林楠 Nan LIN 16/10/25 Palaiseau

Cours 4

Modèle Linéaire Gaussien

$\boxed{\text{Cas fréquentiste}}$

N mesures : $\underset{\in \mathbb{R}^{n_Z}}{Z_k} = H_k \underset{\in \mathbb{R}^{n_\theta}}{\overset{\text{inconnus}}{\theta}} + \underset{\in \mathbb{R}^{n_Z}}{B_k} \sim \mathcal{N}(m_k, R_k)$, $\quad R_k \in \mathcal{M}_{n_Z}$

$k \in \{1 \dots N\}$

$Z_k \sim \mathcal{N}(H_k\theta + m_k, R_k)$ indépendants

对应 σ^{-2}

$f(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$

⚠ 高斯分布公式

Estimateur du maximum de vraisemblance

fréquentiste : 最大化似然，本质是 $f(Z_N)$

$$\ln \mathcal{L}(\theta; Z_1 \dots Z_N) = \sum_{k=1}^{N} \ln f_k(Z_k;\theta)$$
$$= \sum_{k=1}^{N} -\ln (2\pi)^{\frac{n_Z}{2}} \det R_k^{\frac{1}{2}} - \frac{1}{2}(Z_k - (H_k\theta + m_k))^T \cdot R_k^{-1} \cdot (Z_k - (H_k\theta+m_k))$$

Pour maximiser $\ln \mathcal{L}(\theta; Z_1 \dots Z_N)$, on considère :

$$\frac{\partial \ln \mathcal{L}(\theta; Z_1 \dots Z_N)}{\partial\theta} = 0$$

Ce qui implique

$$\sum_{i=1}^{N} \frac{\partial}{\partial\theta}\Big[\cancel{-\tfrac{1}{2}}\underbrace{(Z_k - (H_k\theta+m_k))^T R_k^{-1}(Z_k - (H_k\theta+m_k))}_{A}\Big] = 0$$

$$A = (Z_k - m_k - H_k\theta)^T R_k^{-1}(Z_k - m_k - H_k\theta)$$
$$= \underbrace{(Z_k - m_k)^T R_k^{-1}(Z_k - m_k)}_{A_1} - \underbrace{(Z_k - m_k)^T R_k^{-1} H_k\theta}_{A_2} - \underbrace{\theta^T H_k^T R_k^{-1}(Z_k - m_k)}_{A_3} + \underbrace{\theta^T H_k^T R_k^{-1} H_k\theta}_{A_4}$$

R_k^{-1} 因为协方差是对称的

$$\begin{cases} \frac{\partial}{\partial\theta} A_1 = 0 \\ \frac{\partial}{\partial\theta} A_2 = \left[(Z_k - m_k)^T R_k^{-1} H_k\right]^T = H_k^T R_k^{-1T}(Z_k - m_k) \\ \frac{\partial}{\partial\theta} A_3 = H_k^T R_k^{-1}(Z_k - m_k) \\ \frac{\partial}{\partial\theta} A_4 = H_k^T R_k^{-1} H_k + \left[H_k^T R_k^{-1} H_k\right]^T = 2H_k^T R_k^{-1} H_k\theta \end{cases}$$

$$\boxed{\begin{aligned} \frac{\partial \theta^T A}{\partial\theta} &= A \\ \frac{\partial A^T\theta}{\partial\theta} &= A \\ \frac{\partial \theta^T A\theta}{\partial\theta} &= (A + A^T)\theta \end{aligned}}$$

⚠ 矩阵求导公式

$$\Rightarrow \sum_{k=1}^{N} \frac{\partial}{\partial\theta} A = \sum_{k=1}^{N} 2H_k^T R_k^{-1} H_k \hat{\theta}_N - \sum_{k=1}^{N} 2H_k^T R_k^{-1}(Z_k - m_k)$$

$$\Leftrightarrow \left(\sum_{k=1}^{N} H_k^T R_k^{-1} H_k\right)\hat{\theta}_N = \sum_{k=1}^{N} H_k^T R_k^{-1}(Z_k - m_k)$$

