Cours de probabilités
Prépa Agreg

Clément Pellegrini

- **1) Rappel de théorie de la mesure**
- **2) Variables aléatoires**
- **3) Espérance et Variance**
- **4) Transformées exponentielles**
- **5) Indépendance**
- **6) Conditionnement**
- **7) Convergence**
- **8) Loi des grands nombres et Théorème centrale limite**
- **9) Vecteurs gaussiens**
- **10) Estimateur du maximum de vraisemblance**

1) Rappel de théorie de la mesure

Soit Ω un ensemble

Définition 1. *$\mathcal{A} \subset \mathcal{P}(\Omega)$ est une tribu (ou σ-algèbre) sur Ω si les conditions suivantes sont satisfaites*

1. $\Omega \in \mathcal{A}$
2. *$\mathcal{A}$ est stable par passage au complémentaire : si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$*
3. *$\mathcal{A}$ est stable par réunion dénombrable : soit $(A_n)_n$ une suite d'éléments de $\mathcal{A}$ i.e $A_n \in \mathcal{A}$ pour tout $n \in \mathbb{N}$ alors $\bigcup_n A_n \in \mathcal{A}$*

1. $\{\emptyset, \Omega\}$ s'appelle la **tribu grossière**
2. $\mathcal{P}(\Omega)$ s'appelle la **tribu triviale**, c'est souvent celle considérée lorsque Ω est discret ou dénombrable
3. Lorsque Ω est un espace topologique c'est à dire muni d'une famille d'ouverts, la plus petite tribu contenant ces ouverts est appelée **tribu borélienne**. On la note $\mathcal{B}(\Omega)$. Pourquoi existe t'elle toujours ?

Définition 2. *Soit X et Y deux ensembles munis de tribus $\mathcal{A}$ pour X et $\mathcal{B}$ pour Y. Une application $f : (X, \mathcal{A}) \to (Y, \mathcal{B})$ est dite mesurable si*

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

Rappel : de manière générale **toute application continue** entre espaces topologiques **est mesurable** si on munit ces espaces de leurs tribus boréliennes.

Un ensemble Ω muni d'une tribu $\mathcal{A}$ s'appelle un espace mesurable on le note $(\Omega, \mathcal{A})$

Définition 3. *Une mesure μ sur $(\Omega, \mathcal{A})$ est une application de $\mathcal{A} \to [0, +\infty]$ telle que*

1. $\mu(\emptyset) = 0$
2. *Si $(A_n)_n$ est une suite dénombrable d'ensembles de $\mathcal{A}$ deux à deux disjoints alors*

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

1. Le triplet $(\Omega, \mathcal{A}, \mu)$ est appelé espace mesuré.
2. **Lorsque μ est une mesure de masse** 1 c'est à dire $\mu(\Omega) = 1$ on parle de **mesure de probabilité**. Dans ce cas on note souvent la mesure μ par $\mathbb{P}$.
3. **Un espace de probabilité** est donc un espace mesurable $(\Omega, \mathcal{A})$ muni d'une mesure de probabilité $\mathbb{P}$.

On rappelle que

Proposition 1.
1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
3. *Si $A \subset B$: $\mathbb{P}(A) \leqslant \mathbb{P}(B)$*

4. *Si* (A_n) *est une suite croissante i.e* $A_n \subset A_{n+1}$

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcup_n A_n)$$

5. *Si* (A_n) *est une suite décroissante i.e* $A_{n+1} \subset A_n$

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcap_n A_n)$$

Définition 4. $\mathcal{M} \subset \mathcal{P}(\Omega)$ *est une classe monotone (c.m. en abrégé) si :*

1. $\Omega \in \mathcal{M}$,
2. $\big[A, B \in \mathcal{M}$ et $B \subset A\big] \Rightarrow A \setminus B = A \cap B^c \in \mathcal{M}$,
3. $\big[\forall i \in \mathbb{N},\ A_i \in \mathcal{M}$ et $A_i \subset A_{i+1}\big] \Rightarrow \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{M}$.

— Toute tribu est une classe monotone
— Soient μ et ν deux mesures. L'ensemble $\{A \,|\, \mu(A) = \nu(A)\}$ est une classe monotone

Définition 5. *Une intersection quelconque de c.m. est une c.m.*

Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$. *La classe monotone engendrée par* $\mathcal{E}$, *notée* $\mathcal{M}(\mathcal{E})$, *est par définition l'intersection de toutes les c.m. contenant* $\mathcal{E}$.
C'est la plus petite c.m. sur Ω *qui contient* $\mathcal{E}$.

Proposition 2. *Soit* $\mathcal{M}$ *une classe monotone. Si elle est stable par intersection (ou union) finie, alors c'est une tribu.*

Démonstration. En effet il suffit de voir que si $(A_n)_n$ est une suite d'éléments de $\mathcal{M}$ alors

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} (A_1 \cup \ldots \cup A_n)$$

et si on a stabilité par union finie alors la propriété 3 de la définition ci-dessus permet de conclure. □

Théorème 1 (Classes monotones)**.** *Soit* $\mathcal{E} \subset \mathcal{P}(\Omega)$ *stable par intersection finie. Alors* $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.

Définition 6. *Un ensemble* $\mathcal{H}$ *de fonctions de* $\Omega \to \mathbb{R}$ *est dit stable par convergence monotone bornée si pour toute suite* $(f_n)_{n\in\mathbb{N}}$, *croissante et bornée* ($\exists\, C,\ \forall n,\ \forall \omega,\ |f_n(\omega)| \leqslant C$) *de fonctions de* $\mathcal{H}$, *la limite* $f = \lim f_n$ *est dans* $\mathcal{H}$.

Théorème 2. (Version fonctionnelle du théorème de classe monotone)

Soit $\mathcal{H}$ *un* $\mathbb{R}$*-espace vectoriel de fonctions bornées de* $\Omega \to \mathbb{R}$ *stable par convergence monotone bornée et contenant les constantes. Soit* $\mathcal{E} \subset \mathcal{H}$ *un ensemble de fonctions stable par multiplication.*

Alors $\mathcal{H}$ *contient l'ensemble* $b\left(\sigma(\mathcal{E})\right)$ *des fonctions bornées mesurables pour*

$$\sigma(\mathcal{E}) = \sigma\left(\left\{f^{-1}(B),\ f \in \mathcal{E},\ B \in \mathcal{B}(\mathbb{R})\right\}\right).$$

Théorème 3. (Caractérisation des mesures finies)

Soit $(\Omega, \mathcal{A})$ *un espace mesuré. Soit* $\mathcal{E} \subset \mathcal{A} \subset \mathcal{P}(\Omega)$ *telle que*

— $\Omega \in \mathcal{E}$,

— $A, B \in \mathcal{E} \Rightarrow A \cap B \in \mathcal{E}$,

— $\sigma(\mathcal{E}) = \mathcal{A}$.

Si μ et ν *sont deux mesures finies sur* $(\Omega, \mathcal{A})$ *qui coincident sur* $\mathcal{E}$ *alors* $\mu = \nu$.

2) Variables aléatoires

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité

Définition 7. *On appelle* ***variable aléatoire (v.a) toute application mesurable*** *X d'un espace de probabilité Ω dans $\mathbb{R}$ munit de la tribu borélienne $\mathcal{B}(\mathbb{R})$.*

Si la variable aléatoire est à valeurs dans $\mathbb{R}^d$ on parle de vecteur aléatoire (couple si d=2).

Définition 8. *Soit $X : \Omega \to \mathbb{R}$ une v.a, on appelle loi de X la mesure de probabilité $\mathbb{P}_X$ qui est la mesure image sur $\mathbb{R}$ de $\mathbb{P}$ par X c'est à dire*

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) := \mathbb{P}(X \in A)$$

Pour parler de la loi d'une v.a X on utilise la notation $\sim$. Par exemple $X \sim \mathcal{P}(\lambda)$ pour une loi de Poisson de paramètre λ.

Soit X une v.a.

Définition 9.
- *On appelle atome de X (ou de sa loi) tout $a \in \mathbb{R}$ tel que $\mathbb{P}(X = a) \neq 0$*
- *On appelle* ***support (topologique)*** *d'une v.a X le plus petit fermé F tel que $\mathbb{P}(X \in F) = 1$. On notera $S(X)$ le support d'une v.a X.*

Définition 10. *La tribu engendrée par X, notée $\sigma(X)$, est la plus petite tribu $\mathcal{G}$ sur Ω qui rend l'application $X : (\Omega, \mathcal{G}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})$ mesurable.*

$$\sigma(X) = \{X^{-1}(A), A \in \mathcal{B}(\mathbb{R})\}$$

1. On interprète $\sigma(X)$ comme l'ensemble de tous les évènements accessibles à partir de la connaissance de X. Cela représente en quelque sorte toute l'information nécessaire véhiculée par X.
2. On généralise cela à plusieurs v.a $X_1, \ldots, X_n$ et on note $\sigma(X_1, \ldots, X_n)$ la plus petite tribu rendant les X_i mesurables.

Proposition 3 (Doob)**.** *Soit $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E})$ et $Y : \Omega \to \mathbb{R}^d$ deux v.a.. Alors Y est $\sigma(X)$-mesurable si et seulement s'il existe une fonction $h : E \to \mathbb{R}^d$ borélienne avec $Y = h(X)$.*

Définition 11. *On appelle fonction de répartition d'une v.a $X : \Omega \to \mathbb{R}$ la fonction F_X définie sur $\mathbb{R}$ par*

$$F_X(t) = \mathbb{P}(X \leqslant t) = \mathbb{P}_X(]-\infty, t])$$

Proposition 4.
1. *F_X est croissante*
2. *$\lim_{t\to-\infty} F_X(t) = 0$ et $\lim_{t\to+\infty} F_X(t) = 1$*
3. *F_X est continue à droite et admet une limite à gauche*

4. Si X n'a pas d'atome F_X est continue

Démonstration. Soient $t \leqslant t'$ alors $]-\infty, t] \subset]-\infty, t']$ et donc $F_X(t) \leqslant F_X(t')$ d'où F_X est croissante. Soit (t_n) une suite croissante telle que $t_n \to +\infty$ alors $]-\infty, t_n] \subset]-\infty, t_{n+1}]$ et donc

$$\lim_n F_x(t_n) = P_X(\bigcup]-\infty, t_n]) = P_X(\mathbb{R}) = P(X \in \mathbb{R}) = 1$$

Pour la limite en 0 on utilise une suite décroissante qui tend vers $-\infty$.

Pour la continuité à droite en t_0 on considère (t_n) qui décroit vers t_0 on alors

$$\bigcup]-\infty, t_n] =]-\infty, t_0]$$

et on utilise la limite croissante i.e

$$\lim F_X(t_n) = P_X(\bigcup]-\infty, t_n]) = P_X(]-\infty, t_0]) = F_X(t_0).$$

D'où $F_X(t_0-) = F_X(t_0)$. Pour la limite à gauche soit (t_n) une suite croissante vers t_0 avec $t_n > t_0$ on a donc

$$\bigcup]-\infty, t_n] =]-\infty, t_0[.$$

Par limite décroissante $F_X(t_0+) = P_X(]-\infty, t_0[)$. Ainsi F_X est continue en t_0 ssi $P_X(]-\infty, t_0[) = P_X(]-\infty, t_0])$ i.e $P_X(\{t_0\}) = 0$.

Notons $\mathcal{D}$ l'ensemble des points de discontinuité de F_X. Soient $a > b$ deux points de $\mathcal{D}$. Considérons les intervalles non vides $]F_X(a-), F_X(a+)[$ et $]F_X(b-), F_X(b+)[$. La croissance de F_X implique que ces intervalles sont disjoints. On peut donc sélectionner un rationnel dans et injecter $\mathcal{D}$ dans $\mathbb{Q}$ d'où la dénombrabilité. $\square$

— Les éventuels sauts de F_X qui sont en nombre dénombrable (pourquoi ?) correspondent aux atomes de X

— Le quantile d'ordre α est toute valeur z_α telle que

$$\mathbb{P}(X \leqslant z_\alpha) = \alpha$$

ou

$$\mathbb{P}(X > z_\alpha) = 1 - \alpha$$

— Médiane : tout quantile d'ordre 1/2

— Inverse de la fonction de répartition. Pour $u \in]0, 1[$, on pose

$$G(u) = \inf\{t, |F(t) \geqslant u\}$$

On a que $G(u) \leqslant t \Leftrightarrow F(t) \geqslant u$, cela donne une méthode de simulation.

