Modélisation Statistique

Charles Vin

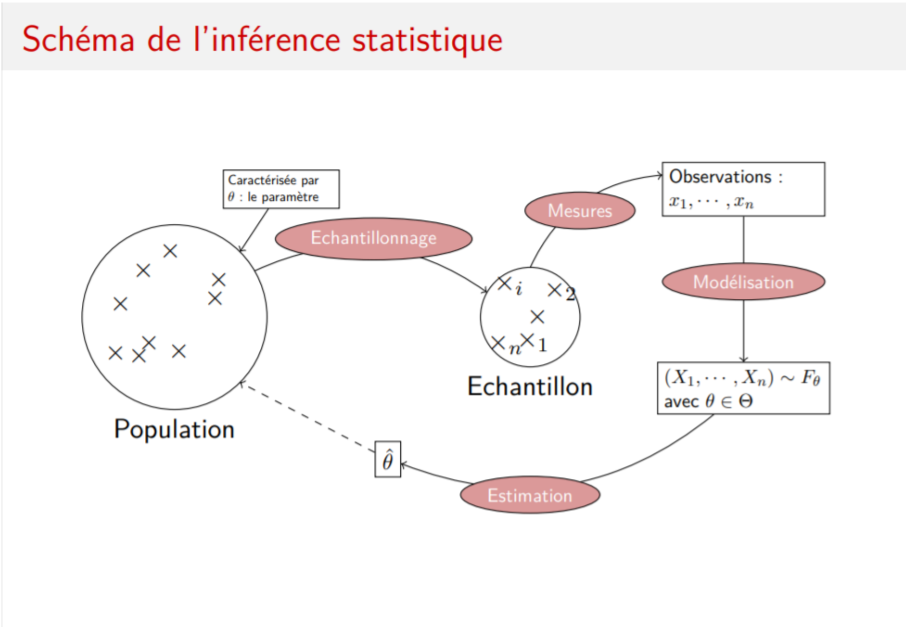
Date

Table des matières

1 Introduction à la modélisation

Un modèle est une simplification de la réalité, formelle grâce à des équations mathématiques conçus dans un but de prédiction.
La modélisation statistique s'appuie sur les résultats d'une expérience. Lorsqu'on répète l'expérience, le résultat peut varier.
Le but est de décrire le phénomène/processus à l'origine des données. Représenté par une variable aléatoire.
Il y a deux rôles

— Descriptif : comprendre comment sont générés les observations, quel lien il peut y avoir entre les variables.
— Prédictif : pouvoir faire une prédiction pour une nouvelle date, lieu, individu...



La modélisation statistique c'est

— Supposer que les observations $x_1, x_2, \ldots, x_n$ sont la réalisation de variables aléatoires $X_1, X_2, \ldots, X_n$.
— Donner une famille de loi possible F pour le vecteur aléatoire $(X_1, X_2, \ldots, X_n)$
  — Si on peut écrire F sous la forme suivante $F = \{F_\theta, \theta n \theta \in \Theta\}$ avec θ de dimension fini, on dira que le modèle est **paramétrique**. Par exemple $F = \{P(\lambda), \lambda, \lambda \in \mathbb{R}^+\}$.
  — On considérera souvent que les variables $X_1, \ldots, X_n$ sont **i.i.d.**.

L'échantillonnage est très important, sans contrôler l'échantillon on peut dire tout et son contraire. On ne vas pas vraiment s'attarder sur les méthodes l'échantillonnage mais il faut garder en tête que si celui-ci est mauvais, tout le reste, toute notre modélisation sera erroné.

2 La régression linéaire simple

2.1 Contexte

On dispose d'un couple de variable aléatoire (X, Y).

Théorème 2.1 (de la variance totale)**.**

$$\underbrace{V(Y)}_{\text{Variance Totale}} = \underbrace{E(V(Y)|X)}_{\text{Variance résiduelle}} + \underbrace{V(E(Y|X))}_{\text{Variance expliquée}} .$$

Donc

$$V(Y) \leq E(V(Y)|X).$$

Le fait de connaître X permet de diminuer l'incertitude sur Y . Ainsi X pourrait servir à prédire Y. On pourrait prédire Y par une fonction de X : $\hat{Y} = f(X)$.

