PC 2 – MAP 433 . 1.9.2023

Objectifs séance :
1. Récap cours et quizz
2. Étude approfondie du MLG
3. Étude des modèles exponentiels → PC3.

Modèle $\left(Z, \mathcal{Z}, \{\mathbb{P}_\theta, \theta \in \Theta\}\right)$

Estimateur : fct° mes à valeurs ds $\mathbb{R}^q$ si $g(\theta) \in \mathbb{R}^q$.

EMM :

$\forall \theta \in \Theta \ \forall j$
$$\mathbb{E}_\theta\left(T_j(X)\right) = e_j(\theta)$$

→ EMM(T) / $e_j(\hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^n T_j(X_i)$

Théorie : LGN / TCL

Z estimateurs : la médiane est un Z estimat par un EMM.

l/EMV modèle à densité

$$\{g_\theta^{\otimes n} \cdot \mu, \ \theta \in \Theta\}$$

$$\mathcal{L}_n(\theta, \mathbb{X}) = g_\theta^{\otimes n}(\mathbb{X}) = \prod_{i=1}^n g_\theta(X_i)$$

EMV $\hat{\theta}_{EMV} = \underset{\theta \in \Theta}{\operatorname{argmax}} \ \mathcal{L}_n(\theta, \mathbb{X}) = \text{———} \ \ell_n(\theta, \mathbb{X})$

s'ils existent.

Bonus EMV $\subset$ M - estimateurs.

$$\frac{1}{n}\sum_{i=1}^n \log g_\theta(X_i) \xrightarrow{\mathbb{P}_{\theta^*}} \mathbb{E}_{\theta^*}\left(\log g_\theta(X)\right)$$

$$\underset{\mathbb{P}_{\theta^*}}{S} \qquad \qquad \| \ \ M(\theta, \theta^*)$$

$\theta^* = \underset{\theta}{\operatorname{argmax}} \ M(\theta, \theta^*)$

$M(\theta,\theta^*) = \text{cte}_{\theta^*} - KL(P_{\theta^*}, P_\theta)$

Pt de vue M estim : fct $M(\theta, \theta^*, X) = \mathbb{E}_{\theta^*}\left(\log g_\theta(X)\right)$

$\theta^* = \operatorname{argmax} M(\theta, \theta^*)$ $\qquad \hat{\theta}_{EMV} = \operatorname{argmax} \frac{1}{n}\sum_{i=1}^n \log g_\theta(X_i)$

Quizz :

× Soit $A \in \mathbb{R}^{n \times d}$ et $X \sim \mathcal{N}(\mu, \Sigma)$ $\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$

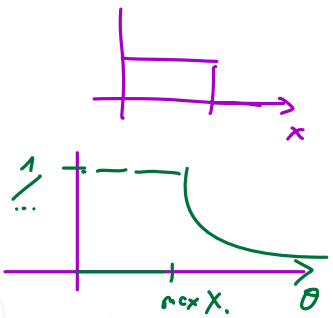
Loi de $AX \sim \mathcal{N}(A\mu, A\Sigma A^T)$

× Donner un exemple de modèle où EMV $\neq$ EMM (T)

$\left(\mathbb{R}^n, -, \{\mathcal{U}[0;\theta]^{\otimes n}, \theta > 0\}\right)$

$p_\theta(x) = \frac{1}{\theta} \mathbb{1}_{0 \leq x \leq \theta}$

EMV : $\max_{i=1..n} X_i = \hat{\theta}_{EMV}$ EMM ($T = Id$) : $\hat{\theta}_n = 2\overline{X}_n$



Preuve : $L_n(\theta, \mathbb{X}) = \prod_{i=1}^{n} p_\theta(X_i) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{X_i \leq \theta}$

$\hat{\theta}_{EMV} = \operatorname{argmax} L_n(\theta, \mathbb{X}) = \max(X_i) = X_{max} = \frac{1}{\theta^n} \mathbb{1}_{\max_{i=1..n} X_i \leq \theta}$

C'est l'objet de l'exo 3 du matin que pour une large classe de modèle (les modèles exponentiels), l'EMV coïncide à l'EMM pour un T naturel.

