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 :

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

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. $X \sim N(\mu,0)$.

高斯退化

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

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

中心极限定理是概率论中最著名的结果之一。
它提出，大量的独立随机变量之和具有近似于正态的分布。

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

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

Si les X_k sont $N(\mu, \sigma^2)$, alors $\frac{S_n}{n}$ est de loi $N\left(\mu, \frac{\sigma^2}{n}\right)$ 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, $\mathbf{S_n = X_1 + \ldots X_n}$ **de loi binomiale B(n, p)** et

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

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

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**

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

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

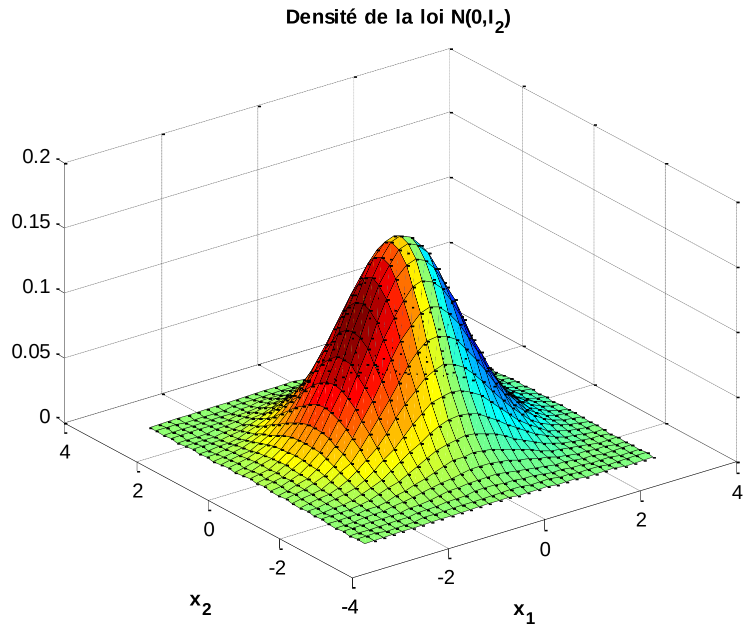
- K = 2 : loi binomiale
- K = 3 : loi trinomiale

☞ elle sera étudiée en TP 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. $(N_0, N_1, \ldots, N_{K-1})$?

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 :

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

☞ cloche de Gauss bi-dimensionnelle

Définition (densité d'un V.a. continu). $X = (X_1, \ldots, X_d)^T \in \mathbb{R}^d$ V.a.r. est de densité $x = (x_1, \ldots, x_d) \in \mathbb{R}^d \to f_X(x) \in \mathbb{R}$ si

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

où [a, b] est 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 $X_1, \ldots, X_d$ et admettent dans le cas continu 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 $\Leftrightarrow f_X(x) = f_{X_1}(x_1)\times\ldots\times f_{X_d}(x_d)$

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

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

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 $Y = AX + b$.

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

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

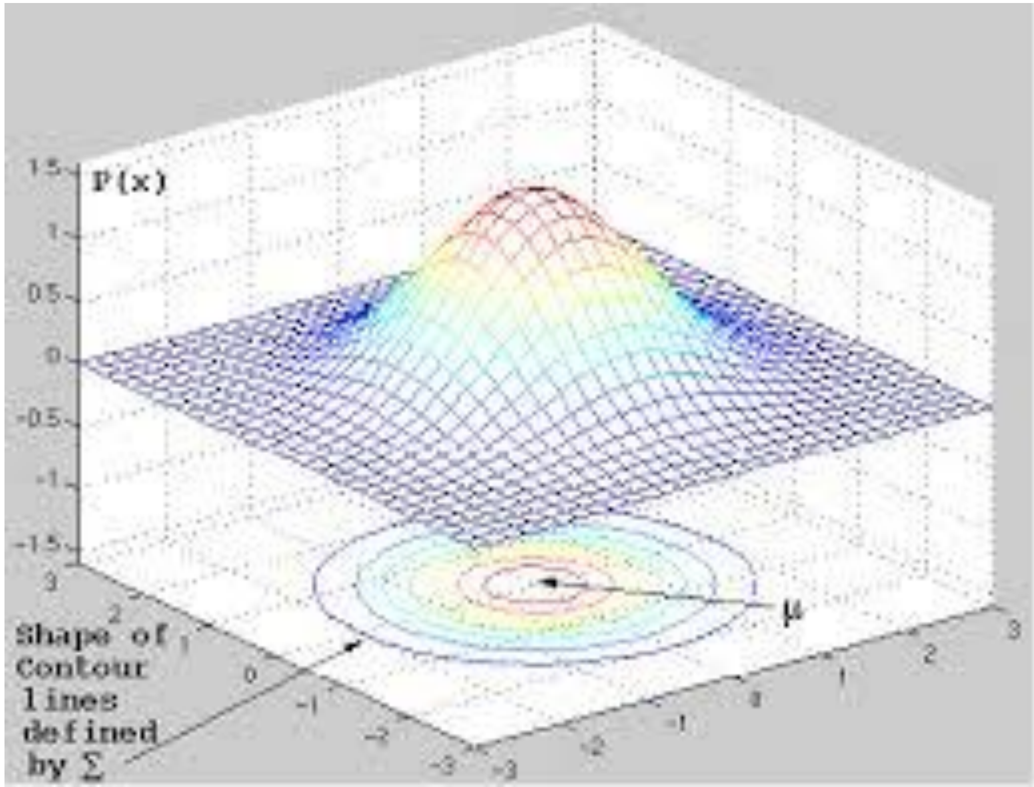
$$\boldsymbol{x} \to \frac{1}{\sqrt{\det(\Gamma)}\ \sqrt{2\pi}^{\,d}} \exp\left(-\frac{1}{2} <\boldsymbol{x}-\mu, \Gamma^{-1}(\boldsymbol{x}-\mu)> \right)$$