Preuve :

$$\begin{aligned} V(Y) &= E(Y^2) - E(Y)^2 \\ &= E(E(Y^2)|X) - E(Y|X)^2 \\ &= E(E(Y^2|X)) - E(E(Y)^2|X) + E(E(Y)^2|X) - E(E(Y)|X)^2 \\ &= E(E(Y^2) - E(Y)^2|X) + Var(E(Y|X)) \\ E(Var(Y)|X) & \end{aligned}$$

□

Corollaire (Résultat important admis)**.**

$$\arg\max E((Y - f(X))^2) = E(Y|X).$$

La meilleurs façon d'utiliser X pour prédire Y est de prendre $f(X) = E(Y|X)$. On va alors poser

$$Y = E(Y|X) + \epsilon.$$

avec ϵ un terme d'erreur aléatoire. On peut montrer que
- $E(\epsilon) = 0$
- $Cov(\epsilon, X) = Cov(\epsilon, E(Y|X)) = 0$
- $V(\epsilon) = (1 - \frac{V(E(Y|X))}{V(Y)})V(Y)$

2.2 Cadre de la régression linéaire

Dans le cadre de la régression linéaire on va poser

$$\begin{aligned} &E(Y|X) = \alpha + \beta X \\ \Leftrightarrow &Y = \alpha + \beta X + \epsilon \end{aligned}$$

Preuve : (je sais pas trop ce qu'elle prouve là) Le point $(E(X), E(Y))$ appartient à la droite d'équation $E(Y|X) = \alpha + \beta X$

$$E(E(Y|X)) = E(\alpha + \beta X) \Rightarrow E(Y) = \alpha + \beta E(X).$$

$$\begin{aligned} E(Y) = \alpha + \beta E(X) \Rightarrow Y - E(Y) & \\ &= \alpha + \beta X + \epsilon - \alpha - \beta E(X) \\ &= \beta(X - E(X) + \epsilon) \Rightarrow (Y - E(Y))(X - E(X)) = \beta(X - E(X)^2) + E(X - E(X)) \\ &\Rightarrow E((Y - E(Y))(X - E(X))) = \beta E((X - E(X))^2) + E(\epsilon(X - E(X)))_{=Cov(\epsilon,X)=0} \\ \beta &= \frac{Cov(X, Y)}{V(X)} = \rho_{X,Y}\frac{\sigma_y}{\sigma_x} \end{aligned}$$

□

2.3 Application à la statistique

On dispose d'un échantillon de taille n sur lequel on mesure deux variables que l'on nommera y_i et x_i.

On suppoera que le couple (x_i, y_i) DIAPO 17

2.4 Exemple des salaires

2.4.1 Le modèle linéaire

Quand elle demande d'écrire le modèle il faut écrire

$$\begin{aligned} &Y_i = \alpha + \beta x_i + E_i \\ &E_i i.i.d., V(E_i) = \sigma^2 \end{aligned}$$

2.4.2 Estimation de α, β, σ par les moindres carrés

La méthode des moindre carrés ordinaire (MCO), consiste à choisir a, b qui minimise $\sum_{i=1}^n (y_i - (\alpha + x_i\beta))^2$. On pose donc la fonction de coût suivante

$$F(\alpha, \beta) = \sum_{i)1}^{n} (y_i - \alpha + \beta x_i)^2.$$

Nouveau cours du 21/01

Note. y'a des belles images dans le diapo

Selon la méthode des moindres carrés ordinaire on estime α et β par

$$(a, b) = \arg \min_{(\alpha,\beta)\in\mathbb{R}^2} F(\alpha, \beta).$$

Pour trouver ça, on utilise la gradient comme l'année dernière :

