

IX - Méthode des Moindres Carrés

Principe

- Soit $y = f(x / \underline{\theta})$ une relation fonctionnelle entre les variables y et x à k paramètres $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ inconnus,
- Soient $\underline{\theta}_0$ les vraies valeurs de $\underline{\theta}$ à estimer
- Soit y_i la mesure de y en x_i
- La f.d.p. de y_i est une $N(f(x_i / \underline{\theta}_0), \sigma_i^2)$
où l'écart type σ_i (erreur de mesure) est connu
- Soit une série de $n > k$ mesures $\underline{y} = (y_1, y_2, \dots, y_n)$
effectuées en $\underline{x} = (x_1, x_2, \dots, x_n)$

Estimation $\hat{\underline{\theta}}$ de $\underline{\theta}$: valeurs de $\underline{\theta}$ qui minimise

$$X^2(\underline{\theta}) = \sum_{i=1}^n \frac{(y_i - f(x_i / \underline{\theta}))^2}{\sigma_i^2} \quad \text{étant donnés } \underline{y}$$

$$X^2(\underline{\theta}) = \sum_{i=1}^n \frac{(y_i - f(x_i / \underline{\theta}))^2}{\sigma_i^2} \quad \text{étant donnés } \underline{y} \text{ et } \underline{\sigma}$$

Somme des carrés des écarts entre le point mesuré \underline{y} et le point espéré $f(\underline{x} / \underline{\theta})$ à $\underline{\theta}$ donné dans l'espace à n dimensions en utilisant $\underline{\sigma}$ comme unité de mesure.

Ecart minimum $\Rightarrow \hat{\underline{\theta}}$

Exemple

Mouvement uniformément accéléré

$$x = x_0 + v_0 t + \frac{a t^2}{2}$$

mesures x_i effectuées à $n > 3$ instants t_i avec une précision σ_i

estimer x_0, v_0 et a et leurs erreurs

Exemple : Mise en histogramme (ou mise en classes)

$$X^2(\underline{\theta}) = \sum_{i=1}^N \frac{(n_i - np_i(\underline{\theta}))^2}{np_i(\underline{\theta})}$$

n_i = nombre observé d'événements dans la boîte $i = [X_i \leq x \leq X_{i+1}]$

$$n = \sum_{i=1}^N n_i$$

$$p_i(\underline{\theta}) = \int_{X_i}^{X_{i+1}} f(x / \underline{\theta}) dx$$

Choix des classes

Contraintes contradictoires:

binomiale \rightarrow Poisson $\Rightarrow p_i$ petits \Rightarrow beaucoup de classes

Poisson \rightarrow gaussienne $\Rightarrow n_i$ grands \Rightarrow beaucoup de mesures par classe

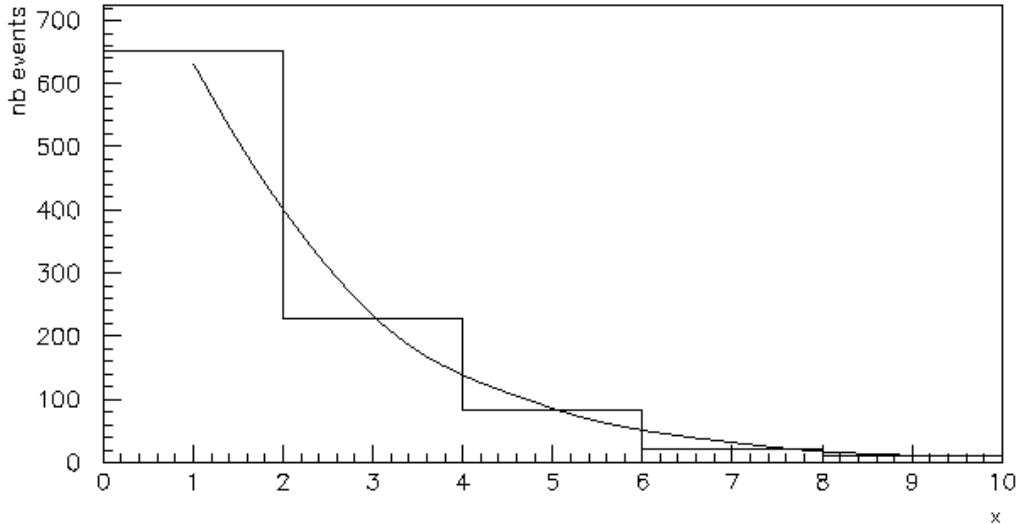
beaucoup de mesures par classe \Rightarrow perte d'information

