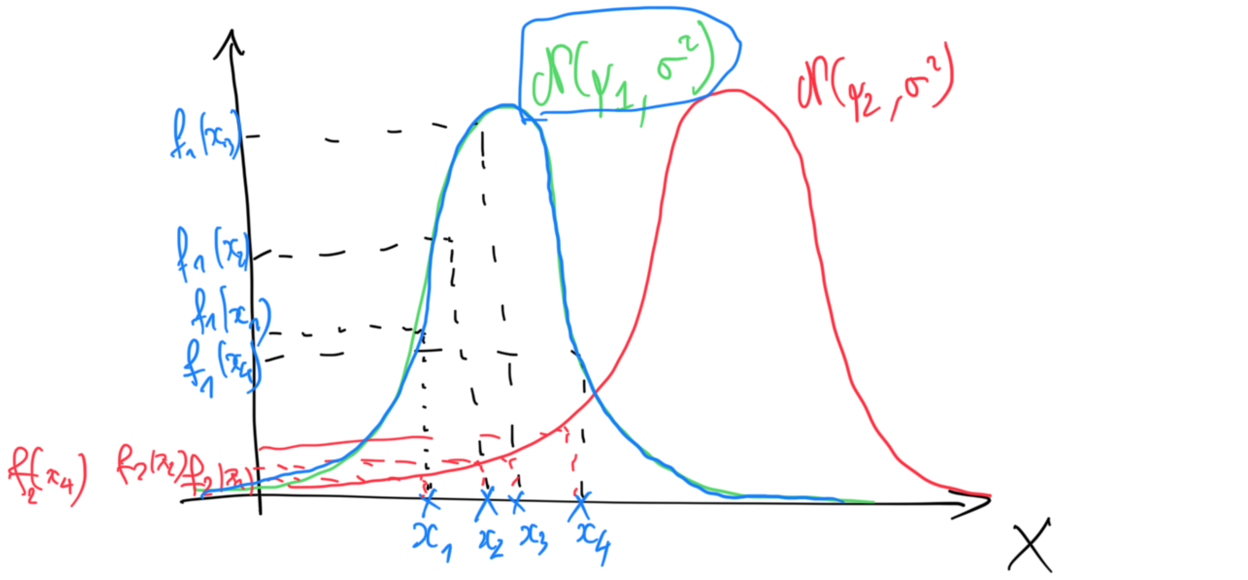
Formation Stat Bayésienne, 22-23 mai 2023



$$\underbrace{f_1(x_1)\times\cdots\times f_1(x_4)}_{\text{vraisemblance des données pour le modèle bleu (1)}} > \underbrace{f_2(x_1)\times\cdots\times f_2(x_4)}_{\text{vraisemblance du modèle (2)}}$$

Vraisemblance (Likelihood)

$$L(x_1,\dots,x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Calcul de la loi a posteriori

$$p(\theta \mid \underline{x}) = \frac{\boxed{\ell(\underline{x} \mid \theta)}\; p(\theta)}{p(\underline{x})}$$

