Notes manuscrites sur le

cours du mardi 18 / sept. / 2020

(slides 18 à 23)

Fin de la partie du cours sur

les Vecteurs Gaussiens

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)* **non dégénéré** (ou 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 :

$E(X) = \mu$

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

Explications (rappels)

→ définition alternative en liaison avec

$$X = \mu + \sigma \varepsilon \quad \text{où } \varepsilon = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

dans le cas $d = 1$

$$\longrightarrow \quad X = \mu + A\varepsilon \iff \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} + \underbrace{A}_{d \times d} \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_d \end{pmatrix}$$

$$\Longrightarrow \quad E(X) = \mu + A E(\varepsilon) = \mu$$

$$\operatorname{Cov}(X) = E\left[(X-\mu)(X-\mu)^T\right] = E\left[A\varepsilon\varepsilon^T A^T\right]$$

$$= A \underbrace{E(\varepsilon\varepsilon^T)}_{I_d} A^T = AA^T$$

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 $\mathbf{X = \mu +} \mathrm{U\Sigma Z}$ avec $\mathbf{Z \sim N(0, I_d)}$ 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**)

Explications $X = \mu + A\varepsilon$ où A inversible, $\varepsilon \sim \mathcal{N}(0, I_d)$

$\Gamma = \mathrm{Cov}(X) = AA^T$ est symétrique définie positive,

donc se diagonalise en base orthonormée :

$\Gamma = U \Lambda U^T$ où $\begin{cases} U^T U = I_d & (U \text{ matrice des V.p en colonne}) \\ \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} & \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0 \ (\text{spectre}) \end{cases}$

On écrit : $\Lambda = \Sigma^2$ où $\Sigma = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_d} \end{pmatrix}$

Alors $\boxed{A = U \Sigma V^T \text{ où : } V^T V = I_d}$

En effet, posons $V = A^T U \Sigma^{-1}$.

Alors $V^T = \Sigma^{-1} U^T A \iff A = U \Sigma V^T$

Puis, $V^T V = \Sigma^{-1} U^T \underbrace{A A^T}_{\Gamma = U \Lambda U^T} U \Sigma^{-1} = \Sigma^{-1} \cancel{U^T U} \Sigma^2 \cancel{U^T U} \Sigma^{-1}$

$= I_d$ //

Ainsi, $X = \mu + A\varepsilon = \mu + U\Sigma \underbrace{(V^T \varepsilon)}_{= Z}$

