

Examen DEA SICOM, Cours Identification
Luc Pronzato, Décembre 2002

Les problèmes 1 et 2 sont indépendants. Dans le problème 2, les parties A et B sont indépendantes. Les parties 2.3, 2.4 et 2.5 peuvent se traiter même si l'on n'a pas su répondre aux questions qui précèdent. On prendra soin de souligner les vecteurs (lettres minuscules en gras dans l'énoncé).

Problème 1 : planification d'expériences

On estime par moindres carrés les paramètres θ d'un modèle

$$y_m(\theta, t) = \mathbf{r}^\top(t)\theta, \theta \in \mathbb{R}^p.$$

Les erreurs de mesure ϵ_i sont de moyenne nulle, non corrélées ($E\{\epsilon_i\epsilon_j\} = 0$ pour $i \neq j$) de variance $E\{\epsilon_i^2\} = \sigma^2$ pour tout i . On notera $\hat{\theta}$ la vraie valeur (inconnue) de θ .

1.1) Donner l'expression de l'estimateur $\hat{\theta}$ en fonction du vecteur \mathbf{y} des observations, $\mathbf{y} = (y_1, \dots, y_N)^\top$, $N \geq p$, et de la matrice \mathbf{R} des régresseurs (matrice dont la ligne i est donnée par $\mathbf{r}^\top(t_i)$, $i = 1, \dots, N$).

1.2) Calculer la moyenne et la matrice de covariance \mathbf{C} de $\hat{\theta}$.

1.3) On souhaite prédire la valeur de la réponse du modèle à un instant t_0 particulier. On utilise le prédicteur $y_m(\hat{\theta}, t_0) = \mathbf{r}^\top(t_0)\hat{\theta}$. Calculer la moyenne et la variance de la prédiction. Comment choisir les temps d'observation pour avoir la prédiction la plus précise possible?

1.4) On s'intéresse à présent aux prédictions obtenues pour tout t tel que $\mathbf{r} = \mathbf{r}(t)$ est dans $\mathcal{B}(\mathbf{0}, R)$, la boule de centre $\mathbf{0}$ de rayon R . Si l'on souhaite obtenir la plus grande précision possible dans le pire cas, c'est-à-dire garantir une précision pour tout \mathbf{r} dans $\mathcal{B}(\mathbf{0}, R)$, comment choisir les temps d'observation? On exprimera le critère à optimiser en fonction des valeurs propres de \mathbf{C} . (On rappelle que pour \mathbf{A} symétrique semi-définie positive,

$$\max_{\{\mathbf{x} / \|\mathbf{x}\|=1\}} \mathbf{x}^\top \mathbf{A} \mathbf{x} = \lambda_{\max}(\mathbf{A}), \text{ plus grande valeur propre de } \mathbf{A}.)$$

Si l'on souhaite obtenir la plus grande précision possible en moyenne, pour \mathbf{r} distribué suivant une loi normale de moyenne $\mathbf{0}$ et de matrice de covariance Ω , comment choisir les temps d'observation? On exprimera le critère à optimiser en fonction de \mathbf{C} et Ω .

Problème 2 : krigeage

On effectue N observations

$$y_i = y(x_i) = \theta + z(x_i) + \epsilon_i$$

où les x_i sont choisis (déterministes), les ϵ_i sont des réalisations de variables aléatoires indépendantes, de moyenne nulle et de variance constante σ^2 , et les $z(x_i)$ sont aussi des réalisations de variables aléatoires, indépendantes des ϵ_i mais dépendantes entre elles. On écrira $z_i = z(x_i)$, et on suppose que $E\{z_i\} = 0$ et $E\{z_i, z_j\} = \omega^2 C(x_i - x_j)$, avec $C(0) = 1$. Le paramètre (scalaire) θ et les variances σ^2 et ω^2 sont des constantes (que l'on essaiera plus loin d'estimer). La fonction $C(\cdot)$ est également une donnée du problème. On supposera plus loin que cette fonction est de forme paramétrique, c'est à dire

$$C(x_i - x_j) = f(\beta, x_i - x_j)$$

avec $f(\cdot, \cdot)$ connue et β des paramètres que l'on essaiera d'estimer.

Partie A Dans un premier temps, θ , σ^2 , ω^2 et $C(\cdot)$ sont supposés connus.

2.1 Après avoir observé $\mathbf{y} = (y_1, \dots, y_N)^\top$, on souhaite prédire la valeur de y en un autre point x , où l'on n'a pas effectué d'observation. On notera $\hat{y}(x)$ cette prédiction, et $y(x)$ la valeur moyenne, par rapport au bruit ϵ , de y en x , c'est-à-dire $y(x) = \theta + z(x)$ (noter que $y(x)$ est inconnu).

On utilise un prédicteur linéaire, c'est-à-dire que l'on impose que $\hat{y}(x)$ est une combinaison linéaire des y_i , $i = 1, \dots, N$: $\hat{y}(x) = \mathbf{c}^\top(x)\mathbf{y}$, avec $\mathbf{c}(x)$ un vecteur de dimension N que l'on se propose de choisir.