Deux méthodes:

classes de largeurs égales : arithmétique est plus simple

classes de contenus égaux: n_i grands (20-25) sans perte d'information

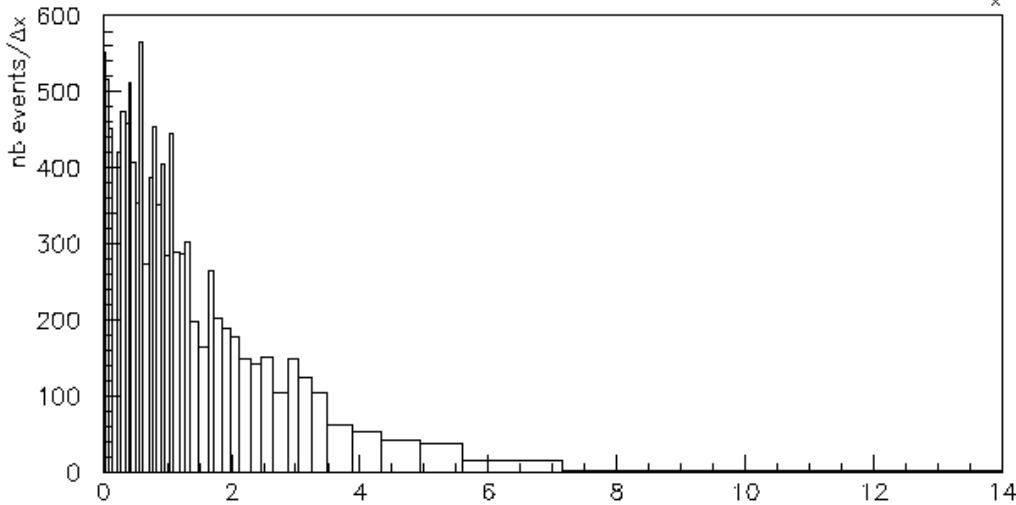
inutile



Classes de largeurs égales

5 classes telles que $n_i \geq 25 \quad \forall i = 1,5$

$n_1 = 650 \Rightarrow$ perte importante d'information



Classes de contenus égaux

40 classes telles que $n_i \approx 25 \quad \forall i = 1,5$

perte d'information minimale

Équivalence entre Moindres Carrés et Maximum de Vraisemblance à l'approximation gaussienne des grands échantillons

Approximation gaussienne:

$$\mathcal{L}(\underline{\theta}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2} \frac{(y_i - f(x_i/\underline{\theta}))^2}{\sigma_i^2}}$$

$$\log \mathcal{L}(\underline{\theta}) = -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - f(x_i/\underline{\theta}))^2}{\sigma_i^2} + Cste$$

$$\log \mathcal{L}^*(\underline{\theta}) = -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - f(x_i/\underline{\theta}))^2}{\sigma_i^2}$$

$$X^2(\underline{\theta}) = \sum_{i=1}^n \frac{(y_i - f(x_i/\underline{\theta}))^2}{\sigma_i^2}$$

$$-2 \log \mathcal{L}^*(\underline{\theta}) = X^2(\underline{\theta})$$

Méthode asymptotiquement efficace et consistante

Détermination d'un domaine de confiance au niveau de confiance α

$$-2 \log \mathcal{L}^*(\underline{\theta}) = X^2(\underline{\theta})$$

Maximum de vraisemblance:

intersection de $\log \mathcal{L}(\underline{\theta})$ avec hyperplan parallèle à $(\underline{\theta})$ à $\frac{r_\alpha^2}{2}$ en dessous de $\text{Max}(\log \mathcal{L}(\underline{\theta}))$

