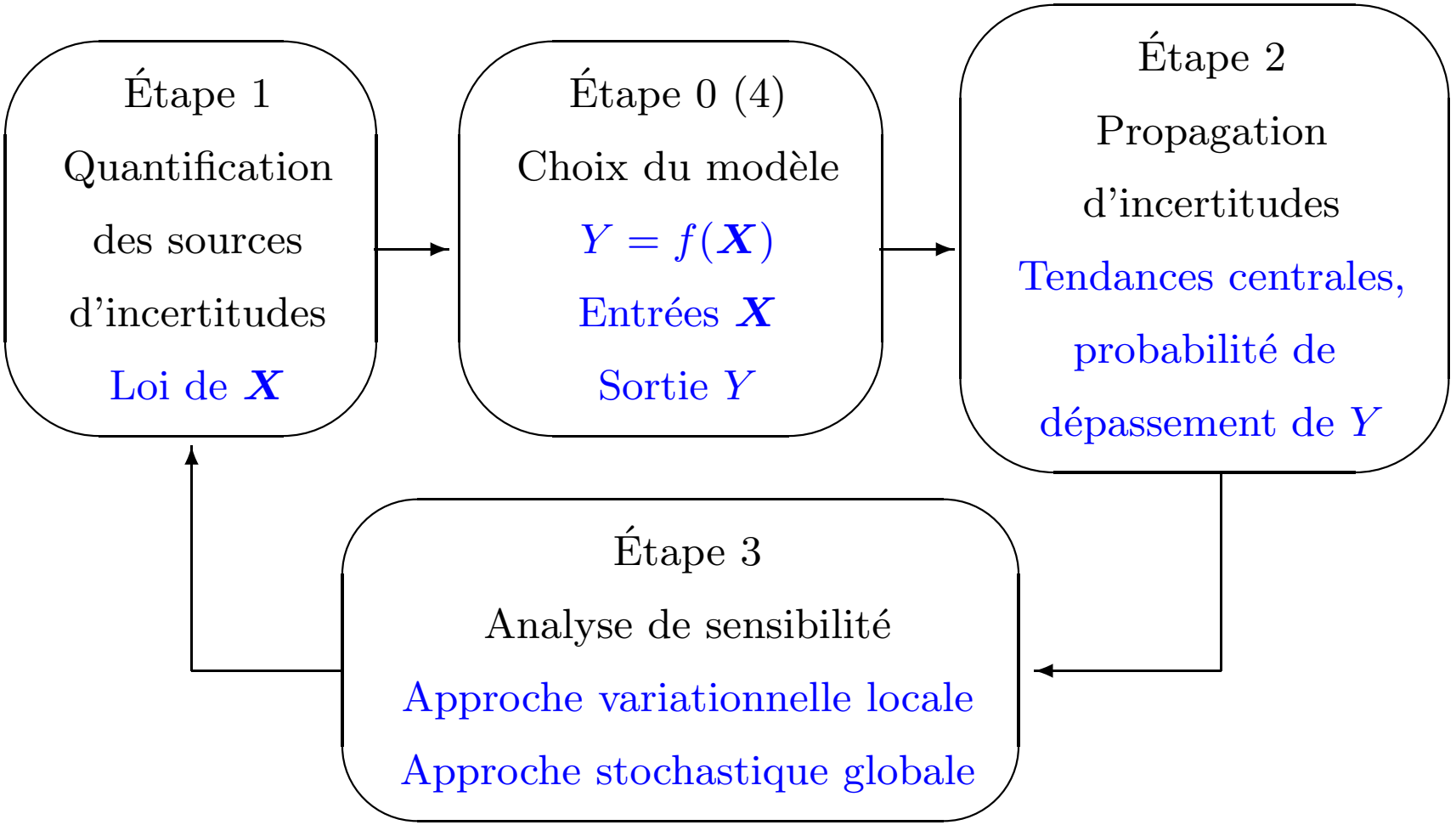
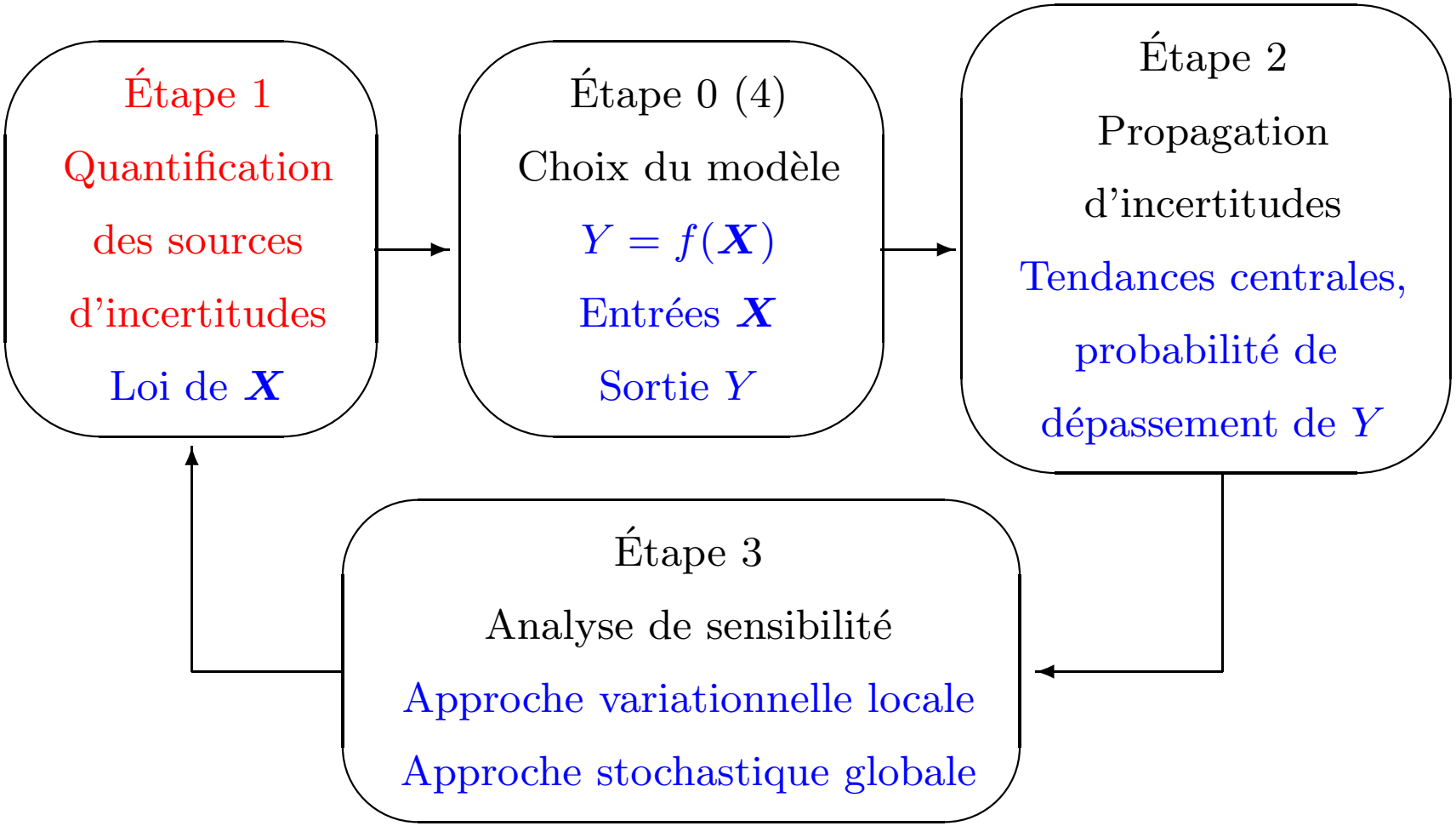
Traitement des incertitudes

Josselin Garnier (Ecole Polytechnique)

• Problème général :
Comment modéliser les incertitudes et leur propagation dans les modèles physiques ou numériques ?
Comment estimer (quantifier) la dispersion de la sortie d'un code ou d'une expérience en fonction de la dispersion des paramètres d'entrée ?
Comment estimer (quantifier) la sensibilité de la sortie d'un code ou d'une expérience vis-à-vis d'un paramètre d'entrée particulier ?
Comment construire un métamodèle de la sortie d'un code ?

• But de l'exposé : faire un rapide état de l'art des différentes approches.


Étape 1
Quantification
des sources
d'incertitudes
Loi de $\boldsymbol{X}$
Étape 0 (4)
Choix du modèle
$Y = f(\boldsymbol{X})$
Entrées $\boldsymbol{X}$
Sortie Y
Étape 2
Propagation
d'incertitudes
Tendances centrales,
probabilité de
dépassement de Y
Étape 3
Analyse de sensibilité
Approche variationnelle locale
Approche stochastique globale


Étape 1
Quantification
des sources
d'incertitudes
Loi de $\boldsymbol{X}$
Étape 0 (4)
Choix du modèle
$Y = f(\boldsymbol{X})$
Entrées $\boldsymbol{X}$
Sortie Y
Étape 2
Propagation
d'incertitudes
Tendances centrales,
probabilité de
dépassement de Y
Étape 3
Analyse de sensibilité
Approche variationnelle locale
Approche stochastique globale

Partie I. Les sources d'incertitudes

- Deux types d'entrée "incertaines" :

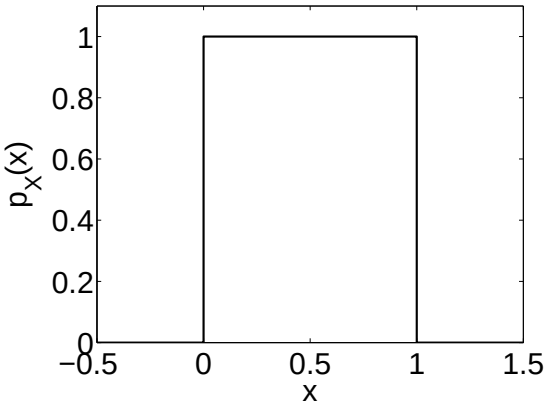
- variables stochastiques : ces variables ont une variabilité naturelle résultant de phénomènes aléatoires (typiquement, une quantité soumise à des fluctuations dans un procédé de fabrication).

- variables épistémiques : ces variables possèdent une valeur mais elle nous est inconnue, à cause d'un manque de connaissance (typiquement, une constante d'une loi physique).

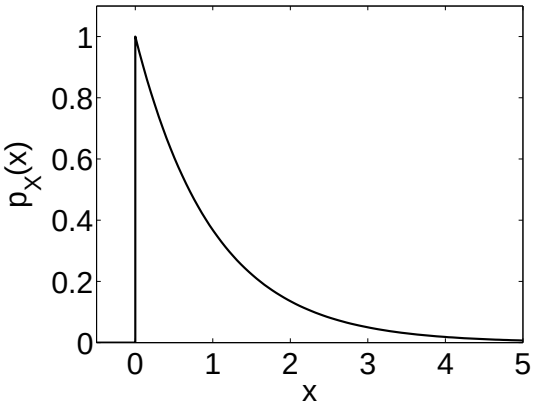
• La modélisation par des lois de probabilités : les variables d'entrées sont traitées comme des variables aléatoires, de lois de probabilités données.

• La loi de probabilité d'une variable aléatoire réelle X caractérise la probabilité $\mathbb{P}(X \in [a, b])$ pour tout $a < b$.
Dans le cas d'une variable aléatoire continue (prend ses valeurs sur $\mathbb{R}$ ou un intervalle de $\mathbb{R}$) : la loi est donnée par la densité $(p(x))_{x \in \mathbb{R}}$

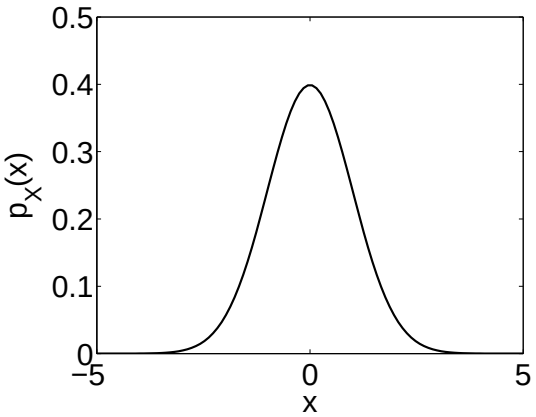
$$\mathbb{P}(X \in [a, b]) = \int_a^b p(x)dx, \qquad \mathbb{P}(X \in [x, x + \delta x]) \simeq p(x)\,\delta x$$



uniforme



exponentielle



gaussienne

• Les variables d'entrées sont traitées comme des variables aléatoires, de lois de probabilités données.

• Pour déterminer la densité de la loi d'une variable aléatoire :
- méthodes non-paramétriques à noyaux,
- méthodes paramétriques d'ajustement à une loi analytique,
- méthodes entropiques.

• Extensions : modèles hiérarchiques (classiques en analyse bayésienne).

• Remarque : Autres modélisations possibles :
- distributions de possibilité (aussi appelées intervalles flous),
- intervalles aléatoires utilisant les fonctions de croyance de Dempster-Shafer.

Cf: T. Hastie, R. Tibshirani et J. Friedman, The Elements of Statistical Learning, Springer, 2001.

- Méthodes à noyaux. Aussi appelée méthode de Parzen-Rozenblatt.
- Méthode non-paramétrique d'estimation de la densité d'une variable aléatoire.

- se base sur un échantillon $(x_i)_{i=1,\ldots,n}$ de réalisations indépendantes de la variable (on suppose la variable d'entrée de dimension un).

- nécessite un nombre suffisant de données.

- permet d'estimer la densité en tout point du support.

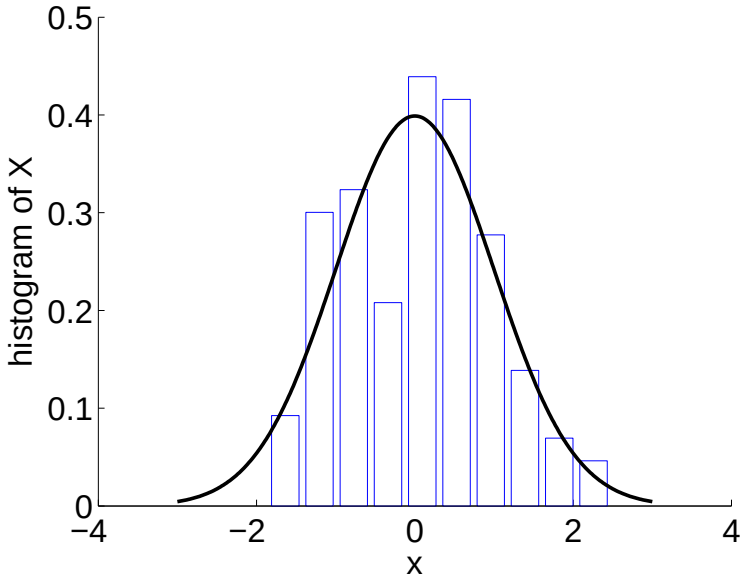
- généralise la méthode d'estimation par un histogramme.