$p(\underline{x} \mid \theta)$

$p(\underline{x})$ ne dépend pas de θ

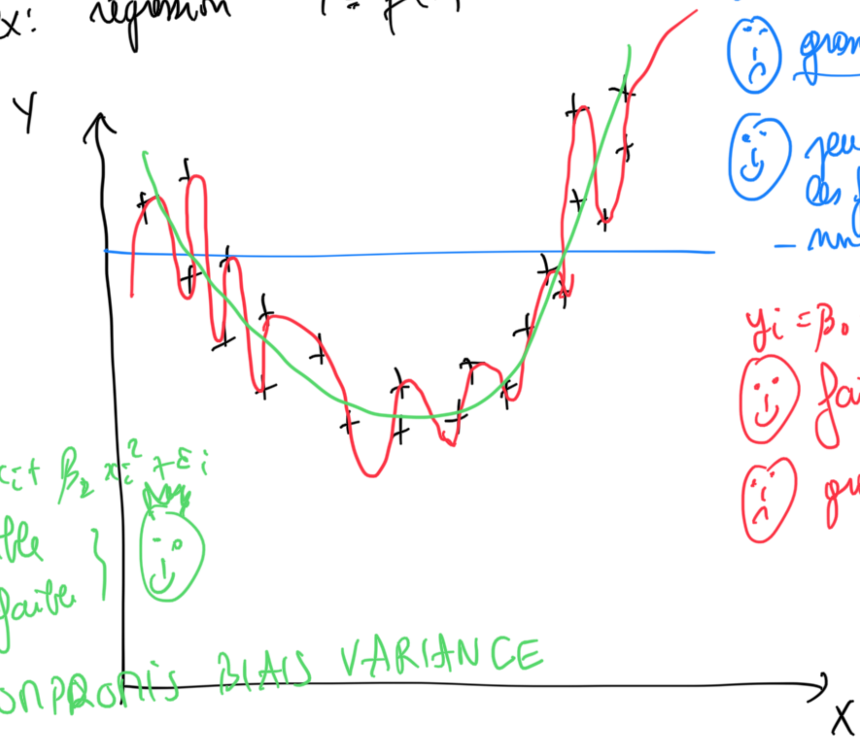
$$p(\underline{x}) = \int p(\underline{x} \mid \theta)\, p(\theta)\, d\theta$$

/ θ

NON calculable analytiquement pour certains $p(X|\theta)$, $p(\theta)$

Complexité du modèle

ex: régression $Y = f(x)$



$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

- grand biais (résidus ε_i grands)
- peu influencé par les fluctuations d'échantillonnage → variance faible

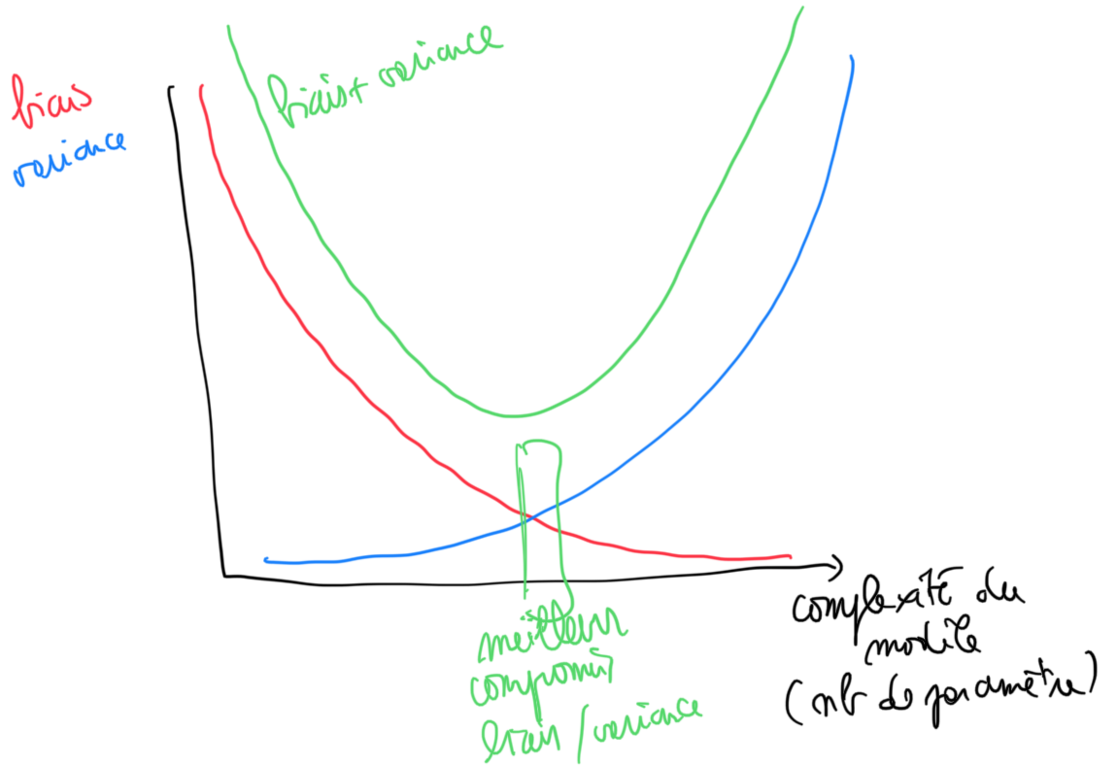
$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_{22} x_i^{22} + \varepsilon_i$

