

Propagation Incertitudes

Fabrice Gaudier, Jean-Marc Martinez

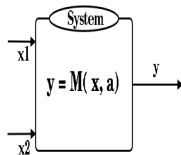
CEA-Saclay, DEN, DM2S, F-91191 Gif-sur-Yvette, France.

ATELIER ESNT
ORME DES MERISIERS, SACLAY
13 Février 2014

Plan

- ▶ Contexte
- ▶ Plan d'Expériences (Design of Experiments "DoE")
 - ▶ plan Factoriel
 - ▶ plan Stochastique (SRS, LHS)
 - ▶ plan à Discrépance Faible (qMC)
- ▶ Analyse de la quantité de la variable d'intérêt
 - ▶ Distribution
 - ▶ Moments de la loi (Moyenne, Ecart type)
 - ▶ Exploration de la relation "Entrée-Sortie"
 - ▶ Quantile
- ▶ Méta-Modèles
 - ▶ Réseaux de neurones
 - ▶ Krigeage
 - ▶ Polynômes de Chaos
- ▶ Présentation d'Uranie et Applications
 - ▶ Exploration et Analyse en grande dimension
 - ▶ Application Andra Stockage des déchets
 - ▶ Application Andra Stockage des déchets / GPU

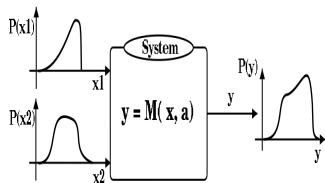
- ▶ **Simulation** de plus en plus de calculs
 - ▶ En optimisation dans les études de conception, de sûreté
 - ▶ Prise en compte des incertitudes épistémiques, stochastiques
 - ▶ En complément aux méthodes de sensibilité locale
- ▶ **Supervision** des codes \sim boîtes noires
 - ▶ Planifier l'exploration, modéliser la variabilité
 - ▶ Plans d'expériences numériques \rightarrow instancier les calculs
 - ▶ Réaliser les calculs
 - ▶ Récupérer, interpréter et analyser les résultats



Modèle

- ▶ x, y : Quantités physiques
- ▶ M : Modèle
- ▶ a : Paramètres des modèles

Expériences fournissent les données $\mathcal{B}(x_i, y_i)$ et les incertitudes $\mathcal{B}(\delta x_i, \delta y_i)$



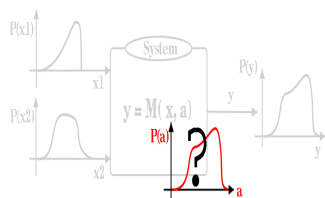
Modèle

- ▶ x, y : Quantités physiques
- ▶ M : Modèle
- ▶ a : Paramètres des modèles

Expériences fournissent les données $\mathcal{B}(x_i, y_i)$ et les incertitudes $\mathcal{B}(\delta x_i, \delta y_i)$

Questions soulevées

1. Quelle est l'incertitude de $y(x)$ sachant l'incertitude de x ?



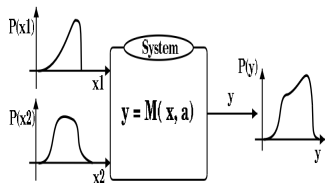
Modèle

- ▶ x, y : Quantités physiques
- ▶ M : Modèle
- ▶ a : Paramètres des modèles

Expériences fournissent les données $\mathcal{B}(x_i, y_i)$ et les incertitudes $\mathcal{B}(\delta x_i, \delta y_i)$

