pierre.ribereau@univ-lyon1.fr
ribereau.pierre@gmail.com

I. Vecteurs aléatoires (X, Y)

II. Fonction caractéristique

III. Vecteur Gaussiens (X, Y) Gaussien ?

IV. Convergence de V.A

$X_1, X_2, X_3, \dots \quad X_n \to X$?

V. Théorème Limites : Centrale limite, loi des grands nombres

VI. Théorème de Radon-Nikodym

X V.A sous $\mathbb{P}$. Loi de X sous $\mathbb{Q}$?

VII. Espérance conditionnelle

si on a un événement A tq $\mathbb{P}(A) \neq 0$ $\quad E(X|A)$

On veut $E(X|Y) \Leftarrow$ V.A qui s'écrit comme $g(Y)$

et $E(X|\mathcal{F})$ où $\mathcal{F} \subset \mathcal{A}$

VIII. Introduction aux martingales

Chapitre 1 : Vecteurs Aléatoires

Definition: Une famille de v.a $X = (X_1, X_2, \dots, X_n)$ définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^n$

Ω : espace fondamental, $\mathcal{A}$: Tribu, $\mathbb{P}$: mesure de Probas

par $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$ pour $\omega \in \Omega$

est un vecteur aléatoire ou une variable aléatoire de dimension n si l'image réciproque de l'intervalle

$I = \{(x_1, \dots, x_n) ; -\infty < x_i \leq a_i , \ a_i \in \mathbb{R}\}$

est dans $\mathcal{A}$. autrement dit

$X^{-1}(I) = \{\omega : X_1(\omega) \leq a_1, \dots, X_n(\omega) \leq a_n\} \in \mathcal{A}$

ou encore $\forall B \in \mathcal{B}(\mathbb{R}^n)$, $X^{-1}(B) \in \mathcal{A}$.

Remarque : Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire alors $X_1, \dots, X_n$ sont des variables aléatoires. Et inversement si $X_1, \dots, X_n$ sont des variables. a sur $(\Omega, \mathcal{A}, \mathbb{P})$ alors $X = (X_1, \dots, X_n)$ est un vecteur aléatoire.

Definition: Fonction de répartition

On appelle Fonction de répartition (fdr ou cdf) d'un V.A $X = (X_1, \dots, X_n)$ la fonction à plusieurs variables F_X définie par

$F(x_1, \dots, x_n) = \mathbb{P}((X_1 \leq x_1) \cap \dots \cap (X_n \leq x_n))$

$F(x_1, x_2) = \mathbb{P}(X_1 \leq x_1 , X_2 \leq x_2)$

Propriétés :

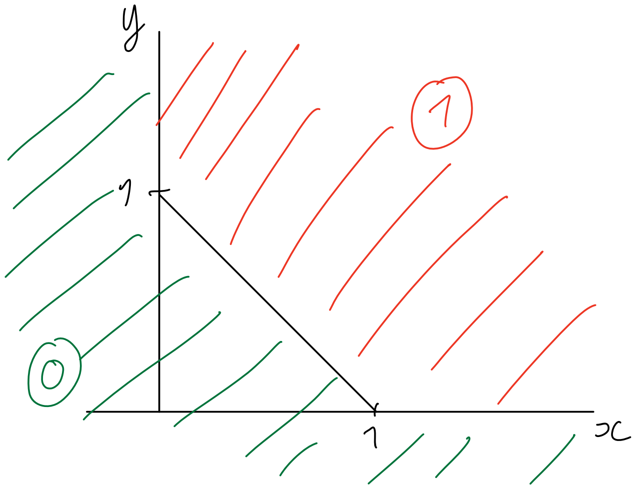
1. $F(x_1, \dots, x_n) \to 0$ quand n'importe quel $x_i \to -\infty$
2. $F(x_1, \dots, x_n) \to 1$ quand tous les $x_i \to +\infty$
3. F est croissante et $\underbrace{\text{cadlag}}_{\text{continue à droite, limite à gauche}}$ par rapport à chacune des coordonnées

Remarque : En dimension 1, si une fct est croissante, de 0 en $-\infty$ à 1 en $+\infty$ et cadlag alors c'est une f.d.r

Mais c'est pas suffisant en dimension supérieure

Exemple : On prend

$$F(x,y) = \begin{cases} 0 & \text{si } x<0 \text{ ou } y<0 \text{ ou } x+y<1 \\ 1 & \text{sinon} \end{cases}$$



On a bien $F(-\infty, x_2) = 0$

et $F(x_1, -\infty) = 0$ et

$F(\infty, \infty) = 1$ et F est croissante

et cadlag par rapport à x et y

⚠ Mais ça n'est pas une f.d.r en dimension 2 car si F était une fdr alors

$$\mathbb{P}\left(\tfrac{1}{3} < X \le 1, \tfrac{1}{3} < Y \le 1\right) = \underbrace{F(1,1)}_{=1} - \underbrace{F(1,\tfrac{1}{3})}_{=1} - \underbrace{F(\tfrac{1}{3},1)}_{=1} + \underbrace{F(\tfrac{1}{3},\tfrac{1}{3})}_{=0}$$

$= -1$ pas possible

plus généralement, si on a une fonction F qui vérifie les propriétés, si on veut que ça soit une f.d.r il faut en plus que

$$\mathbb{P}(x_1 < X \leq x_2, y < Y < y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

soit $\in [0,1]$ $\forall x_1 < x_2$ et $y_1 < y_2$

Caractérisation par la densité

Cas discret : ⚠ on peut dire densité dans le cas discret mais c'est mieux de dire fct de masse

Définition I.3 *Une v.a. bivariée (X, Y) est dite discrète si elle prend ps ses valeurs dans un ensemble dénombrable A de (couples) valeurs. Toute paire (x_i, y_j) qui a une probabilité strictement positive p_{ij} est appelée point de saut de la fonction de répartition, et p_{ij} est appelé le saut en (x_i, y_j). A est le support de la distribution de (X, Y).*

$$\sum_{i,j} p_{ij} = 1 \quad \text{et} \quad F(x,y) = \mathbb{P}(X \leq x, Y \leq y) = \sum_{B} p_{ij}$$

$$B = \{(x_i, y_j) \in A \text{ tq } x_i \leq x \text{ et } y_j \leq y\}$$

Les $p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$ forment la fonction de masse (PMF)

Remarque : Toute famille (p_{ij}) tq $\sum_{i,j} p_{i,j} = 1$ définit une fonction de masse.

Exemple : Distribution Multinomiale

On a n étudiants qui doivent choisir entre 3 BDE. Soit X_1, X_2 et X_3 les V.A qui correspondent au nombre de voix pour le BDE 1, BDE 2, BDE 3. On suppose que les étudiants votent indépendamment les uns des autres avec $p_1 = 0.4$, $p_2 = 0.45$ et $p_3 = 0.15$

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3)$$

$$= \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

ou $x_1, x_2, x_3 \in \{0, 1, \dots, n\}$ et $x_1 + x_2 + x_3 = n$

Cas continu : Un vecteur aléatoire (X, Y) est dit continu si il existe une fonction f positive tq $\forall (x,y) \in \mathbb{R}^2$ on a

$$F(x,y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f(u,v)\, dv \right) du$$

où F est la f.d.r de (X,Y) et f est la densité de (X,Y)

Probabilité de Ledoux et Barbe + exercices corrigés de Carrieu

$$F(\infty, \infty) = \lim_{x,y \to \infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dv\, du = 1$$

Si en plus f est continue en x et y

$$\frac{\partial^2 F}{\partial x \partial y} = f$$

Théorème : Si f est une fonction positive tq

$$\iint_{\mathbb{R}^2} f(x,y)\, dx\, dy = 1$$

alors f définit une densité

Definition : Loi marginale i.e loi de X_i ds $(X_1, \dots, X_n)$

La fonction de répartition marginale de $X = (X_1, \dots, X_n)$ relativement à X_i est def par

$$F_{X_i}(x_i) = F_X(\infty, \dots, \infty, x_i, \infty \dots)$$

$$F_X(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_i \leq x_i, \dots, X_n \leq x_n)$$

C'est la fonction de répartition marginale de X_i. C'est une f.d.r univariée qui correspond à la loi de X_i

De la même façon :

$$F_{X_i, X_j}(x_i, x_j) = F(\infty, \dots \infty, x_i, \infty, \dots, \infty, x_j, \infty, \dots, \infty)$$

est la loi marginale de (X_i, X_j)

Remarque : si la densité f_X du vecteur $(X_1, \dots, X_n) = X$ existe

on a $\int \dots \int f(x_1, \dots, x_n)\, dx_1, \dots, dx_n = 1$

Si on veut récupérer la densité marginale de X_i, on calcule

$$f_{X_i}(x_i) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(x_1, \dots, x_n) \underbrace{dx_1 \dots dx_{i-1}\, dx_{i+1} \dots dx_n}_{\text{on intègre par rapport à toutes les coordonnées sauf } x_i}$$

De manière analogue, on a

$$f_{X_i, X_j}(x_i, x_j) = \underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}}}_{n-2} f_X(x_1, \dots, x_n)\, dx_1 \dots dx_{i-1}\, dx_{i+1} \dots dx_{j-1}\, dx_{j+1} \dots dx_n$$

Résumé en dimension 2 avec (X, Y)

$$f_X(x) = \begin{cases} \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy & \text{si cont} \\ \sum_y \mathbb{P}(X = x, Y = y) & \text{si discret} \end{cases}$$

Indépendance

Définition : Soit X un vect. aléatoire de dimension n, partitionnée en $X = (X_1, X_2)$ avec $X_1 \in \mathbb{R}^{n_1}$, $X_2 \in \mathbb{R}^{n_2}$ et $n_1 + n_2 = n$, alors X_1 est indépendant de X_2 si

$$F_X(\underbrace{x_1}_{\in \mathbb{R}^{n_1}}, \underbrace{x_2}_{\in \mathbb{R}^{n_2}}) = F_X(\underbrace{x_1}_{\in \mathbb{R}^{n_1}}, \underbrace{\infty}_{(\infty,\dots,\infty)}) \times F_X(\infty, \underbrace{x_2}_{\in \mathbb{R}^{n_2}}) \qquad \forall (x_1, x_2) \in \mathbb{R}^{n_1+n_2}$$

En dimension 2, X et Y sont indep si

$$F_{X,Y}(x,y) = F_X(x)\, F_Y(y) = \mathbb{P}(X \leq x)\, \mathbb{P}(Y \leq y) \qquad \forall (x,y) \in \mathbb{R}^2$$
$$= \mathbb{P}(X \leq x, Y \leq y)$$

Ce qui implique que

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) \qquad \forall (x,y) \in \mathbb{R}^2$$

Définition : On définit la densité de Y conditionnelle à $X = x$ par

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \qquad \forall y \in \mathbb{R}$$

Exemple : Le vect. a (X, Y) a comme densité

$$f_{X,Y}(x,y) = \begin{cases} 2xy + \frac{3}{2} y^2 & \text{si } 0 < x < 1 \text{ et } 0 < y < 1 \\ 0 & \text{sinon} \end{cases}$$

1) Vérifier que $f_{X,Y}$ est une densité

2) Calculer f_X, f_Y et $f_{Y|X=x}$

3) $\mathbb{P}((X,Y) \in [0,\frac{1}{2}]^2)$

4) Calculer $\mathbb{P}(X < Y)$

1- Rappel : f est une densité si $f \geq 0$ et $\iint f = 1$

f est positive et

$$\iint_{\mathbb{R}^2} f(x,y)\, dx\, dy = \iint 2xy + \frac{3}{2}y^2\, dx\, dy = \int_{\mathbb{R}} (y + \frac{3}{2}y^2)\, dy = 1$$

donc f est bien une densité

2- Rappel : $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy$

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dx = \int_0^1 (2xy + \frac{3}{2}y^2)\, dx = (y + \frac{3}{2}y^2)\, 𝟙_{[0,1]}(y)$$

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy = \int_0^1 (2xy + \frac{3}{2}y^2)\, dy = (x + \frac{1}{2})\, 𝟙_{[0,1]}(x)$$

On remarque que X et Y ne sont pas indep car

$$f_X(x)\, f_Y(y) = (x + \frac{1}{2})(y + \frac{3}{2}y^2) \neq f_{X,Y}(x,y) = 2xy + \frac{3}{2}y^2$$

Rappel $f_{Y|X=x}(y|x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)} = \dfrac{2xy + \frac{3}{2}y^2}{x + \frac{1}{2}}$

3) On veut $\mathbb{P}((X,Y) \in [0,\frac{1}{2}]^2)$

$$\mathbb{P}((X,Y) \in [0,\tfrac{1}{2}]^2) = \int_0^{1/2}\int_0^{1/2} (2xy + \frac{3}{2}y^2)\, dx\, dy$$

$$= \int_0^{1/2} (\frac{y}{4} + \frac{3}{4}y^2)\, dy = \left[\frac{y^2}{8} + \frac{y^3}{4}\right]_0^{1/2} = \frac{1}{16}$$

4) On veut

$$\mathbb{P}(X<Y) = \int_0^1 \int_0^y \left(2xy + \frac{3}{2}y^2\right) dx\, dy$$

$$= \int_0^1 \left(y^3 + \frac{3}{2}y^3\right) dy = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$$

Rappel : $F_X(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x)$

$$f_X(x_1, \dots, x_n) = \frac{\partial^n F_X(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}$$

Loi marginale : $F_{X_i}(x_i) = F(\infty, \dots, \infty, x_i, \infty, \dots, \infty)$

indépendance : $F_{X,Y}(x,y) = F_{X,Y}(x,\infty) \times F_{X,Y}(\infty, y)$

$$= F_X(x) \cdot F_Y(y)$$

ou $f_{XY}(x,y) = f_X(x)\, f_Y(y)$

II Moments et espaces L^p

Rappel : En dimension 1, on définit le moment d'ordre k d'une V.A X comme :

$$m_k = E(X^k) = \begin{cases} \int x^k f_X(x)\,dx & \text{si densité existe} \\ \sum_{x_i \in \mathcal{X}} x_i^k \,\mathbb{P}(X = x_i) & \text{si discrète} \end{cases}$$

Le moment centré d'ordre k est défini par

$$\bar{m}_k = E\left(\left(X - E(X)\right)^k\right) = E\left((X - m_1)^k\right)$$

A quoi ça sert ?

→ les moments fournissent un résumé numérique de la distribution

$m_1 \simeq$ le "milieu" de la distribution

→ m_2 est la dispersion autour de m_1

→ Si on connait tous les moments $< \infty$ on connait la distribution

→ les moments peuvent ne pas exister à partir d'un certain k_0.

plus k_0 est petit, plus la distribution est dangereuse d'un point de vue actuarielle

Si m_1 existe pas, le risque n'est pas assurable en théorie

Si X admet un moment d'ordre p $(E(X^p) < \infty)$ alors on dit que $X \in L^p$

<u>Rappel sur les espaces L^p</u> :

On appelle $L^p(\Omega, \mathcal{A}, \mathbb{P})$ l'espace vectorielle des v.A réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ qui admettent un moment d'ordre p.

Remarque : Si X admet un moment d'ordre 2 alors X admet un moment d'ordre 1 donc si $X \in L^2$ alors $X \in L^1$

$\Rightarrow L^2(\Omega, \mathcal{A}, \mathbb{P}) \subset L^1(\Omega, \mathcal{A}, \mathbb{P})$

$\Rightarrow L^p(\Omega, \mathcal{A}, \mathbb{P}) \subset L^q(\Omega, \mathcal{A}, \mathbb{P})$ pour $p > q$

Definition : <u>Covariance</u>

Si X et Y sont dans L^2 alors on définit la covariance entre X et Y par

$$\begin{aligned} \operatorname{Cov}(X,Y) &= E\big((X - E(X))(Y - E(Y))\big) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

si X et Y indep alors $\operatorname{cov}(X,Y) = 0$ car $E(XY) = E(X)E(Y)$

⚠ La réciproque est fausse

On definit le coefficient de <u>corrélation linéaire</u> entre X et Y par

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$

∗ Inégalité de Cauchy - Shwartz :

Si X et $Y \in L^2$ alors

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$$

Remarque :

- La covariance est une mesure de la dépendance entre deux V.A
- Le coefficient de corrélation mesure l'intensité d'une relation linéaire entre X et Y
- Par Cauchy - Schwartz on a $|\rho_{XY}| \leq 1$
- Dans le cas où $|\rho_{XY}| = 1$ alors $Y = aX + b$
  signe de a = signe de ρ_{XY}

Definition : Si $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ alors $\|X\|_p = \left(\int |X|^p dP\right)^{1/p}$

En dimension supérieur : $X = (X_1, \dots, X_n)$

Le moment d'ordre 1 de X est le vecteur des moments d'ordre 1 des composantes de X

$\Rightarrow$ Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^n)$. Le moment d'ordre 1 de X noté $\bar{X}$ est le vecteur composé des espérences des composantes de X.

$$\bar{X} = E(X) = E\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

et $X - \bar{X}$ est le vecteur centré.

Remarque : $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^n)$

Si $\lambda \in \mathbb{R}$ et $\mu \in \mathbb{R}^n$ alors

$$E(\lambda X + \mu) = E\begin{pmatrix} \lambda X_1 + \mu_1 \\ \vdots \\ \lambda X_n + \mu_n \end{pmatrix} = \lambda \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

$$= \lambda E(X) + \mu$$

Si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ et $B \in \mathbb{R}^p$ alors $E(AX + B) = A\,E(X) + B$

Remarque : Pour le moment 2, on doit définir une matrice qui contient plus d'information

Définition : On définit la matrice de variance (-covariance) d'un vecteur comme

transposé

$$\Gamma_X = Var(X) = E\left((X - E(X)) \cdot (X - E(X))^t\right)$$

$$= E\left(\begin{pmatrix} x_1 - E(X_1) \\ \vdots \\ x_n - E(X_n) \end{pmatrix} \cdot (X_1 - E(X_1), \dots, X_n - E(X_n))\right)$$

$$\Gamma_X = E\begin{pmatrix} (X_1 - E(X_1))^2 & & (X_i - E(X_i))(X_j - E(X_j)) \\ & \ddots & \\ & & (X_n - E(X_n))^2 \end{pmatrix}$$

$$= \begin{pmatrix} V(X_1) & & cov(X_i, X_j) \\ & \ddots & \\ & & V(X_n) \end{pmatrix} = Cov(X_i, X_j)_{i,j = 1 \dots n}$$

Propriétés $X \in L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^n)$

1- $\Gamma_X = V(X)$ est une matrice symmétrique $n \times n$ définie positive

2- Si $A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n$ alors $V({}^tA\ X) = V(\sum a_i x_i) = {}^tA\, \Gamma_X\, A$

3- Si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ et $B \in \mathbb{R}^p$ alors

$$V(AX+B) = A\, V(X)\, .{}^tA$$

Exemple - exercice : Régression linéaire

Soit (X, Y) un vecteur aléatoire de $L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^2)$

1- Quelle est la constante a qui approxime le mieux Y au sens de L^2 ?

On veut minimiser :

$$\begin{aligned}\Phi(a) &= E((Y-a)^2) = E((Y-E(Y)+E(Y)-a)^2) \\ &= E((Y-E(Y))^2) + (E(Y)-a)^2 + \underbrace{2E((Y-E(Y))(E(Y)-a))}_{=(E(Y)-a)\underbrace{E((Y-E(Y)))}_{=E(Y)-E(Y)}} \\ &= V(Y) + (E(Y)-a)^2\end{aligned}$$

fonction positive en a minimum en 0 quand $a = E(Y)$

$\Rightarrow E(Y)$ est la meilleur approximation de Y par une constante au sens de la norme L^p

2- Quelle est la fonction affine $(ax+b)$ qui approxime le mieux Y au sens L^2 ? i.e déterminer (a,b) qui minimisent $E((Y-aX-b)^2)$

On veut minimiser $\phi(a,b) = E((Y-aX-b)^2)$

$$\phi(a,b) = E\Big(\big((Y-E(Y)) - a(X-E(X)) - b + E(Y) - aE(X)\big)^2\Big)$$

$$= E\Big(\big((Y-E(Y)) - a(X-E(X))\big)^2\Big) + E\big((E(Y)-b-aE(X))^2\big)$$

$$+ 2\underbrace{E\big(((Y-E(Y)) - a(X-E(X))\big)}_{\substack{=E(Y-E(Y))-aE(X-E(X))\\ =0}} \cdot \underbrace{(E(Y)-b-aE(X))}_{\text{constante}})$$

$$= E\Big(\big((Y-E(Y)) - a(X-E(X))\big)^2\Big) + \big(E(Y)-b-aE(X)\big)^2$$

pour a fixé, $\phi(a,b)$ est la somme de deux termes, le premier est une constante, le second est un carré qui est minimal quand il vaut 0

i.e. $E(Y)-b-aE(X)=0$ i.e. $b = E(Y) - aE(X)$

Il reste à minimiser en a la première partie

$$\psi(a) = E\big(((Y-E(Y)) - a(X-E(X)))^2\big)$$

$$= E((Y-E(Y))^2 + a^2 E((X-E(X))^2 - 2a E((X-E(X))(Y-E(Y)))$$

$$= V(Y) + a^2 V(X) - 2a \operatorname{cov}(X,Y)$$

$\psi(a)$ est minimale quand $\psi'(a) = 0$ i.e. $a V(X) - \operatorname{cov}(X,Y) = 0$

$$\Rightarrow a_0 = \frac{\operatorname{cov}(X,Y)}{V(X)} = \rho_{XY} \times \frac{V(Y)^{1/2}}{V(X)^{1/2}}$$

$$\Rightarrow b_0 = E(Y) - a_0 E(X)$$

La v.A $\rho\left(\frac{V(Y)}{V(X)}\right)^{1/2} X + \left(E(Y) - a_0 E(X)\right)$ est la meilleur approximation de Y par une fct affine de X

$\Rightarrow$ Si on connait pas bien Y et qu'on pense qu'il y a une relation linéaire entre X et Y, on peut approcher Y par $\rho\left(\frac{V(Y)}{V(X)}\right)^{1/2} X + \left(E(Y) - a_0 E(X)\right)$

3- Généralisation à $X_1, \dots, X_n$ i.e déterminer $a_1, \dots, a_n, b$ qui minimisent $E\left((Y - a_1 X_1 - \dots - a_n X_n - b)^2\right)$

Remarque : (X,Y) vecteur aléatoire

* $E(Y)$ est la meilleure approximation par une constante de Y
  i.e minimise $E\left((Y-a)^2\right)$
* $E(Y) + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}\left(X - E(X)\right)$ est la meilleure approximation par une fonction linéaire de X, de Y

III - Transformations de vecteurs aléatoires

Rappel : Soit $Y = g(X)$ avec g une fonction connue.
On veut déterminer F_Y et f_Y à partir de F_X et f_X

Soit $g : \mathbb{R} \to \mathbb{R}$, $B \in \mathcal{B}(\mathbb{R})$, $g^{-1}(B) \subset \mathbb{R}$ l'ensemble pour lequel $g\left(g^{-1}(B)\right) = B$

alors $\mathbb{P}(Y \in B) = \mathbb{P}(g(X) \in B) = \mathbb{P}\left(X \in g^{-1}(B)\right)$

en particulier $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(X \in g^{-1}(]-\infty, y])\right)$

Cas des distributions à densité

Théorème de Transfert : On a une v.a à densité X, une fonction g.