Dans ce cas, $EX = \mu$.

On note encore $\mathbf{N(\mu, \Gamma)}$ la loi normale multivariée de moyenne μ et de covariance Γ.

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

I.2. Vecteurs aléatoires réels



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 $\mathbf{S_n = X_1 + \ldots X_n} \in \mathbb{R}^d$.

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

$$\mathbf{P(\; a \leq \frac{S_n - n\mu}{\sqrt{n}} \in [a, b]\;) \to_n \int_{[a, b]} f(x)dx}$$

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

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

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* de *loi* admettant la *densité de probabilité*

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

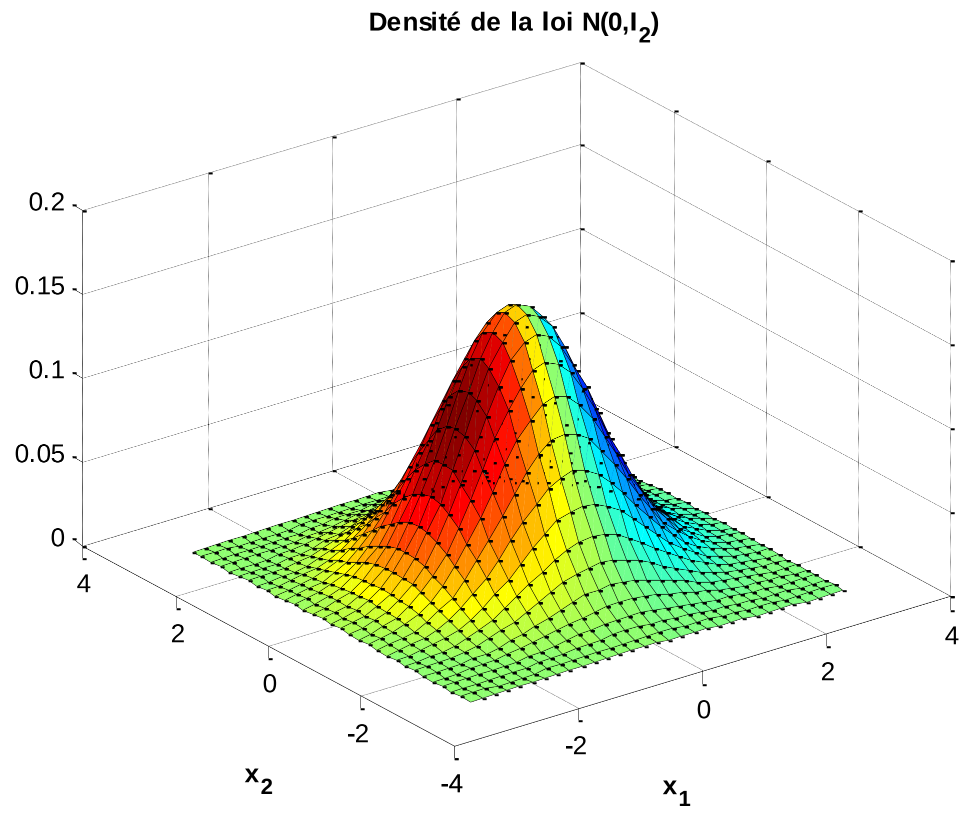
où $\|x\| = \sqrt{x_1^2 + \ldots + x_d^2}$ désigne la norme euclidienne usuelle de $\mathbb{R}^d$.