Moindres carrés:

intersection de $X^2(\underline{\theta})$ avec hyperplan parallèle à $(\underline{\theta})$ à r_α^2 au dessus de $\text{Min}(X^2(\underline{\theta}))$

$$r_\alpha \text{ tel que } \int_0^{r_\alpha^2} \chi_k^2(r^2) dr^2 = \alpha \text{ ou}$$

$$r_\alpha \text{ tel que } \int_{-r_\alpha}^{r_\alpha} N(0,1) dx = \alpha \text{ si } k = 1$$

Résolution analytique du modèle linéaire

- Soit $y = \sum_{l=1}^L a_l(x) \theta_l$ une relation fonctionnelle entre les variables y et x dépendant linéairement de L paramètres $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_L)$ inconnus,
- Soient $\underline{\theta}_0$ les vraies valeurs de $\underline{\theta}$ à estimer
- Soit N mesures $y_n \pm \sigma_n$ indépendantes de y en x_n , $n = 1, N$

$$X^2(\underline{\theta}) = \sum_{n=1}^N \frac{\left(y_n - \sum_{l=1}^L a_l(x_n) \theta_l \right)^2}{\sigma_n^2} = \sum_{n=1}^N \frac{\left(y_n - \sum_{l=1}^L a_{nl} \theta_l \right)^2}{\sigma_n^2} \quad \text{avec } a_{nl} = a_l(x_n)$$

Exemples

- polynôme de degré $L - 1$

$$X^2(\underline{\theta}) = \sum_{n=1}^N \frac{\left(y_n - \sum_{l=1}^L x_n^{l-1} \theta_l \right)^2}{\sigma_n^2} \quad \text{avec } a_l(x_n) = a_{nl} = x_n^{l-1}$$

- développement en série de Fourier

représentation d'une fonction périodique de fréquence f dans une base de fonction exponentielles complexes orthogonales

$$X^2(\underline{\theta}) = \sum_{n=1}^N \frac{\left(y_n - \sum_{l=-\infty}^{\infty} e^{i2\pi l f x_n} \theta_l \right)^2}{\sigma_n^2} \quad \text{avec } a_l(x_n) = a_{nl} = e^{i2\pi l f x_n}$$

Recherche du minimum

$$X^2(\underline{\theta}) = \sum_{n=1}^N \frac{\left(y_n - \sum_{l=1}^L a_{nl} \theta_l \right)^2}{\sigma_n^2} \quad \text{avec } a_{nl} = a_l(x_n)$$

$$X^2(\underline{\theta}) = (\underline{y} - A\underline{\theta})^T V^{-1} (\underline{y} - A\underline{\theta})$$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1L} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NL} \end{pmatrix} \quad \text{et} \quad V = \begin{pmatrix} \sigma_1^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_N^2 \end{pmatrix}$$

$$\frac{\partial X^2(\underline{\theta})}{\partial \underline{\theta}} = -2A^T V^{-1} (\underline{y} - A\underline{\theta}) = -2A^T V^{-1} \underline{y} + 2A^T V^{-1} A \underline{\theta} = \mathbf{0}$$

$$\hat{\underline{\theta}} = (A^T V^{-1} A)^{-1} A^T V^{-1} \underline{y}$$

$$\hat{\underline{\theta}} = C^{-1} \underline{c} \quad C = A^T V^{-1} A$$

$$\underline{c} = A^T V^{-1} \underline{y}$$

Calcul d'erreur

$$\hat{\underline{\theta}} = \left(A^T V^{-1} A \right)^{-1} A^T V^{-1} \underline{y} = f(\underline{y})$$

$$\begin{aligned} V(\hat{\underline{\theta}}) &= \left(\frac{\partial \hat{\underline{\theta}}}{\partial \underline{y}} \right) V \left(\frac{\partial \hat{\underline{\theta}}}{\partial \underline{y}} \right)^T \\ &= \left(\left(A^T V^{-1} A \right)^{-1} A^T V^{-1} \right) V \left(\left(A^T V^{-1} A \right)^{-1} A^T V^{-1} \right)^T \end{aligned}$$

