Introduction à la régression cours n°2 - Estimation

ENSM.SE – 1A
Olivier Roustant

Objectif du cours

- Considérons une variable dépendant linéairement d'une autre : $y \approx \beta_0 + \beta_1 x$

- Quelques méthodes d'estimation de β_0 et β_1:
  - Minimiser la somme des $(y_i-(\beta_0 + \beta_1 x_i))^2$ MOINDRES CARRES
  - Minimiser la somme des $|y_i-(\beta_0 + \beta_1 x_i)|$, des $(y_i-(\beta_0 + \beta_1 x_i))^4$ …

- *On utilise souvent les moindres carrés.*
  - *A-t-on toujours raison de le faire ?*
  - *Quand peut-on le faire ? Quel critère utiliser en général ?*

1ère partie. Estimation. Cadre.
Méthodes de construction d'estimateurs

- Cadre : voir poly p. 54
- Estimation / estimateur : voir poly p. 56
- Méthodes de construction d'estimateurs
  - Méthode des moments : cf. p. 59
  - Maximum de vraisemblance : cf. p. 60 et 61
- Propriétés (biais, variance, risque)
  - Sera vu en TD mardi 30 mai

Exemple

- On dispose d'un échantillon $x_1, \ldots, x_n$ d'une loi normale $N(\mu,\sigma^2)$. Estimation de μ et σ^2?
  - Estimations naturelles :
    - $\mu = E(X_i) \Rightarrow \hat{\mu} = \bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$
    - $\sigma^2 = \mathrm{var}(X_i) \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
  - Remarque : ce sont les estimateurs obtenues par la méthode des moments

Exemple (suite)

- Estimation par maximum de vraisemblance
  - On écrit la vraisemblance :

$$L(x_1,\dots,x_n;\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(x_i-\mu)^2}{2\sigma^2})$$

  - On cherche μ et σ de façon à maximiser L, ou de façon équivalente, à minimiser -2log(L)

$$-2\log(L)=n\ln(\sigma^2)+n\ln(2\pi)+\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Exemple (suite)

- On écrit les conditions du 1er ordre
  - Estimation de μ

$$\frac{\partial}{\partial \mu}(-2\log(L))\Big|_{(\mu^*,\sigma^{*2})} = 0 \iff \frac{\partial}{\partial \mu}\Big(\sum_{i=1}^{n}(x_i-\mu)^2\Big)\Big|_{(\mu^*,\sigma^{*2})} = 0$$

  - Deux remarques :
    - Revient à minimiser la somme des carrés $(x_i - \mu)^2$
    - L'estimation obtenue est l'estimateur usuel de la moyenne

$$\frac{\partial}{\partial \mu}(-2\log(L))\Big|_{(\mu^*,\sigma^{*2})} = 0 \iff \dots \iff \mu^* = \bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$$

Exemple (suite)

- Estimation de σ^2

$$\frac{\partial}{\partial \sigma^2}(-2\log(L))\Big|_{(\mu^*,\sigma^{*2})} = 0 \;\Leftrightarrow\; \ldots \;\Leftrightarrow\; (\sigma^*)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

- On retrouve l'estimateur « naturel » de la variance

➢ Remarque : moindres carrés $\Leftarrow$ loi normale

Exercice

- Considérons un échantillon $x_1, \ldots, x_n$ de la loi de Laplace, définie par sa densité

$$f_{\mu,\sigma}(t) = \frac{1}{2\sigma}\exp(-\frac{|t-\mu|}{\sigma})$$

- Vérifier que :
  - μ est à la fois l'espérance et la médiane de la loi de Laplace
  - L'estimateur de maximum de vraisemblance de μ s'obtient en minimisant la somme des valeurs absolues $| x_i - \mu |$
  - Il s'agit de la médiane des x_i

2ème partie
Application à la régression

- Considérons le modèle linéaire avec 1 prédicteur
  - $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, avec $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ i.i.d $N(0,\sigma^2)$
  - Notons $(y_{i,\, obs})_{1 \leq i \leq n}$ les résultats des expériences x_i

- Trois paramètres à estimer : β_0, β_1 et σ^2
- Estimation usuelle :
  - β_0, β_1 par moindres carrés
  - Ensuite, $\sigma^2 = E(\varepsilon_i^2) = E((y_i-(\beta_0+\beta_1 x_i)^2)$ estimé par

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_{i,obs} - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

Estimation par Maximum de Vraisemblance (EMV)

- Remarque : la v.a y_i s'obtient à partir de ε_i par translation de la quantité fixe $\beta_0 + \beta_1 x_i$

- Conséquences :
  - Les y_i sont aussi indépendantes
  - y_i est de loi $N(\beta_0 + \beta_1 x_i, \sigma^2)$
  - La vraisemblance des observations s'écrit :

$$L(y_{1,obs},\dots,y_{n,obs};\beta_0,\beta_1,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\{y_{i,obs} - (\beta_0 + \beta_1 x_i)\}^2}{2\sigma^2} \right)$$

EMV (suite)

- Minimisation de -2log(L)
  - Estimation de β_0 et β_1:

$$\frac{\partial}{\partial \beta_k}(-2\log(L))\Big|_{(\beta_0^*,\beta_1^*,\sigma^{*2})} = 0 \Leftrightarrow \frac{\partial}{\partial \beta_k}\Big(\sum_{i=1}^{n}\{y_{i,obs} - (\beta_0 + \beta_1 x_i)\}^2\Big)\Big|_{(\beta_0^*,\beta_1^*,\sigma^{*2})} = 0$$

  - On retrouve les moindres carrés
  - C'est dû à l'hypothèse que les ε_i sont de loi <u>normale</u>
    - Si les ε_i étaient supposés de loi de Laplace, on obtiendrait β_0 et β_1 en minimisant la somme des <u>valeurs absolues</u> des écarts

EMV (suite)

- Estimation de σ^2

$$\frac{\partial}{\partial \sigma^2}(-2\log(L))\Big|_{(\beta_0^*,\beta_1^*,\sigma^{*2})} = 0 \Leftrightarrow \dots \Leftrightarrow (\sigma^*)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_{i,obs} - (\beta_0^* + \beta_1^* x_i))^2$$

- On retrouve l'estimateur usuel

Retour sur les objectifs

- Considérons le modèle linéaire avec 1 prédicteur
  - $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, avec $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ i.i.d
- Alors le critère à utiliser pour estimer la droite de régression dépend des hypothèses sur la loi des ε_i
  - ε_i de loi $N(0,\sigma^2)$ $\Rightarrow$ EMV = Moindres carrés
  - ε_i de loi de Laplace $\Rightarrow$ EMV = Moindres valeurs absolues
  - Etc.

Exercice

- Vous pouvez réaliser *n* expériences pour estimer un phénomène linéaire sur [a,b] impliquant 1 prédicteur
  - Comment répartir les expériences dans le domaine expérimental [a,b] de façon à ce que l'estimation soit la plus précise possible ?