$\downarrow$ 堆叠

$$\underbrace{\left(H_{1:N}^T R_{1:N}^{-1} H_{1:N}\right)}_{\text{Matrice normale du problème}} \hat{\theta}_N = \underbrace{H_{1:N}^T R_{1:N}^{-1}(Z_{1:N} - m_{1:N})}_{\text{Équation normale du problème}} \qquad (1)$$

$$\underbrace{\begin{bmatrix} H_1^T & \cdots & H_N^T \end{bmatrix}}_{N n_Z} \Big\updownarrow n_\theta \quad \underbrace{\begin{bmatrix} R_1^{-1} & 0 & \cdots & 0 \\ 0 & R_2^{-1} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & R_N^{-1} \end{bmatrix}}_{(R_{1:N}^{-1}),\ N n_Z} \quad \underbrace{\begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix}}_{n_\theta} \Big\updownarrow N n_Z \ (\updownarrow n_Z) \quad \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix} \quad \begin{bmatrix} m_1 \\ \vdots \\ m_N \end{bmatrix}$$

Rappel Équation normale du problème (1)

$$(H_{1:N}^T R_{1:N}^{-1} H_{1:N})\hat{\theta}_N = H_{1:N}^T R_{1:N}^{-1}(Z_{1:N} - m_{1:N}) \quad \text{avec } H_{1:N} = \begin{bmatrix} H_1 \\ | \\ H_N \end{bmatrix},\ R_{1:N} = \begin{bmatrix} R_1 & & 0 \\ & R_2 & \\ 0 & & \ddots \quad R_N \end{bmatrix}$$

同样的分布

Cas particuliers : ① $\forall k,\ H_k = H,\ m_k = m,\ R_k = R$

Équation normale devient : $(H^T R^{-1} H)\hat{\theta}_N = H^T R^{-1}(\bar{Z} - m)$

退化成只有一个 avec $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$

② $m_k = 0,\ R_k = I_k$ cad $B_k \sim \mathcal{N}(0,1)$

Équation normale devient : $(H_{1:N}^T H_{1:N})\hat{\theta}_N = H_{1:N}^T Z_{1:N}$

si $\forall k,\ H_k = H$ $\Rightarrow$ $(H^T H)\hat{\theta}_N = H^T \bar{Z}$

Résolution du problème – équation (1)

1. Théorème de Gauss-Markov

Conditions : ① Toutes les mesures sont disponibles au moment de l'estimation

② $H_{1:N}^T R_{1:N}^{-1} H_{1:N}$ inversible

Solution est donc
$$\boxed{\hat{\theta}_N = (H_{1:N}^T R_{1:N}^{-1} H_{1:N})^{-1} H_{1:N}^T R_{1:N}^{-1}(Z_{1:N} - m_{1:N})} \quad (2)$$

$\hat{\theta}_N$ est sans biais à minimum de variance, consistant et unique

2. Algorithme récursifs

Les mesures sont arrivés au fur et à mesure ou $H_{1:N}^T R_{1:N}^{-1} H_{1:N}$ n'est pas invisible

$$\begin{cases} \hat{\theta}_0 \\ \hat{\theta}_{k+1} = f(\hat{\theta}_k, Z_{k+1}) \end{cases} \quad \text{avec objectif } \mathcal{L}(\hat{\theta}_{k+1}; Z_1 - Z_{k+1}) \geq \mathcal{L}(\hat{\theta}_k, Z_1 Z_k)$$

