

Introduction au traitement des incertitudes

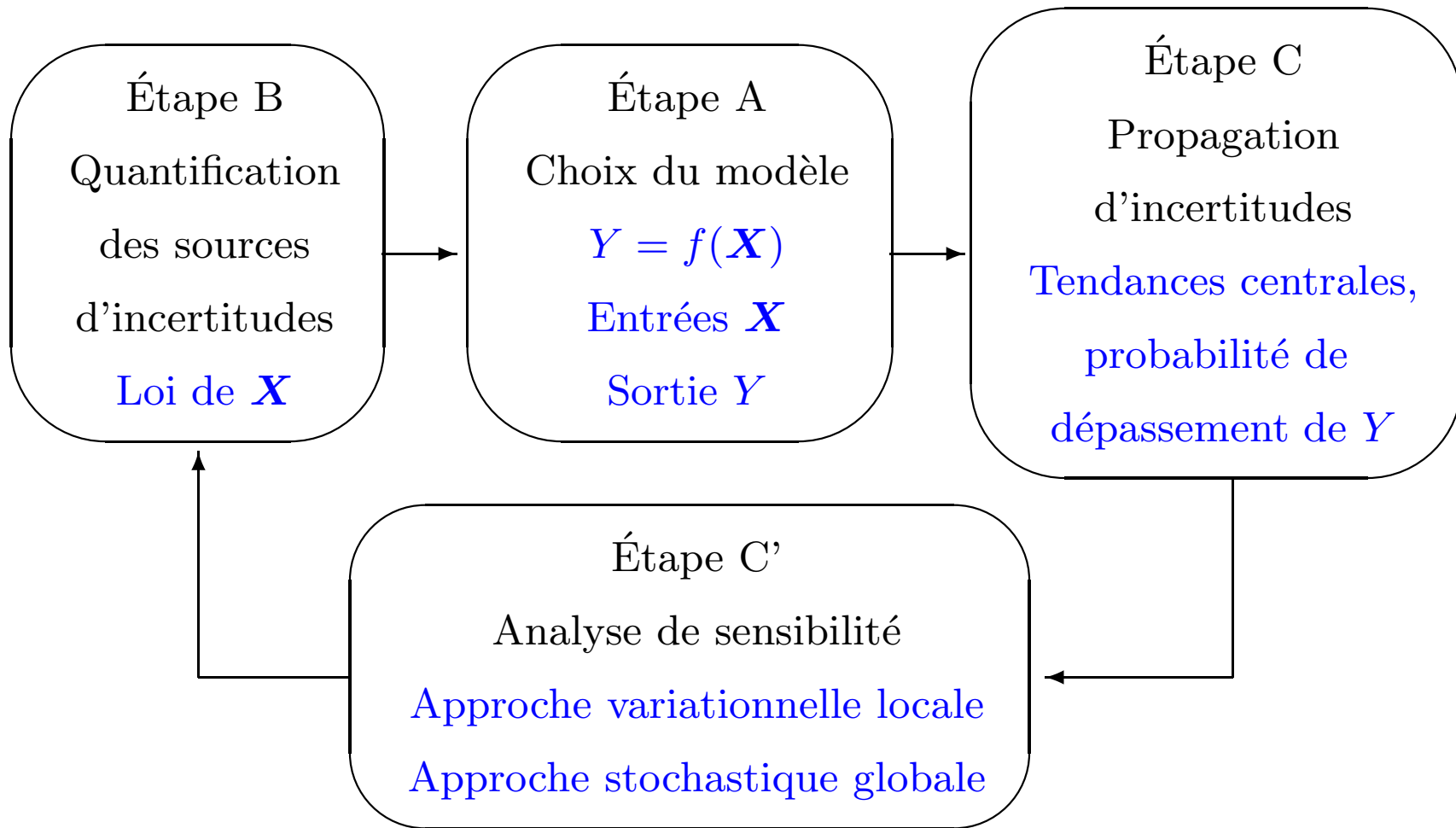
- Problème général :

