

Modélisation probabiliste des incertitudes

G. Perrin

guillaume.perrin@univ-eiffel.fr

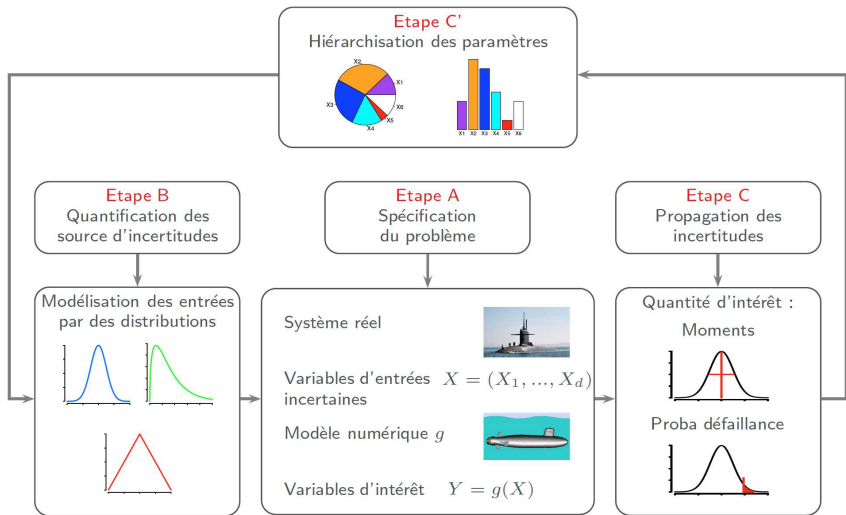
Année 2022-2023



Plan de la séance

- 1 Introduction
- 2 Maximum d'entropie
- 3 Méthode d'ajustement de familles de loi
- 4 Méthode à noyau

Introduction - Schéma général



Introduction - contexte de la séance

- On suppose s'intéresser à un système particulier, dont le comportement peut être décrit par $d \geq 1$ paramètres scalaires regroupés dans le vecteurs $\mathbf{x} = (x_1, \dots, x_d)$.
- Ces entrées sont considérées **incertaines**, au sens où ils ne sont pas parfaitement connus.
- On peut distinguer deux types d'entrées incertaines :
 - les variables **stochastiques** \leftrightarrow incertitude liée à une variabilité naturelle résultant de phénomènes aléatoires (fluctuations aléatoires dans un procédé de fabrication, vent,...),
 - les variables **épistémiques** \leftrightarrow incertitude liée à un manque de connaissance pour des quantités déterministes (grandeurs physiques par ex.).
- Cette distinction est particulièrement importante lorsque l'on s'intéresse à des **séries** de systèmes.
- Dans la suite on modélise \mathbf{x} comme un vecteur aléatoire de PDF $f_{\mathbf{x}}$ à déterminer.

Introduction - problématique

Le choix de f_x est **délicat** :

- un choix arbitraire peut significativement fausser les conclusions d'une quantification d'incertitude,
- l'information disponible est souvent partielle pour construire une loi de probabilité.

⇒ comment construire f_x **le moins subjectivement** possible, à partir de la seule information disponible a priori ?

Quelques approches classiques

- 1 maximum d'entropie (avis d'expert sans observation),
- 2 lois paramétriques (avis d'expert + quelques observations),
- 3 méthode des noyaux (beaucoup d'observations).

Introduction - un exemple

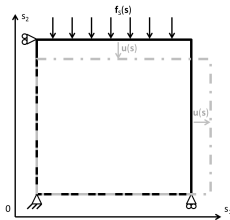
- On cherche à dimensionner un barrage pour un site particulier.
- Un nouveau béton a été mis au point spécialement pour cet objectif, aux propriétés a priori "remarquables" (corrosion, rigidité, coût de fabrication, longévité...).



Introduction - un exemple

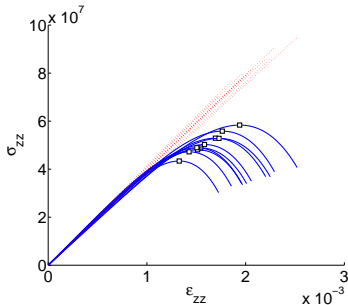
- On dispose d'une loi de comportement pour ce matériau (ex : élasto-plasticité), caractérisée par trois paramètres :
 - un module de Young E ,
 - un coefficient de Poisson ν ,
 - une pression maximale P_{\max} à rupture en compression.
- ⇒ Les valeurs de ces paramètres, regroupés dans $\mathbf{x} = (E, \nu, P_{\max})$, sont a priori **inconnues**.
- ⇒ On aimerait disposer d'une loi pour \mathbf{x} permettant de caractériser les valeurs qu'il peut vraisemblablement prendre.