- faible biais
- grande variance
  OVER-FITTING

$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$

biais faible
variance faible

BON COMPROMIS BIAIS VARIANCE



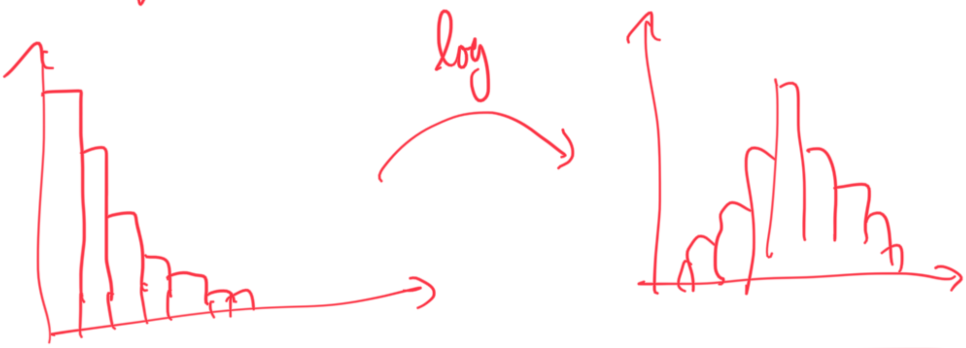
Si X et Y sont deux variables aléatoires gaussiennes et si on cherche à approcher Y par une fonction f de X :

$$Y = f(X)$$

La meilleure f possible (qui minimise $E\|Y - f(X)\|^2$) est la fonction linéaire

$$f(x) = ax + b$$

$\Rightarrow$ Avant d'utiliser un modèle de régression linéaire on s'assurera que Y est approximativement gaussien, quitte à le transformer par une transformation de type logarithmique



$$y \mid \alpha, \beta, \sigma^2 \sim \mathcal{N}(\alpha \mathbb{1}_n + X\beta, \sigma^2 I_n)$$

$$(\alpha \sim \qquad)$$

$$\boxed{\beta \sim \mathcal{N}(\tilde{\beta}, ?)} \qquad \beta = \begin{pmatrix} \beta_1 \\ | \\ \beta_p \end{pmatrix}$$

$\sigma^2 \sim$ non informatif

Espérance $\tilde{\beta}$ de la loi a priori est généralement le terme qui porte l'information a priori

? : $Cor(\hat{\beta})$ = matrice $p \times p$

⟹ Régression bayesienne avec a priori de Zellner

Questionnaire, J2

- La statistique bayesienne permet d'introduire de l'information dont on dispose sur le phénomène étudié en plus de celle contenue dans les données

 oui ☐ non ☐

- Je peux choisir la loi a priori que je souhaite en fonction du résultat que je veux obtenir

 vrai ☐ faux ☐

- Lorsque la taille d'échantillon est grande, l'influence de la loi a priori est faible

 vrai ☐ faux ☐

- La statistique bayesienne est l'approche à privilégier pour les petits échantillons

 vrai ☐ faux ☐

- Je peux utiliser une approche bayesienne sans connaissance a priori vrai ☐ faux ☐

- Pour tester des hypothèses en statistique bayésienne, je mets en compétition des modèles

vrai ☐ faux ☐

- Le terme $p(\theta|\underline{x}) = \dfrac{\ell(\underline{X}|\theta)\, p(\theta)}{p(X)}$ est

la vraisemblance ☐

la loi a priori ☐

la loi a posteriori ☐

- Pour choisir le meilleur parmi deux modèles de régression de complexités différentes (nombre de paramètres), je peux utiliser :
  - la vraisemblance ☐
  - le critère BIC ☐
  - le critère R^2 ☐
  - l'erreur de prédiction sur des données test, non utilisées pour estimer les paramètres des modèles ☐
  - le critère AIC ☐
  - l'erreur de prédiction sur les données d'apprentissage ☐
  - l'erreur de prédiction