Probabilités - 21h00 –

UP1 – Probabilités et Statistique Avancées
Majeure Science des Données 2020-2021

Objectif. Apporter les outils et concepts indispensables à la bonne compréhension et utilisation des méthodes et techniques statistiques qui seront présentées dans les 3 autres UPs de la Majeure « Science des Données » :

UP2 : Apprentissage statistique et automatique
UP3 : Données temporelles et analyse spectrale
UP4 : Exploitation mathématique de simulateurs numériques

Contenu.

Ch. 1 (12h00) : Vecteurs Gaussiens (VGs)
Ch. 2 (9h00) : Lois Conditionnelles – Processus Aléatoires

Vecteurs Gaussiens (algèbre, géométrie, probabilités)

$\Rightarrow$ **Analyse de données, Régression, Plan d'Expériences, …**

Lois conditionnelles $\Rightarrow$ Processus Aléatoires

$\Rightarrow$ **Séries chronologiques, …**

Cours: slides de cours + fiches TD et TP + supports sur Campus en complément

I.1. Loi gaussienne sur $\mathbb{R}$

X v.a.r. est dite de loi *normale* (ou *gaussienne*) de *moyenne* $\mu \in \mathbb{R}$ et de *variance* $\sigma^2 > 0$ (*écart-type* $\sigma > 0$) si sa *loi de probabilité* admet la *densité* suivante :

$$\mathbf{f_X(\mathit{x}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{(\mathit{x} - \mu)^2}{2\sigma^2} \right)}$$

Ainsi,

$$P(X \in [a, b]) = \int_a^b f_X(x)dx$$

On notera $X \sim N(\mu, \sigma^2)$.

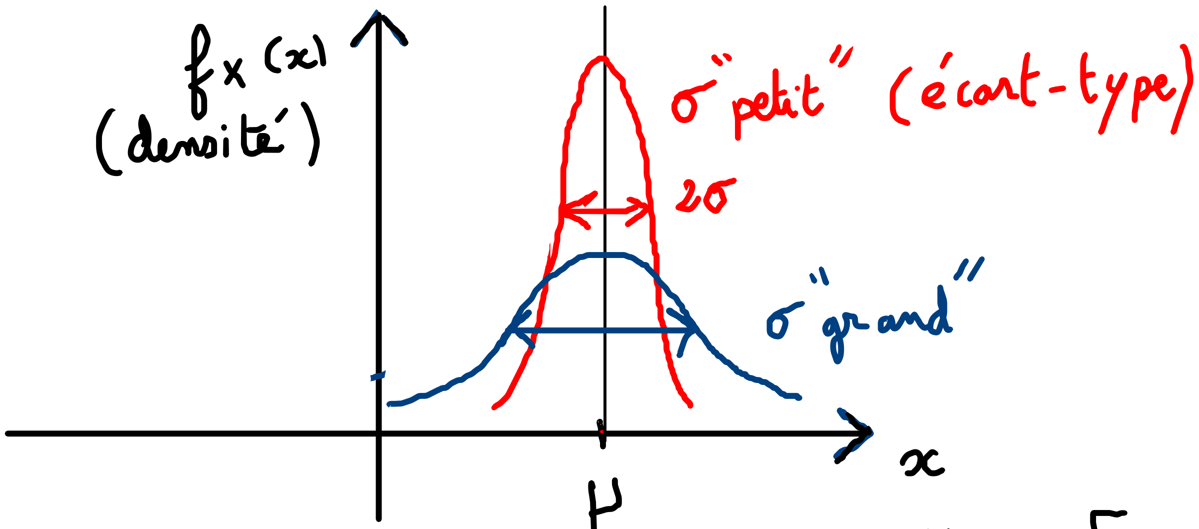
Convention. $X = \mu$ est gaussienne dégénérée, i.e. $X \sim N(\mu,0)$.

A savoir sur la loi $N(\mu,\sigma^2)$:

$$X \sim N(\mu,\sigma^2) \iff X = \mu + \sigma Z \text{ avec } Z \sim N(0,1)$$

$$\mu \to EX \qquad \sigma^2 \to Var(X) = E[(X-EX)^2] = EX^2 - (EX)^2$$

$$Var(X) = 0 \iff X = EX = \mu \quad (X \text{ dégénérée})$$



Intervalle de fluctuations à 95% : $[\mu - 1.96\sigma, \mu + 1.96\sigma] \approx \mu \pm 2\sigma$

De plus : $\mu \pm \sigma \mapsto \approx 68\%$ et $\mu \pm 3\sigma \mapsto \approx 99,7\%$

Importance de la loi normale – Théorème Central Limite (TCL):

$X_1, X_2, \ldots$ **indépendantes et identiquement distribuées** (en abrégé i.i.d.) de moyenne μ et de variance σ^2

Considérons la suite des sommes partielles : $\mathbf{S_n = X_1 + \ldots + X_n}$ $(n \geq 1)$.

Alors, pour tout $a < b$:

$$\mathbf{P\left(a \leq \frac{\sqrt{n}}{\sigma} \times \left(\frac{S_n}{n} - \mu \right) \leq b \right) = P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b \right) \to_n \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} \right) dx}$$

En pratique : $\quad \dfrac{S_n}{n} \approx_{\text{n grand}} N\left(\mu, \dfrac{\sigma^2}{n}\right) \Leftrightarrow S_n \approx_{\text{n grand}} N(n\mu, n\sigma^2)$

Si les X_k sont $N(\mu, \sigma^2)$, alors $\dfrac{S_n}{n}$ est de loi $N\left(\mu, \dfrac{\sigma^2}{n}\right)$ et S_n est de loi $N(n\mu, n\sigma^2)$.

Exemple classique d'application : estimation d'une proportion p

$X_1, X_2, \ldots$ i.i.d. de loi de Bernoulli $B(p)$ avec $0 < p < 1$ (où $p = P(X = 1)$).

Alors, **$S_n = X_1 + \ldots X_n$ de loi binomiale $B(n, p)$** et

- **S_n à peu près de loi $N(np, np(1-p))$ si n grand**
- **$\dfrac{S_n}{n} = p + \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} Z$ où Z est à peu près de loi $N(0, 1)$ si n grand**

Estimation de p : $\dfrac{S_n}{n}$ estimateur naturel de p (estimateur de la moyenne)

$$IC_{95\%} = [\ \hat{p} - 1.96 \times \sqrt{\frac{\hat{p}\,(1-\hat{p}\,)}{n}}\ ;\ \hat{p} - 1.96 \times \sqrt{\frac{\hat{p}\,(1-\hat{p}\,)}{n}}\ ]$$

Détails sur estimation de p (proba/proportion/taux/etc.)

$X \in \{0,1\}$, $p = \mathbb{P}(X=1)$ (et donc $1-p = \mathbb{P}(X=0)$)

$X_1, \dots, X_n$ i.i.d. de même loi que X (n épreuves de Bernoulli)

$$\begin{cases} S_n = X_1 + \dots + X_n = \text{"nombre de fois où 1 apparaît"} \\ \dfrac{S_n}{n} \text{ estimateur naturel de } p \end{cases}$$

Alors
$$\begin{cases} S_n \text{ de loi } \mathcal{B}(n,p) \quad \text{(loi binomiale)} \\ S_n \approx \mathcal{N}(np, np(1-p)) \\ \dfrac{S_n}{n} \approx \mathcal{N}\left(p, \dfrac{p(1-p)}{n}\right) \end{cases}$$ si n assez grand

Remarquer que : $E\left(\dfrac{S_n}{n}\right) = p$ (estimateur sans biais)

et $\mathrm{Var}\left(\dfrac{S_n}{n}\right) = \dfrac{p(1-p)}{n} \xrightarrow[n]{} 0$ (estimateur convergent)

Intérêt pratique

① Intervalle de confiance à 95% sur p

$$IC_{95\%} = \hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

où $\hat{p} := \bar{x} = \dfrac{x_1 + \cdots + x_n}{n}$ (réalisation de $\bar{X} = \dfrac{S_n}{n}$)

désigne l'estimation ponctuelle de p (= nombre calculé à partir de l'échantillon $x_1, x_2, \cdots, x_n$)

$\longrightarrow$ $IC_{95\%}$ estimation par intervalle

ou fourchette d'estimation

② On peut tester l'hypothèse $H_0 : p = p_0$ contre $H_1 : p > p_0$

(penser au COVID et p_0 seuil critique de % de "positifs" dans la population par exemple)

I.2. Vecteurs aléatoires réels

Définition. $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ où $X_1, \ldots, X_d$ sont des v.a. est appelé **vecteur aléatoire réel (V.a.r.) de dimension d (ou V.a.r. d-dimensionnel)**.

Exemple 1 – cas discret : X v.a. discrète à K modalités codées de $k = 0$ à $k = K - 1$

On note $p_k = P(X = k)$ pour $0 \leq k \leq K - 1$ de sorte que

$$p_0 = P(X = 0) = 1 - p_1 - \ldots - p_{K-1}$$

On cherche à estimer les probabilités p_k (proportions) à partir d'un échantillon de taille n de la loi de X. On considère donc $X_1, \ldots, X_n$ i.i.d. de même loi que X.
On pose pour $0 \leq k \leq K - 1$:

$$N_k = \mathbf{1}_{X_1 = k} + \ldots + \mathbf{1}_{X_n = k} \text{ le nombre d'occurrence de la modalité n°k}$$

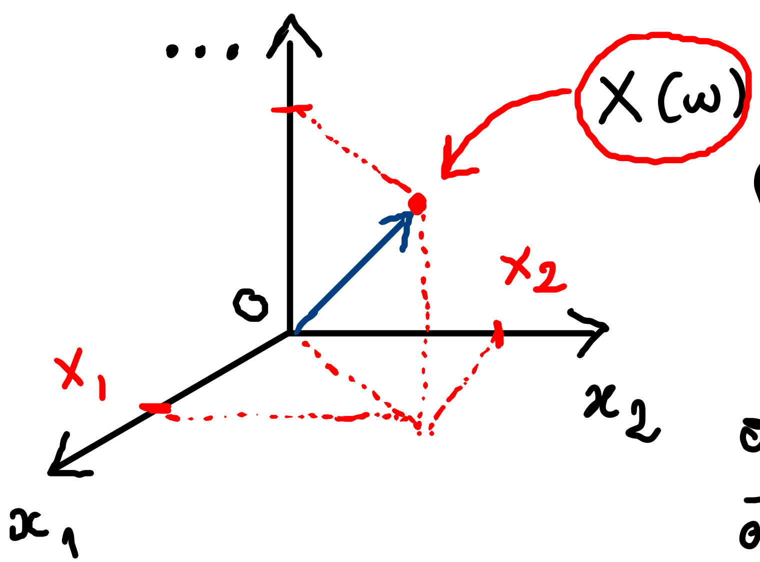
sachant que $N_0 = n - N_1 - \ldots - N_{K-1}$

☞ **$N = (N_1, \ldots, N_{K-1})$ V.a.r. de dimension $(K - 1)$ et discret**

Commentaire sur la terminologie X "vecteur" aléatoire :

Si l'on dispose de l'observation de plusieurs variables $X_1, \dots, X_d$, il est naturel de les considérer conjointement (pour apprendre les dépendances éventuelles)

$$\Longrightarrow \quad X = (X_1, X_2, \dots, X_d) \in \mathbb{R}^d$$



$X(\omega)$ est à la fois un point aléatoire de $\mathbb{R}^d$ (en rouge) et un V. a. (en bleu)

$\mathbb{R}^d$ a une structure d'espace vectoriel où l'addition des éléments (points) correspond à l'addition usuelle des vecteurs ...

Loi de N : c'est la loi dite **multinomiale**

$$P(N = (n_1, \ldots, n_{K-1})) = \frac{n!}{n_1!\times\ldots\times n_{K-1}!(n - n_1 - \ldots - k_{K-1})!} p_1^{n_1} \times\ldots\times p_{K-1}^{n_{K-1}} \times (1 - p_1 - \ldots - p_{K-1})^{n - n_1 - \ldots - n_{K-1}}$$

- $K = 2$: loi binomiale
- $K = 3$: loi trinomiale

☞ loi étudiée au TP1 en liaison avec le TCL multi-dimensionnel

Exercice 1.