Introduction - un exemple



- On prépare alors une série de N échantillons cubiques de ce même matériau, que l'on soumet à une **même** pression verticale lentement croissante $\mathbf{f}_s(t)$ (quasi-statique).
- On mesure les déplacements induits $\mathbf{u}^{(n)}(t)$, ainsi que la pression $P_{\max}^{(n)}$ ayant conduit à la ruine du matériau, $1 \leq n \leq N$.

Introduction - un exemple



Pente à l'origine $\times 10^{10} \text{ Pa}$	$\rho_{\text{max},n}^{\text{mes}} \times 10^7 \text{ Pa}$
3.94 ± 0.01	4.8 ± 0.1
3.96 ± 0.01	4.8 ± 0.1
4.13 ± 0.01	4.7 ± 0.1
3.96 ± 0.01	5.0 ± 0.1
4.08 ± 0.01	4.3 ± 0.1
3.76 ± 0.01	5.8 ± 0.1
3.95 ± 0.01	5.5 ± 0.1
3.90 ± 0.01	5.2 ± 0.1
3.81 ± 0.01	5.2 ± 0.1
4.06 ± 0.01	4.8 ± 0.1

⇒ A partir de ces N résultats, comment estimer f_x ?

! Potentielle présence d'une **double source d'incertitudes** :

- incertitudes de mesure $\leftrightarrow \pm 0.1$ (Incertitude **épistémique** ou "réductible"),
- variabilité "naturelle" $\leftrightarrow \{4.8, 4.8, 4.7, \dots\}$ (Incertitude **aléatoire** ou "irréductible").

Plan de la séance

- 1 Introduction
- 2 Maximum d'entropie
- 3 Méthode d'ajustement de familles de loi
- 4 Méthode à noyau

Maximum d'entropie - principe

Information disponible

- Aucune observation de \mathbf{x} .
- Un avis d'expert portant sur le support et/ou les moments statistiques de \mathbf{x} (ex : moyenne, variance,...).

On suppose alors que cette information peut s'écrire sous la forme d'un sous ensemble \mathcal{A} à $M + 1$ contraintes, tel que :

$$\mathcal{A} = \left\{ \int_{\mathbb{R}^d} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = 1, \int_{\mathbb{R}^d} \mathbf{g}_i(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = \mathbf{h}_i, 1 \leq i \leq M \right\}.$$

Principe

La loi $f_{\mathbf{x}}$ à construire est celle qui doit avoir la **plus grande "incertitude"** sur l'ensemble de toutes les PDFs qui vérifient les contraintes définies par l'information **disponible** (ou **utilisable**).

Maximum d'entropie - formulation du problème

L'entropie statistique S comme mesure de l'incertitude

$$S(f_{\mathbf{x}}) = - \int_{\mathbb{R}^d} f_{\mathbf{x}}(\mathbf{x}) \log(f_{\mathbf{x}}(\mathbf{x})) d\mathbf{x} = -\mathbb{E} [\log(f_{\mathbf{x}}(\mathbf{x}))].$$

On peut notamment montrer que :

- $S(f_{\mathbf{x}}) = -\infty \leftrightarrow$ absence d'incertitude (fonction Dirac).
- Si $f_{\mathbf{x}}$ est à support compact K , alors $f_{\mathbf{x}} = \frac{1}{|K|} \mathbb{I}_K$ maximise S .
- Si \mathbf{X} et \mathbf{Y} sont dépendants, alors $S(f_{\mathbf{XY}}) \leq S(f_{\mathbf{X}}) + S_N(f_{\mathbf{Y}})$, et s'ils sont indépendants, $S(f_{\mathbf{XY}}) = S(f_{\mathbf{X}}) + S(f_{\mathbf{Y}})$.