On note $\mathbf{z} = (z_1, \dots, z_N)^\top$, \mathbf{n} le vecteur $(1, \dots, 1)^\top$ dont les N composantes valent 1, \mathbf{C} la matrice de composantes $[\mathbf{C}]_{i,j} = C(x_i - x_j)$,

$$\mathbf{V} = \sigma^2 \mathbf{I}_N + \omega^2 \mathbf{C}, \text{ avec } \mathbf{I}_N \text{ la matrice identité de dimension } N$$

et $\mathbf{u}(x)$ le vecteur

$$\mathbf{u}(x) = \mathbf{E}\{z(x)\mathbf{z}\}.$$

Calculer l'erreur quadratique moyenne de la prédiction

$$EQM(x) = \mathbf{E}\{[\hat{y}(x) - y(x)]^2\}$$

($\mathbf{E}\{\cdot\}$ désigne la moyenne par rapport à toutes les variables aléatoires présentes, c'est-à-dire les ϵ_i , les z_i et $z(x)$).

2.2 Quelle contrainte imposer à $\mathbf{c}(x)$ pour que $EQM(x)$ ne dépende pas de θ ?

On souhaite utiliser la valeur de $\mathbf{c}(x)$ qui minimise $EQM(x)$ sous la contrainte précédente. Montrer que le Lagrangien du problème s'écrit

$$\mathcal{L}(\mathbf{c}, \lambda) = \omega^2 + \mathbf{c}^\top \mathbf{V} \mathbf{c} - 2\mathbf{u}^\top \mathbf{c} + 2\lambda(\mathbf{c}^\top \mathbf{n} - 1)$$

où $\mathbf{c} = \mathbf{c}(x)$, $\mathbf{u} = \mathbf{u}(x)$ et λ est la variable de Lagrange associée à la contrainte imposée sur \mathbf{c} .

2.3 Écrire les conditions de stationnarité de $\mathcal{L}(\mathbf{c}, \lambda)$. Montrer qu'elles se mettent sous la forme

$$\begin{pmatrix} \mathbf{V} & \mathbf{n} \\ \mathbf{n}^\top & 0 \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{u} \\ 1 \end{pmatrix}$$

Calculer $\mathbf{c}(x)$. (On rappelle que

$$\begin{pmatrix} \mathbf{V} & \mathbf{n} \\ \mathbf{n}^\top & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{V}^{-1} - \frac{\mathbf{V}^{-1} \mathbf{n} \mathbf{n}^\top \mathbf{V}^{-1}}{\mathbf{n}^\top \mathbf{V}^{-1} \mathbf{n}} & \frac{\mathbf{V}^{-1} \mathbf{n}}{\mathbf{n}^\top \mathbf{V}^{-1} \mathbf{n}} \\ \frac{\mathbf{n}^\top \mathbf{V}^{-1}}{\mathbf{n}^\top \mathbf{V}^{-1} \mathbf{n}} & -\frac{1}{\mathbf{n}^\top \mathbf{V}^{-1} \mathbf{n}} \end{pmatrix}.$$

Soit $\hat{\theta}_{MCP}$ l'estimateur de θ pour le critère des moindres carrés pondérés, avec la matrice de pondération \mathbf{V}^{-1} , c'est-à-dire qui minimise

$$(\mathbf{y} - \theta \mathbf{n})^\top \mathbf{V}^{-1} (\mathbf{y} - \theta \mathbf{n}).$$

Donner l'expression de $\hat{\theta}_{MCP}$.

Exprimer la valeur de la prédiction $\hat{y}(x)$ en fonction de $\hat{\theta}_{MCP}$, $\mathbf{u}(x)$, \mathbf{V} et $(\mathbf{y} - \hat{\theta}_{MCP} \mathbf{n})$.

2.4 Calculer $EQM(x)$ pour la valeur de $\mathbf{c}(x)$ trouvée en 2.3.

2.5 On suppose que $\sigma^2 = 0$ (pas de bruit de mesure). Montrer que $\mathbf{V}^{-1} \mathbf{u}(x_i) = \mathbf{e}_i$, i ème vecteur de base, pour tout i de 1 à N .

Calculer alors $EQM(x_i)$, $i = 1, \dots, N$, et commenter ce résultat.

Partie B

2.6 On souhaite à présent estimer les paramètres inconnus du problème, par *maximum de vraisemblance*.

On suppose que $C(x_i - x_j) = f(\beta, x_i - x_j)$, ce qui donne une matrice $\mathbf{C} = \mathbf{C}(\beta)$. Les paramètres du modèle (à estimer) sont donc $\mu = (\theta, \beta, \sigma^2, \omega^2)$.

Donner l'expression de la densité de probabilité $\pi(\mathbf{y}|\mu)$ lorsque les ϵ_i et les z_i sont supposés gaussiens (on rappelle que la somme de 2 variables gaussiennes est encore gaussienne, et pour les variables indépendantes les variances s'ajoutent).

En déduire l'expression de la log-vraisemblance de \mathbf{y} .

2.7 Calculer l'estimateur $\hat{\theta}_{MV}$ de θ au sens du maximum de vraisemblance (on utilisera la stationnarité de la log-vraisemblance à l'optimum). Comparer avec $\hat{\theta}_{MCP}$.

2.8 On pose $\sigma^2 = \alpha \omega^2$, ce qui donne une autre paramétrisation $\mu' = (\theta, \beta, \alpha, \omega^2)$.

Calculer l'estimateur ω_{MV}^2 de ω^2 au sens du maximum de vraisemblance (on utilisera encore la stationnarité de la log-vraisemblance à l'optimum).

Reporter la valeur de ω_{MV}^2 dans la log-vraisemblance pour obtenir enfin l'expression à maximiser par rapport à α et β (en fonction de ω_{MV}^2 , α et $\mathbf{C} = \mathbf{C}(\beta)$).