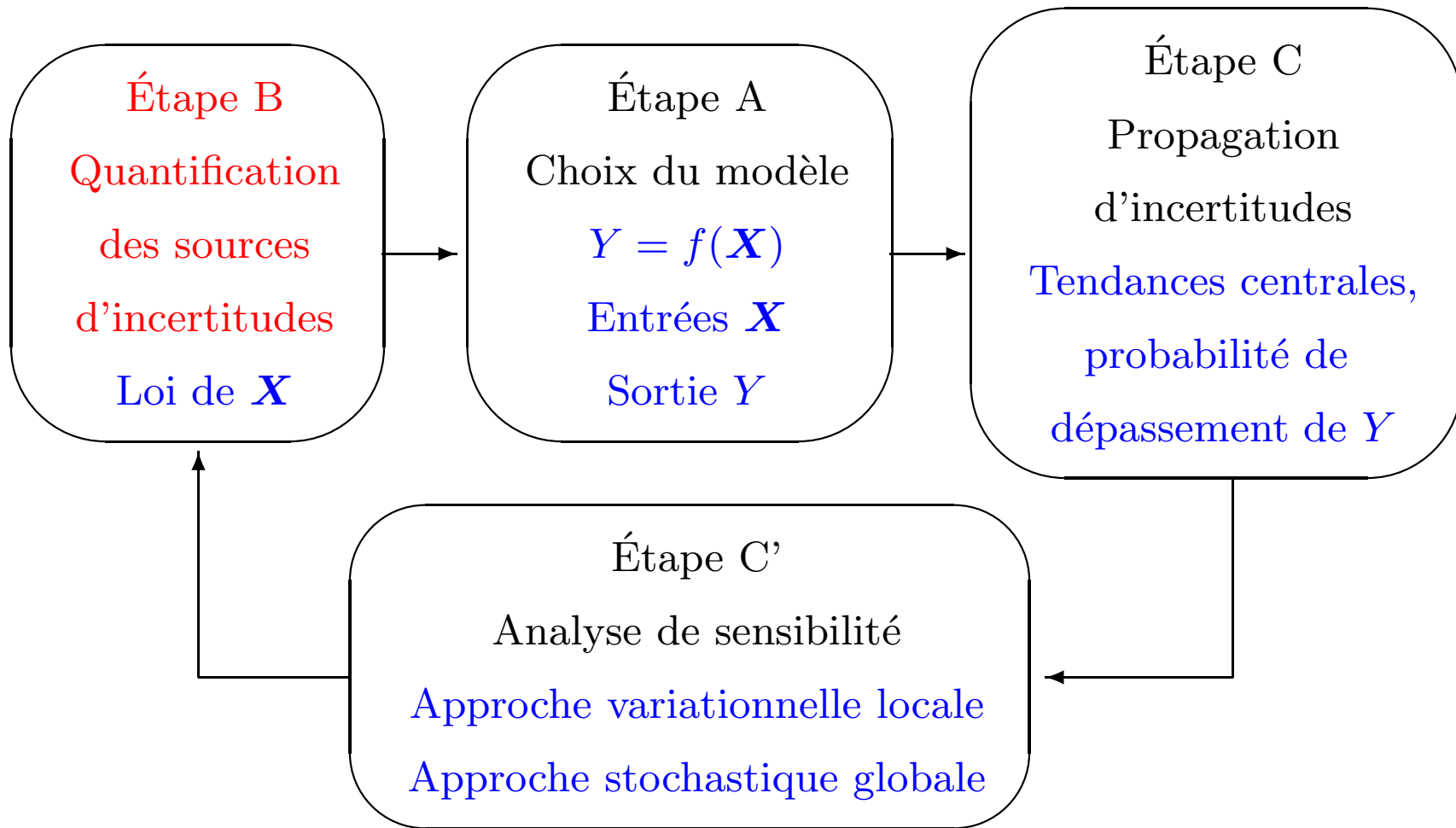
Comment modéliser les incertitudes et leur propagation dans les modèles physiques ou numériques ?

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 ?

- But de l'exposé : faire un rapide survol de l'état de l'art.





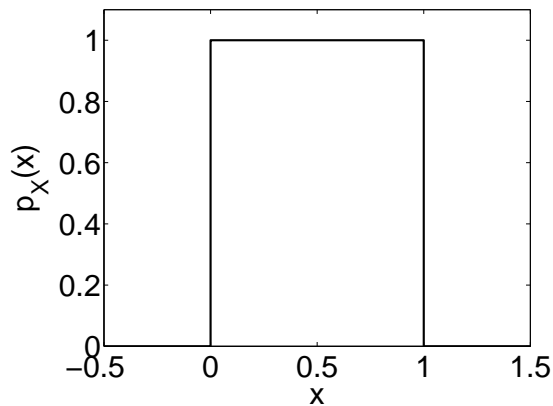
Partie I. Les sources d'incertitudes

- 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).

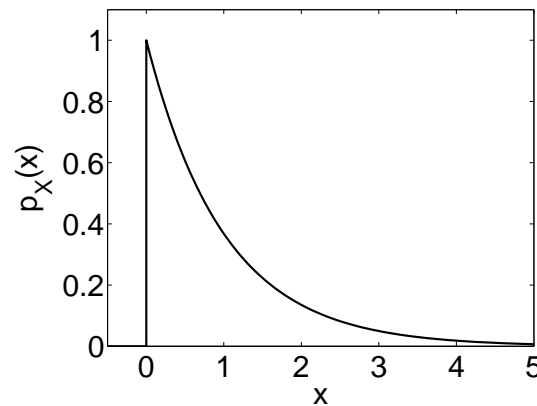
- Modélisation par des variables aléatoires.
- Les variables d'entrées sont traitées comme des variables aléatoires, de lois de probabilités données.
- La loi de probabilité d'une variable aléatoire réelle X caractérise la probabilité $\mathbb{P}(X \in [a, b])$ pour tout $a < b$.

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 $(p(x))_{x \in \mathbb{R}}$

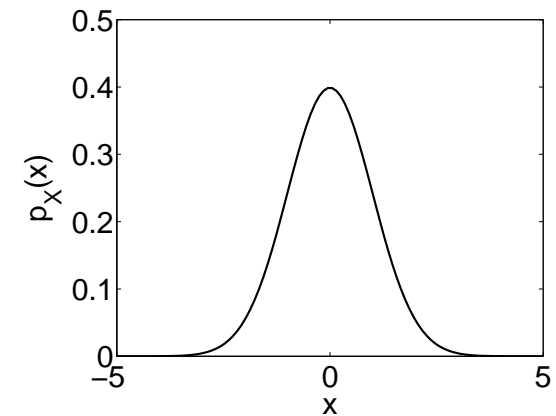
$$\mathbb{P}(X \in [a, b]) = \int_a^b p(x) dx, \quad \mathbb{P}(X \in [x, x + \delta x]) \simeq p(x) \delta x$$



uniforme



exponentielle



gaussienne

- Modélisation par des variables aléatoires.
- 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).
- La modélisation est basée sur la théorie des probabilités.
D'autres modélisations sont possibles; elles sont basées sur des théories de l'incertain dédiées à l'information imprécise :
 - distributions de possibilité, aussi appelées intervalles flous,
 - intervalles aléatoires utilisant les fonctions de croyance de Dempster-Shafer.