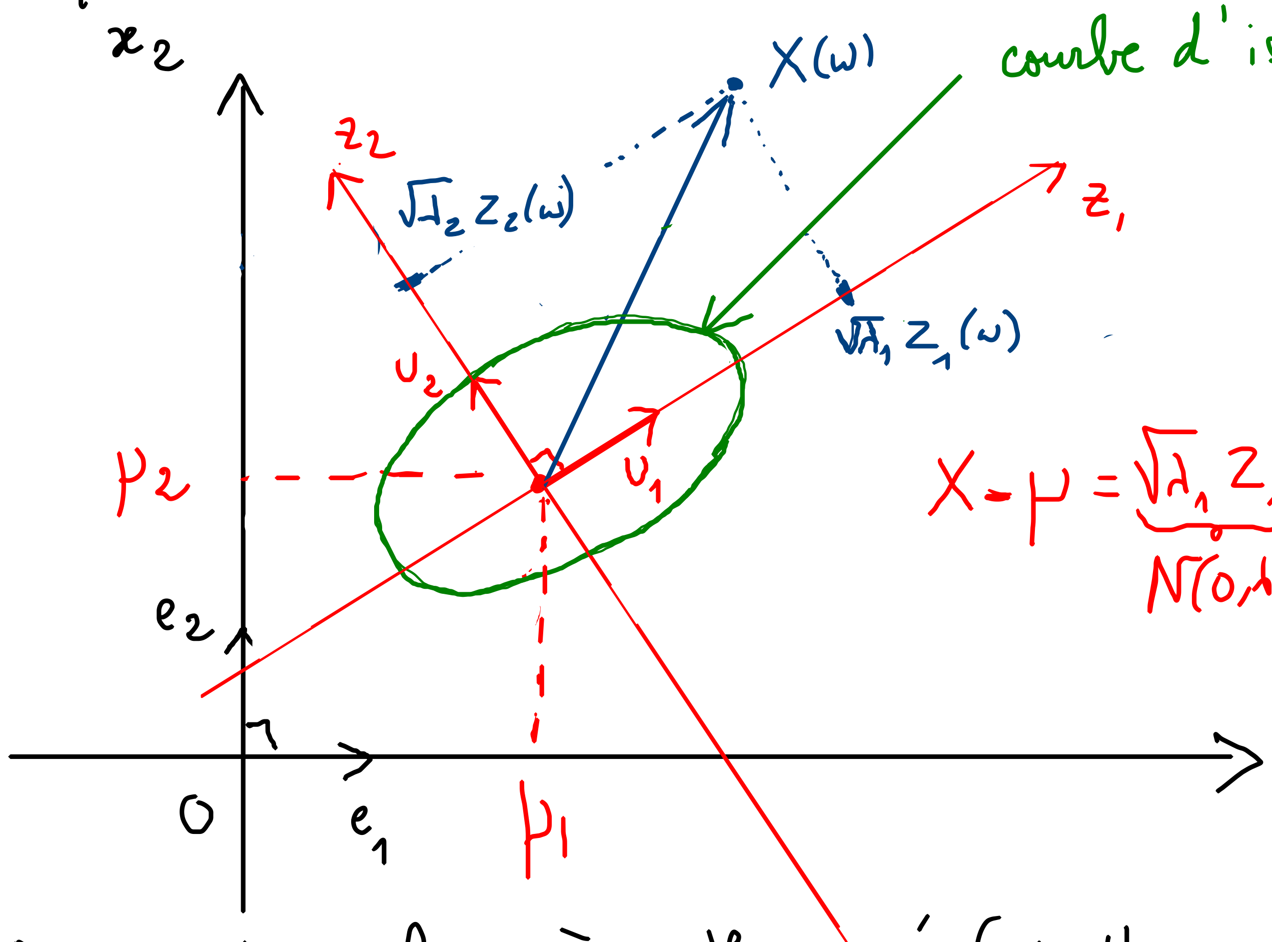
Or $Z \sim \mathcal{N}(0, I_d)$ car $\begin{cases} V^T = V^{-1} \text{ isométrie} \\ \varepsilon \sim \mathcal{N}(0, I_d) \end{cases}$

RQUE : $\varepsilon = \sum_{k=1}^{d} z_k V_k$ donc $(z_1, \dots, z_d)$ sont les coordonnées euclidiennes de ε dans la base orthonormée $(V_1, \dots, V_k)$, donc $Z_1, \dots, Z_d$ i.i.d. $\mathcal{N}(0,1)$.

D'où la décomposition dite de Mahalanobis :

$$\boxed{X(\omega) = \mu + \sum_{k=1}^{d} \underbrace{\sqrt{\lambda_k}\, Z_k(\omega)}_{\text{v.a. indépendantes de lois } \mathcal{N}(0,\lambda_k) \text{ resp}^t} U_k}$$

Le petit dessin en dim $d=2$ pour illustrer :



$$X-\mu = \underbrace{\sqrt{\lambda_1}\, Z_1}_{\mathcal{N}(0,\lambda_1)}\, U_1 + \underbrace{\sqrt{\lambda_2}\, Z_2}_{\mathcal{N}(0,\lambda_2)}\, U_2$$

indépendantes !

