty$ et $\forall \epsilon > 0, \exists \eta > 0, \forall A$ with $P(A) \leq \eta, \forall i \in I, \int_A |X_i| dP \leq \epsilon$.

aussi que si alors la famille $(X_i)_{i\in I}$ est uniformément intégrable.

Proposition 49 *(lien entre les convergences L^1 et en probabilité)*

> *Soit $(X_n)_{n\geq 1}$ une suite de v.a. intégrables. Les deux propositions suivantes sont équivalentes :*
> 1. *$X_n \xrightarrow{P} X$ et la famille $(X_n)_{n\geq 1}$ est équiintégrable.*
> 2. *X est intégrable et $\lim_{n\to\infty} \|X_n - X\|_1 = 0$.*

Remarque : La convergence L^p n'implique pas la convergence p.s., ni l'inverse.

Contre-exemple : convergence dans tous les $L^p, 0 < p < \infty$ mais pas p.s.
Dans l'exemple donné aprés la Proposition 42, la suite $(X_n)_{n\geq 1}$ converge dans tous les $L^p, 0 < p < \infty$ mais pas p.s.

Contre-exemple : convergence p.s. mais dans aucun $L^p, p > 0$.
Soit $(Y_n)_{n\geq 1}$ une suite de v.a. indépendantes, telles que $P(Y_n = e^n) = \frac{1}{n^2}$ et $P(Y_n = 0) = 1 - \frac{1}{n^2}$. On peut vérifier que $Y_n \xrightarrow{p.s.} 0$ mais il n'y a convergence dans aucun $L^p, p > 0$.

4. Convergence en loi

On ne suppose plus nécessairement que les v.a. $(X_n)_{n\geq 1}$ sont définies sur le même espace de probabilités. Ce mode de convergence correspond à la convergence étroite des lois des v.a. $(X_n)_{n\geq 1}$.

4.1. Définition, caractérisation

Théorème 50 *(définition et caractérisation de la convergence en loi)*

> *Soient $(X_n)_{n\geq 1}$ et X des v.a. réelles. On dit que $(X_n)_{n\geq 1}$ converge en loi vers X si l'une de ces trois conditions équivalentes est vérifiée :*
>
> 1. *$\forall \varphi : \mathbb{R} \to \mathbb{R}$ continue bornée,*
> $$\lim_{n\to\infty} E(\varphi(X_n)) = E(\varphi(X)).$$
> 2. *En tout point de continuité t de F_X, on a*
> $$\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$$
> 3. *Il existe un espace de probabilité $(\Omega', \mathcal{F}', P')$ sur lequel sont définies des v.a. $(X'_n)_{n\geq 1}$ et X' de même loi respectivement que $(X_n)_{n\geq 1}$ et X telles que $X'_n \xrightarrow{p.s.} X'$.*

On note alors $X_n \xrightarrow{\mathcal{L}} X$.

> **Aparté 8 : Convergences vague, faible, étroite de mesures**
>
> Soient $(\mu_n)_{n\geq 1}$ une suite de mesures sur $\mathbb{R}$ et μ une mesure sur $\mathbb{R}$.
>
> 1. Si pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$ continue à support compact, $\int \varphi d\mu_n \to \int \varphi d\mu$, on dit que la suite $(\mu_n)_{n\geq 1}$ converge **vaguement** vers μ.
> 2. Si pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$ continue tendant vers zéro à l'infini, $\int \varphi d\mu_n \to \int \varphi d\mu$, on dit que la suite $(\mu_n)_{n\geq 1}$ converge **faiblement** vers μ.
> 3. Si pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$ continue bornée, $\int \varphi d\mu_n \to \int \varphi d\mu$, on dit que la suite $(\mu_n)_{n\geq 1}$ converge **étroitement** vers μ.
>
> Si $(\mu_n)_{n\geq 1}$ **et** μ sont des mesures de probabilités sur $\mathbb{R}$, les trois notions coïncident (mais attention $(\delta_n)_{n\geq 1}$ converge faiblement mais pas étroitement vers la mesure nulle).

On a aussi la caract érisation suivante :

Théorème 51 *(dit de Porte-Manteau)(Admis)*

> *Soient $(X_n)_{n\geq 1}$ et X des v.a. réelles. Les assertions suivantes sont équivalentes :*
>
> 1. *$X_n \xrightarrow{\mathcal{L}} X$.*
> 2. *Pour tout ouvert G de $\mathbb{R}$, $\liminf_{n\to\infty} P_{X_n}(G) \geq P_X(G)$*
> 3. *Pour tout fermé F de $\mathbb{R}^d$, $\limsup_{n\to\infty} P_{X_n}(F) \leq P_X(F)$*
> 4. *Pour tout borélien B de $\mathbb{R}^d$ tel que $P_X(\partial B) = 0$, $\lim_{n\to\infty} P_{X_n}(B) = P_X(B)$.*

Attention, pas d'opérations usuelles sur la convergence en loi.