- Formule

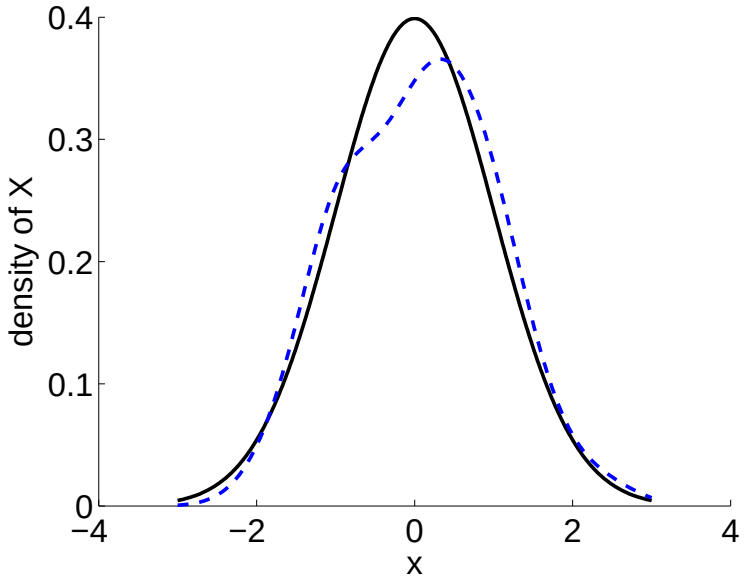
$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big)$$

K est le noyau (souvent gaussien), h la fenêtre (régit le niveau de lissage).

• Exemple. Soit un échantillon de 200 données tirées selon une loi gaussienne :



histogramme



méthode à noyaux

• Résultats :

- Il n'existe pas d'estimateur non-paramétrique qui converge plus vite que l'estimateur à noyaux (avec le choix de la fenêtre $h \simeq n^{-1/5}$).

- La vitesse de convergence ($n^{-4/5}$ pour le risque quadratique) est plus faible que la vitesse typique des méthodes paramétriques (n^{-1}, voire plus rapide encore).

• Méthodes d'ajustement à une loi analytique.

Etant donné un échantillon $(x_i)_{i=1,\ldots,n}$ de réalisations indépendantes de la variable.

On cherche à ajuster les paramètres d'une famille paramétrique de lois de probabilités aux données.
Exemple : famille de loi gaussiennes de densités de probabilités

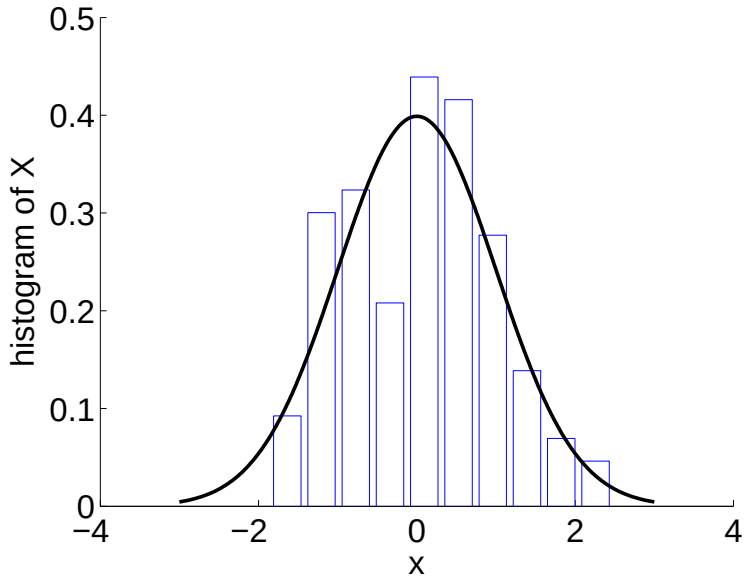
$$p_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2} \Big)$$

(gaussienne de moyenne μ et de variance σ^2).
Nécessite un peu moins de données $(x_i)_{i=1,\ldots,n}$ que la méthode à noyaux.

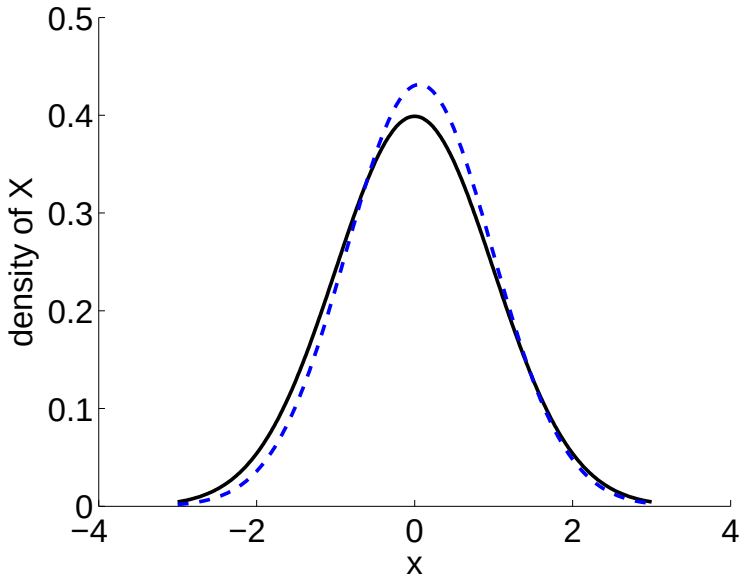
• Méthode des moments : on ajuste les paramètres de la loi analytique pour que ses premiers moments correspondent aux moments empiriques de l'échantillon.
Exemple :

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

Exemple. Soit un échantillon de 200 données tirées selon une loi gaussienne :



histogramme



méthode des moments

Résultat : risque quadratique converge en $1/n$ si la famille paramétrique contient la loi originale.

• Méthode du maximum de vraisemblance : on ajuste les paramètres $\boldsymbol{\theta}$ de la loi analytique pour que la log-vraisemblance soit maximale (en $\boldsymbol{\theta}$):

$$L_{\boldsymbol{x}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} p_{\boldsymbol{\theta}}(x_i), \qquad \ln L_{\boldsymbol{x}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln p_{\boldsymbol{\theta}}(x_i)$$

- Il y a un théorème de Bayes la-dessous : $L_{\boldsymbol{x}}(\boldsymbol{\theta})$ est la vraisemblance des données $\boldsymbol{x} = (x_i)_{i=1,\ldots,n}$ sachant le paramètre $\boldsymbol{\theta}$, et est aussi la vraisemblance du paramètre $\boldsymbol{\theta}$ sachant les données $\boldsymbol{x}$ sans a priori sur $\boldsymbol{\theta}$.
- L'estimateur du maximum de vraisemblance peut exister et être unique, ne pas être unique, ou ne pas exister.
- Dans les bons cas, il converge, son risque quadratique est en $1/n$, il est asymptotiquement normal, asymptotiquement efficace (i.e., on ne peut pas faire mieux, asymptotiquement).
- Exemple : $\boldsymbol{\theta} = (\mu, \sigma^2)$:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

- Possibilité de prendre en compte un a priori (avis d'expert ou étude antérieure) sur $\boldsymbol{\theta}$ ($\rightarrow$ méthode du maximum a posteriori).

• Méthode bayésienne : fournit une distribution sur les paramètres de la loi (et pas seulement une estimation); permet d'incorporer une idée a priori sur les paramètres; on obtient un modèle hiérarchique.

Exemple : $\boldsymbol{x} = (x_k)_{k=1,\dots,n}$ échantillon de loi $\mathcal{N}(\mu, \sigma^2_{\text{mes}})$ avec μ inconnu et σ_{mes} connu :

$$p(\boldsymbol{x}|\mu) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_{\text{mes}}} \exp\Big(- \frac{(x_k - \mu)^2}{2\sigma^2_{\text{mes}}}\Big)$$

$$\hookrightarrow \qquad p_{\text{post}}(\mu|\boldsymbol{x}) \quad \approx \quad \exp\Big(- \sum_{k=1}^{n} \frac{(x_k - \mu)^2}{2\sigma^2_{\text{mes}}}\Big) \approx \exp\Big(- \frac{(\mu - \frac{1}{n}\sum_{k=1}^{n} x_k)^2}{2\frac{\sigma^2_{\text{mes}}}{n}}\Big)$$

la loi a posteriori de μ, étant données les observations $\boldsymbol{x}$, est

$$\mathcal{N}\Big(\frac{1}{n}\sum_{k=1}^{n} x_k, \frac{\sigma^2_{\text{mes}}}{n}\Big)$$

A posteriori :

$$\mathbb{E}[f(X)] = \int \mathbb{E}_{\mathcal{N}(\mu,\sigma^2_{\text{mes}})}[f(X)] \frac{1}{\sqrt{2\pi\frac{\sigma^2_{\text{mes}}}{n}}} \exp\Big(- \frac{(\mu - \frac{1}{n}\sum_{k=1}^{n} x_k)^2}{2\frac{\sigma^2_{\text{mes}}}{n}}\Big) d\mu$$

• Test de qualité d'ajustement permet de dire si les données $(x_i)_{i=1,\dots,n}$ sont compatibles avec la famille paramétrique $p_{\boldsymbol{\theta}}$.
Attention avec les tests : la réponse du test est "je ne peux pas rejeter l'hypothèse que les données sont compatibles avec la famille de lois" ou "je rejette l'hypothèse que les données sont compatibles avec la famille de lois".

• Le choix de la famille est un jugement d'expert; peut-être motivée par des considérations théoriques, un théorème central limite pour justifier une gaussienne, un principe d'entropie, ...

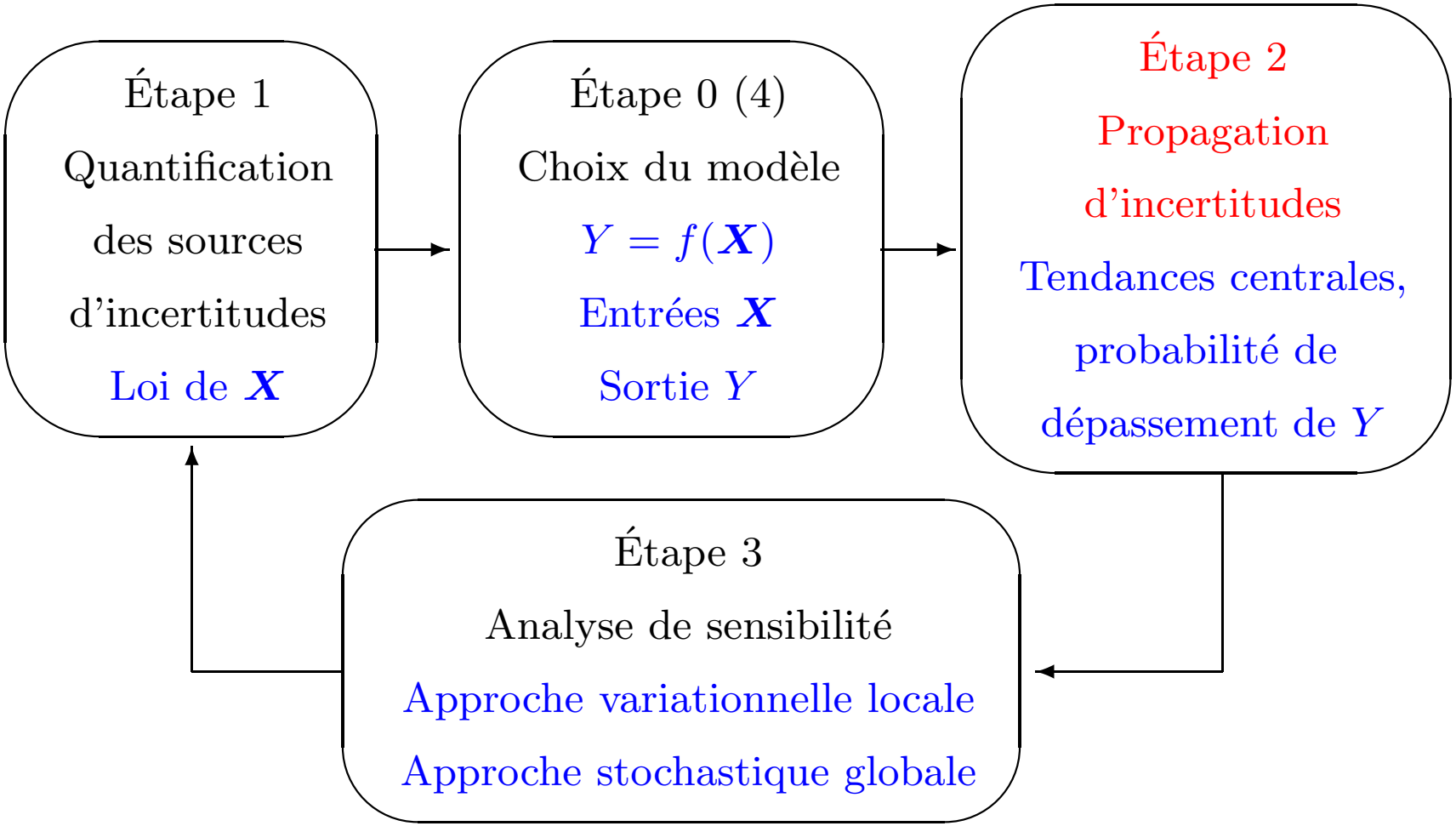
• Méthode du maximum d'entropie.