Questions soulevées

1. Quelle est l'incertitude de $y(x)$ sachant l'incertitude de x ?
2. Quelle est la distribution des paramètres a



Modèle

- ▶ x, y : Quantités physiques
- ▶ M : Modèle
- ▶ a : Paramètres des modèles

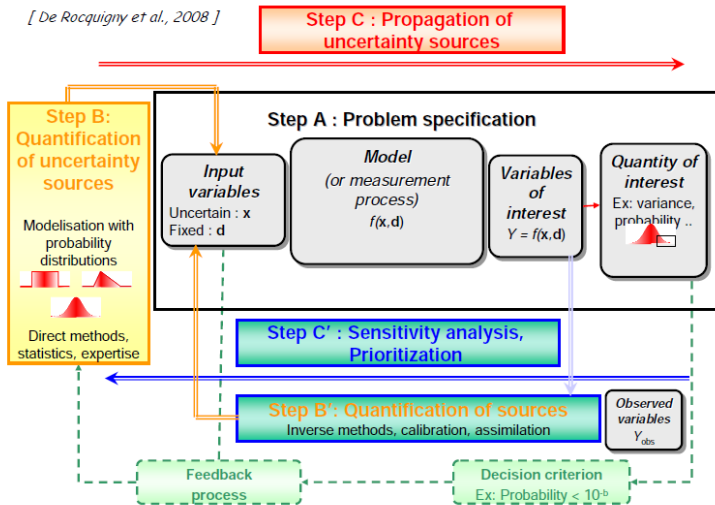
Expériences fournissent les données $\mathcal{B}(x_i, y_i)$ et les incertitudes $\mathcal{B}(\delta x_i, \delta y_i)$

Questions soulevées

1. Quelle est l'incertitude de $y(x)$ sachant l'incertitude de x ?
2. Quelle est la distribution des paramètres a
3. Quelles sont les variables x_k les plus influentes pour $y(x)$?

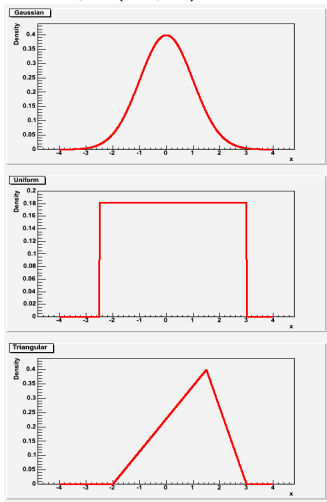
Traitement des Incertitudes

[De Rocquigny et al., 2008]



Modélisation des lois des entrées

Les marginales des Densité de Probabilité des variables d'entrées peuvent être obtenues par (Step B) :



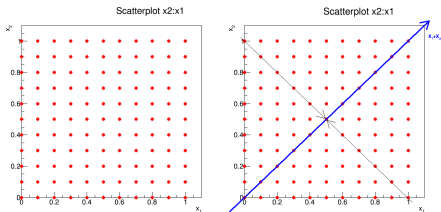
- ▶ Littérature scientifique
- ▶ Bornes physique
- ▶ Jugement d'expert
- ▶ base de données expérimentales (ajustement, tests statistiques)

Plan

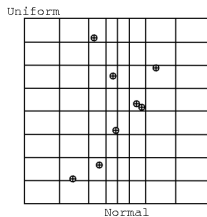
- ▶ Contexte
- ▶ Plan d'Expériences
 - ▶ plan Factoriel
 - ▶ plan Stochastique (SRS, LHS)
 - ▶ plan à Discrépance Faible (qMC)
- ▶ Analyse de la quantité de la variable d'intérêt
- ▶ Méta-Modèles
- ▶ Applications avec Uranie

Plan Factoriel

- ▶ Chaque variable possède sa propre suite de points
- ▶ "Coupler" chaque n-uplet de dimension $n - 1$ avec toutes les valeurs de la dernière composante
- ▶ Taille évolue de manière exponentielle ($\sim n^{nX}$)
- ▶ Pas une bonne solution d'un point de vue statistique (dépendance)

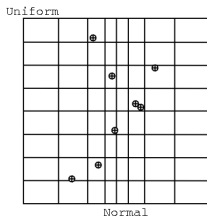


Simple Random Sampling - SRS

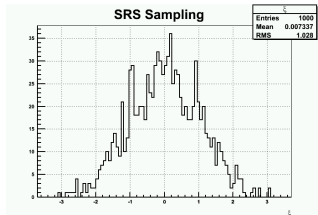


- ▶ Un échantillon x_k est tiré aléatoirement suivant sa densité de probabilité (PDF)
- ▶ Les échantillons x_k sont indépendants les uns des autres

Simple Random Sampling - SRS

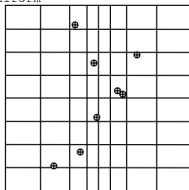


- ▶ Un échantillon x_k est tiré aléatoirement suivant sa densité de probabilité (PDF)
- ▶ Les échantillons x_k sont indépendants les uns des autres