4.2. Lien avec les autres modes de convergence

Proposition 52

> *Si $X_n \xrightarrow{p.s.} X$ alors $X_n \xrightarrow{\mathcal{L}} X$.*
> *Si $X_n \xrightarrow{P} X$ alors $X_n \xrightarrow{\mathcal{L}} X$.*

Les réciproques sont fausses
Exemple. $X_n = (-1)^n X$ avec X de loi $\mathcal{N}(0,1)$ converge en loi mais ne converge pas en probabilité, donc ne converge pas p.s.

On a cependant une réciproque partielle importante (très utilisée en statistiques) :

Proposition 53 *(lemme de Slutsky)*

> *Si X_n converge en loi vers une constante c alors X_n converge en probabilité vers c.*
> *Par conséquent, si X_n converge en loi vers X et Y_n vers c, alors (X_n, Y_n) converge en loi vers (X, c).*

4.3. Convergence en loi pour des v.a. discrètes

Proposition 54

> *Soit $(X_n)_{n\geq 0}$ et X des v.a. à valeurs dans $\mathbb{N}$. Alors $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si pour tout entier $k \in \mathbb{N}$,*
>
> $$\lim_{n\to+\infty} P(X_n = k) = P(X = k).$$

Une application importante est le résultat suivant :

Proposition 55

> *Soit $(X_n)_{n\geq 1}$ des v.a. de loi binomiale $\mathcal{B}(n, p_n)$.*
> *Si $\lim\limits_{n\to+\infty} np_n = \lambda > 0$ alors X_n converge en loi vers une v.a. de loi de Poisson $\mathcal{P}(\lambda)$.*

Conséquence : approximation d'une binomiale par une loi de Poisson.
Il est d'usage de remplacer $\mathcal{B}(n, p)$ par $\mathcal{P}(np)$ quand $p < 0,1$ et n est grand (par exemple $n \geq 100$).

4.4. Une caractérisation très importante de la convergence en loi

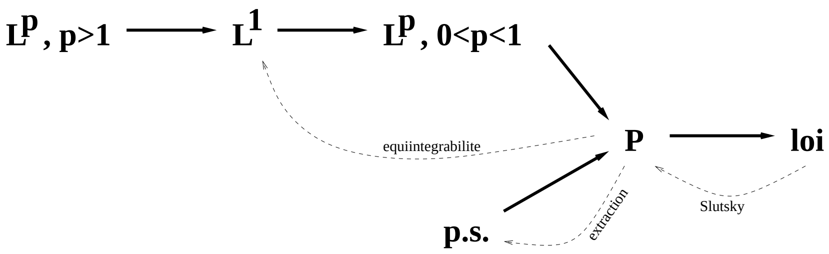
Théorème 56 *(Lévy)*

> *Soient $(X_n)_{n\geq 1}$ et X des v.a. réelles.*
>
> 1. *Si $X_n \xrightarrow{\mathcal{L}} X$, alors $\lim\limits_{n\to+\infty} \varphi_{X_n}(t) = \varphi_X(t)$ pour tout $t \in \mathbb{R}$.*
> 2. *Si $\lim\limits_{n\to+\infty} \varphi_{X_n}(t) = \varphi(t)$ pour tout $t \in \mathbb{R}$ et φ est la fonction caractéristique d'une v.a. X, alors $X_n \xrightarrow{\mathcal{L}} X$.*

Il ne suffit pas d'avoir la convergence des fonctions φ_{X_n}, il faut s'assurer que la limite est la fonction caractéristique d'une certaine loi.
Une condition (admise) pour que φ soit la fonction caractéristique d'une v.a. est que φ soit une limite de fonctions caractéristiques, continue en 0.

5. Résumé des liens entre les différents types de convergence

Sur le schéma ci-dessous, les flèches pleines indiquent les liens qui ont toujours lieu et les flèches en pointillés indiquent les réciproquent partielles.



VI. Loi des grands nombres

On considère $(X_n)_{n\geq 1}$ une suite de v.a. toutes définies sur le même $(\Omega, \mathcal{F}, P)$. On note $S_n = X_1 + \cdots + X_n$, et on s'intéresse à la convergence de la suite $\frac{S_n}{n}$.

On dit que les v.a. $(X_n)_{n\geq 1}$ sont **iid** si elles sont indépendantes et de même loi.

1. Loi faible des grands nombres

On parle de loi faible lorsqu'il y a convergence en probabilité.