On choisit la loi (densité $p(x)$) qui maximise l'entropie (le manque d'information)

$$E(p) = -\int p(x) \ln p(x) dx$$

étant données certaines informations (moyenne, variance, intervalles de définition).
Exemples :
- la loi qui maximise l'entropie d'une variable aléatoire de moyenne et de variance prescrites est une loi gaussienne.
- la loi qui maximise l'entropie d'une variable aléatoire à valeurs dans un intervalle prescrit est une loi uniforme.
- la loi qui maximise l'entropie d'une variable aléatoire à valeurs positives et à moyenne prescrite est une loi exponentielle.

Joue un rôle dans les méthodes bayésiennes pour définir les lois a priori.


Étape 1
Quantification
des sources
d'incertitudes
Loi de $\boldsymbol{X}$
Étape 0 (4)
Choix du modèle
$Y = f(\boldsymbol{X})$
Entrées $\boldsymbol{X}$
Sortie Y
Étape 2
Propagation
d'incertitudes
Tendances centrales,
probabilité de
dépassement de Y
Étape 3
Analyse de sensibilité
Approche variationnelle locale
Approche stochastique globale

Partie II. Propagation d'incertitudes

- Contexte : code de calcul informatique ou expérience modélisé par

$$Y = f(\boldsymbol{X})$$

avec Y =variable de sortie

$\boldsymbol{X} = (X_i)_{i=1,\dots,d}$ variables d'entrée

f = boîte noire déterministe

On se donne : la loi de $\boldsymbol{X}$.

- But : estimation d'une quantité

$$\mathbb{E}[\psi(Y)]$$

avec une barre d'erreur et le minimum de simulations/expériences.

- Exemples (pour Y à valeurs réelles) :

$\psi(y) = y \rightarrow$ moyenne de Y, i.e. $\mathbb{E}[Y]$

$\psi(y) = y^2 \rightarrow$ variance de Y, i.e. $\mathrm{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$

$\psi(y) = \mathbf{1}_{[a,\infty)}(y) \rightarrow$ probabilité de dépasser le seuil a, i.e. $\mathbb{P}(Y \geq a)$.

Méthode analytique

- La quantité à estimer est une intégrale d-dimensionnelle :

$$I = \mathbb{E}[\psi(Y)] = \mathbb{E}[F(\boldsymbol{X})] = \int_{\mathbb{R}^d} F(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}$$

où $p(\boldsymbol{x})$ est la densité de probabilité de $\boldsymbol{X}$ et $F(\boldsymbol{x}) = \psi(f(\boldsymbol{x}))$.

Dans des cas simples (quand la densité p et la fonction F sont connues de manière analytique), on peut parfois évaluer l'intégrale de manière exacte.

Situation "exceptionnelle".

Méthodes de quadrature

• La quantité à estimer est une intégrale d-dimensionnelle :

$$I = \mathbb{E}[\psi(Y)] = \mathbb{E}[F(\boldsymbol{X})] = \int_{\mathbb{R}^d} F(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}$$

où $p(\boldsymbol{x})$ est la densité de probabilité de $\boldsymbol{X}$ et $F(\boldsymbol{x}) = \psi(f(\boldsymbol{x}))$.

Si $p(\boldsymbol{x}) = \prod_{i=1}^{d} p_0(x_i)$, alors on peut appliquer une méthode de quadrature gaussienne avec grille pleine de n^d points :

$$\hat{I} = \sum_{j_1=1}^{n} \cdots \sum_{j_d=1}^{n} \rho_{j_1} \cdots \rho_{j_d} F(\xi_{j_1}, \ldots, \xi_{j_d})$$

avec $(\rho_j)_{j=1,\ldots,n}$ poids et $(\xi_j)_{j=1,\ldots,n}$ nœuds de quadrature gaussienne associée à la densité p_0.

Il existe aussi des méthodes de quadrature avec des grilles creuses (Smolyak).

• Les méthodes de quadrature sont efficaces avec :
- des conditions de régularité sur $\boldsymbol{x} \to F(\boldsymbol{x})$ (et pas seulement f),
- une dimension d petite (même avec des grilles creuses de type Smolyak).
Elles réclament sinon beaucoup d'appels au code.

Méthode de Monte Carlo

On tire un n-échantillon $(\boldsymbol{X}^{(k)})_{k=1,\dots,n}$ (des réalisations indépendantes de $\boldsymbol{X}$).
Estimateur de $I = \mathbb{E}[F(\boldsymbol{X})]$:

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^{n} F(\boldsymbol{X}^{(k)})$$

Estimation non-biaisée :

$$\mathbb{E}[\hat{I}_n] = I \qquad \text{pour tout } n$$

Convergence :

$$\hat{I}_n \overset{n\to\infty}{\longrightarrow} I \quad \text{avec probabilité 1}$$

Erreur (risque quadratique) :

$$\mathbb{E}\big[(\hat{I}_n - I)^2\big] = \mathrm{Var}(\hat{I}_n) = \frac{1}{n}\mathrm{Var}(F(\boldsymbol{X}))$$

Erreur relative :

$$\frac{\mathbb{E}\big[(\hat{I}_n - I)^2\big]^{1/2}}{I} = \frac{1}{\sqrt{n}} \frac{\mathrm{Var}(F(\boldsymbol{X}))^{1/2}}{\mathbb{E}[F(\boldsymbol{X})]}$$

Caractéristiques :

1) possibilité d'obtenir des intervalles de confiance,

2) pas de régularité requise sur F (avec $F(\boldsymbol{x}) = \psi(f(\boldsymbol{x}))$),

3) vitesse de convergence indépendante de la dimension (mais lente).

Intervalles de confiance

Question : A partir de l'échantillon $(\boldsymbol{X}^{(k)})_{k=1,\dots,n}$, l'estimateur $\hat{I}_n$ donne une valeur approchée de I, d'autant meilleure que n est grand. Comment quantifier précisément l'erreur ?

Réponse : On construit un intervalle de confiance au niveau 0.95, i.e. un intervalle $[\hat{a}_n, \hat{b}_n]$ tel que

$$\mathbb{P}\left(I \in [\hat{a}_n, \hat{b}_n]\right) \geq 0.95$$

Construction basée sur le théorème central limite :

$$\mathbb{P}\left(\left|\hat{I}_n - I\right| < c\frac{\mathrm{Var}(F(\boldsymbol{X}))^{1/2}}{\sqrt{n}}\right) \overset{n\to\infty}{\longrightarrow} \frac{2}{\sqrt{2\pi}}\int_0^c e^{-x^2/2}dx$$

Le membre de droite vaut 0.95 si $c = 1.96$. Donc

$$\mathbb{P}\left(I \in \left[\hat{I}_n - 1.96\frac{\mathrm{Var}(F(\boldsymbol{X}))^{1/2}}{\sqrt{n}}, \hat{I}_n + 1.96\frac{\mathrm{Var}(F(\boldsymbol{X}))^{1/2}}{\sqrt{n}}\right]\right) \simeq 0.95$$

$$\mathbb{P}\left(I \in \left[\hat{I}_n - 1.96\frac{\mathrm{Var}(F(\boldsymbol{X}))^{1/2}}{\sqrt{n}}, \hat{I}_n + 1.96\frac{\mathrm{Var}(F(\boldsymbol{X}))^{1/2}}{\sqrt{n}}\right]\right) \simeq 0.95$$

Les bornes sont encore inconnues car on ne connaît pas $\mathrm{Var}(F(\boldsymbol{X}))$! Deux solutions :

- solution conservatrice, du type $\mathrm{Var}(F(\boldsymbol{X})) \leq \|F\|_\infty^2$, et alors

$$\mathbb{P}\left(I \in \left[\hat{I}_n - 1.96\frac{\|F\|_\infty}{\sqrt{n}}, \hat{I}_n + 1.96\frac{\|F\|_\infty}{\sqrt{n}}\right]\right) \geq 0.95$$

- asymptotiquement, on remplace $\mathrm{Var}(F(\boldsymbol{X}))$ dans les bornes par son estimateur empirique $\hat{\sigma}_n^2$:

$$\mathbb{P}\left(I \in \left[\hat{I}_n - 1.96\frac{\hat{\sigma}_n}{\sqrt{n}}, \hat{I}_n + 1.96\frac{\hat{\sigma}_n}{\sqrt{n}}\right]\right) \simeq 0.95$$

où

$$\hat{\sigma}_n = \left(\frac{1}{n}\sum_{k=1}^{n} F(\boldsymbol{X}^{(k)})^2 - \hat{I}_n^2\right)^{1/2}$$

Conclusion : Il n'y a aucun intervalle borné de $\mathbb{R}$ dont on puisse dire avec certitude qu'il contient I, mais il y a des intervalles, dits intervalle de confiance, dont on peut dire qu'ils contiennent I avec une probabilité proche de 1.

Propagation d'incertitudes par métamodèles

On remplace f par un métamodèle (modèle réduit) f_r et on applique une des techniques précédentes (analytique, quadrature, Monte Carlo).

$\rightarrow$ On peut faire beaucoup d'appels au métamodèle.

$\rightarrow$ Le choix du métamodèle est critique.

$\rightarrow$ Le contrôle de l'erreur n'est pas simple.

Développements de Taylor

- On approche la sortie $Y = f(\boldsymbol{X})$ par un développement de Taylor $Y_r = f_r(\boldsymbol{X})$.

- Exemple :

- On souhaite estimer $\mathbb{E}[Y]$ et $\mathrm{Var}(Y)$ pour $Y = f(\boldsymbol{X})$ avec X_i décorrélées, $\mathbb{E}[X_i] = \mu_i$ et $\mathrm{Var}(X_i) = \sigma_i^2$ connus, σ_i^2 petits.
- On approche $Y = f(\boldsymbol{X})$ par $Y_r = f_r(\boldsymbol{X}) = f(\boldsymbol{\mu}) + \nabla f(\boldsymbol{\mu}) \cdot (\boldsymbol{X} - \boldsymbol{\mu})$. On trouve :

$$\mathbb{E}[Y] \simeq \mathbb{E}[Y_r] = f(\boldsymbol{\mu}), \qquad \mathrm{Var}(Y) \simeq \mathrm{Var}(Y_r) = \sum_{i=1}^{d} \partial_{x_i} f(\boldsymbol{\mu})^2 \sigma_i^2$$

On a juste besoin de calculer $f(\boldsymbol{\mu})$ et $\nabla f(\boldsymbol{\mu})$ (calcul du gradient par différences finies ou par différenciation automatique, donc en gros $d+1$ appels à f).

- Rapide, analytique, permet de calculer approximativement des tendances centrales de la sortie (moyenne, variance).

- Convenable pour des petites variations des paramètres d'entrée et un modèle régulier (qu'on peut linéariser).

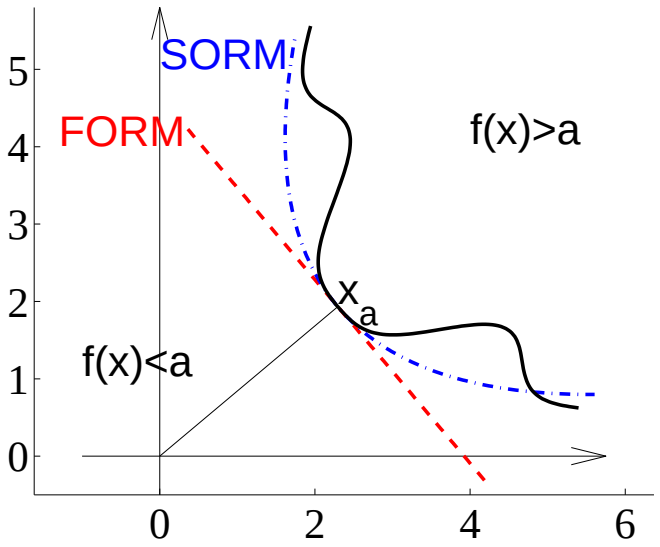
- Approche "locale". En général, pas de contrôle de l'erreur.

FORM-SORM

$$P = \mathbb{P}(f(\boldsymbol{X}) \geq a) = \mathbb{P}(\boldsymbol{X} \in \mathcal{F}) = \int_{\mathcal{F}} p(\boldsymbol{x}) d\boldsymbol{x}, \qquad \mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^d,\ f(\boldsymbol{x}) \geq a\}$$

- Méthode FORM-SORM, analytique mais approchée, sans contrôle d'erreur.
- on suppose que les X_i sont indépendants et de loi gaussienne de moyenne zéro et de variance un (ou on se ramène à ce cas par transformation isoprobabiliste).
- on trouve par optimisation (sous contrainte) le point $\boldsymbol{x}_a$ de dépassement de seuil (i.e. $f(\boldsymbol{x}_a) = a$) le plus proche de l'origine.
- on approche la surface de défaillance $\{\boldsymbol{x} \in \mathbb{R}^d, f(\boldsymbol{x}) = a\}$ par une surface régulière $\hat{\mathcal{F}}$ qui permette de faire un calcul analytique $\hat{P} = \int_{\hat{\mathcal{F}}} p(\boldsymbol{x}) d\boldsymbol{x}$:

- un hyperplan pour FORM
(et alors $\hat{P} = \frac{1}{2}\mathrm{erfc}\left(\frac{|\boldsymbol{x}_a|}{\sqrt{2}}\right)$),
- une forme quadratique pour SORM
(et alors $\hat{P}$ = formule de Breitung).



Cf: O. Ditlevsen et H.O. Madsen,
Structural reliability methods, Wiley, 1996.

Techniques de réduction de variance

Objectif : réduire la variance de l'estimateur de Monte Carlo classique :

$$\mathbb{E}\big[(\hat{I}_n - I)^2\big] = \frac{1}{n}\mathrm{Var}(F(\boldsymbol{X}))$$

Les méthodes
- Echantillonnage préférentiel
- Variables de contrôle
- Variables antithétiques
- Stratification

ont pour but de réduire la constante, en restant proches de l'esprit Monte Carlo (parallélisable).

Les méthodes
- Quasi-Monte Carlo

ont pour but de changer le $1/n$.

Les méthodes
- Systèmes de particules en interaction

s'éloignent de l'esprit Monte Carlo (séquentielle).

Echantillonnage préférentiel (importance sampling)

• Observation : la représentation de I comme espérance n'est pas unique. Si $\boldsymbol{X}$ est à densité p :

$$I = \mathbb{E}_p[F(\boldsymbol{X})] = \int F(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = \int \frac{F(\boldsymbol{x})p(\boldsymbol{x})}{q(\boldsymbol{x})}q(\boldsymbol{x})d\boldsymbol{x} = \mathbb{E}_q\left[\frac{F(\boldsymbol{X})p(\boldsymbol{X})}{q(\boldsymbol{X})}\right]$$

L'utilisateur est libre dans le choix de la densité q.