Simple Random Sampling - SRS

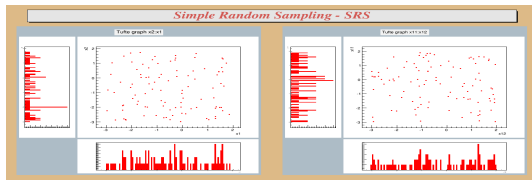
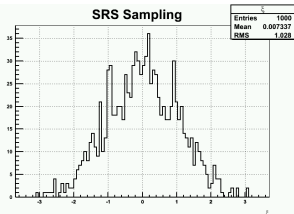
Uniform



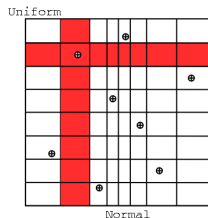
Normal

- ▶ Un échantillon x_k est tiré aléatoirement suivant sa densité de probabilité (PDF)
- ▶ Les échantillons x_k sont indépendants les uns des autres

SRS Sampling

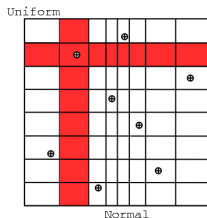


Latin Hypercube Sampling - LHS

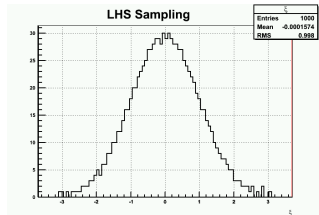


- ▶ Découper la distribution de chaque paramètre en strates S_1, S_2, \dots, S_{nS} de probabilités identiques
- ▶ Générer les échantillons
 - ▶ Une valeur de x_{jk} est aléatoirement sélectionnée par intervalle
 - ▶ Les valeurs de x_1 sont combinées aléatoirement sans remise avec les valeurs de x_2 pour former nS paires
 - ▶ Les paires précédentes sont également combinées aléatoirement sans remise avec les valeurs de x_3 pour former nS triplés
 - ▶ ainsi de suite jusqu'à produire nS échantillons

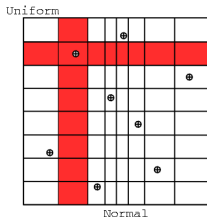
Latin Hypercube Sampling - LHS



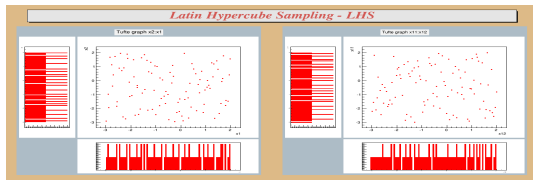
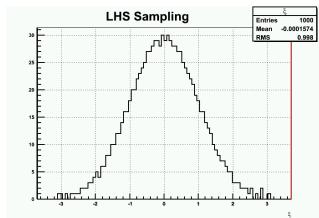
- ▶ Découper la distribution de chaque paramètre en strates S_1, S_2, \dots, S_{nS} de probabilités identiques
- ▶ Générer les échantillons
 - ▶ Une valeur de x_{jk} est aléatoirement sélectionnée par intervalle
 - ▶ Les valeurs de x_1 sont combinées aléatoirement sans remise avec les valeurs de x_2 pour former nS paires
 - ▶ Les paires précédentes sont également combinées aléatoirement sans remise avec les valeurs de x_3 pour former nS triplés
 - ▶ ainsi de suite jusqu'à produire nS échantillons



Latin Hypercube Sampling - LHS

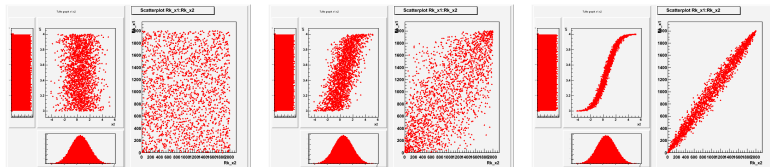


- ▶ Découper la distribution de chaque paramètre en strates S_1, S_2, \dots, S_{nS} de probabilités identiques
- ▶ Générer les échantillons
 - ▶ Une valeur de x_{jk} est aléatoirement sélectionnée par intervalle
 - ▶ Les valeurs de x_1 sont combinées aléatoirement sans remise avec les valeurs de x_2 pour former nS paires
 - ▶ Les paires précédentes sont également combinées aléatoirement sans remise avec les valeurs de x_3 pour former nS triplés
 - ▶ ainsi de suite jusqu'à produire nS échantillons