1. Déterminer les lois marginales du vecteur N
2. Que peut-on dire du V.a. $(N_0, N_1, \ldots, N_{K-1})$?

Cas de la loi trinomiale : $X \in \{0,1,2\}$ à $k=3$ modalités

Exemple : enquête de satisfaction ("marketing")

$$X = \begin{cases} \text{"Pas du tout satisfait"} \longrightarrow 0 : p_0 \\ \text{" Satisfait "} \longrightarrow 1 : p_1 \\ \text{" Très satisfait "} \longrightarrow 2 : p_2 \end{cases}$$

(codage)

→ p_k proportions dans la clientèle que l'on veut estimer/apprendre

→ sondage aléatoire/randomisé de n clients où l'on suppose $n << N$ = taille population globale

⟹ données :

x_1	x_2	...			...	x_n
1	0	2	2	0	...	1

$\boxed{x_1, x_2, \dots, x_n \text{ réalisations de } X_1, \dots, X_n \text{ i.i.d. de loi la loi de } X \text{ ternaire avec } P(X=k) = p_k}$

Comme $p_0 = 1 - p_1 - p_2$, il suffit d'estimer p_1, p_2 et bien sûr :

$$\begin{cases} \hat{p}_1 = \dfrac{N_1}{n} & \text{" fréquence avec laquelle apparaît la modalité 1 "} \\ \hat{p}_2 = \dfrac{N_2}{n} & \text{" ——— 2 "} \end{cases}$$

Formellement, $N_1 = \mathbb{1}_{X_1 = 1} + \cdots + \mathbb{1}_{X_n = 1}$

$=$ " nbre de clients satisfaits "

$N_2 = \mathbb{1}_{X_1 = 2} + \cdots + \mathbb{1}_{X_n = 2}$

$=$ " nbre de clients très satisfaits "

et $N_0 = n - N_1 - N_2 =$ " nbre de clients pas du tout satisfaits "

$$\Longrightarrow \hat{p}_0 = \frac{N_0}{n} = 1 - \hat{p}_1 - \hat{p}_2$$

Loi de (N_1, N_2) = loi trinomiale :

$$\mathbb{P}(N_1 = n_1 \text{ et } N_2 = n_2) = \frac{n!}{\underbrace{n_1!\, n_2!\, (n-n_1-n_2)!}_{\text{coefficient trinomial}}} \underbrace{p_1^{n_1}\, p_2^{n_2}\, (1-p_1-p_2)^{n-n_1-n_2}}_{\text{par indépendance}}$$

où $\begin{cases} 0 \leq n_1 \leq n \\ 0 \leq n_2 \leq n \\ n_1 + n_2 \leq n \end{cases}$

(cf. $(a+b+c)^n = \ \dots$)

En effet : $$\mathbb{P}(N_1 = n_1 \text{ et } N_2 = n_2) = \underbrace{\binom{n}{n_1} \times \binom{n}{n-n_1}}_{= \text{ coeff. trinomial après simplification !}} \times \text{ "}$$

Solution exercice 1 dans le cas $k = 3$ (adapter pour k qcque)

1. loi de N_1 = loi $\mathcal{B}(n, p_1)$

   loi de N_2 = loi $\mathcal{B}(n, p_2)$

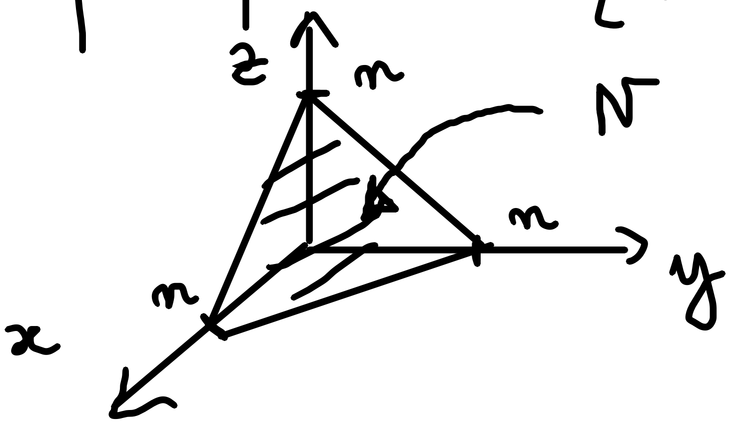
puisque, par exemple,

$$N_1 = \mathbb{1}_{X_1 = 1} + \dots + \mathbb{1}_{X_n = 1}$$

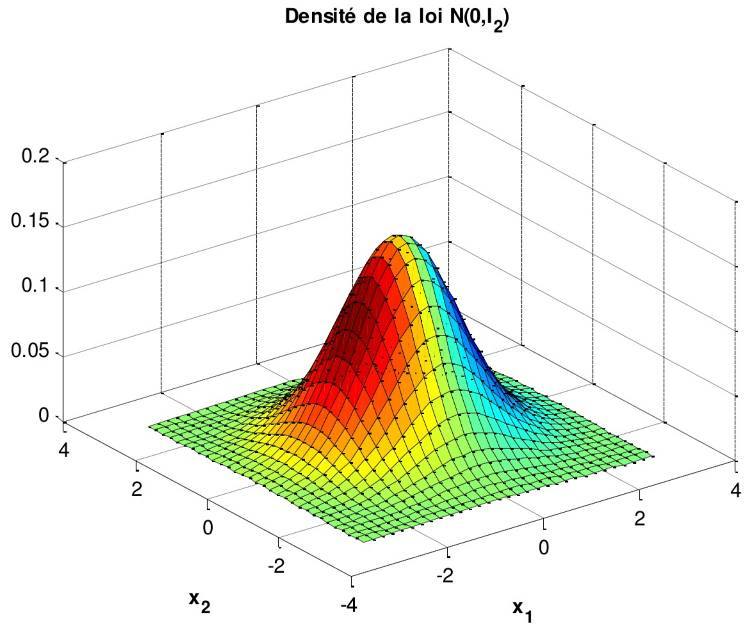
et $\mathbb{1}_{X_i = 1}$ de loi $\mathcal{B}(p_1)$

2. $N = (N_0, N_1, N_2)$ est un v.a. dégénéré de $\mathbb{R}^3$

puisque $N \in \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = n\}$ (plan)



Exemple 2 – cas continu : $X = (X_1, X_2)$ où X_1, X_2 sont i.i.d. de loi $N(0, 1)$



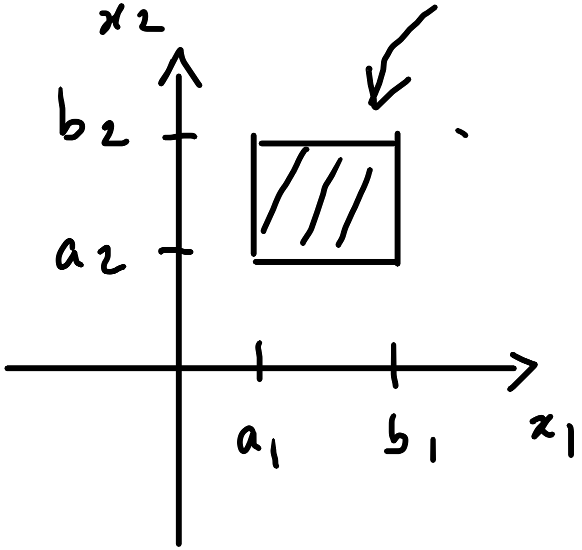
X admet la densité de probabilité suivante :

$$\boldsymbol{x = (x_1, x_2) \to \frac{1}{2\pi} \exp\left(-\frac{\|x\|^2}{2}\right)}$$

☞ cloche de Gauss bi-dimensionnelle

Exemple 2 : quelques explications

$$P\big((X_1,X_2)\in \underbrace{[a_1,b_1]\times[a_2,b_2]}_{\text{rectangle } R \text{ de } \mathbb{R}^2}\big) = P\big(X_1\in[a_1,b_1] \ \underline{\text{et}}\ X_2\in[a_2,b_2]\big)$$



par définition de X_1, X_2 v.a. indépendantes

$$= P\big(X_1\in[a_1,b_1]\big)\times P\big(X_2\in[a_2,b_2]\big)$$

$$= \left(\int_{a_1}^{b_1} \frac{e^{-\frac{x_1^2}{2}}}{\sqrt{2\pi}}\,dx_1\right)\times\left(\int_{a_2}^{b_2} \frac{e^{-\frac{x_2^2}{2}}}{\sqrt{2\pi}}\,dx_2\right)$$

$$= \iint_R \frac{e^{-\frac{\|x\|^2}{2}}}{2\pi}\,dx_1\,dx_2$$

donc $X=(X_1,X_2)$ v.a. continue de densité :

$$x=(x_1,x_2)\longrightarrow \frac{1}{2\pi}\,e^{-\frac{\|x\|^2}{2}} =: f_X(x)$$

On peut vérifier que : $f_X(x)\geq 0$

et $c = \iint_{\mathbb{R}^2} f_X(x) \underbrace{dx}_{= dx_1 dx_2} = 1$ puisque

$$c = \int_0^{+\infty} \int_0^{2\pi} \frac{1}{2\pi} e^{-r^2/2} r\,dr\,d\theta = \int_0^{+\infty} r e^{-r^2/2} dr$$

$$\text{soit } c = \left[- e^{-r^2/2} \right]_0^{+\infty} = 1 \; !$$

Ainsi, on a 2 exemples simples de :

① X V.a. <u>discret</u> avec (N_1, N_2)

② X V.a. <u>continu</u> avec $X = (X_1, X_2)$

où X_1, X_2 iid $\mathcal{N}(0,1)$.

Définition (V.a. continu). $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ V.a.r. est dit continu de densité $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d \to \mathbf{f_X}(\boldsymbol{x}) \geq 0$ si (par définition)

$$\mathbf{P}(\mathbf{X} \in [\mathbf{a_1}, \mathbf{b_1}] \times \ldots \times [\mathbf{a_d}, \mathbf{b_d}]) = \int_{[[\mathbf{a}, \mathbf{b}]]} \mathbf{f_X}(\boldsymbol{x}) \mathbf{d}\boldsymbol{x}$$

où $[[a, b]]$ désigne un hyper-rectangle quelconque $[a_1, b_1] \times \ldots \times [a_d, b_d]$ de $\mathbb{R}^d$.

Lois marginales d'un V.a.r. : ce sont les lois des composantes $X_1, \ldots, X_d$. Dans le cas continu, elles admettent des densités notées respectivement $f_{X_1}, \ldots, f_{X_d}$.

Exercice 2. Exprimer les densités marginales $f_{X_1}, \ldots, f_{X_d}$ en fonction de la densité conjointe $x = (x_1, \ldots, x_d) \to f_X(x)$ du V.a. continu X. Comparer avec le cas discret.

Cas indépendance : $X_1, \ldots, X_d$ indépendantes $\Leftrightarrow f_X(x) = f_{X_1}(x_1) \times \ldots \times f_{X_d}(x_d)$

__Explications__ $f_X(x) = f_X(x_1, \dots, x_d) \geq 0$ et

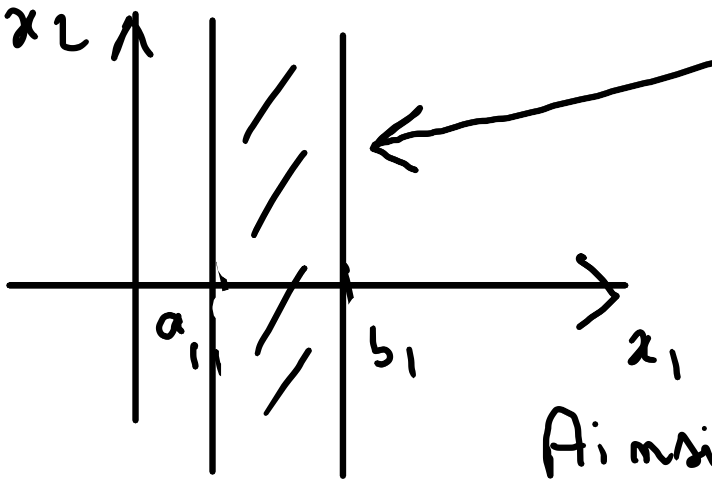
$$\int_{\mathbb{R}^d} f_X(x) \underbrace{dx}_{= dx_1 \cdots dx_d} = \int \cdots \int f_X(x_1, \dots, x_d)\, dx_1 \cdots dx_d = 1$$

pour que f_X soit une densité de probabilité d'1 v.a. continue.

$$\boxed{\text{Lois marginales} = \text{lois des composantes ou marges } X_1, \dots, X_d.}$$

__Solution exercice 2 dans le cas $d = 2$ (généralisation facile) :__

$$P(X_1 \in [a_1, b_1]) = P((X_1, X_2) \in \underbrace{[a_1, b_1] \times \mathbb{R}})$$



$$= \int_{a_1}^{b_1} \underbrace{\left(\int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2)\, dx_2 \right)} dx_1$$

FUBINI

$$= f_{X_1}(x_1) \text{ densité de } X_1$$

Ainsi : $f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_{(X_1, X_2)}(x_1, x_2)\, dx_2$

C'est analogue au cas discret puisque :

$$X \in \{x_1, \cdots, x_K\} \; ; \; p_{k\ell} = \mathbb{P}\big((X,Y) = (x_k, y_\ell)\big)$$
$$Y \in \{y_1, \cdots, y_L\} \qquad = \mathbb{P}(X = x_k \text{ et } Y = y_\ell)$$

($p_{k\ell}$ = loi conjointe)

Alors

$$\mathbb{P}(X = x_k) = \sum_{\ell=1}^{L} \mathbb{P}(X = x_k \text{ et } Y = y_\ell)$$

soit
$$\boxed{\mathbb{P}(X = x_k) = \sum_{\ell=1}^{L} p_{k\ell} \quad \text{pour } 1 \leq k \leq K}$$

A comparer avec $f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\,dy$ <u>en continu</u>.