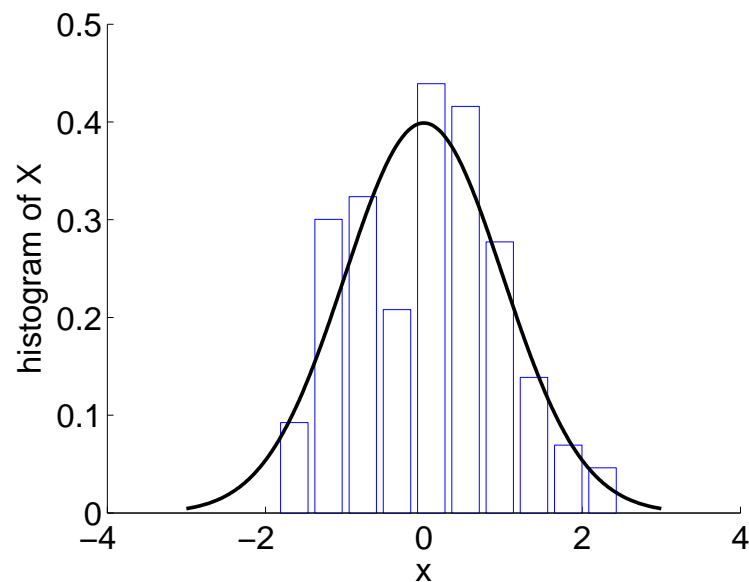
Cf: T. Hastie, R. Tibshirani et J. Friedman, *The Elements of Statistical Learning*, Springer, 2001.

- **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 $(x_i)_{i=1,\dots,n}$ 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

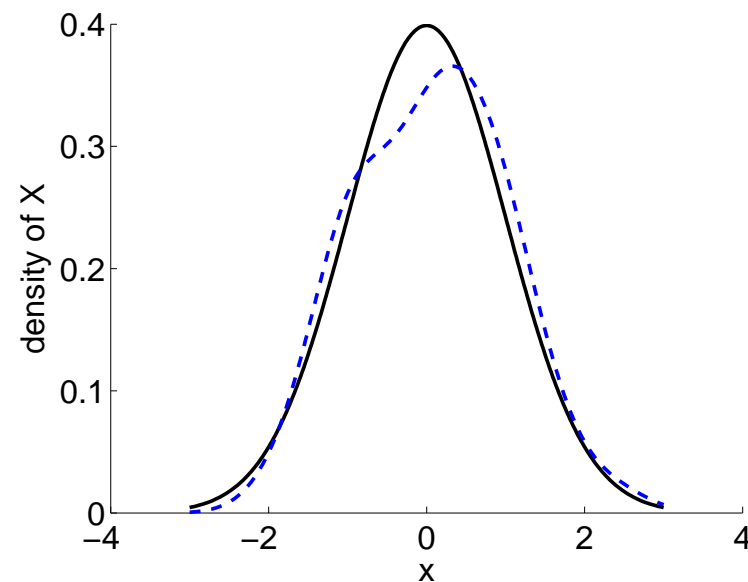
$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

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 $(x_i)_{i=1,\dots,n}$ 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

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

(gaussienne de moyenne μ et de variance σ^2).

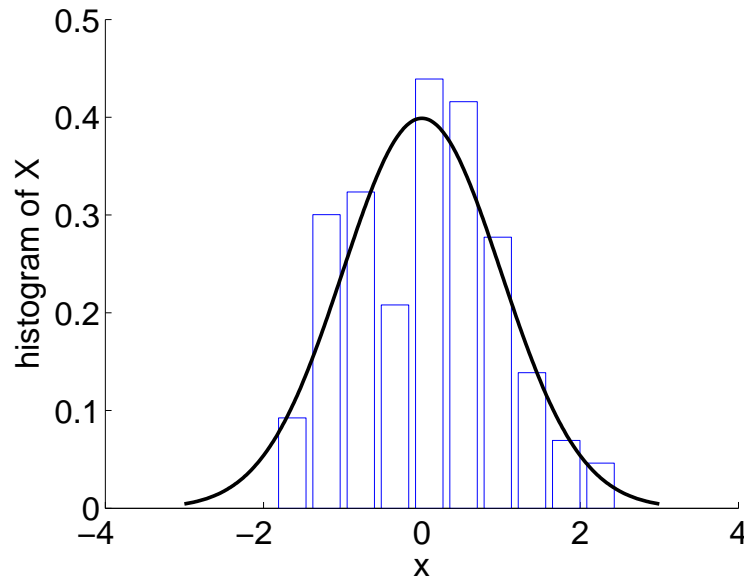
Nécessite un peu moins de données $(x_i)_{i=1,\dots,n}$ que la méthode à noyaux.

