

---

**Sujet 13: Etude de l'estimateur des paramètres et des résidus en régression non-linéaire, comparaison dans le cas des modèles linéairement transformables**

---

De manière générale, le but de tout modèle de régression est de spécifier une relation (de nature stochastique) entre une variable  $Y$  appelée variable à expliquer et un certain nombre de variables explicatives  $X^{(1)}, \dots, X^{(p)}$ . Dans toute la suite, on supposera que les variables explicatives sont de loi continue.

• **Quelques rappels sur la régression linéaire:**

Notons  $(Y_i, X_i^{(1)}, \dots, X_i^{(p)})$  pour  $i = 1, \dots, n$ , les observations correspondant à  $n$  individus que l'on suppose distribuées comme  $(Y, X^{(1)}, \dots, X^{(p)})$ . Le modèle de régression linéaire gaussien standard s'écrit sous la forme générique suivante:

$$Y_i = \beta_0 + \sum_{k=1}^p \beta_k X_i^{(k)} + \varepsilon_i, \quad \text{où } \varepsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2). \quad (1)$$

Le modèle de régression linéaire standard repose ainsi sur les hypothèses suivantes:

**(H<sub>1</sub>) linéarité:** l'espérance conditionnelle de la variable réponse vaut

$$\mathbb{E}[Y_i | X_i^{(1)}, \dots, X_i^{(p)}] = \beta_0 + \sum_{k=1}^p \beta_k X_i^{(k)}.$$

Attention, bien que l'espérance conditionnelle de la variable réponse soit une combinaison affine des covariables, la linéarité s'apprécie relativement à la linéarité de  $\mathbb{E}[Y_i | X_i^{(1)}, \dots, X_i^{(p)}]$  en les paramètres.

On supposera toujours que le modèle est identifiable, à savoir

$$\beta := (\beta_0, \dots, \beta_p)^t = (\beta'_0, \dots, \beta'_p)^t := \beta' \implies \mathbb{E}_\beta[Y_i | X_i^{(1)}, \dots, X_i^{(p)}] = \mathbb{E}_{\beta'}[Y_i | X_i^{(1)}, \dots, X_i^{(p)}].$$

**(H<sub>2</sub>) centrage des erreurs:**  $\mathbb{E}[\varepsilon_i] = 0$  pour  $i = 1, \dots, n$ .

**(H<sub>3</sub>) exogénéité = non-endogénéité:**  $(X_i^{(1)}, \dots, X_i^{(p)})$  et  $\varepsilon_i$  sont indépendants pour  $i = 1, \dots, n$ .

**(H<sub>4</sub>) non-colinéarité des covariables:** les covariables sont non-corrélées entre elles, ce qui se traduit en pratique par le fait que les vecteurs  $(X_1^{(k)}, \dots, X_n^{(k)})$  sont non-colinéaires pour  $k = 1, \dots, p$ . On écartera également le cas trivial où l'un de ces vecteurs serait colinéaire au vecteur de longueur  $n$  dont toutes les composantes sont égales à 1.

**(H<sub>5</sub>) non-corrélation:** les vecteurs  $(Y_i, X_i^{(1)}, \dots, X_i^{(p)})$  sont non-corrélés pour  $i = 1, \dots, n$ .

**(H<sub>6</sub>) homoscedasticité:** les lois conditionnelles de  $Y_i$  sachant  $X_i^{(1)}, \dots, X_i^{(p)}$  ont même variance donc on a  $\sigma_i^2 = \sigma^2$  pour  $i = 1, \dots, n$ .

Le modèle de régression linéaire standard est gaussien lorsqu'on effectue l'hypothèse supplémentaire suivante:

**(H<sub>7</sub>) normalité:** conditionnellement à  $X_i^{(1)}, \dots, X_i^{(p)}$ , la variable de réponse  $Y_i$  suit la loi normale.

Notons  $I_n$ =matrice identité de taille  $n \times n$  et avec

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbb{X} = \begin{pmatrix} 1 & X_1^{(1)} & \dots & X_1^{(p)} \\ \vdots & \vdots & & \vdots \\ 1 & X_n^{(1)} & \dots & X_n^{(p)} \end{pmatrix} = \begin{pmatrix} \mathbb{X}_1 \\ \vdots \\ \mathbb{X}_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Le modèle de régression linéaire homoscédastique se réécrit matriciellement sous la forme:

$$\mathbb{Y} = \mathbb{X}.\beta + \varepsilon, \quad \text{où } \varepsilon \sim (0_n, \sigma^2 I_n) \text{ et } \varepsilon \perp \mathbb{X}.$$

Le modèle de régression linéaire gaussien homoscédastique s'obtient sous forme matricielle lorsque l'on ajoute l'hypothèse  $\varepsilon \sim \mathcal{N}_n(0_n, \sigma^2 I_n)$ .