× On considère le modèle de Poisson : $\left(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathbb{P}_\lambda, \lambda > 0\}\right)$

Proposer plusieurs EMM.

$\mathbb{P}_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$

x Soit $X \in \mathbb{R}^{n\times d}$, $X = \begin{pmatrix} x_1 & x_2 & \dots & x_d \\ \vdots & \vdots & & \vdots \end{pmatrix} \in \mathbb{R}^{n\times d}$ $n \geq d$.

→ Exprimer $\mathrm{Im}\, X$. Majorer $rg(X)$.

→ Mq $\ker X = (\mathrm{Im}(X^T))^\perp$. Montrer que $(x_1 \dots x_d)$ famille libre $\iff$ X^TX inversible.

→ Montrer alors que $\forall z,\ p^\perp_{\mathrm{Im}(X)}(z) = X(X^TX)^{-1}X^Tz$

→ $\mathrm{Im}\, X = \{X\alpha,\ \alpha \in \mathbb{R}^d\} = \{\sum_{i=1}^{d} \alpha_i x_i,\ \alpha \in \mathbb{R}^d\} = \mathrm{Vect}(x_1 \dots x_d)$

$rg(X) = \min(n,d) = d$ $\quad n \geq d$.

$(X\alpha)^T\beta = \alpha^T(X^T\beta)$

→ Mq $\ker(X) = \mathrm{Im}(X^T)^\perp$.

$\alpha \in \ker(X) \iff X\alpha = 0 \iff \forall \beta,\ \langle X\alpha, \beta\rangle = 0 \iff \forall \beta,\ \langle \alpha, X^T\beta\rangle = 0 \iff \alpha \in (\mathrm{Im}\, X^T)^\perp$

→ $(x_1 \dots x_d)$ famille libre $\iff$ X^TX inversible.

$(x_1 \dots x_d)$ fa libre $\iff rg(X) = d \iff \ker(X) = \{0\} \iff X$ inj $\Leftarrow$ X^TX bij

X inj $\Rightarrow$ (X^TX) inv. : $u \in \ker(X^TX)$: $X^TXu = 0 \Rightarrow u^TX^TXu = 0$

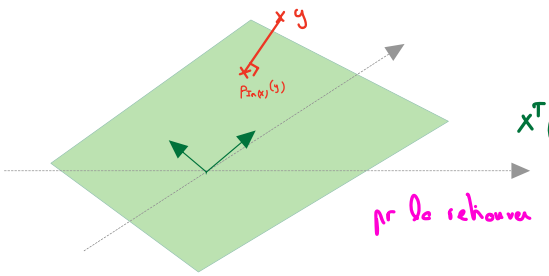
$\Rightarrow \|Xu\|^2 = 0 \Rightarrow Xu = 0$

D'où $\ker(X^TX) = \{0\}$ $\Rightarrow u = 0$

X inj

→ $p^\perp_{\mathrm{Im}(X)}(y)$ est caractérisé par : $p_{\mathrm{Im}(X)}(y) \in \mathrm{Im}(X)$ (α)

$y - p_{\mathrm{Im}(X)}(y) \in (\mathrm{Im}(X))^\perp \overset{(B)}{=} \ker(X^T)$



Formule donnée : $p_{\mathrm{Im}(X)}(y) = X(X^TX)^{-1}X^Ty \overset{\alpha}{\in} \mathrm{Im}(X)$

$X^T(y - X(X^TX)^{-1}X^Ty) \overset{(B)}{=} 0$.