Si pour toute fonction borélienne bornée h on a

$$E(h(g(X))) = \int h(u) f(u)\, du$$

alors f est la densité de $g(X) = Y$

Théorème : Cas bivarié

Soit $X = (X_1, X_2)$ un vecteur aléatoire de loi continue,

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = g(X) = \begin{pmatrix} g_1(X_1, X_2) \\ g_2(X_1, X_2) \end{pmatrix} \text{ où}$$

- Le système d'équations $\begin{cases} y_1 = g_1(x_1, x_2) \\ y_2 = g_2(x_1, x_2) \end{cases}$ peut être résolu pour tout (y_1, y_2) donnant comme solution $x_1 = h_1(y_1, y_2)$
$x_2 = h_2(y_1, y_2)$
- Les fonctions g_1 et g_2 sont continuement différentiable et de matrice Jacobienne

$$J(x_1, x_2) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{pmatrix}$$

alors $f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(x_1, x_2) \left| J(x_1, x_2) \right|^{-1} \Big|_{(x_1, x_2) = h(y_1, y_2)}$

Exemple : Soit (X, Y) un couple de densité

$$f(x,y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

c'est à dire que $X \sim N(0,1)$ et X et Y indépendants

Puis changement de variable

$$X = \underbrace{R\cos\theta}_{h_1(R,\theta)} \quad \text{et} \quad Y = \underbrace{R\sin\theta}_{h_2(R,\theta)}$$ dans le théorème

Cette transformation est bijective de $\mathbb{R}^2 \setminus \{0,0\}$ dans $\mathbb{R}_+^* \times [0, 2\pi[$

$$\operatorname{Jac}(R,\theta) = \begin{pmatrix} \cos\theta & -R\sin\theta \\ \sin\theta & R\cos\theta \end{pmatrix}$$

et le Jacobien vaut $|\operatorname{Jac}(R,\theta)| = R$

Donc la densité de (R,θ) vaut

$$f_{R,\theta}(R,\theta) = \frac{R}{2\pi} e^{-\frac{R^2}{2}} \mathbb{1}_{[0,2\pi[}(\theta)\, \mathbb{1}_{\mathbb{R}_+^*}(R)$$

$$= \underbrace{\frac{1}{2\pi} \mathbb{1}_{[0,2\pi[}(\theta)}_{\substack{\text{distribution uniforme} \\ \text{sur } [0,2\pi[}} \times \underbrace{R e^{-R^2/2} \mathbb{1}_{\mathbb{R}_+^*}(R)}_{\substack{\text{distribution de} \\ \text{Rayleigh}}}$$

Le vecteur (X,Y) (qui est gaussien) est un vecteur de direction θ choisi uniformément sur $[0;2\pi[$ et de Rayon R suivant une distribution de Rayleigh

<u>A quoi ça sert ?</u> Simulation de v.a

Si $X \sim F_X$ avec F_X inversible alors $F_X(x) \sim \mathcal{U}[0,1]$

donc si on prend une loi $U \sim \mathcal{U}[0,1]$ alors $F_X^{-1}(U) \sim F_X$

Donc si on simule des obs $U_1, \dots, U_n$ suivant une $\mathcal{U}[0,1]$

alors $F_X^{-1}(U_1), \dots, F_X^{-1}(U_n)$ sont des réalisations issues d'une v.a $X \sim F_X$. $\mathbb{P}(F_X(X) \leq u) = \mathbb{P}(X \leq F^{-1}(u)) = F_X(F_X^{-1}(u)) = u$

In other words :

Si X une v.A de distribution $\mathbb{P}_X$ et de f.d.r F_X ($X \sim \mathbb{P}_X$) alors $F_X(X) \sim \mathcal{U}[0,1]$ et donc si $U \sim \mathcal{U}[0,1]$ alors $F_X^{-1}(U) \sim \mathbb{P}_X$

Soit la f.d.r de $F_X^{-1}(U)$ est F_X

Par contre on ne sait pas inverser la f.d.r d'une loi $\mathcal{N}(0,1)$.

Or, si on a (U_1, U_2) deux v.A indep tq. $U_1 \sim \mathcal{U}[0,1]$ et $U_2 \sim \mathcal{U}[0,1]$ on pose

$$\underbrace{R = \sqrt{-2\ln(1-U)}}_{\substack{F_X(x) = e^{-R^2/2} = u \\ F_R^{-1}(u) = \sqrt{-2\ln(1-u)}}} \quad \text{et} \quad \underbrace{\theta = (2\pi)\, U_2}_{\text{suit loi } \mathcal{U}[0,2\pi]}$$

et donc $R\cos\theta \sim \mathcal{N}(0,1)$ et $R\sin(\theta) \sim \mathcal{N}(0,1)$ independantes

Somme de variables aléatoires indépendantes

Théorème : Soit X et Y deux v.A independantes et $S = X+Y$ alors la densité de S est le produit de convolution $f_X \star f_Y$ des densités f_X et f_Y

$$f_S(s) = f_X \star f_Y\,(s) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} f_X(x)\, f_Y(s-x) \\ \displaystyle\sum_{x \in \mathrm{Supp}(X,Y)} f_X(x)\, f_Y(s-x) \end{cases}$$

$\leftarrow \{(x,y) ; f_X(x) > 0 \text{ et } f_Y(y) > 0\}$

Chapitre 2 : fonction caractéristiques

La fonction caractéristique d'une V.A est une fonction de $\mathbb{R}$ ou $\mathbb{R}^n$ dans $\mathbb{C}$ définie à partir de la transformée de Fourier de sa distribution.

4 propriétés :

- Injectivité : on peut caractériser une loi par sa f.c
- La f.c d'une somme de V.A indep est le produit des f.c
- On a une formule d'inversion. A partir de la f.c on peut retrouver la densité ou la f.d.r
- On peut retrouver tous les moments en dérivant la f.c

I - Transformée de Fourier - fonction caractéristique

Definition : Soit $\mathbb{P}$ une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
On définit la transformée de Fourier de la mesure $\mathbb{P}$ par la fonction $\phi : \mathbb{R} \to \mathbb{C}$ tq

$$\phi(t) = \int e^{itx} \, d\mathbb{P}(x)$$

Definition : Soit X une V.A. Réelle. On définit la f.c de X comme la transformée de Fourier de $\mathbb{P}_X$ (mesure image de $\mathbb{P}$ par X) et on note

$$\phi_X(t) = \int e^{itx} \, d\mathbb{P}_X(x)$$

En particulier si X est abs. cont. par rapport à λ, de densité f_X on a $\phi_X(t) = \int e^{itx} f_X(x)\,dx$

Si $X = (x_1, \dots, x_n)$ est un vecteur aléatoire, la f.c est définie par :

$$\forall t = (t_1, \dots, t_n) \in \mathbb{R}^n, \quad \phi_X(t) = E\left(e^{i\langle t, x\rangle}\right) = E\left(e^{i\sum_{j=1}^{n} t_j x_j}\right)$$

Propriétés :

- $\phi_X(0) = 1$
- $|\phi_X(t)| \leq 1$
- ϕ_X est unif continue
- $\forall a, b \in \mathbb{R} \qquad \phi_{aX+b}(t) = e^{itb}\phi_X(at)$ en dimension 1

  $\forall A \in \mathcal{M}_{k\times m}(\mathbb{R}),\ B \in \mathbb{R}^k, \quad \phi_{AX+B}(t) = e^{i\langle t, B\rangle}$

II - Utilité de la f.c

Théorème : Injectivité

Soit $\mathbb{P}$ et Q deux mesures de proba sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si $\forall t \in \mathbb{R}$

$$\int e^{itx}\,d\mathbb{P}(x) = \int e^{itx}\,dQ(x)$$

alors $\mathbb{P} = Q$

Corrolaire : Si X et Y sont deux v.A sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $\forall t \in \mathbb{R}$ $\quad \phi_X(t) = \phi_Y(t)$ alors $\mathbb{P}_X = \mathbb{P}_Y$

Conséquence : La fonction caractéristique caractérise entièrement la distribution

Il y a autant d'info sur la loi de X dans sa f.d.r ou sa f.c

Corollaire 2 : Si X et Y sont deux vecteurs dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ et $\forall t \in \mathbb{R}^n$ $\quad \phi_X(t) = \phi_Y(t)$ alors $\mathbb{P}_X = \mathbb{P}_Y$

Proposition : Soit X une v.A à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ alors $\mathbb{P}_X = \mathbb{P}_{-X} \iff \phi_X$ est réelle

Théorème : (formule d'inversion)

Soit $\mathbb{P}$ une proba sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ de f.c $\phi(t) = \int e^{itx} d\mathbb{P}(x)$. On note $F(x) = \mathbb{P}(]-\infty, x])$ sa f.d.r. Alors si F est cont en a et b on a

$$F(b) - F(a) = \frac{1}{2\pi} \lim_{c \to \infty} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \phi(t)\, dt$$

On aura jamais besoin de calculer ça en exo.

Remarque : On a des résultats analogues pour la densité (Fast Fourier Transform).

Théorème : La transformée de Fourier du produit de convolution de deux mesures μ_1 et μ_2 sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est le produit des transformées de Fourier

$\Rightarrow$ La f.c de la somme de 2 v.A indep est le produit des f.c. $\forall t \in \mathbb{R}^d \quad \phi_{X+Y}(t) = \phi_X(t)\,\phi_Y(t)$

Théorème : Critère d'indépendance

Soit $X = (X_1, X_2)$ un v.A à valeurs dans $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$

Alors X_1 et X_2 sont indépendants ssi

$$\phi_{(X_1,X_2)}(t_1, t_2) = \phi_{X_1}(t_1)\,\phi_{X_2}(t_2)$$

Définition - Propriété : Soit $X = (X_1, X_2)$ à valeurs dans $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. On définit les f.c marginales de X_1 et X_2 par

$\forall t_1 \in \mathbb{R}^{d_1}$

$$\boxed{\begin{gathered} \phi_{X_1}(t_1) = \phi_{(X_1,X_2)}(t_1, 0) \\ \phi_{X_2}(t_2) = \phi_{(X_1,X_2)}(0, t_2) \end{gathered}}$$

Conséquence : On définit un autre critère d'indépendance
X_1 et X_2 indépendants ssi

$$\phi_{(X_1,X_2)}(t_1,t_2) = \phi_{(X_1,X_2)}(t_1,0)\cdot\phi_{(X_1,X_2)}(0,t_2)$$

Rappel : fonction caractéristique

- Transformée de Fourier d'une mesure $\mathbb{P} = \int e^{itx}\,d\mathbb{P}(x)$
- Fc de X = Transf de Fourier de $\mathbb{P}_X$

$$\phi_X = \int e^{itx}\,d\mathbb{P}(x) = E^{\mathbb{P}}(e^{itx}) \text{ si } X \text{ dans } (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

si $X = (X_1, \dots, X_n)$, $\quad \phi_X(\underbrace{t}_{\in\mathbb{R}^n}) = E^{\mathbb{P}}(e^{i\langle t,x\rangle}) = E^{\mathbb{P}}(e^{i\sum_j^n t_j x_j})$

Remarque : $\phi_X(t) = \phi_{\underbrace{t}_{\text{transposé}} X}(1)$

Convolution : $\phi_{X+Y}(t) = \phi_X(t)\cdot\phi_Y(t)$ si X et Y indep

Indépendance : $X \perp Y$, $\phi_{(X,Y)}(t_1,t_2) = \phi_X(t_1)\cdot\phi_Y(t_2) = \phi_{(X,Y)}(t_1,0)\cdot\phi_{(X,Y)}(0,t_2)$

Inversion : A partir de ϕ on peut retrouver F ou f

Moments : A partir de ϕ on peut trouver tous les moments de X

Théorème : Si X est une v.A de f.c ϕ_X

- Si X admet des moments d'ordre n, alors ϕ_X est de classe $\mathcal{C}^n$ et

$$\forall 1 \leq k \leq n, \quad \forall t \in \mathbb{R}, \quad \phi_X^{(k)}(t) = i^k \int x^k e^{itx} \, d\mathbb{P}_X(x)$$

En particulier $\phi_X^k(0) = i^k E^{\mathbb{P}}(X^k)$

- inversement, si ϕ_X est n fois dérivable alors X admet des moments d'ordre $2\left[\frac{n}{2}\right]$

Exemple : Soit (X_1, X_2) deux v.a.r indépendantes de loi de Laplace i.e de densité

$$f_X(x) = \frac{1}{2} e^{-|x|}$$

On pose $Y_1 = X_1 - X_2$ et $Y_2 = X_1 + X_2$

1- f.c de X_1, X_2 et X

2- f.c de Y_1 et Y_2

3- Montrer que Y_1 et Y_2 ont même loi

4- Mq $cov(Y_1, Y_2) = 0$

5- Y_1 et Y_2 indépendants ?

1) X_1 a même distribution que X_2 donc même f.c

$$\phi_{X_1}(t) = E(e^{itX}) = \int e^{itx} \frac{1}{2} e^{-|x|} \, dx$$

$$= \frac{1}{2}\left[\int_{-\infty}^{0} e^{x(1+it)}\,dx + \int_{0}^{+\infty} e^{x(it-1)}\,dx\right]$$

$$= \frac{1}{2}\left(\left[\frac{e^{x(1+it)}}{1+it}\right]_{-\infty}^{0} + \left[\frac{e^{x(it-1)}}{it-1}\right]_{0}^{+\infty}\right) = \frac{1}{2}\left(\frac{1}{it+1} + \frac{1}{1-it}\right)$$

$$\phi_{X_1}(t) = \frac{1}{1+t^2}$$

Comme X_1 et X_2 sont indep,

$$\phi_X(t_1,t_2) = \phi_{(X_1,X_2)}(t_1,t_2) = \phi_{X_1}(t_1)\,\phi_{X_2}(t_2) =$$

$$= \frac{1}{1+t_1^2}\,\frac{1}{1+t_2^2}$$

2) On a $Y_1 = X_1 - X_2$ et $Y_2 = X_1 + X_2$

Rappel : Si X est un vecteur aléatoire à valeurs dans $\mathbb{R}^n$

$\forall A \in \mathcal{M}_{k\times n}(\mathbb{R})$ et $\forall B \in \mathbb{R}^k$ alors

$$\phi_{AX+B}(t) = e^{i\langle t,B\rangle}\,\phi_X(A^t \cdot t)$$

Ici on a $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

$A^t = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

$A^t\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{matrix} t_1+t_2 \\ t_1-t_2 \end{matrix}$ et $B=0$ donc

$$\phi_{(Y_1,Y_2)}(t_1,t_2) = \frac{1}{1+(t_1+t_2)^2}\cdot\frac{1}{1+(t_2-t_1)^2}$$

$$\left.\begin{aligned} \phi_{Y_1}(t) &= \phi_{(Y_1,Y_2)}(t,0) = \frac{1}{(1+t^2)^2} \\ \phi_{Y_2}(t) &= \phi_{(Y_1,Y_2)}(0,t) = \frac{1}{(1+t^2)^2} \end{aligned}\right\} \phi_{Y_1}(t) = \phi_{Y_2}(t) \quad \forall t \in \mathbb{R}$$

donc Y_1 et Y_2 ont même loi

4) On veut calculer $\mathrm{cov}(Y_1, Y_2)$

On calcule : $E(Y_1 Y_2) = E((X_1 - X_2)(X_1 + X_2))$
$= E(X_1^2) - E(X_2^2) = 0$ car X_1 et X_2 ont même loi

$E(Y_2) = E(Y_1) = E(X_1) - E(X_2) = 0$

Donc $\mathrm{cov}(Y_1, Y_2) = E(Y_1 Y_2) - E(Y_1)E(Y_2) = 0$

5) Par contre Y_1 et Y_2 ne sont pas indépendants car

$$\phi_{(Y_1, Y_2)}(t_1, t_2) = \frac{1}{1+(t_1+t_2)^2} \cdot \frac{1}{1+(t_2-t_1)^2} \neq \frac{1}{(1+t_1^2)^2} \frac{1}{(1+t_2^2)^2} = \phi_{Y_1}(t_1) \cdot \phi_{Y_2}(t_2)$$

Chapitre III : Vecteurs gaussiens

I - Définition: Une v.a X réelle définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ est dite Gaussienne ou Normale si sa loi image $\mathbb{P}_X$ a une densité par rapport à la mesure de Lebesgue de la forme

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \text{avec } \mu \in \mathbb{R} \text{ et } \sigma \in \mathbb{R}_+^*$$

On note $X \sim \mathcal{N}(\mu, \sigma^2)$ ou $\mathbb{P}_X = \mathcal{N}(\mu, \sigma^2)$

Si $\mu = 0$ et $\sigma^2 = 1$ on dit que X est une loi Gaussienne centrée réduite

Propriétés :

- Si $X \sim \mathcal{N}(0,1)$ alors $\phi_X(t) = e^{-\frac{t^2}{2}}$
- Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $\phi_X(t) = e^{it\mu} e^{-\frac{t^2\sigma^2}{2}}$
- Si $X \sim \mathcal{N}(0,1)$, alors $aX+b \sim \mathcal{N}(b, a^2)$ $\forall a,b \in \mathbb{R}$

Rappel : Si X admet la f.c $\phi_X(t)$ alors $\phi_{aX+b} = e^{itb}\phi(ta)$

Ici:

$$\phi_{aX+b}(t) = e^{itb}\phi_X(at) = e^{itb} e^{-\frac{a^2t^2}{2}} \Rightarrow aX+b \sim \mathcal{N}(b, a^2)$$

- Si X_1 et X_2 sont indépendants et $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$

alors $\phi_{X_1+X_2}(t) = \phi_{X_1}(t) \cdot \phi_{X_2}(t) = e^{itm_1} e^{\frac{-t^2\sigma_1^2}{2}} e^{itm_2} e^{-\frac{t^2\sigma_2^2}{2}}$

$$= \underbrace{e^{it(m_1+m_2)} e^{-\frac{t^2}{2}(\sigma_1^2+\sigma_2^2)}}_{\text{f.c d'une } \mathcal{N}(m_1+m_2,\ \sigma_1^2+\sigma_2^2)}$$

$$\Rightarrow X_1 + X_2 \sim \mathcal{N}(m_1+m_2,\ \sigma_1^2+\sigma_2^2)$$

- Si $X \sim \mathcal{N}(0,1)$ alors on a

$$E(X^n) = \begin{cases} 0 & \text{si } r \text{ est impair} \\ (r-1)(r-3) & \text{si } r \text{ est pair} \end{cases}$$

Vecteurs Gaussiens : plusieurs caractéristiques

Definition : Caractérisation par sa loi

$X = (X_1, X_2)$ un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, X est un vecteur gaussien si il admet comme densité par rapport à Lebesgue sur $\mathbb{R}^n$

$$f(x) = \frac{1}{(2\pi)^{\frac{1}{2}} (\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2} {}^t(x-\mu) \Sigma^{-1} (x-\mu)\right)$$

avec $\mu \in \mathbb{R}^n$ et Σ une matrice réelle symétrique définie positive

On a $\mu = E(X)$ et Σ est la matrice de variance

Definition : Caractérisation par stabilité par combinaison linéaire

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire réel

X est Gaussien ssi $\forall \alpha \in \mathbb{R}^n \quad {}^t\alpha X = \sum_{i=1}^{n} \alpha_i X_i$ est Gaussienne

on considère que les constantes sont des Gaussienne (dégénérées) de variance nulle