Afin de maximiser l'incertitude (et donc de minimiser les biais potentiels, on cherche alors $f_{\mathbf{x}}$ comme solution de :

$$f_{\mathbf{x}} = \arg \max_{f \in \mathcal{A}} S(f).$$

Maximum d'entropie - solution

La **solution analytique** du problème précédent, pour une fonction à support sur K , s'écrit :

$$f_{\mathbf{x}}(\mathbf{x}) = \mathbb{I}_K(\mathbf{x}) \exp \left(-\lambda_0 - \sum_{m=1}^M \langle \boldsymbol{\lambda}_m, \mathbf{g}_m(\mathbf{x}) \rangle \right),$$

où les $M + 1$ multiplicateurs de Lagrange $\lambda_0, \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_M$ sont à déterminer à partir des contraintes précédentes, ce qui peut se récrire sous la forme d'une minimisation d'une **fonctionnelle convexe** :

$$\boldsymbol{\Lambda} = \arg \min_{\boldsymbol{\Lambda}} \left[\langle \boldsymbol{\Lambda}, \mathbf{H} \rangle_{\mathbb{R}^D} + \int_K \exp(-\langle \boldsymbol{\Lambda}, \mathbf{G}(\mathbf{x}) \rangle) d\mathbf{x} \right],$$

$$\boldsymbol{\Lambda} = (\lambda_0, \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_M) \in \mathbb{R}^D, \quad \mathbf{H} = (1, \mathbf{h}_1, \dots), \quad \mathbf{G} = (1, \mathbf{g}_1, \dots).$$

Maximum d'entropie - Résolution analytique en 1D

- ❶ Support compact $[a, b]$ connu :

$$f_{\theta}(\theta) = \mathbb{I}_{[a,b]}(\theta).$$

- ❷ Support $[0, +\infty[$ et moyenne m_{θ} connus :

$$f_{\theta}(\theta) = \mathbb{I}_{[0,+\infty[}(\theta) \frac{1}{m_{\theta}} e^{-\theta/m_{\theta}}.$$

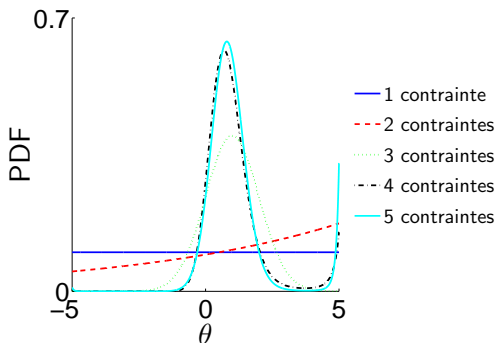
- ❸ Support \mathbb{R} , moyenne μ et écart type σ connus :

$$f_{\theta}(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma} \right)^2 \right).$$

Les distributions classiques uniforme, exponentielle et gaussienne sont des solutions particulières du principe du maximum d'entropie.

Maximum d'entropie - Résolution numérique en 1D

Contraintes : Support $[-5, 5]$, $\mathbb{E}[\theta] = 1$, $\mathbb{E}[(\theta - \mu)^2] = 1$,
 $\mathbb{E}[(\theta - \mu)^3] = 2$, $\mathbb{E}[(\theta - \mu)^4] = 10$.



Plan de la séance

- 1 Introduction
- 2 Maximum d'entropie
- 3 Méthode d'ajustement de familles de loi**
- 4 Méthode à noyau

Méthode d'ajustement de familles de loi - principe

Information disponible

- n réalisations indépendantes de \mathbf{x} , regroupées dans $\mathcal{S}_n := \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$.
- Un avis d'expert supposant que $f_{\mathbf{x}} \in \mathcal{P} := \{p_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^q\}$, ce qui revient à dire supposer qu'il existe $\boldsymbol{\theta}^*$ tel que $f_{\mathbf{x}} = p_{\boldsymbol{\theta}^*}$ (ex : lois uniformes, gaussiennes, de Poisson...).

Principe

Construire $f_{\mathbf{x}}$ revient à chercher la valeur de $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ la plus vraisemblable vis-à-vis de l'information disponible contenue dans \mathcal{S}_n :

- méthode des moments,
- méthode du maximum de vraisemblance,
- approches bayésiennes.

Méthode d'ajustement de familles de loi - méthode des moments

- ❶ Chercher une transformation F telle que $\boldsymbol{\theta} = \mathbb{E}_{\boldsymbol{\theta}} [F(\mathbf{x})]$.
- ❷ L'estimateur empirique de $\boldsymbol{\theta}$ s'écrit alors : $\hat{\boldsymbol{\theta}} := \frac{1}{n} \sum_{i=1}^n F(\mathbf{x}_i)$.