下面找到从 $\hat{\theta}_k \to \hat{\theta}_{k+1}$ 的方法

Noté $\begin{cases} Q_k = H_{1:k}^T R_{1:k}^{-1} H_{1:k} = \sum_{i=1}^{k} H_i^T R_i^{-1} H_i, \text{ donc } Q_{k+1} = Q_k + H_{k+1}^T R_{k+1}^{-1} H_{k+1} & ① \\ b_k = H_{1:k}^T R_{1:k}^{-1}(Z_{1:k} - m_{1:k}) = \sum_{i=1}^{k} H_i^T R_i^{-1}(Z_i - m_i) & ③ \end{cases}$

donc $b_{k+1} = b_k + H_{k+1}^T R_{k+1}^{-1}(Z_{k+1} - m_{k+1})$ ④

L'équation (1) devient

pour $N = k$, pour $N = k+1$
$$\begin{cases} Q_k \hat{\theta}_k = b_k \\ Q_{k+1}\hat{\theta}_{k+1} = b_{k+1} \end{cases} \Rightarrow \begin{cases} \hat{\theta}_k = Q_k^{-1} b_k & ⑤ \\ \hat{\theta}_{k+1} = Q_{k+1}^{-1} \underbrace{b_{k+1}}_{= b_k + \ldots} & ⑥ \end{cases}$$

D'après le lemme de Schur et ②, 试图寻找 Q_{k+1}^{-1} 和 Q_k^{-1} 关系

$$Q_{k+1}^{-1} = (\underbrace{Q_k}_{A} + \underbrace{H_{k+1}^T}_{B} \underbrace{R_{k+1}^{-1}}_{C} \underbrace{H_{k+1}}_{D})^{-1}$$

$$= \underbrace{Q_k^{-1}}_{P_k} - \boxed{\underbrace{Q_k^{-1}}_{P_k} H_{k+1}^T \underbrace{(H_{k+1} Q_k^{-1} H_{k+1}^T + R_{k+1})^{-1}}_{S_{k+1}}}^{K_{k+1}} H_{k+1} Q_k^{-1} \quad ⑦$$

Noté $\forall k$,
$$\begin{cases} P_k = Q_k^{-1} & ⑧ \\ S_{k+1} = H_{k+1} Q_k^{-1} H_{k+1}^T + R_{k+1} & ⑨ \\ K_{k+1} = P_k H_{k+1}^T S_{k+1}^{-1} & ⑩ \end{cases}$$

⑧⑨⑩

$$\Leftrightarrow P_{k+1} = P_k - K_{k+1} H_{k+1} P_k \quad ⑪$$

Donc, (b) $\rightarrow \hat{\theta}_{k+1} = P_{k+1} b_{k+1}$

(4) (11) $\rightarrow$

$$= (P_k - K_{k+1} H_{k+1} P_k)(b_k + H_{k+1}^T R_{k+1}^{-1} (z_{k+1} - m_{k+1}))$$

$$= \underbrace{P_k b_k}_{⑤\ \hat{\theta}_k} - K_{k+1} H_{k+1} \underbrace{P_k b_k}_{⑤\ \hat{\theta}_k} + (P_k - K_{k+1} H_{k+1} P_k) H_{k+1}^T R_{k+1}^{-1} (z_{k+1} - m_{k+1})$$

$$= \hat{\theta}_k - K_{k+1} H_{k+1} \hat{\theta}_k + \underbrace{\left[P_k H_{k+1}^T R_{k+1}^{-1} - K_{k+1} H_{k+1} P_k H_{k+1}^T R_{k+1}^{-1}\right]}_{(B)} (z_{k+1} - m_{k+1}) \quad (12)$$

On voudrait calculer (B)

À partir de (9), on obtient

$$(S_{k+1} - H_{k+1} Q_k^{-1} H_{k+1}^T) R_{k+1}^{-1} = R_{k+1} R_{k+1}^{-1} = I$$

$$S_{k+1}^{-1} (S_{k+1} - H_{k+1} Q_k^{-1} H_{k+1}^T) R_{k+1}^{-1} = S_{k+1}^{-1}$$

$$[I - S_{k+1}^{-1} H_{k+1} Q_k^{-1} H_{k+1}^T) R_{k+1}^{-1} = S_{k+1}^{-1} \quad (13)$$

Donc $(B) = \underbrace{P_k H_{k+1}^T}\left(R_{k+1}^{-1} - S_{k+1}^{-1} H_{k+1} \underbrace{P_k H_{k+1}^T}_{(10)} R_{k+1}^{-1}\right) \overset{(13)}{=} P_k H_{k+1}^T S_{k+1}^{-1} \overset{(10)}{=} K_{k+1}$