• Idée : Dans la situation où on sait que $F(\boldsymbol{X})$ est surtout sensible à certaines valeurs de $\boldsymbol{X}$, au lieu de tirer les $\boldsymbol{X}^{(k)}$ selon la densité originale $p(\boldsymbol{x})$ de $\boldsymbol{X}$, on les tire selon une densité "biaisée" $q(\boldsymbol{x})$ qui favorise les valeurs de $\boldsymbol{X}$ dans la zone d'importance.

• En considérant la représentation

$$I = \mathbb{E}_p[F(\boldsymbol{X})] = \mathbb{E}_q\left[F(\boldsymbol{X})\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right]$$

on propose l'estimateur :

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} F(\boldsymbol{X}^{(k)})\frac{p(\boldsymbol{X}^{(k)})}{q(\boldsymbol{X}^{(k)})}$$

où $(\boldsymbol{X}^{(k)})_{k=1,\ldots,n}$ est un n-échantillon tiré selon q.

• Estimateur non-biaisé : $\mathbb{E}_q[\hat{I}_n] = I$.

- Estimateur convergent :

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} F(\boldsymbol{X}^{(k)})\frac{p(\boldsymbol{X}^{(k)})}{q(\boldsymbol{X}^{(k)})} \overset{n\to\infty}{\longrightarrow} \mathbb{E}_q\left[F(\boldsymbol{X})\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right] = \mathbb{E}_p\left[F(\boldsymbol{X})\right] = I$$

- Variance de l'estimateur :

$$\mathrm{Var}(\hat{I}_n) = \frac{1}{n}\mathrm{Var}_q\left(F(\boldsymbol{X})\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right) = \frac{1}{n}\left(\mathbb{E}_p\left[F(\boldsymbol{X})^2\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right] - \mathbb{E}_p\left[F(\boldsymbol{X})\right]^2\right)$$

- En choisissant bien q, on peut fortement réduire la variance. En fait, en choisissant (en supposant $F \geq 0$)

$$q_{opt}(\boldsymbol{x}) = \frac{F(\boldsymbol{x})p(\boldsymbol{x})}{I}$$

on trouve

$$\mathrm{Var}(\hat{I}_n) = 0$$

Mais $q_{opt}(\boldsymbol{x})$ dépend de I ! ($\hookrightarrow$ méthodes séquentielles/adaptatives).

- Points pratiques importants pour pouvoir implémenter la méthode :

- Il faut savoir simuler $\boldsymbol{X}$ de loi de densité q.

- Il faut savoir calculer le rapport de vraisemblance $\dfrac{p(\boldsymbol{x})}{q(\boldsymbol{x})}$.

• Exemple : On veut estimer

$$I = \mathbb{E}[F(X)]$$

avec $X \sim \mathcal{N}(0,1)$ et $F(x) = \mathbf{1}_{[4,\infty[}(x)$.

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathbf{1}_{[4,\infty[}(x) e^{-\frac{x^2}{2}}\, dx = \frac{1}{2}\mathrm{erfc}\Big(\frac{4}{\sqrt{2}}\Big) \simeq 3.17\,10^{-5}$$

Monte Carlo :

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4}, \quad X_k \sim \mathcal{N}(0,1)$$

On a $\mathrm{Var}(\hat{I}_n) = \frac{1}{n} 3.17\,10^{-5}$.

Echantillonnage préférentiel : on tire X_k selon la loi $\mathcal{N}(4,1)$.

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4} \frac{e^{-\frac{X_k^2}{2}}}{e^{-\frac{(X_k-4)^2}{2}}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4} e^{-4X_k+8}, \quad X_k \sim \mathcal{N}(4,1)$$

On a $\mathrm{Var}(\hat{I}_n) = \frac{1}{n} 5.53\,10^{-8}$.

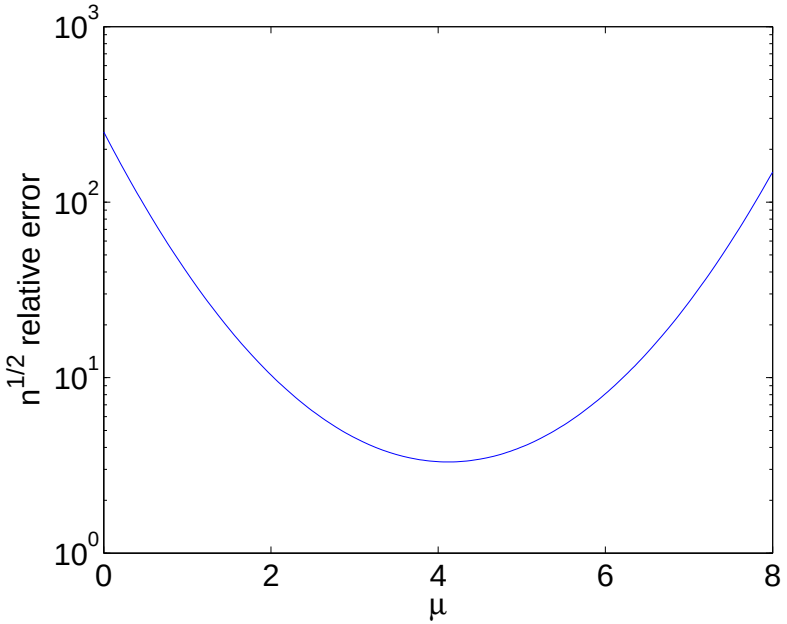
Il faut 1000 fois moins de simulations avec la méthode IS pour atteindre la même précision !

Attention cependant, il ne faut pas trop pousser.

Echantillonnage préférentiel : on tire X_k selon la loi $\mathcal{N}(\mu, 1)$.

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4} \frac{e^{-\frac{X_k^2}{2}}}{e^{-\frac{(X_k-\mu)^2}{2}}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4} e^{-\mu X_k + \frac{\mu^2}{2}}, \quad X_k \sim \mathcal{N}(\mu, 1)$$

On a $\mathrm{Var}(\hat{I}_n) = \frac{1}{n}\frac{e^{\mu^2}}{2}\mathrm{erfc}\left(\frac{4+\mu}{\sqrt{2}}\right) - \frac{1}{n}I^2$, ce qui donne pour l'erreur relative normalisée $\sqrt{n}\mathbb{E}[(\hat{I}_n - I)^2]^{1/2}/I$:



Si on pousse trop, les fluctuations des rapports de vraisemblance deviennent trop grandes.

Variables de contrôle

• Rappel : Le but est d'estimer $I = \mathbb{E}[F(\boldsymbol{X})]$ pour $\boldsymbol{X}$ un vecteur aléatoire et $F(\boldsymbol{x}) = \psi(f(\boldsymbol{x}))$ une fonction déterministe.

• Dans la situation où on dispose d'un modèle réduit $f_{\mathrm{r}}(\boldsymbol{x})$.

• Méthode d'échantillonnnage préférentiel : on calcule (on approche) la densité optimale $q_{opt,\mathrm{r}}(\boldsymbol{x}) = \frac{\psi(f_{\mathrm{r}}(\boldsymbol{x}))p(\boldsymbol{x})}{I_{\mathrm{r}}}$, avec $I_{\mathrm{r}} = \int \psi(f_{\mathrm{r}}(\boldsymbol{x}))p(\boldsymbol{x})d\boldsymbol{x}$, puis on l'utilise comme densité d'importance.

• Méthode de variables de contrôle :
On note $F(\boldsymbol{x}) = \psi(f(\boldsymbol{x}))$, $F_{\mathrm{r}}(\boldsymbol{x}) = \psi(f_{\mathrm{r}}(\boldsymbol{x}))$.
On suppose qu'on connaît $I_{\mathrm{r}} = \mathbb{E}[F_{\mathrm{r}}(\boldsymbol{X})]$.

En considérant la représentation

$$I = \mathbb{E}[F(\boldsymbol{X})] = I_{\mathrm{r}} + \mathbb{E}[F(\boldsymbol{X}) - F_{\mathrm{r}}(\boldsymbol{X})]$$

on propose l'estimateur :

$$\hat{I}_n = I_{\mathrm{r}} + \frac{1}{n}\sum_{k=1}^{n} F(\boldsymbol{X}^{(k)}) - F_{\mathrm{r}}(\boldsymbol{X}^{(k)}),$$

où $(\boldsymbol{X}^{(k)})_{k=1,\ldots,n}$ est un n-échantillon (tiré selon p).

Estimateur :

$$\hat{I}_n = I_{\mathrm{r}} + \frac{1}{n}\sum_{k=1}^{n} F(\boldsymbol{X}^{(k)}) - F_{\mathrm{r}}(\boldsymbol{X}^{(k)})$$

Cet estimateur est sans biais et convergent.

Sa variance est :

$$\mathrm{Var}(\hat{I}_n) = \frac{1}{n}\mathrm{Var}\big(F(\boldsymbol{X}) - F_{\mathrm{r}}(\boldsymbol{X})\big)$$

$\hookrightarrow$ Cette méthode peut réduire la variance.

- Exemple : on souhaite estimer

$$I = \mathbb{E}[f(X)]$$

avec $X \sim \mathcal{U}(0,1)$, $f(x) = \exp(x)$.
Résultat : $I = e - 1 \simeq 1.72$.

Monte Carlo.

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^{n} \exp[X^{(k)}]$$

Variance de l'estimateur MC $= \frac{1}{n}(2e - 1) \simeq \frac{1}{n} 4.44$.

Variable de contrôle. Modèle réduit : $f_{\mathrm{r}}(x) = 1 + x$ (ici $I_{\mathrm{r}} = \frac{3}{2}$). Estimateur VC :

$$\hat{I}_n = I_{\mathrm{r}} + \frac{1}{n} \sum_{k=1}^{n} \left\{ \exp[X^{(k)}] - 1 - X^{(k)} \right\}$$

Variance de l'estimateur VC $= \frac{1}{n}(3e - \frac{e^2}{2} - \frac{53}{12}) \simeq \frac{1}{n} 0.044$.

Il faut donc 100 fois moins de simulations avec l'estimateur VC !

• Application : Méthodes de Romberg statistiques pour l'estimation de

$$I = \mathbb{E}[\psi(f(\boldsymbol{X}))]$$

On dispose d'un code léger f_{r} en plus du code lourd f. Le rapport du coût calcul entre un appel à f et un appel à f_{r} est $q > 1$.

Estimateur

$$\hat{I}_n = \frac{1}{n_{\mathrm{r}}} \sum_{i=1}^{n_{\mathrm{r}}} F_{\mathrm{r}}(\tilde{\boldsymbol{X}}^{(k)}) + \frac{1}{n} \sum_{i=1}^{n} F(\boldsymbol{X}^{(k)}) - F_{\mathrm{r}}(\boldsymbol{X}^{(k)})$$

avec $n_{\mathrm{r}} \gg n$, $F(\boldsymbol{x}) = \psi(f(\boldsymbol{x}))$, $F_{\mathrm{r}}(\boldsymbol{x}) = \psi(f_{\mathrm{r}}(\boldsymbol{x}))$.
Allocation entre appels au code lourd et appels au code léger à optimiser sous la contrainte $n_{\mathrm{r}}/q + n(1 + 1/q) = n_{\mathrm{tot}}$.
Compromis classique entre erreur d'approximation et erreur d'estimation.

• Utilisé dans le cas où $f(\boldsymbol{X})$ est la solution d'une équation différentielle discrétisée finement, avec $f_{\mathrm{r}}(\boldsymbol{X})$ la solution avec un schéma de discrétisation grossier (MultiLevel Monte Carlo).
↪ Il ne faut pas utiliser le code avec la précision maximale pour tous les appels dans un problème de propagation d'incertitudes (ou d'analyse de sensibilité) !

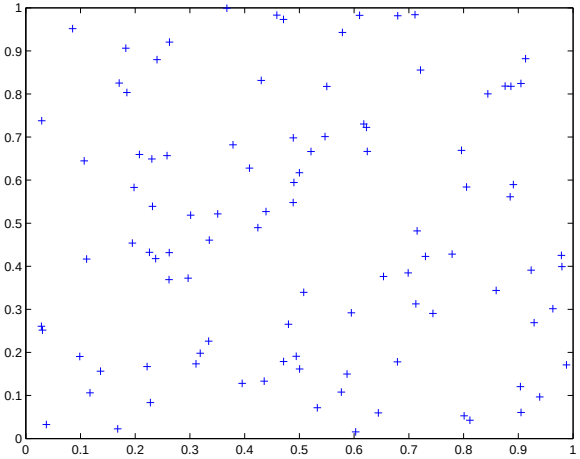
Suite à discrépance faible (quasi Monte Carlo)

on tire l'échantillon de manière moins aléatoire que MC, pour combler les trous qui se forment naturellement dans un échantillon aléatoire.

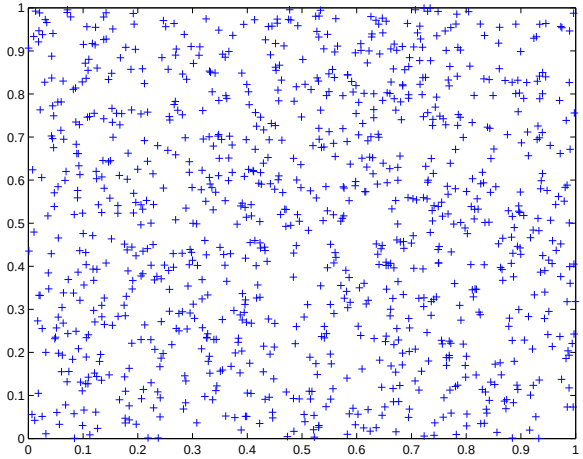
Cette technique
- réduit la variance si f a un peu de régularité et/ou de monotonie; on peut aller jusqu'à une variance en $C_d(\log n)^{s(d)}/n^2$,
- marche en dimension pas trop grande,
- représente un intermédiaire entre MC et quadrature usuelle,
- en compétition avec des méthodes de quadrature creuse (Smolyak).

Attention :
- il n'est pas facile de rajouter des points,
- on n'a pas d'estimée d'erreur (sans hypothèse supplémentaire),
- la méthode n'est pas adaptée pour l'estimation de probabilité d'événement rare.