Définition 12. ***Une variable aléatoire X est dite discrète*** *si elle est à valeurs dans un ensemble discret ou dénombrable. En particulier $S(X)$ est fini ou dénombrable.*

On pose $S(X) = \{a_i, i \in I\}$ où I est un ensemble fini ou dénombrable. Si on pose

$$p_i = \mathbb{P}(X = a_i), i \in I$$

La loi de X est

$$P_X = \sum_{i \in I} p_i \delta_{a_i}$$

Rq : Si (X, Y) est un couple de v.a discrète on obtient la loi de X par la formule

$$P_X(x) = \sum_{y \in S(Y)} \mathbb{P}_{(X,Y)}(x, y), \forall x \in S(X)$$

— Les v.a constantes. Si $X = c$ presque sûrement alors $S(X) = \{c\}$ et

$$\mathbb{P}_X = \delta_c, \quad \mathbb{P}(X = c) = 1$$

— Si $S(X) = \{a_1, \ldots, a_n\}$ et X prend la valeur a_i avec probabilité p_i on a

$$\mathbb{P}_X = \sum_{i=1}^{n} p_i \delta_{a_i}, \quad \mathbb{P}(X = a_i) = p_i$$

On dit que X suit la loi uniforme (ou équirépartie) sur $\{a_1, \ldots, a_n\}$ si $p_i = \frac{1}{n}$ pour tout $i \in \{1, \ldots, n\}$ dans ce cas

$$\mathbb{P}_X = \sum_{i=1}^{n} \frac{1}{n} \delta_{a_i}, \quad \mathbb{P}(X = a_i) = \frac{1}{n}$$

(exemple du dé à six face)

— Si X modélise un succés (obtenir 1) ou un échec (obtenir 0). Son support est $S(X) = \{0, 1\}$ et sa loi est

$$\mathbb{P}_X = p\delta_1 + (1 - p)\delta_0, \quad \mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0)$$

où p est la probabilité de succés. Il s'agit de la loi de **Bernoulli** $X \sim \mathcal{B}(p)$

— Si la v.a X modélise le nombre de succès dans une suite de n expériences identiques et indépendantes avec chacune une probabilité p de succès alors $S(X) = \{0, \ldots, n\}$ et on a

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

Dans ce cas on dit que X suit une loi **binomiale de paramètre** $\underline{p}$, la probabilité de succès, et $\underline{n}$, le nombre d'expérience, : $X \sim \mathcal{B}(n, p)$.

— La variable aléatoire X qui modélise le rang du premier succès dans une suite d'expériences indépendantes et identiques a pour support $S(X) = \mathbb{N}^*$ et on a

$$\mathbb{P}_X = \sum_{k=1}^{\infty} (1 - p)^{k-1} p \, \delta_k, \quad \mathbb{P}(X = k) = (1 - p)^{k-1} p$$

Dans ce cas on dit que X suit une **loi géometrique** de paramètre $\underline{p}$: $X \sim \mathcal{G}(p)$

— On dit que la variable aléatoire X suit une **loi de Poisson** de paramètre $\underline{\lambda > 0}$: $X \sim \mathcal{P}(\lambda)$ si $S(X) = \mathbb{N}$ et

$$\mathbb{P}_X = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k, \quad \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Une telle loi sert à modéliser des petits effectifs

Proposition 5. *Soit X une v.a discrète de support $S(X) = \{a_i, i \in I\}$ avec les probabilités $p_i, i \in I$ associées. La fonction de répartition est constante par morceaux avec des sauts de taille p_i en a_i*

$$F_X(t) = \sum_{j \leqslant i} p_j, \quad \text{si } t \in [a_i, a_{i+1}[, i \in I$$

Définition 13. *Une v.a X est dite à densité f si pour tout borélien A*

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

— Dans ce cas sa loi $\mathbb{P}_X$ est absolument continue par rapport à la mesure de Lebesgue et sa **dérivée de Radon Nykodim** est $\underline{f}$

— On a nécessairement que $f \geqslant 0$

— $\int_{\mathbb{R}} f(x) dx = 1$

— Pour un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ on a pour tout borélien A de $\mathbb{R}^d$:

$$\mathbb{P}_X(A) = \int_A f(x_1, \ldots, x_d) dx_1 \ldots dx_d$$

— Loi uniforme sur $[0, 1]$: $X \sim \mathcal{U}([0, 1])$ la densité

$$f(x) = \mathbf{1}_{[0,1]}(x)$$

— Loi uniforme sur $[a, b]$: $X \sim \mathcal{U}([a, b])$ la densité

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

— Loi exponentielle de paramètre $\lambda > 0$: $X \sim \mathcal{E}(\lambda)$ est de densité

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

— Loi normale centrée réduite $X \sim \mathcal{N}(0, 1)$ est de densité

$$f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

— Loi normale de moyenne m et de variance σ^2 : $X \sim \mathcal{N}(m, \sigma^2)$ est de densité

$$f(x) = \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

— Loi de Cauchy de paramètre $a > 0$ $X \sim \mathcal{C}(a)$ est de densité

$$f(x) = \frac{a}{\pi(a^2 + x^2)}$$

— Loi Gamma de paramètres $n \in \mathbb{N}^*, \lambda > 0$: $X \sim \gamma(n, \lambda)$ est de densité

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

Proposition 6. *Si X est une v.a de densité f, sa fonction de répartition F_X vérifie*

1. $\forall t \in \mathbb{R},\ F_X(t) = \int_{-\infty}^{t} f(x)dx$
2. *F_X est continue sur $\mathbb{R}$*
3. *Si f est continue au point t_0, alors F_X est dérivable en t_0 de dérivée $F'_X(t_0) = f(t_0)$*

— Rq : $\mathbb{P}(X = a) = \int_a^a f(x)dx = 0$. Cela implique que dans le calcul des probabilités, on peut considérer indifféremment des inégalités larges ou strictes.

— Si (X, Y) est un couple de densité $f(x, y)$, alors X est à densité et celle-ci s'obtient par

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

3) Espérance et Variance

Soit X une v.a (ou vecteur), on rappelle que la loi de X est décrite à travers la formule $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$, pour tout A borélien. On pose alors

$$\mathbb{E}[\mathbf{1}_A(X)] = \mathbb{P}_X(A) = \mathbb{E}[\phi(X)],$$

avec $\phi(x) = \mathbf{1}_A(x)$. On généralise cette formule pour toute fonction ϕ continue et bornée

Théorème 4. *Soit X une v.a, pour toute fonction continue bornée ϕ*

$$\mathbb{E}[\phi(X)] = \int \phi(X(\omega))d\mathbb{P}(\omega) = \int \phi(x)d\mathbb{P}_X(x)$$

Démonstration. On suppose d'abord $\varphi \geqslant 0$. Dans le cas ou φ est étagée on a que

$$\varphi(X) = \sum_i \alpha_i \mathbf{1}_{B_i}(X)$$

Ainsi

$$\begin{aligned}
E\varphi(X) &= \sum_i \alpha_i E\mathbf{1}_{B_i}(X) \\
&= \sum_i \alpha_i E\mathbf{1}_{B_i}(X) \\
&= \sum_i \alpha_i P(X \in B_i) \\
&= \sum_i \alpha_i P_X(B_i) \\
&= \int \sum_i \alpha_i \mathbf{1}_{B_i} dP_X
\end{aligned}$$

Sinon φ est limite croissante de fonctions étagées i.e

$$\varphi = \lim_n \varphi_n$$

Donc par convergence monotone 2 fois

$$\begin{aligned}
E\varphi(X) &= E \lim_n \varphi_n(X) \\
&= \lim_n E\varphi_n(X) \\
&= \lim_n \int \varphi_n dP_X \\
&= \int \lim_n \varphi_n dP_X \\
&= \int \varphi dP_X
\end{aligned}$$

Dans le cas général on décompose $\varphi = \varphi^+ - \varphi^-$. Puis φ est intégrable ssi $\varphi+$ et φ^- le sont. $\square$

- Dans le cas discret $\mathbb{E}[\phi(X)] = \sum_{x\in S(X)} \phi(x)\mathbb{P}[X = x]$
- Dans le cas continu $\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x)f(x)dx$
- Rq : X a pour densité f ssi $\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x)f(x)dx$ pour tout ϕ continue bornée
- Soit X une v.a à densité f et g une fonction bijective de $\mathbb{R}$ dans $\mathbb{R}$ de classe C^1. On pose $Y = g(X)$. Alors la densité de Y est

$$f_Y(y) = f_X(g^{-1}(y))|(g^{-1})'(y)|$$

- Si $Z \sim \mathcal{U}([-1, 1])$ alors $|Z| \sim \mathcal{U}([0, 1])$
- Si $X \sim \mathcal{N}(m, \sigma^2)$ alors

$$\frac{X-m}{\sigma} \sim \mathcal{N}(0,1)$$

- Si $X \sim \mathcal{N}(0, 1)$ quelle est la loi de $Y = X^2$
- Pour des vecteurs aléatoires on utilise le même type de calcul mais il faut utiliser un changement de variable en dimension supérieure.

Définition 14. *Une v.a X à un moment d'ordre $p \geqslant 1$ si*

$$\mathbb{E}(|X|^p) < +\infty$$

Si X a un moment d'ordre 1 on appelle espérance la quantité

$$\mathbb{E}[X]$$

Si X a un moment d'ordre 2 on appelle ***variance*** *la quantité*

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

On appelle ***écart type*** *de X la quantité*

$$\sigma_X = \sqrt{Var(X)}$$

- Inégalité de Markov

$$\mathbb{P}(|X| \geqslant t) \leqslant \frac{\mathbb{E}(|X|)}{t}$$

- Inégalité de Bienaymé-Tchebychev

$$\mathbb{P}(|X - \mathbb{E}(X)| \geqslant t) \leqslant \frac{Var(X)}{t^2}$$

- Comme l'espace est de masse finie si une v.a X admet un moment d'ordre $p \geqslant 1$ alors elle admet un moment d'ordre q pour tout q tel que $p \geqslant q \geqslant 1$.
- L'espérance est linéaire.
- $Var(aX) = a^2Var(X)$.
- $Var(X) = 0$ ssi X est constante

- L'espace vectoriel $L^p(\Omega, \mathcal{A}, \mathbb{P}) = \{X \ v.a, \mathbb{E}(|X|^p) < +\infty\}$ est **un espace de Banach**.
- $X \sim \mathcal{B}(p)$ alors $\mathbb{E}[X] = p$, $Var(X) = p(1-p)$
- $X \sim \mathcal{B}(n,p)$ alors $\mathbb{E}[X] = np$, $Var(X) = np(1-p)$
- $X \sim \mathcal{G}(p)$ alors $\mathbb{E}[X] = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$
- $X \sim \mathcal{P}(\lambda)$ alors $\mathbb{E}[X] = \lambda$, $Var(X) = \lambda$
- $X \sim \mathcal{U}([a,b])$ alors $\mathbb{E}[X] = \frac{b+a}{2}$, $Var(X) = \frac{(a-b)^2}{12}$
- $X \sim E(\lambda)$ alors $\mathbb{E}[X] = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$
- $X \sim C(a)$ alors $\mathbb{E}[|X|] = +\infty$
- $X \sim \mathcal{N}(m, \sigma^2)$ alors $\mathbb{E}(X) = m$, $Var(X) = \sigma^2$

4) Transformée exponentielle

Définition 15. ***La fonction caractéristique*** *d'une v.a X est la fonction définie par*

$$\phi_X(t) = \mathbb{E}[e^{itX}], \forall t \in \mathbb{R}$$

La fonction caractéristique d'un vecteur aléatoire *X à valeurs dans $\mathbb{R}^d$ est la fonction définie par*

$$\phi_X(u) = \mathbb{E}[e^{i<u,X>}], \forall u \in \mathbb{R}^d,$$

où $<,>$ désigne le produit scalaire euclidien sur $\mathbb{R}^d$.