Ainsi, dans le repère orthonormé $(\mu; u_1, \dots, u_d)$, la densité du V.a. X s'écrit :

$$z \longrightarrow \frac{1}{\sqrt{\lambda_1}\cdots\sqrt{\lambda_d}}\, \frac{1}{\sqrt{2\pi}^{\,d}}\, e^{-\frac{1}{2}\sum_{i=1}^{d} \frac{z_i^2}{\lambda_i}}$$

(produit des densités marginales par indépendance)

Sous forme matricielle, cette densité s'écrit :

$$z \longrightarrow \frac{1}{\sqrt{\det(\Gamma)}\,\sqrt{2\pi}^{d}}\, e^{-\frac{1}{2} z^{\top} \Lambda^{-1} z}$$

En revenant au repère initial de $\mathbb{R}^2$: $(0\,;\, e_1, e_2)$,

on obtient vu que $z = U^{\top}(x-\mu)$:

$$f_X(x) = \frac{1}{\sqrt{\det(\Gamma)}\,\sqrt{2\pi}^{d}}\, e^{-\frac{1}{2}(x-\mu)^{\top} \underbrace{U \Lambda^{-1} U^{\top}}_{\Gamma^{-1}} (x-\mu)}$$

car : $\Gamma = U \Lambda U^{\top}$

donc $X = \mu + A\varepsilon$ a bien la densité de probabilité attendue en liaison avec la définition usuelle de la loi $\mathcal{N}(\mu, \Gamma)$ à partir de sa densité !

Conséquences : $X = \mu + A\mathcal{E} \implies$

(1) $x \longrightarrow \dfrac{1}{\sqrt{\det(\Gamma)}\,\sqrt{2\pi}^{\,d}}\, e^{-\frac{1}{2}(x-\mu)^T \Gamma^{-1}(x-\mu)}$ est bien une densité de probabilité (où $\Gamma = AA^T$) entièrement caractérisée donc par μ = moyenne et $\Gamma = \mathrm{Cov}(X)$ matrice des variances-covariances du vecteur $(X_1, \dots, X_d)$

(2) $X_1, \dots, X_d$ sont elles-mêmes gaussiennes car (par exemple)

$X_1 = \mu_1 + A_{11}\,\varepsilon_1 + \dots + A_{1d}\,\varepsilon_d$ est gaussienne

par stabilité de la loi normale ! (cf. $\varepsilon_1, \dots, \varepsilon_d$ indépendantes)

donc $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ où $\sigma_i^2 = \Gamma_{ii}$

(3) $\begin{cases} B \text{ inversible} \\ b \in \mathbb{R}^d \end{cases} \implies Y = b + BX = b + B\mu + BA\mathcal{E}$

est encore un V.G. de loi $\mathcal{N}(b + B\mu ; BAA^TB^T)$

COROLLAIRE. ☺

Si $X \in \mathbb{R}^d$ est un vecteur gaussien non dégénéré, 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 $X = \mu + S\varepsilon$

Sur le corollaire : $X_1, \dots, X_d$ indépendantes $\Longrightarrow$ Γ diagonale

est toujours vrai !

C'est la réciproque dans le cas $X = (X_1, \dots, X_d)$ V.G. qui est intéressante :

X V.G avec $\mathrm{Cov}(X_i, X_j) = 0$ si $i \neq j$

$\Longrightarrow$ $X_1, \dots, X_d$ v.a. indépendantes !

En effet, il suffit de remarquer que

$$f_X(x) = \prod_{i=1}^{d} f_{X_i}(x_i)$$

dans le cas Γ diagonale

Retenir donc que dans le cas d'1 Vecteur Gaussien X :

non corrélation linéaire entre les composantes 2 à 2

$\Longrightarrow$ indépendance !

Sur la simulation d'1 VG en pratique :

décomposition de Cholesky de Γ = décomposition LU dans le cas symétrique : $U = L^T$ avec L triangulaire inférieure

On note souvent $S = L$, d'où

$$\Gamma = S S^T$$

En prenant $\varepsilon \sim \mathcal{N}(0, I_d)$, on a bien

$$X = \mu + \overset{A}{\widehat{S}}\, \varepsilon \quad \text{de loi } \mathcal{N}(0, \Gamma)$$

puisque : $E(X) = \mu$ et $Cov(X) = S S^T = \Gamma$

<u>Exemple</u> $(d=2)$: $X \sim \mathcal{N}(\mu, \Gamma)$; $\Gamma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