• **Estimation de l'effet des variables explicatives:**

Dans le cadre du modèle linéaire gaussien standard présenté ci-dessus, on se demande si les variables explicatives introduites dans le modèle ont un effet ou non sur la variable à expliquer. Cet effet est quantifié par le coefficient de régression correspondant.

L'estimateur des moindres carrés ordinaires de  $\beta$  est  $\hat{\beta} = (\mathbb{X}^t.\mathbb{X})^{-1}.\mathbb{X}^t.\mathbb{Y}$ . Sous les hypothèses  $(H_1) - (H_4)$ , cet estimateur est sans biais

$$\mathbb{E}[\hat{\beta}|\mathbb{X}] = \beta.$$

Sous les hypothèses  $(H_1) - (H_6)$ , la variance de  $\hat{\beta}$  est

$$\text{Var}(\hat{\beta}|\mathbb{X}) = \sigma^2(\mathbb{X}^t.\mathbb{X})^{-1}.$$

Notons  $\hat{\sigma}^2$  l'estimateur de la variance de  $\sigma^2$  défini par:

$$\hat{\sigma}^2 = \frac{1}{n - (p + 1)} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{\hat{\varepsilon}^t.\hat{\varepsilon}}{n - (p + 1)} \quad (2)$$

en notant

$$\hat{\varepsilon}_i = Y_i - \mathbb{X}_i.\hat{\beta}$$

et

$$\hat{\varepsilon} = \mathbb{Y} - \mathbb{X}.\hat{\beta}.$$

Sous les hypothèses  $(H_1)-(H_6)$ , l'estimateur  $\hat{\sigma}^2$  est sans biais pour  $\sigma^2$ .

Introduisons une hypothèse supplémentaire:

**(H<sub>8</sub>):**  $\frac{\mathbb{X}^t.\mathbb{X}}{n} \xrightarrow{\mathbb{P}} Q$  lorsque  $n \rightarrow \infty$  où  $Q$  est une matrice symétrique définie positive.

Sous les hypothèses  $(H_1)-(H_6)$  et  $(H_8)$ , la convergence  $\hat{\beta} \xrightarrow{\mathbb{P}} \beta$  a lieu lorsque  $n \rightarrow \infty$ .

Sous les hypothèses  $(H_1) - (H_7)$ , la loi de l'estimateur  $\widehat{\beta}$  de  $\beta$  est

$$\widehat{\beta}_{EMV} \sim \mathcal{N}_{p+1}(\beta, \sigma^2(\mathbb{X}^t \cdot \mathbb{X})^{-1}),$$

et, pour  $j = 0, 1, \dots, p$ , on a

$$\frac{(\widehat{\beta}^{(j)} - \beta^{(j)})}{\sqrt{\widehat{\sigma}^2[(\mathbb{X}^t \cdot \mathbb{X})^{-1}]_{j,j}}} \sim T(n - (p + 1)).$$

• **Résidus standardisés:**

Considérons les résidus (bruts) que l'on a défini pour  $i = 1, \dots, n$  par  $\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i$ . Lorsque la valeur  $|\widehat{\varepsilon}_i|$  est "anormalement" élevée, cela indique que le  $i^{\text{ème}}$  réponse a été mal reconstituée par le modèle. Afin de pouvoir trancher si une observation est "anormalement" élevée, il faut être en mesure de pouvoir comparer des choses comparables. Or, bien que l'on ait travaillé sous l'hypothèse d'homoscédasticité qui stipule que  $\text{Var}(\varepsilon_i) = \sigma^2$  pour  $i = 1, \dots, n$ , il n'en va pas de même pour les résidus bruts. On a en effet établi que  $\text{Var}(\widehat{\varepsilon}_i) = \sigma^2(1 - h_i)$  pour  $i = 1, \dots, n$ . Pour rendre les amplitudes comparables entre les différents résidus, ie afin d'obtenir des résidus de variances égales, on normalise chacun des résidus (bruts) par son écart-type estimé pour ne plus s'intéresser qu'aux résidus normalisés. Les résidus standardisés sont des résidus normalisés définis de la façon suivante:

$$\widehat{\varepsilon}_i^{\text{Stand}} = \frac{\widehat{\varepsilon}_i}{\sqrt{\widehat{\sigma}^2(1 - h_i)}}$$

où  $\widehat{\sigma}^2$  est l'estimateur sans biais de  $\sigma^2$  défini en (2).

On peut montrer que, pour  $n$  assez grand, la distribution des résidus standardisés est approximativement la loi gaussienne centrée réduite. Lorsque le modèle linéaire gaussien s'ajuste bien aux données, pour  $n$  assez grand, on s'attend donc à ce qu'à peu près 95% des résidus standardisés se trouvent entre les quantiles d'ordre respectif 2.5% et 97.5% de la loi  $\mathcal{N}(0, 1)$ .