La fonction caractéristique est continue bornée par 1 et vaut 1 en 0.

Théorème 5. *$\phi_X = \phi_Y$ ssi $P_X = P_Y$*

Démonstration. Pour cela on va utiliser la version fonctionnelle des classes monotones. On considère

$$\mathcal{H} = \{\phi \text{ borélienne bornée} | E(\phi(X)) = E(\phi(Y))\}$$

C'est un e.v monotone par convergence dominé. On a les égalités suivante pour tout x et a

$$\begin{aligned} \cos(\langle a, x\rangle) &= \frac{e^{i\langle a,x\rangle} + e^{-i\langle a,x\rangle}}{2} \\ \sin(\langle a, x\rangle) &= \frac{e^{i\langle a,x\rangle} - e^{-i\langle a,x\rangle}}{2i}. \end{aligned}$$

Les fonctions qui en découlent sont donc des éléments de $\mathcal{H}$. Donc $\mathcal{H}$ contient l'espace vectoriel $\mathcal{C}$ engendré par les fonctions

$$\{\cos(\langle a, x\rangle), \sin(\langle a, x\rangle), a \in \mathbb{R}\}.$$

L'e.v $\mathcal{C}$ est stable par multiplication (formule de trigo !), donc $\mathcal{H}$ contient les fonctiosn mesurables pour $\sigma(\mathcal{C})$. pour conclure, il suffit de montrer que

$$\mathcal{B}(\mathbb{R}^d) \subset \sigma(\mathcal{C}).$$

Considérons l'application

$$\begin{array}{ccc} \mathbb{R}^d & \to & \mathbb{R} \\ (x_i) & \mapsto & n\sin(\frac{x_i}{n}) \end{array}$$

qui est $\sigma(\mathcal{C}))$ mesurable. Par limite la fonction $\pi_i : (x_i) \leftarrow x_i$ est alors $\sigma(\mathcal{C})$ mesurable. Ainsi soit B^i un borélien on a que

$$\pi_i^{-1}(B_i) = \mathbb{R} \times \ldots \times B_i \times \ldots \times \mathbb{R} \in \sigma(\mathcal{C}).$$

Il s'agit d'un système de générateur de $\sigma(\mathbb{R}^d)$ et donc on a le résultat. □

Démonstration. Voilà une preuve alternative en dimension 1. Soit X une variable aléatoire réelle. On note P_X sa loi, Φ_X sa fonction caractéristique et F_X sa fonction de répartition.

1. Pour tous $a \neq b \in \mathbb{R}$ et $T > 0$, on a

$$\begin{aligned}\frac{1}{2\pi}\int_{-T}^{T} & \frac{e^{-ita}-e^{-itb}}{it}\Phi_X(t)\,dt \\ &= \int_{\mathbb{R}}\left(\frac{1}{\pi}\int_0^T \frac{\sin\big(t(x-a)\big)}{t}dt - \frac{1}{\pi}\int_0^T \frac{\sin\big(t(x-b)\big)}{t}dt\right) P_X(dx).\end{aligned}$$

2. Par une IPP on montre que $\lim_{T\to\infty}\int_0^T \frac{\sin t}{t}dt = \int_0^{+\infty}\frac{1-\cos t}{t^2}\,dt \in]0,+\infty[$. On admettra pour la suite que cette limite vaut $\frac{\pi}{2}$. On en déduit que $x \in \mathbb{R}$,

$$\lim_{T\to\infty}\int_0^T \frac{\sin(tx)}{t}\,dt = \frac{\pi}{2}\text{signe}(x),$$

où $\text{signe}(x) = 1$ si $x > 0$, $\text{signe}(x) = -1$ si $x < 0$ et $\text{signe}(0) = 0$.

3. Pour $a < b$ on a donc

$$\lim_{T\to+\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\Phi_X(t)\,dt = P_X\big(]a,b[\big) + \frac{P_X(\{a\})+P_X(\{b\})}{2}.$$

4. Ainsi si F_X est continue en a et b, on a $P_X(\{a\}) = P_X(\{b\}) = 0$ et par suite

$$F_X(b) - F_X(a) = \lim_{T\to+\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\Phi_X(t)\,dt.$$

Ainsi en tout point de continuité de F_X et F_Y on a que

$$F_X(b) - F_X(a) = F_Y(b) - F_Y(a)$$

Ainsi $F_X = F_Y$ en tout point de continuité de F_X et F_Y. Les points de discontinuité de F_X ou F_Y sont en nombre dénombrable donc si t_0 est l'un de ces points on peut trouver une suite (t_n) avec $t_n > t_0$ pour tout n et (t_n) sont des points de continuité de F_X et F_Y. On a donc par continuité à droite

$$F_X(t_0) = \lim F_X(t_n) = \lim F_Y(t_n) = F_Y(t_0).$$

Donc $F_X = F_Y$ sur tout $\mathbb{R}$ et donc $P_X = P_Y$.

□

Théorème 6. *Soit X v.a. à valeurs dans $\mathbb{R}^d$. On note $\lambda_{\mathbb{R}^d}$ la mesure de Lebesgue sur $\mathbb{R}^d$. Si $\Phi_X \in L^1(\mathbb{R}^d, \lambda_{\mathbb{R}^d})$ alors $P_X \ll \lambda_{\mathbb{R}^d}$ et sa densité est donnée par*

$$f_X(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle\xi,x\rangle}\Phi_X(\xi)d\xi.$$

Démonstration. On va considérer l'ensemble des fonctions continues à support compact et montrer que pour tout ϕ dans cet ensemble

$$\int \phi(t)dP_X(t) = \int \phi(t)f_X(t)dt$$

ce qui par la version fonctionnelle des c.m suffit à conclure. Introduisons $\gamma_{n,t}$ la densité d'une gaussienne de variance $\frac{1}{n^2}$ centrée en t i.e

$$\gamma_{n,t}(y) = \frac{1}{(\sqrt{2\pi}/n)^d} e^{\frac{-|y-t|^2}{2/n^2}}$$

On a que (changement de varaiable plus convergence dominée)

$$\phi(t) = \lim_n \int \phi(y)\gamma_{n,t}(y)dy$$

Ainsi

$$\begin{aligned}
\int \phi(t)dP_X(t) &= \int \lim_n \int \phi(y)\gamma_{n,t}(y)dydP_X(t) \\
&= \lim_n \int \int \phi(y)\gamma_{n,t}(y)dydP_X(t) \quad \left(\left|\int \phi(y)\gamma_{n,t}(y)dy\right| \leqslant \|\phi\|_\infty\right) \\
&= \lim_n \int \phi(y) \int \gamma_{n,t}(y)dP_X(t)dy \quad \text{(Fubini avec la même raison que ci-dessus)}
\end{aligned}$$

A ce stade on veut à nouveau intervertir limite et intégrale. Pour cela on utilise l'inversion de la transformée de Fourier d'une gaussienne c'est à dire

$$\gamma_{n,t}(y) = \frac{1}{(2\pi)^d} \int e^{i\langle\zeta,t-y\rangle} e^{-|\zeta|^2/2n^2} d\zeta$$

(ce résultat est une généralisation multi d de la fonction caractéristique de la gaussienne $\mathcal{N}(0,1)$).

Maintenant on peut observer que

$$\begin{aligned}
\int \gamma_{n,t}(y)dP_X(t) &= \int \frac{1}{(2\pi)^d} \int e^{i\langle\zeta,t-y\rangle} e^{-|\zeta|^2/2n^2} d\zeta dP_X(t) \\
&= \int \frac{1}{(2\pi)^d} \int e^{i\langle\zeta,t-y\rangle} e^{-|\zeta|^2/2n^2} dP_X(t)d\zeta \quad (Fubini) \\
&= \frac{1}{(2\pi)^d} \int e^{-i\langle\zeta,y\rangle} e^{-|\zeta|^2/2n^2} \Phi_X(\zeta)d\zeta
\end{aligned}$$

Par suite on a que

$$\left|\phi(y) \int \gamma_{n,t}(y)dP_X(t)\right| \leqslant \frac{1}{(2\pi)^d}|\phi(y)| \int |\Phi_X(\zeta)|d\zeta$$

Cette borne est indépendante de n et on a que Φ_X est supposé L^1 et ϕ est à support compact donc on peut appliquer le théorème de convergence dominée. On a

$$\begin{aligned}
\int \phi(t) dP_X(t) &= \int \phi(y) \lim_n \int \gamma_{n,t} dP_X(t) dy \\
&= \int \phi(y) \lim_n \frac{1}{(2\pi)^d} \int e^{-i\langle \zeta, y\rangle} e^{-|\zeta|^2/2n^2} \Phi_X(\zeta) d\zeta dy \\
&= \int \phi(y) \frac{1}{(2\pi)^d} \int e^{-i\langle \zeta, y\rangle} \Phi_X(\zeta) d\zeta d
\end{aligned}$$

d'où le résultat □

— $X \sim \mathcal{B}(p)$ alors $\phi_X(t) = 1 - p + pe^{it}$
— $X \sim \mathcal{B}(n,p)$ alors $\phi_X(t) = (1 - p + pe^{it})^n$
— $X \sim \mathcal{P}(\lambda)$ alors $\phi_X(t) = exp(\lambda(e^{it} - 1))$
— $X \sim \mathcal{U}([a,b])$ alors $\phi_X(t) = \frac{e^{ibt} - e^{iat}}{(b-a)it}$
— $X \sim \mathcal{E}(\lambda)$ alors $\phi_X(t) = \frac{\lambda}{\lambda - it}$
— $X \sim C(a)$ alors $\phi_X(t) = exp(-a|t|)$
— $X \sim \mathcal{N}(m, \sigma^2)$ alors $\phi_X(t) = exp(imt - \frac{\sigma^2 t^2}{2})$

Proposition 7. *Si une v.a X admet un moment d'ordre p alors sa fonction caractéristique est dérivable p fois et on a*

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

Une réciproque partielle est si ϕ_X est $p \geqslant 2$ fois dérivable en 0 alors X admet des moments d'ordre $2[p/2]$

Démonstration. Pour montrer cela on utilise le théorème de continuité et de dérivation sous le signe intégrale.

Pour la réciproque, faisons le cas $n = 2$ on suppose donc que $\Phi_X(t)$ est 2 fois dérivable en 0. On peut donc écrire

$$\begin{aligned}
\Phi_X(t) &= 1 + t\Phi_X'(0) + \frac{t^2}{2}\Phi_X''(0) + \circ(t^2) \\
\Phi_X(-t) &= 1 - t\Phi_X'(0) + \frac{t^2}{2}\Phi_X''(0) + \circ(t^2)
\end{aligned}$$

Donc on a

$$\Phi_X(t) + \Phi_X(-t) = E(2\cos(Xt)) = 2 + t^2\Phi_X''(0) + \circ(t^2)$$

Ainsi on a que pour t suffisamment petit

$$E(2 - \frac{2\cos(Xt)}{t^2}) = -\Phi_X''(0) + \circ(1)$$

On peut alors appliquer le Lemme de Fatou avec une suite (t_n) qui tend vers 0.

Proposition 8. *Soit* $(E, \mathcal{A}, \mu)$ *un espace mesuré et soit* (f_n) *une suite de fonctions positives alors*

$$\int \liminf f_n d\mu \leqslant \liminf \int f_n d\mu$$

On a alors

$$E\left[\liminf \frac{2-2\cos(Xt_n)}{t_n^2}\right] \leqslant \liminf E\left[\frac{2\cos(Xt_n)-2}{t_n^2}\right] = -\Phi_X''(0)$$

Or on a que

$$\liminf \frac{2-2\cos(Xt_n)}{t_n^2} = X^2$$

et donc

$$EX^2 \leqslant -\Phi_X''(0)$$

□

— Fonction génératrice des moments pour une v.a X tel que $S(X) \subset \mathbb{N}$ et $p_k = \mathbb{P}(X = k)$

$$G_X(t) = \mathbb{E}[t^X] = \sum_k p_k t^k$$

Cette fonction est C^∞ sur $[0, 1[$ et dérivable à l'ordre p en 1 si $\mathbb{E}[X^p] < +\infty$

$$G_X^{(k)}(0) = k! p_k, k \in \mathbb{N}$$

De plus si le moment d'ordre 1 existe $G_X'(1) = \mathbb{E}(X)$

— Par unicité du DSE en 0 la fonction génératrice des moments caractérise la loi.