Même formule pour la loi marginale associée à Y !

Cas de l'indépendance : précisons l' $\Longleftrightarrow$:

① si $X_1, \dots, X_d$ sont des v.a. continues, de densités respectives $f_{X_1}(x_1), \dots, f_{X_d}(x_d)$ alors $X = (X_1, \dots, X_d)$ V.a. continu de densité

$$f_X(x) = f_{X_1}(x_1) \times \dots \times f_{X_d}(x_d)$$

En effet, $\mathbb{P}(X \in [[a,b]]) = \prod_{i=1}^{d} \mathbb{P}(X_i \in [a_i, b_i])$ par indépendance

$$= \prod_{i=1}^{d} \int_{a_i}^{b_i} f_{X_i}(x_i)\, dx_i$$

$$= \int_{[[a,b]]} \underbrace{f_{X_1}(x_1) \dots f_{X_d}(x_d)}_{= f_X(x)} dx_1 \dots dx_d$$

Donc X continu de densité $f_X(x)$!

② Si X V.a continue de densité

$$f_X(x) = f_{X_1}(x_1) \cdots f_{X_d}(x_d)$$

alors $X_1, \dots, X_d$ indépendants !

En effet, $\displaystyle \mathbb{P}(X \in [[a,b]]) = \int_{[[a,b]]} f_X(x)\,dx$

$$= \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f_{X_1}(x_1) \cdots f_{X_d}(x_d)\,dx_1 \dots dx_d$$

$$= \prod_{i=1}^{d} \int_{a_i}^{b_i} f_{X_i}(x_i)\,dx_i$$

$$= \prod_{i=1}^{d} \mathbb{P}(X_i \in [a_i, b_i]) \quad //$$

Définition – moyenne et matrice de (variance-)covariance d'un V.a.r. X d-dimensionnel :

- $\mu = (EX_1, \ldots, EX_d) \in \mathbb{R}^d$ **moyenne de X : $\mu = EX$**
- $\Gamma = E[\,(X-\mu)(X-\mu)^T\,] = (\,E[(X_i-\mu_i)(X_j-\mu_j)]\,)_{i,\,j}$ matrice de taille $d\times d$ appelée **covariance de X : $\Gamma = \mathrm{Cov}(X)$**

Attention à l'abus de notation : Cov(X) = matrice de covariance

Propriété. La matrice Γ de terme général $\mathbf{\Gamma_{ij} = Cov(X_i, X_j)}$ est une matrice symétrique positive. Sa diagonale est formée des variances individuelles des variables : $\mathbf{\Gamma_{ii} = Cov(X_i, X_i) = Var(X_i)}$.

Exercice 3. Calculer moyenne et covariance du vecteur (N_1, N_2) associé à une loi trinomiale. En déduire la corrélation entre N_1 et N_2.

Exercice 4. Si A est une application linéaire de $\mathbb{R}^d$ dans $\mathbb{R}^d$ (identifiée à une matrice $d\times d$) et b un vecteur de $\mathbb{R}^d$, calculer la moyenne et covariance du vecteur aléatoire $Y = AX + b$.

Explications sur moyenne et covariance d'un v.a. $X=(X_1,\dots,X_d)$:

On souhaite généraliser les notions de moyenne et variance de X v.a. $\in \mathbb{R}$ $(d=1)$

D'où $EX = (EX_1,\dots,EX_d) \iff \mu = (\mu_1,\dots,\mu_d)$

$\rightsquigarrow \mu$ vecteur des moyennes individuelles

Pour définir la covariance, on utilise le calcul matriciel avec la convention usuelle :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \text{ vecteur colonne ou matrice } d\times 1$$

De même donc $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix}$ et $X-\mu = \begin{pmatrix} X_1-\mu_1 \\ \vdots \\ X_d-\mu_d \end{pmatrix}$

Ainsi : $\underset{(d\times 1)\times(1\times d)}{(X-\mu)(X-\mu)^T} = \begin{pmatrix} X_1-\mu_1 \\ \vdots \\ X_d-\mu_d \end{pmatrix} \begin{pmatrix} X_1-\mu_1 & \dots & X_d-\mu_d \end{pmatrix}$

$$= \begin{pmatrix} (X_i-\mu_i)(X_j-\mu_j) \end{pmatrix}_{i,j} \text{ de taille } d\times d$$

De là $\mathrm{Cov}(X) = E\left[(X-\mu)(X-\mu)^T\right] = \left(E\left[(X_i-\mu_i)(X_j-\mu_j)\right]\right)_{i,j}$.

sachant que l'on définit

$$E(M) = \left(EM_{ij}\right)_{i,j}.$$

matrice dont les termes sont des v.a.r.

(généralise $E\left[\begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}\right] = \begin{pmatrix} EX_1 \\ \vdots \\ EX_d \end{pmatrix}$)

On voit donc que

$$\mathrm{Cov}(X) = \left(\mathrm{Cov}(X_i, X_j)\right)_{1 \leq i,j \leq d}$$

où $\mathrm{Cov}(X_i, X_j)$ est la covariance entre les deux v.a.r. X_i et X_j.

Rappel : Si U, V sont deux v.a. réelles, alors

$$\mathrm{Cov}(U,V) = E\big[(U - EU)(V - EV)\big] = E(UV) - E(U) \times E(V)$$

$$\mathrm{Cov}(U,U) = \mathrm{Var}(U) \;!$$

De plus, $\rho_{U,V} = \dfrac{\mathrm{Cov}(U,V)}{\underbrace{\sqrt{\mathrm{Var}(U)}}_{\sigma_U}\underbrace{\sqrt{\mathrm{Var}(V)}}_{\sigma_V}} \in [-1 ; +1]$ coefficient de corrélation linéaire

Bien remarquer que $(U,V) \longrightarrow \mathrm{Cov}(U,V) \in \mathbb{R}$

est une forme bilinéaire symétrique positive !

(la variance est la forme quadratique associée)

Retenir que

$$Cov(X) = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots \\ Cov(X_2, X_1) & Var(X_2) & \\ \vdots & & \ddots \\ & & & Var(X_d) \end{pmatrix}$$

est la matrice des variances-covariances

du V. a. $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$.

C'est la généralisation de $\sigma^2 = Var(X)$ du cas $d = 1$ au cas $d \geqslant 2$: $Cov(X)$ intègre les corrélations linéaires entre les composantes 2 à 2 puisque :

$$\mathrm{Cov}(X) = \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & & \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_d^2 \end{pmatrix}$$

En notant $\begin{cases} \sigma_i^2 = \mathrm{Var}(X_i) \\ \rho_{i,j} = \mathrm{Cor}(X_i, X_j) = \dfrac{\mathrm{Cov}(X_i, X_j)}{\sigma_i \sigma_j} \in [-1, +1] \end{cases}$

<u>Sur la propriété</u> : Cov(X) matrice symétrique <u>positive</u>

Résulte de : $0 \leq \mathrm{Var}(a_1 X_1 + \dots + a_d X_d) = a^T \Gamma \times a$

en notant $a = \begin{pmatrix} a_1 \\ \vdots \\ a_d \end{pmatrix}$

Solution exercice 3 : $N = (N_1, N_2)$ où (rappel)

$$\begin{cases} N_1 = \mathbb{1}_{X_1=1} + \cdots + \mathbb{1}_{X_n=1} \sim \mathcal{B}(n, p_1) \\ N_2 = \mathbb{1}_{X_1=2} + \cdots + \mathbb{1}_{X_n=2} \sim \mathcal{B}(n, p_2) \end{cases}$$

Donc $\mathrm{Var}(N_1) = n p_1(1-p_1)$ et $\mathrm{Var}(N_2) = n p_2(1-p_2)$

Puis, $\mathrm{Cov}(N_1, N_2) = E(N_1 N_2) - (E N_1) \times (E N_2)$

On a
$$E(N_1 N_2) = \sum_{i,j} E\left[\mathbb{1}_{X_i=1} \mathbb{1}_{X_j=2}\right]$$
$$= \sum_{i=1}^{n} E\Big[\underbrace{\mathbb{1}_{X_i=1}\mathbb{1}_{X_i=2}}_{=0}\Big] + \sum_{i \neq j}$$
$$= \sum_{i \neq j} \mathbb{P}(X_i = 1)\, \mathbb{P}(X_j = 2) \quad \text{(par indépendance)}$$

De là,
$$\mathrm{Cov}(N_1, N_2) = (\cancel{n^2} - n)\, p_1 p_2 - (\cancel{n p_1})(\cancel{n p_2})$$
$$= - n p_1 p_2$$

et donc :

$$\mathrm{Cov}(N) = \begin{pmatrix} n p_1 (1-p_1) & -n p_1 p_2 \\ & \\ -n p_1 p_2 & n p_2 (1-p_2) \end{pmatrix}$$

On en déduit $\rho_{N_1, N_2} = -\sqrt{\dfrac{p_1 p_2}{(1-p_1)(1-p_2)}}$

Dans le cas $p_0 = 0$, $p_1 + p_2 = 1$

et $\rho_{N_1, N_2} = -1$

ce qui est logique puisque alors $N_1 + N_2 = n$
$(N_0 = 0)$.

Solution exercice 4

(connaître le résultat ou savoir refaire ce type de calcul)

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix} = \underbrace{A}_{\substack{\text{matrice } d\times d \\ \text{déterministe}}} X + b \quad \text{où } X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} ; \ b = \begin{pmatrix} b_1 \\ \vdots \\ b_d \end{pmatrix} \in \mathbb{R}^d$$

Y = transformation linéaire (affine) du v.a. X

On a
$$\begin{cases} \mu_Y = A\mu_X + b \\ Cov(Y) = A \, Cov(X) \, A^T \end{cases}$$

En effet, comme l'espérance est linéaire, on a

$$\begin{cases} EY_1 = a_{11}\mu_1 + \cdots + a_{1d}\mu_d + b_1 \\ \vdots \\ EY_d = a_{d1}\mu_1 + \cdots + a_{dd}\mu_d + b_d \end{cases} \quad \text{si } \mu = \mu_X$$

Soit donc $\mu_Y = A\mu_X + b$

On dit que l'opérateur E "traverse" la matrice A

puisque $E(AX+b) = A\,EX + b$!

Puis, $Y - \mu_Y = AX + \cancel{b} - (A\mu_X \cancel{+b})$

soit $Y - \mu_Y = A(X-\mu)$

partie aléatoire

DONC $(Y-\mu_Y)(Y-\mu_Y)^T = A\,\underbrace{\overbrace{(X-\mu)(X-\mu)^T}}\,A^T$

et, de même, $\mathrm{Cov}(Y) = E[\ \ ] = A\,E[\ \ ]\,A^T$

soit $\mathrm{Cov}(Y) = A\,\mathrm{Cov}(X)\,A^T$ //

Petite illustration : cf. TP 0 (étude de la loi normale bidimensionnelle)

$$\begin{cases} X = aU + bV \\ Y = cU + dV \end{cases} \quad \text{avec } U, V \text{ i.i.d. } \mathcal{N}(0,1)$$

Alors $\begin{pmatrix} X \\ Y \end{pmatrix} = \underbrace{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}_{= A} \begin{pmatrix} U \\ V \end{pmatrix}$ avec $\mathrm{Cov}\left(\begin{pmatrix} U \\ V \end{pmatrix}\right) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$

$$\mathrm{Cov}\left(\begin{pmatrix} X \\ Y \end{pmatrix}\right) = A I_2 A^T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a^2+b^2 & ac+bd \\ ca+db & c^2+d^2 \end{pmatrix}$$

$$\Longrightarrow \mathrm{Cov}\left(\begin{pmatrix} X \\ Y \end{pmatrix}\right) = \begin{pmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{pmatrix} \quad //$$

Le calcul direct donne bien en effet $\mathrm{Var}(X) = a^2+b^2$, $\mathrm{Var}(Y) = c^2+d^2$ et $\mathrm{Cov}(X,Y) = ac+bd$ ...