Corrélations

- ▶ Importance des corrélations entre les variables
 - ▶ Variables corrélées : corrélations appropriées
 - ▶ Variables indépendantes : corrélations nulles
- ▶ Difficulté d'établir une matrice de corrélation complexe
- ▶ Méthode d'Iman & Conover : corrélation sur les rangs
 - ▶ Indépendante des distributions
 - ▶ S'applique pour LHS et SRS



Comparaison des techniques d'échantillonnage

SRS

- ▶ Facile à développer et expliquer
- ▶ Estimateurs non-biaisés pour la moyenne, l'écart-type, *etc*
- ▶ Échantillons de grande taille non réalisables

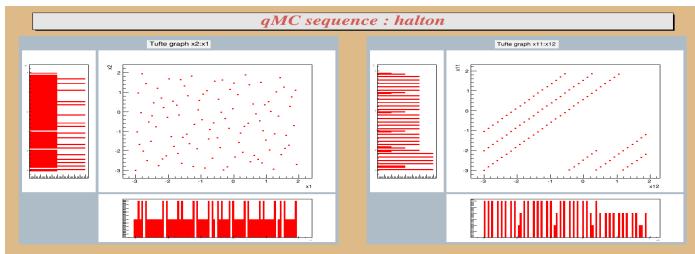
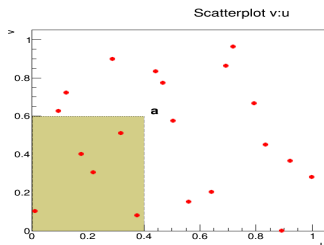
LHS

- ▶ Estimateurs non-biaisés pour la moyenne, l'écart-type , *etc*
- ▶ Stratification de chaque variable incertaine
- ▶ Utilisé lorsque des échantillons de grande taille sont non réalisables et l'estimation des quantiles élevés non exigés
- ▶ Résultats robustes pour des échantillons de relative petite taille ($nS = 50$ à 200)

quasi Monte-Carlo

- ▶ Générateurs déterministes utilisant la théorie des nombres
- ▶ Parfois possède une meilleure vitesse de convergence que MC ($n^{-1/2}$)
- ▶ Séquences à "discrépance faible"
Discrépance "étoile" notée D_n^* :

$$D_n^*(x_1, \dots, x_n) = \sup_{a \in [0,1]^n} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{0 \leq x_i \leq a} - |[0, a]| \right|$$

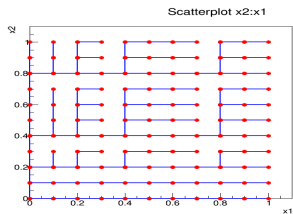
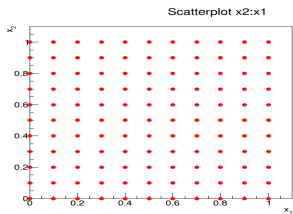


Classification des Plans d'Expériences Numériques

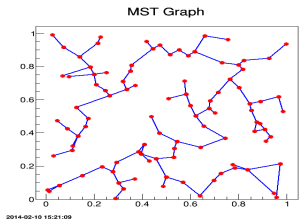
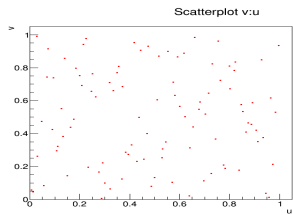
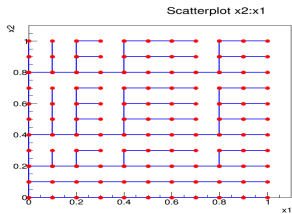
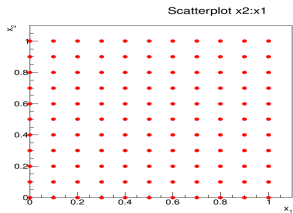
- ▶ Mesure de l'uniformité (Discrépance) d'un DoE ne suffit pas à elle seule pour mesurer la qualité d'un plan
- ▶ Besoin de mesurer également le remplissage de l'espace par le plan
- ▶ Difficile en grande dimension
- ▶ utilisation de critère synthétique
- ▶ Théorie des graphes et arbres de longueurs minimales (*Minimal Spanning Tree*, "MST") pour qualifier la répartition des points dans un espace multi-dimensionnel
- ▶ Utilise des méthodes issus de la théorie des graphes
- ▶ Permettra de regrouper les DoE suivant leurs structures (grille régulière, aléatoire, quasi Monte-Carlo,...)