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

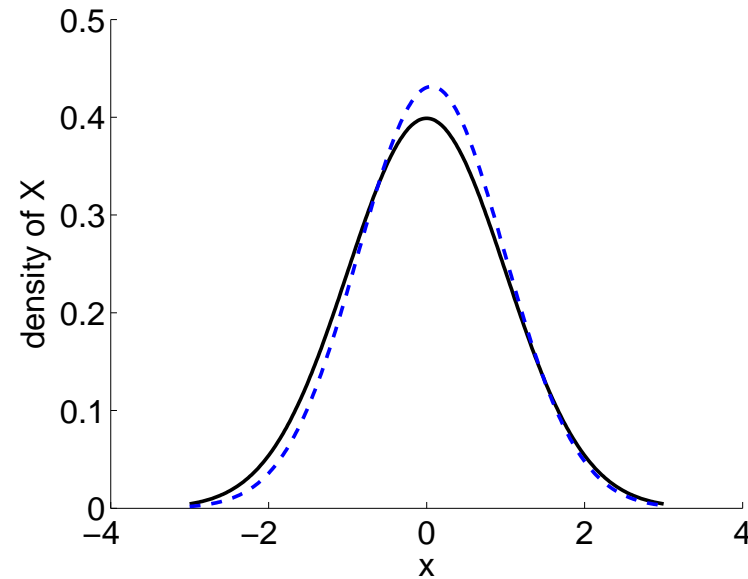
Exemple :

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

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 $\boldsymbol{\theta}$):

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

- $L_{\mathbf{x}}(\boldsymbol{\theta})$ est la vraisemblance des données $\mathbf{x} = (x_i)_{i=1, \dots, n}$ sachant le paramètre $\boldsymbol{\theta}$. Par le théorème de Bayes, c'est aussi la vraisemblance du paramètre $\boldsymbol{\theta}$ sachant les données \mathbf{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 cas standards, 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 : famille gaussienne $\boldsymbol{\theta} = (\mu, \sigma^2)$:

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

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

- On termine par un test de qualité d'ajustement, qui permet de dire si les données $(x_i)_{i=1,\dots,n}$ sont compatibles avec la famille paramétrique p_{θ} .

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)

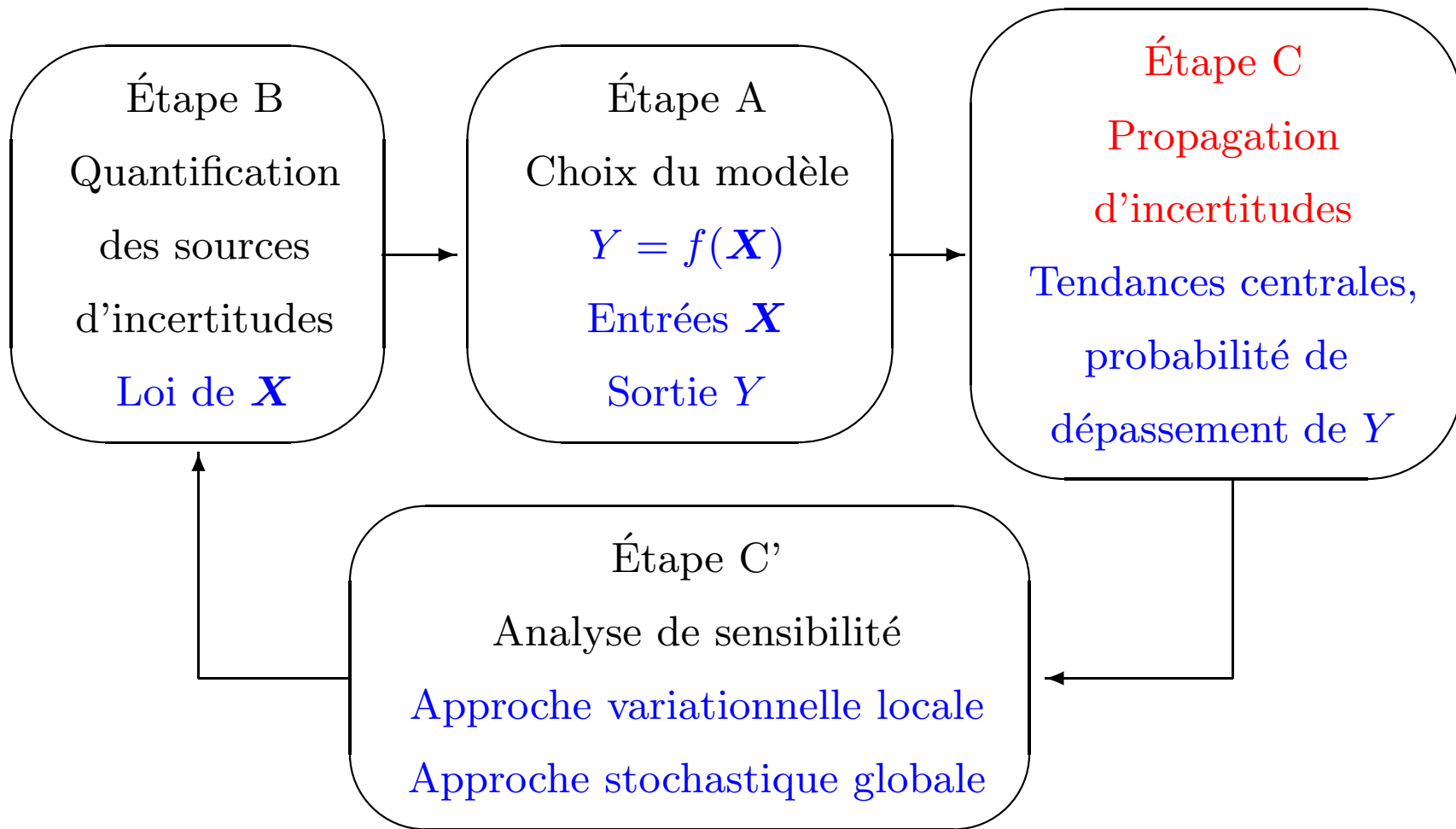
$$E(p) = - \int p(x) \ln p(x) dx$$

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.