Définition (première approche) : Un vecteur aléatoire réel X = $(X_1, \ldots, X_d)^T \in \mathbb{R}^d$ est dit ***gaussien*** *(ou multi-gaussien)* non dégénéré si

- sa matrice de covariance Γ est ***inversible*** (ou *non dégénérée*)
- et s'il admet la densité dite (multi-)gaussienne ou normale suivante :

$$x \to \frac{1}{\sqrt{\det(\Gamma)}\ \sqrt{2\pi}^{\,d}} \exp\Big(-\frac{1}{2} \underbrace{< x-\mu \mid \Gamma^{-1}(x-\mu) >}_{Q(x) = (x-\mu)^T \Gamma^{-1} (x-\mu)}\Big)$$

On vérifie que cette expression définit bien une densité associée à un vecteur aléatoire de moyenne μ et de covariance Γ.

$d=1: \quad Q(x) = \dfrac{(x-\mu)^2}{\sigma^2}$

On note encore **N(μ, Γ)** la loi normale multivariée de moyenne μ et de covariance Γ.

Comment retenir / retrouver cette expression analytique

Prenons $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \dots, X_d \sim \mathcal{N}(\mu_d, \sigma_d^2)$ <u>indépendants</u>

Alors

$$f_X(x) = \prod_{i=1}^{d} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x_i - \mu_i)^2}{\sigma_i^2}}$$

$$f_X(x) = \frac{1}{(\sigma_1 \cdots \sigma_d) \sqrt{2\pi}^d} e^{-\frac{1}{2} (x-\mu)^T \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_d^2 \end{pmatrix} (x-\mu)}$$

Or $\Gamma = \mathrm{Cov}(X) = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix}$

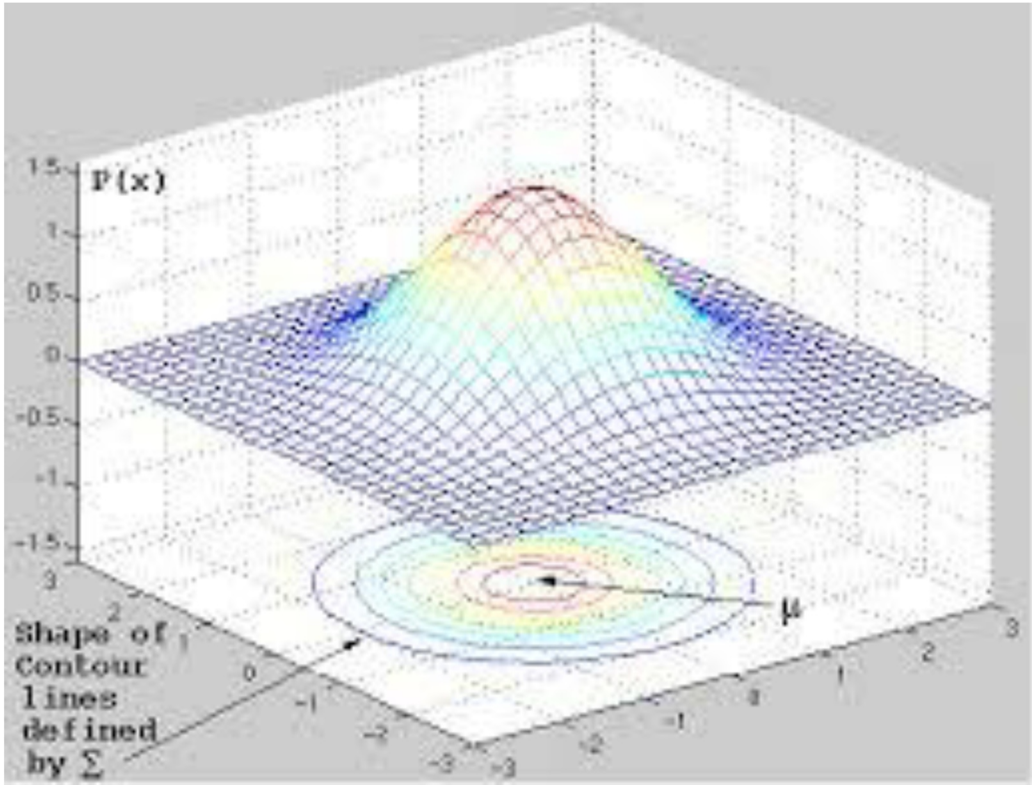
donc $$\boxed{f_X(x) = \frac{1}{\sqrt{\det(\Gamma)} \sqrt{2\pi}^d} e^{-\frac{1}{2} (x-\mu)^T \Gamma^{-1} (x-\mu)}}$$

Ce calcul montre qu'un v.a. $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ dont les composantes sont des v.a.r. gaussiennes indépendantes (et non dégénérées : $\sigma_i > 0$) est bien un vecteur gaussien (VG)

Par ailleurs dans le cas $d = 1$ où $X \sim \mathcal{N}(\mu, \sigma^2)$, on a $\operatorname{Cov}(X) = \sigma^2$ (scalaire) et

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

Exercice 5. Expliciter la densité d'un vecteur gaussien de dimension 2 et représenter sa densité.



Solution exercice 5

$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ vecteur gaussien non dégénéré

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{et} \quad \Gamma = \mathrm{Cov}(X) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

où $\sigma_1^2 = \mathrm{Var}(X_1)$; $\sigma_2^2 = \mathrm{Var}(X_2)$; $\rho = \mathrm{Cor}(X_1, X_2)$

On a $\det(\Gamma) = \sigma_1^2\sigma_2^2(1-\rho^2) \geq 0$ car $|\rho| \leq 1$

Γ inversible $\Longleftrightarrow$ $|\rho| < 1 \Longleftrightarrow \rho \neq +1$ et -1

On a :

$$\Gamma^{-1} = \frac{1}{\det(\Gamma)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

et donc

$$f_X(x) = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2} \times 2\pi} e^{-\frac{1}{2} Q(x)}$$

où $Q(x) = \frac{1}{\det(\Gamma)} \begin{pmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{pmatrix} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}$

En développant (forme quadratique en x_1, x_2) :

$$Q(x) = \frac{1}{\det(\Gamma)} \left\{ \sigma_2^2 (x_1 - \mu_1)^2 + \sigma_1^2 (x_2 - \mu_2)^2 - 2\rho\sigma_1\sigma_2 (x_1 - \mu_1)(x_2 - \mu_2) \right\}$$

et (rappel)

$$f_X(x) = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2} \times 2\pi} e^{-\frac{1}{2} Q(x)}$$

Tout cela est compliqué du fait de Γ^{-1} que l'on ne calcule jamais dès que $d \geq 3$ (sauf cas simples ...).

Néanmoins, il est facile de se représenter cette densité par ses lignes/courbes de niveau :

$$f_X(x) = \text{cste} \quad \text{ssi} \quad (x-\mu)^T \Gamma^{-1} (x-\mu) = \text{cste}$$

OR : Γ symétrique positive se diagonalise en base orthonormée

$$\Gamma U = U \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \iff \Gamma = U \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} U^T$$

où $U = (\underbrace{U_1 \, U_2}_{\text{vecteurs colonne}})$ = matrice des vecteurs propres (normés)

vérifie $U^T U = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

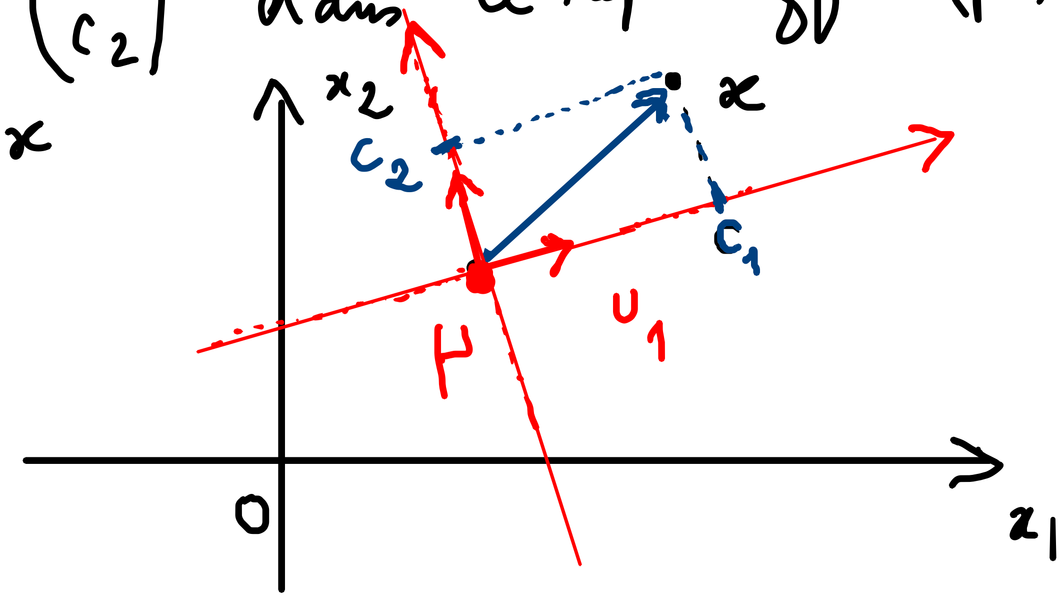
$(U^{-1} = U^T \iff U$ matrice orthogonale$)$

De là, $\Gamma^{-1} = U \begin{pmatrix} 1/\lambda_1 & 0 \\ 0 & 1/\lambda_2 \end{pmatrix} U^T$

et :

$$(x-\mu)^T \Gamma^{-1} (x-\mu) = \text{cste} \iff \left[U^T(x-\mu)\right]^T \begin{pmatrix} 1/\lambda_1 & 0 \\ 0 & 1/\lambda_2 \end{pmatrix} \left[U^T(x-\mu)\right]$$

Or $U^T(x-\mu) = \begin{pmatrix} U_1^T(x-\mu) \\ U_2^T(x-\mu) \end{pmatrix}$ est le vecteur des coordonnées $\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ dans le repère affine $(\mu ; U_1, U_2)$ du point x

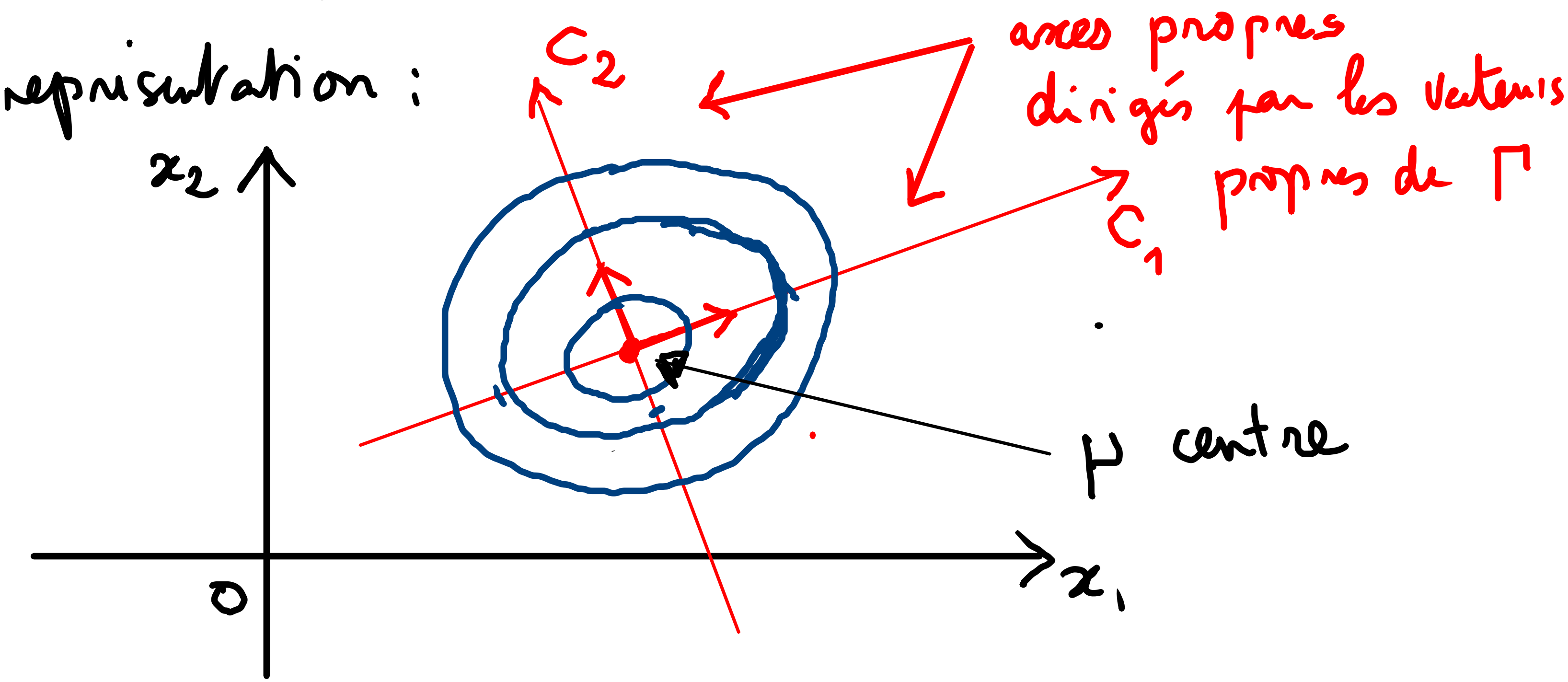


Dans ce nouveau repère, on a donc

$$\frac{c_1^2}{\lambda_1} + \frac{c_2^2}{\lambda_2} = c \qquad (c > 0 \text{ cte})$$

$\Longrightarrow$ ellipses de demi-grand axe $a = \sqrt{\lambda_1}\sqrt{c}$
et demi-petit axe $b = \sqrt{\lambda_2}\sqrt{c}$
si l'on convient de choisir au départ $\lambda_1 \geq \lambda_2$

D'où la représentation :



Théorème Central Limite multidimensionnel

Soit X_1, X_2, … une suite de vecteurs aléatoires **indépendants et identiquement distribués** dans $\mathbb{R}^d$ de vecteur moyenne μ et de matrice de covariance Γ supposée **inversible**. Pour n ≥ 1, on note encore $\mathbf{S_n = X_1 + \ldots + X_n} \in \mathbb{R}^d$.