Théorème 57 *(loi faible - v.a. L^2 non corrélées)*

> *Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles dans L^2, centrées, non corrélées deux à deux (c'est-à-dire $\mathrm{Cov}(X_i, X_j) = 0$ si $i \neq j$).*
> *S'il existe C tel que $\forall n \geq 1,\ \mathrm{Var}(X_n) \leq C$, alors $\frac{S_n}{n}$ converge vers 0 en probabilité et dans L^2.*

2. Lois fortes des grands nombres

[Ouvrard 2 chap 10]

On parle de loi forte des grands nombres lorsqu'il y a convergence p.s.

On peut renforcer le théorème précédent de la façon suivante :

Théorème 58 *(loi forte - v.a. iid intégrables)*

> *Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles, iid. Les deux conditions suivantes sont équivalentes :*
> 1. $E(|X_1|) < \infty$
> 2. $\dfrac{S_n}{n} \xrightarrow{p.s.} E(X_1)$.

On a l'interprétation suivante en termes de fréquence empirique (cf l'énoncé de la loi des grands nombres dans les programmes de lycée) :

Proposition 59

Si $(A_n)_{n\geq 1}$ sont des événements indépendants de même probabilité, on a

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{A_n} \xrightarrow{p.s.} P(A_1).$$

Pour montrer la loi forte ci-dessus, on a besoin de la version plus facile suivante :

Proposition 60 *(loi forte - v.a. indépendantes, L^4)*

Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles L^4, centrées, indépendantes. On suppose qu'il existe C tel que $\forall n \geq 1,\ E((X_n)^4) \leq C$. Alors $\frac{S_n}{n}$ converge vers 0 p.s. et dans L^4.

On a aussi une version plus forte de la proposition précédente :

Théorème 61 *(loi forte - indépendantes, L^2)(Admis)*

Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles L^2, centrées, indépendantes. Supposons que $\sum_{n\geq 1}\frac{\mathrm{Var}(X_n)}{n^2} < +\infty$. Alors $\frac{S_n}{n}$ converge vers 0 p.s. et dans L^2.

3. Convergence de la fonction de répartition empirique

On énonce maintenant un prolongement important de la loi des grands nombres, particulièrement utile en statistiques.

Soit $(X_n)_{n\geq 1}$ des v.a. réelles. La fonction de répartition empirique de $X_1, \ldots, X_n$ est :

$$F_n(x,\omega) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{]-\infty,x]}(X_k(\omega)).$$

Théorème 62 *(de Glivenko-Cantelli)*

Si $(X_n)_{n\geq 1}$ est une suite de v.a. réelles iid et F la fonction de répartition commune des X_n, on a que pour presque tout ω,

$$\sup_{x\in\mathbb{R}}|F_n(x,\omega) - F(x)| \xrightarrow{p.s.} 0$$

VII. Théorème Central Limite (TCL)

1. Enoncés

Soit $(X_n)_{n\geq 1}$ des v.a. L^2, indépendantes, de même loi, et $S_n = X_1 + \cdots + X_n$. Par la loi des grands nombres, $\frac{S_n}{n}$ converge p.s. vers $E(X_1)$. Autrement dit, p.s. $\dfrac{S_n(\omega)}{n} = E(X_1) + o(1)$. Le théorème central limite précise le $o(1)$

Théorème 63 *(central limite (TCL))*

> *Soit $(X_n)_{n\geq 1}$ des v.a. L^2, indépendantes, de même loi, et $S_n = X_1 + \cdots + X_n$. Alors*
>
> $$\frac{S_n - nE(X_1)}{\sqrt{n\mathrm{Var}(X_1)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \;\text{ quand } n \to +\infty.$$

Il est équivalent de dire que

$$\sqrt{n}\left(\frac{S_n}{n} - E(X_1)\right) = \frac{S_n - nE(X_1)}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{Var}(X_1)).$$

Une conséquence du théorème central limite est la suivante : pour tout $a < b$,

$$P\left(a \leq \frac{S_n - nE(X_1)}{\sqrt{n\mathrm{Var}(X_1)}} \leq b\right) \longrightarrow \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}dt.$$

Le théorème de de Moivre (historiquement plus ancien) est le TCL pour des lois de Bernoulli.

Théorème 64 *(de de Moivre)*

> *Si S_n suit une loi $\mathcal{B}(n,p)$ alors*
>
> $$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \longrightarrow \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}dt.$$

Il est d'usage d'approcher $\frac{S_n - np}{\sqrt{np(1-p)}}$ par $\mathcal{N}(0,1)$ dès que $np(1-p) > 10$. La convergence est plus rapide si p est proche de $1/2$. Attention à la correction de continuité !

On a aussi le résultat suivant :

Théorème 65 *(TCL poissonien)*

> *Pour tout $\lambda > 0$, soit X_λ une v.a. de loi de Poisson $\mathcal{P}(\lambda)$. Alors*
>
> $$\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{loi} \mathcal{N}(0,1) \text{ quand } \lambda \to +\infty.$$