Remarque : Il y a une (grosse) différence entre un Vecteur Gaussien et un vecteur composé de variables Gaussiennes. Par contre si $X = (X_1, \dots, X_n)$ est Gaussien alors $\forall i = 1, \dots, n \quad X_i$ est Gaussien

(on utilise la stabilité par C.L avec $\alpha = (0, \dots, 0, 1, 0, 0, \dots, 0)$
i = position

X Gaussien $\Rightarrow$ X_i est Gaussien $\forall i = 1, \dots, n$

$\Leftarrow$ Faux

Exemple : $X \sim \mathcal{N}(0,1)$. $\mathcal{E}$ est une v.a indep de X tq $\mathbb{P}(\mathcal{E} = 1) = \mathbb{P}(\mathcal{E} = -1) = \frac{1}{2}$

On pose $Y = \mathcal{E} X$. On veut la loi de Y

$$\phi_Y(t) = \phi_{\mathcal{E}X}(t) = E\left(e^{it\mathcal{E}X}\right) = \iint_{\mathbb{R}^2} e^{ityx} \overbrace{d\mathbb{P}_{(\mathcal{E},X)}(y,x)}^{d\mathbb{P}_{\mathcal{E}}(y)\, d\mathbb{P}_X(x) \ \perp}$$

$$= \int_{\mathbb{R}} e^{itx}\, \mathbb{P}(\mathcal{E} = 1)\, d\mathbb{P}_X(x) + \int_{\mathbb{R}} e^{-itx}\, \mathbb{P}(\mathcal{E} = -1)\, d\mathbb{P}_X(x)$$

$$\phi_Y(t) = \frac{1}{2}\int e^{itx}\, d\mathbb{P}_X(x) + \frac{1}{2}\int e^{-itx}\, d\mathbb{P}_X(x)$$

$$= \frac{1}{2}\phi_X(t) + \frac{1}{2}\phi_{-X}(t) = \phi_X(t)$$

Donc $\mathcal{E}X$ et X ont même distribution et $\mathcal{E}X \sim \mathcal{N}(0,1)$

Or $(X, \mathcal{E}X)$ n'est pas un vecteur Gaussien

$1^{ère}$ méthode : $\phi_{X+\mathcal{E}X}(t) \neq$ f.c d'une loi normale

$2^{ème}$ méthode : Stabilité par combinaison linéaire

$(X, \mathcal{E}X)$ n'est pas gaussien car $X + \mathcal{E}X$ n'est pas Gaussien

$\mathbb{P}(X+\mathcal{E}X=0) = \mathbb{P}(\mathcal{E}=-1) = \frac{1}{2}$ or si $Z \sim \mathcal{N}(\mu, \sigma^2)$ alors

$\mathbb{P}(Z=z) = 0 \quad \forall z \in \mathbb{R}$ ou $\mathbb{P}(Z=z)=1$ si $\sigma^2=0$

et $E(Z) = z$

Proposition : Si X_1 et X_2 sont deux v.a Gaussiennes indépendantes alors (X_1, X_2) est un vecteur Gaussien $\forall a,b,c,d \in \mathbb{R}$

Définition : Caractérisation par la f.c

$X = (X_1, \dots, X_n)$ est un vecteur Gaussien si et seulement il existe un matrice Σ (symétrique, définie positive) (Matrice de Variance) et un vecteur $\mu \in \mathbb{R}^n$ tq la f.c s'écrit
$$\phi_X(t) = e^{i\langle t,\mu\rangle} e^{-\frac{{}^t t \Sigma t}{2}}$$

Rappel : Vecteurs Gaussiens.

$X = (X_1, \dots, X_n)$ est un vecteur gaussien

1) Densité sous la forme

$$f_X(x) = \frac{1}{(2\pi)^{n/2} (\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2} {}^t(x-\mu) \Sigma^{-1} (x-\mu)\right)$$

avec μ = vecteur des espérances = $\begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$ et $\Sigma = (\mathrm{cov}(X_i, X_j))_{i,j=1,\dots,n}$

matrice de cov

2) Stabilité par combinaison linéaire

$\forall \alpha_1, \dots, \alpha_n \in \mathbb{R}$, $\sum_{i=1}^{n} \alpha_i X_i$ est une v.a Gaussienne.

3) Fonction Caractéristique

$$\phi_X(t) = e^{i\langle t, \mu \rangle - \frac{1}{2} {}^t t \Sigma t}$$

pour montrer que c'est Gaussien

Conséquences

- Si $X = (X_1, \dots, X_n)$ est Gaussien alors $\forall i = 1, \dots, n$ X_i est Gaussien

  Réciproque fausse

- Toute la distribution est cardérisée par μ (le vecteur d'espérence) et Σ la matrice de covariance

  contient toute l'info sur la dépendance entre les X_i

Théorème de Sklar

$X = (X_1, \dots, X_d)$ de f.d.r F_X et de f.d.r marginales F_{X_i} alors il existe une fonction C de répartition sur $[0,1]^d$ de lois marginales uniformes sur $[0,1]$ tq

$$F_X(x_1, \dots, x_d) = C(F_{X_1}(x_1), \dots, F_{X_d}(x_d))$$

La fonction C est la fonction Copule ou la Copule. C'est elle qui définit la dépendance entre les X_i ou qui contient toute l'information sur la dépendance entre les X_i.

Propriétés :

$C(u_1, \dots, u_d) = 0$ si un des u_i vaut 0

$C(1, \dots, u_i, 1, \dots, 1) = u_i$

C est d croissante (croissante sur chacune des coordonnées)

- Si les marginales F_{X_i} sont continues alors C est unique
- Dans les vecteurs Gaussiens, toute la structure de dépendance est contenue dans Σ.

$$\Sigma \iff \text{Copule (dans les vecteurs Gaussiens)}$$

Théorème (indépendance des vecteurs Gaussiens)

Soit $X = (X_1, \dots, X_n)$ un vecteur Gaussien. Les composantes de X sont indépendantes ssi elles sont non corrélées.

i.e ssi $\forall j \neq k$ $\quad cov(X_j, X_k) = 0$. autrement dit ssi la matrice de covariance est diagonale

Preuve :

- Indépendance $\Rightarrow cov(X_j, X_k) = 0$ triviale
- On considère un vecteur gaussien avec une matrice de covariance Σ diagonale

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix} \text{ et } \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

Chaque X_i est tq $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$

Si on a $\phi_X(t_1, \dots, t_n) = \prod_{i=1}^n \phi_{X_i}(t_i)$ alors les X_i sont indépendants

On a :

$$\phi_X(t_1, \dots, t_n) = e^{i\sum_{j=1}^n t_j\mu_j - \frac{1}{2}\underbrace{{}^t t \Sigma t}}$$

$$(t_1, \dots, t_n)\begin{pmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n^2} \end{pmatrix}\begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} = \sum \frac{t_j^2}{\sigma_j}$$

$$\phi_X(t_1, \dots, t_n) = e^{i\sum_{j=1}^n t_j\mu_j} e^{-\frac{1}{2}\sum_{j=1}^n \sigma_j^2 t_j^2}$$

$$= e^{it_1\mu_1} e^{-\frac{\sigma^2 t_1^2}{2}} \times \dots \times e^{it_n\mu_n - \frac{\sigma_n^2 t_n^2}{2}}$$

$= \phi_{X_1}(t_1) \times \dots \times \phi_{X_n}(t_n)$ avec ϕ_{X_i} f.c d'une $\mathcal{N}(\mu_i, \sigma_i^2)$

alors les X_i sont indépendants

Remarque : C'est la seule distribution pour laquelle on a équivalence entre indépendance et covariance nulle

Remarque : Si $X = (X_1, \dots, X_d)$ est Gaussien alors si $\exists i, j$ tq $cov(X_i, X_j) = 0$ on a X_i indépendant de X_j

Remarque : Ça marche parce que X est un vecteur Gaussien et pas parce que les X_i sont Gaussiennes

Exemple : Soit $X \sim \mathcal{N}(0,1)$ et $\mathcal{E}$ tq $\mathbb{P}(\mathcal{E}=1) = \mathbb{P}(\mathcal{E}=-1) = \frac{1}{2}$ et X indep de $\mathcal{E}$. On pose $Y = \mathcal{E}X$

On a vu que $Y \sim \mathcal{N}(0,1)$ mais (X,Y) n'est pas Gaussien

On a

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = E(\mathcal{E}X^2) - E(X)E(\mathcal{E}X)$$
$$= E(\mathcal{E})E(X^2) - E(\mathcal{E})E(X)^2 = 0$$

On a $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$ et $\mathrm{cov}(X,Y) = 0$ mais X et Y ne sont pas indépendants.

Si X et Y étaient indépendants alors si g est une fct mesurable, $g(X)$ et $g(Y)$ seraient indépendants. En particulier si on prend $g(u) = u^2$ on aura X^2 indep de $Y^2 = (\mathcal{E}X)^2 = X^2 \Rightarrow$ impossible

Exemple : On prend une loi normale centrée réduite i.e $X \sim \mathcal{N}(0,1)$

On prend $a \in \mathbb{R}_+^*$ et on définit

$$Y_a = -X \mathbb{1}_{\{|X| \leq a\}} + X \mathbb{1}_{\{|X| > a\}}$$

1) Mq Y_a a la même distribution que X (ça marche si la distribution de X est symétrique)

2) Mq (X, Y_a) n'est pas un vecteur Gaussien

3) $\exists a$ tq $\mathrm{cov}(X, Y_a) = 0$

1. Loi de Y_a : Théorème de transfert i.e si $E(g(Y_a)) = \int g(y) f(y)\, dy$ alors f est la densité de Y

$$E(g(Y_a)) = \int_{|x|\leq a} g(-x) f_X(x)\, dx + \int_{|x|>a} g(x) f_X(x)\, dx$$

$$= \int_{|y|\leq a} g(y) f_X(y)\, dy + \int_{|x|>a} g(x) f_X(x)\, dx = \int g(y) f_X(y)\, dy$$

Donc X et Y_a ont même distribution i.e $Y_a \sim \mathcal{N}(0,1)$

2. On veut montrer qu'on n'a pas (X, Y_a) Gaussien

On a $X + Y_a$ n'est pas Gaussien car

$$\mathbb{P}(X + Y_a = 0) = \mathbb{P}(|X| \leq a) = \int_{-a}^{a} f_X(x)\, dx \in]0,1[$$

$\Rightarrow X + Y_a$ pas Gaussien $\Rightarrow (X, Y_a)$ pas Gaussien

3. Mq $\exists a$ tq $\operatorname{cov}(X, Y_a) = 0$

$$\operatorname{Cov}(X, Y_a) = E(XY_a) - E(X)E(Y_a) = E(XY_a) = \underbrace{E(X^2(2\mathbb{1}_{\{|X|>a\}} - 1))}_{E(2X^2\mathbb{1}_{\{|X|>a\}} - X^2)}$$

$$= \underbrace{2E(X^2\mathbb{1}_{\{|X|>a\}})}_{= 2\int_{|x|>a} x^2 f_X(x)\, dx} - \underbrace{E(X^2)}_{= 1 \text{ car } E(X^2) = V(X) = 1}$$

On a $\operatorname{cov}(X, Y_a) = 0$ ssi $\int_{|x|>a} x^2 f_X(x)\, dx = \frac{1}{2}$

Si $\phi(a) = \int_{|x|>a} x^2 f_X(x)\, dx$ alors ϕ est continu, décroissante (strict) tq $\phi(0) = 1$ et $\lim_{a \to +\infty} \phi(a) = 0$

$\Rightarrow \exists\, a$ tq $\int_{|x|>a} x^2 f(x)\,dx = \frac{1}{2}$ et donc $cov(X, Y_a) = 0$

II - transformation de loi Gaussienne

Loi du χ^2

Si $X \sim \mathcal{N}(0, \sigma^2)$ alors $X^2 \sim \Gamma\left(\frac{1}{2\sigma^2}, \frac{1}{2}\right)$

En particulier si $\sigma^2 = 1$ alors $X^2 \sim \chi^2_1$ (khi deux à 1 degré de liberté)

ie $\chi^2_1 = \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$

Si on a $X_1, \dots, X_n$ des v.a.i.i.d tq $X_i \sim \mathcal{N}(0,1)$
alors $\sum_{i=1}^{n} X_i^2 \sim \chi^2_n$ (khi 2 à n degrès de liberté)

Rappel : Si $Y_1, \dots, Y_n$ sont indep tq $Y_i \sim \Gamma(a_i, b)$
alors $\sum Y_i \sim \Gamma\left(\sum_{i=1}^{n} a_i, b\right)$
On a donc $\chi^2_i = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$

La densité de la χ^2_n est donné par

$$f(x) = \frac{\left(\frac{1}{2}\right)^{n/2} x^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} \mathbb{1}_{\{x>0\}}$$

Si $X \sim \chi^2_n$ alors $E(X) = n$ et $V(X) = 2n$

Loi Student

Soit X et Y independants tq $X \sim \mathcal{N}(0,1)$ et $Y \sim \chi^2_n$

alors $T_n = \dfrac{X}{\sqrt{Y/n}} \sim t_n$ (loi de Student à n degrès de liberté).

Loi de Fischer

Si X et Y independantes tq $X \sim \chi^2_n$ et $Y \sim \chi^2_m$

alors $F = \dfrac{X/n}{Y/m} \sim F(n,m)$ (loi de Fischer à n et m degrès de libertés)

Resumé :

- Loi Normale $\mathcal{N}(m, \sigma^2)$
  - $\mathcal{N}(m_1, \sigma_1^2) + \mathcal{N}(m_2, \sigma_2^2) = \mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$
  - $a\mathcal{N}(0,1) + b = \mathcal{N}(b, a^2)$

$\downarrow$

$\chi^2_1 = (\mathcal{N}(0,1))^2$ $\qquad \chi^2_n = \sum \mathcal{N}(0,1)$

- Khi Deux χ^2_n
  - $\chi^2_n + \chi^2_m = \chi^2_{n+m}$

$\downarrow$

$\chi^2_n = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$

- Loi Gamma $\Gamma(n,p)$
  - $\Gamma(a_1, b) + \Gamma(a_2, b) = \Gamma(a_1+a_2, b)$

$$\downarrow$$

$$\mathcal{E}(\lambda) = \Gamma(1, \lambda)$$

$$\downarrow$$

$$\mathcal{E}\left(\frac{1}{2}\right) = \chi^2_2 = \Gamma\left(1, \frac{1}{2}\right)$$

$$\downarrow$$

$$\frac{\chi^2_n / n}{\chi^2_m / m} = \mathcal{F}(n, m) \quad \text{loi de Fischer}$$

$$\downarrow$$

$$\mathcal{F}(1, n) = t^2_n \quad \text{loi de Student}$$

$$\text{et} \quad \frac{\mathcal{N}(0,1)}{\sqrt{\chi^2_n / n}} = t_n$$

Application de la Statistique Gaussienne

Soit $X_1, \dots, X_n$ des v.a. i.i.d tq $X_i \sim \mathcal{N}(m, \sigma^2)$

On pose $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ et $V_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

Alors

1 - $\bar{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$

2 - $\frac{n}{\sigma^2} V_n \sim \chi^2_{n-1}$

3 - $\bar{X}_n$ et V_n sont indépendants

4 - $\dfrac{\bar{X}_n - m}{\sqrt{V_n/(n-1)}} \sim t_{n-1}$

Preuves

1) $X_1, \dots, X_n$ v.a.i.i.d tq $X_i \sim \mathcal{N}(m, \sigma^2)$

$\Rightarrow \frac{1}{n} \sum_{i=1}^{n} X_i$ est Gaussienne car combinaison linéaire de v.a Gaussienne indep

$$\left.\begin{aligned} E(\bar{X}_n) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = m \\ V(\bar{X}_n) &= V\left(\frac{1}{n}\sum_i X_i\right) = \frac{1}{n^2}\sum_i V(X_i) = \frac{\sigma^2}{n} \end{aligned}\right\} \bar{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$$

2) On veut la loi de $\frac{n}{\sigma^2} V_n = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X}_1)^2}{\sigma^2} = \sum_{i=1}^{n} \left(\dfrac{X_i - \bar{X}_n}{\sigma}\right)^2$

On a $\sum_{i=1}^{n} \Big(\underbrace{\dfrac{X_i - m}{\sigma}}_{\mathcal{N}(0,1)}\Big)^2 \sim \chi^2_n$

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

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

$$\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i - m)^2}_{\sim \chi^2_n} = \frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i - \bar{X}_n) + \underbrace{\frac{n}{\sigma^2}(\bar{X}_n - m)^2}_{\sim \chi^2_1}$$

Si $\sum (X_i - \bar{X}_n)^2$ est indep de $(\bar{X}_n - m)^2$ alors $\frac{1}{\sigma^2} \sum (X_i - \bar{X})^2 \sim \chi^2_{n-1}$

Or $(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n, \bar{X}_n - m)$ est un vecteur Gaussien car toute CL d'éléments du vecteur s'écrit comme CL de $X_1, \dots, X_n$

On calcule

$$\mathrm{cov}(X_i - \bar{X}_n, X_n) = \mathrm{cov}(X_i, \bar{X}_n) - \mathrm{cov}(\bar{X}_n, \bar{X}_n)$$
$$= \frac{1}{n} \sum_{j=1}^{n} \mathrm{cov}(X_i, X_j) - V(\bar{X}_n) = 0$$

$\Rightarrow \forall i = 1, \dots, n \quad X_i - \bar{X}_n$ est indep de $\bar{X}_n - m$

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

Séance 8 février

Rappel : Vecteurs Gaussiens

$X = (X_1, \dots, X_n)$

3 caractérisation :

- Densité : $f(x) = \frac{1}{(\det \Sigma)^{1/2} (2\pi)^{n/2}} \exp(\dots)$
- $\forall \alpha_1, \dots, \alpha_d \in \mathbb{R}$ $\quad \sum \alpha_i X_i$ est Gaussienne (en dimension 1) — Pour montrer que pas Gaussien
- Fonction caractéristique $\phi(t) = e^{i\langle t, \mu \rangle - \frac{1}{2} t^t \Sigma t}$ — Pour montrer que oui

> **Proposition**
>
> Si $X = (X_1, \dots, X_d)$ est vecteur Gaussien
>
> $\mathrm{cov}(X_i, X_j) = 0 \iff X_i$ indep de X_j

Remarque : C'est le seul cas où on a ça

Chapitre IV : Une Convergence de variables aléatoires

On a $(X_n)_n$ une suite de v.a.