Attention aux notations :

$$X_j = \begin{pmatrix} X_{1,j} \\ \vdots \\ X_{d,j} \end{pmatrix} \in \mathbb{R}^d$$

$j \geq 1$

Alors, pour tout hyper-rectangle [[a, b]] de $\mathbb{R}^d$

$$\mathbf{P}\left(\frac{\mathbf{S_n - n\mu}}{\sqrt{\mathbf{n}}} \in [[\mathbf{a, b}]) \to_{\mathbf{n}} \int_{[[\mathbf{a, b}]]} \mathbf{f_X}(x)\mathbf{d}x$$

où f est la densité de la loi gaussienne multivariée N(0, Γ).

En pratique : $$\frac{\mathbf{S_n}}{\mathbf{n}} \approx \mathbf{N}(\mu, \frac{\mathbf{1}}{\mathbf{n}}\Gamma)$$

Exemple : voir TP n°1

Explications TCL multidimensionnel :

$S_n = X_1 + \cdots + X_n$ est un vecteur aléatoire

- de moyenne $\boxed{ES_n = n\mu}$ où $\mu = EX_1 = \dots = EX_n$
- de covariance $\boxed{\mathrm{Cov}(S_n) = n\Gamma}$ si $\Gamma = \mathrm{Cov}(X_1) = \dots = \mathrm{Cov}(X_n)$

En effet : $\mathrm{Cov}(S_n) = E\left[(X_1 - \mu + \dots + X_n - \mu)(X_1 - \mu + \dots + X_n - \mu)^T\right]$

$$= \sum_{i,j} E\left[(X_i - \mu)(X_j - \mu)^T\right]$$

$$= \sum_{i=1}^{n} E\left[(X_i - \mu)(X_i - \mu)^T\right] + \sum_{i \neq j} 0$$

par indépendance

$$= n\Gamma$$

Le TCL dit que $S_n \approx \mathcal{N}(n\mu, n\Gamma)$

ou encore que $\frac{S_n}{n} \approx \mathcal{N}(\mu, \frac{1}{n}\Gamma)$

Illustrons avec l'exemple du TP1 (loi trinomiale)

$$N = \begin{pmatrix} N_1 \\ N_2 \end{pmatrix} \text{ où } \begin{cases} N_1 = \mathbb{1}_{X_1 = 1} + \dots + \mathbb{1}_{X_n = 1} \sim \mathcal{B}(n, p_1) \\ N_2 = \mathbb{1}_{X_1 = 2} + \dots + \mathbb{1}_{X_n = 2} \sim \mathcal{B}(n, p_2) \end{cases}$$

$N = \varepsilon_1 + \dots + \varepsilon_n$ où $\varepsilon_i = \begin{pmatrix} \mathbb{1}_{X_i = 1} \\ \mathbb{1}_{X_i = 2} \end{pmatrix}$ sont des vecteurs i.i.d. de moyenne $\mu = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$

et de covariance $\Gamma = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 \\ -p_1 p_2 & p_2(1-p_2) \end{pmatrix}$

car $\mathrm{Cov}(\mathbb{1}_{X_i=1}, \mathbb{1}_{X_i=2}) = E[\underbrace{\mathbb{1}_{X_i=1} \mathbb{1}_{X_i=2}}_{=0}] - p_1 p_2$

DONC : $\begin{pmatrix} N_1/n \\ N_2/n \end{pmatrix} \approx \mathcal{N}\left(\begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \begin{pmatrix} p_1(1-p_1)/n & -\frac{p_1 p_2}{n} \\ -\frac{p_1 p_2}{n} & \frac{p_2(1-p_2)}{n} \end{pmatrix} \right)$

Commentaire général sur la loi $\mathcal{N}(\mu, \Gamma)$:

Le TCL explique en grande partie toute l'importance de la loi $\mathcal{N}(\mu, \Gamma)$ en pratique.

Mais, son expression analytique est compliquée et ne permet pas de manipuler facilement des Vecteurs Gaussiens.

Besoin d'une définition alternative plus facile à appréhender :

$d = 1$

idée $\quad X \sim \mathcal{N}(\mu, \sigma^2) \iff X = \mu + \sigma \varepsilon$ avec $\varepsilon \sim \mathcal{N}(0,1)$

d'où X vecteur gaussien si $X = \mu + A\varepsilon$ (transformation linéaire)

avec $\varepsilon \sim \mathcal{N}(0, I_d)$

$\longrightarrow$ notion de bruit blanc gaussien

I.3. Bruit blanc gaussien

Définition. X est dit *gaussien* (ou *normal*) *centré-réduit* si **ses composantes $X_1, \ldots, X_d$ sont indépendantes et de loi N(0, 1)**. C'est un vecteur aléatoire *continu* admettant la *densité de probabilité*

$$\mathbf{x} = (x_1, \ldots, x_d)^T \rightarrow \frac{1}{\sqrt{2\pi}^{\,d}} \exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right)$$

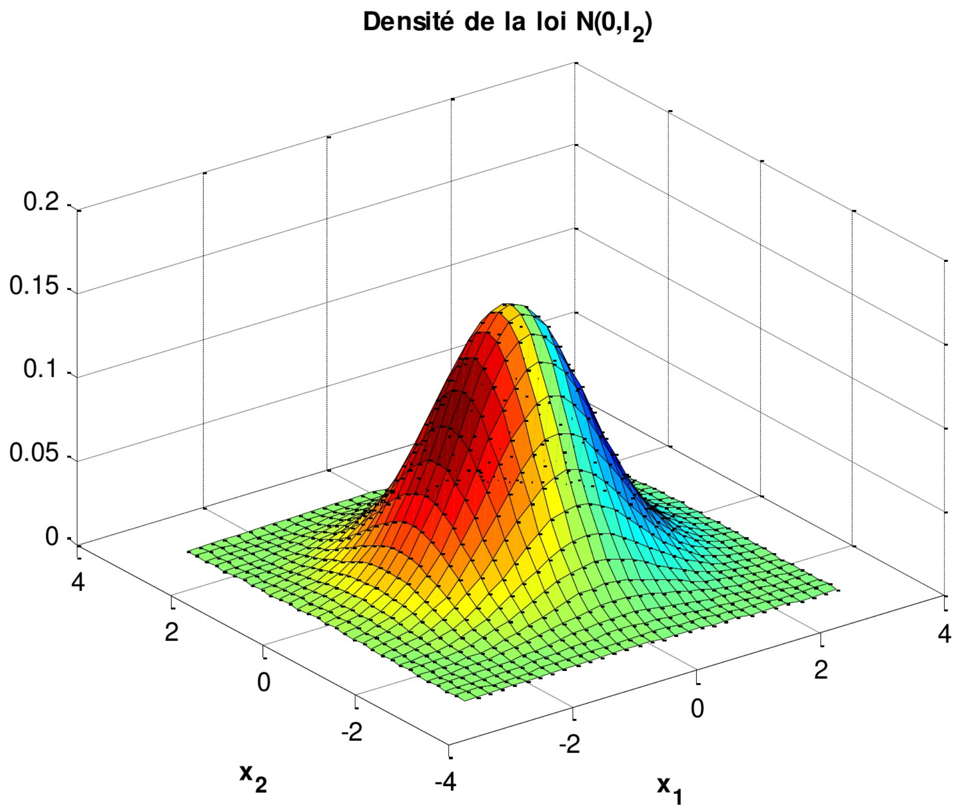
où $\|x\| = \sqrt{x_1^2 + \ldots + x_d^2}$ désigne la norme euclidienne usuelle de $\mathbb{R}^d$.

Interprétation. Les composantes du vecteur **X** forment un bruit « blanc », ce qui renvoie aussi à la terminologie de **bruit blanc gaussien** (standard).

On a **: $X \sim N(0, I_d)$** où I_d désigne la matrice *identité* de taille d, donc X est bien un vecteur gaussien au sens de la définition 1 (slide 11).

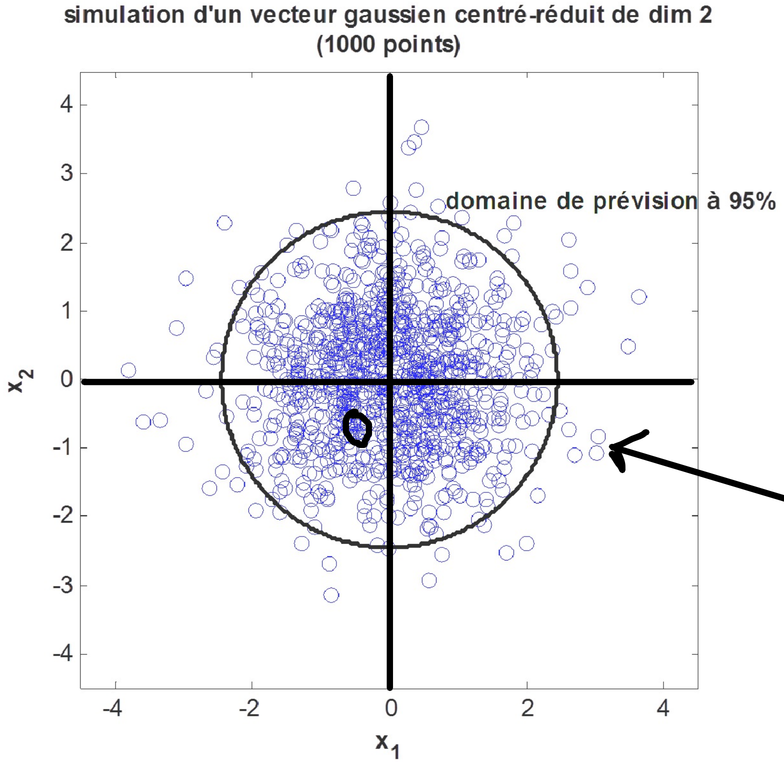
Dans toute la suite, ε désignera un vecteur aléatoire de loi **$N(0, I_d)$**.

Cas de la dimension d = 2 :



I.3. Bruit blanc gaussien

Simulation de $n = 1000$ points selon la loi $\mathcal{N}(0, I_2)$



$(d = 2)$

environ 5% des 1000 points sont en dehors

$$\varepsilon \sim N(0, I_d) \Rightarrow U\varepsilon \sim N(0, I_d) \text{ dès que } U^T U = I_d$$

U matrice orthogonale ⇔ U isométrie ⇔ U associée à une base orthonormée

Propriété fondamentale pour un V.G. standard :

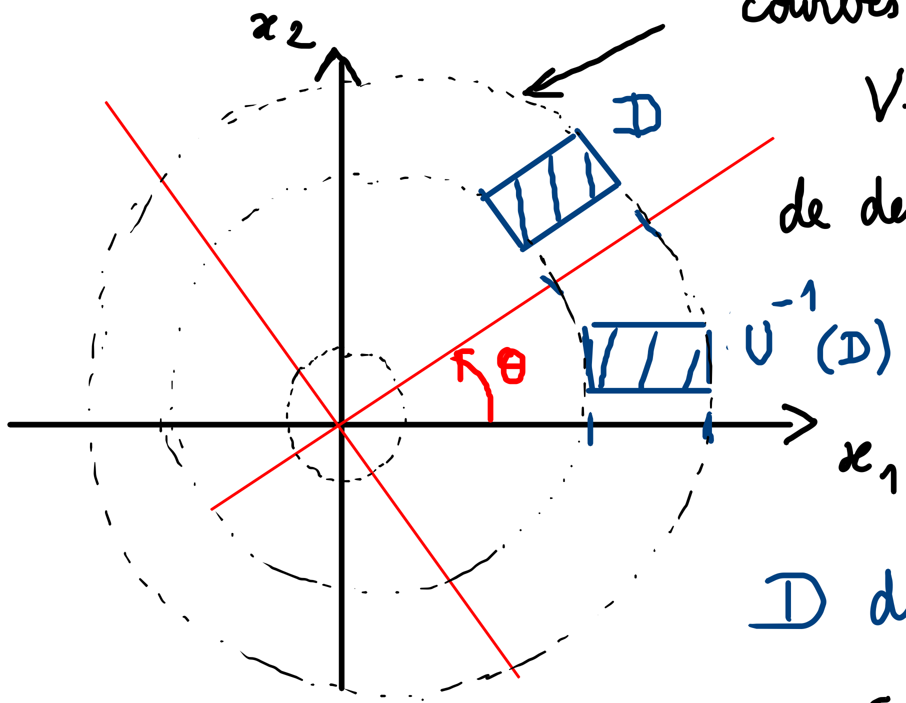
$\varepsilon \sim N(0, I_d) \Rightarrow U\varepsilon \sim N(0, I_d)$ dès que $U^TU = I_d$

Rappelons que U matrice orthogonale correspond à une <u>isométrie</u> puisque $\|Ux\|^2 = \langle Ux \mid Ux \rangle = \langle U^TUx \mid x \rangle = \|x\|^2$

Dans le cas $d=2$, U est la matrice d'une rotation (isométrie directe) ou une symétrie par rapport à une droite (isométrie indirecte).