Partie II. Propagation d'incertitudes

- Contexte : code de calcul informatique ou expérience modélisé par

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

avec Y =variable de sortie

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

f = boîte noire déterministe

On se donne : la loi de \mathbf{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. $\text{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$

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

Méthodes de quadrature

La quantité à estimer est une intégrale d -dimensionnelle :

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

Les méthodes de quadrature réclament

- $\mathbf{x} \rightarrow \psi(f(\mathbf{x}))$ régulier,
- une dimension d pas trop grande.

Méthodes très gourmandes en nombres d'appels au code.

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

$$\hat{I} = \sum_{j_1=1}^n \cdots \sum_{j_d=1}^n w_{j_1} \cdots w_{j_d} \psi(f(x_{j_1}, \dots, x_{j_d}))$$

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

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

Développements de Taylor - cumul quadratique

- 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.
- Approche "locale". En général, pas de contrôle de l'erreur.
- Exemple : on veut estimer $\mathbb{E}[Y]$ et $\text{Var}(Y)$ avec X_i décorrélées, $\mathbb{E}[X_i] = \mu_i$ et $\text{Var}(X_i) = \sigma_i^2$ connus, σ_i^2 petits. Alors on a juste besoin de $f(\boldsymbol{\mu})$ et $\nabla f(\boldsymbol{\mu})$

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

Preuve : $Y = f(\boldsymbol{\mu}) + \nabla f(\boldsymbol{\mu}) \cdot (\mathbf{X} - \boldsymbol{\mu}) + \dots$.

Calcul du gradient par différences finies, ou par différentiation automatique, ou par calcul adjoint.

Méthode de Monte Carlo

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

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n \psi(f(\mathbf{X}^{(k)}))$$

Estimation non-biaisée :

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

Convergence :

$$\hat{I}_n \xrightarrow[n \rightarrow \infty]{} I \quad \text{avec probabilité 1}$$

Erreur (risque quadratique) :

$$\mathbb{E}[(\hat{I}_n - I)^2] = \text{Var}(\hat{I}_n) = \frac{1}{n} \text{Var}(\psi(Y))$$

Erreur relative :

$$\frac{\mathbb{E}[(\hat{I}_n - I)^2]^{1/2}}{I} = \frac{1}{\sqrt{n}} \frac{\text{Var}(\psi(Y))^{1/2}}{\mathbb{E}[\psi(Y)]}$$

Avantages :

- 1) possibilité d'obtenir des intervalles de confiance,
- 2) pas de régularité requise sur f , ψ ,
- 3) vitesse de convergence indépendante de la dimension (mais lente).

Méthodes de Monte Carlo accéléré

- **Techniques de réduction de variance**

On cherche à réduire la constante dans l'erreur relative (qui reste en $1/\sqrt{n}$).

- Echantillonnage préférentiel (importance sampling) : on tire les $\mathbf{X}^{(k)}$ selon une densité "biaisée" qui favorise les valeurs de \mathbf{X} dans la zone d'importance.
- Variables de contrôle : on dispose d'un modèle réduit et on travaille sur la différence.
- Stratification : on force l'échantillon à respecter exactement des proportions imposées dans certaines "strates".

- **Suite à discrépance faible** (quasi Monte Carlo).

On cherche à diminuer le facteur $1/\sqrt{n}$ de l'erreur.

Principe : On tire l'échantillon de manière moins aléatoire que Monte Carlo, pour combler les trous qui se forment naturellement dans un échantillon aléatoire.

- réduit l'erreur si $\mathbf{x} \rightarrow \psi(f(\mathbf{x}))$ a un peu de régularité; on peut aller jusqu'à une erreur en $C_d(\log n)^{d/2}/n$,
- marche en dimension pas trop grande,
- représente un intermédiaire entre Monte Carlo et quadrature usuelle,
- ne donne pas d'intervalle de confiance.

Cf: R. Y. Rubinstein, Simulation and the Monte Carlo Method, Wiley, 1981.

Estimation particulière : probabilité de défaillance

On cherche à estimer

$$p_s = \mathbb{P}(Y \geq y_s)$$

avec y_s grand si bien que $p_s \ll 1$.

• Possible par Monte Carlo :

$$\hat{p}_s = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{f(\mathbf{X}^{(k)}) \geq y_s}$$

Erreur relative :

$$\frac{\mathbb{E}[(\hat{p}_s - p_s)^2]^{1/2}}{p_s} = \frac{\sqrt{1 - p_s}}{\sqrt{p_s} \sqrt{n}}$$

Il faut donc $np_s > 1$ pour avoir une erreur relative plus petite que 1.

↪ Utiliser une technique de réduction de variance.