$$\frac{\partial F}{\partial \alpha}(\alpha; \beta) = \sum_{i=1}^{n} 2(-1)(y_i - \alpha - \beta x_i) = -2(n\bar{y} - n\alpha - \beta n\bar{x})$$
$$\frac{\partial F}{\partial \beta}(\alpha; \beta) = \sum_{i=1}^{n} 2(-x_i)(y_i - \alpha - \beta x_i) = -2(n\bar{xy} - n\alpha\bar{x} - \beta n\bar{x^2})$$

avec $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i, \bar{y} = \frac{1}{n}\sum_{i=1}^n y_i, \bar{x^2} = \frac{1}{n}\sum_{i=1}^n x_i^2$

$$\nabla F(\alpha, \beta) = 0$$
$$\Leftrightarrow \begin{cases} \frac{\partial F}{\partial \alpha}(\alpha \\ \quad \beta) = 0 \\ \frac{\partial F}{\partial \beta}(\alpha \\ \quad \beta) = 0 \end{cases} \Leftrightarrow \begin{cases} \bar{y} = & \alpha + \beta\bar{x} = 0 \\ \bar{xy} = \alpha\bar{x} + \beta\bar{x^2} = 0 \end{cases} \Leftrightarrow \begin{cases} a = & \bar{y} - b\bar{x} \\ \bar{xy} = (\bar{y} - b\bar{x})^{\bar{x}} + b\bar{x^2} \end{cases}$$

Focus sur la deuxième équation

$$\bar{x}\bar{y} = (\bar{y} - b\bar{x})^{\bar{x}} + b\bar{x^2}$$
$$\Leftrightarrow \bar{xy} - \bar{y}\bar{x} = b(\bar{x^2} - \bar{x}^2)$$
$$\Leftrightarrow b = \quad = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Remarque.
— b est la version empirique de $\frac{Cov(X,Y)}{V(X)}$
— On pose $y_i^* = a + bx_i = \bar{y} + b(x_i - \bar{x})$, c'est l'estimation de $E(Y_i|X = x_i)$

2.4.3 Variance des estimateurs

Proposition 2.2. *a, b, y_i^* sont des estimation sans biais de, respectivement, $\alpha, \beta, E(Y|X = x_i) = \alpha + \beta x_i$*

Preuve : ...j'ai un tableau de retard... j'ai pas $E(B)$ gros gros calcul
a est une réalisation de $A = \bar{Y} - B\bar{X}$

$$E(A) = E(\bar{Y} - E(B)\bar{x}) = E(\bar{Y} - E(B)\bar{x}) = \alpha + \beta\bar{x} - \beta\bar{x} = \alpha$$
$$E(Y_i^*) = E(A + Bx_i) = E(A) + E(B)x_i = \alpha + \beta x_i$$

□

Proposition 2.3 (Gauss-Markov)**.** *Parmi les estimateur linéaires des Y_i, A, B sont de variance minimales. On dit que ce sont des estimateur BLUE (Best linear Unbiased Estimates). Ce sont les meilleurs estimateur linéaire que l'ont peut trouver en terme de variance.*

Calcul de la variance des estimateurs

$$V(B) = V(\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2})$$
$$\frac{1}{(2(x_i-\bar{x}))^2}V(\sum_{i=1}^{n}(Y_i\bar{y})(x_i-\bar{x}))$$

2 tableaux de calculs horribles, voir moodle

$$\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$
$$V(A) = V(\bar{Y}+\beta\bar{x}) = V(\bar{Y}) + \bar{x}^2V(B) - \bar{x}Cov(\beta,\bar{Y})_{=0}$$
$$= \frac{\sigma^2}{n} + \bar{x}^2\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$
$$= \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2})$$
$$V(Y_i^*) = V(A+\beta x_i) = V(A) + x_i^2V(\beta) + 2x_iCov(A,B)$$
$$= \sigma^2(\frac{1}{n} + \frac{(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2})$$