Ainsi la densité $x \longrightarrow \frac{1}{\sqrt{2\pi}^d} e^{-\frac{\|x\|^2}{2}}$ d'un V.G. standard est invariante pour toute isométrie (le groupe des isométries).

La propriété en résulte, ce que l'on peut illustrer facilement dans le cas $d=2$:



courbes d'isoprobabilité de la v.a. $\varepsilon \sim N(0, I_2)$

de densité $f(x) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)}$

U rotation d'angle θ (par exemple)

D domaine quelconque

Alors $P(U\varepsilon \in D) = P(\varepsilon \in U^{-1}(D))$

OR $P(\varepsilon \in U^{-1}(D)) = P(\varepsilon \in D)$

car c'est le même volume sous la cloche de Gauss (invariante par rotation) qui s'appuie sur D ou $U^{-1}(D)$

(se représenter le volume en question).

DONC $\mathbb{P}(U\varepsilon \in D) = \mathbb{P}(\varepsilon \in D)$

ce qui signifie que $U\varepsilon$ a même loi que ε !

Formellement, on peut écrire dans le cas général :

$$\mathbb{P}(U\varepsilon \in D) = \mathbb{P}(\varepsilon \in U^{-1}(D))$$
$$= \int_{U^{-1}(D)} e^{-\frac{1}{2}\|x\|^2}\, dx$$

Changement de variables : $x = U^{-1} y = U^T y$

Alors $dx = dx_1 \dots dx_d = |\det(U^{-1})|\, dy_1 \dots dy_d$

soit $dx = dy$

car $|\det(U^{-1})| = 1$ (cf. U^{-1} isométrie).

De là, $\mathbb{P}(U\varepsilon \in D) = \int_D e^{-\frac{1}{2}\|U^{-1}y\|^2}\, dy$

$$= \int_D f(y)\, dy = \mathbb{P}(\varepsilon \in D) \quad //$$

Retenir donc que $U\mathcal{E} \sim \mathcal{N}(0, I_d)$ si $\mathcal{E} \sim \mathcal{N}(0, I_d)$

et U isométrie $(U^T U = I_d)$

En particulier, si l'on choisit une base orthonormée quelconque de $\mathbb{R}^d$ de la forme $(\underbrace{U_1, U_2, \dots, U_d}_{\text{vecteurs colonne de } \mathbb{R}^d})$

alors $\mathcal{E} = \sum_{j=1}^{d} \langle \mathcal{E} | U_j \rangle U_j$ où donc

coordonnées dans la nouvelle base $\rightarrow$ $\begin{pmatrix} \langle \varepsilon | U_1 \rangle \\ \vdots \\ \langle \varepsilon | U_d \rangle \end{pmatrix} = \begin{pmatrix} U_1^T \varepsilon \\ \vdots \\ U_d^T \varepsilon \end{pmatrix} = U^T \mathcal{E} \sim \mathcal{N}(0, I_d)$

$\uparrow$ isométrie réciproque de $U = (U_1 \dots U_d)_{d \times d}$

Donc, si ε est un VG standard ($\sim N(0, I_d)$),
alors ses coordonnées dans n'importe quelle base orthonormée de $\mathbb{R}^d$ forment encore un bruit blanc gaussien (des v.a. i.i.d. $N(0,1)$)

On en déduit facilement les 2 conséquences qui suivent :

Conséquence 1 - La loi normale est stable pour l'addition : $\forall n \geq 1$, on a :

$X_1, \ldots, X_n$ v.a.r. gaussiennes **indépendantes** $\Rightarrow X_1 + \ldots + X_n$ v.a.r. gaussienne

Conséquence 2 (THÉORÈME DE COCHRAN*)

Soit P un projecteur orthogonal, i.e. $P^2 = P$ et $P^T = P$. Alors :

- **$P\varepsilon$ et $(I - P)\varepsilon$** sont **indépendants**, **gaussiens centrés-réduits de degrés de liberté (ou de dimensions respectives) p et d – p où p = dim(ImP**).
- Les v.a. $\|P\varepsilon\|^2$ et $\|(I - P)\varepsilon\|^2$ sont **indépendantes** de lois respectives

$$\|P\varepsilon\|^2 \sim \chi_p^2 \text{ et } \|(I - P)\varepsilon\|^2 \sim \chi_{d-p}^2$$

(*William Gemmel Cochran, 1909-1980)

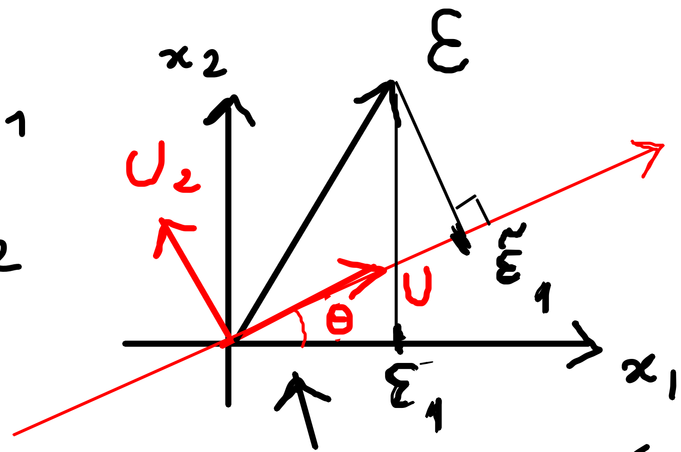
<u>Conséquence 1</u> : il suffit de traiter le cas $n=2$ (puis récurrence sur n).

Soient donc $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ indépendantes

Écrivons que : $X_1 = \mu_1 + \sigma_1 \varepsilon_1$, $X_2 = \mu_2 + \sigma_2 \varepsilon_2$ avec donc $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \mathcal{N}(0, I_2)$

Alors : $X_1 + X_2 = \mu_1 + \mu_2 + \sqrt{\sigma_1^2 + \sigma_2^2}\, \tilde{\varepsilon}_1$

où $\tilde{\varepsilon}_1 = \dfrac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}} \varepsilon_1 + \dfrac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} \varepsilon_2$



Or $U_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ et $U_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$ nouvelle <u>base orthonormée</u>

donc (en particulier), $\tilde{\varepsilon}_1 = \langle \varepsilon | U_1 \rangle \sim \mathcal{N}(0,1)$

Comparaison avec le calcul direct dans le cas simple $X_1 \sim N(0,1)$ et $X_2 \sim N(0,1)$ indépendantes :

$X_1 + X_2$ est de loi de densité (produit de convolution)

$$x \longrightarrow f(x) = \int_{-\infty}^{+\infty} f_{X_1}(x_1)\, f_{X_2}(x - x_1)\, dx_1$$

$$\left(= \int_{-\infty}^{+\infty} f_{X_1}(x - x_2)\, f_{X_2}(x_2)\, dx_2 \text{ par symétrie} \right)$$

Donc, $$f(x) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \times \frac{1}{\sqrt{2\pi}} e^{-(x-x_1)^2/2}\, dx_1$$

Or : $$x_1^2 + (x - x_1)^2 = 2x_1^2 - 2x x_1 + x^2$$
$$= 2\left(x_1 - \frac{x}{2}\right)^2 + \frac{x^2}{2}$$

D'où : $$f(x) = \left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\left(x_1 - \frac{x}{2}\right)^2} dx_1 \right) \times \frac{e^{-\frac{x^2}{4}}}{\sqrt{2\pi}}$$

Par un petit changement de variable ($x_1 \leftarrow x_1 - \frac{x}{2}$) :

$$f(x) = \underbrace{\left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x_1^2} \, dx_1\right)}_{= \frac{1}{\sqrt{2}}} \times \frac{e^{-\frac{1}{2}\frac{x^2}{2}}}{\sqrt{2\pi}}$$

$= \frac{1}{\sqrt{2}}$ (poser $x_1 = \frac{\tilde{x}_1}{\sqrt{2}}$)

donc $$f(x) = \frac{1}{\underbrace{\sqrt{2}}_{\sigma}\sqrt{2\pi}} \, e^{-\frac{1}{2}\frac{x^2}{\underbrace{2}_{\sigma^2}}}$$

c'st bien la loi $\mathcal{N}(0, 1+1) = \mathcal{N}(0, 2)$

On mesure sur cet exemple l'intérêt à raisonner de manière géométrique ...

Conséquence 2 (preuve)

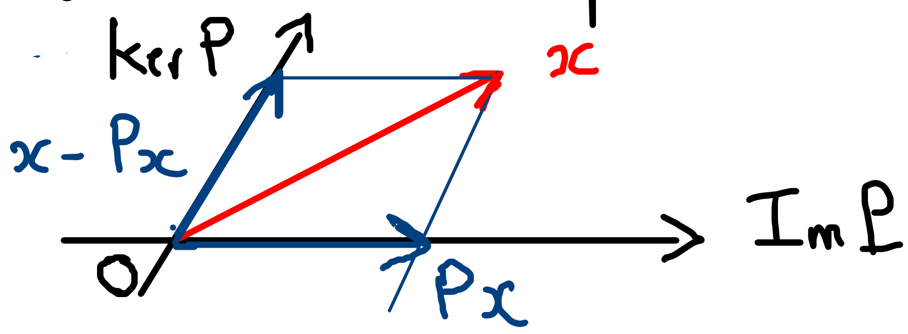
Rappel $P : \mathbb{R}^d \to \mathbb{R}^d$ linéaire (P identifié à sa matrice dans la base canonique : $P(x) = Px$) est un projecteur si $P^2 = P$ ($P \circ P = P$)

Alors $\mathbb{R}^d = \mathrm{Im}\,P \oplus \ker P$ puisque

$$x = \underbrace{Px}_{\in \mathrm{Im}\,P} + \underbrace{x - Px}_{\in \ker P}$$

(décomposition unique)

et (géométriquement) P correspond à la projection sur le sous-espace vectoriel $\mathrm{Im}\,P$ parallèlement à $\ker P$.



P est un projecteur orthogonal si, de plus, $\boxed{\operatorname{Im} P \perp \operatorname{Ker} P}$

Cela équivaut à $P = P^T$ (P symétrique)

En effet, si $\boxed{P = P^T}$, alors

$$\langle Pu | v \rangle = u^T P^T v = u^T P v = \langle u | Pv \rangle$$

et, si $v \in \operatorname{Ker} P$, on a bien $\langle Pu | v \rangle = 0$ pour tout u,

soit $\operatorname{Ker} P \perp \operatorname{Im} P$

Dans l'autre sens, si $\operatorname{Ker} P \perp \operatorname{Im} P$, alors

$$\langle Pu | v \rangle = \langle \underbrace{Pu}_{\in \operatorname{Im} P} | Pv + \underbrace{v - Pv}_{\in \operatorname{Ker} P} \rangle = \underbrace{\langle Pu | Pv \rangle}_{\text{expression symétrique en } u \text{ et } v}$$

donc $\langle Pu | v \rangle = \langle Pv | u \rangle = \langle u | Pv \rangle$

et $P = P^T$ $\left(u^T P^T v = u^T P v, \forall u, v \right)$

Avec ces rappels le théorème de Cochran devient clair.
En effet, prenons $(U_1, \dots, U_p)$ une base orthonormée de $\operatorname{Im} P$
$(U_{p+1}, \dots, U_d)$ une base de $\ker P$ (de dim $d-p$).
Alors $(U_1, \dots, U_p, U_{p+1}, \dots, U_d)$ base orthonormée de $\mathbb{R}^d$.
Puis,
$$\varepsilon = \underbrace{\sum_{i=1}^{p} \langle \varepsilon | U_i \rangle U_i}_{= P\varepsilon} + \underbrace{\sum_{i=p+1}^{d} \langle \varepsilon | U_i \rangle U_i}_{= \varepsilon - P\varepsilon}$$

Comme les coordonnées $(\langle \varepsilon | U_i \rangle)_i$ sont i.i.d. $\mathcal{N}(0,1)$, les vecteurs $P\varepsilon$ et $\varepsilon - P\varepsilon$ sont bien des V. gaussiens standard dans $\operatorname{Im} P \simeq \mathbb{R}^p$ et $\ker P \simeq \mathbb{R}^{d-p}$ respectivement.