• **Régression non-linéaire:**

Notons  $(Y_i, X_i^{(1)}, \dots, X_i^{(p)})$  pour  $i = 1, \dots, n$ , les observations correspondant à  $n$  individus que l'on suppose distribuées comme  $(Y, X^{(1)}, \dots, X^{(p)})$ . Notons  $\mathbb{X}_i = (X_i^{(1)}, \dots, X_i^{(p)})$ . Soit  $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$  un vecteur de paramètres de dimension  $p$ . Le modèle de régression non-linéaire gaussien

homoscédastique s'écrit sous la forme générique suivante:

$$Y_i = f(\mathbb{X}_i; \beta) + \varepsilon_i, \quad \text{où } \varepsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2). \quad (3)$$

Le modèle de régression non-linéaire homoscédastique repose ainsi sur les hypothèses suivantes:

**(H<sub>1</sub>) non-linéarité:** l'espérance conditionnelle de la variable réponse vaut

$$\mathbb{E}[Y_i | X_i^{(1)}, \dots, X_i^{(p)}] = f(\mathbb{X}_i; \beta)$$

où  $\beta \rightarrow f(x; \beta)$  est  $c^2$  pour tout  $x$  de  $\mathbb{R}^q$  et où au moins d'une des dérivées partielles de  $\beta \rightarrow f(x; \beta)$  par rapport à l'une des composantes de  $\beta$  dépend au moins d'une des composantes de  $\beta$ . On supposera toujours que le modèle est identifiable, à savoir

$$\beta := (\beta_1, \dots, \beta_p)^t = (\beta'_1, \dots, \beta'_p)^t := \beta' \implies \mathbb{E}_\beta[Y_i | X_i^{(1)}, \dots, X_i^{(p)}] = \mathbb{E}_{\beta'}[Y_i | X_i^{(1)}, \dots, X_i^{(p)}].$$

**(H<sub>2</sub>) centrage des erreurs:**  $\mathbb{E}[\varepsilon_i] = 0$  pour  $i = 1, \dots, n$ .

**(H<sub>3</sub>) exogénéité = non-endogénéité:**  $(X_i^{(1)}, \dots, X_i^{(p)})$  et  $\varepsilon_i$  sont indépendants pour  $i = 1, \dots, n$ .

**(H<sub>4</sub>) non-corrélation:** les vecteurs  $(Y_i, X_i^{(1)}, \dots, X_i^{(p)})$  sont non-corrélés pour  $i = 1, \dots, n$ .

**(H<sub>5</sub>) homoscedasticité:** les lois conditionnelles de  $Y_i$  sachant  $X_i^{(1)}, \dots, X_i^{(p)}$  ont même variance donc on a  $\sigma_i^2 = \sigma^2$  pour  $i = 1, \dots, n$ .

Le modèle de régression linéaire standard est gaussien lorsqu'on effectue l'hypothèse supplémentaire suivante:

**(H<sub>6</sub>) normalité:** conditionnellement à  $X_i^{(1)}, \dots, X_i^{(p)}$ , la variable de réponse  $Y_i$  suit la loi normale.

Notons  $V$  la matrice de taille  $n \times p$  des dérivées partielles suivante:

$$V(\beta) = \left[ \frac{\partial f(\mathbb{X}_i; \beta)}{\partial \beta_j} \right]_{1 \leq i \leq n, 1 \leq j \leq p}$$

et, pour  $i = 1, \dots, n$ , notons  $W_i$  la matrice de taille  $p \times p$  des dérivées secondes suivante:

$$W_i(\beta) = \left[ \frac{\partial^2 f(\mathbb{X}_i; \beta)}{\partial \beta_j \partial \beta_k} \right]_{1 \leq j, k \leq p}.$$

L'estimateur des moindres carrés ordinaires de  $\beta$  est donné par

$$\hat{\beta} \in \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (Y_i - f(\mathbb{X}_i; \beta))^2.$$

Ce problème d'optimisation n'est pas résoluble analytiquement. Un algorithme itératif permet de fournir une estimation numérique. A distance finie, l'estimateur  $\hat{\beta}$  de  $\beta$  est biaisé. Sous  $(H_1)$ - $(H_5)$  ainsi que sous quelques conditions techniques assurant le "bon comportement" de la matrice  $(V(\beta)^t \cdot V(\beta))^{-1}$ , lorsque  $n$  tend vers  $\infty$ , l'estimateur  $\hat{\beta}$  est consistant et la loi de  $\hat{\beta}$  s'approche de la loi  $\mathcal{N}_p(\beta, \sigma^2(V(\beta)^t \cdot V(\beta))^{-1})$ . Un estimateur consistant de  $\sigma^2$  est