On écrit : $\begin{cases} X_1 = \mu_1 + \sigma_1 Y_1 \\ X_2 = \mu_2 + \sigma_2 Y_2 \end{cases}$

$$\Longrightarrow Y \sim \mathcal{N}\left(0 ; \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Décomposition de Cholesky de $\Gamma_Y = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$:

on écrit $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = S S^T = \begin{pmatrix} s_{11} & 0 \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} s_{11} & s_{21} \\ 0 & s_{22} \end{pmatrix} = \begin{pmatrix} s_{11}^2 & \cdots \\ s_{21} s_{11} & s_{21}^2 + s_{22}^2 \end{pmatrix}$

puis on identifie successivement s_{11}, s_{21} et s_{22} :

$1 = s_{11}^2$ et on prend $s_{11} = \sqrt{1} = 1$

$\rho = s_{21} s_{11}$ d'où $s_{21} = \rho$

$1 = s_{21}^2 + s_{22}^2$ d'où $s_{22}^2 = 1 - \rho^2$

et on prend $s_{22} = \sqrt{1-\rho^2}$

$$\boxed{S = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}}$$

De là : $Y = S\varepsilon \iff \begin{cases} Y_1 = \varepsilon_1 \\ Y_2 = \rho\,\varepsilon_1 + \sqrt{1-\rho^2}\,\varepsilon_2 \end{cases}$

et donc $\begin{cases} X_1 = \mu_1 + \sigma_1 \varepsilon_1 \\ X_2 = \mu_2 + \sigma_2 \rho\,\varepsilon_1 + \sigma_2 \sqrt{1-\rho^2}\,\varepsilon_2 \end{cases}$

jeu de données avec p variables et n individus } situation courante en pratique ! p peut être grand

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

$$X = \mu + A\varepsilon \iff \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} + \underbrace{\begin{bmatrix} A \end{bmatrix}}_{p \times q} \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_q \end{pmatrix}$$

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

$A : \mathbb{R}^q \longrightarrow \mathbb{R}^p$

Moyenne et matrice de covariance du vecteur X :

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

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

$AA^T : (p \times q) \times (q \times p) = p \times p$

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

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

RQUE : nécessairement, $q \geq p$

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

$$\mathbf{X = \mu + \sqrt{\lambda_1}\, Z_1 U_1 + \ldots + \sqrt{\lambda_d}\, Z_d U_d \ ; \ Z \sim N(0, I_d)}$$

cf. décomposition de Mahalanobis adaptée à cette situation

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

$$\Lambda = \left(\begin{array}{ccc|c} \lambda_1 & & & \\ & \ddots & & 0 \\ & & \lambda_d & \\ \hline & 0 & & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{array}\right)$$

0 est v.p.

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 ~~Y~~ 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 <u>encore</u> que

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

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.

<u>Détaillons l'exemple qui nous a servi de "fil rouge" dans ce cours :</u>

$$Z = \begin{pmatrix} Z_0 \\ Z_1 \\ Z_2 \end{pmatrix} \in \mathbb{R}^{d=3} \quad \text{avec} \quad \begin{cases} Z_k = \dfrac{1}{\sqrt{p_k}} \left(\mathbb{1}_{X=k} - p_k \right) \\ X \in \{0,1,2\} \text{ variable ternaire} \end{cases}$$

Un calcul simple (déjà vu) donne :

$$\mu = EZ = 0 \quad \text{et} \quad \Gamma = \mathrm{Cov}(Z) = \begin{pmatrix} 1-p_0 & -\sqrt{p_0p_1} & -\sqrt{p_0p_2} \\ -\sqrt{p_0p_1} & 1-p_1 & -\sqrt{p_1p_2} \\ -\sqrt{p_0p_2} & -\sqrt{p_1p_2} & 1-p_2 \end{pmatrix}$$

Γ est dégénérée (non inversible) car

$$\Gamma \begin{pmatrix} \sqrt{p_0} \\ \sqrt{p_1} \\ \sqrt{p_2} \end{pmatrix} = 0 \implies \ker\Gamma \text{ de dim } \geq 1.$$

$$\text{TCL multidimensionnel} \implies \frac{1}{\sqrt{n}} \left\{ Z^{(1)} + \dots + Z^{(n)} \right\} \underset{n \text{ grand}}{\sim} \mathcal{N}(0, \Gamma)$$

où donc <u>Γ est non inversible</u> ! (pas de densité !)