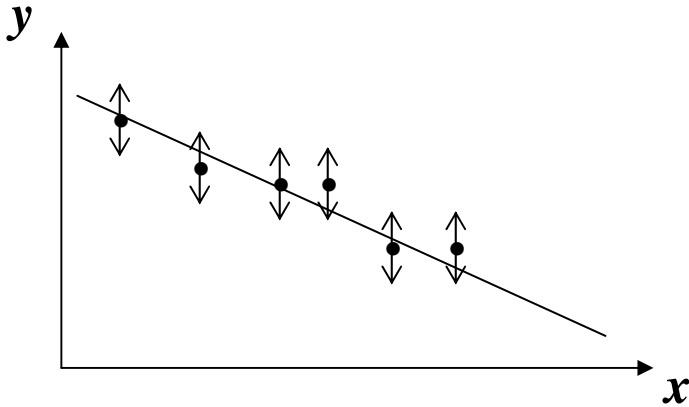
$$V(\hat{\underline{\theta}}) = \left(A^T V^{-1} A \right)^{-1} = C^{-1}$$

Généralisation: mesures corrélées

Méthode valable si corrélations entre les y_n

$$V = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_N^2 \end{pmatrix}$$

Exemple: ajustement d'une droite ($L=2$) par N points



$$X^2(a, b) = \sum_{n=1}^N \frac{(y_n - ax_n - b)^2}{\sigma_n^2}$$

$$\theta_1 = b, \theta_2 = a$$

$$\frac{\partial X^2}{\partial a} = -2 \sum_{n=1}^N \frac{x_n}{\sigma_n^2} (y_n - ax_n - b) = 0$$

$$\frac{\partial X^2}{\partial b} = -2 \sum_{n=1}^N \frac{1}{\sigma_n^2} (y_n - ax_n - b) = 0$$