2.5 Le modèle linéaire Gaussien

On pense que les erreur suivent une gaussienne pour pouvoir faire des tests

2.5.1 Exemple des salaires

Voir diapo
Nouveau cours du 28/01

2.5.2 Décomposition de la variance

$$SCT = SCM + SCR$$
$$\sum_{i=1}^{n}(y_i-\bar{y}) = \sum_{i=1}^{n}((y_i-y_i^*)+(y_i^*-\bar{y}))^2$$
$$\sum_{i=1}^{n}(y_i-\bar{y}) = \sum_{i=1}^{n}(y_i-y_i^*)^2 + \sum_{i=1}^{n}(y_i^*-\bar{y})^2 + 2\sum_{i=1}^{n}(y_i-y_i^*)(y_i^*-\bar{y})$$

Or $2\sum_{i=1}^{n}(y_i-y_i^*)(y_i^*-\bar{y}) = 0$ (remplacer y_i^* par $a+bx_i$ et a et b par leur estimation $\rightarrow$ Pythagore

2.5.3 Coeffficient de détermination R^2

$$\begin{aligned}
SCR &= \sum_{i=1}^{n}(y_i - y_i^*)^2 = \sum_{i=1}^{n}(y_i - a - bx_i)^2 \\
&= \sum_{i=1}^{n}(y_i - (\bar{y} - b\bar{x}) - bx_i)^2 \\
&= \sum_{i=1}^{n}((y_i - \bar{y}) - b(x_i - \bar{x}))^2 \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 + b^2\sum_{i=1}^{n}(x_i - \bar{x})^2 - 2b\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\
&= S_y^2 + b^2S_x^2 + 2bS_{xy}
\end{aligned}$$

Or $b = \frac{S_{xy}}{S_x^2}$, donc

$$\begin{aligned}
SCR &= S_y^2 + \frac{S_{xy}^2}{(S_x^2)^2}S_x^2 - 2\frac{S_{xy}}{S_x^2}S_{xy} \\
&= S_y^2(1 - \frac{S_{xy}}{S_x^2S_y^2}) \\
&= SCT(1 - R^2) \Rightarrow \frac{SCR}{SCT} = 1 - R^2
\end{aligned}$$

Or $SCM = SCT - SCR \Rightarrow 1 - R^2 = 1 - r^2 \Rightarrow R^2 = r^2$

2.5.4 Solutions

Soient deux variable X et Y définies sur (Ω, t, P)

$$\rho_{X,Y} = \frac{E((X - E(X))(Y - E(Y)))}{\sqrt{V(X)V(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}}.$$

On dispose de n couple (x_i, y_i) iid et on suppose que (x_i, y_i) est une réalisation de (X, Y). L'estimation de Pearson de $\rho_{X,Y}$ est

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x - \bar{x})(y - \bar{y})}{\sqrt{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)(\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2)}}$$

2.5.5 Test d'analyse de la variance

2.5.6 Intervalle de confiance pour $E(Y_j|X = x_j)$

$$\begin{aligned}
Y_j^* &= A + Bx_j \\
V(Y_j^*) &= V(A + Bx_j) \\
&= V(\bar{Y} - B\bar{x} + Bx_j) \\
&= V(\bar{Y} + B(x_i - \bar{x})) = V(\bar{Y}) + (x_j - \bar{x})^2V(B) \text{ car } B \text{ est non corrélé à } Y \\
&= \frac{\sigma^2}{n} + (x_j - \bar{x})\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sigma^2(\frac{1}{n} + \frac{(x_j - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2})
\end{aligned}$$

2.5.7 Intervalle de prédiction pour une nouvelle donnée

2.5.8 Validation des hypothèses du modèle

— $y_i = \alpha + \beta x_i + E_i$ "les erreurs sont indépendantes
— ....