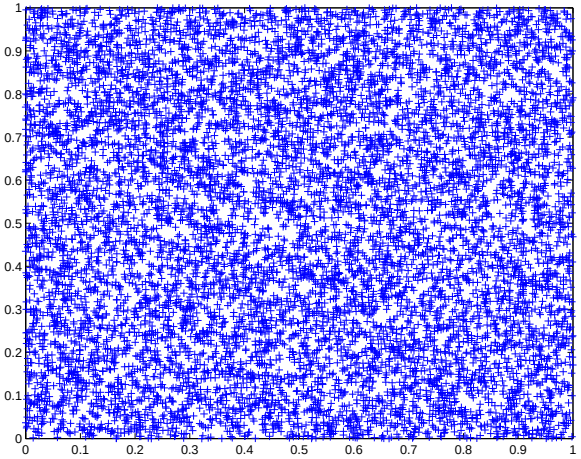
Exemple : échantillon Monte Carlo.



$n = 100$

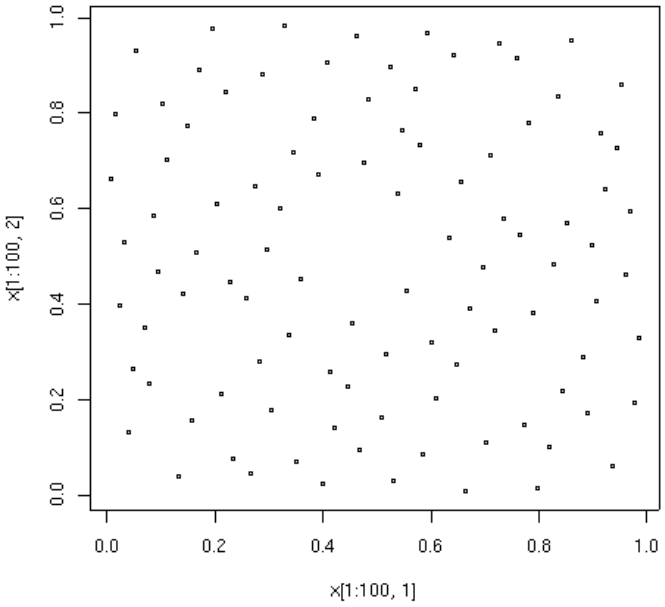


$n = 1000$

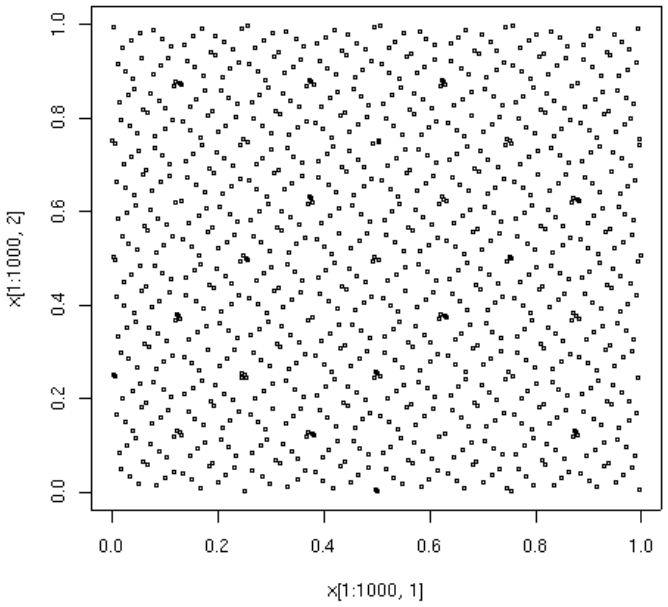


$n = 10000$

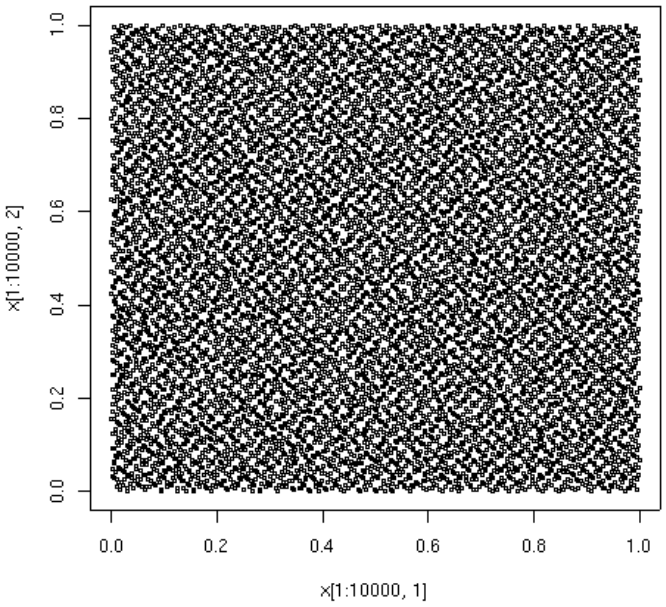
Exemple : suites de Sobol en dimension 2.



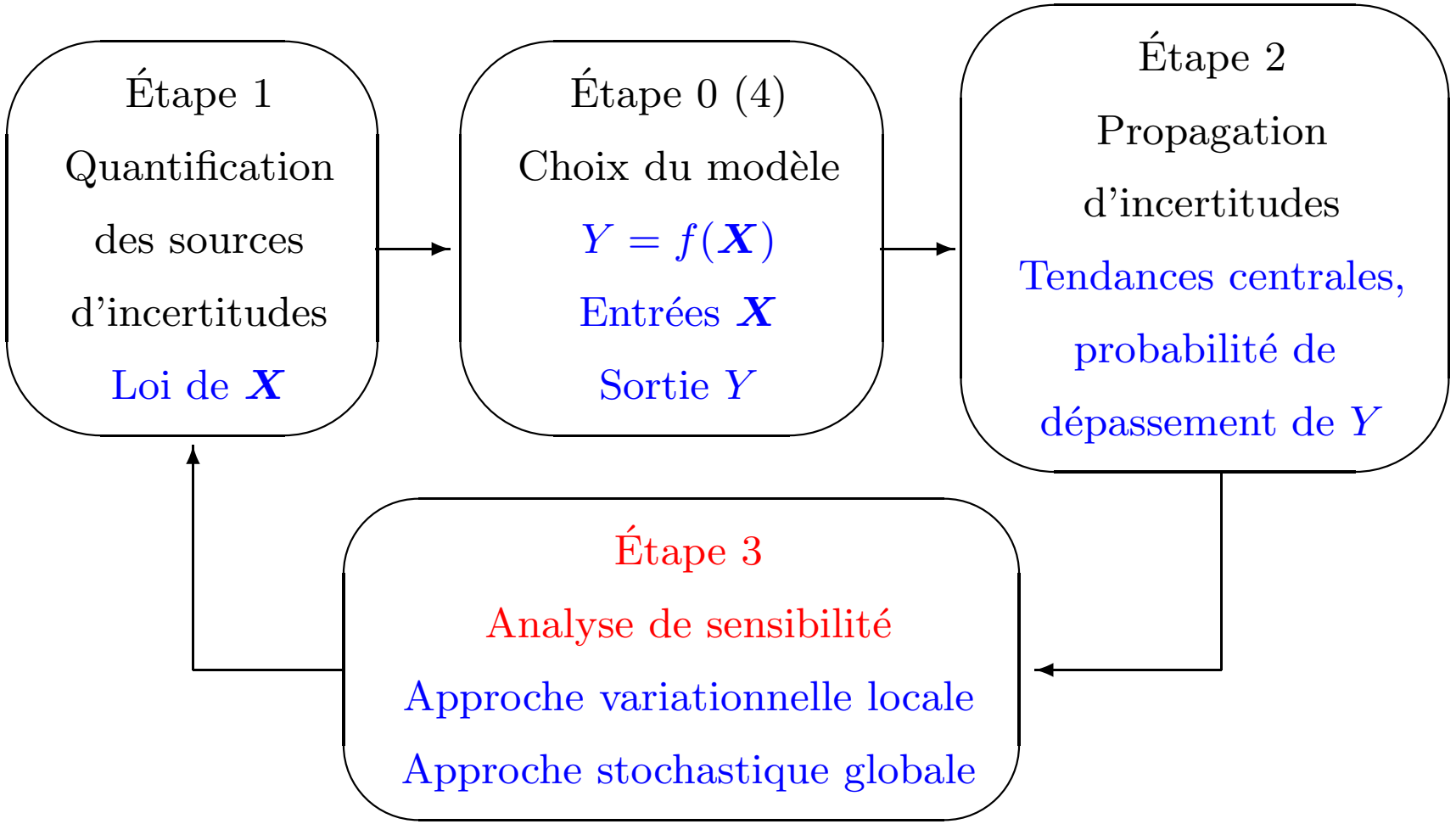
$n = 100$



$n = 1000$



$n = 10000$


Étape 1
Quantification
des sources
d'incertitudes
Loi de $\boldsymbol{X}$
Étape 0 (4)
Choix du modèle
$Y = f(\boldsymbol{X})$
Entrées $\boldsymbol{X}$
Sortie Y
Étape 2
Propagation
d'incertitudes
Tendances centrales,
probabilité de
dépassement de Y
Étape 3
Analyse de sensibilité
Approche variationnelle locale
Approche stochastique globale

Partie III. Analyse de sensibilité

- Contexte : code de calcul informatique ou expérience modélisé par

$$Y = f(\boldsymbol{X})$$

avec Y =variable de sortie réelle

$\boldsymbol{X} = (X_1, ..., X_d)$ variables d'entrée

f = boîte noire déterministe

- But : expliquer la variabilité de la réponse Y en fonction des X_i.

- Objectifs principaux :

- améliorer la compréhension du phénomène.
- réduire l'incertitude d'un modèle, en identifiant les variables les plus influentes dans un domaine de valeurs de la sortie ($\rightarrow$ il devient prioritaire de réduire la variabilité de ces entrées).
- simplifier ou alléger le modèle, en fixant les variables les moins influentes.

On se donne : la loi des X_i (gaussienne, uniforme,...).

But de l'analyse de sensibilité :

- qualitative : identifier les paramètres importants parmi les nombreux paramètres (screening ou criblage).

- quantitative : déterminer la part de la variance de la sortie Y due à une variable d'entrée ou un sous-ensemble de variables d'entrée X_i.

Références :

A. Saltelli, K. Chan et E. M. Scott, Sensitivity analysis, Wiley, 2001

A. Saltelli et S. Tarantola, Sensivity analysis in practice, Wiley, 2004

Criblage

Un modèle comportant beaucoup de variables d'entrée est difficile à explorer.
Souvent, seulement quelques entrées sont influentes.
Objectif : identifier rapidement les quelques h entrées influentes parmi les d entrées, avec n calculs.

Méthodes possibles :
- méthode OAT (One factor At a Time) : calcul de gradient avec $n = d + 1$.
- méthode de Morris avec $n = R(d + 1)$.
- criblage par groupes, bifurcation séquentielle (avec des hypothèses de monotonie) avec $n < d$ et $h \ll d$.

• Méthode OAT.

- On fait $d+1$ appels au code en $\boldsymbol{x}_0$ et $\boldsymbol{x}_0 + \delta \boldsymbol{e}_j$, avec $\boldsymbol{x}_0$ un point de référence, δ un pas, et $(\boldsymbol{e}_j)_{j=1,\dots,d}$ base canonique de $\mathbb{R}^d$.

- Indice de sensibilité :

$$S_j^{\boldsymbol{x}_0} = \frac{f(\boldsymbol{x}_0 + \delta \boldsymbol{e}_j) - f(\boldsymbol{x}_0)}{\delta}$$

Une valeur importante de $|S_j^{\boldsymbol{x}_0}|$ est révélateur d'un modèle sensible aux variations de l'entrée X_j.

Inconvénient : indice local, dépend du choix de $\boldsymbol{x}_0$.

• Méthode de Morris. Indices de sensibilité globaux :

$$\mu_j = \mathbb{E}[|S_j^{\boldsymbol{X}}|], \qquad \sigma_j^2 = \mathrm{Var}(S_j^{\boldsymbol{X}})$$

Estimation des indices :

- On tire R points $(\boldsymbol{x}_k)_{k=1,\dots,R}$ (avec Monte Carlo, ou bien avec quasi-Monte Carlo; par exemple, avec des X_i uniformes sur $[0,1]$ et indépendants, on tire les $\boldsymbol{x}_k$ avec un hypercube latin).

- On fait $R(d+1)$ appels au code en $\boldsymbol{x}_k$ et $\boldsymbol{x}_k + \delta \boldsymbol{e}_j$, $j = 1, \dots, d$, $k = 1, \dots, R$.

$$\hat{\mu}_j = \frac{1}{R} \sum_{k=1}^{R} |S_j^{\boldsymbol{x}_k}|, \qquad \hat{\sigma}_j^2 = \frac{1}{R-1} \sum_{k=1}^{R} \left(S_j^{\boldsymbol{x}_k} - \frac{1}{R} \sum_{l=1}^{R} S_j^{\boldsymbol{x}_l} \right)^2$$

$$\mu_j = \mathbb{E}[|S_j^{\boldsymbol{X}}|] \text{ estimé par } \hat{\mu}_j = \frac{1}{R}\sum_{k=1}^{R} |S_j^{\boldsymbol{x}_k}|$$

$$\sigma_j^2 = \mathrm{Var}(S_j^{\boldsymbol{X}}) \text{ estimé par } \hat{\sigma}_j^2 = \frac{1}{R-1}\sum_{k=1}^{R}\left(S_j^{\boldsymbol{x}_k} - \frac{1}{R}\sum_{l=1}^{R} S_j^{\boldsymbol{x}_l}\right)^2$$

- μ_j est une mesure de la sensibilité.

Une valeur importante traduit des effets importants (en moyenne) et un modèle sensible aux variations de l'entrée X_j.

- σ_j est une mesure des interactions et des effets non-linéaires.

Une valeur importante traduit des effets différents les uns des autres, des effets qui dépendent de la valeur :

- soit de l'entrée elle-même : effet non-linéaire
- soit des autres entrées : interaction

(la méthode ne permet pas de distinguer les deux cas)

Criblage : Méthode de Morris

• Permet de classer les entrées en trois groupes selon leurs effets :

- effets négligeables
- effets linéaires et sans interaction
- effets non-linéaires et/ou avec interactions

• En théorie, pas d'hypothèse sur le modèle; en pratique, la méthode marche mieux avec une certaine régularité.