Transformée de Laplace. Pour une v.a X, on appelle transformée de Laplace de X

$$L_X(t) = \mathbb{E}[e^{\langle t, X\rangle}]$$

Théorème 7. *Soient* X, Y *deux v.a. à valeurs dans* $\mathbb{R}^d$*. Soit* $\epsilon > 0$*. Si pour tout* s *tel que* $\|s\| \leqslant \epsilon$ *on a* $L_X(s) = L_Y(s) < +\infty$*, alors* $P_X = P_Y$*.*

Démonstration. Traitons le cas de la dimension $d = 1$. On peut remarquer que les fonctions

$$s \mapsto E(e^{sX}), \quad s \mapsto E(e^{sY})$$

sont définies pour tout complexe s tels que $|Re(s)| < \epsilon$.

Considérons s_0 tel que $Re(s_0) = 0$ et s tel que $|s| < \epsilon$. En remarquant que $|e^{s_0 X}| = 0$ et $|e^{sX}| = |e^{|\Re(s)X|}|$ est inégrable, on a en utilisant Fubini

$$E[e^{(s_0+s)X}] = \sum_n \frac{s^n}{n!} E(e^{s_0 X} X^n).$$

Il s'agit donc d'un développement en série entière valable sur $\mathring{B}(0, \epsilon)$. On a donc que $s \mapsto E(e^{sX})$ est analytique sur $\bigcup_{s_0 \in i\mathbb{R}} \mathring{B}(s_0, \epsilon) = \{s \in \mathbb{C} : |Re(s)| < \epsilon\}$. On a également que $s \mapsto E(e^{sY})$ est analytique. On en déduit que $s \mapsto E(e^{sY})$ et $s \mapsto E(e^{sX})$ coincident sur $\{s \in \mathbb{C} : |Re(s)| < \epsilon\}$ et donc pour $s = it$ avec $t \in \mathbb{R}$ on a que $\Phi_X = \Phi_Y$. □

5) Indépendance

Définition 16. — *Deux évènements A et B d'une tribu $\mathcal{A}$ sont indépendants si*

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

— *Deux variables aléatoires sont dites indépendantes si pour tous boréliens A et B on a*

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

— *Deux tribus $\mathcal{F}$ et $\mathcal{G}$ sont dites indépendantes si pour tous événements $F \in \mathcal{F}$ et $G \in \mathcal{G}$, F et G sont indépendants.*
On voit que X indépendantes de Y ssi $\sigma(X)$ indépendante de $\sigma(Y)$

Proposition 9. *Deux v.a X et Y sont indépendantes ssi la loi $\mathbb{P}_{X,Y}$ est la loi produit des marginales $\mathbb{P}_X$ et $\mathbb{P}_Y$ i.e*

$$\mathbb{P}_{X,Y} = \mathbb{P}_X \otimes \mathbb{P}_Y$$

— Dans le cas discret cela signifie que pour tout $x \in S(X)$ et pour tout $y \in S(Y)$ on a

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

— Dans le cas continu

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Théorème 8. *Des v.a. $X_1, \ldots, X_n$ sont indépendants si et seulement si*

$$P_{(X_1,\ldots,X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}.$$

Dans ce cas :

— *$\forall f_i$ mesurables positives : $E\big(\prod_{i=1}^{n} f_i(X_i)\big) = \prod_{i=1}^{n} E f_i(X_i)$.*

— *Si f_i mesurable de signe quelconque, l'égalité précédente est vérifiée si*

$$E\Big(\prod_{i=1}^{n} |f_i(X_i)|\Big) = \prod_{i=1}^{n} E|f_i(X_i)| < \infty.$$

Démonstration. On vérifie l'égalité des lois $P_{(X_1,\ldots,X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}$ sur les pavés et ensuite on utilise le théorème de caractérisation des mesures finies. □

Théorème 9. *Deux v.a X et Y sont indépendantes ssi*

$$\phi_{X,Y}(s, t) = \phi_X(s)\phi_Y(t)$$

pour tout $s, t \in \mathbb{R}$.

Exemple Soit X et Y deux v.a $\mathcal{N}(0,1)$ alors $U = X+Y$ et $V = X-Y$ sont indépendantes. Pour cela utilisons deux méthodes.

La première utilise la fonction caractéristique. Soit $(u,v) \in \mathbb{R}^2$ on a

$$\begin{aligned}
\phi_{(U,V)}(u,v) &= E(e^{i(uU+vV)}) \\
&= E(e^{i((u+v)X+(u-v)Y)}) \\
&= E(e^{i((u+v)X)}E(e^{i((u-v)Y)} \quad \text{par indépendance} \\
&= e^{-(u+v)^2/2}e^{-(u-v)^2/2} \\
&= e^{-u^2}e^{-v^2}
\end{aligned}$$

La fonction caractéristique est à variables séparées donc les deux v.a sont indépendantes. Or la fonction $u \mapsto e^{-u^2}$ est la fonction caractéristique de $\mathcal{N}(0,2)$. On a donc que U et V ont même loi et que $U \sim \mathcal{N}(0,2)$.

Exemple Soit X et Y deux v.a $\mathcal{N}(0,1)$ alors il existe des unique v.a $R > 0$ et $\Theta \in]0,2\pi[$, telles que

$$X = R\cos(\Theta), \quad Y = R\sin(\Theta)$$

Alors R et Θ sont indépendantes. On note $\phi(r,\theta) = (r\cos(\theta); r\sin(\theta))$ le changement de coordonnée polaire qui est un isomorphisme de $\mathbb{R}_+^* \times]0,2\pi]$ dans $\mathbb{R}^2 \setminus (0,0)$. Soit $f(r,\theta) = h(r)g(\theta)$ avec h et g deux fonctions continues bornées, on a

$$\begin{aligned}
E(f(R,\Theta)) &= E(f \circ \phi^{(-1)}(X,Y)) \\
&= \int_{\mathbb{R}^2} f \circ \phi^{(-1)}(x,y)\frac{1}{2\pi}e^{-(x^2+y^2)/2}dxdy \\
&= \int_0^{\infty}\int_0^{2\pi} f(r,\theta)\frac{1}{2\pi}e^{-r^2/2}rdrd\theta \quad \text{changement de v.a } (r,\theta) = \phi^{(-1)}(x,y) \\
&= \int_0^{\infty} f(r)e^{-r^2/2}rdr \int_0^{2\pi}\frac{1}{2\pi}g(\theta)d\theta
\end{aligned}$$

On voit donc que R et Θ sont indépendantes et que la densité de R est $r \mapsto e^{-r^2/2}r\mathbf{1}_{\mathbb{R}+}$ et que Θ suit une loi uniforme sur $[0,2\pi]$.

Proposition 10. *Si X et Y sont deux v.a indépendantes alors la loi de $X+Y$ est donnée par $\mathbb{P}_X * \mathbb{P}_Y$.*

En particulier si X et Y ont pour densité f et g on a

$$f_{X+Y}(x) = \int f(y)g(x-y)dy$$

On a également

$$\phi_{X+Y}(.) = \phi_X(.)\phi_Y(.)$$

(Lois classiques dont les convolées sont très simples)

— Lois binomiales : pour $n \in \mathbb{N}^*$ et $p \in (0,1)$,

$$\mathcal{B}(n,p) := \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} \delta_k$$

On a $\mathcal{B}(n,p) * \mathcal{B}(n',p') = \mathcal{B}(n+n',p)$. En d'autres termes, si $X \sim \mathcal{B}(n,p)$ et $Y \sim \mathcal{B}(n',p')$ sont indépendantes alors :

$$X + Y \sim \mathcal{B}(n+n',p).$$

On appelle loi de Bernoulli avec probabilité de succès $p \in [0,1]$, la loi

$$b(p) := \mathcal{B}(1,p) = (1-p)\delta_0 + p\delta_1.$$

— Lois binomiales négatives : $\mathcal{B}(-m,p)$ avec $p \in (0,1)$ et $m \in \mathbb{N}^*$, définies par

$$\mathcal{B}(-m,p) := \sum_{k \geqslant 0} C_{m+k-1}^{m-1} \, p^m (1-p)^k \, \delta_k.$$

Pour $m, n \in \mathbb{N}^*$, on a $\mathcal{B}(-m,p) * \mathcal{B}(-n,p) = \mathcal{B}\big(-(m+n),p\big)$. On appelle loi géométrique $\mathcal{G}(p) := \mathcal{B}(-1,p) = \sum_{k \geqslant 0} p(1-p)^k \, \delta_k$.

— Loi de Poisson : $\mathcal{P}(\lambda) := \sum_{k \geqslant 0} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k$ avec $\lambda > 0$. Elle vérifie

$$\mathcal{P}(\lambda) * \mathcal{P}(\lambda') = \mathcal{P}(\lambda + \lambda').$$

— Lois $\gamma(t,a)$ avec $t > 0$ et $a > 0$ de densité $\frac{a^t}{\Gamma(t)} x^{t-1} e^{-ax} 1_{x>0}$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$. On rappelle que pour $t > 0$, $\Gamma(t) = \int_0^{+\infty} e^{-x} x^{t-1} dx$. Si $s, t > 0$,

$$\gamma(t,a) * \gamma(s,a) = \gamma(t+s,a).$$

On appelle loi exponentielle de paramètre $a > 0$ la mesure de probabilité $\mathcal{E}xp(a) := \gamma(1,a)$.

— Lois gaussiennes : $\mathcal{N}(m,\sigma^2), m \in \mathbb{R}, \sigma \geqslant 0$. Elles sont définies par $\mathcal{N}(m,0) := \delta_m$ et pour $\sigma > 0$

$$\mathcal{N}(m,\sigma^2)(dx) := e^{\frac{-(x-m)^2}{2\sigma^2}} \frac{dx}{\sigma\sqrt{2\pi}}, \; x \in \mathbb{R}.$$

On a $\mathcal{N}(m,\sigma^2) * \mathcal{N}(m',\sigma'^2) = \mathcal{N}(m+m',\sigma^2+\sigma'^2)$.

Définition 17. *Une suite de v.a (X_n) est dite indépendante si pour tout $i_1, \ldots, i_k$, $i_1 < i_2 < \ldots < i_k$ et pour tout $k \in \mathbb{N}$*

$$\mathbb{P}(X_{i_1} \in A_1, \ldots X_{i_1} \in A_k) = \mathbb{P}(X_{i_1} \in A_1) \ldots \mathbb{P}(X_{i_1} \in A_k),$$

pour tout $A_1, \ldots A_k$ dans la tribu.

En particulier pour le cas d'une famille finie $X_1, \ldots, X_N$, cette famille est dite indépendante si :

$$\mathbb{P}(X_1 \in A_1, \ldots X_N \in A_N) = \mathbb{P}(X_1 \in A_1) \ldots \mathbb{P}(X_N \in A_N),$$

pour tout $A_1, \ldots A_N$ dans la tribu

— Soit $X_1, \ldots, X_n$ une famille de v.a.i.i.d de loi de $\mathcal{B}(p)$ alors

$$X_1 + \ldots + X_n \sim \mathcal{B}(n,p)$$

6) Conditionnement

Définition 18. *Soit B un évènement de probabilité non nulle $\mathbb{P}(B) \neq 0$. Pour tout évènement A on définit la probabilité conditionnelle de A sachant B par*

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

— $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$
— L'application $\mathbb{P}(\cdot|B)$ définit une mesure $(\Omega, \mathcal{A})$
— Si A est indépendant de B alors $\mathbb{P}(A|B) = \mathbb{P}(A)$
— Probabilité totale :

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$$

— Deux joueurs A et B avec respectivement une fortune a et b joue à pile ou face avec où pile arrive avec probabilité p. Le joeur A gagne 1 euro que lui donne B si la pièce tombe sur pile et perd 1 euro qu'il donne à B si la pièce tombe sur face. On pose u_a la probabilité que A soit ruiné. On a

$$u_a = pu_{a+1} + (1-p)u_{a-1}$$

— Formule de Bayes :

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

Dans la pratique on utilise souvent la formule des probabilités totales pour calculer $\mathbb{P}(A)$

Proposition 11. *Soit $A_1, \dots, A_N$ une partition de Ω alors*

$$\mathbb{P}(A) = \sum_{i=1}^{N} \mathbb{P}(A|A_i)\mathbb{P}(A_i)$$

$$\mathbb{P}(A_i|A) = \frac{\mathbb{P}(A|A_i)\mathbb{P}(A_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|A_i)\mathbb{P}(A_i)}{\sum_{i=1}^{N} \mathbb{P}(A|A_i)\mathbb{P}(A_i)}$$

— Soient X et Y deux variables aléatoires. Le théorème de désintégration des mesures nous dit que

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

— Attention la quantité $\mathbb{P}(Y \in A|X = x)$ est une notation cela correspond à la densité de Radon Nykodym
— La famille $(\mathbb{P}(Y \in \cdot|X = x)_{x \in \mathbb{R}}$ s'appelle la famille des probabilités conditionnelles de Y sachant X.