2.5.9 Graphique des résidus vs valeurs prédites

Nouveau cours du 04/02
full diapo sur les espaces euclidiens
Dessin de projection dans onenote
Toutes la fin dans OneNote

3 Régression linéaire multiple

— Est-ce que nos point sont aligné sur un plain
— Même hypothèses qu'avant
— On peut écrire $\mathbb{Y}$ sont plusieurs forme (qu'on a vu avec la régression en DS)
— On utilise la méthode des moindres carrées again

Exemple 3.1 (Exercice diapo)**.** 1. X matrice avec n lignes et $p+1$ colonnes. Si le modèle est bien posé, toutes les colonnes de X sont linéairement indépendantes : $rg(X^TX) = rg(X) = p+1$

2. $X^TX \to (p+1 \times n)(n \times p+1)$.. $(X^TX)^T = X^TX =$ Symétrique.
   Soit $u \in \mathbb{R}^{p+1}, u^TX^TXu = (Xu)^T(Xu) =< Xu, Xu >= \|Xu\|^2 > 0$
3. Si $p = 1$:

$$X^TX = \begin{pmatrix} 1 & \dots & 1 \\ x_{11} & \dots & x_{n1} \end{pmatrix} \begin{pmatrix} 1 & X_{11} \\ \vdots & \vdots \\ 1 & x_{n1} \end{pmatrix} = \begin{pmatrix} n & \sum x_{i1} \\ \sum x_{i1} & \sum x_{i1}^2 \end{pmatrix}$$

Si $p = 3$

$$X^TX = \begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} & \sum x_{i3} \\ & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \sum x_{i1}x_{i3} \\ & & \sum x_{i2}^2 & \sum x_{i2}x_{i3} \\ & & & \sum x_{i3}^2 \end{pmatrix}.$$

$$X^TY = \begin{pmatrix} 1 & \dots & 1 \\ x_{11} & \dots & x_{n1} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \left(\sum\right)$$

elle a effacé bruh, pourtant j'ai pas trainé

ESTIMATEUR DE β A RETENIR PAR COEUR

$$\hat{\beta} = (X^TX)^{-1}X^TY.$$

Sa variance :

$$\begin{aligned} V(\hat{\beta}) = V((X^TX)^{-1}X^TY) &= ((X^TX)^{-1}X^T)V(Y)((X^TX)^{-1}X^T)^T \\ &= ((X^TX)^{-1}X^T)\sigma I_n(X(X^TX)^{-1}) \text{ car } V(AX) = AV(X)A^T \\ &= \sigma^2(X^TX)^{-1}(X^TX)(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1} \end{aligned}$$

3.1 Propriété de l'EMC

Théorème 3.1 (Gauss-)**.** *Point (ii) : On montre que $\forall\tilde{\beta}, \forall c \in \mathbb{R}^{p+1}, V(c\tilde{\beta} \leq V(c\tilde{\beta}))$ avec $\tilde{\beta}$ un autre estimateur linéaire sans biais de β*

— *$\hat{\beta}$ est un estimateur linéaire $\hat{\beta} = (X^TX)^{-1}X^TY \to \hat{\beta}$ est bien conbinaison linéaire de Y*
— *$\hat{\beta}$ est sans biais $E(\hat{\beta}) = E((X^TX)^{-1}X^TY) = (X^TX)^{-1}X^TE(Y) = (X^TX)^{-1}X^TX\beta = \beta$ car $Y = X\beta + E, E \sim \mathcal{N}(0, \sigma^2 I_n), E(Y) = E(X\beta) + E(E) = X\beta$*

Soit $\tilde{\beta}$ un autre estimateur linéaire sans biais de $\beta, \exists A : \tilde{\beta} = AY$ comme $\tilde{\beta}$ sans biais $AE(Y) = AX\beta = \beta \rightarrow AX = I_{p+1}$

$$\begin{aligned} V(\tilde{\beta}) &= V(\tilde{beta} - \hat{\beta} + \hat{\beta}) = V(\tilde{\beta} - \hat{\beta}) + V(\hat{\beta}) + 2Cov(\tilde{beta} - \hat{\beta}, \hat{\beta}) \\ Cov(\tilde{beta} - \hat{\beta}, \hat{\beta}) &= Cov(\tilde{\beta}, \hat{\beta}) - V(\hat{\beta}) = ACov(Y, Y)((X^TX)^{-1}X^T)^T - V(\hat{\beta}) \\ &= \sigma^2 AX(X^2X)^{-1} - V(\hat{\beta}) = 0 \end{aligned}$$