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

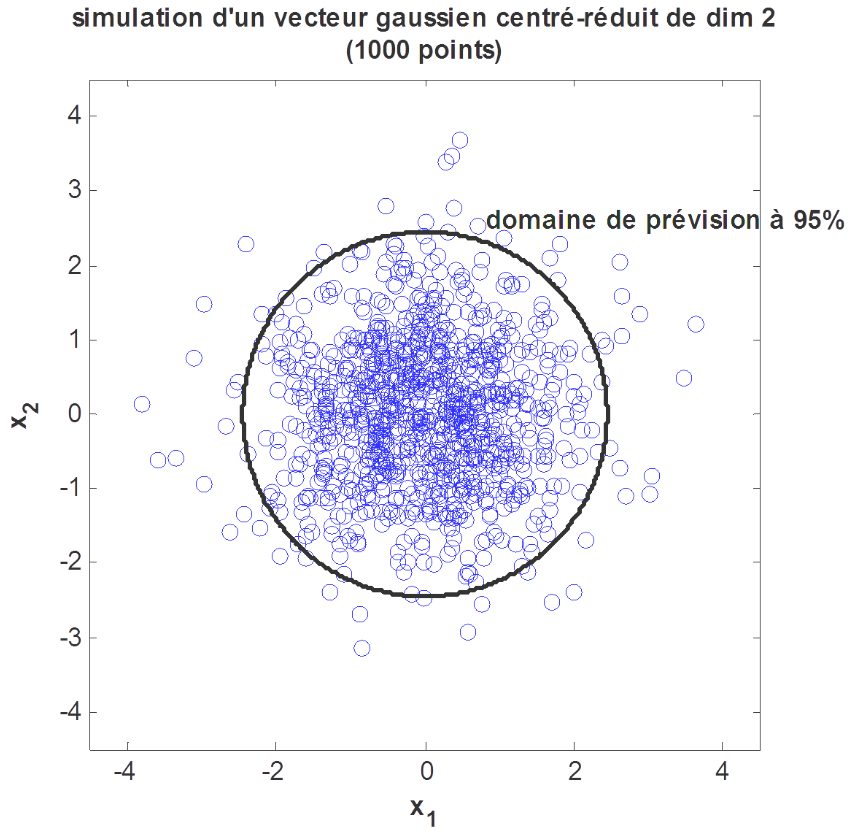
On a **: $X \sim N(0, I_d)$** où $\mathbf{I_d}$ désigne la matrice *identité* de taille d, donc X est bien un vecteur gaussien au sens de la définition 1 (slide 11).

$$\boxed{\text{Dans toute la suite, } \varepsilon \text{ désignera un vecteur aléatoire de loi } \mathbf{N(0, I_d)}.}$$

Cas de la dimension d = 2 :



I.3. Bruit blanc gaussien



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

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

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)

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

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

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 :

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

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

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

- 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 $N(\mu, \Gamma)$.

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.

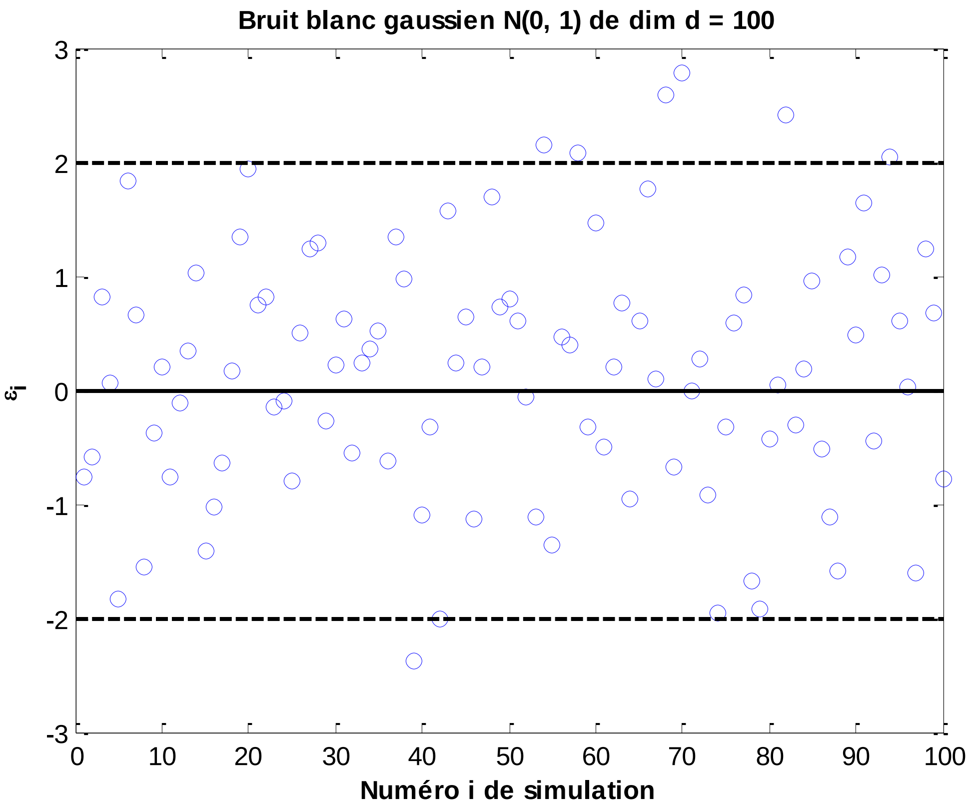
Par construction, $\mu = E(Z) = 0$.