Si \mathcal{P} est régulière vis-à-vis de $\boldsymbol{\theta}$, et **si $f_{\mathbf{x}}$ est bien dans \mathcal{P}** , alors $p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}') \approx p_{\boldsymbol{\theta}^*}(\mathbf{x}') + \frac{\partial p_{\boldsymbol{\theta}^*}}{\partial \boldsymbol{\theta}}(\mathbf{x}')(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ et l'erreur quadratique moyenne est elle aussi en $1/n$:

$$\begin{aligned}
 \mathcal{E} &= \mathbb{E} \left[\int_{\mathbb{R}^d} (p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}') - p_{\boldsymbol{\theta}^*}(\mathbf{x}'))^2 d\mathbf{x}' \right] \\
 &\approx \int_{\mathbb{R}^d} \frac{\partial p_{\boldsymbol{\theta}^*}}{\partial \boldsymbol{\theta}}(\mathbf{x}') \mathbb{E} \left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T \right] \left(\frac{\partial p_{\boldsymbol{\theta}^*}}{\partial \boldsymbol{\theta}}(\mathbf{x}') \right)^T d\mathbf{x}' \\
 &\approx \frac{1}{n} \int_{\mathbb{R}^d} \frac{\partial p_{\boldsymbol{\theta}^*}}{\partial \boldsymbol{\theta}}(\mathbf{x}') \mathbf{R} \left(\frac{\partial p_{\boldsymbol{\theta}^*}}{\partial \boldsymbol{\theta}}(\mathbf{x}') \right)^T d\mathbf{x}', \quad \mathbf{R} := \text{Cov}(F(\mathbf{x})).
 \end{aligned}$$

Méthode d'ajustement de familles de loi - maximum de vraisemblance

On ajuste θ pour que la log-vraisemblance soit maximale (en θ) :

$$L_n(\theta) := \prod_{i=1}^n p_{\theta}(\mathbf{x}_i), \quad \ell_n(\theta) := \log(L_n(\theta)),$$

$$f_{\mathbf{x}} \approx p_{\theta^{\text{MV}}}, \quad \theta^{\text{MV}} \in \arg \max_{\theta \in \Theta} \ell_n(\theta).$$

- $L_n(\theta)$ est la vraisemblance des données \mathcal{S}_n sachant le paramètre θ (qui s'écrit sous forme produit dans le cas de données indépendantes).
- l'estimateur du maximum de vraisemblance peut exister et être unique, ne pas être unique, ou ne pas exister.
- Sous des hypothèses de régularité sur \mathcal{P} , θ^{MV} converge vers θ^* , avec un risque quadratique en $1/n$, est asymptotiquement normal et efficace (i.e. on ne peut pas faire mieux asymptotiquement).
- Méthodes des moments et du max. de vrais. conduisent souvent à des

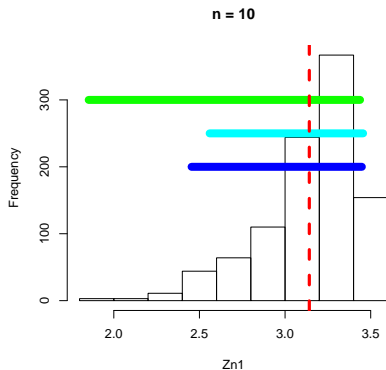
Méthode d'ajustement de familles de loi - exemple

Exercice :

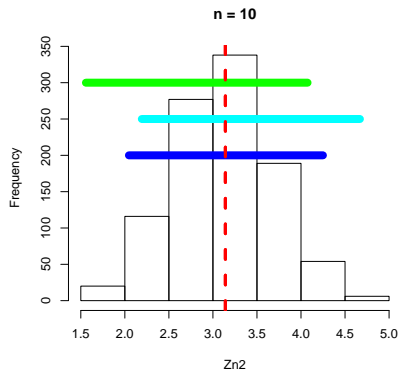
- On suppose que x suit une loi uniforme $\mathcal{U}(0, \theta)$.
- Proposer un estimateur de θ par la méthode des moments et par le maximum de vraisemblance.
- Comparer les estimateurs obtenus en terme de risque quadratique.

$X \sim \mathcal{U}(0, \theta)$ ($\theta = \pi$ pour l'illustration graphique)

- $Z_n^{(1)} = \frac{n+1}{n} \max_{1 \leq i \leq n} X_i$, $R(Z_n^{(1)}, \theta) = \frac{\theta^2}{n(n+2)}$
- $Z_n^{(2)} = 2\bar{X}_n$, $R(Z_n^{(2)}, \theta) = \frac{\theta^2}{3n}$.