• Généralise la méthode OAT (One factor At a Time).

Indices de sensibilité pour un modèle linéaire

- Soit le modèle linéaire

$$Y = f(X_1, ..., X_d) = \alpha_0 + \sum_{i=1}^{d} \alpha_i X_i$$

avec $X_i \in \mathbb{R}$ décorrélés.

- La variance de Y s'écrit

$$\mathrm{Var}(Y) = \sum_{i=1}^{d} \alpha_i^2 \mathrm{Var}(X_i)$$

La part de variance de Y due à X_i est $\alpha_i^2 \mathrm{Var}(X_i)$.
$\rightarrow$ On obtient une décomposition de la variance de Y en fonction des X_i.

- La sensibilité de Y à X_i est quantifiée par l'indice SRC (Standardized Regression Coefficient) :

$$\mathrm{SRC}_i = \frac{\alpha_i^2 \mathrm{Var}(X_i)}{\mathrm{Var}(Y)}$$

- Avec le modèle linéaire $Y = \alpha_0 + \sum_{i=1}^{d} \alpha_i X_i$,

$$\mathrm{SRC}_i = \frac{\alpha_i^2 \mathrm{Var}(X_i)}{\mathrm{Var}(Y)}$$

de manière équivalente :

$$\mathrm{SRC}_i = \frac{\mathrm{Cov}(X_i, Y)^2}{\mathrm{Var}(X_i)\mathrm{Var}(Y)}$$

- Estimation :

- estimation des α_i (OAT différences finies, régression linéaire).

- Monte Carlo : on tire $\boldsymbol{X}^{(k)}$, on calcule $Y^{(k)} = f(\boldsymbol{X}^{(k)})$, et on estime

$$\widehat{\mathrm{SRC}}_i = \frac{\left(\sum_{k=1}^{n} (X_i^{(k)} - \hat{X}_i)(Y^{(k)} - \hat{Y})\right)^2}{\sum_{k=1}^{n} (X_i^{(k)} - \hat{X}_i)^2 \sum_{k=1}^{n} (Y^{(k)} - \hat{Y})^2},$$

$$\hat{X}_i = \frac{1}{n} \sum_{k=1}^{n} X_i^{(k)}, \qquad \hat{Y} = \frac{1}{n} \sum_{k=1}^{n} Y_i^{(k)},$$

Indices de sensibilité pour un modèle monotone

• Soit le modèle monotone

$$Y = f(X_1, ..., X_d)$$

avec $X_i \in \mathbb{R}$ décorrélés et f monotone.

• Coefficient de corrélation basé sur les rangs :

- On tire avec MC $\boldsymbol{X}^{(k)}$ et on calcule $Y^{(k)} = f(\boldsymbol{X}^{(k)})$, $k = 1, \ldots, n$.

- A chaque k on associe son rang $r_Y^{(k)}$ selon la valeur $Y^{(k)}$ (rang 1 attribué à la plus petite valeur, rang n associé à la plus grande valeur) et son rang $r_{X_i}^{(k)}$ selon la valeur de $X_i^{(k)}$.

- La sensibilité de Y à X_i est quantifiée par l'indice SRRC (Standardized Regression Rank Coefficient) qui est le SRC pour les rangs :

$$\widehat{\mathrm{SRRC}}_i = \frac{\left(\sum_{k=1}^{n}(r_{X_i}^{(k)} - \frac{n+1}{2})(r_Y^{(k)} - \frac{n+1}{2})\right)^2}{\sum_{k=1}^{n}(r_{X_i}^{(k)} - \frac{n+1}{2})^2 \sum_{k=1}^{n}(r_Y^{(k)} - \frac{n+1}{2})^2}$$

Indices de sensiblité pour un modèle arbitraire (Sobol)

• Lorsqu'il n'est pas possible de faire d'hypothèse sur le modèle (cas général), on définit des indices de sensibilité à partir d'une décomposition de la variance de Y.

• Soit le modèle $Y = f(X_1, ..., X_d)$ avec $X_i \in \mathbb{R}$ et $X_i \perp X_j$, $i \neq j$.

• Le modèle peut se décomposer (Sobol) en

$$f(X_1, ..., X_d) = f_0 + \sum_{i=1}^{d} f_i(X_i) + \sum_{1 \leq i < j \leq d} f_{ij}(X_i, X_j) + \cdots + f_{1...d}(X_1, ..., X_d)$$

avec

$$\mathbb{E}[f_{i_1...i_s}(X_{i_1}, ..., X_{i_s}) f_{j_1...j_t}(X_{j_1}, ..., X_{j_t})] = 0 \text{ si } (i_1...i_s) \neq (j_1...j_t)$$

En fait, un seul choix possible :

$$\begin{aligned} f_0 &= \mathbb{E}[Y] \\ f_i(X_i) &= \mathbb{E}[Y|X_i] - f_0 \\ f_{ij}(X_i, X_j) &= \mathbb{E}[Y|X_i, X_j] - f_i(X_i) - f_j(X_j) - f_0 \\ &\vdots \end{aligned}$$

En utilisant cette décomposition, la variance de Y se décompose en :

$$\mathrm{Var}(Y) = D = \sum_{i=1}^{d} D_i + \sum_{1 \leq i < j < \leq p} D_{ij} + \cdots + D_{1\ldots p}$$

où

$$\begin{aligned} D_i &= \mathrm{Var}(E[Y|X_i]) \\ D_{ij} &= \mathrm{Var}(E[Y|X_i, X_j] - E[Y|X_i] - E[Y|X_j]) \end{aligned}$$

- D_i est la sensibilité au 1er ordre de Y par rapport à la variable X_i (la part de la variance de Y expliquée par les fluctuations de X_i)
- D_{ij} est la sensibilité de Y par rapport à l'interaction des variables X_i et X_j (la part de la variance Y expliquée par l'interaction de X_i et X_j) .
- Indices de sensibilité :

$$S_i = \frac{D_i}{D}\,, \qquad S_{ij} = \frac{D_{ij}}{D}$$

avec

$$S_i \geq 0\,, \quad S_{ij} \geq 0\,, \qquad \text{et} \qquad \sum_{i=1}^{d} S_i + \sum_{1 \leq i < j < \leq p} S_{ij} + \cdots + S_{1\ldots p} = 1$$

Le nombre d'indices de sensibilité $2^d - 1$ devient vite grand avec d.

On introduit alors l'indice de sensibilité total

$$S_{Ti} = \text{ somme de tous les indices relatifs à } X_i$$

qui exprime la sensibilité de Y à X_i sous toutes ses formes, i.e. à X_i seule et en interaction avec d'autres variables.

On donne en général l'indice du premier ordre S_i et l'indice total S_{Ti}.
- Si S_{Ti} petit : la variable X_i a des effets négligeables,
- Si S_i grand : la variable X_i a des effets propres importants,
- Si S_i petit et S_{Ti} grand : la variable X_i a des effets importants en interaction avec d'autres variables.

• Méthodes d'estimation des indices de Sobol.

• Méthodes particulières :

- Mc Kay
- FAST (Fourier Amplitude Sensitivity Test)

• Méthode de Sobol

→ basée sur des estimations de Monte Carlo,

→ des améliorations existent (avec des techniques de réduction de variance) :

- échantillonnage stratifié,
- échantillonnage par plans space-filling (Hypercubes Latins),
- suites à discrépance faible (suites de Sobol LPτ).

• Estimations gourmandes en nombre d'appels.

$\hookrightarrow$ Utilisation d'un métamodèle (en particulier, polynômes de chaos). Mais on estime alors la sensibilité du métamodèle.

- Estimations de Sobol.

Indice de premier ordre : $S_i = \dfrac{D_i}{D} = \dfrac{\mathrm{Var}(E[Y|X_i])}{\mathrm{Var}(Y)}$.

$f_0 = \mathbb{E}[Y]$ et $D = \mathrm{Var}(Y)$ sont estimés classiquement par

$$f_0 \simeq \frac{1}{n}\sum_{k=1}^{n} f(X_1^{(k)}, ..., X_{i-1}^{(k)}, X_i^{(k)}, X_{i+1}^{(k)}, ..., X_d^{(k)})$$

$$D \simeq \frac{1}{n}\sum_{k=1}^{n} f(X_1^{(k)}, ..., X_{i-1}^{(k)}, X_i^{(k)}, X_{i+1}^{(k)}, ..., X_d^{(k)})^2 - f_0^2$$

où $(X_1^{(k)}, ..., X_d^{(k)})_{k=1...n}$ est un n-échantillon des variables d'entrée.

$D_i = \mathrm{Var}(E[Y|X_i])$ est estimée par

$$D_i \simeq \frac{1}{n}\sum_{k=1}^{n} f(X_1^{(k)}, ..., X_{i-1}^{(k)}, X_i^{(k)}, X_{i+1}^{(k)}, ..., X_d^{(k)}) f(\tilde{X}_1^{(k)}, ..., \tilde{X}_{i-1}^{(k)}, X_i^{(k)}, \tilde{X}_{i+1}^{(k)}, ..., \tilde{X}_d^{(k)}) - f_0^2$$

où $(\tilde{X}_1^{(k)}, ..., \tilde{X}_d^{(k)})_{k=1...n}$ est un second échantillon des variables d'entrée.

Questions complémentaires (plus ou moins ouvertes)

Incertitude de modèle :
Que deviennent les indices de sensibilité estimés si le modèle change (mutation) ?
$\mapsto$ quelques résultats sur quelques modèles de mutation (additifs).
Comment prendre en compte, dans les résultats de sensibilité, l'utilisation d'un modèle simplifié ?
$\mapsto$ pour l'estimation Monte Carlo des indices, pour des approches multi-fidélité.

Modèles à entrées corrélées :
Comment réaliser (interpréter) une analyse de sensibilité lorsque les variables d'entrée ne sont pas indépendantes ?
$\mapsto$ se ramener à des variables décorrélées.

Références :
A. Saltelli, K. Chan et E. M. Scott, Sensitivity analysis
A. Saltelli et S. Tarantola, Sensivity analysis in practice

Partie IV. Plans d'expériences et métamodèles

Planification d'expériences : Mise au point d'une suite d'expériences (de simulations numériques) en fonction d'un but :
- criblage (détermination des paramètres d'entrée importants)
- étude quantitative de l'influence des paramètres d'entrée
- construction d'une surface de réponse, optimisation, ...

Référence :
J. J. Droesbeke, J. Fine et G. Saporta, Plans d'expériences, application à l'entreprise.

Ici : construction d'un métamodèle (ou surface de réponse).

Type de métamodèles (fonction f_r destinée à approcher la fonction f) :

• Polynômes

• Splines (fonctions définies par morceaux par des polynômes)

• Modèles linéaires généralisés

• Polynômes de chaos

• Krigeage (la fonction f est représentée comme une réalisation d'un processus gaussien; utilise la théorie des processus gaussiens)

• Réseaux de neurones (optimisés par méthodes d'apprentissage à des fins de mémorisation et de généralisation).
mémorisation : le fait d'assimiler des exemples éventuellement nombreux,
généralisation : le fait d'être capable, grâce aux exemples appris, de traiter des exemples distincts, encore non rencontrés, mais similaires.

• Machines à vecteurs de support (en anglais Support Vector Machine SVM) sont destinées à résoudre des problèmes de discrimination (à l'origine, classifieur linéaire).

$\hookrightarrow$ Chaque méthode a un (des) plan(s) d'expériences approprié(s).

Plans pour des surfaces de réponse polynomiales

• Supposons $\boldsymbol{X}$ de dimension d à moyenne nulle. La "vraie" fonction est :

$$Y = f(\boldsymbol{X}) = \beta_0 + \sum_{j=1}^{d} \beta_j X_j + \sum_{1 \leq i < j \leq d} \beta_{ij} X_i X_j + \cdots$$

$\beta_j X_j$: effet principal; $\beta_{ij} X_i X_j$: interaction d'ordre 2; etc.

Modèle polynomial (degré 1) :

$$f_r(\boldsymbol{X}) = b_0 + \sum_{j=1}^{d} b_j X_j$$

Modèle polynomial (degré 2) :

$$f_r(\boldsymbol{X}) = b_0 + \sum_{j=1}^{d} b_j X_j + \sum_{1 \leq i < j \leq d} b_{ij} X_i X_j$$

• Résolution : une méthode est dite de résolution L si aucune interaction d'ordre m n'est confondue avec des interactions d'ordre $L - m - 1$.

Résolution III : aucun effet principal (d'ordre 1) n'est confondu (aliasing) avec un autre effet principal (mais un effet principal peut être confondu avec un effet d'interaction d'ordre 2).

Plackett-Burman design $2^3//4$

"Vraie" fonction : $Y = \beta_0 + \sum_{j=1}^{3} \beta_j X_j + \sum_{1 \leq i < j \leq 3} \beta_{ij} X_i X_j + \cdots$

Modèle postulé : $f_r(\boldsymbol{X}) = b_0 + \sum_{j=1}^{3} b_j X_j = (1, \boldsymbol{X}^T)\boldsymbol{b}$

Plan :

n	X_0	X_1	X_2	X_3
1	+1	-1	-1	+1
2	+1	+1	-1	-1
3	+1	-1	+1	-1
4	+1	+1	+1	+1

Le plan forme une matrice $\mathbf{M}$, de taille 4×4.
On effectue les 4 simulations $\rightarrow \boldsymbol{Y}_{obs}$, on résoud $\boldsymbol{Y}_{obs} \simeq \mathbf{M}\boldsymbol{b}$, qui donne $\boldsymbol{b} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\boldsymbol{Y}_{obs}$.

$$\begin{aligned} b_0 &= \beta_0 + \cdots \\ b_1 &= \beta_1 + \beta_{23} + \cdots \\ b_2 &= \beta_2 + \beta_{13} + \cdots \\ b_3 &= \beta_3 + \beta_{12} + \cdots \end{aligned}$$

Résolution R_{III}.

• Résolution : une méthode est dite de résolution L si aucune interaction d'ordre m n'est confondue avec des interactions d'ordre $L - m - 1$.

Résolution III : aucun effet principal (d'ordre 1) n'est confondu (aliasing) avec un autre effet principal (mais un effet principal peut être confondu avec un effet d'interaction d'ordre 2).