pr la retrouver

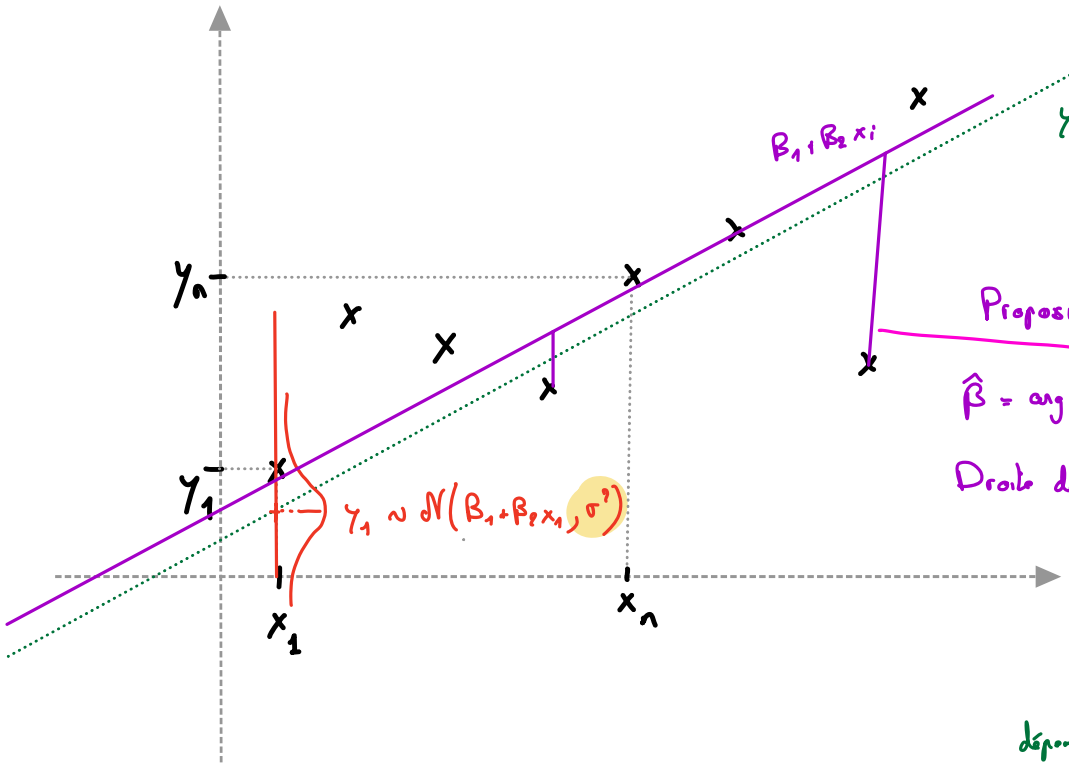
$p_{\mathrm{Im}X}(y) = Xz$?? avec z / $(y - Xz) \in (\mathrm{Im}\, X)^\perp$

$\Rightarrow \forall u\ \langle y - Xz, Xu\rangle = 0 \Rightarrow (X^Ty - X^TXz) = 0$

$\Rightarrow z = (X^TX)^{-1}X^Ty$.

$\Rightarrow p_{\mathrm{Im}(X)}(y) = X(X^TX)^{-1}X^Ty$

Exo 3 PC 1 et exo 1 PC 2



$y = \beta_1 + \beta_2 x$

$\beta_1, \beta_2, \sigma^2$ inconnus

Proposit° d'estimateur $\hat{\beta}$.

$\hat{\beta} = \operatorname{argmin} \sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 x_i)^2$

Droite des moindres carrés ordinaires.

On considère le modèle $\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{P_\theta \cdot Leb^{\otimes n} = \bigotimes_{i=1}^{n} \mathcal{N}(\beta_1 + \beta_2 x_i, \sigma^2)\}$;

dépend de i !

$\theta := (\beta_1, \beta_2, \sigma^2) \in \mathbb{R}^2 \times \mathbb{R}_+^*$

⚠ x_i ne s'[illegible] pas les obs, ils s'[illegible] fixes et connus.

Les obs s'[illegible] notées $(Y_1 \dots Y_n)^T = Y \longrightarrow (Y_i) \overset{\bigotimes_{i=1}^{n}}{\longrightarrow}$ "indep"

$\longrightarrow$ non identiq distribué.

$Y_i \sim \mathcal{N}(\beta_1 + \beta_2 x_i, \sigma^2)$.

♡ de l'exo : On va mq la droite des moindres carrés :

(qu'on aime pcq "pondère les grdes valeurs", "différ. et calcul explicite")

est exactement l'EMV du le modèle gaussien homoscédastiq

σ^2 ne dép. pas de i.