- La loi conditionnelle de Y sachant X est noté $\mathbb{P}(Y \in \cdot|X)$
- Dans le cas discret on obtient facilement les probabilités conditionnelles. En particulier
$$\mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)}$$
On a alors
$$\mathbb{P}(Y \in \cdot|X) = \sum_{x \in S(X)} \mathbb{P}(Y = y|X = x)\mathbf{1}_{X=x}$$
- Dans le cas continu on parle de densité conditionnelle et on pose
$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}\mathbf{1}_{f_X(x)>0}$$
avec
$$f_X(x) = \int f(x, y)dy$$

7) Convergence stochastique

Définition 19. *Soit (X_n) une suite de v.a et X une v.a. On dit que (X_n) converge vers X*

— ***Presque sûrement p.s*** *si* $\mathbb{P}(\lim X_n = X) = 1$
on note
$$X_n \xrightarrow{p.s} X$$
— ***En norme*** L^p *si* $\lim\limits_{n\to+\infty} \mathbb{E}[|X_n - X|^p] = 0$
on note
$$X_n \xrightarrow{L^p} X$$
— ***En probabilité*** *si* $\forall \epsilon > 0, \lim\limits_{n\to+\infty} \mathbb{P}[|X_n - X| > \epsilon] = 0$
on note
$$X_n \xrightarrow{\mathbb{P}} X$$
— ***En loi*** *si pour toute fonction continue bornée f on a* $\lim\limits_{n\to+\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$

on note $X_n \xrightarrow{Loi} X$

On suppose (X_n) à valeurs dans un espace métrique

Théorème 10. *Les assertions suivantes sont équivalentes*

— *(X_n) converge en loi vers X*
— *pour toute fonction φ bornée et uniformément continue sur E,*
$$\lim_n \mathbb{E}\left[\varphi(X_n)\right] \;=\; \mathbb{E}\left[\varphi(X)\right]$$
— *Pour tout fermé F on a* $\limsup_n \mathbb{P}(X_n \in F) \leqslant \mathbb{P}(X \in F)$
— *Pour tout ouvert O on a* $\liminf_n \mathbb{P}(X_n \in O) \geqslant \mathbb{P}(X \in F)$
— *Pour tout borélien A dont la frontière ∂A vérifie $\mathbb{P}(X \in \partial A) = 0$ on a*
$$\lim_n \mathbb{P}\left(X_n \in A\right) \;=\; \mathbb{P}\left(X \in A\right).$$

Proposition 12. *Soient μ_n, μ des mesures de probabilité sur $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. Les assertions suivantes sont équivalentes :*

1. $\mu_n \xrightarrow{étroit} \mu$.
2. *Pour tout ouvert $G \subset \mathbb{R}^d$,* $\liminf\limits_{n\to\infty} \mu_n(G) \geqslant \mu(G)$.
3. *Pour tout fermé $F \subset \mathbb{R}^d$,* $\limsup\limits_{n\to\infty} \mu_n(F) \leqslant \mu(F)$.
4. *Pour tout borélien $B \subset \mathbb{R}^d$ tel que $\mu(\partial B) = 0$ on a* $\lim\limits_{n\to+\infty} \mu_n(B) = \mu(B)$.

Démonstration. 2) $\Leftrightarrow$ 3) vient du fait que le complémentaire d'un fermé est un ouvert et vice versa.

Montrons que 1) $\Rightarrow$ 2). Soit O un ouvert. On considère la fonction
$$\phi_k(x) = \min(1, kd(x, O^c))$$

qui est une fonction continue bornée. Ainsi $\phi_k \leqslant \mathbf{1}_O$ et $\phi_k \to_k \mathbf{1}_O$. On a donc pour tout n

$$\mu_n(O) \geqslant \int \phi_k d\mu_n$$

et donc

$$\liminf \mu_n(O) \geqslant \liminf \int \phi_k d\mu_n = \int \phi_k d\mu$$

Ensuite par convergence dominée en faisant tendre k vers l'infini on a

$$\liminf \mu_n(O) \geqslant \liminf \int \phi_k d\mu_n = \int \mathbf{1}_O d\mu = \mu(O)$$

Montrons que 2) et 3) implique 4). Si $\mu(\partial B) = 0$ alors

$$\mu(B) = \mu(\mathring{B}) = \mu(\bar{B})$$

donc

$$\begin{aligned}
\liminf \mu_n(B) &\geqslant \liminf \mu_n(\mathring{B}) \geqslant \mu(\mathring{B}) = \mu(B) && (1)\\
\limsup \mu_n(B) &\leqslant \limsup \mu_n(\bar{B}) \leqslant \mu(\bar{B}) = \mu(B) && (2)
\end{aligned}$$

Donc

$$\liminf \mu_n(B) = \limsup \mu_n(B) = \lim \mu_n(B) = \mu(B)$$

Montrons enfin que 4) $\Rightarrow$ 1). Soit $\phi \in C_b(\mathbb{R}^d)$. On peut supposer quitte à considérer la partie positive puis ensuite négative que $\phi \geqslant 0$. On a que

$$\begin{aligned}
\int_{\mathbb{R}^d} \phi d\mu_n &= \int_{\mathbb{R}^d} \left(\int_0^{\|\phi\|_\infty} \mathbf{1}_{y \leqslant \phi(x)} dy \right) d\mu_n(x)\\
&= \int_0^{\|\phi\|_\infty} \mu_n(\{x : \phi(x) \geqslant y\}) dy \quad (Fubini)
\end{aligned}$$

A ce stade considérons le fermé $F_y = \{x : \phi(x) \geqslant y\}$ qui contient l'ouvert $\{x : \phi(x) > y\}$. On a donc

$$\partial F_y \subset F_y \setminus \{x : \phi(x) > y\} = \phi^{-1}(\{y\}).$$

On a donc que $\mu(\partial F_y) = 0$ pour presque tout y. En effet

$$\{y : \mu(\partial F_y) > 0\} \subset \{y : \mu(\phi^{-1}(\{y\})) > 0\}$$

qui sont les atomes de la mesure image de μ par ϕ. Donc $\{y : \mu(\partial F_y) > 0\}$ est au plus dénombrable. Ainsi pour presque tout y

$$\mu_n(F_y) \to \mu(F_y)$$

et donc par convergence dominée

$$\int_0^{\|\phi\|_\infty} \mu_n(\{x : \phi(x) \geqslant y\}) dy \to \int_0^{\|\phi\|_\infty} \mu(\{x : \phi(x) \geqslant y\}) dy = \int_{\mathbb{R}^d} \phi d\mu$$

et le résultat en découle. $\square$

Pour une v.a on note F_X sa fonction de répartition et ϕ_X sa fonction caractéristique

Théorème 11. *(X_n) converge en loi vers X si et seulement si*

$$F_{X_n}(t) \to F_X(t)$$

en tout point de continuité de F_X i.e en tout t tel que $\mathbb{P}(X = t) = 0$

Démonstration. L'implication $\Rightarrow$ est une conséquence du théorème du portemanteau. Pour la réciproque on va montrer le point 2) du théorème du portemanteau. Ainsi considérons un ouvert $O =]a, b[$ (le cas général s'en déduit par union dénombrable). Comme le nombre de points de discontinuité de F_X est dénombrable on peut trouver une suite (a_k) décroissante et convergeante vers a telle que F_X est continue en a_k pour tout k et une suite (b_k) croissante qui converge vers b telle que F_X est continue en b_k. Par continuité à droite et à gauche, on a que

$$\begin{aligned} \limsup F_{X_n}(a) &\leqslant \limsup F_{X_n}(a_k) = F_X(a_k) \to F_X(a) \\ \liminf F_{X_n}(b^-) &\geqslant \liminf F_{X_n}(b_k) = F_X(b_k) \to F_X(b-) \end{aligned}$$

Ainsi

$$\begin{aligned} \liminf P_{X_n}(]a, b[) &= \liminf F_{X_n}(b^-) - F_{X_n}(a) \\ &\geqslant F_X(b^-) - F_X(a) = P_X(]a, b[) \end{aligned}$$

et donc on a le résultat attendu. □

Théorème 12. (Lévy).

1. *Si $X_n \xrightarrow{\mathcal{L}} X$ alors $\forall \xi \in \mathbb{R}^d$, $\lim_{n\to\infty} \Phi_{X_n}(\xi) = \Phi_X(\xi)$.*
2. *S'il existe une fonction $\Phi : \mathbb{R}^d \to \mathbb{C}$ continue en 0 telle que pour tout $\xi \in \mathbb{R}^d$,*

$$\lim_{n\to\infty} \Phi_{X_n}(\xi) = \Phi(\xi),$$

alors il existe une mesure de probabilité μ sur $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$ telle que $\Phi = \hat{\mu}$. De plus $\mu_n \xrightarrow{\text{étroit}} \mu$ et si X est une v.a. de loi μ on a $X_n \xrightarrow{\mathcal{L}} X$.

Démonstration. Le point 1) est évident en considérant la partie réelle et imaginaire de $x \mapsto e^{ix}$.

Pour le point 2) il suffit de montrer que μ_n est tendue. En effet le théorème de Prokhorov nous donnera l'existence d'une sous suite convergente vers une mesure μ. Ainsi Φ est n :'ecessairement la fonction caractéristique de μ. De cette manière toutes les valeurs d'adhérence seront égale à μ (car elles auront même fonction caractéristique).

Pour montrer la tension, on va montrer que $P_{X_n}([-M, M])$ converge uniformément vers 1 lorsque M tend vers l'infini. Sur l'ensemble $[-M, M]^c$ la quantité

$$1 - \frac{\sin(X/M)}{X/M}$$

est minorée par une certaine constante α indépendante de M. On a donc

$$\begin{aligned} P_{X_n}([-M,M]^c) &= E\left(\mathbf{1}_{[-M,M]^c}(X_n)\right) \\ &\leqslant \frac{1}{\alpha}E\left(1-\frac{\sin(X_n/M)}{X_n/M}\right) \\ &\leqslant \frac{1}{\alpha}E\left(\frac{M}{2}\int_{-1/M}^{1/M}(1-e^{iX_nt})dt\right) \\ &\leqslant \frac{1}{\alpha}\frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi_{X_n}(t))dt \end{aligned}$$

Lorsque n tend vers l'infini, par convergence dominée, on a

$$\frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi_{X_n}(t)) \to \frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi(t))$$

Donc

$$\limsup_n P_{X_n}([-M,M]^c) \leqslant \frac{1}{\alpha}\frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi(t))$$

Ainsi il existe N telle que

$$\sup_{n\geqslant N} P_{X_n}([-M,M]^c) \leqslant \frac{1}{\alpha}\frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi(t))$$

Or

$$\lim_M \frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi(t)) = 1-\Phi(0) = 0$$

donc $P_{X_n}([-M,M]^c)$ converge uniformmment vers 0 et le résultat est obtenu. □

Théorème 13. (Lévy)(Version faible). $X_n \xrightarrow{\mathcal{L}} X$ *ssi* $\forall \xi \in \mathbb{R}^d$, $\lim_{n\to\infty}\Phi_{X_n}(\xi) = \Phi_X(\xi)$.