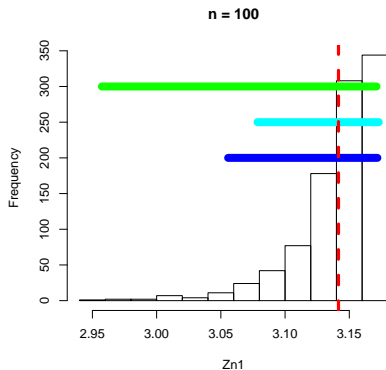
(a) $Z_n^{(1)}, n = 10$



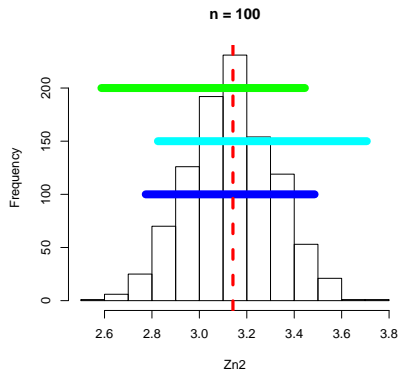
(b) $Z_n^{(2)}, n = 10$

$X \sim \mathcal{U}(0, \theta)$ ($\theta = \pi$ pour l'illustration graphique)

- $Z_n^{(1)} = \frac{n+1}{n} \max_{1 \leq i \leq n} X_i$, $R(Z_n^{(1)}, \theta) = \frac{\theta^2}{n(n+2)}$
- $Z_n^{(2)} = 2\bar{X}_n$, $R(Z_n^{(2)}, \theta) = \frac{\theta^2}{3n}$.



(c) $Z_n^{(1)}, n = 100$



(d) $Z_n^{(2)}, n = 100$

Méthode d'ajustement de familles de loi - approche bayésienne

- La méthode d'ajustement bayésienne est une extension de l'approche du max. de vrais. pour laquelle on suppose disposer d'une loi a priori sur θ .
- ⇒ on obtient alors un modèle **hiérarchique** (la loi de \mathbf{x} suit une loi donnée, paramétrée par θ qui suit lui même une loi donnée).

Hypothèses : $\mathbf{x}|\theta \sim p_\theta$, $\theta \sim f_\theta$.

Modèle déduit :

$$f_{\mathbf{x}}(\mathbf{x}) = \int_{\Theta} p_\theta(\mathbf{x}) f_{\theta|\mathcal{S}_n}(\theta) d\theta,$$

$$f_{\theta|\mathcal{S}_n}(\theta) \propto L_n(\theta) f_\theta(\theta) \quad (\text{formule de Bayes}).$$

Plan de la séance

- 1 Introduction
- 2 Maximum d'entropie
- 3 Méthode d'ajustement de familles de loi
- 4 Méthode à noyau**

Méthode à noyau - principe

$$f_{\mathbf{X}}(\mathbf{x}) \approx \frac{1}{N} \sum_{n=1}^N K(\mathbf{x} - \mathbf{X}(\theta_n))$$

Même si les approches à noyaux sont qualifiées de non-paramétriques, leur efficacité résulte de la bonne adéquation du noyau K à la distribution $f_{\mathbf{X}}$. Le noyau K doit ainsi être adapté au nombre de réalisations disponibles, N , ainsi qu'à la dimension de \mathbf{X} , d .

Exercice : proposer des fonctions K pour la reproduction de $f_{\mathbf{X}}$ dans le cas où X est une v.a. :

- uniforme $\mathcal{U} \{[0, 1]\}$,
- gaussienne $\mathcal{N}(0, 1)$.