Alors,

(12) $\rightarrow \quad \hat{\theta}_{k+1} = \hat{\theta}_k - K_{k+1} H_{k+1} \hat{\theta}_k + K_{k+1} (z_{k+1} - m_{k+1})$

$$= \hat{\theta}_k + K_{k+1} (z_{k+1} - m_{k+1} - H_{k+1} \hat{\theta}_k) = \hat{\theta}_{k+1} \quad (3\text{-}1)$$

avec

$$K_{k+1} = P_k H_{k+1}^T S_{k+1}^{-1} \quad (3\text{-}2)$$

$$S_{k+1} = H_{k+1} Q_k^{-1} H_{k+1}^T + R_{k+1} \quad (3\text{-}3)$$

$$P_{k+1} = P_k - K_{k+1} H_{k+1} P_k \quad (3\text{-}4)$$

(3)

Algorithme des Moindres Carrés Récursifs (MCR)

1. Initialisation (e.g. $\theta_0 = 0_{n_\theta,1}$, $P_0 = \alpha I_{n_\theta}$)

   θ_0 est l'estimé initial

   对初始估计值的不确定程度

   P_0 est la covariance de l'estimé initial.

2. Itération pour $k \in \{0, .. N-1\}$. Voir (3).

Position du problème : • $\hat{\theta}_{k+1} = \hat{\theta}_k + \underbrace{K_{k+1}}_{\text{Gain}} \underbrace{(z_{k+1} - m_{k+1} - H_{k+1} \hat{\theta}_k)}_{\text{Innovation}}$

新数据不再带来改善 • De moins en moins en faveur des nouvelles données car $tr(P_{k+1})$ décroit.

Algorithmes adaptifs

Conserver les influences de donnés les plus récentes

$\rightarrow$ Moindre Carrés adaptifs à oubli exponentiel.

$$(1) \rightarrow \left(\sum_{k=1}^{N} \lambda^{N-k} (H_k^T R_k^{-1} H_k)\right) \hat{\theta}_N = \sum_{k=1}^{N} \lambda^{N-k} H_k^T R_k^{-1} (z_k - m_k)$$

avec $0 < \lambda \leq 1$ le facteur d'oubli (4)

Résolution : (3-4) devient $P_{k+1} = \frac{1}{\lambda} (P_k - K_{k+1} H_{k+1} P_k)$

(3-1) ~ (3-3) inchangées

增加不确定性，从而也会导致权重 K_{k+1} 变大

. λ ne doit pas prendre une valeur trop faible

会造成 instabilité

Modèle Linéaire Gaussien

Cas Bayésien

N mesures : $k \in \{1, \dots, N\}$, $\underset{\mathbb{R}^{n_z}}{Z_k} = \underset{M_{n_z, n_\theta}}{H_k} \overset{\text{inconnue}}{\underset{\mathbb{R}^{n_\theta}}{\theta}} + \underset{\mathbb{R}^{n_z}}{B_k}$ et $B_k \sim \mathcal{N}(\underset{\mathbb{R}^{n_z}}{m_k}, \underset{M_{n_z}}{R_k})$

Bayésien不同的地方

Xi.

Supposons que $\theta \sim \mathcal{N}(m_\theta, R_\theta)$ ① et $Z_k | \theta \sim \mathcal{N}(H_k\theta + m_k, R_k)$ ②

Équation normale

$$\ln f(\theta | Z_1 - Z_N) = \ln\left(\frac{f(Z_1 - Z_N|\theta) f(\theta)}{f(Z_1 - Z_N)}\right) = \underbrace{\ln f(Z_1 - Z_N|\theta)}_{A} + \underbrace{\ln f(\theta)}_{B} - \underbrace{\ln f(Z_1 - Z_N)}_{C}$$

$$A \overset{②}{=} \sum_{k=1}^{N} \ln f_k(Z_k ; \theta) = \sum_{k=1}^{N} -\ln\sqrt{(2\pi)^{n_z}\det(R_k)} - \frac{1}{2}(Z_k - H_k\theta - m_k)^T R_k^{-1}(Z_k - H_k\theta - m_k)$$

$$B \overset{①}{=} -\ln\sqrt{(2\pi)^{n_\theta}\det(R_\theta)} - \frac{1}{2}(\theta - m_\theta)^T R_\theta^{-1}(\theta - m_\theta)$$