Soit $u \in \mathbb{R}^{p+1}$

$$\begin{aligned} V(u\tilde{\beta}) &= uV(\tilde{\beta})u^T \\ & u(V(\tilde{\beta} - \hat{\beta}) + V(\hat{\beta}))u^T \\ &= V(u(\tilde{\beta} - \hat{\beta})) + V(u\hat{\beta}) \end{aligned}$$

3.2 Etude des erreurs

démo SCT = SCR + SCM

$$\begin{aligned} \left\|Y - \hat{Y}1_n\right\|^2 = SCT &= (Y - \hat{Y})^T(Y - \hat{Y}1_n) \\ &= ((Y - \hat{Y}) + (\hat{Y} - \hat{Y}1_n))^T((Y - \hat{Y}) + (\hat{Y} - \hat{Y}1_n)) \\ &= (Y - \hat{Y})^T - (Y - \hat{Y}) + (\hat{Y} - \hat{Y}1_n)^T(\hat{Y} - \hat{Y}1_n) + 2(\hat{Y} - \hat{Y}1_n)^T(Y - \hat{Y}) \\ &= \left\|Y - \hat{Y}\right\|^2 + \left\|\hat{Y} - \hat{Y}1_n\right\|^2 + 2(X\hat{\beta} - \hat{Y}1_n)^2(Y - X\hat{\beta}) \\ &= SCR + SCM + 2(X\hat{\beta} - \hat{Y}1_n)^2(Y - X\hat{\beta}) \end{aligned}$$

Exemple 3.2. $H_x =$ hat matrix : $H_xY = \hat{Y} = X(X^TX)^{-1}X^TY$

1.

$$\begin{aligned} H_x * H_x &= X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T \\ &= X(X^TX)^{-1}I_{p+1}X^T = H_x \rightarrow \text{ projecteur} \\ H_x^T &= (X(X^TX)^{-1}X^T)^T = X(X^TX)^{-1}X^T \rightarrow \text{ orthogonal} \\ rg(H_x) &= J_n(H_x) \\ &= T_n(X(X^TX)^{-1}X^T) \\ &= T_n(X^TX(X^TX)^{-1}) = T_nI_{p+1} = p + 1 \end{aligned}$$

2. $H_xX = X(X^TX)^{-1}X^TX = X$
   Chaque vecteur colonne de X appartient à $Im(H_x)$. Or $\dim(Im(H_x)) = p+1$ et les $p+1$ vecteurs colonnes de X étant indépendants, ils forment une base de $Im(H_x)$. H_x est le projecteur orthogonal dans le sous espace vectoriel engendré par les colonnes de X, on le note $\mathcal{L}(X)$
   Ainsi $H_xY = \hat{Y}$ donc $\hat{Y}$ est le projeté orthogonal de Y dans l'espace engendré par les colonnes de X
3. Calcul matriciel

$$\begin{aligned} (Y - \hat{Y})^T(\hat{Y} - \bar{Y}1_n) &= ((1_n - H_x)Y)^T(H_x - J_n)Y \\ &= Y^T(1_n - H_x)^T(H_x - J_n)Y \end{aligned}$$

   Or soit $u \in \mathbb{R}^n$

$$J_nu = \ldots, J_nu \in Im(H_x) = \mathcal{L}(X)$$

   Car la première colonne de X est le vecteur $(1_n, 1_n \ldots, 1_n)$. Donc $(1_X - J_n)Y \in \mathcal{L}(X)$ et $(1_n - H_x)Y \in \mathcal{L}(X)$ (J'suis pas sur de bien lire ce qu'elle écrit entre les 1 et les J)