The concept of regression comes from genetics and was popularized by Sir Francis Galton during the late 19th century with the publication of *Regression towards mediocrity in hereditary stature*.[12] Galton observed that extreme characteristics (e.g., height) in parents are not passed on completely to their offspring. Rather, the characteristics in the offspring *regress* towards a *mediocre* point (a point which has since been identified as the mean). By measuring the heights of hundreds of people, he was able to quantify regression to the mean, and estimate the size of the effect. Galton wrote that, "the average regression of the offspring is a constant fraction of their respective mid-parental deviations". This means that the difference between a child and its parents for some characteristic is proportional to its parents' deviation from typical people in the population. If its parents are each two inches taller than the averages for men and women, then, on average, the offspring will be shorter than its parents by some factor (which, today, we would call one minus the regression coefficient) times two inches. For height, Galton estimated this coefficient to be about 2/3: the height of an individual will measure around a midpoint that is two thirds of the parents' deviation from the population average.

Galton coined the term "regression" to describe an observable fact in the inheritance of multi-factorial quantitative genetic traits: namely that the offspring of parents who lie at the tails of the distribution will tend to lie closer to the centre, the mean, of the distribution. He quantified this trend, and in doing so invented linear regression analysis, thus laying the groundwork for much of modern statistical modelling. Since then, the term "regression" has taken on a variety of meanings, and it may be used by modern statisticians to describe phenomena of sampling bias which have little to do with Galton's original observations in the field of genetics.

Wiki: À l'instar de Galton auquel le lie une indéfectible amitié, Pearson pense qu'il est possible et même éminemment souhaitable d'améliorer la race humaine en sélectionnant et favorisant les plus doués de ses représentants comme le fait la sélection naturelle pour les animaux.

L'eugénisme n'est pas une théorie scientifique, en relevant plutôt de l'idéologie scientifique[16]. Cependant en ayant recours à des méthodes statistiques pour évaluer les problèmes de son époque (dégénérescence et transmission de caractères mesurables dans une population), Karl Pearson et Walter Weldon fondent la biométrie. Les eugénistes anglais ont ainsi conçu et amélioré des outils mathématiques qui seront largement repris par les généticiens[14].

$$\left(\beta_1 + \beta_2 x_i\right)_{i=1}^{n}$$

$$Y = (Y_1 \dots Y_n), \quad X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \Longrightarrow \boxed{Y \sim \underbrace{\mathcal{N}}_{\text{gauss indép}}\left(X\beta, \sigma^2 I_n\right)}$$

modèle équivalent

1) Mq $(X^T X)$ est bijective

$(X^T X)$ bij $\overset{\text{rappel}}{\Longleftrightarrow}$ X inj $\Longleftrightarrow$ $\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}x_1\\\vdots\\x_n\end{pmatrix}$ fi libre $\Longleftrightarrow$ $(x_i)_{i=1}^n$ non tous égaux !

Rq : si X est non inject. : q se passe-t-il.

X matrice des var. explicatives : si les colonnes sont lin dép. je ne peux pas identifier $\beta_1 \dots \beta_p$.

Modèle est non identifiable : $\Theta \longrightarrow \mathcal{N}(X\beta, \sigma^2 I)$ non inj

si $\beta \to X\beta$ non inj !

2) Déterminer l'EMV :

Sol 1 $\quad L_n(\theta, Y) = \overset{\text{indép}}{\prod_{i=1}^{n}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - \beta_1 - \beta_2 x_i)^2}{2\sigma^2}}$

densité de Y_i / Leb.

$$L_n(\theta, Y) = \frac{1}{\sqrt{(2\pi)^n \sigma^{2n}}} \exp - \frac{\sum_{i=1}^{n} (Y_i - (\beta_1 + \beta_2 x_i))^2}{2\sigma^2}$$

$l_n(\theta, Y) = \dots \longrightarrow$ on différencie en $\beta_1, \beta_2, \sigma^2$

Sol 2 = sol pro $\quad Y \sim \mathcal{N}(X\beta, \sigma^2 I)$

densité de Y. $\quad L_n(\theta, Y) = \frac{1}{\sqrt{(2\pi)^n \underbrace{\det \Sigma}_{(\sigma^2)^n}}} \exp - \frac{(Y - X\beta)^T \Sigma^{-1} (Y - X\beta)}{2}$

$$\exp\left(-\frac{\|Y - X\beta\|^2}{2\sigma^2}\right)$$

(c'est pareil en matriciel)

Me donne $\ell_n(\theta, Y) = -\log\left((2\pi)^{n/2}\right) - \frac{n}{2}\log(\sigma^2) - \frac{\|Y - X\beta\|^2}{2\sigma^2}$

Que dire de $\underset{\theta=(\beta,\sigma^2)}{\operatorname{argmax}}\ \ell_n(\theta, Y)$.