Soit $\alpha = \sum_{n=1}^N \frac{x_n}{\sigma_n^2}$

$$\beta = \sum_{n=1}^N \frac{1}{\sigma_n^2}$$

$$\gamma = \sum_{n=1}^N \frac{y_n}{\sigma_n^2}$$

$$\delta = \sum_{n=1}^N \frac{x_n^2}{\sigma_n^2}$$

$$\varepsilon = \sum_{n=1}^N \frac{x_n y_n}{\sigma_n^2}$$

$$D = \delta\beta - \alpha^2$$

$$a = \frac{(\varepsilon\beta - \gamma\alpha)}{D}$$

$$\sigma_a^2 = \frac{\beta}{D}$$

$$b = \frac{(\gamma\delta - \varepsilon\alpha)}{D}$$

$$\sigma_b^2 = \frac{\delta}{D}$$

$$\sigma_{ab} = -\frac{\alpha}{D}$$

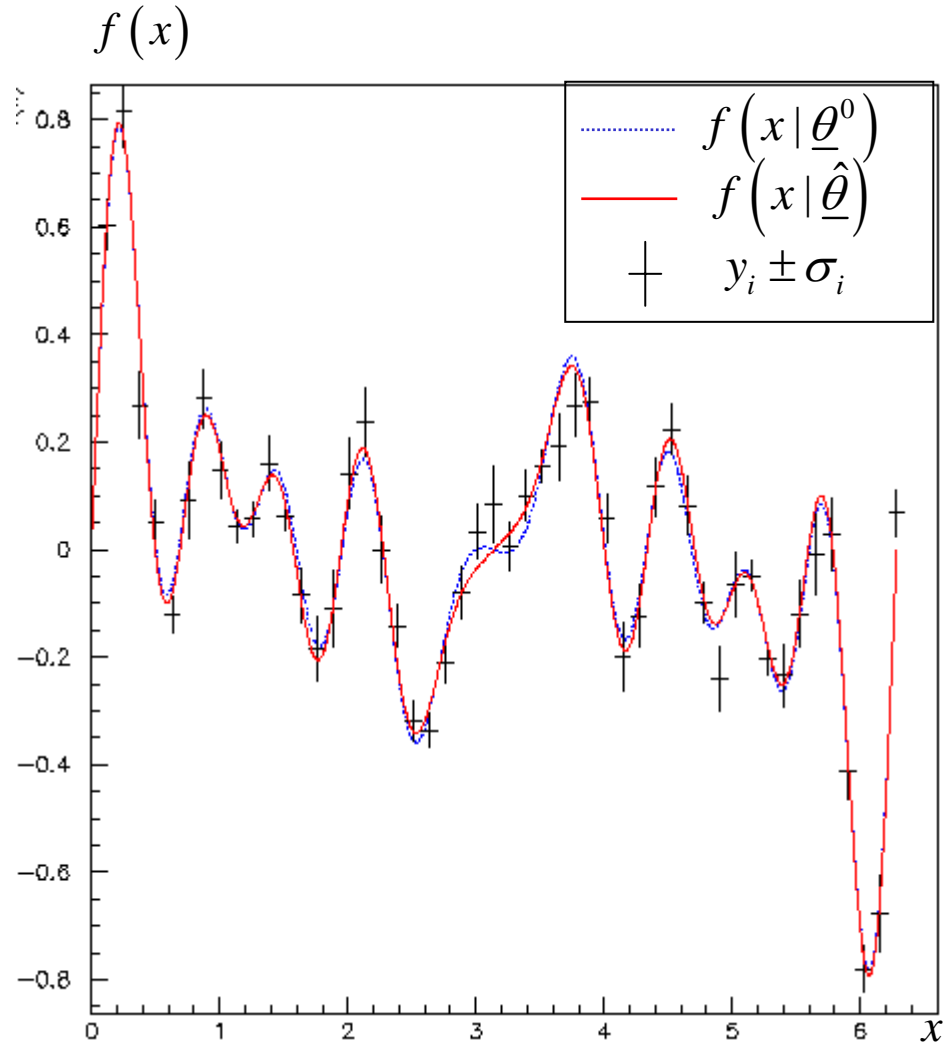
Exemple: fonction oscillante superposition de 10 sinusoides de fréquences connues

- $f(x | \underline{\theta}^0) = \sum_{l=1}^{10} \sin(\omega_l x) \theta_l^0 \quad L = 10$

- $y_i \pm \sigma_i, i = 1, N = 50$:

mesures de $f(x)$ en 50 points

ω	$\underline{\theta}^0$	$\hat{\underline{\theta}}$	$\sigma(\hat{\underline{\theta}})$
1.0	0.040	0.034	0.010
2.0	0.182	0.175	0.010
3.0	0.026	0.028	0.010
4.0	0.123	0.126	0.009
5.0	0.041	0.028	0.010
6.0	0.116	0.130	0.010
7.0	0.174	0.178	0.009
8.0	0.032	0.037	0.010
9.0	0.158	0.149	0.009
10.0	0.107	0.116	0.010



Généralisation: le modèle non linéaire avec contraintes

Système à $M = N + L$ variables, dont

- **N variables mesurables** $\underline{\eta} = (\eta_1, \eta_2, \dots, \eta_N)$
estimations (mesures) $\underline{\hat{\eta}} = (\hat{\eta}_1, \hat{\eta}_2, \dots, \hat{\eta}_N)$ et $V_{\hat{\eta}}$ la matrice de covariance
- **L variables non mesurables** $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_L)$
- **$K > L$ équations de contrainte entre les $M = N + L$ variables** $f(\underline{\eta}, \underline{\theta}) = 0$