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n \left( Y_i - f(\mathbb{X}_i; \hat{\beta}) \right)^2.$$

Considérons les résidus (bruts) que l'on a défini pour  $i = 1, \dots, n$  par  $\hat{\varepsilon}_i = Y_i - f(\mathbb{X}_i; \hat{\beta})$ . On dit que ces résidus sont biaisés au sens où

$$\mathbb{E}[\hat{\varepsilon}_i | \mathbb{X}_i] \neq 0.$$

On peut montrer que

$$\mathbb{E}[\hat{\varepsilon}_i | \mathbb{X}_i] \approx (I_p - H(\beta)) \cdot U(\beta)$$

en notant  $H(\beta) = V(\beta) \cdot (V(\beta)^t \cdot V(\beta))^{-1} \cdot V(\beta)$  et  $U(\beta)$  le vecteur-colonne de longueur  $n$  dont la  $i^{\text{ème}}$  composante est égale à  $-\frac{1}{2} \sigma^2 \text{Tr}((V(\beta)^t \cdot V(\beta))^{-1} \cdot W_i(\beta))$ . Une normalisation fréquente de ces résidus consiste à considérer les résidus standardisés suivants:

$$\frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}^2}}$$

Une autre possibilité est de considérer les résidus standardisés suivants:

$$\frac{\widehat{\varepsilon}_i}{\sqrt{\widehat{\sigma}^2(1 - H_{ii}(\widehat{\beta}))}}$$

• **Implémentation au moyen du logiciel R:**

Le logiciel R permet d'ajuster simplement un modèle de régression à des données. Pour ajuster un modèle linéaire sur un échantillon  $(Y_i, X_i^{(1)}, \dots, X_i^{(p)})_{i=1, \dots, n}$ , on stockera les  $n$  valeurs des  $Y_i$  dans un vecteur  $\mathbf{y}$ , puis, pour  $k = 1, \dots, p$ , on stockera les  $n$  valeurs des  $X_i^{(k)}$  dans un vecteur  $\mathbf{xk}$ . Le modèle linéaire est alors ajusté au moyen de l'instruction suivante (où  $p = 3$ )

```
mylm <- lm(y ~ x1 + x2 + x3)
```

et le résultat est stocké dans l'objet `mylm`. L'instruction suivante permet d'accéder à l'ensemble des quantités calculées:

```
summary(mylm)
```

Sont notamment calculées, les estimations des  $\beta_j$  pour  $j = 0, \dots, p$  et les estimations de leurs écarts-types. L'objet `summary(mylm)` est une liste dont on peut récupérer chacune des composantes qui nous intéresse. Pour cela, l'instruction `str(summary(mylm))` permet de voir le nom et la structure des différentes composantes de cette liste.

La fonction `nls` permet d'ajuster un modèle de régression non-linéaire. Le package `nlstools` contient également une fonction qui permet d'ajuster un modèle de régression non-linéaire ainsi que certaines fonctions de diagnostics.

**Exercice 1.**

Illustrer de manière empirique à partir de données simulées le comportement de l'estimateur des coefficients de régression, en termes de biais, variance, écart quadratique moyen, consistance et distribution de l'estimateur  $\widehat{\beta}$ , ainsi que le comportement des résidus

1. en ajustant un modèle de régression non-linéaire gaussien,
2. en ajustant la version linéairement transformable gaussienne,

sur des données simulées avec l'un puis l'autre des modèles. Vous veillerez à faire varier également:

- la taille de l'échantillon  $n$  simulé,
- la variance de l'erreur résiduelle simulée.

Voici quelques exemples de modèles non-linéaires et leur version linéairement transformable:

modèle non-linéaire	version linéarisée
$f(x; \beta) = \frac{1}{\beta_1 + \beta_2 \exp(-x)}$	$\frac{1}{f(x; \beta)} = \beta_1 + \beta_2 \exp(-x)$
$f(x; \beta) = \frac{\beta_1 x}{\beta_2 + x}$	$\frac{1}{f(x; \beta)} = \frac{1}{\beta_1} + \frac{\beta_2}{\beta_1 x}$
$f(x; \beta) = \beta_1 x^{\beta_2}$	$\log(f(x; \beta)) = \log(\beta_1) + \beta_2 \log(x)$
$f(x; \beta) = \exp(-\beta_1 x_1 \exp(-\beta_2 x_2))$	$\log \log(f(x; \beta)) = \log(-\beta_1) + \log(x) - \frac{\beta_2}{x_2}$
$f(x; \beta) = \beta_1 (x_1)^{\beta_2} (x_2)^{\beta_3} (x_3)^{\beta_4}$	$\log(f(x; \beta)) = \log(\beta_1) + \beta_2 \log(x_1) + \beta_3 \log(x_2) + \beta_4 \log(x_3)$