Résolution IV : aucun effet principal n'est confondu avec un autre effet principal ou un effet d'ordre 2 (mais un effet principal peut être confondu avec un effet d'interaction d'ordre 3, ou bien un effet d'ordre 2 peut être confondu avec un autre effet d'ordre 2).

Résolution V : aucun effet principal n'est confondu avec un autre effet principal ou un effet d'ordre 2 ou 3, aucun effet d'ordre 2 n'est confondu avec un autre effet principal ou un effet d'ordre 2.

- Plackett-Burman (d paramètres d'entrée à deux niveaux)

- $n \geq d+1$, $n \equiv 0[4]$, R_{III}

- $n \geq 2(d+1)$, $n \equiv 0[4]$, R_{IV}

- Plackett-Burman (q_j niveaux pour le paramètre j, $j = 1, \ldots, d$)

- $n \geq 1 + \sum_{j=1}^{d}(q_j - 1)$, R_{III}

- Design de Rechtschaffner (d paramètres d'entrée à deux niveaux).

- $n = 1 + d(d-1)/2$, $R_V \rightarrow$ évalue tous les effets principaux et toutes les interactions d'ordre 2

Plans pour des surfaces de réponse polynomiales

• Pas de plan d'expérience universel !

Dépend de la forme de la surface de réponse et du domaine de variation (loi) des X_i.

• Plusieurs sens d'optimalité sont possibles (souvent : minimiser la variance de prédiction intégrée sur le domaine).

• Nécessité de valider sur des points de validation (ou de test), et donc de construire la surface de réponse avec seulement une partie des simulations/des expériences à disposition.

Plans space-filling

Quand la forme de la surface n'est pas a priori connue (en particulier, pour des surfaces de réponse non-paramétriques de type krigeage) : plans space-filling (pour balayer le domaine d'intérêt).

• Typiquement hypercubes latins (HL) (domaine d'intérêt : hypercube) :
- Chaque facteur a le même nombre de niveaux (n).
- Chacun des niveaux est pris une fois et une seule par chaque facteur.

• Critères de sélection parmi tous les HL $(\boldsymbol{x}_j)_{j=1,\ldots,n}$:

- Remplissage (maximin) : indicateur $D_{min} = \min_{1\leq i\neq j\leq n} d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ à maximiser.

- Indépendance des facteurs : déterminant R de la matrice de corrélation à maximiser.

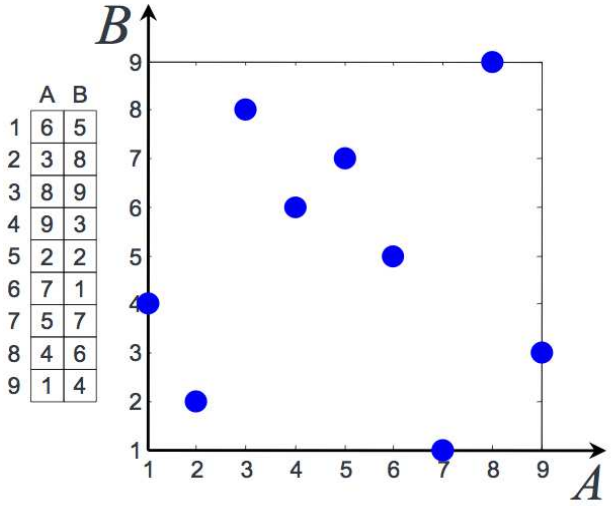
- Uniformité (discrépance) : distance (centered L^2 discrepancy) à la répartition uniforme à minimiser :

$$\begin{aligned} CL_2 = & \left(\frac{13}{12}\right)^d - \frac{2}{n}\sum_{i=1}^{n}\prod_{k=1}^{d}\left(1+\frac{|x_{ik}-0.5|}{2}-\frac{|x_{ik}-0.5|}{2}\right) \\ & +\frac{1}{n^2}\sum_{i,j=1}^{n}\prod_{k=1}^{d}\left(1+\frac{|x_{ik}-0.5|}{2}+\frac{|x_{jk}-0.5|}{2}-\frac{|x_{ik}-x_{jk}|}{2}\right) \end{aligned}$$

• Méthodes de sélection :

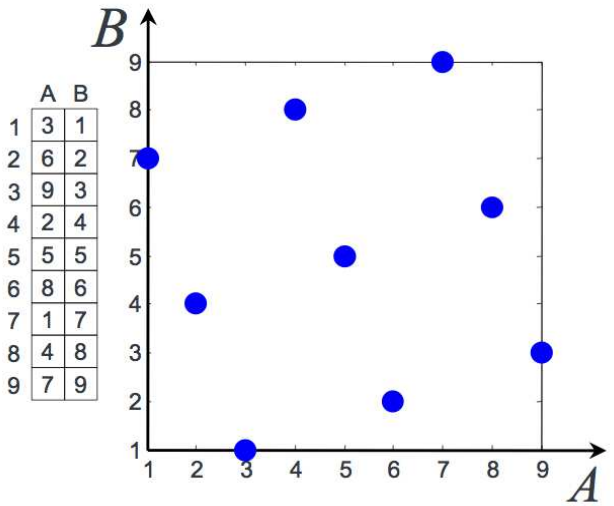
- Méthode exploratoire : On génère un grand nombre de plans et on retient le meilleur selon le critère retenu.

- Méthodes évoluées : Il existe dans la littérature différentes heuristiques (recuit simulé, algorithmes d'échanges) pour optimiser les plans selon le critère que l'on souhaite.



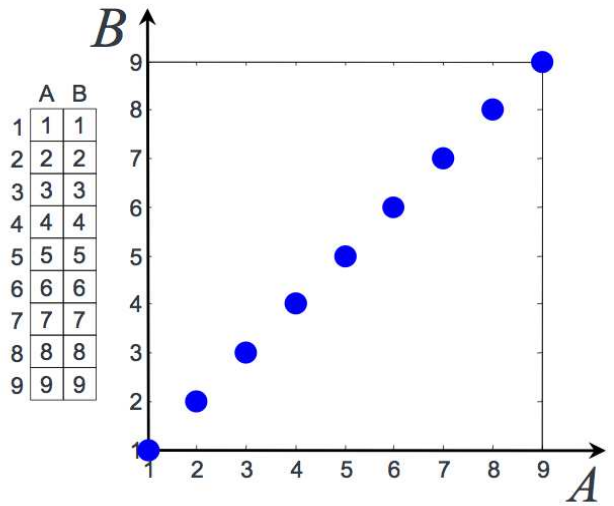
	A	B
1	6	5
2	3	8
3	8	9
4	9	3
5	2	2
6	7	1
7	5	7
8	4	6
9	1	4

$D_{min} = 1.4142$
$R = 0.9989$
$CL_2 = 4.19e\text{-}3$



	A	B
1	3	1
2	6	2
3	9	3
4	2	4
5	5	5
6	8	6
7	1	7
8	4	8
9	7	9

$D_{min} = 3.1623$
$R = 1$
$CL_2 = 3.79e\text{-}3$



	A	B
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9

$D_{min} = 1.4142$
$R = 0$
$CL_2 = 13.68e\text{-}3$

$\hookrightarrow$ on recommande souvent le critère maximin.

Polynômes de chaos

• Représentation de la solution $Y = f(\boldsymbol{X})$ sous la forme d'une série $\sum_j a_j \Phi_j(\boldsymbol{X})$ construite grâce à une projection sur une base spectrale; les coefficients a_j sont calculés par le biais de projections.

R. Ghanem et P. Spanos, Stochastic finite elements: a spectral approach, Dover Publications, 2004.

• Polynôme de chaos (de Wiener).

Supposons que $\boldsymbol{X} = (X_i)_{i=1,\dots,d}$ avec X_i variables aléatoires gaussiennes centrées réduites indépendantes.

Les polynômes d'Hermite (tensorisés) $\phi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \phi_{\alpha_1}(x_1)\cdots\phi_{\alpha_d}(x_d)$ sont orthonormaux dans $L^2(\mathbb{P})$:

$$\mathbb{E}\left[\phi_{\boldsymbol{\alpha}}(\boldsymbol{X})\phi_{\boldsymbol{\alpha}'}(\boldsymbol{X})\right] = \begin{cases} 1 \text{ si } \boldsymbol{\alpha} = \boldsymbol{\alpha}' \\ 0 \text{ sinon} \end{cases}$$

La variable aléatoire $Y = f(\boldsymbol{X})$ (sous réserve que $\mathbb{E}[Y^2] < \infty$) s'écrit

$$Y = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d} a_{\boldsymbol{\alpha}}\phi_{\boldsymbol{\alpha}}(\boldsymbol{X})$$

$\rightarrow$ La variable aléatoire Y est caractérisée par les constantes $a_\cdot$.

En renumérotant les polynômes $(\Phi_j)_{j\in\mathbb{N}} = (\phi_{\alpha_1,\dots,\alpha_d})_{\alpha_1,\dots,\alpha_d\in\mathbb{N}}$:

$$Y = \sum_{j=0}^{\infty} a_j\Phi_j(\boldsymbol{X}) \simeq \sum_{j=0}^{M} a_j\Phi_j(\boldsymbol{X})$$

d : dimension de l'espace probabiliste
q : ordre le plus élevé du polynôme pour représenter Y
$M+1 = (d+q)!/d!q!$ termes dans la somme

- Estimation des coefficients des polynômes de chaos.

$$Y = f(\boldsymbol{X})$$

1/ On suppose que les entrées aléatoires sont des variables aléatoires (gaussiennes) indépendantes $(X_i)_{i=1}^d$ (ou bien on se ramène à ce cas).

2/ On écrit la solution sous forme d'une somme finie de PC :

$$Y = \sum_{i=0}^{M} a_i \Phi_i(\boldsymbol{X})$$

on substitue dans l'équation, et on projette sur la base des polynômes orthogonaux :

$$\mathbb{E}\left[\Phi_k(\boldsymbol{X}) \sum_{i=0}^{M} a_i \Phi_i(\boldsymbol{X})\right] = \mathbb{E}\left[\Phi_k(\boldsymbol{X}) f(\boldsymbol{X})\right] , \qquad k = 0, ..., M$$

$$a_k = \mathbb{E}\left[\Phi_k(\boldsymbol{X}) f(\boldsymbol{X})\right] , \qquad k = 0, ..., M$$

$\rightarrow$ estimation des a_k par régression ou par quadratures numériques (déterministes ou Monte Carlo ou quasi Monte Carlo).

- Estimation des coefficients des polynômes de chaos par moindres carrés.

$Y = f(\boldsymbol{X})$ à approcher par $\boldsymbol{a}^T \boldsymbol{\Phi}(\boldsymbol{X})$, avec $\boldsymbol{\Phi}(\boldsymbol{x}) = \big(\Phi_i(\boldsymbol{x})\big)_{i=0,\dots,M}$, $\quad \boldsymbol{a} = (a_i)_{i=0,\dots,M}$

Idée : minimisation de l'erreur résiduelle :

$$\hat{\boldsymbol{a}}_{LS} = \underset{\boldsymbol{a}\in\mathbb{R}^{M+1}}{\operatorname{argmin}} \ \mathbb{E}\big[\big(\boldsymbol{a}^T \boldsymbol{\Phi}(\boldsymbol{X}) - f(\boldsymbol{X})\big)^2\big]$$

Si on a un n-échantillon $(\boldsymbol{X}^{(k)})_{k=1,\dots,n}$, estimation par

$$\hat{\boldsymbol{a}} = \underset{\boldsymbol{a}\in\mathbb{R}^{M+1}}{\operatorname{argmin}} \ \frac{1}{n}\sum_{k=1}^{n} \big(\boldsymbol{a}^T \boldsymbol{\Phi}(\boldsymbol{X}^{(k)}) - f(\boldsymbol{X}^{(k)})\big)^2$$

c'est-à-dire :

$$\hat{\boldsymbol{a}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{F}, \qquad A_{ki} = \Phi_i(\boldsymbol{X}^{(k)}), \quad F_k = f(\boldsymbol{X}^{(k)})$$

• Estimation de l'erreur.

L'erreur de généralisation du modèle $\hat{\boldsymbol{a}}^T \boldsymbol{\Phi}(\boldsymbol{x})$ obtenu avec $(\boldsymbol{X}^{(k)})_{k=1,\dots,n}$ est :

$$\mathcal{E}_{gen} = \int_{\mathbb{R}^d} \left(\hat{\boldsymbol{a}}^T \boldsymbol{\Phi}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2 \mathbb{P}(d\boldsymbol{x})$$

⚠ Elle est très mal estimée (risque d'overfitting) par l'erreur empirique utilisant l'ensemble d'apprentissage :

$$\mathcal{E}_{emp} = \frac{1}{n} \sum_{k=1}^{n} \left(\hat{\boldsymbol{a}}^T \boldsymbol{\Phi}(\boldsymbol{X}^{(k)}) - f(\boldsymbol{X}^{(k)})\right)^2$$

↪ Validation avec l'erreur empirique sur un ensemble test, indépendant de l'ensemble d'apprentissage (bien, mais cher) :

$$\mathcal{E}_{pre} = \frac{1}{n} \sum_{k=1}^{n} \left(\hat{\boldsymbol{a}}^T \boldsymbol{\Phi}(\widetilde{\boldsymbol{X}}^{(k)}) - f(\widetilde{\boldsymbol{X}}^{(k)})\right)^2$$

↪ Validation par validation croisée (Leave-One Out) (moins bien, mais gratuit) :

$$\begin{aligned} \mathcal{E}_{LOO} &= \frac{1}{n} \sum_{k=1}^{n} \left(\hat{\boldsymbol{a}}[\boldsymbol{X}^{(-k)}]^T \boldsymbol{\Phi}(\boldsymbol{X}^{(k)}) - f(\boldsymbol{X}^{(k)})\right)^2 \\ &= \frac{1}{n} \sum_{k=1}^{n} \left(\frac{\hat{\boldsymbol{a}}^T \boldsymbol{\Phi}(\boldsymbol{X}^{(k)}) - f(\boldsymbol{X}^{(k)})}{1 - q^{(k)}}\right)^2 \text{ avec } q^{(k)} = \left(\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\right)_{kk} \end{aligned}$$

• Quelques remarques.

• Décomposition bien adaptée pour représenter les entrées ou une combinaison linéaire des entrées, mais pas forcément une fonction non-linéaire de ceux-ci. La décomposition converge d'autant plus vite que la fonction est régulière.