- ▶ un **Graphe** $G = (X, U)$ est un couple constitué :
 - ▶ d'un ensemble de points $X = (x_1, \dots, x_n)$ appelés **nœuds** ;
 - ▶ d'une famille $U = (u_1, \dots, u_m)$ d'éléments de $X \times X = \{(x, y) | x \in X, y \in X\}$
- ▶ **Définitions**
 - ▶ un graphe **simple** s'il ne contient pas de boucles : les arcs sont alors appelés des **arêtes**
 - ▶ Un **chemin** d'un graphe simple est une séquence $\mu = (u_1, u_2, \dots, u_q)$ d'arêtes contiguës
 - ▶ un graphe est **connexe** s'il existe un chemin pour toute paire de points $(x, y) \in X \times X$
 - ▶ un **arbre** est un graphe simple, connexe *ne contenant pas de circuits*
 - ▶ un arbre est dit **maximal** s'il contient tous les points
 - ▶ un **arbre de longueur maximal** (*Minimal Spanning Tree, "MST"*) est un arbre maximal pondéré dont la longueur totale (somme de toutes ses arêtes) est minimale de tous les arbres maximaux.

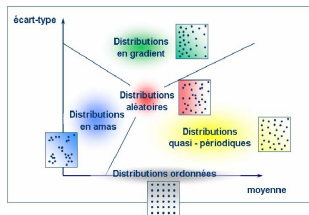
Représentations Graphiques



Représentations Graphiques



- ▶ Longueur d'un MST : Expression analytique (en asymptotique)
- ▶ Utilisation des algorithmes de Kruskal (1956) ou Prim (1957) pour rechercher un MST
- ▶ Existence de plusieurs solutions possibles
- ▶ Unicité de l'histogramme des longueurs d'arêtes (démonstré)
- ▶ Caractérisation par exemple par les deux premiers moments ; moyenne μ et écart-type σ
- ▶ Dans ce plan (μ, σ) , tout DoE occupe une position déterminée en fonction de sa structure



J. Franco, 2008

Supervision du Code de Calcul

Lancer de manière automatique le code de calcul sur des ressources informatiques

- ▶ séquentiel sur une machine
- ▶ distribué "HPC - *Embarrassingly Parallel*"

▶ Exemples d'outil informatique :

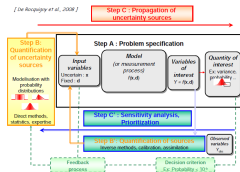
Salomé
Uranie
OpenTURNS
Dakota

▶ Principe :

Substitution dans le Jeu de Données (JDD) du code des valeurs courantes

Exécution du code de calcul avec ce JDD

Récupération des valeurs (scalaires ou temporelles) des variables de sorties



Plan

- ▶ Contexte
- ▶ Plan d'Expériences
- ▶ Analyse de la variable d'intérêt
 - ▶ Distribution
 - ▶ Moments de la loi (Moyenne, Ecart type)
 - ▶ Exploration de la relation "Entrée-Sortie"
 - ▶ Quantiles
- ▶ Méta-Modèles
- ▶ Applications avec Uranie

Propagation des Incertitudes

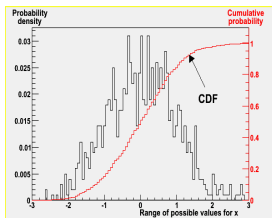
Question importante : Sous quelle forme est attendu le résultat ?