On distingue deux façons de voir les v.a :

1 - Une v.a X définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ peut être vue comme une fonction (mesurable) de Ω dans $\mathbb{R}$. — CV presque sûre et en proba

2 - Une v.a peut être vue comme un objet mathématique caractérisé par $\mathbb{P}_X$ i.e par sa distribution sur $\mathbb{R}$ $\Rightarrow$ convergence en distribution

⚠ 1 et 2 très séparés

I - Convergence presque sûre et en probabilité

Définition : Soit $(X_n)_n$ une suite de v.a définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$

- On dit que X_n converge presque sûrement vers X (et on note $X_n \xrightarrow{p.s} X$) ssi

$$\underbrace{\mathbb{P}(\omega : X_n(\omega) \longrightarrow X(\omega)) = 1}$$

L'ensemble des points de Ω pour lesquels on n'a pas la convergence est négligeable

- On dit que X_n converge en probabilité vers X (noté $X_n \xrightarrow{\mathbb{P}} X$) ssi : $\forall \varepsilon > 0 \quad \mathbb{P}(\| X_n - X \| > \varepsilon) \longrightarrow 0$

ou $\quad \mathbb{P}(|X_n - X| < \varepsilon) \longrightarrow 1$

Remarque : On note $|x|$ la norme euclidienne de x i.e $\left(\sum_{i=1}^{d} x_i^2\right)^{1/2}$ si $x = (x_1, \dots, x_d)$

Remarque : Si on enlève un ensemble négligeable (celui pour lequel on n'a pas la convergeance), on peut dire que $X_n \xrightarrow{p.s} X$ ssi X_n converge ponctuellement vers X en tant que fonctions de Ω ds $\mathbb{R}^d$

Rappel : Soit $(A_n)_n$ une suite de parties de Ω. On définit les deux événements limites suivants :

- $\limsup_n A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{p \geq n} A_p = \overline{\lim} A_n$
- $\liminf_n A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{p \geq n} A_p = \underline{\lim} A_n$

L'ensemble $\limsup A_n$ représente l'ensemble des $\omega \in \Omega$ qui sont dans une infinité de A_n et $\liminf A_n$ represente l'ensemble des $\omega \in \Omega$ qui, à partir d'un certain rang sont dans tous les A_n

Propriétés :

- $\liminf A_n \subset \limsup A_n$
- $(\limsup A_n)^c = \liminf A_n^c$
- $(\liminf A_n)^c = \limsup A_n^c$

Lemme de Borel-Cantelli

Soit $(A_n)_n$ une suite d'événements de Ω alors on a :

- Si $\sum_n \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}(\limsup A_n) = 0$ $\otimes$
- Si les $(A_n)_n$ sont indépendants et $\sum \mathbb{P}(A_n) = \infty$ alors $\mathbb{P}(\limsup A_n) = 1$

Lemme : Soit $(X_n)_n$ une suite de v.a sur $(\Omega, \mathcal{A}, \mathbb{P})$ alors

$$X_n \xrightarrow{P.S} X \iff \mathbb{P}(\limsup \{|X_n - X| > \varepsilon\}) = 0 \quad \otimes$$

Preuve : $X_n \xrightarrow{P.S} X \iff \forall \varepsilon > 0 \;\exists n \text{ tq } \forall k \geq n \;\; |X_k - X| < \varepsilon$

$$X_n(\omega) \longrightarrow X(\omega) \Leftrightarrow \omega \in \left\{ \bigcup_n \bigcap_{k \geq n} |X_k(\omega) - X(\omega)| < \varepsilon \right\} \quad \forall \varepsilon > 0$$

Rappel : $X_n \xrightarrow{P.S} X$ ssi $\mathbb{P}(X_n(\omega) \longrightarrow X(\omega)) = 1$

$\Rightarrow X_n \xrightarrow{P.S} X$ ssi $\mathbb{P}\left(\bigcup_n \bigcap_{k \geq n} \{|X_k - X| < \varepsilon\}\right) = 1 \quad \forall \varepsilon > 0$

ou $\mathbb{P}\left(\bigcap_n \bigcup_{k \geq n} \{|X_k - X| \geq \varepsilon\}\right) = 0$

ou $\mathbb{P}\left(\limsup \{|X_k - X| \geq \varepsilon\}\right) = 0$

Lemme de Borel-Cantelli V2

Soit $(X_n)_n$ une suite de V.A sur $(\Omega, \mathcal{A}, \mathbb{P})$

- Si $\forall \varepsilon > 0 \quad \sum \mathbb{P}(|X_n - X| > \varepsilon) < \infty$ alors $X_n \xrightarrow{P.S} X$
- Si les $(X_n)_n$ sont indépendants alors

$$X_n \xrightarrow{P.S} 0 \Rightarrow \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| > \varepsilon) < \infty \quad \forall \varepsilon > 0$$

Rappel : $X_n \xrightarrow{\mathbb{P}} X \Leftrightarrow \forall \varepsilon > 0 \quad \mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0$

Théorème : Soit $(X_n)_n$ une suite de V.A sur $(\Omega, \mathcal{A}, \mathbb{P})$ alors

$$X_n \xrightarrow{P.S} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$$

Preuve : $X_n \xrightarrow{P.S} X \Rightarrow \mathbb{P}\left(\bigcup_{k \geq n} \{|X_k - X| > \varepsilon\}\right) \to 0$

or $\underbrace{\mathbb{P}\left(\bigcup_{k \geq n} \{|X_k - X| > \varepsilon\}\right)}_{\downarrow \, 0} \geq \underbrace{\mathbb{P}(|X_n - X| \geq \varepsilon)}_{\downarrow \, 0} \Rightarrow X_n \xrightarrow{\mathbb{P}} X$

Remarque : Quand on regarde la CV P.S ou la CV en proba, il faut que

$$\mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0 \quad \forall \varepsilon > 0$$

Ça suffit à avoir la CV en proba

Si on veut la C.V P.S il faut que la convergence soit suffisamment rapide pour que la série converge.

On a $X_n \xrightarrow{P.S} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$

Réciproque fausse

Exemple : Soit $(A_n)_n$ une suite d'événements indépendants.

On note $X_n = \mathbb{1}_A$ et $p_n = \mathbb{P}(A_n)$ et $X = 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n| > \varepsilon) = \begin{cases} 0 & \text{si } \varepsilon \geq 1 \\ \mathbb{P}(X_n = 1) = \mathbb{P}(A_n) = p_n & \text{si } \varepsilon < 1 \end{cases}$$

On a alors $X_n \xrightarrow{\mathbb{P}} 0$ ssi $p_n \longrightarrow 0$

Convergence p.s ?

Si $\sum p_n < \infty$ alors $X_n \xrightarrow{P.S} 0$

Si $\sum p_n = +\infty$ alors $X_n \cancel{\xrightarrow{P.S}} 0$ par exemple si $p_n = \frac{1}{n}$

La difference c'est la vitesse à laquelle $p_n \to 0$

Théorème : Soit $(X_n)_n$ une suite de v.a sur $(\Omega, \mathcal{A}, \mathbb{P})$

- Si $X_n \xrightarrow{\mathbb{P}} X$ alors il existe une sous suite $(X_{n_k})_{n_k}$ qui converge p.s vers X
- $X_n \xrightarrow{\mathbb{P}} X \Leftrightarrow$ De toutes sous suites de $(X_n)_n$ on peut extraire une sous suite qui converge P.S
- $X_n \xrightarrow{\mathbb{P}} X \Leftrightarrow (X_n)_n$ est de Cauchy i.e $\forall \varepsilon > 0$

$$\lim_{n,m} \mathbb{P}(|X_n - X_m| > \varepsilon) = 0$$

Propriétés : Soit $(X_n)_n$ et $(Y_n)_n$ deux suites de v.a sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^k$ et $\mathbb{R}^l$ avec $X_n \xrightarrow{\mathbb{P}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$.

- Si f est mesurable de $\mathbb{R}^{k+l}$ dans $\mathbb{R}^m$ alors $f(X_n, Y_n) \longrightarrow f(X, Y)$ en particulier $(X_n, Y_n) \xrightarrow{\mathbb{P}} (X, Y)$

II - Inégalités de Probas

Théorème : Inégalité de Markov

$$\forall \alpha > 0 \quad \mathbb{P}(|X| > \alpha) \leq \frac{E(|X|^k)}{\alpha^k}$$

Preuve :

$$E(|X|^k) = \int |X|^k d\mathbb{P} = \int_{\{|X| > \alpha\}} |X|^k d\mathbb{P} + \int_{\{|X| \leq \alpha\}} |X|^k d\mathbb{P}$$

$$\geqslant \int_{|X| > \alpha} |X|^k d\mathbb{P} \geqslant \int_{|X| > \alpha} \alpha^k d\mathbb{P} = \alpha^k \mathbb{P}(|X| > \alpha)$$

donc $\mathbb{P}(|X| > \alpha) \leq \dfrac{E(|X|^k)}{\alpha^k}$

Théorème : Inégalité de Bienayme - Tchebicher

On suppose de $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ ie $\underbrace{V(X)}_{=\sigma^2} < \infty$ et $E(X) = \mu$ alors

$$\forall \varepsilon > 0 \quad \mathbb{P}(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

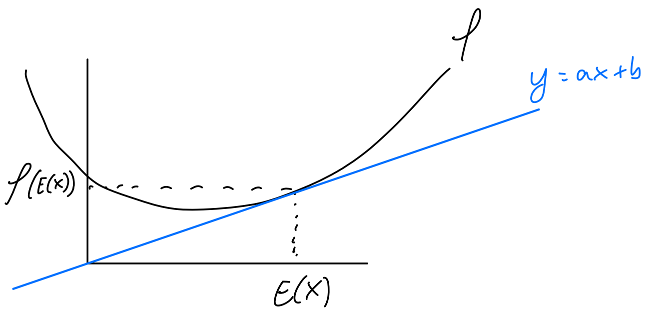
Preuve : on applique Markov à $X - E(X)$ et $k = 2$

Théorème : Inégalité de Jensen

Soit f une fonction convexe sur un intervalle I et soit X une v.a à valeurs dans I tq X et $f(X)$ sont intégrables alors $f(E(X)) \leq E(f(X))$

preuve :

- Si f est convexe alors il existe une droite de la forme $y = ax+b$ qui passe par $(E(X), f(E(X)))$ tq tout le graphe de f est au dessus de la droite



ie $f(x) \geq ax+b$

on passe a l'espérence

$E(f(x)) \geq aE(X)+b = f(E(x))$

Théorème : Inégalité de Holder

Soit p et q tq $\frac{1}{p} + \frac{1}{q} = 1$

alors $E(|XY|) \leq E(|X|^p)^{1/p} \, E(|Y|^q)^{1/q}$

Théorème : Inégalité de Cauchy - Schwartz

Si X et Y sont dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ alors

$cov(X,Y) \leq \sqrt{V(X) \cdot V(Y)}$

Preuve : On applique Jensen avec $X - E(X)$, $Y - E(Y)$ et $f(\cdot) = |\cdot|$ convexe

et Hölder avec $p = q = 2$

Théorème : Inégalité de Lyapounov

On prend $0 < \alpha < \beta$ alors
$$E(|X|^\alpha)^{1/\alpha} \leq E(|X|^\beta)^{1/\beta}$$

Exemple : Soit $(X_n)_n$ une suite de v.a non corrélées tq $E(X_i) = \underset{\overset{\shortparallel}{m}}{0}$ et $V(X_i) = \sigma^2 < \infty$. On pose $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

On veut montrer que $\bar{X}_n \xrightarrow{\mathbb{P}} 0$ ie $\forall \varepsilon > 0 \quad \mathbb{P}(|\bar{X} - 0| \geq \varepsilon) \to 0$

Or $E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = 0 = m$ et

$V(\bar{X}_n) = V\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{\sigma^2}{n}$

On applique l'inégalité de Tchebichev

$\forall \varepsilon > 0$ fixé,
$$\mathbb{P}(|\bar{X}_n - \underset{\overset{\shortparallel}{m}}{0}| \geq \varepsilon) \leq \frac{V(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \to \infty} 0$$

$$\Rightarrow \bar{X}_n \xrightarrow{\mathbb{P}} \underset{\overset{\shortparallel}{m}}{0}$$

Rappel : $(X_n)_n$ suite de v.a

- $X_n \xrightarrow{P.S} X$ ssi $\mathbb{P}(\omega : X_n(\omega) \longrightarrow X(\omega)) = 1$
- $X_n \xrightarrow{\mathbb{P}} X$ ssi $\forall \varepsilon > 0 \quad \mathbb{P}(|X_n - X| > \varepsilon) \to 0$

  ou $\mathbb{P}(|X_n - X| < \varepsilon) \longrightarrow 1$

Borell Cantelli

$$\sum \mathbb{P}(|X_n - X| > \varepsilon) < \infty \Rightarrow X_n \xrightarrow{P.S} X$$

Remarque :

La principale différence entre les deux types de convergeance est à quelle vitesse $\mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0$

Propriété :

$$X_n \xrightarrow{P.S} X \implies X_n \xrightarrow{\mathbb{P}} X$$

Théorème : Si $X_n \xrightarrow{\mathbb{P}} X$ alors $\exists\, (X_{n_k})_{n_k}$ une sous suite tq $X_{n_k} \xrightarrow{P.S} X$

Tchebishev + Markov les inégalités les plus importantes

III - Convergeance dans L^p

Soit $(X_n)_n$ une suite de v.a. On dit que X_n CV dans L^p ou CV en moyenne d'ordre p vers X (noté $X_n \xrightarrow{L^p} X$) si $|X|^p$ est intégrable et

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

Remarque :

- $\|X\|_p = E(|X|^p)^{1/p}$ est une norme sur L^p

  On a $\|X_n\|_p \leq \underbrace{\|X\|_p}_{\substack{< \infty \\ \text{par définition}}} + \underbrace{\|X_n - X\|_p}_{\substack{\text{tend vers } 0 \\ \text{si } X_n \xrightarrow{L^p} X}}$

$\Rightarrow$ Si $X_n \xrightarrow{L^p} X$ alors X_n admet des moments d'ordre p à partir d'un certain rang

- Si $p=2$ on parle de convergeance en moyenne quadratic

  Si $p=1$ on parle de convergence en moyenne simple

- Si $r<p$, par Lyapounov on a $\|X\|_r < \|X\|_p$

  donc

  $\|X_n - X\|_p \to 0 \Rightarrow \|X_n - X\|_r \to 0$

  donc

  $X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{L^\lambda} X$

  En particulier on a si $p>1$

  $X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{L^1} X$

<u>Proposition :</u> (CV L^1 $\Rightarrow$ CV $\mathbb{P}$)

Soit $(X_n)_n$ une suite de v.a tq $X_n \xrightarrow{L^1} X$ alors $X_n \xrightarrow{\mathbb{P}} X$

<u>Preuve :</u> Si $X_n \xrightarrow{L^1} X$ alors $E(|X_n - X|) \to 0$

On veut montrer que $X_n \xrightarrow{\mathbb{P}} X$ i.e $\forall \varepsilon > 0$ $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$

On applique Markov a $|X_n - X|$, $\alpha = \varepsilon$ et $k=1$

Rappel (Markov) : $\forall \alpha > 0$ $\mathbb{P}(|X| \geq \alpha) \leq \frac{E(|X|)^k}{\alpha^k}$

Ici : $\underbrace{\mathbb{P}(|X_n - X| \geq \varepsilon)}_{\to 0} \leq \underbrace{\frac{E(|X_n - X|)}{\varepsilon}}_{\to 0}$ car $X_n \xrightarrow{L^1} X$

$\Rightarrow X_n \xrightarrow{\mathbb{P}} X$

Proposition : Soit $(X_n)_n$ une suite de v.a. On suppose que tous les X_n sont p.s majorées par une constante K (i.e il existe une constante K tq $|X_n| \leq K$ p.s) et que $X_n \xrightarrow{\mathbb{P}} X$ alors

$X_n \xrightarrow{L^1} X$

Proposition : $X_n \xrightarrow{L^1} X$ ssi on a les 2 conditions suivantes :

1. $X_n \xrightarrow{\mathbb{P}} X$
2. La suite $(X_n)_n$ est uniformément intégrable i.e

$$\forall \varepsilon > 0 \quad \exists K > 0 \quad \forall n \in \mathbb{N}$$

$$E\left(|X_n| \mathbb{1}_{(|X_n| > K)}\right) < \varepsilon$$

Résumé :

$$\boxed{\begin{array}{c} X_n \xrightarrow{P.S} X \\ \mathbb{P}(\omega : X_n(\omega) \to X(\omega)) = 1 \end{array}}$$

$$\boxed{X_n \xrightarrow{L^q} X}$$

$\Downarrow$ $q > p$

$X_n \xrightarrow{L^p} X$

$E(|X_n - X|^p)$

IV Convergence en loi

Ce qu'on a fait sur les convergences en $\mathbb{P}$, p.s ou L^p, c'est qu'on a regardé les suites de v.a comme des suites de fonctions de Ω dans $\mathbb{R}^d$.
Maintenant on va regarder les v.a uniquement sur la caractérisation sur $\mathbb{R}^k$ i.e la proba image, la f.d.r, la densité, la f.c ...

Définition (convergence étroite)

On dit qu'une suite de mesures de probabilité $(\mathbb{P}_n)_{n\in\mathbb{N}}$ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ converge étroitement vers une mesure $\mathbb{P}$ (de probabilité) et on note $P_n \xrightarrow{e} \mathbb{P}$ si quelque soit la fonction réelle continue et bornée f définie sur $\mathbb{R}^d$ on a

$$\int_{\mathbb{R}^d} f \, d\mathbb{P}_n \longrightarrow \int_{\mathbb{R}^d} f \, d\mathbb{P} \qquad n \to \infty$$

Définition (Propriété) (convergence complète)

Soit $(\mathbb{P}_n)_n$ une suite de mesures de proba sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ de f.d.r $(F_n)_n$ (i.e. $F_n(n) = \mathbb{P}_n(]-\infty, n])$) alors

$$\mathbb{P}_n \xrightarrow{e} \mathbb{P} \iff F_n(n) \longrightarrow F(n) \text{ en tout point de continuité de } F$$

f.d.r de $\mathbb{P}$

On dit alors que F_n converge complètement vers F et on note $F_n \xrightarrow{c} F$

Exemple : On prend une variable aléatoire qui prend les valeurs $-\frac{1}{n}$ et $\frac{1}{n}$ avec probabilité $\frac{1}{2}$ i.e

$$\mathbb{P}(X_n = -\tfrac{1}{n}) = \mathbb{P}(X_n = \tfrac{1}{n}) = \tfrac{1}{2}$$

Si on regarde la mesure de probabilité associé, on a

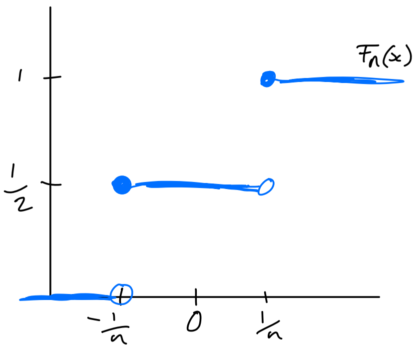
$$\mathbb{P}_n = \frac{1}{2}\left(\delta_{-\frac{1}{n}} + \delta_{\frac{1}{n}}\right)$$

Rappel : δ_a est la masse de Dirac en a

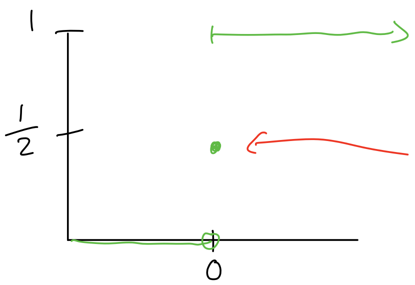
et est tq $\forall A \subseteq \mathcal{B}(\mathbb{R})$ $\begin{cases} \delta_a(A) = 1 & \text{si } a \in A \\ \delta_a(A) = 0 & \text{si } a \notin A \end{cases}$

La f.d.r associé à X_n est donnée par

$$F_n(x) = \begin{cases} 0 & \text{si } x < -\frac{1}{n} \\ \frac{1}{2} & \text{si } -\frac{1}{n} \leq x \leq \frac{1}{n} \\ 1 & \text{si } x \geq \frac{1}{n} \end{cases}$$



$$F_n(x) \longrightarrow \tilde{F}(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{2} & \text{si } x = 0 \\ 1 & \text{si } x > 0 \end{cases}$$



$\tilde{F}$ n'est pas une f.d.r
(pas continu à droite en 0)

$\tilde{F}$ pas f.d.r mais c'est pas grave

Car on définit F tq

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$$

F est une f.d.r associé à la v.a $X = 0$ p.s ou à la mesure de proba $\mathbb{P} = \delta_0$. On a que, $\forall$ point de continuité de F (i.e $\mathbb{R} \setminus \{0\}$)

on a $F_n(x) \longrightarrow F(x)$.

On a donc $F_n \xrightarrow{\mathcal{L}} F$ et donc $\underset{\frac{1}{2}(\delta_{-\frac{1}{n}} + \delta_{\frac{1}{n}})}{\mathbb{P}_n} \xrightarrow{e} \underset{\delta_0}{\mathbb{P}}$

i.e si on prend n'importe quelle fonction continue et bornée f on a

$$\int f \, d\mathbb{P}_n = \frac{1}{2}\left(f\left(-\frac{1}{n}\right) + f\left(\frac{1}{n}\right)\right)$$
$$\downarrow$$
$$f(0) = \int f \, d\mathbb{P}$$

Rappel : $\int f \, d\delta_a = f(a)$

<u>Remarque</u> : On peut faire le même raisonnement avec $X_n \sim \mathcal{N}(0, \sigma_n^2)$ avec $\sigma_n^2 \longrightarrow 0$

Propriété : (CV des densités $\Rightarrow$ CV $\mathbb{P}_n$)

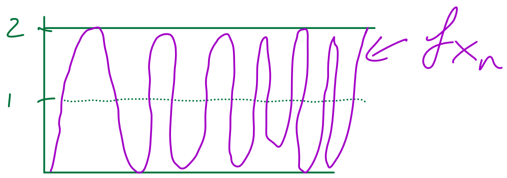
Soit $(\mathbb{P}_n)_n$ une suite de mesures de probas sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ de densités f_n par rapport à Lebesgue sur $\mathbb{R}^k$. Si la suite f_n converge presque partout vers une densité f alors $\mathbb{P}_n \xrightarrow{e} \mathbb{P}$ où $\mathbb{P}$ est la mesure de probabilité de densité f.

C'est à dire

$$\boxed{\text{Si } \forall x \quad f_n(x) \longrightarrow f(x) \Rightarrow \mathbb{P}_n \xrightarrow{e} \mathbb{P}}$$

⚠ réciproque fausse

Exemple : Soit $\forall n \in \mathbb{N}^*$ la v.a X_n admettant une densité f_{X_n} définie $\forall x \in \mathbb{R}$ par $f_X(x) = (1 - \cos(2\pi n x)) \mathbb{1}_{]0,1]}(x)$

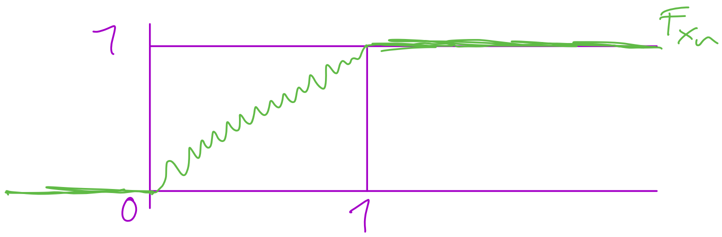


f_{X_n} oscille beaucoup. La suite f_{X_n} ne converge pas et même elle diverge en tout point

La suite de densité ne converge pas.