Démonstration. Naturellement c'est une conséquence du théorème précédent mais on peut directement le résultat sans faire apppel à la tension. Le sens $\Rightarrow$ est immédiat car $x \mapsto e^{itx}$ est continue et bornée (au pire on prend la partie imaginaire et réelle). Pour le sens $\Leftarrow$, on considère une fonction $\phi \in L^1$ on considère

$$\xi \mapsto f(\xi) = \int e^{it\xi}\phi(t)dt$$

qui est une fonction C_0. Alors

$$\begin{aligned} E(f(X_n)) &= E(\int e^{itX_n}\phi(t)dt) \\ &= \int E(e^{itX_n}\phi(t)) \quad (Fubini) \\ &= \int \phi(t)\Phi_{X_n}(t)dt \\ &\to_n \int \phi(t)\Phi_X(t) = E(f(X)) \quad \text{(convergence dominée)} \end{aligned}$$

En notant C_c^∞ les fonctions C^∞ à support compact on sait que C_c^∞ est dense dans C_0. Or $C_c^\infty \subset \mathcal{S}$ où $\mathcal{S} = \{f \in \mathcal{C}^\infty(\mathbb{R}^d) \mid \forall(\alpha,\beta)\ \|f\|_{\alpha,\beta} < +\infty\}$ désigne l'espace de Schwarz avec $\|f\|_{\alpha,\beta} = \|x^\alpha D^\beta f\|_\infty$. Or la transformée de Fourier est un automorphisme sur l'espace de Schwarz donc pour tout $f \in C_c^\infty$ il existe $\phi \in \mathcal{S}$ telle que $\xi \mapsto f(\xi) = \int e^{it\xi}\phi(t)dt$. Ainsi pour tout $f \in C_c^\infty$

$$\lim_n Ef(X_n) = Ef(X).$$

En utilisant une proposition ci dessus on obtient le résultat car C_c^∞ est dense dans C_0 pour la norme $\|.\|_\infty$. □

Rappel des convergences classiques en termes probabilistes

Théorème 14. — *Soit (X_n) une suite croissante de v.a positives et $\lim_n X_n = X$ alors*

$$\lim_n \mathbb{E}[X_n] = \mathbb{E}[X]$$

— *Soit (X_n) une suite croissante de v.a positives*

$$\mathbb{E}[\liminf_n X_n] \leqslant \liminf_n \mathbb{E}[X_n]$$

— *Soit (X_n) une suite de v.a telle que X_n converge p.s vers X. Soit Y tel que $\mathbb{E}[|Y|] < \infty$ et $|X_n| < Y|$ alors*

$$\lim_n \mathbb{E}[X_n] = \mathbb{E}[\lim_n X_n]$$

On a les liens suivants

— Convergence ps $\Longrightarrow$ Convergence en probabilité
— Convergence L^p $\Longrightarrow$ Convergence L^1 $\Longrightarrow$ Convergence en probabilité
— Toute Convergence $\Longrightarrow$ Convergence en loi
— Convergence p.s + domination $\Longrightarrow$ Convergence L^1
— Convergence L^1 $\Longrightarrow$ Convergence p.s pour une sous-suite

Montrer des convergences presque sûres ? Le Lemme de Borel Canteli

Proposition 13. *Soit $(A_n)_{n\in\mathbb{N}}$ une suite dévènements*
— *Si $\sum \mathbb{P}(A_n) < +\infty$ alors*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) := \mathbb{P}(\bigcap_{n\geqslant 0}\bigcup_{k\geqslant n} A_k) = 0$$

En d'autres termes seuls un nombre fini d'évènements A_n sont réalisés.
— *Si $\sum \mathbb{P}(A_n) = +\infty$ et les A_n sont indépendants alors*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) := \mathbb{P}(\bigcap_{n\geqslant 0}\bigcup_{k\geqslant n} A_k) = 1$$

En d'autres termes une infinité d'évènements A_n sont réalisés.

Proposition 14. *(X_n) converge en probabilité vers X si et seulement si de toute suite extraite on peut extraire une sous suite qui converge presque sûrement vers X*

— La distance $d(X,Y) = \mathbb{E}[|X-Y| \wedge 1]$ métrise la convergence en probabilité
— La topologie de la convergence presque sûre n'est pas métrisable

Définition 20. *Soit $L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$ l'ensemble des v.a de $(\Omega, \mathcal{A}, P)$ dans $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ quotienté par la relation d'égalité p.s.. L'application $d(X,Y) := E\min\big(1, |X-Y|\big)$ est une distance sur cet espace et*

$$X_n \xrightarrow{P} X \quad \Leftrightarrow \quad \lim_{n\to\infty} d(X_n, X) = 0.$$

On dit que la convergence en probabilité est métrisable.

Démonstration. Il est facile de vérifier que d définie une distance. Montrons l'équivalence. De droite à gauche, on a que pour tout $1 > \varepsilon > 0$

$$E(\min(1, |X-Y|)) \geqslant \varepsilon P(\min(1, |X-Y|) \geqslant \varepsilon) = \varepsilon P(|X-Y| \geqslant \varepsilon)$$

Ainsi si $d(X, X_n) \longrightarrow 0$ alors $P(|X - X_n| \geqslant \varepsilon) \longrightarrow 0$ pour tout ε.

Pour la réciproque on observe que

$$E(\min(1, |X-X_n|)) \leqslant \varepsilon + P(\min(1, |X-X_n|) \geqslant \varepsilon)$$

et donc on a que

$$\limsup E(\min(1, |X-X_n|)) \leqslant \varepsilon$$

pour tout ε. Ainsi

$$\limsup E(\min(1, |X-X_n|)) = 0$$

et le résultat en découle. □

Proposition 15. *Les assertions suivantes sont équivalentes*

— $X_n \xrightarrow{P} X$,
— *de toute sous-suite $(X_{n_k})_{k\geq 0}$ on peut extraire une sous-suite $\big(X_{n_{k_j}}\big)_{j\geq 0}$ telle que $X_{n_{k_j}} \xrightarrow[j\to\infty]{p.s.} X$.*

Démonstration. De la première assertion à la deuxième on que si $X_n \xrightarrow{P} X$ alors $d(X_n, X) \to 0$ ainsi pour toute sous suite $d(X_{n_k}, X) \to 0$ et on peut alors extraire de cette sous suite une sous suite convergente presque sûrement par la proposition ci-dessus.

Dans l'autre sens, supposons par l'absurde que $d(X, X_n)$ ne converge pas vers 0, il existe donc $\varepsilon > 0$ et une sous suite telle que $d(X_{n_k} X) \geqslant \varepsilon$. De cette sous suite on peut extraire une sous suite qui converge p.s donc en proba et donc $d(X_{n_{k_j}}, X)$ converge vers 0 mais on a aussi $d(X_{n_{k_j}} X) \geqslant \varepsilon$ ce qui est impossible. □

Des exemples et des contre-exemples

- Pour tout n on considère la variable aléatoire X_n qui prend les valeurs $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$ avec la probabilité uniforme alors (X_n) converge en loi vers $X \sim \mathcal{U}([0,1])$.
- Une suite de variables aléatoires (X_n) telle que $X_n \sim \mathcal{B}(p_n)$ pour tout $n \in \mathbb{N}$ (fournit des exemples et des contrexemples...)
- Soit (X_n) une suite de v.a.i.i.d telle $X_0 \sim \mathcal{B}(\frac{1}{2})$, alors

$$\sum_{n=0}^{\infty} \frac{X_n}{2^n}$$

converge presque sûrement vers une v.a $X \sim \mathcal{U}([0,1])$. Réciproque ?
- (X_n) telle que $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n}$ et $\mathbb{P}[X_n = a_n] = \frac{1}{n}$
- (X_n) converge en loi vers X et (Y_n) converge en loi vers Y n'implique pas que le couple converge en loi vers (X, Y).

Proposition 16. *(X_n) converge en loi vers une constance c alors (X_n) converge en probabilité vers c*

- Si (X_n) converge en loi vers X et (Y_n) converge en loi vers c alors (X_n, Y_n) converge en loi vers (X, c) (Slutsky)

8) Loi des grands nombres et Théorème centrale limite

Le but de cette partie est de décrire le comportement de somme de v.a qui la plupart du temps seront i.i.d.

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - m)$$

Théorème 15. *Soit (X_n) une suite de v.a.i.i.d et L^2 alors*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}[X_0]$$

Démonstration. Pour tout $n \in \mathbb{N}^*$ on a $E(\bar{X}_n) = E(X_1)$. Ainsi

$$\begin{aligned} E\left(\left|\bar{X}_n - E(X_1)\right|^2\right) &= Var(\bar{X}_n) \\ &= \frac{1}{n} Var(X_1) \\ &\to_n 0 \end{aligned}$$

□

- Soit (X_n) une suite de v.a.i.i.d $\mathcal{B}(p)$ alors $M_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} p$
- Estimation d'une proportion inconnue

Théorème 16. *Soit (X_n) une suite de v.a.i.i.d et L^1 alors*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}[X_0]$$

Théorème 17. ***Loi Forte des Grands Nombres :*** *Soit (X_n) une suite de v.a.i.i.d et L^1 alors*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s} \mathbb{E}[X_0]$$

Démonstration. **Etape 1) :** On peut supposer que $X_i \geqslant 0$ en décomposant en partie positive et négative.

Etape 2) : On peut remplacer X_i par $Y_i = X_i \mathbf{1}_{X_i \leqslant i}$. En effet les suites (X_n) et (Y_n) sont égales à partir d'un rang (aléatoire). En effet on a que

$$\begin{aligned}
\sum_{i=1}^{\infty} P[X_i \neq Y_i] = &= \sum_{i=1}^{\infty} P(X_i > i) \\
&= \sum_{i=1}^{\infty} P(X_1 > i) \\
&\leqslant \sum_{i=1}^{\infty} \int_{i-1}^{i} P(X_1 > x) dx \\
&= \int_{0}^{\infty} P(X_1 > x) dx \\
&= E\left(\int_0^{\infty} \mathbf{1}_{X_1 > x} dx \right) = E(X_1) < +\infty
\end{aligned}$$

Ainsi par Borel Cantelli

$$P\left(\bigcup_{n \in \mathbb{N}} \bigcap_{k \geqslant n} \{X_k = Y_k\} \right) = 1$$

Ainsi pour ω dans un ensemble de probabilité il existe $n \in \mathbb{N}$ telle que pour tout $k \geqslant n$ on a $X_k(\omega) = Y_k(\omega)$. On va donc démontrer que p.s

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \to E[X_1].$$

Etape 3) : On fixe $\alpha > 1$ et on définit $k_n = \lfloor \alpha^n \rfloor$. On va montrer que

$$\bar{Y}_{k_n} \to E(X_1)$$

on comblera les trous ensuite. Par convergence monotone on a que

$$E(Y_i) = E(X_1 \mathbf{1}_{X_1 \leqslant i}) \to E(X_1).$$

On a donc que

$$E(\bar{Y}_{k_n}) = \frac{1}{k_n} \sum_{i=1}^{k_n} E(X_1 \mathbf{1}_{X_1 \leqslant i}) \to E(X_1)$$

par Césaro. On est donc ramener à montrer que p.s

$$\bar{Y}_{k_n} - E(\bar{Y}_{k_n}) \to 0$$

Utilisons pour cela le critère de Borel Cantelli. Soit $\epsilon > 0$ on a

$$\begin{aligned}
\sum_{n=1}^{\infty} P(|\bar{Y}_{k_n} - E(\bar{Y}_{k_n}| \geqslant \epsilon) &\leqslant \sum_{n=1}^{\infty} \frac{Var(\bar{Y}_{k_n})}{\epsilon^2} \\
&\leqslant \sum_{n=1}^{\infty} \frac{1}{\epsilon^2} \sum_{i=1}^{k_n} \frac{Var(Y_i)}{k_n^2}.
\end{aligned}$$