- ▶ Distribution de probabilité :
Simulation + ajustement + tests statistiques
- ▶ Premiers moments statistiques
 - ▶ Composition de la variance

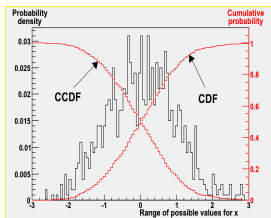
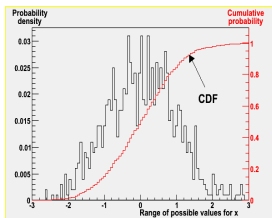
$$\sigma_{f(x_1, \dots, x_n)}^2 = \sum_i^n \left(\frac{\partial f}{\partial x_i} \sigma_{x_i} \right)^2 + 2 \sum_i^n \sum_{j=i+1}^n \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial f}{\partial x_j} \right) \text{cov}(x_i, x_j)$$

- ▶ Statistiques sur l'échantillon (convergence, bootstrap)
- ▶ Estimation par Intervalle de Confiance $\mathbb{P} [\mathbb{P} [m \leq Y \leq M] \geq \alpha] \geq \beta$
 - ▶ À partir de la distribution de probabilité
 - ▶ Formule de Wilks $1 - \alpha^N - N(1 - \alpha)\alpha^{N-1} \geq \beta$
 - ➡ N Calculs : m=min, M=max

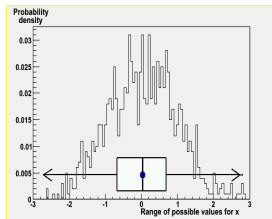
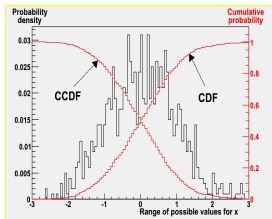
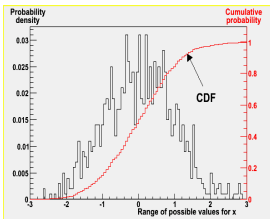
Estimation de la Distribution de Probabilité



Estimation de la Distribution de Probabilité



Estimation de la Distribution de Probabilité



Statistiques sur l'échantillon

- ▶ valeurs scalaires de la sortie $y_k = y(x_k)$, $k = 1, 2, \dots, nS$
- ▶ Moyenne

$$\mu_y = \sum_{k=1}^{nS} y_k / nS$$

- ▶ Écart-type

$$\sigma_y = \sqrt{\sum_{k=1}^{nS} (y_k - \mu_y)^2 / (nS - 1)}$$

- ▶ le coefficient d'asymétrie ou **skewness**

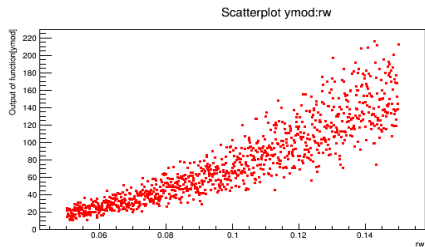
$$sk_y = \sum_{k=1}^{nS} (y_k - \mu_y)^3 / nS * \sigma_y^3$$

- ▶ le coefficient d'aplatissement ou **kurtosis**

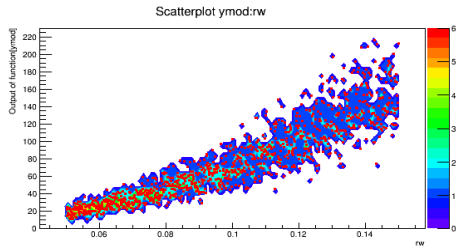
$$k_y = \sum_{k=1}^{nS} (y_k - \mu_y)^4 / nS * \sigma_y^4$$

- ▶ extrêmes (minimum et maximum), quantiles

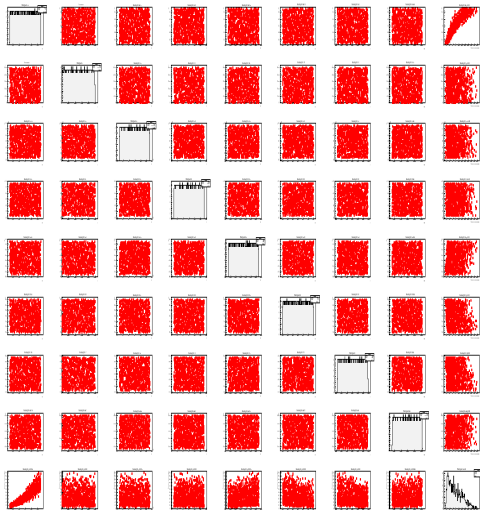
Exploration de la relation "Entrée-Sortie"