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

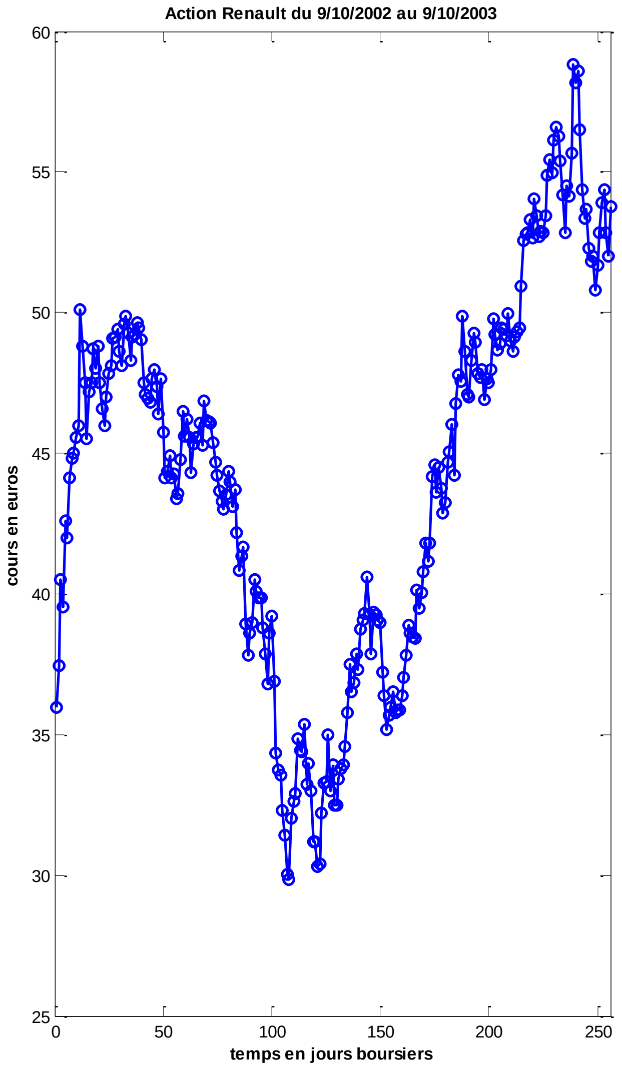
$$\frac{\mathbf{1}}{\sqrt{\mathbf{n}}}\left(\mathbf{Z^{(1)} + \ldots + Z^{(n)}}\right) \to_{\mathbf{n}} \mathbf{N(0, \Gamma)}$$