Mais si on regarde la f.d.r on a

$$F_{X_n}(x) = \begin{cases} 0 & \text{si } x < 0 \\ x - \dfrac{\sin(2\pi n x)}{2\pi n} & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$



On a alors

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

On a donc $F_{X_n} \xrightarrow{\mathcal{L}} F_X$

f.d.r d'une $U[0,1]$

Conclusion : les densités ne convergent pas mais les lois convergent

Théorème : Soit $(\phi_n)_n$ une suite de transformées de Fourier de mesures de Proba $\mathbb{P}_n$ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Si ϕ_n converge en tout point de $\mathbb{R}^d$ vers ϕ alors ϕ est la transformée de Fourier d'une mesure de Proba $\mathbb{P}$ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et on a

$$\mathbb{P}_n \xrightarrow{\mathcal{L}} \mathbb{P}$$

Définition : Soit $(X_n)_n$ une suite de vecteurs aléatoires à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On dit que X_n converge en loi vers un vecteur aléatoire X et on note $X_n \xrightarrow{\mathcal{L}} X$ si

$$\mathbb{P}_{X_n} \xrightarrow{e} \mathbb{P}_X \text{ ou } F_{X_n} \xrightarrow{c} F_X$$

Propriété : Il y a équivalence entre

- convergence en loi
- convergeance étroite des mesures de probas
- convergeance complète des f.d.r
- convergence des transformés de Fourier

Propriété :

Soit $(X_n)_n$ une suite de vecteur aléatoires de $\mathbb{R}^d$ alors

$X_n \xrightarrow{\mathcal{L}} X$ ssi $\forall \lambda \in \mathbb{R}^d$ on a

$$\underbrace{{}^t\lambda \cdot X_n}_{\text{C.L des coordonnés de } X_n} \xrightarrow{\mathcal{L}} \lambda^t \cdot X$$

Conséquence :

- Pas mal parce que ça permet de passer d'un critère en dimension d à un critère en dimension 1
- Si on prend $\lambda = (0, \dots, 0, 1, 0, \dots, 0)$ on a que la convergence en loi du vecteur implique la convergence en loi de chacune des composantes du vecteur.

i.e si $X_n = \begin{pmatrix} X_n^1 \\ \vdots \\ X_n^d \end{pmatrix}$ et $X = \begin{pmatrix} X^1 \\ \vdots \\ X^d \end{pmatrix}$ et $X_n \xrightarrow{\mathcal{L}} X$ alors $\forall i = 1, \dots, d$

on a $X_n^i \xrightarrow{\mathcal{L}} X^i$

Réciproque fausse

Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$ alors

$$\cancel{(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)}$$

Exemple : on prend $X \sim \mathcal{N}(0,1)$

$$X_n = X \quad \forall n \quad \text{et} \quad Y_n = (-1)^n X$$

On a $X_n \xrightarrow{\mathcal{L}} X$

$Y_n \xrightarrow{\mathcal{L}} X$

mais $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \not\xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ X \end{pmatrix}$ car $X_n + Y_n = \begin{cases} 0 \text{ si } n \text{ est impair} \\ 2X \text{ si } n \text{ est pair} \end{cases}$

donc pas de convergence en loi

Résumé : $X = X_1, \dots, X_n$ suite de v.a

Convergence p.s :

$$X_n \xrightarrow{p.s} X \quad \text{ssi} \quad \mathbb{P}(\omega : X_n(\omega) \to X(\omega)) = 1$$

différence de vitesse

Convergence en Proba :

$$X_n \xrightarrow{\mathbb{P}} X \quad \text{ssi} \quad \forall \varepsilon > 0 \quad \mathbb{P}(\|X_n - X\| > \varepsilon) \to 0$$

$$\mathbb{P}(|X_n - X| < \varepsilon) \to 1$$

Convergence L^p

$$X_n \xrightarrow{L^p} X \quad \text{ssi} \quad E(|X_n - X|^p) \to 0$$

Convergence en loi

$X_n \xrightarrow{\mathcal{L}} X$ ssi loi de $X_n \longrightarrow$ loi de X

ou

$\mathbb{P}_{X_n} \xrightarrow{e} \mathbb{P}_X$

$F_{X_n} \xrightarrow{c} F_X$ ou $\phi_{X_n} \longrightarrow \phi_X$

Lemme de Slutsky

Soit $(X_n)_n$ et $(Y_n)_n$ deux suites de v.a et

si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} a$ ← constante alors

$$X_n Y_n \xrightarrow{\mathcal{L}} aX$$

La convergence en loi est bcp plus faible que la convergence en proba.

On prend $X \sim \mathcal{N}(0,1)$ et on pose $X_n = (-1)^n X$

On a $X_n \xrightarrow{\mathcal{L}} X$ et même loi que X_n = loi de X ($X_n \overset{d}{=} X$)

Mais on ne peut pas avoir la convergence en proba car

$$\mathbb{P}(|X_n - X| > \varepsilon) = \begin{cases} \mathbb{P}(|2X| > \varepsilon) & \text{si } n \text{ impair} \not\to 0 \\ \mathbb{P}(0 > \varepsilon) & \text{sinon} \end{cases}$$

Théorème :

$$X_n \xrightarrow{\mathbb{P}} X \Longrightarrow X_n \xrightarrow{\mathcal{L}} X$$

Théorème

$$X_n \xrightarrow{\mathcal{L}} a \Rightarrow X_n \xrightarrow{\mathbb{P}} a$$

preuve : On a $\mathcal{L}(X_n) \longrightarrow \mathcal{L}(X)$ et on veut $\mathbb{P}(|X_n - a| > \varepsilon) \to 0$

Si on a $X_n \xrightarrow{\mathcal{L}} a$ alors $\mathbb{P}_{X_n} \xrightarrow{e} \delta_a$

et on veut mq $\mathbb{P}(|X_n - a| \leq \varepsilon) \to 1$

Rappel

$$\int g(x) \, d\delta_c(x) = g(c)$$

On a

$$\mathbb{P}(|X_n - a| \leq \varepsilon) = E(\mathbb{1}_{\{|X_n - a| \leq \varepsilon\}})$$

$$= \int_{\mathbb{R}} \mathbb{1}_{\{|x-a| \leq \varepsilon\}} \, d\mathbb{P}_{X_n}(x)$$

Remarque : Ça marche parce qu'on considère la constante comme une v.a définie sur n'importe quel espace

Résumé :

$X_n \xrightarrow{p.s} X$

$\mathbb{P}(\omega : X_n(\omega) \to X(\omega)) = 1$

$\exists$ une sous suite $(X_{n_k})_{n_k}$

$X_n \xrightarrow{\mathbb{P}} X$

$\forall \varepsilon > 0 \quad \mathbb{P}(|X_n - X| > \varepsilon) \to 0$

$X_n \xrightarrow{L^p} X$

$E(|X_n - X|^p) \to 0$

$p > 1$

si $|X_n| \leq K$ p.s

$X_n \xrightarrow{L^1} X$

$E(|X_n - X|) \to 0$

$r \geq 1$

$X_n \xrightarrow{L^r} X$

$E(|X_n - X|^r) \to 0$

Les v.a sont des fonctions de Ω dans $\mathbb{R}^k$

si $X = a$ constante

Les v.a sont caractérisé par leur comportement sur $\mathbb{R}^k$

$X_n \xrightarrow{\mathcal{L}} X$

$\mathcal{L}(X_n) \to \mathcal{L}(X)$

$\phi_{X_n}(x) \to \phi_X(x)$

$F_{X_n} \xrightarrow{c} F_X$

$F_{X_n}(x) \to F_X(x)$ en tout point de continuité de F

$f_{X_n}(x) \to f_X(x)$

$\mathbb{P}_{X_n} \xrightarrow{e} \mathbb{P}_X$

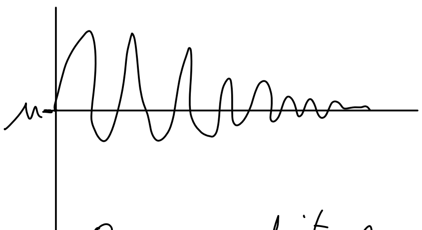
$\int g \, d\mathbb{P}_{X_n} \to \int g \, d\mathbb{P}_X$

On a : Si $X_1, \dots, X_d$ i.i.d tq $E(X_i) = \mu$ et $V(X_i) = \sigma^2$

alors $\frac{1}{n} \sum X_i = \overline{X}_n \xrightarrow{\mathbb{P}} \mu$

On simule n réalisations de loi Normale $x_1, \dots, x_n$ (rnorm(n,0,1) sur R)

On calcule la suite des $\overline{X}_n(w)$ i.e $\left(x_1, \frac{x_2+x_1}{2}, \dots, \frac{\sum x_i}{n}\right)$



traduit la convergence en proba

On a dit que si on pose $X_n = (-1)^n X$ avec $X \sim \mathcal{N}(0,1)$

on a $X_n \xrightarrow{\mathcal{L}} X$

$$x_n^* = (-1)^n x_n$$

<u>Théorème</u> : On prend une suite $(X_n)_n$ de v.a Gaussiennes alors il y a équivalence entre

- $\sum X_k \xrightarrow{p.s} X$
- $\sum X_k \xrightarrow{\mathbb{P}} X$
- $\sum X_k \xrightarrow{L^p} X$
- $\sum X_k \xrightarrow{\mathcal{L}} X$

Chapitre V : Théorèmes Limites

I - Loi des grands nombres

Proposition : Si X est une v.a admettant un moment d'ordre 2 (i.e $V(X) < \infty$) alors la fonction caractéristique de X admet un développement limité à l'ordre 2 en 0 de la forme

$$\phi_X(t) = 1 + it\,E(X) - \frac{t^2}{2}E(X^2) + o(t^2)$$

Preuve : On applique Taylor et le fait que $\phi^{(k)}(0) = i^k E(X^k)$

Théorème : Loi faible des grands nombres ou théorème de Khintchine.

Soit $(X_n)_n$ une suite de v.a définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ independantes, de même loi et intégrables avec $E[X_i] = \mu$ alors $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mu$

Preuve : Comme μ est une constante, si on veut montrer que $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$ il suffit de montrer que $\overline{X}_n \xrightarrow{\mathcal{L}} \mu$

ou $\phi_X(t) \longrightarrow \phi_\mu(t) = e^{it\mu}$ = f.c de la constante μ

$$\phi_{\overline{X}_n}(t) = \phi_{\frac{1}{n}\Sigma X_i}(t) = \phi_{\Sigma X_i}\left(\frac{t}{n}\right) = \prod_{i=1}^{n} \phi_{X_i}\left(\frac{t}{n}\right) = \phi_X\left(\frac{t}{n}\right)^n$$

Or X admet un moment d'ordre 1 donc sa f.c admet un développement limité en 0 à l'ordre 1 avec

$$\phi_X(t) = 1 + itE(X) + o(t) = 1 + it\mu + o(t)$$

donc

$$\phi_X(t) = \phi_X\left(\frac{t}{n}\right)^n = \left(1 + i\frac{t}{n}\mu + o\left(\frac{t}{n}\right)\right)^n$$

$$= \exp\left(n \ln\left(1 + i\frac{t}{n}\mu + o\left(\frac{t}{n}\right)\right)\right)$$

$$\phi_{\bar{X}_n}(t) = \exp\Big(n \underbrace{\ln\left(1 + i\frac{t}{n}\mu + o\left(\frac{t}{n}\right)\right)}_{\sim i\frac{t}{n}\mu}\Big)$$

$$\sim e^{it\mu} = \text{f.c de } \mu$$

$$\Rightarrow \bar{X}_n \xrightarrow{\mathcal{L}} \mu \Rightarrow \bar{X}_n \xrightarrow{\mathbb{P}} \mu$$

<u>Théorème</u> : Loi forte des grands nombres ou théorème de Kolmogorov.

Sous les mêmes hypothèses que précédemment, on a $\bar{X}_n \xrightarrow{p.s} \mu$

<u>Théorème</u> : LGN dans L^2

Soit $(X_n)_n$ une suite de v.a non corrélées, de carré intégrable, de même espérance μ et de même variance $\sigma^2 < \infty$ alors

$$\bar{X}_n \xrightarrow{L^2} \mu$$

Remarque : il existe plein de LGN en modifiant les hypothèses (X_n faiblement dépendants ou on a $\frac{1}{n}\sum_{i=1}^{n} E(X_i) \to \mu$) ou les types de convergence.

Rappel : Théorèmes Limites

- Loi des grands nombres.

Soit $(X_n)_n$ une suite de v.a indep, de même loi, intégrables tq $E(X_i)=\mu$ alors

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P.S} \mu$$

<u>Application :</u> Justification de l'approche fréquentiste

On répète une expérience aléatoire.

Les répétitions des expériences sont indépendantes les unes des autres.

On peut voir les résultats de l'expérience comme les réalisations d'une suite de v.a.i.i.d

Soit A un évènement résultat de l'expérience

On pose $X_n = \begin{cases} 1 & \text{si } A \text{ est réalisé} \\ 0 & \text{sinon} \end{cases}$

On pose $p = \mathbb{P}(A)$. On a $X_n \sim Bern(p)$ et $E(X_n) = p$

On pose $h(A_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ = la proportion parmi les n expériences qui ont donné A comme résultat

On a alors $h(A_n) \xrightarrow{P.S} p = \mathbb{P}(A)$

Traduction : Même si on ne connait pas la proba d'un évènement, on peut la trouver (si assez d'obs)

Exemple : On veut savoir quelle est la proba d'avoir un sinistre auto supérieur à 100 000 €. Si Y_i est le i^e sinistre, on pose

$$X_i = \mathbb{1}_{\{Y_i > 100000\}} \text{ alors}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P.S} \mathbb{P}(Y_i > 100000)$$

Application : Convergence de la f.d.r empirique

On a vu que $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$

Si $(X_n)_n$ est une suite de v.a.i.i.d de f.d.r F
et

$$\lim_{n\to\infty} \sup_{x} |F_n(x,\cdot) - F(x)| = 0 \qquad \mathbb{P}.\ p.s$$

Si on pose $Y_i = \mathbb{1}_{\{X_i \leq x\}}$ alors $Y_i \sim Bern(p)$
avec $p = \mathbb{P}(X_i \leq x) = F(x)$
et par la LGN

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i = F_n(x) \xrightarrow{P.S} F(x)$$

Corollaire : $a < b$

$$F_n(b) - F_n(a) = \frac{1}{n}\{\#\text{obs entre } a \text{ et } b\} \to F(b) - F(a)$$

Application : Théorème de Weierstrass

On prend $(X_n)_n$ une suite de v.a indep tq
$X_i \sim Bern(x)$ avec $x \in [0,1]$. On a $E(X_i) = x$
et $\sum_{i=1}^{n} X_i \sim B(n,x)$

On prend une fonction f continue sur $[0,1]$. On calcule
$E\left(f\left(\frac{X_1+\dots+X_n}{n}\right)\right) = E\left(f\left(\frac{S_n}{n}\right)\right) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\mathbb{P}(S_n = k)$

où $\mathbb{P}(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k}$

$$E\left(f\left(\frac{S_n}{n}\right)\right) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

Par la loi des grands nombres on a $f(\bar{X}_n) \xrightarrow{P.S} f(x)$
et par le théorème de convergence dominée on a

$$\sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} \longrightarrow f(x)$$

i.e toute fonction continue sur $[0,1]$ peut être approchée par une suite de polynômes

Théorème Centrale Limite

Soit $(X_n)_n$ une suite de v.a.i.i.d tq $E(X_i) = m$
et $V(X) = \sigma^2 < \infty$ alors

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

Traduction : Le comportement moyen de mon risque (caractérisé par $\bar{X}_n$) peut être considéré comme Gaussien i.e $\bar{X}_n \approx \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$

On considère qu'à partir de $n=30$, on a une bonne approximation de la loi de $\bar{X}_n$ par une Gaussienne.

Sous R :

```
X = c()
for(i in 1:1000) {X[i] = mean(rexp(30,2))}
x = sqrt(30)(X-2)/sqrt(4); hist(x),
curve(dnorm(x), add=true)
```

$\mathcal{E}(\frac{1}{2})$

Remarque : Ça justifie l'importance de la loi Normale

Preuve : Les X_i sont i.i.d tq $E(X_i) = m$ et $V(X_i) = \sigma^2 < \infty$

Donc $\frac{X_i - m}{\sigma}$ est d'espérance 0 et Var 1

et de f.c ϕ

$$\phi_{\frac{1}{\sqrt{n}} \frac{X_i - m}{\sigma}}(t) = \phi\left(\frac{t}{\sqrt{n}}\right)$$

donc $\phi_{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m}{\sigma}}(t) = \phi\left(\frac{t}{\sqrt{n}}\right)^n$

Rappel : Si X admet un moment d'ordre 2 et

$$\phi_X(t) = 1 + itE(X) - \frac{t^2}{2}E(X^2) + o(t^2)$$

$$\phi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t}{n}\right)$$

$$\phi\left(\frac{t}{\sqrt{n}}\right)^n = \exp\left(n \ln\left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)\right) \sim \exp\left(-\frac{t^2}{2}\right)$$

f.c d'une $\mathcal{N}(0,1)$

Application : Monte-Carlo

En finance, actuariat, en économie, on est souvent amené à calculer $E(f(X))$ (le prix d'un produit dérivé, le montant d'une prime, l'utilité espérée...)

$\Rightarrow$ On est pas toujours capable de le calculer

On va simuler des réalisations de $X = x_1, \dots, x_n$

On va approcher $E(f(X))$ par $\frac{1}{n} \sum_{i=1}^{n} f(x_i)$

(la loi GN dit $\frac{1}{n} \sum_{i=1}^{n} f(x_i) \xrightarrow{PS} E(f(X))$)

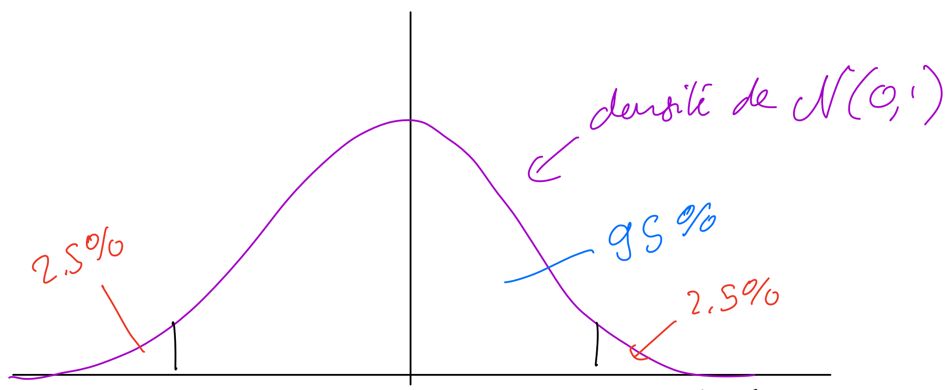
A chaque fois que je simule $f(x_i)$ on obtient une approximation différente de $E(f(X))$

On va rajouter un intervalle de confiance : on va donner un intervalle qui contient $E(f(X))$ avec une proba élevée

On note $E(f(X)) = \mu_f$ et $V(f(X)) = \sigma_f^2$, le TCL dit

$$\xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Si $Z \sim \mathcal{N}(0,1)$



$z_{0.025} = -z_{0.975}$

$z_{0.975}$ = le réel tq $P(\mathcal{N}(0,1)) = 0.975$

qnorm$(0.975) \approx 1.96$

$$\Rightarrow \mathbb{P}(-z_{0.975} \leq Z \leq z_{0.975}) = 0.95$$

On a

ce qui nous intéresse

$$\mathbb{P}\left(-z_{0.975} \leq \sqrt{n}\, \frac{\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \mu_f}{\sigma_f} \leq z_{0.975}\right) \longrightarrow 0.95$$

$$\Rightarrow \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i) - z_{0.975}\frac{\sigma f}{\sqrt{n}} \leq \mu f \leq \frac{1}{n}\sum f(x_i) + z_{0.975}\frac{\sigma f}{\sqrt{n}}\right) \approx 0.95$$

l'intervalle $\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i) \pm z_{0.975}\frac{\sigma f}{\sqrt{n}}\right]$

contient $\mu f = E(f(x))$ avec une probabilité proche de 95 %

Comme l'intervalle est centré sur $\frac{1}{n}\sum_{i=1}^{n} f(x_i)$, on peut avoir confiance que $\frac{1}{n}\sum f(x_i)$ est une bonne approx de $E(f(x))$

$\Rightarrow$ Si l'intervalle est grand, on augmente n pour le rendre plus petit

Théorie des valeurs extrêmes

Théorème (Fischer - Typpett)

On a une suite $(X_n)_n$ de v.a.i.i.d. On pose

$M_n = \max_{i=1,\dots,n} X_i$ si il existe deux suites c_n et $d_n > 0$

tq $\mathbb{P}\left(\frac{M_n - c_n}{d_n} \leq x\right) \longrightarrow G(x)$ (non dégénérée)

Alors G est la distribution des extrêmes généralisées (GEV) de paramètres μ, σ et ξ tq

$$G(x) = \begin{cases} \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) & \text{si } \xi = 0 \\ \exp\left(-\left(1 - \xi\left(\frac{x-\mu}{\sigma}\right)\right)_+^{-\frac{1}{\xi}}\right) & \text{si } \xi \neq 0 \end{cases}$$

Remarque : Le ξ contient toute l'info sur les valeurs élevées de la distribution de X

$$\begin{cases} \text{Si } \xi < 0, \text{ la distribution de } X \text{ est bornée} \\ \text{Si } \xi > 0, \text{ la dist de } X \text{ va décroître de manière exponentielle} \end{cases}$$

ex : loi Normale

Si $\xi > 0$, on a $E(X^a) = \infty$ ssi $a > \frac{1}{\xi}$.