- Nuages de points y_j versus x_i

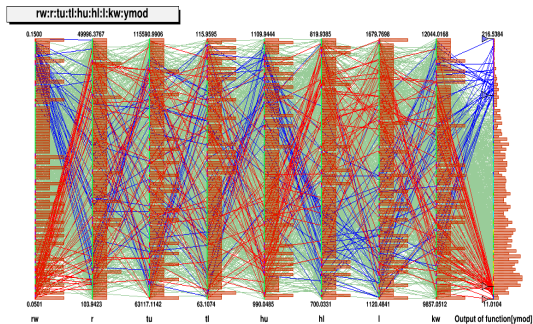


Exploration de la relation "Entrée-Sortie"



- ▶ Nuages de points y_j versus x_i
- ▶ Matrice de scatterplots "drawPairs"

Exploration de la relation "Entrée-Sortie"



- ▶ Nuages de points y_j versus x_i ;
- ▶ Matrice de scatterplots "drawPairs"
- ▶ graphique "CobWeb"

Introduction à l'estimation de quantile

- ▶ Etudes de sûreté : calculer la probabilité d'un scénario *critique* défini par le dépassement d'une valeur seuil particulière
- ▶ Par exemple pour une perte du réfrigérant primaire dans une centrale nucléaire quelle est la valeur de la température du combustible correspondant à la proportion des scénarii inférieurs à ce seuil critique ?

Soit α une probabilité critique donnée
calculer y_α tel que $\mathbb{P}(Y \leq y_\alpha) = \alpha$
par définition y_α est le α -quantile de Y

- ▶ On note $Y_{\alpha,n}$ l'estimateur du α -quantile y_α .
- ▶ Utilisation des statistiques d'ordre pour construire des intervalles de confiance de $Y_{\alpha,n}$
- ▶ Nous allons commencer par donner la loi asymptotique de $Y_{\alpha,n}$

Quantile empirique

- ▶ L'estimateur est basé sur les statistiques d'ordre.

$$(Y_1, Y_2, \dots, Y_n) \rightarrow (Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}) \text{ avec } Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$$

- ▶ En notant $\lceil x \rceil$ le plus petit entier $\geq x$, l'estimateur du α -quantile

$$Y_{\alpha, n} = Y_{(\lceil \alpha n \rceil)}$$

- ▶ En notant f_Y la densité de probabilité et y_α le α -quantile de Y , on a le résultat asymptotique :

$$Y_{\alpha, n} \xrightarrow{n \rightarrow \infty} \mathcal{N}\left(y_\alpha, \frac{\alpha(1-\alpha)}{n \times f_Y^2(y_\alpha)}\right)$$

- ▶ Ecart-type inversement proportionnel à \sqrt{n} et à la densité de probabilité $f_Y(y_\alpha)$, valeur en général très petite en queue de distribution mais surtout inconnue dans le cas général.

Introduction

- ▶ On peut donc songer à utiliser les méthodes basées sur les intervalles de confiance car l'estimateur $Y_{\alpha,n}$ tend vers une loi normale
- ▶ Malheureusement la variance de $Y_{\alpha,n}$ dépend de la valeur $f_Y(\alpha)$ inconnue et difficile à évaluer notamment en queue de distribution
- ▶ Pour les études de sûreté on préfère utiliser la méthode proposée par Wilks très pratique.
- ▶ Asymptotiquement on a autant de chance de surestimer que de sous estimer la valeur du quantile (car l'estimateur $Y_{[\alpha n]}$ tend vers une loi normale centrée en y_α)
- ▶ La méthode de Wilks impose un niveau de confiance β sur l'estimateur du quantile : **pour plus de sûreté**, on se prémunit d'une sous estimation du quantile

$$\mathbb{P}(Y_{(r)} \geq y_\alpha) \geq \beta$$

- ▶ Donc en prenant $\beta > 0.5$ nous allons estimer le quantile avec moins de chance de le sous estimer. En général pour les études de sûreté on prend $\beta = 0.95$.
- ▶ Pour ne pas surdimensionner le système, on retiendra la plus petite valeur r telle que $Y_{(r)}$ vérifie le niveau de confiance β