Méthode à noyau et fonction caractéristique

- Par définition de la fonction caractéristique, pour tout $\mathbf{t} \in \mathbb{R}^d$:

$$\begin{aligned}\Phi_{\mathbf{X}}(\mathbf{t}) &:= \mathbb{E} [\exp(i \langle \mathbf{X}, \mathbf{t} \rangle_d)], \\ &= \int_{\mathbb{R}^d} \exp(i \langle \mathbf{x}, \mathbf{t} \rangle_d) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x},\end{aligned}$$

- On peut alors définir son approximation empirique :

$$\Phi_{\mathbf{X}}(\mathbf{t}) \approx \hat{\Phi}_{\mathbf{X}}(\mathbf{t}) := \frac{1}{N} \sum_{n=1}^N \exp(i \langle \mathbf{X}(\omega_n), \mathbf{t} \rangle_d), \quad \mathbf{t} \in \mathbb{R}^d.$$

Méthode à noyau et fonction caractéristique

$f_{\mathbf{X}}$ étant la transformée de Fourier inverse de $\Phi_{\mathbf{X}}$, l'idée des approches à noyaux repose sur l'approximation suivante :

$$f_{\mathbf{X}}(\mathbf{x}) \approx \hat{f}_{\mathbf{X}}(\mathbf{x}; h, \mathcal{S}(N)),$$

$$\begin{aligned}\hat{f}_{\mathbf{X}}(\mathbf{x}; h, \mathcal{S}(N)) &\propto \int_{\mathbb{R}^d} \hat{\Phi}_{\mathbf{X}}(\mathbf{t}) \exp\left(-\frac{1}{2}h^2 \mathbf{t}^T [\hat{R}_{\mathbf{X}}] \mathbf{t}\right) \exp(-i \langle \mathbf{x}, \mathbf{t} \rangle_d) d\mathbf{t}, \\ &= \frac{1}{N} \sum_{n=1}^N \phi\left(\mathbf{x}; \mathbf{X}(\omega_n), h^2 [\hat{R}_{\mathbf{X}}]\right).\end{aligned}$$

où h est un paramètre de lissage, et où $\phi(\mathbf{x}; \boldsymbol{\mu}, [C])$ désigne la PDF gaussienne multidimensionnelle de moyenne $\boldsymbol{\mu}$ et de covariance $[C]$.

Exercice : montrer l'égalité précédente.

Méthode à noyau - cas gaussien

$$f_{\mathbf{X}}(\mathbf{x}) \approx \hat{f}_{\mathbf{X}}(\mathbf{x}; h, \mathcal{S}(N)) := \frac{1}{N} \sum_{n=1}^N \phi\left(\mathbf{x}; \mathbf{X}(\omega_n), h^2[\hat{R}_{\mathbf{X}}]\right)$$

- Par construction, $f_{\mathbf{X}}$ est approchée comme une somme de N PDF gaussiennes centrées au niveau des réalisations de \mathbf{X} .
- La covariance de ces N gaussiennes est égale à $h^2 \times [\hat{R}_{\mathbf{X}}]$.

Exercice : commenter l'influence de h sur la construction de $f_{\mathbf{X}}$.

Méthode à noyau - propriété

Si,

- $[\hat{R}_X]^{1/2}$ est une matrice symétrique telle que $[\hat{R}_X]^{1/2}[\hat{R}_X]^{1/2} = [\hat{R}_X]$,
- Q est une v.a. uniformément distribuée sur $\{1, \dots, N\}$,
- ξ est un vecteur Gaussian centré réduit de composantes indépendantes,
- Q et ξ sont indépendants,

alors la PDF du vecteur

$$\hat{\mathbf{X}}(Q, \xi) := \mathbf{X}(\omega_Q) + h \times [\hat{R}_X]^{1/2} \xi$$

est égale à $\hat{f}_X(\cdot; h, \mathcal{S}(N))$.

Exercices : montrer la propriété précédente. En déduire comment générer des réalisations indépendantes de $\hat{\mathbf{X}}(Q, \xi)$.

Méthode à noyau - propriétés

$$\hat{\mathbf{X}}(Q, \xi) := \mathbf{X}(\omega_Q) + h \times [\hat{R}_{\mathbf{X}}]^{1/2} \xi$$

Exercice : calculer la moyenne et la covariance de $\hat{\mathbf{X}}(Q, \xi)$.

Si la PDF de $\tilde{\mathbf{X}}$ est

$$\begin{aligned} \tilde{f}_{\mathbf{X}}(\cdot; h, \mathcal{S}(N)) &:= \frac{1}{N} \sum_{n=1}^N \phi(\mathbf{x}; \alpha \mathbf{X}(\omega_n) + \beta, [Q]), \\ \alpha^2 &:= (1 - h^2) \frac{N}{N-1}, \quad \beta := (1 - \alpha) \hat{\mu}, \quad [Q] := h^2 [\hat{R}_{\mathbf{X}}], \end{aligned}$$

alors les moyenne et covariance de $\tilde{\mathbf{X}}$ sont égales à $\hat{\mu}$ et $[\hat{R}_{\mathbf{X}}]$ respectivement.