Solution exercice : analyse de la loi $\mathcal{N}(0;\Gamma)$ où

$$\Gamma = \begin{pmatrix} 1-p_0 & -\sqrt{p_0p_1} & -\sqrt{p_0p_2} \\ -\sqrt{p_0p_1} & 1-p_1 & -\sqrt{p_1p_2} \\ -\sqrt{p_0p_2} & -\sqrt{p_1p_2} & 1-p_2 \end{pmatrix}$$

On voit qu'en réalité $\Gamma = \underline{P}^2 = \underline{P}$ où $\underline{P}$ est un projecteur orthogonal.

Calculons par exemple $[P^2]_{11}$ si $P = \Gamma$:

$$[P^2]_{11} = (1-p_0)^2 + p_0p_1 + p_0p_2 = (1-p_0)^2 + p_0(1-p_0) = 1-p_0$$

donc $[P^2]_{11} = P_{11} = \Gamma_{11}$. De m̂ : $[P^2]_{ij} = P_{ij}$.

Par ailleurs, $P^T = \Gamma^T = P \Longrightarrow P$ projecteur orthogonal.

Or tout projecteur orthogonal se diagonalise sous la forme simple suivante :

$$P = U \begin{pmatrix} \underbrace{\begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix}}_{\dim(\mathrm{Im}\, P)} & \underbrace{0}_{} \\ 0 & \underbrace{\begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix}}_{\dim(\ker P)} \end{pmatrix} U^T$$

si l'on choisit $\begin{cases} (U_1, \dots, U_r) \text{ b.o.n de } \mathrm{Im}\, P \\ (U_{r+1}, \dots, U_d) \text{ b.o.n de } \ker P \end{cases}$ (cf. $\mathbb{R}^d = \mathrm{Im}\, P \overset{\perp}{\oplus} \ker P$)

Dans notre cas, $d=3$ et :

$$Tr(P) = Tr(\Gamma) = 1-p_0 + 1-p_1 + 1-p_2 = 3-1$$

soit $Tr(P) = 2$.

Par ailleurs, $Tr(P) = \sum_{k=1}^{d} \underbrace{\lambda_k}_{v.p. \in \{0,1\}} = dim(Im P)$ $\boxed{r = dim(Im P) = 2}$

Donc, P est le projecteur orthogonal correspondant à la projection orthogonale sur le plan vectoriel

$$\{x \in \mathbb{R}^3 \mid \langle x | u \rangle = 0\} \quad \text{où} \quad u = \begin{pmatrix} \sqrt{p_0} \\ \sqrt{p_1} \\ \sqrt{p_2} \end{pmatrix} \in \ker P$$

parallèlement à $\mathbb{R}.u$

De plus, la loi $\mathcal{N}(0;\Gamma)$ est la loi du V.a.

$$P\,\mathcal{E} \quad \text{où} \quad \mathcal{E} \sim \mathcal{N}(0, I_2)$$

puisque $Cov(P\mathcal{E}) = PP^T = P^2 = \Gamma$

Le test classique d'adéquation à une loi discrète ayant $k=3$ modalités :

jeu de données : $x_1 \quad x_2 \quad \dots \quad x_n$ réalisations indépendantes de $X \in \{0,1,2\}$

$$\begin{matrix} x_1 & x_2 & & x_n \\ \downarrow & \downarrow & & \downarrow \\ X_1 & X_2 & & X_n \end{matrix} \mapsto \text{i.i.d.}$$

$$1 \leq i \leq n : Z_k^{(i)} = \frac{1}{\sqrt{p_i}} \left(\mathbb{1}_{X_i = k} - p_k \right) \rightsquigarrow Z^{(i)} = \begin{pmatrix} Z_0^{(i)} \\ Z_1^{(i)} \\ Z_2^{(i)} \end{pmatrix}$$