$\Rightarrow$ On peut avoir des extrêmes très élevées

Remarque : Si on a assez d'observation on a

$$M_n \sim GEV(\mu, \sigma, \xi)$$

Théorème : C'est équivalent de dire que

- $\exists c_n$ et $d_n > 0$ tq $\mathbb{P}\left(\frac{M_n - c_n}{d_n}\right) \longrightarrow G(x)$

- $\exists$ une fonction positive σ tq

$$\lim_{u \to x_F} \sup_{u<x<x_F} \left| F_u(x) - G_{\xi,\sigma(u)}(x) \right| = 0$$

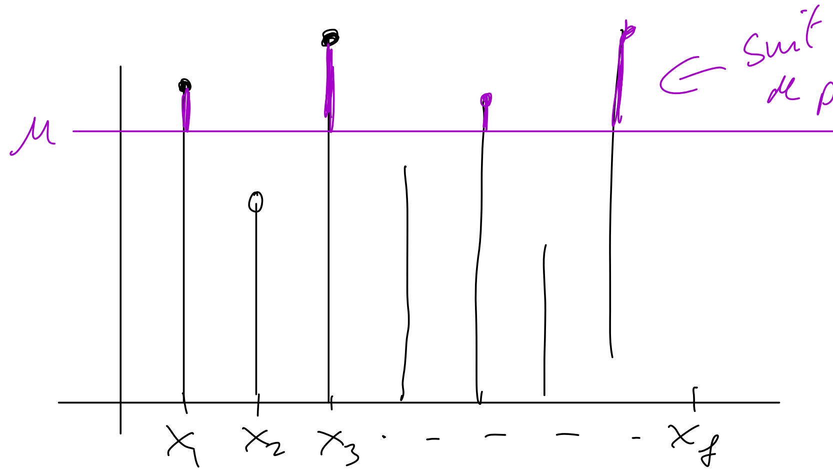
point terminal i.e

$$x_F = \min\left(x : F(x) = 1\right)$$

$F_u(x) = \mathbb{P}(X-u \leq x \mid X>u)$

distribution conditionnelle de $X-u$ sachant que $X>u$

Traduction Statistique



suit (à peu près) une loi de Pareto Généralisé

Traduction Actuariat

En réassurances, ce qui dépasse un certain seuil est pris en charge par la réassurance. Ces montants peuvent être modélisés par une pareto Généralisée

Séance 22 Mars

Rappel : Théorèmes limites

- La Loi des grands nombres

$X_1, X_2, \ldots$ tq $E(X_i) = \mu$

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

- Théorème Centrale Limite

$X_1, \ldots, X_n$ $\quad E(X_i) = \mu, \quad V(X_i) = \sigma^2 < \infty$

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

Traduction Stat : $\bar{X}_n \approx \mathcal{N}$

Traduction actuariat : Le comportement moyen suit (à peu près) une Gaussienne

- Valeurs extrêmes : Si $M_n = \max X_i \quad i=1,\ldots,n$

$$\mathbb{P}\left(\frac{M_n - a_n}{b_n} \leq x\right) \longrightarrow G_{\mu,\sigma,\xi}(x)$$

$\leftarrow$ f.d.r d'une $GEV(\mu, \sigma, \xi)$

$\uparrow$ Generalized extreme Value

- $$\limsup_{u \to x_F} \left| F_u(x) - GPD(x) \right| = 0$$

$\uparrow$ $\mathbb{P}(X - u \leq x \mid X > \mu)$

$\nwarrow$ f.d.r d'une GPD

Remarque : Théorème central limite multivariée

Soit $(X_i)_{i\geq 1}$ suite de vecteurs aléatoires $\overset{i.i.d}{\vee}$ dans $\mathbb{R}^d$ tq $E[X_i] = \mu \in \mathbb{R}^d$ et de matrice de covariance Γ alors

$$\sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

Approximation de lois

Théorème : Approximation de la loi Binomiale par une Gaussienne

Soit $(X_n)_n$ une suite de v.a tq $X_n \sim \mathcal{B}(n,p)$ alors $\dfrac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$

"preuve" : On pose $(Y_i)_{i\geq 1}$ une suite de v.a i.i.d tq $Y_i \sim \mathrm{Bern}(p)$

On a $X_n \overset{d}{=} \sum_{i=1}^{n} Y_i$ et on applique le TCL

$$\sqrt{n}\,\frac{\bar{Y}_n - p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

$$\frac{X_n - np}{\sqrt{p(1-p)n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Théorème : Approximation d'une loi Binomiale par une loi Poisson

Soit $(S_n)_n$ une suite de v.a independantes tq $S_n \sim \mathcal{B}(n, p_n)$

Si $\lim_{n\to\infty} np_n = \lambda$ alors $S_n \xrightarrow{\mathcal{L}} \mathrm{Poi}(\lambda)$

Application : Modélisation de la charge sinistre

- Modèle individuel : On a un portefeuille avec n contrats

Dans le modèle individuel on modélise la charge total générée par chaque contrat

$$S^{indiv}_{total} = \sum_{i=1}^{n} I_i U_i$$ ← $U_i \sim$ Gama ou LN qui modélise le montant du sinistre du i contrat

$I_i \sim Bern(p)$ (ça me dit si le contrat va avoir un sinistre)

(I_i) v.a.i.i.d et $(U_i)_i$ v.a.i.i.d

- Modèle collectif : On modélise les pertes globales sur le portefeuille

$$S^{coll}_{tot} = \sum_{i=1}^{N} U_i$$

N : v.a.r qui modélise le nombre de sinistres sur le portefeuille

U_i : même chose que dans le cas indiv

Si on veut relier les modèles entre eux alors

$$N \sim B(n, p)$$

n : #contrats ; p : proba qu'un contrat ait un sinistre

Dans la réalité, on prend $N \sim Poi(\lambda)$

Remarque : On peut faire l'approximation uniquement dans le cas où p est petit car on a l'hypothèse $\lim_n np_n = \lambda$.

La Loi de Poisson modélise le nb d'évènements rares qui se réalisent

Utilité : Erreur numérique associé à la loi Binomiale > Erreur d'approx par la loi de Poisson

si n est grand et p très petit

Exemple sur R : $n = 10\,000$ et $p = 0.002$

$\mathbb{P}(B(10000, 0.002) = 10) =$ dbinom(10, 10000, 0.002) $= 0.005790$

$\mathbb{P}(Poi(20) = 10) =$ dpois(10, 20) $= 0.005816$

Chapitre 6 : Théorème de Radon-Nykodym

Idée : On a une v.a X qui suit une certaine loi sous $\mathbb{P}$, on change la mesure $\mathbb{P}$ par $\mathbb{Q}$ et on veut voir l'impact sur la loi de X

I - Rappel

Définition : On a un ensemble Ω. Une tribu $\mathcal{A}$ de l'espace Ω est un ensemble de sous parties de Ω qui vérifient

- $\emptyset \in \mathcal{A}$ et $\Omega \in \mathcal{A}$
- Stable par passage au complémentaire

  Si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$
- Stable par union dénombrables

  Si $(A_n)_n$ suite d'évènements avec $A_n \in \mathcal{A}$ alors

  $$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$$

$(\Omega, \mathcal{A})$ est un espace mesurable

Definition : Une mesure positive est une application

$m : \mathcal{A} \longrightarrow \mathbb{R}^+$ tq

- $m(\emptyset) = 0$
- σ-additivité : $\forall (A_n)_n \in \mathcal{A}$ une suite d'évènements disjoints deux à deux de mesures finies

$$m\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} m(A_n)$$

C'est une probabilité si $m(\Omega) = 1$

Définition : Soit $\mathbb{P}$ et Q deux mesures positives définies sur $(\Omega, \mathcal{A})$

On dit que Q est absolument continue par rapport à $\mathbb{P}$ ($Q \ll \mathbb{P}$)

ou que $\mathbb{P}$ domine Q si

$$\forall A \in \mathcal{A} \text{ tq } \mathbb{P}(A) = 0 \Rightarrow Q(A) = 0$$

Si $\mathbb{P} \ll Q$ et $Q \ll \mathbb{P}$ alors $\mathbb{P}$ et Q sont des mesures équivalentes. ($\mathbb{P} \sim Q$)

Remarque : Si $Q \ll \mathbb{P}$, les ensembles négligeables pour $\mathbb{P}$ le sont aussi pour Q

Si $\mathbb{P} \sim Q$, $\mathbb{P}$ et Q ont même ensemble négligeable

Définition : Deux probas $\mathbb{P}$ et Q sont étrangères (noté $\mathbb{P} \perp Q$) si

$$\exists A \in \mathcal{A} \text{ tq } \mathbb{P}(A) = 0 \text{ et } Q(A) = 1$$

Exemple : $X \sim \mathcal{N}(\mu, \sigma^2)$ avec $0 < \sigma^2 < \infty$

On considère $\mathbb{P}_X$ alors $\mathbb{P}_X \ll \lambda$

II Théorème de Randon-Nikodym

Proposition : Soit X une v.a.r positive, définie sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$, d'espérance $E^{\mathbb{P}}(X) = 1$

L'application Q définie sur $\mathcal{A}$ par

$$Q(A) = E^{\mathbb{P}}(X \mathbb{1}_A)$$

définie une nouvelle mesure de probabilité abs. cont par rapport à $\mathbb{P}$

Preuve : On veut montrer que Q est une mesure de probabilité

- $Q(\emptyset) = E^{\mathbb{P}}(X \mathbb{1}_{\emptyset}) = 0$

  $Q(\Omega) = E^{\mathbb{P}}(X \mathbb{1}_{\Omega}) = E^{\mathbb{P}}(X) = 1$

- σ-additivité : Soit $(A_i)_{i\in\mathbb{N}}$ suite d'évènements disjoints (on veut

  $Q(\bigcup_{i\in\mathbb{N}} A_i) = \sum Q(A_i)$

  $Q(\bigcup_{i\in\mathbb{N}} A_i) = E^{\mathbb{P}}(X \mathbb{1}_{\cup A_i}) = E^{\mathbb{P}}(\sum_{i\in\mathbb{N}} X \mathbb{1}_{A_i})$

  Soit $X_n = \sum_{i=1}^{n} X \mathbb{1}_{A_i}$ est une suite croissante de v.a positive

  qui converge vers $X \mathbb{1}_{\bigcup_{i\in\mathbb{N}} A_i}$

  on a $E^{\mathbb{P}}(X_n) = \sum_{i=1}^{n} E^{\mathbb{P}}(X \mathbb{1}_{A_i}) = \sum_{i=1}^{n} Q(A_i)$

  par théorème de convergence monotone,

  $\lim_{n\to\infty} E^{\mathbb{P}}(X_n) = E^{\mathbb{P}}(\lim_{n\to\infty} X_n)$

  $\Rightarrow Q(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} Q(A_i)$

- On veut montrer que $Q \ll \mathbb{P}$ (i.e si $\mathbb{P}(A) = 0 \Rightarrow Q(A) = 0$)

  Si A est négligeable pour $\mathbb{P}$ alors $X \mathbb{1}_A$ est $\mathbb{P}$ p.s nulle

  donc $E^{\mathbb{P}}(X \mathbb{1}_A) = 0$

Définition : Soit $(\Omega, \mathcal{A})$ un espace mesurable et $\mathbb{P}$ une mesure σ-finie. Soit X une v.a.r positive sur $(\Omega, \mathcal{A})$ tq $E^{\mathbb{P}}(X) = 1$ (i.e $\int X(\omega)\, d\mathbb{P}(\omega) = 1$)

Soit Q la mesure définie sur $(\Omega, \mathcal{A})$ par $\forall A \in \mathcal{A}$

$$Q(A) = E^{\mathbb{P}}(X \mathbb{1}_A)$$

On appelle X la densité de probabilité de Q par rapport à $\mathbb{P}$

et on note $X = \frac{dQ}{d\mathbb{P}}$

Remarque : On peut réécrire la def dans le cadre de $\mathbb{R}$

On prend $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on prend la mesure λ

On prend une fonction définie sur $\mathbb{R}$ (une v.a définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

positive et tq $\int f(x)\,dx = 1$

Soit Q la mesure définie par $\mathbb{P}(A) = \int_A f(x)\,dx$ alors $\mathbb{P} \ll \lambda$ et

f est la densité de proba de $\mathbb{P}$ par rapport à λ

Théorème : Radon-Nikodim

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et Q une mesure de probabilité sur $(\Omega, \mathcal{A})$ tq $Q \ll \mathbb{P}$, alors il existe une unique (classe de) variable aléatoire X positive tq $E^{\mathbb{P}}(X) = 1$ et

$$\forall A \in \mathcal{A} \quad Q(A) = E^{\mathbb{P}}(X \mathbb{1}_A)$$

X est appelée la dérivée de Radon-Nikodym de Q par rapport à $\mathbb{P}$ et

on note $X = \frac{dQ}{d\mathbb{P}}$

Remarque (version mesure)

Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on prend $\mathbb{P}$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tq $\mathbb{P} \ll \lambda$ alors il existe une fonction f positive tq $\int f(x)\,dx = 1$ et

$\forall B \in \mathcal{B}(\mathbb{R}) \quad \mathbb{P}(B) = \int_B f(x)dx$ et f est la densité de $\mathbb{P}$ par rapport à λ

Remarque: En finance, on calcule le prix d'un produit dérivé comme une espérence sous une certaine mesure de proba (proba risque neutre sous laquelle on sait faire les calculs). Au départ, on se place sous une proba quelconque (proba historique) et on fait un changement de mesure de proba.

Proposition: Calcul d'espérence sous une mesure différente

Soit $(\Omega, \mathcal{A}, \mathbb{P})$, soit Q une mesure de proba tq $Q \ll \mathbb{P}$

Soit Y une v.a intégrable par rapport à Q alors

$Y\frac{dQ}{d\mathbb{P}}$ est intégrable par rapport à $\mathbb{P}$ et on a

$$E^Q(Y) = E^{\mathbb{P}}\left(Y\frac{dQ}{d\mathbb{P}}\right)$$

$$\int Y(\omega)dQ(\omega) = \int Y(\omega)\frac{dQ}{d\mathbb{P}}(\omega)d\mathbb{P}(\omega)$$

Rappel: On a une mesure $\mathbb{P}$ σ-finie $(\Omega, \mathcal{A})$ et X une v.a tq $\mathbb{E}^{\mathbb{P}}(X) = 1$ alors Q définie par $Q(A) = E^{\mathbb{P}}(X\mathbb{1}_A)$ est une proba tq $Q \ll \mathbb{P}$

Radon-Nikodym : Si $Q \ll \mathbb{P}$, $\exists X$ v.a tq $E^{\mathbb{P}}(X) = 1$

et $Q(A) = E^{\mathbb{P}}(X \mathbb{1}_A)$

et X est la densité de Q par rapport à $\mathbb{P}$

Calcul d'espérance :

On a $Q \ll \mathbb{P}$ donc $\exists \frac{dQ}{d\mathbb{P}}$.

Si X est intégrable par rapport à Q (i.e $E^{Q}(X) < \infty$)

alors $X \frac{dQ}{d\mathbb{P}}$ est intégrable par rapport à $\mathbb{P}$ et

$E^{Q}(X) = E^{\mathbb{P}}\left(X \frac{d\mathbb{P}}{dQ}\right)$ ou $\int X(\omega)\, dQ(\omega) = \int X(\omega) \frac{dQ}{d\mathbb{P}}(\omega)\, d\mathbb{P}(\omega)$

Remarque : Une densité sur $\mathbb{R}$ (fonction positive tq $\int f d\lambda = 1$) est une v.a réelle sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

On peut définir une nouvelle mesure de probabilité $\mathbb{P}$ tq $\mathbb{P}(A) = \int_A f(x)\, dx$ $\forall A \in \mathcal{B}(\mathbb{R})$ et $\mathbb{P} \ll \lambda$. On retombe sur la notion de v.a à densité sur $\mathbb{R}$

Propriétés :

- On a un espace probabilisable $(\Omega, \mathcal{A})$; $\mathbb{P}$ et Q deux mesures sur $(\Omega, \mathcal{A})$ tq $\mathbb{P} \sim Q$ alors

$$\frac{d\mathbb{P}}{dQ} = \frac{1}{\frac{dQ}{d\mathbb{P}}}$$

preuve: On a $\mathbb{P} \sim \mathbb{Q}$ i.e $\mathbb{P} \ll \mathbb{Q} \overset{R.N}{\Longrightarrow} \frac{d\mathbb{P}}{d\mathbb{Q}}$ existe

et $\mathbb{Q} \ll \mathbb{P} \Longrightarrow \frac{d\mathbb{Q}}{d\mathbb{P}}$ existe

$$\forall A \in \mathcal{A} \quad \mathbb{Q}(A) = E^{\mathbb{Q}}(\mathbb{1}_A)$$

$$\underset{\mathbb{Q} \ll \mathbb{P}}{=} E^{\mathbb{P}}\left(\mathbb{1}_A \frac{d\mathbb{Q}}{d\mathbb{P}}\right)$$

$$= E^{\mathbb{Q}}\left(\mathbb{1}_A \underbrace{\frac{d\mathbb{Q}}{d\mathbb{P}} \times \frac{d\mathbb{P}}{d\mathbb{Q}}}_{=1}\right) \quad \text{car } \frac{d\mathbb{P}}{d\mathbb{Q}} = \frac{1}{\frac{d\mathbb{Q}}{d\mathbb{P}}}$$

- On prend un espace proba $(\Omega, \mathcal{A})$ avec 3 mesures $\mathbb{P}, \mathbb{Q}, R$ tq $R \ll \mathbb{Q}$ et $\mathbb{Q} \ll \mathbb{P}$ donc $R \ll \mathbb{P}$ et $\frac{dR}{d\mathbb{P}} = \frac{dR}{d\mathbb{Q}} \cdot \frac{d\mathbb{Q}}{d\mathbb{P}}$

Preuve: $R \ll \mathbb{P}$

$$\forall A \in \mathcal{A} \text{ tq } \mathbb{P}(A) = 0 \overset{\mathbb{Q} \ll \mathbb{P}}{\Longrightarrow} \mathbb{Q}(A) = 0 \Rightarrow R(A) = 0$$

$$\Rightarrow R \ll \mathbb{P}$$

$$\forall A \in \mathcal{A}, \quad R(A) = E^{R}(\mathbb{1}_A) \overset{R \ll \mathbb{Q} \Rightarrow \exists \frac{dR}{d\mathbb{Q}}}{=} E^{\mathbb{Q}}\left(\mathbb{1}_A \frac{dR}{d\mathbb{Q}}\right) \overset{\mathbb{Q} \ll \mathbb{P}}{=} E^{\mathbb{P}}\left(\mathbb{1}_A \underbrace{\frac{dR}{d\mathbb{Q}} \cdot \frac{d\mathbb{Q}}{d\mathbb{P}}}\right)$$

ou

$$\overset{R \ll \mathbb{P} \Rightarrow \exists \frac{dR}{dP}}{=} E^{\mathbb{P}}\left(\mathbb{1}_A \frac{dR}{dP}\right)$$

$=$

donc $\frac{dR}{d\mathbb{P}} = \frac{dR}{d\mathbb{Q}} \frac{d\mathbb{Q}}{d\mathbb{P}}$

Exercice : Soit U une v.a de Bernoulli sous la mesure $\mathbb{P}$

i.e $\mathbb{P}(U=1)=p$ et $\mathbb{P}(U=0)=1-p=q$

On pose $Y=\lambda U+\mu(1-U)$

1) Sous quelles conditions (sur λ et μ) a-t-on $E^{\mathbb{P}}(Y)=1$

2) Sous ces conditions, on a $\frac{dQ}{d\mathbb{P}}=Y$. Quelle est la loi de U sous Q ?

Solutions :

1) $E^{\mathbb{P}}(Y)=E^{\mathbb{P}}(\lambda U+\mu(1-U))$

$=\lambda E^{\mathbb{P}}(U)+\mu E^{\mathbb{P}}(1-U)$

$=\lambda p+\mu(1-p)=1$

Il faut que $\lambda p+\mu(1-p)=1$

2) On suppose $\lambda p+\mu(1-p)=1$. On a donc $E^{\mathbb{P}}(Y)=1$

Y définie une nouvelle mesure de probabilité Q tq $\frac{dQ}{d\mathbb{P}}=Y$

On veut $Q(U=1)$ et $Q(U=0)$

$Q(U=1)=E^{Q}(\mathbb{1}_{(U=1)})=E^{\mathbb{P}}\left(\mathbb{1}_{(U=1)}\frac{dQ}{d\mathbb{P}}\right)$

$=E^{\mathbb{P}}(\mathbb{1}_{(U=1)}\times(\lambda U+\mu(1-U)))$

$=\lambda\,\mathbb{P}(U=1)=\lambda p$

$$Q(U=0) = E^{Q}(\mathbb{1}_{(U=0)}) = E^{\mathbb{P}}\left(\mathbb{1}_{(U=0)} \frac{dQ}{d\mathbb{P}}\right)$$
$$= E^{\mathbb{P}}\left(\mathbb{1}_{(U=0)} \times (\lambda U + \mu(1-U))\right)$$
$$= \mu(1-p)$$

On a bien $Q(U=0) + Q(U=1) = \lambda p + \mu(1-p) = 1$

Sous Q, U suit une loi de Bernoulli de paramètre λp

Exercice : Sous la mesure $\mathbb{P}$ on a $U \sim \mathcal{N}(m_{\mathbb{P}}, \sigma_{\mathbb{P}}^2)$

On pose $Y = \exp\left(\lambda(U - m_{\mathbb{P}}) - \frac{\lambda^2}{2}\sigma_{\mathbb{P}}^2\right)$

1) Mq Y est une v.a positive tq $E^{\mathbb{P}}(Y) = 1$

2) Y définit une nouvelle proba Q tq $Y = \frac{dQ}{d\mathbb{P}}$

Quelle est la loi de U sous Q ?