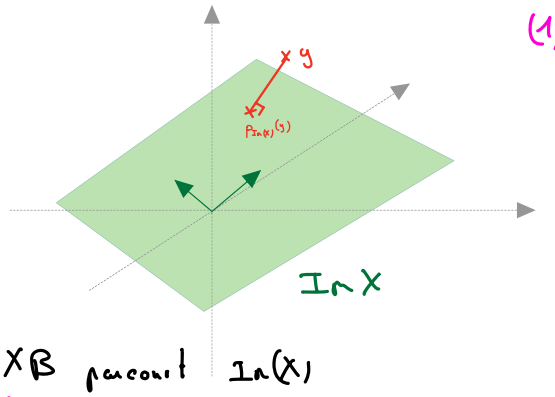
Rq : β correspond à la droite des moindres carrés !! $\sum_{i=1}^{n}(Y_i - (\beta_1 + \beta_2 x_i))^2$

2/ argmin en β ne dépend de la valeur de σ :

$$\begin{cases} \hat\beta_{EMV} \overset{(1)}{=} \underset{\beta\in\mathbb{R}^2}{\operatorname{argmin}} \|Y - X\beta\|^2 \\ \hat\sigma^2_{EMV} = \underset{\sigma}{\operatorname{argmax}} \ -\frac{n}{2}\log\sigma^2 - \frac{\|Y - X\hat\beta_{EMV}\|^2}{2\sigma^2} \end{cases}$$

Idée : Il faire en matriciel

Idée : ne pas différencier trop vite



$X\beta$ parcourt $Im(X)$

(1) $\overset{\text{Lemme}}{\Longrightarrow}$ $X\hat\beta_{EMV} = P_{Im(X)}(Y) = X(X^TX)^{-1}X^TY$

$\Longrightarrow$ $\hat\beta_{EMV} = (X^TX)^{-1}X^TY$.

X inj

Lemme

$\underset{z\in Im(X)}{\operatorname{argmin}} \|Y - z\|^2 = P_{Im(X)}(Y)$

3/ Loi de $\hat\beta_{EMV}$?

Sous P_θ ; $Y \sim \mathcal{N}(X\beta, \sigma^2 I)$.

Lemme
$Z \sim \mathcal{N}(\mu, \Sigma)$
$AZ \sim \mathcal{N}(A\mu, A\Sigma A^T)$

$\hat\beta_{EMV} \sim \mathcal{N}\left(\underbrace{(X^TX)^{-1}X^T X\beta}_{\beta} ; \sigma^2 (X^TX)^{-1}X^T X (X^TX)^{-1}\right)$

Sous P_θ $\Longrightarrow$ $\hat\beta_{EMV} \sim \mathcal{N}(\beta ; \sigma^2(X^TX)^{-1})$

$\forall\theta$; $\mathbb{E}_\theta(\hat\beta_{EMV}) = \beta$: c'est un est. sans biais !!

4/ Mq $\hat\beta_{EMV,1} \perp \hat\beta_{EMV,2} \iff \sum_{i=1}^{n} x_i = 0$

$\hat\beta_{EMV,1} \perp \hat\beta_{EMV,2} \iff (X^TX)^{-1}$ diago $\iff (X^TX)$ diago $\iff \sum_{i=1}^{n} x_i = 0$

$(X^TX) = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$

5/ $\hat\sigma^2_{EMV} = \underset{}{\operatorname{argmax}} \underbrace{-\frac{n\log\sigma^2}{2} - \frac{\|Y - X\hat\beta_{EMV}\|^2}{2\sigma^2}}_{g(\sigma^2)}$

$$\frac{\partial}{\partial \sigma^2} g(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{\|Y - X\hat{B}\|^2}{2\sigma^4} = 0$$

$$\Leftrightarrow \quad \hat{\sigma}^2_{EMV} = \frac{\|Y - X\hat{B}_{EMV}\|^2}{n} = \frac{\|Y - P_{Im(X)}(Y)\|^2}{n} = \frac{\|P_{Im(X)^\perp}(Y)\|^2}{n}$$