$$\text{On a : } \frac{1}{\sqrt{n}} \left(Z^{(1)} + \dots + Z^{(n)} \right) = \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{N_0 - np_0}{\sqrt{p_0}} \\ \frac{N_1 - np_1}{\sqrt{p_1}} \\ \frac{N_2 - np_2}{\sqrt{p_2}} \end{pmatrix}$$

$$\text{TCL} \Rightarrow \frac{1}{\sqrt{n}} \left(Z^{(1)} + \dots + Z^{(n)} \right) \underset{n \text{ grand}}{\approx} \mathcal{N}(0, \Gamma) = \text{loi de } PE$$

$$\text{DONC : } \sum_{k=0}^{2} \frac{(N_k - np_k)^2}{np_k} \underset{n \text{ gd}}{\approx} \|PE\|^2 \text{ de loi } \chi^2(d=2)$$

car P projecteur de rang 2 !

On vient d'élucider le "fameux" test du χ^2 :

$$S := \sum_{k=0}^{2} \frac{(N_k - n p_k)^2}{n p_k}$$ statistique de test

effectif observé sur l'échantillon de la modalité n° k ($k = 0, 1, 2$)

effectif théorique moyen

Mise en oeuvre du test d'hypothèses :

H_0 : $p_0 = p_0^{(0)}$; $p_1 = p_1^{(0)}$; $p_2 = p_2^{(0)}$ contre H_1 = "non H_0"

(par exemple avec $p_0^{(0)} = 50\%$, $p_1^{(0)} = 25\%$; $p_2^{(0)} = 25\%$)

On calcule la valeur s de S sous H_0 et

si $s \geqslant$ seuil, on rejette H_0

Le calcul du seuil se fait en écrivant que

$$P(\text{"rejet de } H_0 \text{ à tort"}) = P_{H_0}(S \geq \text{seuil}) = \alpha$$

avec (par exemple) le risque $\alpha = 5\%$ (risque de 1ère espèce).

On $S \underset{n \text{ gd}}{\simeq} \chi^2(d=2)$ sous H_0 (si H_0 "vraie")

donc le seuil vérifie (à peu près) l'équation :

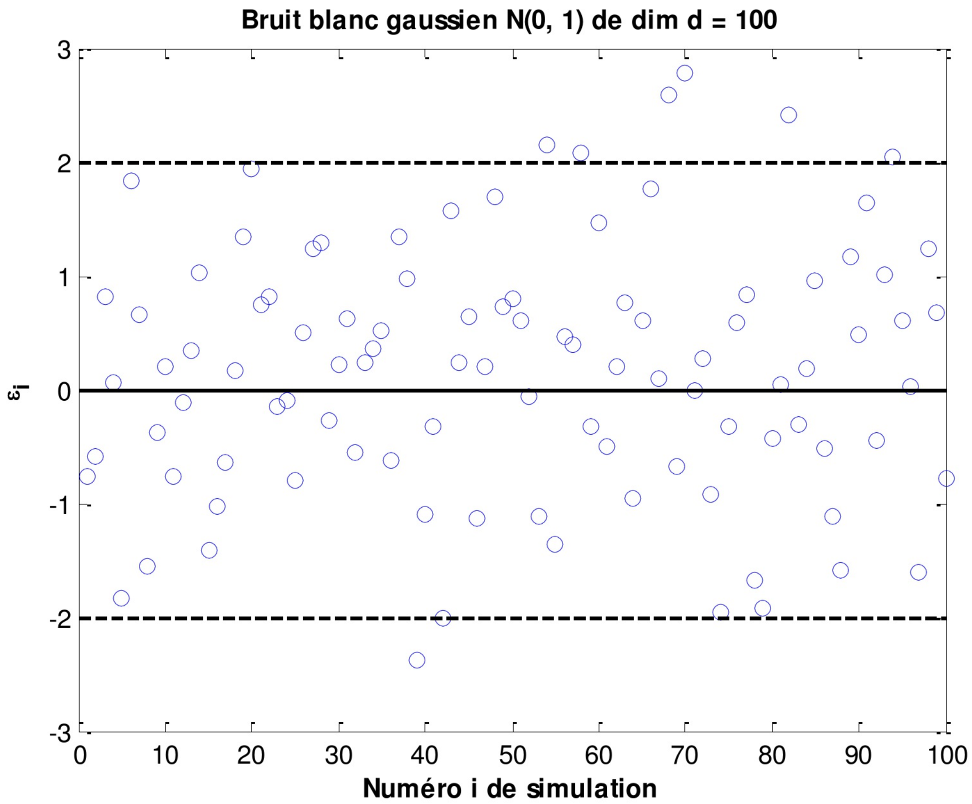
$$P(\chi^2(d=2) \geq \text{seuil}) = \alpha$$