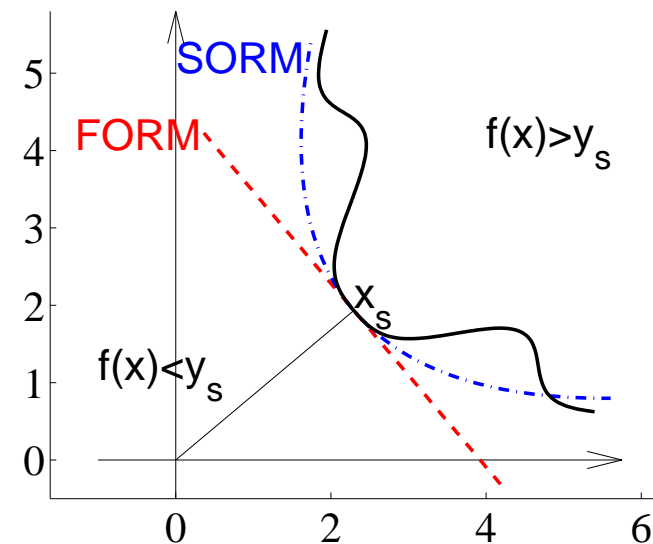
Probabilité à évaluer :

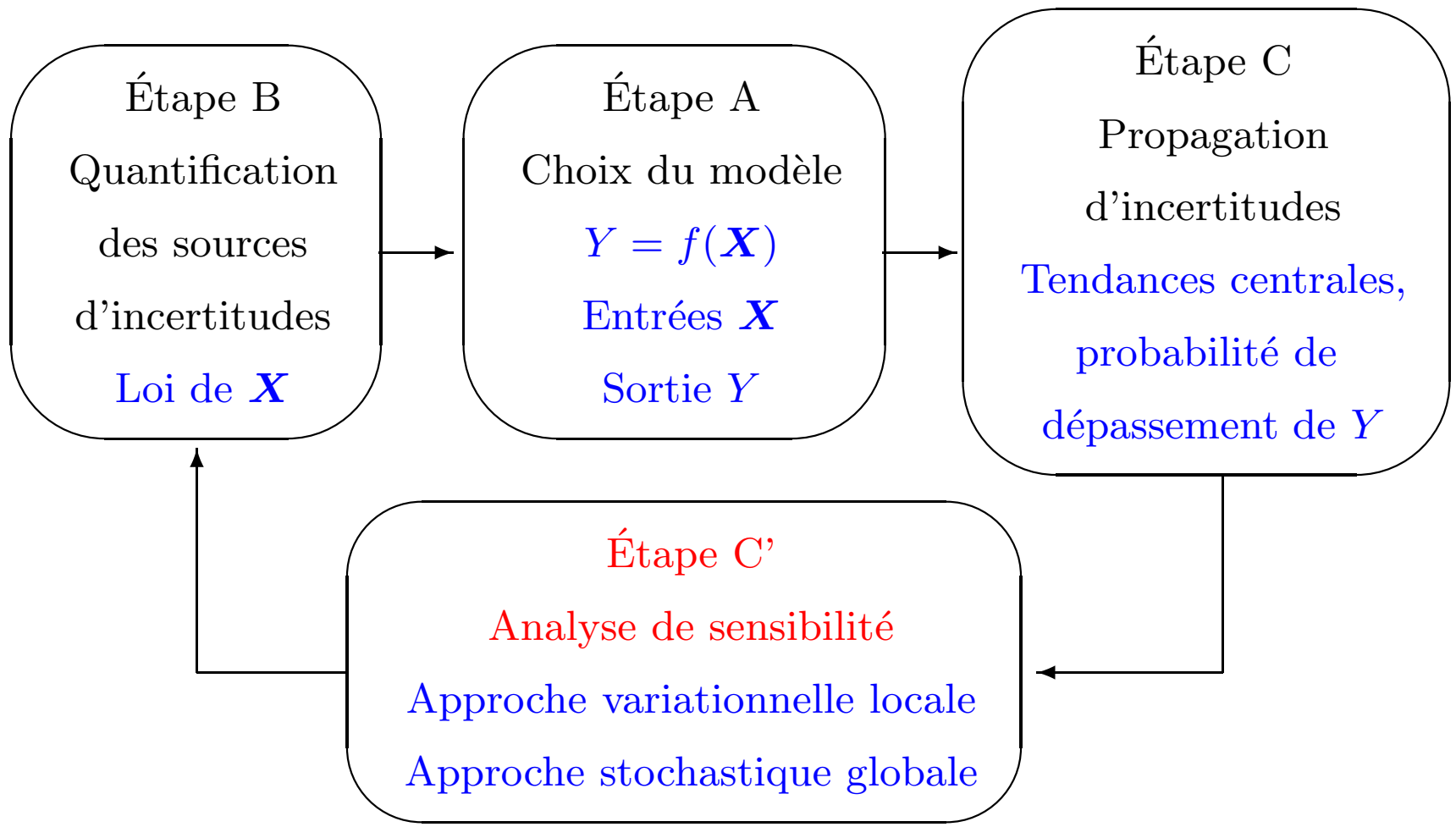
$$p_s = \mathbb{P}(Y \geq y_s) = \mathbb{P}(\mathbf{X} \in F) = \int_F p(\mathbf{x})d\mathbf{x}, \quad F = \{\mathbf{x} \in \mathbb{R}^d, f(\mathbf{x}) \geq y_s\}$$

- 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 \mathbf{x}_s de dépassement de seuil (i.e. $f(\mathbf{x}_s) = y_s$) le plus proche de l'origine.
- on approche la surface de défaillance $\{\mathbf{x} \in \mathbb{R}^d, f(\mathbf{x}) = y_s\}$ par une surface régulière \hat{F} qui permette de faire un calcul analytique $\hat{p}_s = \int_{\hat{F}} p(\mathbf{x})d\mathbf{x}$:

- un hyperplan pour FORM
(et alors $\hat{p}_s = \Phi(-|\mathbf{x}_s|)$),
- une forme quadratique pour SORM
(et alors $\hat{p}_s \simeq$ Breitung's formula).