Exercice : montrer la propriété précédente.

Méthode à noyau - choix de la valeur de h

- La valeur de h doit être optimisée pour minimiser la différence entre $f_{\mathbf{X}}$ and $\widehat{f}_{\mathbf{X}}(\cdot; h, \mathcal{S}(N))$. Le critère basé sur l'erreur intégrée (MISE) est généralement utilisé pour caractériser cette différence :

$$\text{MISE}(h; d, N) = \mathbb{E}_{\mathcal{S}(N)} \left[\int_{\mathbb{R}^d} \left(f_{\mathbf{X}}(\mathbf{x}) - \widehat{f}_{\mathbf{X}}(\mathbf{x}; h, \mathcal{S}(N)) \right)^2 d\mathbf{x} \right].$$

- Sous certaines conditions de régularité sur $f_{\mathbf{X}}$, ce critère est remplacé par le critère asymptotique (AMISE) suivant :

$$\text{MISE}(h; d, N) = \text{AMISE}(h; d, N) + o(h^4 + N^{-1}h^{-d}),$$

$$\text{AMISE}(h; d, N) = \frac{h^4}{4} f_1(f_{\mathbf{X}}, [\nabla^2 f_{\mathbf{X}}]) + \frac{h^{-d}}{Nd} f_2(f_{\mathbf{X}}),$$

où f_1 et f_2 sont deux fonctions qui ne dépendent pas de h mais dépendent de $f_{\mathbf{X}}$ et de sa matrice hessienne $[\nabla^2 f_{\mathbf{X}}]$.

Méthode à noyau - Choix de la valeur de h

$$\text{AMISE}(h; d, N) = \frac{h^4}{4} f_1(f_{\mathbf{X}}, [\nabla^2 f_{\mathbf{X}}]) + \frac{h^{-d}}{Nd} f_2(f_{\mathbf{X}})$$

- En effet, le critère AMISE peut être explicitement minimisé par rapport à h :

$$\begin{aligned} h^{\text{AMISE}}(d, N) &:= \left(\frac{1}{N} \frac{f_2(f_{\mathbf{X}})}{f_1([\nabla^2 f_{\mathbf{X}}], f_{\mathbf{X}})} \right)^{1/(d+4)} \\ &= \arg \min_{h>0} \text{AMISE}(h; d, N). \end{aligned}$$

Méthode à noyau - Choix de la valeur de h

- Sous une hypothèse gaussienne pour $f_{\mathbf{X}}$, on peut montrer que :

$$h^{\text{AMISE}}(d, N) \approx h^{\text{Silv}}(d, N) := \left(\frac{1}{N} \frac{4}{(d+2)} \right)^{1/(d+4)},$$

où $h^{\text{Silv}}(d, N)$ est appelée "longueur de bande de Silverman".

Exercice : calculer la limite de $h^{\text{Silv}}(d, N)$ quand $N = q \times d$ et d tend vers $+\infty$. Qu'implique une telle valeur pour la longueur de bande ?

Méthode à noyau - Choix de la valeur de h

- La valeur de $h^{\text{Silv}}(d, N)$ ne dépend que de d et N , et n'est plus véritablement adaptée à $f_{\mathbf{X}}$.
- Pour des PDFs concentrées sur des sous espaces de \mathbb{R}^d , ou présentant des structures de dépendances particulières, cette valeur a souvent tendance à sur-évaluer la dispersion de $f_{\mathbf{X}}$.

⇒ d'autres approches, basées notamment sur la vraisemblance $\mathcal{L}(\mathcal{S}(N)|h)$ de h pour expliquer $\mathcal{S}(N)$, ont ainsi été proposées,

$$\mathcal{L}(\mathcal{S}(N)|h) := \prod_{n=1}^N \hat{f}_{\mathbf{X}}(\mathbf{X}(\omega_n); h, \mathcal{S}(N)) = \frac{1}{N^N} \prod_{n=1}^N \sum_{m=1}^N \phi_{n,m}(h),$$

$$\phi_{n,m}(h) := \phi\left(\mathbf{X}(\omega_n); \mathbf{X}(\omega_m), h^2[\hat{R}_{\mathbf{X}}]\right), \quad 1 \leq n, m \leq N.$$

Problème : pour N fixé, calculer $\lim_{h \rightarrow 0} \mathcal{L}(\mathcal{S}(N)|h)$.