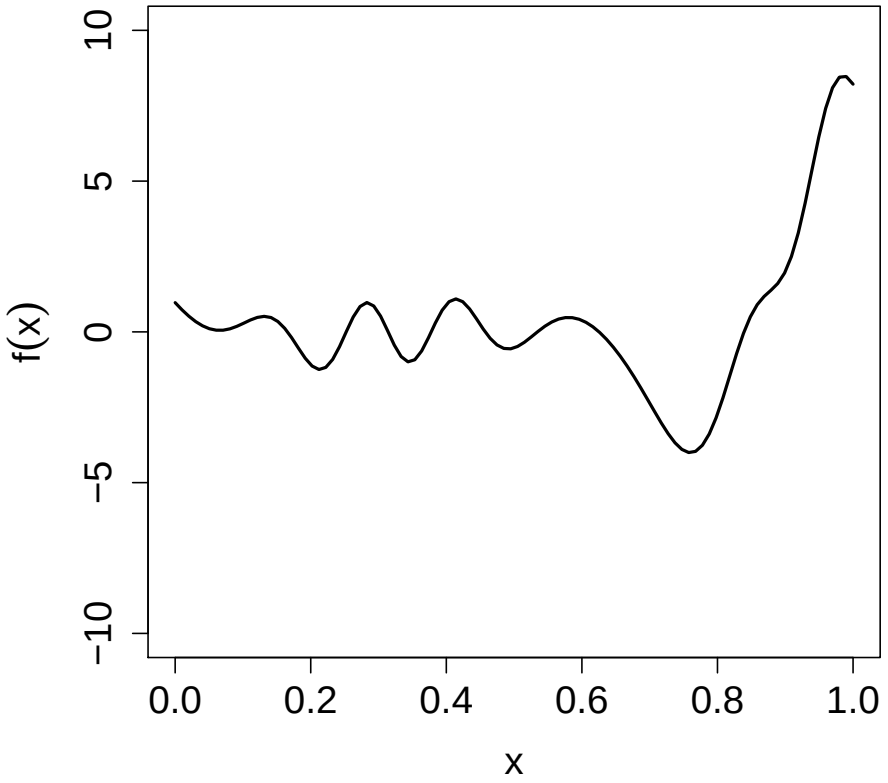
• Travail consiste à estimer les coefficients $(a_j)_j$ par méthodes intrusives ou non-intrusives.

• Généralisation à des distributions autres que gaussiennes. Les polynômes $\phi_.$ sont alors modifiés.

• Avantages :
- Calcul efficace de la sensibilité de la solution aux paramètres d'entrée incertains
- Obtention d'une forme explicite de la solution (calcul facile de moments et de densité de probabilité)

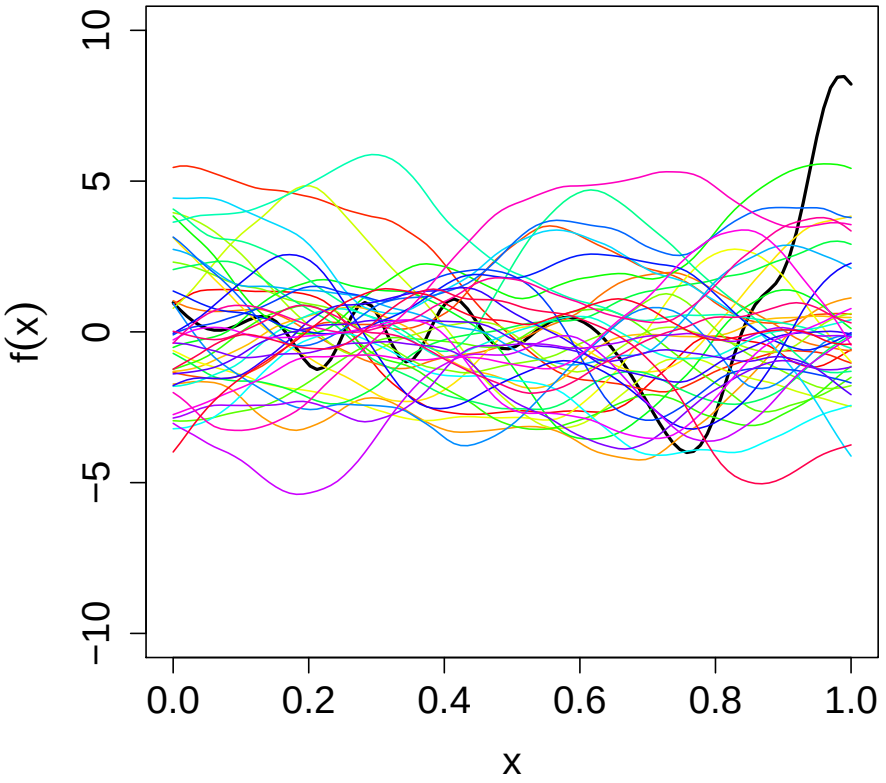
• Inconvénients.
- Mauvais contrôle de l'erreur (de troncation).
- Méthode mal adaptée à des cas où le nombre de paramètres d'entrée incertains est grand.

Krigeage



- But : construire un métamodèle de $f(x)$.

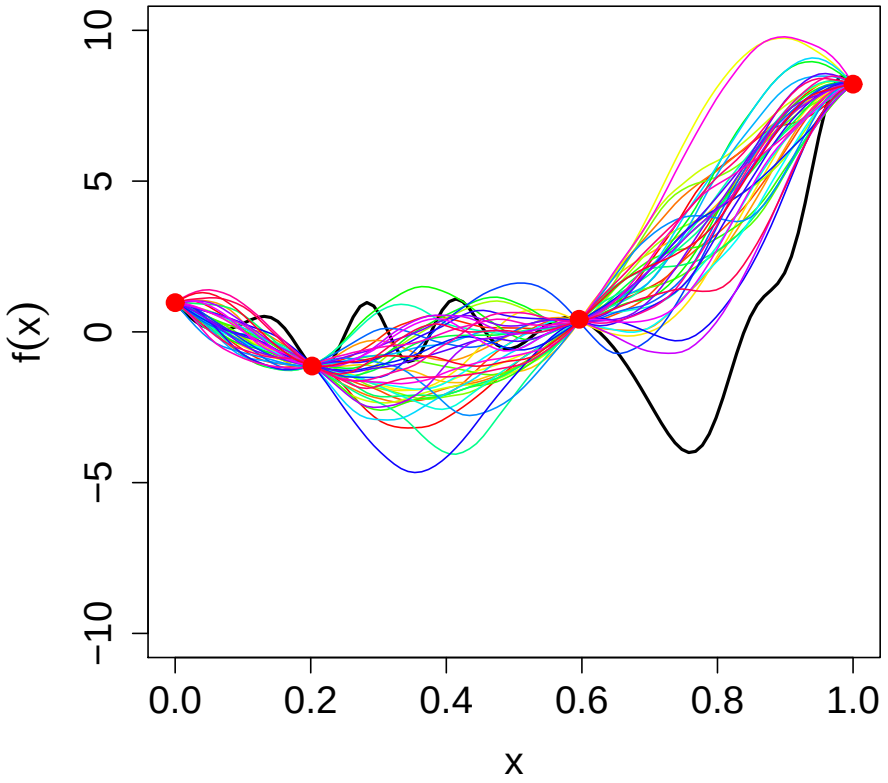
Krigeage



- Idée : on suppose que $f(x)$ une réalisation d'un processus gaussien $Z(x)$ avec moyenne $m(x)$ et covariance $k(x, x')$.

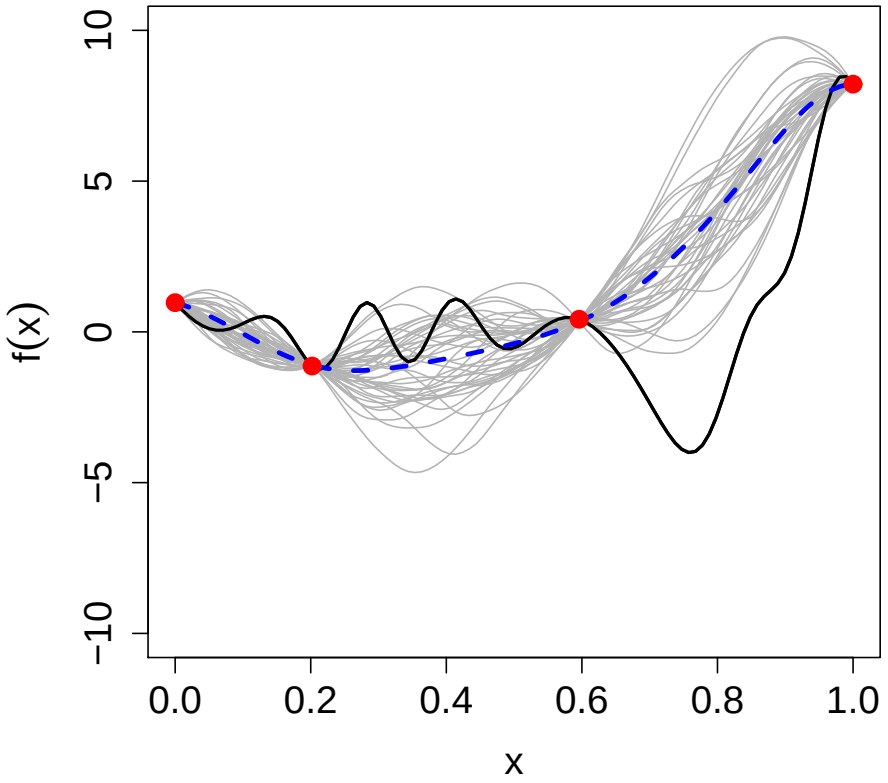
Par exemple $m(x) = \beta_0$ et $k(x, x') = \sigma^2 \exp(-\frac{(x-x')^2}{\theta^2})$.

Krigeage



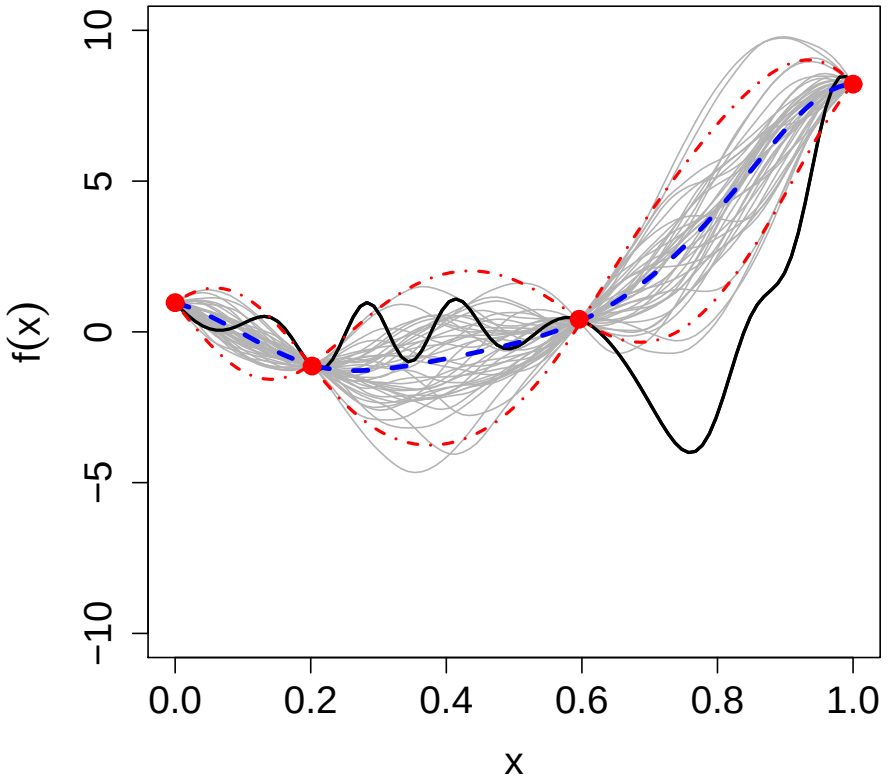
- On suppose qu'on observe $Z(x)$ aux points $x_1, \ldots, x_n \in \mathbb{R}$.
- On veut prédire $Z(x)$ étant données les observations $Z(x_1), \ldots, Z(x_n)$.

Krigeage



- On utilise le Best Linear Unbiased Predictor:
  (BLUP) : $\hat{f}(x) = m(x) + \boldsymbol{k}(x)^T \mathbf{K}^{-1} \boldsymbol{z}^n$.
  où $\boldsymbol{k}(x) = [k(x, x_i)]_{i=1,\ldots,n}$,
  $\mathbf{K} = [k(x_i, x_j)]_{i,j=1,\ldots,n}$,
  $\boldsymbol{z}^n = (f(x_1), \ldots, f(x_n))$.
- Ligne bleue : BLUP

Krigeage



:

- Le BLUP est le meilleur parce qu'il minimise l'erreur quadratique moyenne

$$\hat{\sigma}^2(x) = k(x,x) - \boldsymbol{k}(x)^T \mathbf{K}^{-1} \boldsymbol{k}(x)$$

- Ligne bleue : BLUP
- Lignes rouges : intervalles de confiance

• Quelques remarques.

• Il faut déterminer la tendance $m(x)$ et la fonction de covariance $k(x, x')$.
- Pour la tendance, on considère $m(x) = \sum_{j=1}^{p} \beta_j f_j(x)$ avec f_j données (monômes par exemple) et on estime les β_j (avec leur incertitude).
- Pour la fonction de covariance, on suppose qu'elle appartient à une famille paramétrique $k(x, x'; \theta)$ et on estime θ par maximum de vraisemblance (méthode plug-in) ou bien on utilise une approche bayésienne complète (lourd).

• Avantages :
- Fournit des intervalles de confiance, des erreurs de généralisation.
- Obtention d'une forme explicite du BLUP.
- Permet l'implémentation de stratégie d'échantillonnage séquentielle, pour l'optimisation par exemple.
- Permet de traiter des cas bruités ou des fonctions peu régulières (la famille de fonctions de covariance peut comporter un paramètre de régularité, par exemple la famille de Matérn).

• Inconvénients :
- Méthode mal adaptée lorsque le nombre de paramètres d'entrée est grand.
- Méthode lourde lorsque le nombre de points est grand.

C. E. Rasmussen and C. Williams, Gaussian Processes for Machine Learning.