Solutions :

$$E^{\mathbb{P}}(Y) = E^{\mathbb{P}}\left(\exp\left(\lambda(U - m_{\mathbb{P}}) - \frac{\lambda^2}{2}\sigma_{\mathbb{P}}^2\right)\right)$$
$$= \int \exp\left(\lambda(U - m_{\mathbb{P}}) - \frac{\lambda^2}{2}\sigma_{\mathbb{P}}^2\right) \frac{1}{\sqrt{2\pi}\,\sigma_{\mathbb{P}}} \exp\left(-\frac{1}{2\sigma_{\mathbb{P}}^2}(x - m_{\mathbb{P}})^2\right) dx$$

$$E^{\mathbb{P}}(Y) = \int \frac{1}{\sqrt{2\pi}\,\sigma_{\mathbb{P}}} \exp\left(-\frac{1}{2\sigma^2}\left((x^2 - 2x m_{\mathbb{P}} + m_{\mathbb{P}}^2) - 2\sigma^2\lambda(x - m_{\mathbb{P}}) + \lambda^2\sigma_{\mathbb{P}}^4\right)\right) dx$$

$$= \underbrace{\int \frac{1}{\sqrt{2\pi}\,\sigma_{\mathbb{P}}} \exp\left(-\frac{1}{2\sigma^2}\left(x - (m_{\mathbb{P}} + \lambda\sigma_{\mathbb{P}}^2)\right)^2\right)}_{\text{densité d'une loi normale } \mathcal{N}(m_{\mathbb{P}} + \lambda\sigma_{\mathbb{P}}^2,\ \sigma_{\mathbb{P}}^2)} dx$$

donc $= 1$

Loi de U sous Q ?

On prend une fonction quelconque f et on calcule

$$E^{Q}(f(U)) = \cdots = \int f(x) \overbrace{g(x)}^{\text{densité de } f(U) \text{ sous } Q} dx$$

$$= E^{\mathbb{P}}\left(f(U) \times \frac{dQ}{d\mathbb{P}}\right)$$

$$= E^{\mathbb{P}}\left(f(U) \times \exp\left(\lambda(U - m_{\mathbb{P}}) - \frac{\lambda^2}{2}\sigma_{\mathbb{P}}^2\right)\right)$$

$$= \int f(x) \exp\left(\lambda(x - m_{\mathbb{P}}) - \frac{\lambda^2}{2}\sigma_{\mathbb{P}}^2\right) \times \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - m_{\mathbb{P}})^2\right) dx$$

$$E^{Q}(f(U)) = \int f(x) \times \underbrace{\frac{1}{\sqrt{2\pi}\sigma_{\mathbb{P}}} \exp\left(-\frac{1}{2\sigma^2}\left(x - (m_{\mathbb{P}} + \lambda\sigma_{\mathbb{P}}^2)\right)^2\right)}_{\text{densité de } U \text{ sous } Q} dx$$

sous Q, $U \sim \mathcal{N}(m_{\mathbb{P}} + \lambda\sigma_{\mathbb{P}}^2, \sigma_{\mathbb{P}}^2)$

Conclusion : Sous la mesure $\mathbb{P}$, la v.a $U + \lambda\sigma_{\mathbb{P}}^2$ est une v.a Gaussienne (i.e $\mathcal{N}(m_{\mathbb{P}} + \lambda\sigma_{\mathbb{P}}^2, \sigma_{\mathbb{P}}^2)$) qui a même distribution que U sous la mesure Q

Exemple / exercice

Le prix d'une action sous la proba historique $\mathbb{P}$ est

$$S_t = x e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t}$$

où $x = 1$, $\mu \in \mathbb{R}$, $\sigma > 0$ et $W_t \sim \mathcal{N}(0, t)$

Soit r le taux sans risque. En finance, on sait faire des calculs sous la proba risque neutre sauf laquelle S_t suit une loi $\mathcal{LN}(r, \sigma^2)$

$$\frac{dQ}{d\mathbb{P}} = e^{\frac{\mu - r}{\sigma} W_t - \frac{1}{2}\left(\frac{\mu - r}{2}\right)^2 t}$$

sous Q, $S_t \sim \mathcal{LN}(r, \sigma^2)$

Chapitre 7 : Conditionnement

Objectif : On a $(\Omega, \mathcal{A}, \mathbb{P})$, X et Y deux v.a, $\mathcal{F}$ une sous tribu de $\mathcal{A}$, on veut définir $E(X|\mathcal{F})$ et $E(X|Y)$

Etapes : 1) On définit $E(X|A)$ ← évènement

2) ——— $E(X|\mathcal{F})$

3) ——— $E(X|Y) = E(X|\sigma(Y)) = h(Y)$

I - Rappels sur les probabilités conditionnelles

Définition : Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, B un évènement non négligeable (i.e. $\mathbb{P}(B) \neq 0$), on définit la probabilité d'un évènement A sachant B ou conditionnellement à B comme

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

Remarques

- Comme $\mathbb{P}(A|B) \leq \mathbb{P}(B)$ car $A \cap B \subseteq B$
  on a $0 \leq \mathbb{P}_B(A) \leq 1$

- L'application $\mathbb{P}(\cdot|B)$ tq $A \longrightarrow \mathbb{P}(A|B)$ définit bien une mesure de proba sur $(\Omega, \mathcal{A}) \Rightarrow \mathbb{P}_B$

- $\mathbb{P}_B$ est une mesure de proba absolument continue par rapport à $\mathbb{P}$. En effet $\forall A \in \mathcal{A}$ tq $\mathbb{P}(A) = 0$ alors $\mathbb{P}(A \cap B) = \mathbb{P}_B(A) = 0$ donc $\mathbb{P}_B \ll \mathbb{P}$

On peut calculer la densité de Radon Nykodym.

$$\mathbb{P}(A|B) = E^{\mathbb{P}_B}(\mathbb{1}_A) = E^{\mathbb{P}}\left(\mathbb{1}_A \frac{d\mathbb{P}_B}{d\mathbb{P}}\right) = \frac{P(A\cap B)}{\mathbb{P}(B)}$$

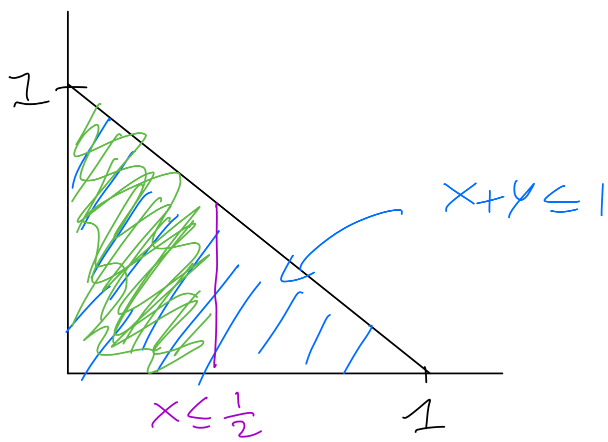
$$= \frac{E^{\mathbb{P}}(\mathbb{1}_{A\cap B})}{\mathbb{P}(B)} = \frac{E(\mathbb{1}_{A\cap B})}{\mathbb{P}(B)} = \frac{E^{\mathbb{P}}(\mathbb{1}_A \mathbb{1}_B)}{\mathbb{P}(B)}$$

$$= \frac{d\mathbb{P}_B}{d\mathbb{P}} = \frac{\mathbb{1}_B}{\mathbb{P}(B)}$$

Exercice : Soit (X, Y) un vecteur aléatoire de densité

$$f(x,y) = (x+y)\mathbb{1}_{[0,1]^2}(x,y)$$

On cherche $\mathbb{P}(2X \leq 1 \mid X+Y \leq 1)$
trouver $\mathbb{P}(X+Y \leq 1)$ et $\mathbb{P}((X+Y \leq 1) \cap (2X \leq 1))$



$(X+Y\leq 1)\cap(X\leq \frac{1}{2})$

$$\mathbb{P}(X+Y\leq 1)=\int_0^1\int_0^{1-x}(x+y)\,dy\,dx$$

$$=\int_0^1\left[xy+\frac{y^2}{2}\right]_0^{1-x}dx$$

$$=\int_0^1\left(x(1-x)+\frac{(1+x)^2}{2}\right)dx$$

$$\mathbb{P}(X+Y\leq 1)=\int_0^1\frac{1-x^2}{2}\,dx$$

$$=\left[\frac{x}{2}-\frac{x^3}{6}\right]_0^1$$

$$=\frac{1}{3}$$

$$\mathbb{P}((X+Y\leq 1)\cap(2X\leq 1))=\int_0^{\frac{1}{2}}\frac{1-x^2}{2}\,dx=\left[\frac{x}{2}-\frac{x^3}{6}\right]_0^{\frac{1}{2}}=\frac{11}{48}$$

$$\mathbb{P}(2X\leq 1\mid X+Y\leq 1)=\frac{11/48}{1/3}=\frac{11}{16}$$

Exercice : Soit $X\sim\mathcal{E}(\lambda)$ et x et y deux réels positifs

On veut $\mathbb{P}(X>x+y\mid X>x)=\dfrac{\mathbb{P}((X>x+y)\cap(X>x))}{\mathbb{P}(X>x)}$

$$=\frac{\mathbb{P}(X>x+y)}{\mathbb{P}(X>x)}=\frac{e^{-\lambda(x+y)}}{e^{-\lambda x}}=e^{-\lambda y}=\mathbb{P}(X>y)$$

C'est une propriété de la loi exponentielle, elle est sans mémoire. Quelque soit l'instant considéré, l'évolution future suit une loi exponentielle de même paramètre

Définition : Soit X une v.a définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On considère un évènement B non négligeable (i.e $\mathbb{P}(B) \neq 0$)
On définit la loi conditionelle de X sachant B comme la proba image de $\mathbb{P}_B(\cdot)$ par X définie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et qui à tout évènement $A \in \mathcal{B}(\mathbb{R}^d)$ associe

$$\mathbb{P}_{X|B}(A) = \mathbb{P}(X^{-1}(A) | B) = \mathbb{P}(X \in A | B)$$

Remarque : C'est comparable à la loi de la v.a qui est la mesure image de $\mathbb{P}$ par X.

$\mathbb{P}_{X|B}$ est définie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ alors que $\mathbb{P}$ et $\mathbb{P}_B$ sont définies sur $(\Omega, \mathcal{A})$
=> On a la mesure, on peut définir la f.d.r et la densité.

Definition: On définit la f.d.r de X sachant B

$$F_{X|B}(x) = \mathbb{P}_{X|B}(]-\infty, x]) = \mathbb{P}(X \leq x \mid B)$$

Si en plus $\mathbb{P}_{X|B}$ est abs. cont° par rapport à λ
(i.e $\mathbb{P}_{X|B} \ll \lambda$)
alors la densité conditionelle de X sachant B
$f_{X|B}$ existe et

$$\mathbb{P}(X \in A \mid B) = \mathbb{P}_{X|B}(A) = \int_A f_{X|B}(x)$$

On définit l'espérence conditionelle de X sachant B comme l'espérence de X sous la proba conditionelle sachant B

$$E(X|B) = E^{\mathbb{P}_B}(X) = \int x f_{X|B}(x)\,dx$$

$$= E^{\mathbb{P}}\left(X \frac{d\mathbb{P}_B}{d\mathbb{P}}\right) = E^{\mathbb{P}}\left(X \frac{\mathbb{1}_B}{\mathbb{P}(B)}\right)$$

$$= \frac{E^{\mathbb{P}}(X \mathbb{1}_B)}{\mathbb{P}(B)}$$

Rappel : On a un évènement B non négligeable, on définit $\mathbb{P}(\cdot \mid B)$ comme $\forall A \in \mathcal{A} \quad A \longrightarrow \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

On a $\mathbb{P}(\cdot \mid B) \ll \mathbb{P}$ et on a $\frac{d\mathbb{P}(\cdot \mid B)}{d\mathbb{P}(B)} = \frac{\mathbb{1}_B}{\mathbb{P}(B)}$

On peut alors calculer la loi d'une v.a X conditionnellement à B comme la mesure image de $\mathbb{P}$ par X

$F_{X|B}(x) = \mathbb{P}(X \leq x \mid B)$ et on peut calculer si elle existe, la densité $f_{X|B}(x)$

$$E(X \mid B) = E^{\mathbb{P}(\cdot \mid B)}(X) = E^{\mathbb{P}}\left(X \frac{\mathbb{1}_B}{\mathbb{P}(B)}\right) = \frac{E^{\mathbb{P}}(X \mathbb{1}_B)}{\mathbb{P}(B)}$$

Remarque : En particulier, si on prend un couple de v.a (X, Y) avec Y discrète tq $\mathbb{P}(Y = y_i) \neq 0$, on peut calculer

$$E(X \mid Y = y_j) = \begin{cases} \int x \, f_{X \mid Y = y_j}(x) \, dx \\ \sum_{x \in \mathcal{X}} \mathbb{P}(X = x \mid Y = y_j) \end{cases}$$

$\mathcal{X}$ = support de X i.e $\{x \in \mathbb{R} : \mathbb{P}(X = x) > 0$

On parle d'espérence conditionelle de X sachant $Y = y_j$

On peut construire l'espérence conditionelle de $X \mid Y$ comme

$$\underbrace{E(X \mid Y)}_{\text{devient une v.a}} = \sum_{y \in \text{Supp}(Y)} E(X \mid Y = y) \underbrace{\mathbb{1}_{\{Y = y\}}}_{\text{aléatoire}}$$

II. Espérence Conditionnelle

Idée : On sait comment construire l'espérence conditionnelle par rapport à un événement → on va essayer par rapport à une famille d'événements

Remarque : Si on a une variable aléatoire X (qu'on ne connait pas bien), la meilleure approximation possible de X par un réel est $E(X)$ (i.e le α qui minimise $E((X-a)^2)$)

Si on prend une tribu $\mathcal{F}$ (famille d'événements), on va construire l'espérence conditionelle de X sachant $\mathcal{F}$ comme la meilleure approx de X qui soit $\mathcal{F}$ mesurable

info en plus sur X

Exemple : Soit Ω un univers, $\mathcal{F}$ tribu, $\mathbb{P}$ proba

On prend $(B_i)_{i\in\mathbb{N}}$ un système complet d'événements mesurables (ie. $\bigcup B_i = \Omega$, $\forall i \neq j$ $B_i \cap B_j = \emptyset$)

On note $\sigma(\{B_i\}_{i\in\mathbb{N}})$ la tribu engendrée par $\{B_i\}_{i\in\mathbb{N}}$

Toute variable aléatoire $\sigma(\{B_n\})$ mesurable est de la forme

$$\sum_{n\in\mathbb{N}} b_n \mathbb{1}_{B_n}$$

La variable aléatoire est constante sur chaque B_n

On note $\mathbb{P}_n = \mathbb{P}(\cdot \mid B_n)$

On prend une v.a Y définie sur $(\Omega, \mathcal{A})$ et on veut la meilleur approx de Y qui soit $\sigma(\{B_n\}_n)$ mesurable

On cherche la v.a $\sigma(\{B_n\}_n)$ mesurable qui minimise $E((Y-X)^2)$

$$\Rightarrow E\Big(\big(Y - \sum_{n\in\mathbb{N}} b_n \mathbb{1}_{B_n}\big)^2\Big) = E\Big(\big(\sum (Y-b_n)\mathbb{1}_{B_n}\big)^2\Big)$$

$$\underset{\text{disjoints}}{\overset{B_n \text{ sont}}{=}} \sum_{n\in\mathbb{N}} E\big((Y-b_n)^2 \mathbb{1}_{B_n}\big)$$

$$= \sum_{n\in\mathbb{N}} \mathbb{P}(B_n) \times E^{\mathbb{P}_n}\big((Y-b_n)^2\big)$$

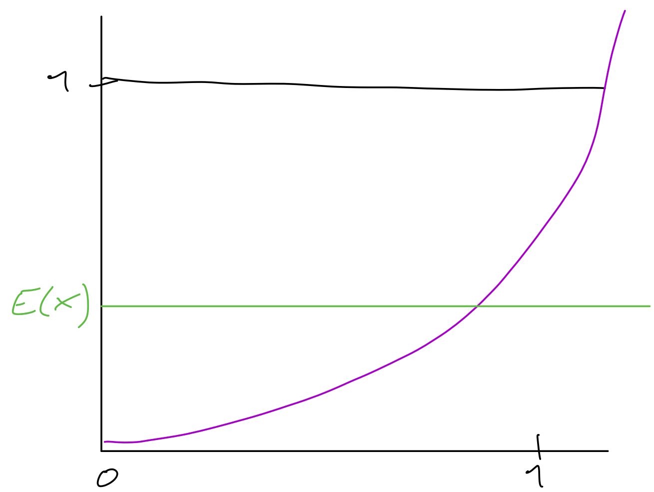
On a vu que le b_n qui minimise $E^{\mathbb{P}_n}((Y-b_n)^2)$ est

$$E^{\mathbb{P}_n}(Y) = \frac{E(Y\mathbb{1}_{B_n})}{\mathbb{P}(B_n)}$$

La meilleure approx de Y par une v.a $\sigma(\{B_n\}_n)$ mesurable est $\sum_{n\in\mathbb{N}} E^{\mathbb{P}_n}(Y)\, \mathbb{1}_{B_n}$

Ca définit $E(Y \mid \sigma(\{B_n\}_n))$ et on peut le voir comme la projection orthogonale de Y sur le sous espace vectoriel des v.a $\sigma(\{B_n\})$ mesurables.