De plus, $P\varepsilon$ et $(I-P)\varepsilon$ sont indépendants et

$$\|P\varepsilon\|^2 = \sum_{i=1}^{p} \langle \varepsilon | U_i \rangle^2 \sim \chi^2(p)$$

$$\|(I-P)\varepsilon\|^2 = \sum_{i=p+1}^{d} \langle \varepsilon | U_i \rangle^2 \sim \chi^2(d-p)$$

par définition de la loi du χ^2 à d degrés de liberté :

$$\chi^2(d) = \text{loi de } z_1^2 + \dots + z_d^2 = \|Z\|^2 \text{ où } Z \sim \mathcal{N}(0, I_d)$$

Remarquer enfin que $P\varepsilon$ est bien un vecteur à p degrés de liberté (il est dans $\operatorname{Im} P$), donc dégénéré si $p < d$.

De plus : $\operatorname{Cov}(P\varepsilon) = P I_d P^T = P^2 = P$ //

(qui est bien non inversible si $p < d$)

On donnera au final une définition d'un V.G. qui englobe le cas du théorème de Cochran où :

$$\begin{cases} P\,\varepsilon & \text{V.G.} \in \text{sous-espace de } \mathbb{R}^d \text{ de dim } p \\ (I-P)\,\varepsilon & \text{V.G.} \in \text{sous-espace de } \mathbb{R}^d \text{ de dim } d-p \end{cases}$$

Mais, on regarde d'abord le cas non dégénéré !

Définition d'un vecteur gaussien (seconde approche).

Un vecteur aléatoire $X = (X_1, \ldots, X_d)^T \in \mathbb{R}^d$ est dit *gaussien (ou multi-gaussien)* de **dimension réelle d** s'il existe $\mu = (\mu_1, \ldots, \mu_d) \in \mathbb{R}^d$, une matrice A de taille $d \times d$ **inversible** et $\varepsilon \sim N(0, I_d)$ tels que

$$X = \mu + A\varepsilon$$

Dit autrement, X est l'image par une **transformation affine** d'un bruit blanc gaussien centré-réduit.

Moyenne et matrice de covariance :

$E(X) = \mu$

$\mathrm{Cov}(X) = AA^T$

Définition d'un vecteur gaussien (seconde approche).

Un vecteur aléatoire $X = (X_1, \ldots, X_d)^T \in \mathbb{R}^d$ est dit *gaussien (ou multi-gaussien*) de **dimension réelle d** s'il existe $\mu = (\mu_1, \ldots, \mu_d) \in \mathbb{R}^d$, une matrice A de taille $d \times d$ **inversible** et $\varepsilon \sim N(0, I_d)$ tels que

$$X = \mu + A\varepsilon$$

Dit autrement, X est l'image par une **transformation affine** d'un bruit blanc gaussien centré-réduit.

Moyenne et matrice de covariance :

$E(X) = \mu$

$\mathrm{Cov}(X) = AA^T$

Densité et décomposition de Mahalanobis*.

Soit X un vecteur de la forme $X = \mu + A\varepsilon \in \mathbb{R}^d$ avec A inversible et $\varepsilon \sim N(0, I_d)$. On a $\Gamma = \mathrm{Cov}(X) = AA^T$ *inversible* et X admet la densité

$$x \to \frac{1}{\sqrt{\det(\Gamma)}\ \sqrt{2\pi}^{\,d}} \exp\left(-\frac{1}{2} <x - \mu, \Gamma^{-1}(x - \mu)> \right)$$

En bref, **$X \sim N(\mu, \Gamma)$ et X gaussien au sens de la première définition.**

On a encore la décomposition $\mathbf{X = \mu + U\Sigma\varepsilon}$ avec $\boldsymbol{\varepsilon \sim N(0, I_d)}$ et où les matrices Σ et U sont obtenues par diagonalisation en base orthonormée de la matrice de covariance : $\Gamma = U\Sigma^2U^T$ avec $\mathbf{U^TU = I_d}$.

(* Prasanta Chandra Mahalanobis, 1893-1972)

COROLLAIRE. ☺

Si $X \in \mathbb{R}^d$ est un vecteur gaussien de dimension réelle d, alors

$$\mathbf{X_1, \ldots, X_d \text{ indépendantes} \Leftrightarrow Cov(X_i, X_j) = 0 \text{ pour } i \neq j}$$
$$\mathbf{\Leftrightarrow \Gamma \text{ matrice diagonale}}$$

SIMULATION DE $X \sim N(\mu, \Gamma)$.

- Faire la décomposition de Cholesky de Γ: $\Gamma = SS^T$ avec S triangulaire inférieure (et diagonale positive)
- Simuler $\varepsilon \sim N(0, I_d)$
- Calculer $X = \mu + S\varepsilon$

Définition d'un vecteur gaussien (cas général).

Un vecteur aléatoire $X = (X_1, \ldots, X_p)^T \in \mathbb{R}^p$ est dit *gaussien (ou multi-gaussien)* s'il existe $\mu = (\mu_1, \ldots, \mu_p) \in \mathbb{R}^p$, une matrice A de taille $p\times q$ et $\varepsilon \sim N(0, I_q)$ tels que

$$X = \mu + A\varepsilon$$

X image par une **transformation affine** de $\mathbb{R}^q$ dans $\mathbb{R}^p$ d'un bruit blanc gaussien centré-réduit de dimension q.

Moyenne et matrice de covariance du vecteur X :

$E(X) = \mu \in \mathbb{R}^p$

$\Gamma = Cov(X) = AA^T$ de taille $p\times p$

Deux cas sont à envisager :

- $\Gamma = \mathrm{Cov}(X) = AA^T$ est **inversible**, alors X est de dimension réelle d = p et admet la densité

$$\boldsymbol{x \to \frac{1}{\sqrt{\det(\Gamma)}\ \sqrt{2\pi}^{\,p}} \exp(-\frac{1}{2} <x-\mu, \Gamma^{-1}(x-\mu)>)}$$

- Γ est **non inversible**, X est dit « dégénéré » et « vit » en réalité dans un sous-espace affine de dimension $d < p$ (et $d \leq q$). De manière plus précise, on a

$$\mathbf{X = \mu + \sqrt{\lambda_1}\, \varepsilon_1 U_1 + \ldots + \sqrt{\lambda_d}\, \varepsilon_d U_d\ ;\ \varepsilon \sim N(0, I_d)}$$

en écrivant que $\Gamma = U\Lambda U^T$ avec $U^TU = I_p$ et $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d > 0$ les valeurs propres non nulles de Γ.

Dans les deux cas, on dira que X est de loi N(μ, Γ).

Exemple (cf. TP1) : soit X une variable aléatoire à 3 modalités 0, 1 ou 2.

Notons $Z_0 = \frac{\mathbf{1}}{\sqrt{\mathbf{p_0}}}(1_{X=0} - p_0)$, $Z_1 = \frac{\mathbf{1}}{\sqrt{\mathbf{p_1}}}(1_{X=1} - p_1)$, $Z_2 = \frac{\mathbf{1}}{\sqrt{\mathbf{p_2}}}(1_{X=2} - p_2)$

puis Z le vecteur dans $\mathbb{R}^3$ de composantes Z_0, Z_1 et Z_2.

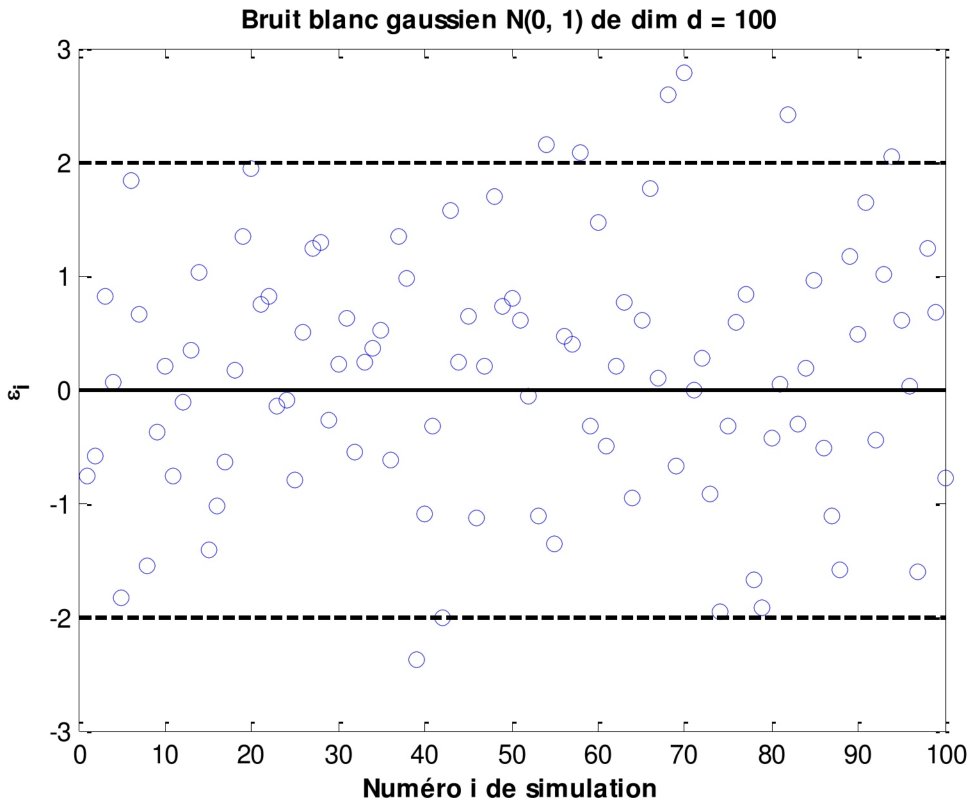
Par construction, $\mu = E(Z) = 0$.

Alors, $\Gamma = \mathrm{Cov}(Z)$ est une matrice **non inversible** et le théorème central limite affirme que

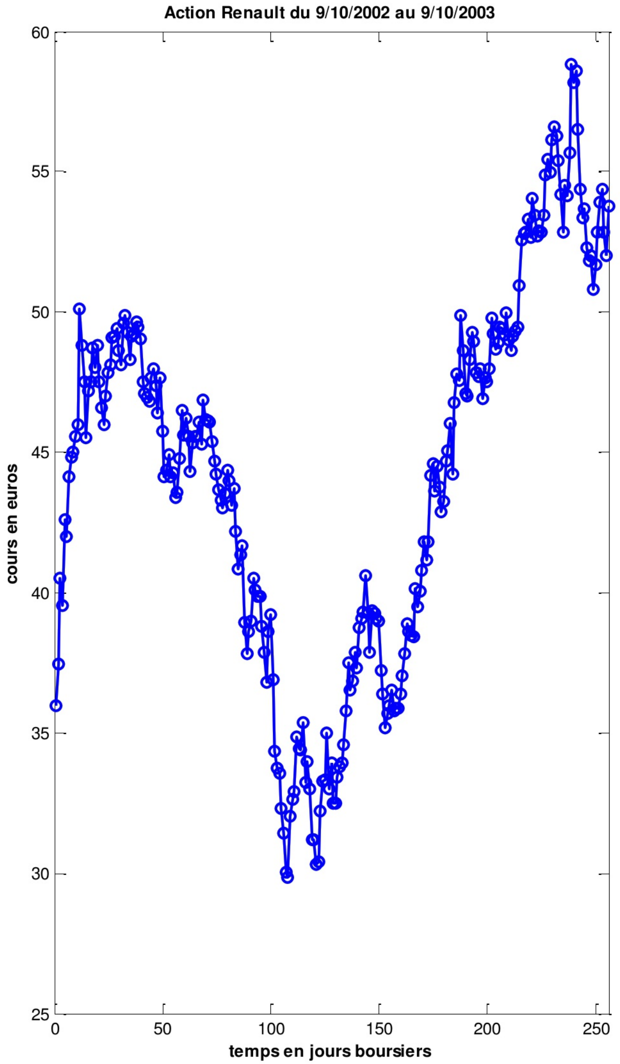
$$\frac{\mathbf{1}}{\sqrt{\mathbf{n}}}(\mathbf{Z}^{(1)} + \ldots + \mathbf{Z}^{(n)}) \rightarrow_{\mathbf{n}} \mathbf{N(0, \Gamma)}$$