2 questions : loi de $\hat{\sigma}^2_{EMV}$; m.q $\hat{\sigma}^2_{EMV} \perp\!\!\!\perp \hat{B}_{EMV}$

Théorème : Théorème de Cochran

Soit X un vecteur aléatoire gaussien de $\mathbb{R}^n$ de loi $\mathcal{N}(0_{\mathbb{R}^n}, \sigma^2 \mathrm{Id}_n)$. Soit F un sev de $\mathbb{R}^n$ de dimension d, $F^\perp$ son orthogonal et $P_F, P_{F^\perp}$ les matrices des projections orthogonales sur F et $F^\perp$. Alors :

- les vecteurs aléatoires $P_F X, P_{F^\perp} X$ sont indépendants et de lois respectives $\mathcal{N}(0_{\mathbb{R}^n}, \sigma^2 P_F), \mathcal{N}(0_{\mathbb{R}^n}, \sigma^2 P_{F^\perp})$;
- les variables aléatoires réelles $\|P_F X\|^2, \|P_{F^\perp} X\|^2$ sont indépendantes et de lois respectives $\sigma^2\chi^2(d), \sigma^2\chi^2(n-d)$.

Preuve $Z \sim \mathcal{N}(0, \sigma^2 I_n)$

Cas n°1 $F = \mathrm{Vect}(e_1 \dots e_d)$; $P_F(z_1 \dots z_n) = (z_1, \dots, z_d, 0, \dots 0)$

$P_{F^\perp}(\dots) = (0, \dots 0, z_{d+1} \dots z_n)$

$(Z_1 \dots Z_n)$ i.i.d. $\Rightarrow P_F(Z) \perp\!\!\!\perp P_{F^\perp}(Z)$!

$$\|P_F(Z)\|^2 = \sum_{i=1}^{d} z_i^2 \sim \sigma^2 \chi^2(d)$$

↑ def.

Cas général je considère une BON adaptée à F et $F^\perp$ et U la matrice de chgt de base.

soit $W = UZ$. $\longrightarrow P_F(Z) = (W_1 \dots W_d, 0 \dots 0)$

$W \sim \mathcal{N}(0, \sigma^2 U^T U) = \mathcal{N}(0, \sigma^2 I_n)$.

On se range au cas n°1.

On a $X\hat{\beta}_{EMV} = P_{Im(X)}(Y)$ et $\hat{\sigma}^2_{EMV} = \dfrac{\| P_{Im(X)^\perp}(Y) \|^2}{n}$

$\Rightarrow \hat{\beta}_{EMV} \perp\!\!\!\perp \hat{\sigma}_{EMV}$

↑ lemme de cochet.

Loi de $\hat{\sigma}^2_{EMV} = \dfrac{\| P_{Im(X)^\perp}(Y) \|^2}{n} = \dfrac{\| P_{Im(X)^\perp}(Y - X\beta) \|^2}{n} \sim \sigma^2 \chi^2(?)$

$\sim \mathcal{N}(0, \sigma^2 I)$

⚠ $Y \sim \mathcal{N}(X\beta ; \sigma^2 I)$ (↑ non centrée)

$\dim(Im\, X^\perp) = n - 2$

$$\hat{\sigma}_{EMV} \sim \frac{\sigma^2 \chi^2(n-2)}{n}$$

Gosset : cas particulier avec 1 seule v.a explicative

$$X = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \qquad \beta = \mu$$

$Y \sim \mathcal{N}(\mu \vec{1}, \sigma^2 Id) \longrightarrow \hat{\mu}_n = \bar{Y}_n = \underbrace{(1^T 1)^{-1}}_{1/n} \underbrace{1^T Y}_{\sum Y_i}$

$\hat{\sigma}^2_{n\,EMV} = \frac{1}{n} \sum (Y_i - \bar{Y}_n)^2$

On retrouve l'exo 2 PC1 !