Ⓒ inchangé

$$= \sum_{k=1}^{N}\left[-\ln\sqrt{(2\pi)^{n_z}\det(R_k)} - \frac{1}{2}(Z_k - H_k\theta - m_k)^T R_k^{-1}(Z_k - H_k\theta - m_k)\right] - \ln\sqrt{(2\pi)^{n_\theta}\det R_\theta} - \frac{1}{2}(\theta - m_\theta)^T R_\theta^{-1}(\theta - m_\theta) - \ln f(Z_1 - Z_N)$$

La recherche du maximum amène :

$$\frac{\partial \ln f(\theta|Z_1 - Z_N)}{\partial\theta} = 0 \Rightarrow \frac{\partial}{\partial\theta}\underbrace{\sum_{k=1}^{N}(Z_k - H_k\theta - m_k)^T R_k^{-1}(Z_k - H_k\theta - m_k)}_{A} + \frac{\partial}{\partial\theta}\underbrace{(\theta - m_\theta)^T R_\theta^{-1}(\theta - m_\theta)}_{B} = 0$$

$$\frac{\partial}{\partial\theta} A = \sum_{k=1}^{N} 2(H_k^T R_k^{-1} H_k)\hat{\theta}_N - \sum_{k=1}^{N} 2 H_k^T R_k^{-1} Z_k$$

第1页已有推导

$$\frac{\partial}{\partial\theta} B = \frac{\partial}{\partial\theta}\left[\hat{\theta}_N^T R_\theta^{-1}\hat{\theta}_N - \hat{\theta}_N^T R_\theta^{-1} m_\theta - m_\theta^T R_\theta^{-1}\hat{\theta}_N + m_\theta^T R_\theta^{-1} m_\theta\right]$$

$$= (R_\theta^{-1} + R_\theta^{-1T})\hat{\theta}_N - R_\theta^{-1} m_\theta - R_\theta^{-1T} m_\theta + 0 = 2R_\theta^{-1}\hat{\theta}_N - 2R_\theta^{-1} m_\theta$$

L'équation donne :

$$\boxed{\left(R_\theta^{-1} + \sum_{k=1}^{N}(H_k^T R_k^{-1} H_k)\right)\hat{\theta}_N = R_\theta^{-1} m_\theta + \sum_{k=1}^{N} H_k^T R_k^{-1}(Z_k - m_k)}$$

ou

$$\underbrace{\underbrace{(R_\theta^{-1} + H_{1:N}^T R_{1:N}^{-1} H_{1:N})}_{\text{Matrice normale du problème}}\hat{\theta}_N = R_\theta^{-1} m_\theta + H_{1:N}^T R_{1:N}^{-1}(Z_{1:N} - m_{1:N})}_{\text{L'équation normale du problème}}$$

Ainsi on obtient l'estimateur de Gauss-Markov :

$$\hat{\theta}_N = (R_\theta^{-1} + H_{1:N}^T R_{1:N}^{-1} H_{1:N})^{-1}(R_\theta^{-1} m_\theta + H_{1:N}^T R_{1:N}^{-1}(Z_{1:N} - m_{1:N}))$$

* La matrice normale est toujours inversible !

R_θ^{-1} 正定 $H_{1:N}^T R_{1:N}^{-1} H_{1:N}$ 半正定 两者之和一定正定.

Moindres carrés récursifs dans le cadre bayésien

Identique au cas fréquentiste (voir (31)), mais avec les valeurs initiales $P_0 = R_\theta$, $\hat{\theta}_0 = m_\theta$