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





Partie III. Analyse de sensibilité

- Contexte : code de calcul informatique ou expérience modélisé par

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

avec Y =variable de sortie réelle

$\mathbf{X} = (X_1, \dots, X_d)$ 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 $n = d + 1$.

Rapide mais dangereux.

- méthode de Morris avec $n = R(d + 1)$. 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).

- criblage par groupes, bifurcation séquentielle (avec des hypothèses de monotonie) avec $n < d$ et $h \ll d$. Permet d'identifier les quelques h paramètres importants en très grande dimension.

Indices de sensibilité pour un modèle linéaire

- Soit le modèle linéaire

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

avec $X_i \in \mathbb{R}$ décorrélés.

- La variance de Y s'écrit

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

La part de variance de Y due à X_i est $\alpha_i^2 \text{Var}(X_i)$.

→ 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) :

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

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

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

de manière équivalente :

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

- Estimation :

- estimation des α_i (OAT différences finies, régression linéaire).
- Monte Carlo : on tire $\mathbf{X}^{(k)}$, on calcule $Y^{(k)} = f(\mathbf{X}^{(k)})$, et on estime

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

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

Indices de sensibilité pour un modèle monotone

- Soit le modèle monotone

$$Y = f(X_1, \dots, X_d)$$

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

- Coefficient de corrélation basé sur les rangs :

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

- 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 :

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

Indices de sensibilité 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 $Y = f(X_1, \dots, X_d)$ avec $X_i \in \mathbb{R}$ et $X_i \perp X_j, i \neq j$.
- Le modèle peut se décomposer (Sobol) en

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

avec

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

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 :

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

où

$$D_i = \text{Var}(E[Y|X_i])$$

$$D_{ij} = \text{Var}(E[Y|X_i, X_j] - E[Y|X_i] - E[Y|X_j])$$

• Indices de sensibilité :

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

avec

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

• S_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)

• S_{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) .

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

On introduit alors l'indice de sensibilité total

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

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_{T_i} .

- Si S_{T_i} 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_{T_i} 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 discrédance faible (suites de Sobol $LP\tau$).
- Estimations gourmandes en nombre d'appels.
 - ↔ Utilisation d'un métamodèle (en particulier, polynômes de chaos). Mais on estime alors la sensibilité du métamodèle.

- Questions complémentaires (largement ouvertes).

Incertitude de modèle :

Que deviennent les indices de sensibilité estimés si le modèle change (mutation) ?

↳ quelques résultats sur quelques modèles de mutation (additifs).

Comment prendre en compte, dans les résultats de sensibilité, l'utilisation d'un modèle simplifié ?

↳ 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 ?

↳ se ramener à des variables décorréées.

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. Dreesbeke, 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).

↪ Chaque méthode a un (des) plan(s) d'expériences approprié(s).

Plans pour des surfaces de réponse polynomiales

- Supposons \mathbf{X} de dimension d à moyenne nulle. La "vraie" fonction est :

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

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

Modèle polynomial (degré 1) :

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

Modèle polynomial (degré 2) :

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

- 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).

Plans pour des surfaces de réponse paramétriques

Pas de plan d'expérience universel !

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

Plusieurs sens d'optimalité sont possibles (minimiser la variance d'estimation des coefficients du métamodèle, minimiser la variance de prédiction intégrée sur un domaine).

Nécessité de valider sur des points de validation (et donc de construire la surface de réponse avec seulement une partie des simulations/des expériences à disposition, ou bien d'utiliser des méthodes de validation croisée).

Attention aux extrapolations.

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).

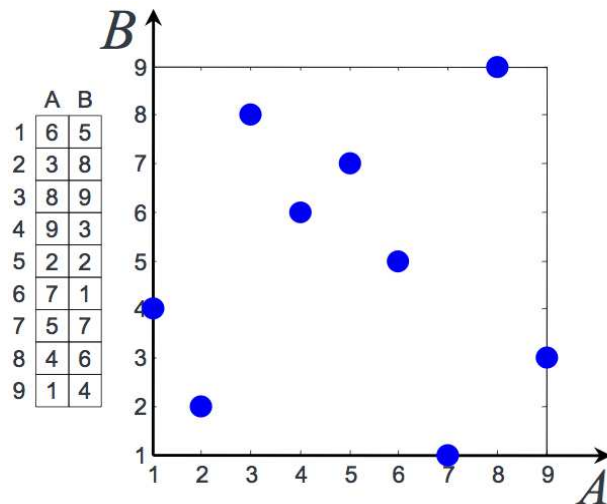
- Typiquement hypercubes latins (HL) (domaine d'intérêt : hypercube) :
 - Chaque facteur a le même nombre de niveaux (n).
 - Chacun des niveaux est pris une fois et une seule par chaque facteur.
- Critères de sélection parmi les plans space-filling $(\mathbf{x}_j)_{j=1,\dots,n}$:
 - Remplissage (maximin) : indicateur $D_{min} = \min_{1 \leq i \neq j \leq n} d(\mathbf{x}_i, \mathbf{x}_j)$ à maximiser.
 - Indépendance des facteurs : déterminant R de la matrice de corrélation à maximiser.
 - Uniformité (discrépance) : distance à la répartition uniforme à minimiser; discrédance L^2 à la loi uniforme sur l'hypercube $[0, 1]^d$:

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

- Méthodes de sélection :

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

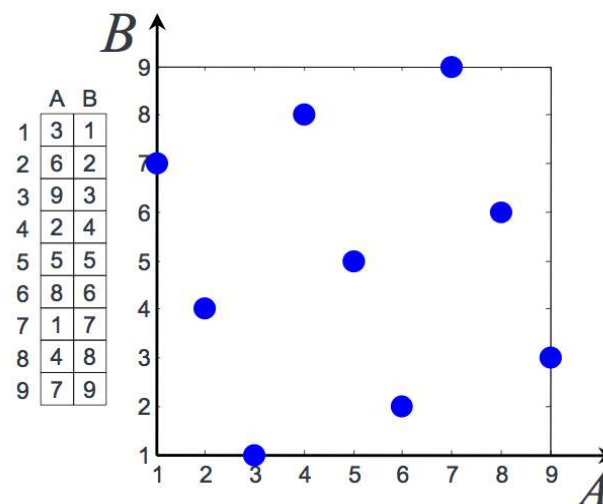
- Méthodes évoluées : Il existe dans la littérature différentes heuristiques (recuit simulé, algorithmes d'échanges) pour optimiser les plans selon le critère que l'on souhaite.



$$D_{min} = 1.4142$$

$$R = 0.9989$$

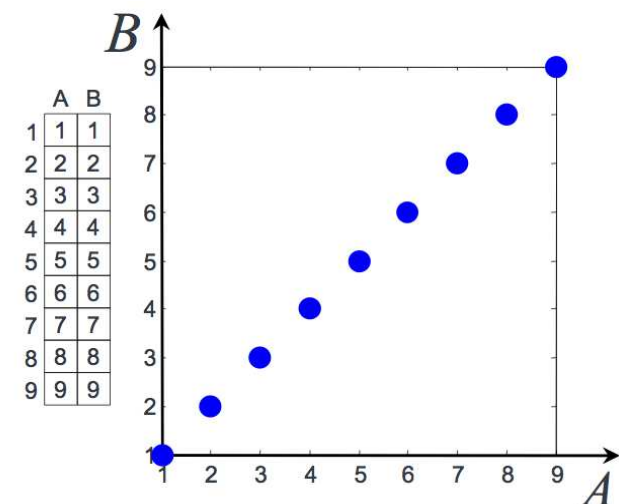
$$CL_2 = 4.19e-3$$



$$D_{min} = 3.1623$$

$$R = 1$$

$$CL_2 = 3.79e-3$$



$$D_{min} = 1.4142$$

$$R = 0$$

$$CL_2 = 13.68e-3$$

↪ on recommande souvent le critère maximin.

Polynômes de chaos

- Représentation de la solution $Y = f(\mathbf{X})$ sous la forme d'une série $\sum_j a_j \Phi_j(\mathbf{X})$ 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*, Springer Verlag, 1991. (Dover Publications, 2004).

- Polynôme de chaos (de Wiener).

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

Les polynômes d'Hermite ϕ , sont orthonormaux:

$$\mathbb{E}[\phi_{i_1\dots i_s}(X_{i_1}, \dots, X_{i_s})\phi_{j_1\dots j_t}(X_{j_1}, \dots, X_{j_t})] = \begin{cases} 1 & \text{si } (i_1\dots i_s) = (j_1\dots j_t) \\ 0 & \text{sinon} \end{cases}$$

La variable aléatoire $Y = f(\mathbf{X})$ s'écrit

$$Y = a_0\phi_0 + \sum_{i=1}^d a_{i_1}\phi_{i_1}(X_{i_1}) + \sum_{1 \leq i_1 \leq i_2 \leq d} a_{i_1 i_2}\phi_{i_1 i_2}(X_{i_1}, X_{i_2}) + \dots$$

→ La variable aléatoire Y est caractérisée par les constantes a .

En renumérotant les polynômes $(\Phi_j)_{j \geq 0} = (\phi_{i_1, \dots, i_n})_{i_1, \dots, i_n}$:

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

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

- 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.
- Le travail consiste à estimer les coefficients $(a_j)_j$ par méthodes intrusives ou non-intrusives.
- Généralisation à des distributions autres que gaussiennes. Les polynômes ϕ . 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 de moments et de densité de probabilité)
- Inconvénients.
 - Pas de contrôle de l'erreur de troncation.
 - Méthode mal adaptée à des cas où le nombre de paramètres d'entrée incertains est grand.

- Estimation des coefficients des polynômes de chaos.

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

1) On suppose que les entrées aléatoires sont des variables aléatoires (gaussiennes) indépendantes $(X_i)_{i=1}^d$.

2) On écrit la solution sous forme de sommes finies de PC :

$$Y = \sum_{i=1}^M a_i \Phi_i(\mathbf{X})$$

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

$$\mathbb{E} \left[\Phi_k(\mathbf{X}) \sum_{i=1}^M a_i \Phi_i(\mathbf{X}) \right] = \mathbb{E} [\Phi_k(\mathbf{X}) f(\mathbf{X})] , \quad k = 0, \dots, M$$

$$a_k = \mathbb{E} [\Phi_k(\mathbf{X}) f(\mathbf{X})] , \quad k = 0, \dots, M$$

→ estimation des a_k par régression ou par quadratures numériques (déterministes ou Monte Carlo ou quasi Monte Carlo).