$\downarrow$

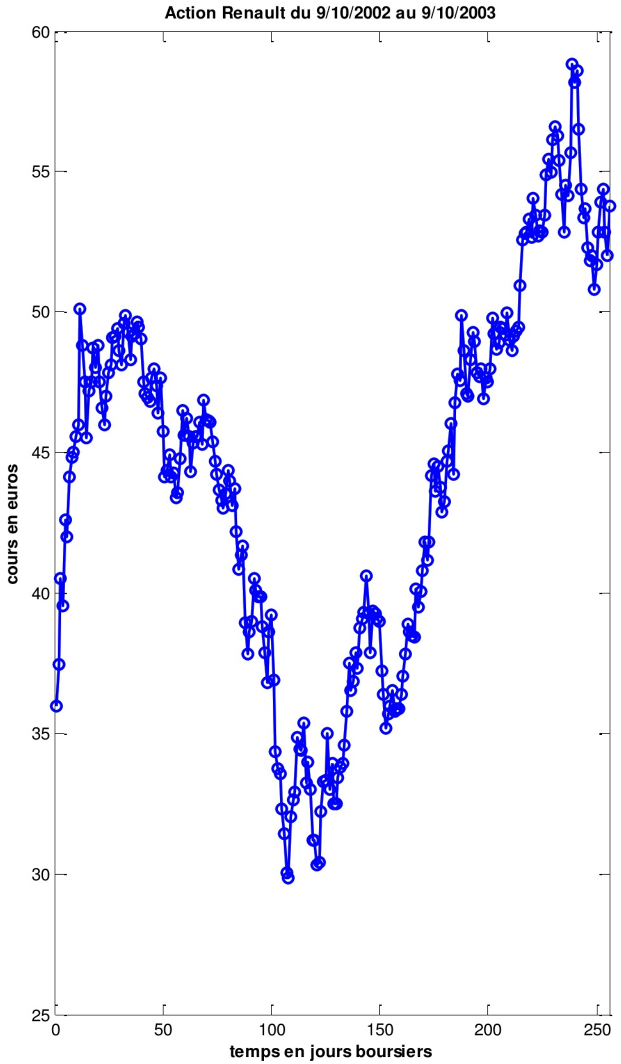
quantile d'ordre $1-\alpha$ de la loi $\chi^2(d=2)$