Utiliser les mesures et leurs erreurs et les contraintes pour:

- **réestimer les N variables mesurables** : $\underline{\hat{\eta}}$ plus précises que $\underline{\hat{\eta}}$
- **estimer les L variables non-mesurables** $\underline{\hat{\theta}}$
- **calculer la matrice des variances et covariances sur les $M=N+L$ variables**

Exemple

Analyse cinématique de la réaction I corps $\rightarrow F$ corps

$I = 1$: désintégration

$I = 2$: interaction

Nombre total de variables cinématiques: $M = (I + F) \times 4$

quadri-vecteurs $(\underline{p}_j, E_j), j = 1, M$

ou $(p_j, \theta_j, \phi_j, E_j), j = 1, M$

$K = 4$ équations de contrainte **non linéaires**: conservation de l'énergie-impulsion

$$\sum_{i=1}^I p_i \sin \theta_i \cos \phi_i - \sum_{f=1}^F p_f \sin \theta_f \cos \phi_f = 0$$

$$\sum_{i=1}^I p_i \sin \theta_i \sin \phi_i - \sum_{f=1}^F p_f \sin \theta_f \sin \phi_f = 0$$

$$\sum_{i=1}^I p_i \cos \theta_i - \sum_{f=1}^F p_f \cos \theta_f = 0$$

$$\sum_{i=1}^I E_i - \sum_{f=1}^F E_f = 0$$

Si le nombre de variables non mesurables

$L > K$: système indéterminé

$L = K$: système soluble

$L < K$: système ajustable par la méthode des moindres carrés

Méthode des multiplicateurs de Lagrange

Les nouvelles estimations $\underline{\hat{\eta}}$ de $\underline{\eta}$ minimisent $X^2 = (\underline{\hat{\eta}} - \underline{\eta})^T V_{\hat{\eta}}^{-1} (\underline{\hat{\eta}} - \underline{\eta})$ } (1)
 tout en vérifiant le système de K équations $\underline{f}(\underline{\eta}, \underline{\theta}) = \mathbf{0}$

Ce système est remplacé par la minimisation de la combinaison linéaire

$$X'^2 = (\underline{\hat{\eta}} - \underline{\eta})^T V_{\hat{\eta}}^{-1} (\underline{\hat{\eta}} - \underline{\eta}) + 2\underline{\lambda}^T \underline{f}(\underline{\eta}, \underline{\theta}) \quad (2)$$

avec K nouveaux paramètres inconnus: multiplicateurs de Lagrange $\underline{\lambda}$

Minimisation de X'^2 :

$$\left. \begin{aligned} \frac{\partial X'^2}{\partial \underline{\eta}} &= -2V_{\hat{\eta}}^{-1} (\underline{\hat{\eta}} - \underline{\eta}) + 2f_{\eta}^T \underline{\lambda} = \mathbf{0} & (3) \\ \frac{\partial X'^2}{\partial \underline{\theta}} &= 2f_{\theta}^T \underline{\lambda} = \mathbf{0} & (4) \\ \frac{\partial X'^2}{\partial \underline{\lambda}} &= 2\underline{f} = \mathbf{0} & (5) \end{aligned} \right\} \begin{array}{l} N + K + L \text{ équations normales} \\ \text{à } N + K + L \text{ inconnues} \\ \underline{\eta}, \underline{\theta}, \underline{\lambda} \end{array}$$

avec $\left\{ \begin{array}{l} (f_{\eta})_{kn} = \frac{\partial f_k}{\partial \eta_n} \text{ matrice } K \times N \text{ des dérivées des contraintes par les } \underline{\eta} \\ (f_{\theta})_{kl} = \frac{\partial f_k}{\partial \theta_l} \text{ matrice } K \times L \text{ des dérivées des contraintes par les } \underline{\theta} \end{array} \right.$