Par ailleurs on a que

$$Var(Y_i) \leqslant E(|X_i|^2 \mathbf{1}_{X_i \leqslant i}) \leqslant E(|X_1|^2 \mathbf{1}_{X_1 \leqslant k_n})$$

ainsi

$$\begin{aligned}
\sum_{n=1}^{\infty} P(|\bar{Y}_{k_n} - E(\bar{Y}_{k_n}| \geqslant \epsilon) &\leqslant \sum_{n=1}^{\infty} \frac{1}{\epsilon^2} \sum_{i=1}^{k_n} \frac{E(|X_1|^2 \mathbf{1}_{X_1 \leqslant k_n})}{k_n^2} \\
&\leqslant \frac{1}{\epsilon^2} E\left(\sum_{n=1}^{\infty} \frac{\mathbf{1}_{X_1 \leqslant k_n}}{k_n} |X_1|^2 \right)
\end{aligned}$$

Comme $k_n \geqslant \alpha^n - 1 \geqslant (1 - 1/\alpha)\alpha^n$ on a que

$$\sum_{n=1}^{\infty} \frac{\mathbf{1}_{X_1 \leqslant k_n}}{k_n} \leqslant \sum_{n : X_1 \leqslant k_n} \frac{1}{(1 - 1/\alpha)\alpha^n}.$$

De plus si $X_1 \leqslant k_n$ alors $n \geqslant \frac{\log(X_1}{\log \alpha}$. Donc

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{\mathbf{1}_{X_1 \leqslant k_n}}{k_n} &\leqslant \frac{1}{1 - \frac{1}{\alpha}} \sum_{n : n \geqslant \frac{\log(X_1}{\log \alpha}} \frac{1}{\alpha^n} \\
&\leqslant \frac{1}{1 - \frac{1}{\alpha}} \frac{\alpha^{-\log(X_1)/\log(\alpha)}}{1 - \frac{1}{\alpha}} \\
&\leqslant \frac{X_1^{-1}}{(1 - 1/\alpha)^2}
\end{aligned}$$

Ainsi

$$E\left(\sum_{n=1}^{\infty} \frac{\mathbf{1}_{X_1 \leqslant k_n}}{k_n} |X_1|^2 \right) \leqslant \frac{EX_1}{(1 - 1/\alpha)^2}$$

On peut donc appliquer le critère de Borel Cantelli et conclure que

$$\bar{Y}_{k_n} - E(\bar{Y}_{k_n}) \to 0.$$

Etape 4) : Pour $n \geqslant 1$, on note $m(n)$ l'entier tel que

$$\lfloor \alpha^{m(n)} \rfloor = k_{m(n)} \leqslant n \leqslant \lfloor \alpha^{m(n+1)} \rfloor.$$

On a que

$$\begin{aligned}
\liminf \frac{\sum_{i=1}^{n} Y_i}{n} &\geqslant \liminf \frac{\sum_{i=1}^{k_{m(n)}} Y_i}{k_{m(n+1)}} \\
&= \liminf \bar{Y}_{k_{m(n)}} \frac{k_{m(n)}}{k_{m(n+1)}} \\
&= \frac{EX_1}{\alpha}
\end{aligned}$$

Ici il faut voir que

$$\frac{\lfloor \alpha^n \rfloor}{\lfloor \alpha^{n+1} \rfloor} \to \frac{1}{\alpha}$$

De manière similaire on montre que

$$\limsup \frac{\sum_{i=1}^n Y_i}{n} \leqslant \alpha E(X_1)$$

Comme α est arbitrairement proche de 1 on conclut que

$$E(X_1) \leqslant \liminf \frac{\sum_{i=1}^n Y_i}{n} \leqslant \limsup \frac{\sum_{i=1}^n Y_i}{n} \leqslant E(X_1)$$

et on obtient le résultat attendu. □

- Une réciproque.
- Application : méthode de Monte Carlo. Soit f une fonction mesurable telle que $f(X_0)$ soit L^1

$$\frac{1}{n}\sum_{i=1}^n f(X_i) \overset{p.s}{\to} \mathbb{E}[f(X_0)]$$

  Rq : ici aucune régularité n'est demandé pour f ce qui n'est pas le cas dans un certain nombre de méthodes classiques d'approximation d'intégrale.
- $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2$ sont des estimateurs de la moyenne et de la variance

Le theoreme central limite est le suivant

Théorème 18. *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi avec $E(X_1^2) < \infty$. Soit $m = EX_1$ et $\sigma^2 = \mathrm{Var}(X_1)$, alors*

$$\sqrt{n}\,\frac{\overline{X}_n - EX_1}{\sigma} = \frac{\sum_{i=1}^n X_i - nEX_1}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Soit $(X_n)_{n\geqslant 1}$ une suite de v.a. à valeur dans $\mathbb{R}^d$, indépendantes et de même loi. On suppose que $E\big(|X_1|^2\big) < \infty$ et on pose $\Gamma = \mathrm{Cov}(X_1)$. Dans ce cas

$$\frac{X_1 + \cdots + X_n - nEX_1}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d(0,\Gamma).$$

Démonstration. Quitte à centrer et à réduire les v.a on peut supposer que $E[X_1] = 0$ et $\sigma = 1$. Soit $T \in \mathbb{R}$, on a

$$\begin{aligned}
\Phi_{\sqrt{n}\bar{X}_n}(t) &= E(e^{it\sqrt{n}\bar{X}_n}) \\
&= E(\prod_{i=1}^n e^{it/\sqrt{n}X_i}) \\
&= \Big(E(e^{it\sqrt{n}X_1})\Big)^n \\
&= \big(\Phi_{X_1}(t\sqrt{n})\big)^n
\end{aligned}$$

Or $E(X^2) \leqslant \infty$ donc $\Phi_{X_1} \in C^2$ et on a

$$\Phi_{X_1}(t) = 1 - \frac{t^2}{2} + \circ(t^2)$$

Donc

$$\Phi_{\sqrt{n}\bar{X}_n}(t) = \left(1 - \frac{t^2}{2n} + \circ(\frac{t^2}{n})\right)^n.$$

Pour $z \in \mathbb{C}$ on a

$$\begin{aligned}
\left|e^z - \left(1 + \frac{z}{n}\right)^n\right| &= \left|\sum_{k=0}^{\infty} \frac{z^k}{k!} - \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \frac{z^k}{n^k}\right| \\
&\leqslant \sum_{k=0}^{n} \left|\frac{1}{k!} - \frac{n!}{k!(n-k)!n^k}\right| |z|^k + \sum_{k \geqslant n+1} \frac{|z|^k}{k!}
\end{aligned}$$

On remarque que la quantité

$$\frac{1}{k!} - \frac{n!}{k!(n-k)!n^k} = \frac{1}{k!}\left(1 - \frac{n(n-1)\dots(n-(k-1))}{n^k}\right) \geqslant 0$$

$$\begin{aligned}
\left|e^z - \left(1 + \frac{z}{n}\right)^n\right| &\leqslant \sum_{k=0}^{n} \left(\frac{1}{k!} - \frac{n!}{k!(n-k)!n^k}\right) |z|^k + \sum_{k \geqslant n+1} \frac{|z|^k}{k!} \\
&= e^{|z|} - \left(1 + \frac{|z|}{n}\right)^n \to_n 0
\end{aligned}$$

Ainsi

$$\Phi_{\sqrt{n}\bar{X}_n}(t) \to e^{-t^2/2}$$

qui est la fonction caractéristique de la $\mathcal{N}(0,1)$ et on applique le théorème de Levy. $\square$

— Affine la LFGN : vitesse de convergence en loi.
— Intervalle de confiance

9) Vecteurs Gaussiens

Définition 21. *Un vecteur aléatoire* $X = (X_1, \dots, X_d)^t$ *est appelé vecteur gaussien si toutes combinaisons linéaires de ces coordonnés est une v.a gaussienne, c'est à dire que pour tout* $a \in \mathbb{R}^d$ *la v.a*

$$< a, X >= \sum_{i=1}^{d} a_i X_i$$

est une v.a gaussienne

— Si X est un vecteur gaussien alors pour toute matrice A le vecteur AX est encore un vecteur gaussien

Définition 22. *Soit* $X = (X_1, \dots, X_d)^t$ *un vecteur aléatoire gaussien on note* K *sa matrice de covariance définie par*

$$K_{i,j} = Cov(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j],$$

pour tout $i, j = 1, \dots, d$. *On notera également*

$$m = \mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^t$$

le vecteur de la moyenne. On notera

$$X \sim \mathcal{N}_d(m, K)$$

— La matrice K est positive, $K = E((X - m)^t(X - m))$
— $\mathbb{E}[< a, X >] =< a, \mathbb{E}[X] >$
— $Var(< a, X >) = Var(\sum_{i=1}^{d} a_i X_i) = \sum_{i,j=1}^{d} a_i a_j Cov(X_i, X_j) = a^t K a =< a, Ka >$
— On a en particulier

$$\phi_{<a,X>}(t) = \exp\left(i < a, m > t - \frac{1}{2} a^t K a\, t^2\right)$$

— $\phi_X(\xi) = \mathbb{E}[e^{i<\xi,X>}] = \phi_{<\xi,X>}(1)$

Proposition 17. *La fonction caractéristique d'un vecteur Gaussien est donnée par*

$$\phi_X(\xi) = \exp\left(i < \xi, m > -\frac{1}{2}\xi^t K \xi\right)$$

— Les coordonnées d'un vecteur gaussien sont indépendantes si et seulement si sa matrice de covariance est diagonale

Proposition 18. *Soit* $X \sim \mathcal{N}_d(m, K)$ *alors pour toute matrice* $A \in \mathbb{M}_{p,d}(\mathbb{R})$ *alors*

$$AX \sim \mathcal{N}_p(AX, AKA^t)$$

— Si $X \sim \mathcal{N}_d(0, I_d)$ alors la loi de X est invariante par toute rotation.
— On dira qu'un vecteur gaussien X est dégénéré si sa matrice de covariance K est non inversible
— Dans le cas dégénéré, il existe donc a tel que $Ka = 0$ ce qui implique que

$$Var(< a, X >) = 0$$

et donc $< a, X >= b$ p.s donc X vit dans l'espace affine $\{< a, x >= b, x \in \mathbb{R}^d\}$
— Si K est inversible alors $\sqrt{K}^{-1}(X - m) \sim \mathcal{N}(0, I_d)$
— Si $N \sim \mathcal{N}(0, I_d)$ alors $X = \sqrt{K}N + m \in \mathcal{N}(m, K)$
— Si $X \sim \mathcal{N}_d(0, I_d)$ alors les coordonnées $(X_i)_{i=1,\dots,d}$ sont i.i.d et $X_1 \sim \mathcal{N}(0, 1)$. Ainsi la densité de X est donnée par le produit des densités i.e

$$f_X(x_1, \dots, x_d) = \frac{1}{\sqrt{2\pi}^d} \exp(-\frac{1}{2} \sum_{i=1}^{d} x_i^2)$$

— Dans le cas où K est inversible on a

$$f_X(x_1, \dots, x_d) = \frac{1}{\sqrt{(2\pi)^d \det K}} \exp(-\frac{1}{2} < (x - m), K^{-1}(x - m) >)$$

— Rq : si X est gaussien toutes ses coordonnées le sont, l'inverse n'est pas vrai en général.

Théorème 19. *Soit $X^{(n)}$ une suite de vecteurs aléatoires de $\mathbb{R}^d$ i.i.d et L^2 de vecteur moyen m et de matrice de covariance K. On pose $S^{(n)} = \sum_{i=1}^n X^{(i)}$ alors on a*

$$n^{-1/2}\sqrt{K}^{-1}(S^{(n)} - nm) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, I_d)$$

ou

$$n^{-1/2}(S^{(n)} - nm) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, K)$$

Soit (X_n) i.i.d de moyenne m et de variance σ^2
— Le TCL nous dit

$$\sqrt{n}(\bar{X}_n - m) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

— Si $\mathbb{E}[X_1^4] < \infty$, en posant $\mu^4 = \mathbb{E}(X_0 - m)^4$ alors on a

$$\sqrt{n}(\bar{S}_n^2 - nm) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu^4 - \sigma^4)$$

— Soit $X = (X_1, \dots, X_d)$ un vecteur gaussien où les (X_i) sont i.i.d de loi $\mathcal{N}(0, 1)$ alors

$$Z = \sum_{i=1}^{d} X_i^2$$

est une variable aléatoire appelé v.a $\chi^2(d)$ où d est appelé degré de liberté

— Sa densité est donnée par

$$f_Z(z) = \frac{1}{2\Gamma(k/2)} z^{\frac{k}{2}-1} e^{-\frac{z}{2}} \mathbf{1}_{z\geqslant 0}$$

où Γ est la fonction gamma.