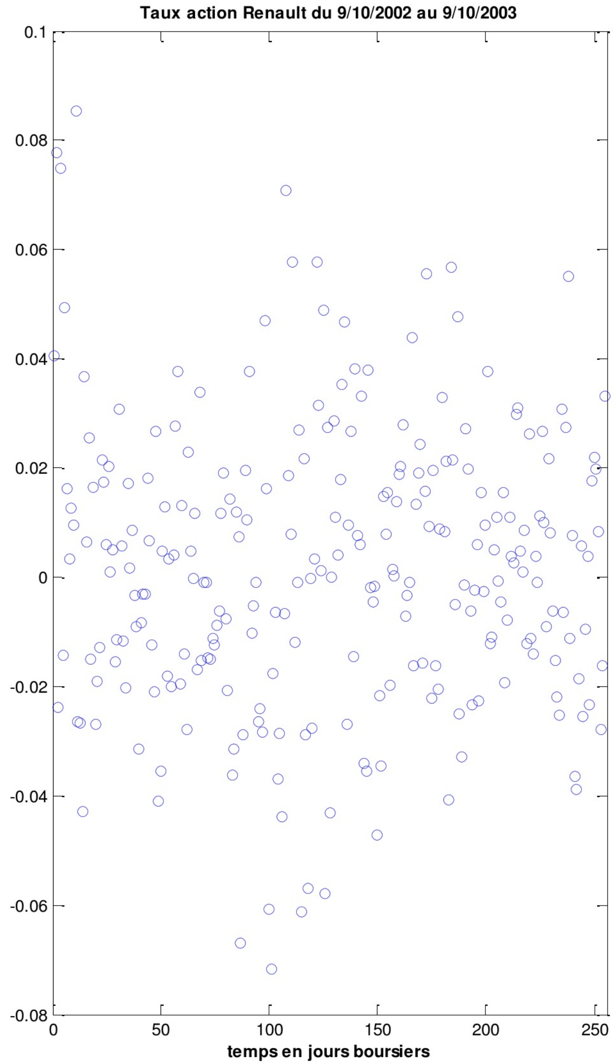
(loi exponentielle dans ce cas, vu en exercice)

Cas général : idem sauf que $S \underset{n \text{ gd}}{\simeq} \chi^2(k-1)$



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

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

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 (!)

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

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

II.2. Lois conditionnelles dans le cas continu

$(X, Y) \in \mathbb{R}^2$ 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

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, …, 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, …, 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.

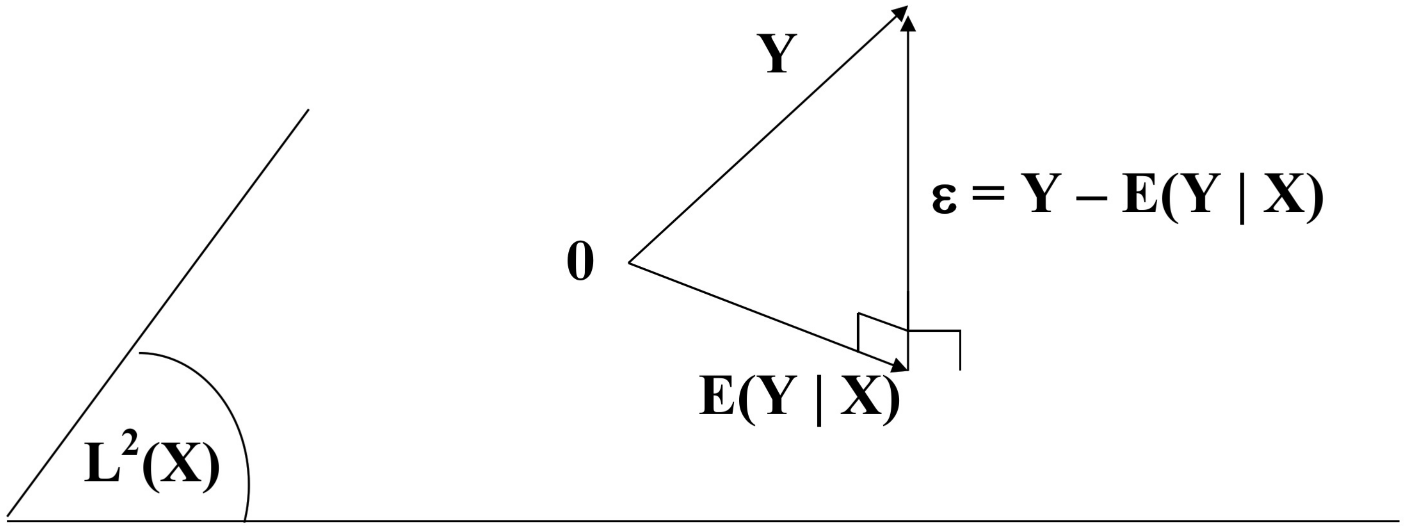
☺☺☺ On introduit pour cela l'espace vectoriel $L^2(P)$ des v.a. admettant une variance et muni du produit scalaire suivant :

$$<U, V> = E(UV)$$

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

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



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**