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

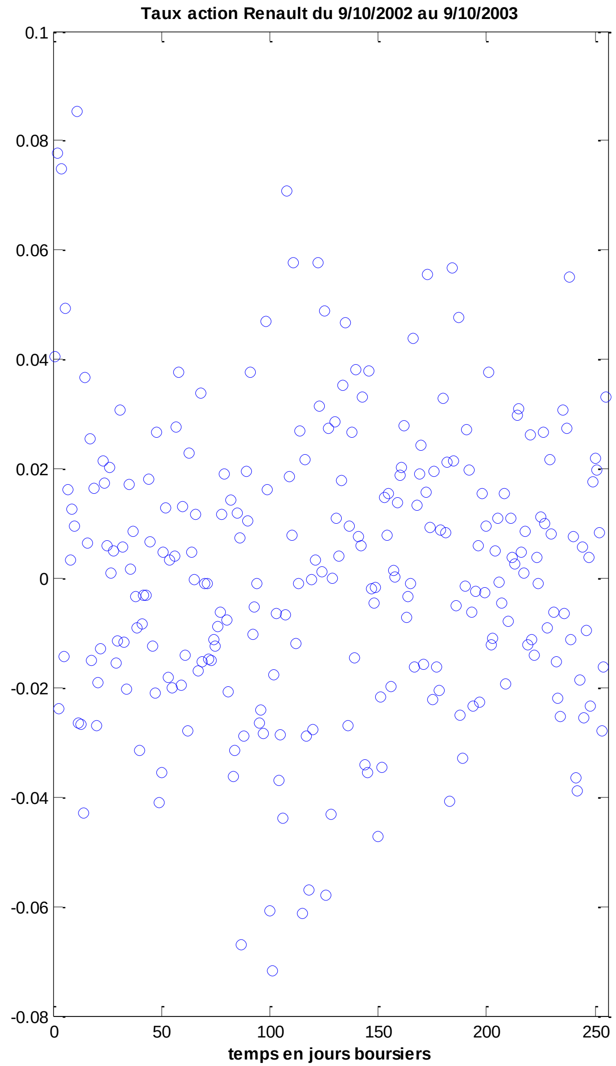
Exercice : analyser la loi en question et en déduire le test classique d'adéquation à une loi discrète ayant 3 modalités. Cas général.



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



Taux action Renault du 9/10/2002 au 9/10/2003
temps en jours boursiers

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** $\mathbf{\Lambda_p} = \Sigma\Sigma^\mathbf{T}$ **diagonale** p×p

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

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

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 $\boldsymbol{\mu_{Y|X=x_k}}$ définie par

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

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

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 (!)

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

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

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.

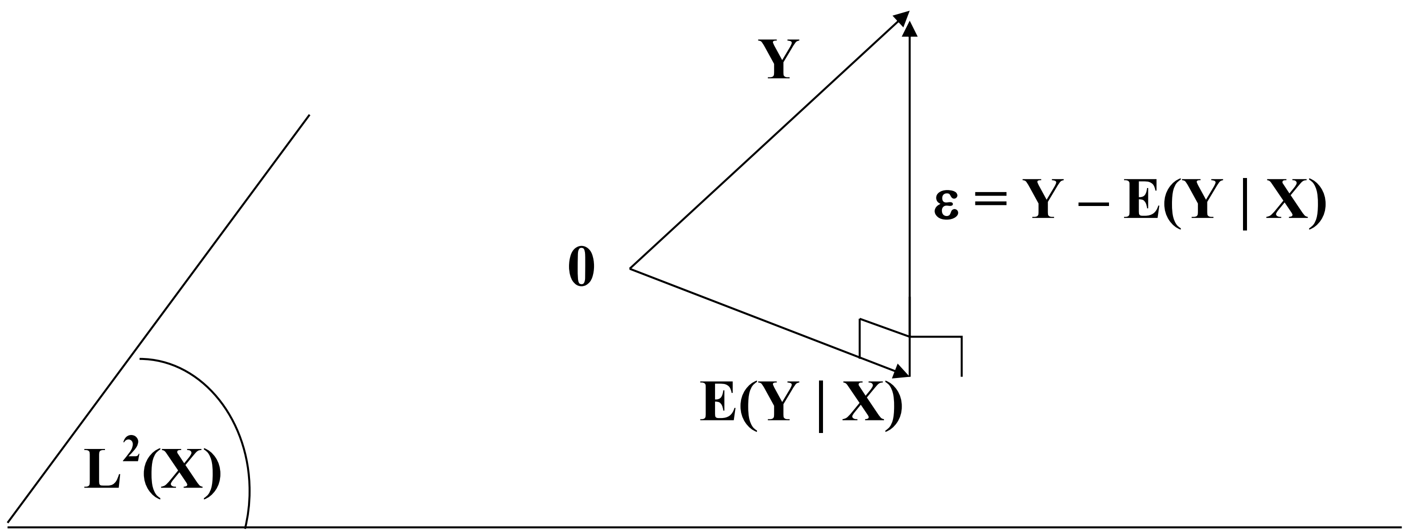
☺☺☺ On introduit pour cela l'espace vectoriel $L^2(P)$ des v.a. admettant une variance et muni du produit scalaire suivant :

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

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

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



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**

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

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**