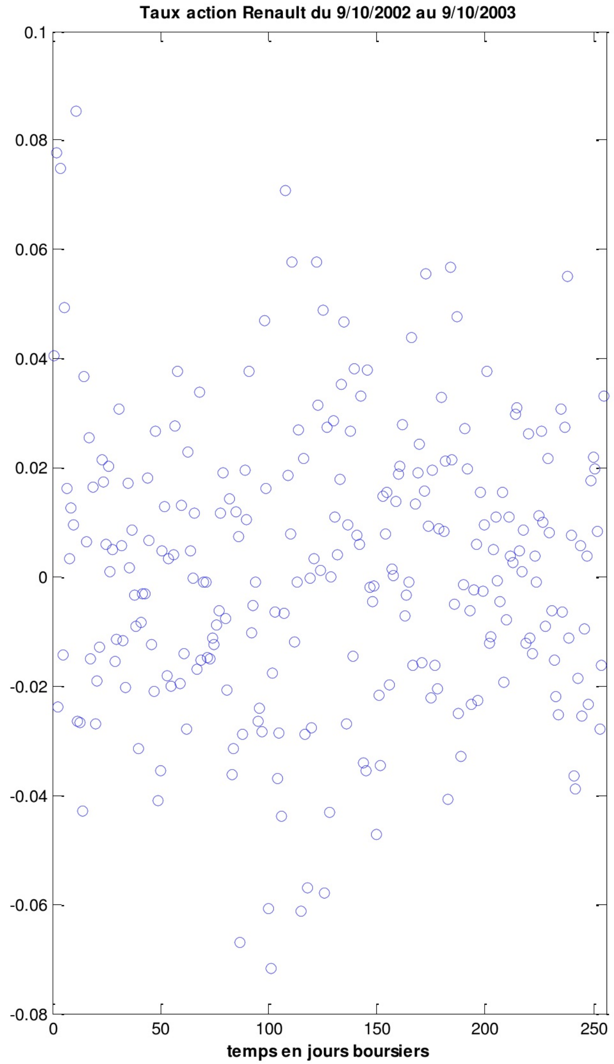
si les $Z^{(k)}$ $(k \geq 1)$ sont des vecteurs aléatoires indépendants de même loi que $Z \in \mathbb{R}^3$.

Exercice : analyser la loi en question et en déduire le test classique d'adéquation à une loi discrète ayant 3 modalités. Cas général.



Une simulation d'un bruit blanc gaussien de taille 100 visualisé par un tracé séquentiel de ses composantes


Action Renault du 9/10/2002 au 9/10/2003
cours en euros
temps en jours boursiers
60
55
50
45
40
35
30
25
0
50
100
150
200
250


Taux action Renault du 9/10/2002 au 9/10/2003
temps en jours boursiers
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
0
50
100
150
200
250

Décomposition en valeurs singulières (DVS ou SVD):

A étant donnée de taille p×q, il existe une matrice p×q Σ diagonale à termes diagonaux ≥ 0 telle que

$$\mathbf{A = U\Sigma V^T} \quad \text{avec} \qquad \mathbf{V^T V = I_q} \text{ et } \mathbf{U^T U = I_p}$$

Ainsi, $\quad \mathbf{\Gamma = AA^T = U\Lambda_p U^T} \qquad$ **avec** $\mathbf{\Lambda_p} = \Sigma\Sigma^T$ **diagonale** p×p

$$\mathbf{A^T A = V\Lambda_q V^T} \qquad \textbf{avec } \mathbf{\Lambda_q} = \Sigma^T\Sigma \textbf{ diagonale } q\times q$$

Les termes diagonaux de Σ sont appelés ***valeurs singulières*** de la matrice A et sont le plus souvent rangées par ordre décroissant.

II.1. Lois conditionnelles dans le cas discret

Soient Y v.a.r. et X v.a. **discrète** $X \in \{ x_1, x_2, \ldots, x_p\}$.

On appelle *loi conditionnelle de Y sachant* $X = x_k$ la loi de probabilité notée $\mu_{Y|X=x_k}$ définie par

$$\mu_{Y|X=x_k}([a, b]) = P(Y \in [a, b] \mid X = x_k) = \frac{P(Y \in [a, b] \text{ et } X = x_k)}{P(X = x_k)}$$

L'ensemble des p lois de probabilité $\mu_{Y|X=x_k}$ $(1 \leq k \leq p)$ est appelé *système des lois conditionnelles de Y sachant X.*

Exemple 1. Loi du pile ou face deux fois : Y = nombre total de pile, X résultat du premier lancer

II.1. Lois conditionnelles dans le cas discret

Cas où Y est également discrète : $Y \in \{ y_1, y_2, \ldots, y_q\}$

les p lois conditionnelles de Y sachant X s'écrivent

$$\mu_{Y|X=x_i} = \sum_{j=1}^{q} p_{j|i}\, \delta_{y_j} \text{ où } p_{j|i} = \frac{p_{ij}}{p_{i.}} = \frac{P(Y = y_j \text{ et } X = x_i)}{P(X = x_i)}$$

On a (!)

X, Y indépendantes $\Leftrightarrow \mu_{Y|X=x_i} = \mu_Y$ **pour tout i**

$$\Leftrightarrow p_{j|i} = \frac{P(Y = y_j \text{ et } X = x_i)}{P(X = x_i)} = P(Y = y_j) = p_{\bullet j}$$

II.2. Lois conditionnelles dans le cas continu

$$\boxed{(X, Y) \in \mathbb{R}^2 \text{ est un vecteur continu de densité } (x, y) \to f_{(X, Y)}(x, y)}$$

La loi conditionnelle de Y sachant X = x est définie par la *densité conditionnelle* suivante

$$\boxed{f_{Y \mid X = x}(y) = \frac{f_{(X, Y)}(x, y)}{f_X(x)}}$$

pour tout x tel que $f_X(x) = \int_{-\infty}^{+\infty} f_{(X, Y)}(x, y)dy \neq 0.$

Exemple 2. $Y = X + \varepsilon$ où $X \sim N(\mu_X, \sigma_X^2)$ et $\varepsilon \sim N(0, \sigma^2)$ sont indépendantes

II.2. Lois conditionnelles dans le cas continu

On a (!)

$$\textbf{X, Y indépendantes} \Leftrightarrow \mu_{Y|X=x} = \mu_Y \Leftrightarrow f_{Y\,|\,X=x}(y) = \frac{f_{(X,\,Y)}(x, y)}{f_X(x)} = f_Y(y)$$

FORMULE DE BAYES.

$$f_{X\,|\,Y=y}(x) = \frac{f_{Y\,|\,X=x}(y)\times f_X(x)}{f_Y(y)}$$

sachant que $f_Y(y) = \int_{-\infty}^{+\infty} f_{Y\,|\,X=x}(y)\times f_X(x)dx$ **est simplement une constante de normalisation.**

II.3. Espérance conditionnelle

On a vu que l'information apportée par $X = (X_1, \ldots, X_p)$ modifie les probabilités que l'on a sur Y avec la notion de loi conditionnelle.

Il est tout autant naturel de se demander en quoi la connaissance des variables $X_1, \ldots, X_p$ réduit ou non l'incertitude que l'on a sur Y si l'on mesure cette incertitude en terme de dispersion usuelle, i.e. de variance.

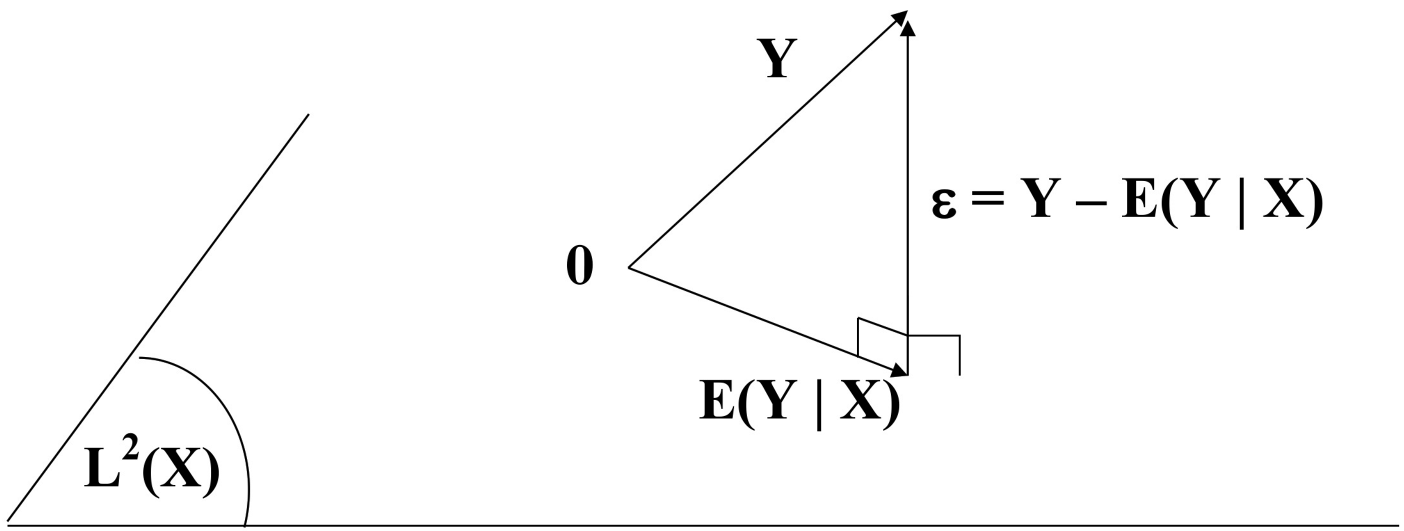
☺☺☺ On introduit pour cela l'espace vectoriel $L^2(P)$ des v.a. admettant une variance et muni du produit scalaire suivant :

$$\langle U, V \rangle = E(UV)$$

Définition de l'*espérance conditionnelle* E(Y | X) :

E(Y | X) est **la projection orthogonale de Y sur le sous-espace L^2(X)** de L^2(P) défini par

$$\mathbf{L^2(X) = \{\ Z = \varphi(X) \in L^2(P)\ \ \text{avec}\ \varphi : \mathbb{R}^p \rightarrow \mathbb{R}\}.}$$



Bien noter que E(Y | X) est une variable aléatoire!

Formules de l'espérance totale et de la variance totale

$$\mathbf{Y = E(Y \mid X) + \varepsilon} \qquad \text{avec} \quad \mathbf{E(\varepsilon) = 0} \text{ et :}$$

$$\mathbf{Cov(\varepsilon, \varphi(X)) = < \varepsilon, \varphi(X) > = 0 \text{ si } \varphi(X) \in L^2(X)}$$

En particulier : $\mathbf{E(E(Y \mid X)) = E(Y)}$ **et** $\mathbf{Var(Y) = Var(E(Y \mid X)) + Var(\varepsilon)}$
Dans cette décomposition, **ε est de variance minimale** : on ne peut pas réduire plus la variance résiduelle!

Définition. On définit le pourcentage de la variance de Y expliquée par X selon

$$\textbf{\% de variance expliquée} = 100 \times \frac{\mathbf{Var(E(Y \mid X))}}{\mathbf{Var(Y)}} \ \%$$

$$\textbf{\% de variance non expliquée ou résiduelle} = 100 \times \frac{\mathbf{Var(\varepsilon)}}{\mathbf{Var(Y)}} \ \%$$

II.3. Espérance conditionnelle

Espérance conditionnelle comme moyenne de la loi conditionnelle. On a simplement

$$\mathbf{E(Y \mid X) = m(X)}$$

où la fonction m vérifie

$$\mathbf{m(x) = \int_{-\infty}^{+\infty} y \times f_{Y \mid X = x}(y)dy = E(Y \mid X = x)}$$

i.e. m (x) est la moyenne de la loi conditionnelle de Y sachant X = x.

De plus, on a **la formule de transfert conditionnel**

$$\mathbf{E(\, f(Y) \mid X) = \varphi(X) \text{ où } \varphi \text{ se calcule selon } \varphi(x) = \int f(y)d\mu_{Y \mid X = x}(y)}$$

II.3. Espérance conditionnelle

On termine par le cas gaussien.

Si $(X, Y) = (X_1, \ldots, X_p, Y)$ est un <u>vecteur gaussien</u>, alors :

- **$E(Y \mid X) = E_L(Y \mid X)$** est une **fonction linéaire (affine)** des X_k
- Le résidu ε est **indépendant** des X_k et de loi **$N(0, Var(\varepsilon))$ où**

$$\mathbf{Var(\varepsilon) = Var(Y) - Var(E(Y \mid X))}$$

- **La loi conditionnelle de Y sachant X = x est gaussienne de variance indépendante de x et égale à la variance du résidu ε**