Exemple : On considère $\Omega = (0,1]$, muni de la mesure de Lebesgue et on prend la v.a X tq $X(\omega) = \omega^2$



Si on a aucune info sur ω et on doit prédire $X(\omega)$, la meilleure approx est $E(X) = \int \omega^2 d\omega = \frac{1}{3}$

On prend la tribu $\mathcal{F}_n = \sigma\left(\left(\frac{j}{2^n}, \frac{j+1}{2^n}\right], j = 0, \dots, 2^n - 1\right)$

On veut calculer $E(X \mid \mathcal{F}_n)$. Comme les $\left(\frac{j}{2^n}, \frac{j+1}{2^{n+1}}\right]$ forment un système complet, on a

$$E(X \mid \mathcal{F}_n) = \sum_j \frac{E\left(X \mathbb{1}_{\left(\frac{j}{2^n}, \frac{j+1}{2^n}\right]}\right)}{\mathbb{P}\left(\left(\frac{j}{2^n}, \frac{j+1}{2^n}\right]\right)} \mathbb{1}_{\left(\frac{j}{2^n}, \frac{j+1}{2^n}\right]} \qquad \leftarrow = \int_{\frac{j}{2^n}}^{\frac{j+1}{2^n}} \omega^2 d\omega$$

longueur de l'intervalle

Si $n = 1$, $\mathcal{F}_1 = \sigma\left(\left(0, \tfrac{1}{2}\right], \left(\tfrac{1}{2}, 1\right]\right)$

$$E(X \mid \mathcal{F}_1) = \frac{E\left(X \mathbb{1}_{(0, \frac{1}{2}]}\right)}{\mathbb{P}\left((0, \frac{1}{2}]\right)} \mathbb{1}_{(0, \frac{1}{2}]} + \frac{E\left(X \mathbb{1}_{(\frac{1}{2}, 1]}\right)}{\mathbb{P}\left((\frac{1}{2}, 1]\right)} \mathbb{1}_{(\frac{1}{2}, 1]}$$

$$= \frac{1}{12} \mathbb{1}_{(0, \frac{1}{2}]} + \frac{7}{12} \mathbb{1}_{(\frac{1}{2}, 1]}$$

$$E(X \mid \mathcal{F}_2) = \frac{E\left(X \mathbb{1}_{(0, \frac{1}{4}]}\right)}{\mathbb{P}\left((0, \frac{1}{4}]\right)} \mathbb{1}_{(0, \frac{1}{4}]} + \dots + \frac{E\left(X \mathbb{1}_{(\frac{3}{4}, 1]}\right)}{\mathbb{P}\left((\frac{3}{4}, 1]\right)} \mathbb{1}_{(\frac{3}{4}, 1]}$$

$$= \frac{1}{48} \mathbb{1}_{(0, \frac{1}{4}]} + \dots + \frac{37}{48} \mathbb{1}_{(\frac{3}{4}, 1]}$$

Définition : Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $\mathcal{F}$ une sous tribu de $\mathcal{A}$, X une v.a.r définie sur $(\Omega, \mathcal{A})$. $\mathbb{P}$ intégrable (i.e $E^{\mathbb{P}}(X) < \infty$)

On définit l'espérance conditionnelle de X sachant $\mathcal{F}$ comme la v.a Y qui vérifie (et on note $Y = E(X \mid \mathcal{F})$

1- Y est mesurable par rapport à $\mathcal{F}$

ie $\forall B \in \mathcal{B}(\mathbb{R}) \quad Y^{-1}(B) \in \mathcal{F}$

2- $\forall A \in \mathcal{F}$ on a $\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$

ou

$$E^{\mathbb{P}}(X \mathbb{1}_A) = E^{\mathbb{P}}(Y \mathbb{1}_A)$$

$E(X \mid \mathcal{F})$ est la meilleure approx de X (X et $E(X \mid \mathcal{F})$ sont égaux en moyenne sur tous les éléments de $\mathcal{F}$) qui soit $\mathcal{F}$ mesurable

Remarque : $\mathcal{F}$ doit être vue comme une information additionelle et $E(X \mid \mathcal{F})$ est la meilleure approximation de X qui en tient compte

Proposition :

Si $\mathcal{A} = \mathcal{F}$ alors $E(X \mid \mathcal{F}) = X$ ou si X est $\mathcal{F}$ mesurable alors $E(X \mid \mathcal{F}) = X$

Si $\mathcal{F} = \{\emptyset, \Omega\}$ on a $E(X|\mathcal{F}) = E(X)$

Si $E(X) = E(E(X|\mathcal{F}))$

Si X est constante, $E(X|\mathcal{F}) = E(X) = X$

Proposition : (Existence et unicité de $E(X|\mathcal{F})$

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathcal{F}$ sous tribu de $\mathcal{A}$
Alors

- $E(X|\mathcal{F})$ existe
- Si Z et Z' sont deux espérances conditionnelles de X sachant $\mathcal{F}$ alors $Z = Z'$ $\mathbb{P}$-p.s

Preuve (existence) :

Soit $X > 0$, Soit $Q : \mathcal{F} \to \mathbb{R}$ tq $\forall B \in \mathcal{F}$

$Q(B) = E^{\mathbb{P}}(X \mathbb{1}_B)$

Q est une mesure positive et $Q \ll \mathbb{P}$

Q est une mesure bornée car $Q(\Omega) = E^{\mathbb{P}}(X) < \infty$

D'après Radon-Nykodym, il existe une densité $\frac{dQ}{d\mathbb{P}}$
i.e une variable aléatoire positive tq

$\forall B \in \mathcal{F}$ $E^{\mathbb{P}}(X \mathbb{1}_B) = E(\frac{dQ}{d\mathbb{P}} \mathbb{1}_B)$

De plus $\frac{dQ}{d\mathbb{P}}$ est $\mathcal{F}$ mesurable
donc $\frac{dQ}{d\mathbb{P}}$ est une espérance conditionnelle de X sachant $\mathcal{F}$.

Propriétés : Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace proba, $X, Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathcal{F}$ sous tribu de $\mathcal{A}$

- Linéarité

$$E(aX+bY|\mathcal{F}) = aE(X|\mathcal{F}) + bE(Y|\mathcal{F})$$

- Positivité

Si $X \geq 0$ p.s alors $E(X|\mathcal{F}) \geq 0$ p.s

- Croissance

Si $X \leq Y$ alors $E(X|\mathcal{F}) \leq E(Y|\mathcal{F})$

- $(E(X|\mathcal{F}))_+ \leq E(X_+|\mathcal{F})$

Proposition : Soit $(\Omega, \mathcal{A}, \mathbb{P})$ espace proba, $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et $\{B_n\}_{n\in\mathbb{N}}$ un système complet et $\mathcal{F} = \sigma(\{B_n\}_{n\in\mathbb{N}})$

$$E(X|\mathcal{F}) = \sum_{n\in\mathbb{N}} \frac{E(X \mathbb{1}_{B_n})}{\mathbb{P}(B_n)} \mathbb{1}_{B_n}$$

Proposition : Soit $X, Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, $\mathcal{F}$ ss tribu de $\mathcal{A}$

1- Si XY est intégrable et si X est $\mathcal{F}$ mesurable

$$E(XY|\mathcal{F}) = X E(Y|\mathcal{F})$$

2- Jensen : Si f est convexe et $f(x)$ intégrable

$$f(E(X|\mathcal{F})) \leq E(f(X)|\mathcal{F})$$

3. Si $\mathcal{F} \subset \mathcal{G} \subset \mathcal{A}$ alors

$$E(E(X|\mathcal{F})|\mathcal{G}) = E(X|\mathcal{F})$$

$$E(E(X|\mathcal{G})|\mathcal{F}) = E(X|\mathcal{F})$$

Exemple / exo :

Soit X v.a définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ tq $X \sim Poi(\lambda)$

On pose $Y = 2\lfloor \frac{X}{2} \rfloor$ (si $X=1 \Rightarrow Y=0$, $X=2 \Rightarrow Y=2$)

Calculer $E(X|\sigma(Y))$ et $E(Y|\sigma(X)) = Y$

1) On cherche les atomes de $\sigma(Y)$

Si B est un atome alors si $A \subset B$, $A = \emptyset$ ou $A = B$

Les atomes ici sont $B_n = \{Y = 2n\}$ avec $n \in \mathbb{N}$

$\Rightarrow$ Ils forment un système complet de $\sigma(Y)$

On a alors $E(X|\sigma(Y)) = \sum_{n\in\mathbb{N}} \frac{E(X 1\!\!1_{B_n})}{\mathbb{P}(B_n)} 1\!\!1_{B_n} = \sum \frac{E(X 1\!\!1_{(Y=2n)})}{\mathbb{P}(Y=2n)} 1\!\!1_{(Y=2n)}$

On doit calculer

$$E(X 1\!\!1_{(Y=2n)}) = E(X 1\!\!1_{(X=2n)} + X 1\!\!1_{(X=2n+1)})$$

$1\!\!1_{((X=2n)\cup(X=2n+1))}$

$$= 2n\,\mathbb{P}(X=2n) + (2n+1)\,\mathbb{P}(X=2n+1)$$

$$= 2n \frac{\lambda^{2n}}{(2n)!} e^{-\lambda} + (2n+1) \frac{\lambda^{2n+1}}{(2n+1)!} e^{-\lambda}$$

$$\mathbb{P}(B_n) = \mathbb{P}(Y=2n) = \mathbb{P}(X=2n) + \mathbb{P}(X=2n+1)$$

$$= \frac{\lambda^{2n}}{(2n)!} e^{-\lambda} + \frac{\lambda^{2n+1}}{(2n+1)!} e^{-\lambda} = \frac{\lambda^{2n}}{(2n)!} e^{-\lambda} \left(\frac{2n+1+\lambda}{2n+1}\right)$$

Donc $\frac{E(X 1\!\!1_{B_n})}{\mathbb{P}(B_n)} = \frac{(2n+\lambda)(2n+1)}{(2n+1+\lambda)}$

$$E(X \mid \sigma(Y)) = \sum_{n \in \mathbb{N}} \frac{(2n+\lambda)(2n+1)}{(2n+1+\lambda)} \mathbb{1}_{(Y=2n)}$$

$$= \sum_{y \in \mathrm{Supp}(Y)} \frac{(y+\lambda)(y+1)}{(y+1+\lambda)} \mathbb{1}_{(Y=y)}$$

$$= \frac{(Y+\lambda)(Y+1)}{(Y+1+\lambda)}$$

Rappel : On a $(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathcal{F}$ une sous tribu et une v.a $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$

On définit $E(X \mid \mathcal{F})$ comme la v.a Y tq

1) $\forall A \in \mathcal{F} \quad E(X \mathbb{1}_A) = E(Y \mathbb{1}_A)$ ou $\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$

2) Y est $\mathcal{F}$ mesurable i.e $\forall B \in \mathcal{B}(\mathbb{R})$, $Y^{-1}(B) \in \mathcal{F}$

- $E(X \mid \mathcal{F})$ est la meilleure approximation de X qui soit $\mathcal{F}$ mesurable

Remarque : $\mathcal{F}$ doit être vue comme de l'info additionnelle

Définition : Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et $Y_1, \dots Y_n$ sont définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. On définit l'espérance conditionnelle de X sachant $Y_1, \dots, Y_n$ comme $E[X \mid Y_1, \dots, Y_n]$

$$= E[X \mid \sigma(Y_1, \dots, Y_n)]$$

Proposition : (Lemme de Doob)

Si on a X une v.a.r Y-mesurable (i.e mesurable par rapport à la tribu engendrée par Y) alors il existe une fonction mesurable g tq $X = g(Y)$

Or comme $E(X \mid Y_1, \dots, Y_n)$ est $\sigma(Y_1, \dots, Y_n)$ mesurable, il existe une fonction mesurable g tq
$E(X \mid Y_1, \dots, Y_n) = g(Y_1, \dots, Y_n)$

Remarque : Si $\forall y \in \mathbb{R}$ on a $E(X \mid Y = y) = g(y)$ alors
$E(X \mid Y) = g(Y)$

Si X et Y sont indépendants, $E[X \mid Y] = E[X]$

Propriétés : Soit (X, Y) une v.a.r φ une fct mesurable sur $\mathbb{R}^2$ et h une fct mesurable sur $\mathbb{R}$
Si $\varphi(X,Y)$ et $h(Y)\varphi(X,Y)$ sont intégrables on a

$$E(E(\varphi(X,Y) \mid Y)) = E(\varphi(X,Y))$$

$$E(h(Y)\varphi(X,Y) \mid Y) = \underbrace{h(y)}_{\sigma(Y) \text{ mesurable}} E(\varphi(X,Y) \mid Y)$$

Définition : Variance conditionnelle

On a $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathcal{F}$ ss tribu de $\mathcal{A}$. On définit la variance conditionelle de X sachant $\mathcal{F}$ comme

$$V(X|\mathcal{F}) = E((X - E(X|\mathcal{F}))^2 | \mathcal{F}))$$

Proposition : $V(X|\mathcal{F}) = E(X^2|\mathcal{F}) - E(X|\mathcal{F})^2$

preuve : $V(X|\mathcal{F}) = E((X - E(X|\mathcal{F}))^2|\mathcal{F})$

$= E((X^2 - 2XE(X|\mathcal{F}) + E(X|\mathcal{F})^2|\mathcal{F})$

$= E(X^2|\mathcal{F}) - 2E(XE(X|\mathcal{F})|\mathcal{F}) + E(E(X|\mathcal{F})^2|\mathcal{F})$

$= E(X^2|\mathcal{F}) - 2E(X|\mathcal{F})E(X|\mathcal{F}) + E(X|\mathcal{F})^2$

$= E(X^2|\mathcal{F}) - E(X|\mathcal{F})^2$

Formule de décomposition de la variance

Soit (X, Y) un v.a sur $(\Omega, \mathcal{A}, \mathbb{P})$, X de carré intégrable alors

$$V(X) = E(V(X|Y)) + V(E(X|Y))$$

Preuve : $V(E(X|Y)) = E(E(X|Y)^2) - \overbrace{E(E(X|Y))^2}^{= E(X)^2}$

$E(V(X|Y)) = \underbrace{E(E(X^2|Y))}_{E(X^2)} - E(E(X|Y)^2)$

$\leftarrow$ Rappel $E(E(f(X)|Y)) = E(f(X))$

$V(E(X|Y)) + E(V(X|Y)) = E(X^2) - E(X)^2 = V(X)$

Application : On a n contrats homogènes, chaque contrat a la même proba d'avoir un sinistre et chaque sinistre suit la même loi

Les pertes sur le portefeuille sont modélisés par

$$Z = \sum_{i=0}^{N} Y_i \qquad N \leftarrow N \sim \mathcal{P}(\lambda)$$

$\hookleftarrow$ Y_i suite de v.a.i.i.d Eq

$Y_i \sim \Gamma(a,b)$ ou $LN(\mu, \sigma^2)$

et Y_i indep de N

On veut $E(Z)$ et $V(Z)$. Si on fixe $N=n$, on a

$$E(Z|N=n) = E\left(\sum Y_i\right) = n\,E(Y_i)$$

Si $Y_i \sim \Gamma(a,b)$, $E(Y_i) = \frac{a}{b}$ et $V(Y_i) = \frac{a}{b^2}$

$$E(Z|N=n) = n\frac{a}{b}, \quad E(Z|N) = N \cdot \frac{a}{b}$$

$$E(Z) = E(E(Z|N)) = E\left(N\frac{a}{b}\right) = \frac{a}{b}E(N) = \frac{a}{b}\lambda$$

Pareil : $V(Z|N=n) = V\left(\sum Y_i\right) = n\frac{a}{b^2}$

$V(Z|N) = N\frac{a}{b^2}$

$V(Z) = V(E(Z|N)) + E(V(Z|N)) = V\left(N\frac{a}{b}\right) + E\left(N\frac{a}{b^2}\right)$

$= \frac{a^2}{b^2}\lambda + \frac{a}{b^2}\lambda = \frac{\lambda a(a+1)}{b^2}$

Calcul de esperences conditionnelles

Définition : (Lois de probas conditionelles)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, X et Y deux v.a.r sur $(\Omega, \mathcal{A}, \mathbb{P})$. On appelle loi conditionnelle de X sachant $Y=y$ la loi de proba sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ qui a tout borélien B associe $\mathbb{P}(X^{-1}(B) \mid Y=y)$

Propriété : Soit (X,Y) un couple de v.a.r de densité $f_{X,Y}(x,y)$ par rapport a Lebesque sur $\mathbb{R}^2$ et de marginales f_X et f_Y on note

$$f_{X|Y=y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Alors $\forall A \in \mathcal{A}$. $\mathbb{P}(X \in A \mid Y=y) = \int_A f_{X|Y=y}(x|y)\,dx$

Soit g une fct mesurable tq $g(X,Y)$ est $\mathbb{P}$ intégrable alors

$$E(g(X,Y) \mid Y=y) = \underbrace{\int_{\mathbb{R}} g(x,y) f_{X|Y=y}(x|y)\,dx}_{\text{dépend uniquement de } y}$$

en particulier

$$E(X|Y=y) = \int_{\mathbb{R}} x f_{X|Y=y}(x|y)\,dx$$

Si en plus $\forall y \in \mathbb{R}$, on a $E(X|Y=y) = h(y)$ alors $E(X|Y) = h(Y)$

Cas discret : (X, Y) discrète

On a $\mathbb{P}(X = x_i \mid Y = y) = \dfrac{\mathbb{P}(X = x_i, Y = y)}{\mathbb{P}(Y = y)}$

et donc $E(X|Y=y) = \sum\limits_{\substack{x_i \in supp(X) \\ \text{support}}} x_i \times \mathbb{P}(X = x_i \mid Y = y)$

Exemple exercice :

On prend $U_1 \sim \mathcal{U}[0,1]$ et $U_2 \sim \mathcal{U}[0,1]$ et $U_1 \perp U_2$

On pose $X = \min(U_1, U_2)$ et $Y = \max(U_1, U_2)$

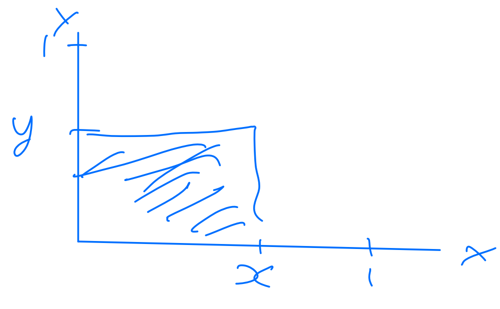
On veut calculer $E(X|Y)$ (et éventuellement $E(Y|X)$)

Il faut calculer $E(X|Y=y)$, donc il faut $f_{X|Y=y}(x|y)$ et $f_{X,Y}$ et f_Y. Attention $\mathcal{L}(X) \neq \mathcal{L}(U_1)$

Loi de Y

$$\mathbb{P}(Y \leq y) = \mathbb{P}(\max(U_1, U_2) \leq y) = \mathbb{P}(U_1 \leq y, U_2 \leq y)$$
$$= \mathbb{P}(U_1 \leq y)\,\mathbb{P}(U_2 \leq y)$$
$$= \mathbb{P}(U_1 \leq y)^2 = y^2$$

$f_Y(y) = 2y \, \mathbb{1}_{(0 \leq y \leq 1)}$



Loi de (X, Y)

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

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

$$= y^2 - \underbrace{\mathbb{P}(\min(U_1,U_2) > x, \max(U_1,U_2) \leq y)}_{\text{i.e } U_1 \text{ et } U_2 \text{ sont entre } x \text{ et } y}$$

$$= \mathbb{P}(x < U_1 \leq y, x < U_2 \leq y) = (y-x)^2$$

$$\Rightarrow \mathbb{P}(X \leq x, Y \leq y) = y^2 - (y-x)^2 \quad \text{si } y \geq x$$

$$f_{X,Y}(x,y) = 2 \mathbb{1}_{(y > x)}$$

$$f_{X|Y=y}(x|y) = \frac{1}{y} \mathbb{1}_{(y>x)}$$

$$E(X|Y=y) = \int x f_{X|Y=y}(x|y)\, dx = \int \frac{x}{y} \mathbb{1}_{(y>x)}\, dx$$

$$= \int_0^y \frac{x}{y}\, dx = \left[\frac{x^2}{2y}\right]_0^y = \frac{y}{2}$$

$$E(X|Y) = \frac{Y}{2}$$

On prend $U_1, \dots, U_n$ i.i.d tq $U_i \sim \mathcal{U}[0,1]$

$X_n = \min\limits_{i=1..n}$ $\qquad Y_n = \max\limits_{i=1,\dots,n} U_i$

Calculer $E(X_n|Y_n) = \frac{Y_n}{n}$ à vérifier

Cas particulier: Vecteurs gaussiens

Proposition: Soit (X,T) un vecteur Gaussien de $\mathbb{R}^n$ ($X \in \mathbb{R}^p$, $T \in \mathbb{R}^q$ et $p+q=n$). On suppose que la matrice de variance de T notée Γ_T est inversible alors

$$E(X|T) = E(X) + \mathrm{cov}(X,T)\,\Gamma_T^{-1}(T-E(T))$$

$E(X|T)$ est Gaussienne

Théorème: On suppose que le vecteur $(X, Y_1, \dots, Y_n)$ est gaussien alors $\exists\, a, b_1, \dots, b_n \in \mathbb{R}$ tq

$$E(X|Y_1,\dots,Y_n) = a + \sum_{i=1}^{n} b_i Y_i$$

et la loi conditionnelle de X sachant $Y_1=y_1 \dots Y_n=y_n$ est Gaussienne

Remarque: On a que si X est de carré intégrable alors $Z = E(X|Y)$ est la projection orthogonale de X sur le ss espace vectoriel des v.a $\sigma(Y)$ mesurable. donc $X-Z$ est orthogonale à Y et donc $E((X-Z)Y)=0$

En pratique, si on a v.a centrées, $X, Y_1, \dots, Y_n$ pour calculer $Z = E(X \mid Y_{1,\dots,n})$ on pose $Z = \sum a_i Y_i$

Les a_i sont calculés à partir des conditions d'orthogonalité

$$\begin{cases} Z = \sum_{i=1}^{n} a_i Y_i \\ E((X-Z)Y_i) = 0 \quad \forall i = 1, \dots, n \end{cases}$$

Les valeurs $E(XY_i)$ et $E(Y_i Y_j)$ sont données par les matrices de covariance de $(X, Y_1, \dots, Y_n)$

Exemple : Soit (X, Y_1, Y_2) un vecteur Gaussien centré de matrice de covariance $\begin{pmatrix} 5 & 0 & 3 \\ 0 & 1 & -1 \\ 3 & -1 & 4 \end{pmatrix}$

et on veut calculer $E(X \mid Y_1, Y_2)$

On pose $X = a_1 Y_1 + a_2 Y_2$ et on résoud

$$\begin{cases} E((X - a_1 Y_1 - a_2 Y_2) Y_1) = 0 \\ E((X - a_1 Y_1 - a_2 Y_2) Y_2) = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} E(XY_1) - a_1 E(Y_1^2) - a_2 E(Y_1 Y_2) = 0 \\ E(XY_2) - a_1 E(Y_1 Y_2) - a_2 E(Y_2^2) = 0 \end{cases}$$

$E(XY_1)$ $E(XY_2)$ $E(Y_1^2)$

$$\begin{array}{c|ccc} & X & Y_1 & Y_2 \\ \hline X & & & \\ Y_1 & & \begin{pmatrix} 5 & 0 & 3 \\ 0 & 1 & -1 \\ 3 & -1 & 4 \end{pmatrix} & \\ Y_2 & & & \end{array}$$

Car centré !

$$\begin{cases} -a_1 + a_2 = 0 \\ 3 + a_1 - 4a_2 = 0 \end{cases}$$

$$\Rightarrow \begin{cases} a_1 = a_2 \\ a_1 = 1 \end{cases}$$

$$E(X \mid Y_1, Y_2) = Y_1 + Y_2 = \sum_{i=1}^{2} a_i Y_i = 1(Y_1) + 1(Y_2)$$