Recherche des estimations par itération

Développement en série de $\underline{f}^{(v+1)} \equiv f\left(\underline{\eta}^{(v+1)}, \underline{\theta}^{(v+1)}\right)$ à l'itération $v + 1$

autour de $\underline{f}^{(v)} \equiv f\left(\underline{\eta}^{(v)}, \underline{\theta}^{(v)}\right)$ à l'itération v

$$\left. \begin{aligned} V_{\hat{\eta}}^{-1} \left(\underline{\eta}^{(v+1)} - \hat{\eta} \right) + f_{\eta}^{(v)T} \underline{\lambda}^{(v+1)} &= \mathbf{0} & (3') \\ f_{\theta}^{(v)T} \underline{\lambda}^{(v+1)} &= \mathbf{0} & (4') \\ \underline{f}^{(v+1)} = \underline{f}^{(v)} + f_{\eta}^{(v)} \left(\underline{\eta}^{(v+1)} - \underline{\eta}^{(v)} \right) + f_{\theta}^{(v)} \left(\underline{\theta}^{(v+1)} - \underline{\theta}^{(v)} \right) & & (5') \end{aligned} \right\} \begin{array}{l} N + K + L \text{ équations normales} \\ \text{à } N + K + L \text{ inconnues} \\ \underline{\eta}, \underline{\theta}, \underline{\lambda} \end{array}$$

inconnues : $\underline{\eta}^{(v+1)}, \underline{\theta}^{(v+1)}, \underline{\lambda}^{(v+1)}$

valeurs de départ, itération 0 :

$$\underline{\eta}^{(0)} = \hat{\eta}$$

$\underline{\theta}^{(0)}$: résolution d'un sous-ensemble de L parmi les K équations de contrainte

$$\underline{\lambda}_k^{(0)} = \mathbf{1} \quad \forall k = 1, K$$

Résolution du processus itératif

- extraire $\underline{\eta}^{(v+1)}$ en fonction de $\underline{\lambda}^{(v+1)}$ de (3')
- remplacer $\underline{\eta}^{(v+1)}$ par sa fonction en $\underline{\lambda}^{(v+1)}$ dans (5')
- extraire $\underline{\lambda}^{(v+1)}$ en fonction de $\underline{\theta}^{(v+1)}$ de (5')
- remplacer $\underline{\lambda}^{(v+1)}$ dans (4') pour obtenir $\underline{\theta}^{(v+1)}$
- utiliser la relation entre $\underline{\lambda}^{(v+1)}$ et $\underline{\theta}^{(v+1)}$ pour obtenir $\underline{\lambda}^{(v+1)}$
- obtenir $\underline{\eta}^{(v+1)}$ en fonction de $\underline{\lambda}^{(v+1)}$ de (5') modifié

$$\left. \begin{aligned}
 \underline{\theta}^{(v+1)} &= \underline{\theta}^{(v)} - H^{-1} f_{\theta}^{(v)T} S^{-1} \underline{r} \\
 \underline{\lambda}^{(v+1)} &= S^{-1} \left(\underline{r} - f_{\theta}^{(v)} \left(\underline{\theta}^{(v+1)} - \underline{\theta}^{(v)} \right) \right) \\
 \underline{\eta}^{(v+1)} &= \underline{\hat{\eta}} - V_{\hat{\eta}} f_{\eta}^{(v)T} \underline{\lambda}^{(v+1)}
 \end{aligned} \right\} \text{ avec } \left. \begin{aligned}
 S &= \left(f_{\eta}^{(v)} V_{\hat{\eta}} f_{\eta}^{(v)T} \right) \text{ matrice } K \times K \\
 H &= \left(f_{\theta}^{(v)T} S^{-1} f_{\theta}^{(v)} \right) \text{ matrice } L \times L \\
 \underline{r} &= \underline{f}^{(v)} + f_{\eta}^{(v)} \left(\underline{\hat{\eta}} - \underline{\eta}^{(v)} \right) \text{ vecteur } K
 \end{aligned} \right\} (6)$$