— Soit $X \sim \mathcal{N}(0,1)$ et $Z \sim \chi^2(k)$ alors la v.a

$$T = \frac{X}{\sqrt{Z/k}}$$

est une variable alétoire dite de Student de degré k

— Sa densité est donnée par

$$f_T(t) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} (1 + \frac{t^2}{2})^{-\frac{k+1}{2}}$$

Proposition 19. *Soit $X \sim \mathcal{N}_d(0, I_d)$ et $\mathbb{R}^d = F_1 \oplus \ldots \oplus F_k$ une décomposition en sous espaces orthogonaux avec $dim(F_i) = d_i$. On note $P_{F_i}, i = 1, \ldots, k$ les projecteurs orthogonaux associés aux espaces $F_i, i = 1 \ldots, k$. Dans ce cas les vecteurs $P_{F_1}(X), \ldots, P_{F_k}(X)$ sont des vecteurs aléatoires gaussiens indépendants. On a également*

$$\|P_{F_i}(X)\|^2 \sim \chi^2(d_i), i = 1, \ldots, k$$

— C'est de l'algèbre linéaire

— On peut exprimer un résultat plus général lorsque $X \sim \mathcal{N}(0, K)$ avec K non dégénérée en introduisant le produit scalaire relié à K i.e $< a, b >_K = < a, Kb >$.

— Considérons

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i)^2 - (\bar{X}_n)^2$$

— Par la loi forte des grands nombres cette quantité converge vers la variance de X_1

— Commençons par calculer

$$\begin{aligned}
\mathbb{E}[\sigma_n^2] &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i^2] - \mathbb{E}[(\bar{X}_n)^2] \\
&= \mathbb{E}(X_1^2) - \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)^2\right] \\
&= \mathbb{E}(X_1^2) - \frac{1}{n^2} \sum_{i,j} \mathbb{E}[X_i X_j] \\
&= \mathbb{E}(X_1^2) - \frac{1}{n^2} \left(\sum_{i=j} \mathbb{E}[(X_i)^2] + \sum_{i \neq j} \mathbb{E}[X_i]\mathbb{E}[X_j] \right) \\
&= \mathbb{E}(X_1^2) - \frac{1}{n}\mathbb{E}[X_1^2] - \frac{1}{n^2} \sum_{i \neq j} \mathbb{E}[X_1]^2 \\
&= \frac{n-1}{n}\mathbb{E}[X_1^2] - \frac{n-1}{n}\mathbb{E}[X_1]^2 = \frac{n-1}{n} Var(X_1)
\end{aligned}$$

— On a donc

$$\mathbb{E}[\sigma_n^2] \quad = \quad \frac{n-1}{n} Var(X_1)$$

— On préfère considérer

$$S_n^2 = \frac{n}{n-1}\sigma_n^2 = \frac{1}{n-1}\sum_i^n (X_i - \bar{X}_n)^2,$$

qui est encore un estimateur de la variance mais qui est non biaisé.
— Supposons que $(X_1, \dots, X_n)^t$ soit un vecteur gaussien indépendant de loi $\mathcal{N}(m, \sigma^2)$. Nous avons

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

— En effet on pose $Y = \frac{1}{\sigma}(X_1 - m, \dots, X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
— Posons $F = Vect(1_n)$ avec $1_n = (1, \dots, 1)^t$. On voit que $dim(F) = 1$ et $dim(F^{\perp}) = n-1$.
— Alors maintenant soit $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X \right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m, \dots \bar{X}_n - m)^t$ et donc

$$P_{F^{\perp}}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)^t$$

— Le théorème de Cochran nous dit que $\|P_{F^{\perp}}(X)\|^2 \sim \chi^2(n-1)$. Il est facile de voir que

$$\|P_{F^{\perp}}(X)\|^2 = \frac{n-1}{\sigma^2} S_n^2$$

— Cela permet entre autre de construire des intervalles de confiance pour la variance d'une loi gaussienne. On pose $\chi^k_{1-\alpha}$ le quantile de la loi de $\chi^2(k)$ c'est à dire si $T \sim \chi^2(k)$ alors

$$\mathbb{P}[\chi^k_{\alpha/2} \leqslant T \leqslant \chi^k_{1-\alpha/2}] = 1 - \alpha$$

— On a alors

$$\mathbb{P}\left[\chi^{n-1}_{\alpha/2} \leqslant \frac{n-1}{\sigma^2} S_n^2 \leqslant \chi^{n-1}_{1-\alpha/2}\right] = 1 - \alpha$$

— Cela implique

$$\mathbb{P}\left[\frac{n-1}{\chi^{n-1}_{1-\alpha/2}} S_n^2 \leqslant \sigma^2 \leqslant \frac{n-1}{\chi^{n-1}_{\alpha/2}} S_n^2\right] = 1 - \alpha$$

et l'intervalle

$$\left[\frac{n-1}{\chi^{n-1}_{1-\alpha/2}} S_n^2, \frac{n-1}{\chi^{n-1}_{\alpha/2}} S_n^2\right]$$

est un intervalle de confiance au degré α pour la variance σ^2 of X_1.
— Quand $X_1, \dots, X_n$ est Gaussien indépendant $\mathcal{N}(m, \sigma^2)$ alors

$$\sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right) \sim \mathcal{N}(0, 1)$$

alors si σ^2 est connu cela permet de construire un IC pour μ

— Si σ^2 est inconnu on remplace σ par S_n et on a

$$\sqrt{n}\left(\frac{\bar{X}_n - m}{S_n}\right) \sim \mathcal{T}_{n-1}$$

où $\mathcal{T}_{n-1}$ suit une loi de Student avec $n-1$ degré de liberté.

Test d'adéquation du χ^2 :

— On observe une variable aléatoire X de support connu $S(X) = \{a_1, \ldots, a_r\}$ et $p_j = \mathbb{P}(X = a_j) = Q(\{a_j\}), j = 1, \ldots, r$ inconnues. On note $p = (p_1, \ldots, p_r)$ le vecteur des probabilités

— On pose $Q_0 = \sum_i \pi_i \delta_i$ de même support mais avec un vecteur $\pi = (\pi_1, \ldots, \pi_r)$ connu où $\pi_i > 0$

— On cherche à tester l'hypothèse $H_0 : Q = Q_0$ contre $H_1 : Q \neq Q_0$.

— On observe un vecteur (X_n) une suite de v.a.i.i.d de loi Q. Pour $n \in \mathbb{N}$, on pose

$$N_j = \sum_{i=1}^{n} \mathbf{1}_{X_i = a_j}$$

— On peut voir que le vecteur $N = (N_1, N_2, \ldots, N_r)^t$ suit une loi mutinomiale $\mathcal{M}(n, p_1, \ldots, p_r)$ i.e

$$\mathbb{P}(N_1 = n_1, \ldots, N_r = n_r) = \frac{n!}{n_1! \ldots n_r!} p_1^{n_1} \ldots p_r^{n_r}, \quad n_1 + \ldots + n_r = n$$

— On pose

$$T_n = \sum_{j=1}^{r} \frac{(N_j - n\pi_j)^2}{n\pi_j}$$

— On voit que sous H_0 cette quantité va être proche de 0 alors que sous H_1 elle va exploser.

Théorème 20. — *Sous l'hypothèse H_0 on a*

$$T_n \xrightarrow{\mathcal{L}} \chi^2(r-1)$$

— *Sous l'hypothèse H_1 on a*

$$T_n \xrightarrow{p.s} +\infty$$

— Test d'homogénéité, test d'indépendance

— Soit (X_n) une séquence de v.a i.i.d L^2. Notons $\theta = \mathbb{E}[X_1]$ et $\sigma^2 = Var(X_1)$. Rappelons que si

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

alors le TCL dit

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

— Dans certaine situation on considère f et on veut regarder

$$\sqrt{n}\left(f(\bar{X}_n) - f(\theta)\right)$$

— On utilise la méthode appelée Delta méthode
— Gardons à l'esprit le résultat du TCL

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

— Considérons $f(x) = ax + b$ alors on a

$$\sqrt{n}\left(a\bar{X}_n - a\theta)\right) \xrightarrow{\mathcal{L}} a\mathcal{N}(0, \sigma^2) = \mathcal{N}(0, a^2\sigma^2)$$

— Supposons que f est differentiable en θ on peut écrire $f(x) = f(\theta) + f'(\theta)(x - \theta) + \circ(|x - \theta|)$. Puisque $\bar{X}_n - \theta$ converge vers 0 presque surement, on a convergence en proba vers 0. On peut alors écrire

$$f(\bar{X}_n) = f(\theta) + f'(\theta)(\bar{X}_n - \theta) + \circ_{\mathbb{P}}(|\bar{X}_n - \theta|)$$

— On peut alors écrire

$$f(\bar{X}_n) = f(\theta) + f'(\theta)(\bar{X}_n - \theta) + \circ_{\mathbb{P}}(|\bar{X}_n - \theta|)$$

into $\sqrt{n}(f(\bar{X}_n) - f(\theta))$, et on a

$$\sqrt{n}(f(\bar{X}_n) - \theta) = \sqrt{n}f'(\theta)(\sqrt{n}(\bar{X}_n - \theta))(1 + \circ_{\mathbb{P}}(1))$$

— Maintenant $1 + \circ_{\mathbb{P}}(1)$ converge vers 1 en proba et donc en loi. On peut alors utiliser le lemme de Slutsky et on a

$$\sqrt{n}(f(\bar{X}_n) - f(\theta)) \xrightarrow{\mathcal{L}} f'(\theta)\mathcal{N}(0, \sigma^2) = \mathcal{N}(0, f'(\theta)^2\sigma^2)$$

10) Estimateur du maximum de vraisemblance

Le contexte est le suivant

— On observe un échantillon $(X_1, \ldots, X_n)$ de loi $\mathcal{L}(\theta)$. On connaît à priori la forme des lois $\mathcal{L}(\theta)$ mais on ne connaît pas θ (exemple paramètre d'une loi de Bernoulli).

— On cherche alors à définir un estimateur de ce paramètre c'est à dire une fonction

$$\hat{\theta}_n : (x_1, \ldots, x_n) \to \hat{\theta}_n(x_1, \ldots, x_n)$$

tel que

$$\lim_{n\to\infty} \hat{\theta}_n(X_1, \ldots, X_n) = \theta \ p.s$$

— Il peut exister plusieurs estimateurs satisfaisant cette contrainte, on va chercher celui qui maximise la vraisemblance

$$L_n(x_1, \ldots, x_n, \theta)$$

Cas discret : soit $(X_1, \ldots, X_n)$ un échantillon de loi $\mathbb{P}_\theta$

— Dans le cas discret la vraisemblance est décrite par

$$L_n(x_1, \ldots, x_n, \theta) = \mathbb{P}_\theta(X_1 = x_1) \ldots \mathbb{P}_\theta(X_n = x_n)$$

— On pose alors estimateur du maximum de vraisemblance

$$\hat{\theta}_n = argmax_\theta(L_n(x_1, \ldots, x_n, \theta))$$

— On est alors amené souvent à déterminer ce maximum en résolvant

$$\frac{d}{d\theta} L_n(x_1, \ldots, x_n, \theta) = 0$$

— Exemple : $\mathcal{B}(p)$, $\mathcal{P}(\lambda)$

Cas continu : soit $(X_1, \ldots, X_n)$ un échantillon de densité $f_\theta(.)$

— Dans le cas discret la vraisemblance est décrite par

$$L_n(x_1, \ldots, x_n, \theta) = f_\theta(x_1) \ldots f_\theta(x_n)$$

— On pose alors estimateur du maximum de vraisemblance

$$\hat{\theta}_n = argmax_\theta(L_n(x_1, \ldots, x_n, \theta))$$

— On est alors amené souvent à déterminer ce maximum en résolvant

$$\frac{d}{d\theta} L_n(x_1, \ldots, x_n, \theta) = 0$$

— Exemple : $\mathcal{U}([0, a])$, $\mathcal{E}(\lambda)$, $\mathcal{N}(m, \sigma^2)$.