Méthode à noyau - Choix de la valeur de h

- La fonction $\mathcal{L}(\mathcal{S}(N)|h)$ utilise deux fois la même information (pour calculer $\hat{f}_{\mathbf{X}}(\cdot; h, \mathcal{S}(N))$ et pour l'évaluer).
- Afin d'éviter ce phénomène, il est alors préférable de s'orienter vers une vraisemblance "LOO", $\mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|h)$, telle que :

$$\mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|h) := \frac{1}{(N-1)^N} \prod_{n=1}^N \sum_{m \neq n} \phi_{n,m}(h).$$

- En se donnant cette approximation de la vraisemblance par validation croisée, et une distribution a priori pour h , f_h , on déduit :

$$f_h(h|\mathcal{S}(N)) \propto \mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|h)f_h(h), \quad h > 0.$$

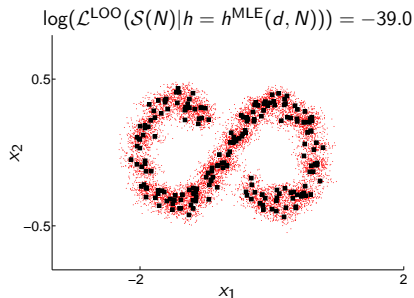
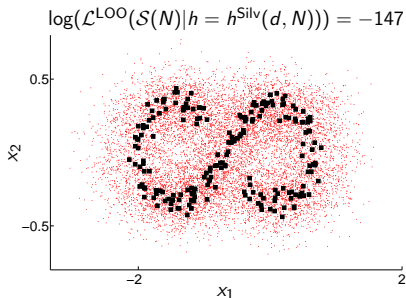
- En pratique, la fonction $\mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|h)$ est souvent très piquée, et pour des valeurs de $d > 1$, un bon compromis est donné par :

$$h^{\text{MLE}}(d, N) := \arg \max_{h>0} \mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|h).$$

Méthode à noyau - exemples

Illustration en 2D

- Points noirs : information disponible.
- Points rouges : nouveaux points générés.

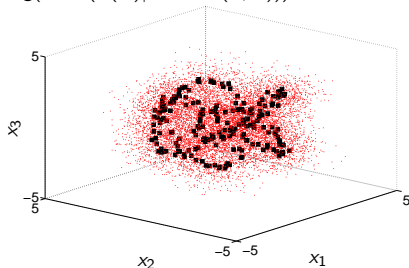


Méthode à noyau - exemples

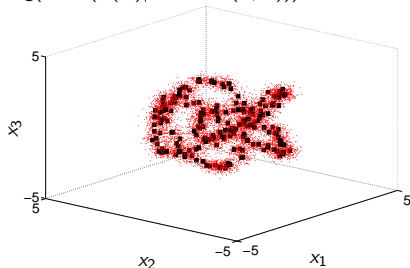
Illustration en 3D

- Points noirs : information disponible.
- Points rouges : nouveaux points générés.

$$\log(\mathcal{L}^{\text{LOO}}(S(N)|h = h^{\text{Silv}}(d, N))) = -1.102 \times 10^3$$



$$\log(\mathcal{L}^{\text{LOO}}(S(N)|h = h^{\text{MLE}}(d, N))) = -8.00 \times 10^2$$



Méthode à noyau - remarques finales

- La méthode à noyau peut être utilisée pour tout type de loi, mais il est particulièrement recommandé lorsqu'il est difficile d'intuiter une forme paramétrique pour les CDFs et surtout le copule de \mathbf{X} .
- Pour être efficace, cette technique nécessite des valeurs de N relativement grandes devant d .
- L'approche par maximum de vraisemblance LOO peut être généralisée à n'importe quel noyau (ici seul le noyau gaussien a été détaillé).
- Lorsque d devient grand, il est souvent pertinent de chercher à diviser \mathbf{X} en blocs indépendants avant d'effectuer cette modélisation.
- Cette approche peut être vue comme une généralisation continue des histogrammes, en optimisant non plus le nombre de seuils, mais les paramètres du noyau, pour maximiser l'objectivité de l'objet renvoyé.

Plan de la séance

- 1 Introduction
- 2 Maximum d'entropie
- 3 Méthode d'ajustement de familles de loi
- 4 Méthode à noyau