Arrêt du processus itératif

$$\begin{array}{l} \hat{\underline{\theta}} = \underline{\theta}^{(v+1)} \\ \hat{\underline{\eta}} = \underline{\eta}^{(v+1)} \end{array} \quad (7) \quad \text{si} \quad \left\{ \begin{array}{l} \left| \theta_l^{(v+1)} - \theta_l^{(v)} \right| \text{ ou } \frac{\left| \theta_l^{(v+1)} - \theta_l^{(v)} \right|}{\theta_l^{(v)}} \leq \varepsilon_l \quad \forall l = 1, L \\ \left| \eta_n^{(v+1)} - \eta_n^{(v)} \right| \text{ ou } \frac{\left| \eta_n^{(v+1)} - \eta_n^{(v)} \right|}{\eta_n^{(v)}} \leq \varepsilon'_n \quad \forall n = 1, N \\ f_k \left(\underline{\eta}^{(v+1)}, \underline{\theta}^{(v+1)} \right) \leq \varepsilon''_k \quad \forall k = 1, K \end{array} \right.$$

Calcul de la matrice des variances/covariances

Relations (7) :

$$\underline{\hat{\eta}} = \underline{g}(\underline{\hat{\eta}})$$

$$\hat{\theta} = h(\underline{\hat{\eta}})$$

et donc, à l'approximation linéaire:

$$\left. \begin{aligned} V_{\hat{\eta}} &= \left(\frac{\partial \underline{g}}{\partial \underline{\hat{\eta}}} \right) V_{\hat{\eta}} \left(\frac{\partial \underline{g}}{\partial \underline{\hat{\eta}}} \right)^T = V_{\hat{\eta}} \left(\underline{I}_N - (G - FH^{-1}F^T) \right) V_{\hat{\eta}} \\ V_{\hat{\theta}} &= \left(\frac{\partial h}{\partial \underline{\hat{\eta}}} \right) V_{\hat{\eta}} \left(\frac{\partial h}{\partial \underline{\hat{\eta}}} \right)^T = H^{-1} \\ Cov_{\hat{\eta}\hat{\theta}} &= \left(\frac{\partial \underline{g}}{\partial \underline{\hat{\eta}}} \right) V_{\hat{\eta}} \left(\frac{\partial h}{\partial \underline{\hat{\eta}}} \right)^T = V_{\hat{\eta}} FH^{-1} \end{aligned} \right\} \text{ avec } \begin{cases} S = f_{\eta}^{(v)} V_{\hat{\eta}} f_{\eta}^{(v)T} \\ H = f_{\theta}^T S^{-1} f_{\theta} \\ G = f_{\eta}^T S^{-1} f_{\eta} \\ F = f_{\eta}^T S^{-1} f_{\theta} \end{cases}$$

Niveau de confiance

Les méthodes d'estimation (maximum de vraisemblance, moindres carrés) sont des algorithmes qui permettent de déterminer les meilleures valeurs des paramètres d'un modèle étant donné les mesures. Elles ne garantissent pas que le modèle est en adéquation avec les données. Elles peuvent être utilisées pour calculer « la meilleure exponentielle décroissante » en accord avec un échantillon issu d'une distribution normale...

L'adéquation entre modèle et données est vérifiée après estimation par un test d'hypothèse ou, de manière équivalente, par le calcul d'un niveau de confiance, suivant la méthode en χ^2 de Pearson. La statistique X^2 de Pearson est distribuée suivant une χ^2_{N-1} si N mesures et le modèle est complètement spécifié.

Que devient le nombre de degrés de liberté si L paramètres $\theta_1 \dots \theta_L$ définissant le modèle sont estimés à partir des mesures?

On montre que, pour N mesures et L paramètres estimés à partir des mesures:

-par moindres carrés : nombre de degrés de liberté = $N-1- L$

-par maximum de vraisemblance : nombre de degrés de liberté indéterminé entre $N-1- L$ et $N-1$.