Quantile de Wilks

- ▶ La méthode de Wilks impose un niveau de confiance $P(Y_{(r)} > y_\alpha) > \beta$:

$$\sum_{k=r}^n C_n^k \alpha^k (1-\alpha)^{n-k} < 1-\beta$$

- ▶ **Propriété remarquable** : expression où tout est connu et indépendant de la loi de Y
- ▶ On doit donc rechercher la valeur de r minimale vérifiant l'inéquation. Il faut au moins que pour $r = n$ le niveau de confiance soit vérifié. Ce qui impose un nombre minimal de réalisations

$$\alpha^n < 1-\beta \Rightarrow n > \frac{\ln(1-\beta)}{\ln \alpha}$$

- ▶ Par exemple si on estime le quantile à 95% avec un niveau de confiance de 95%, il faut faire au moins 59 réalisations de Y ($59 > \ln(1-0.95)/\ln 0.95$).
- ▶ Et pour 59 tirages, le quantile de Wilks à 95% avec un niveau de confiance de 95% est $Y_{(59)}$, c'est à dire la valeur maximale des 59 réalisations.

Méthode de Wilks

- ▶ Le quantile de Wilks est donc la valeur $Y_{(r)}$ où le rang r est obtenu par la résolution :

$$\mathbb{P}(Y_{(r)} > y_\alpha) > \beta \quad \text{avec } r = \min_{i \in [0, n]} : \sum_{k=i}^n C_n^k \alpha^k (1-\alpha)^{n-k} < 1 - \beta$$

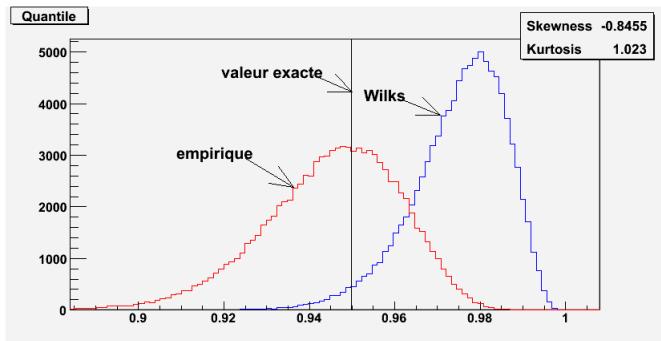
- ▶ En notant $F_{\mathcal{B}(n, \alpha)}$ la fonction de répartition de la loi binomiale $\mathcal{B}(n, \alpha)$, il faut donc trouver la valeur de r minimale (dans $[0, n]$) qui vérifie

$$\begin{aligned} r &= \min(i \in [0, n]) \text{ tel que } F_{\mathcal{B}(n, \alpha)}(i-1) > \beta \\ \Rightarrow r &= \lceil F_{\mathcal{B}(n, \alpha)}^{-1}(\beta) \rceil + 1 \end{aligned}$$

- ▶ Propriété : le quantile de Wilks ne dépend que de α, β, n et pas de la loi de Y .

Exemple $Y \sim \text{Loi Uniforme}$

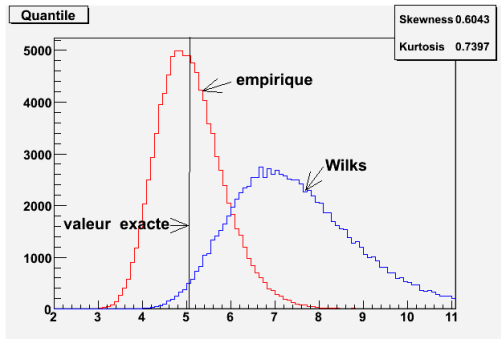
Soit $Y \sim U[0, 1] \Rightarrow y_{0.95} = 0.95$. Echantillon de taille 200 \Rightarrow estimateur du 0.95-quantile est la 191^{ème} valeur, et l'estimateur de Wilks au niveau 95% est la 196^{ème} valeur. Histogrammes via $10^6 \times 200$ -échantillons.



Le quantile de Wilks donne une valeur plus sûre dans 95% mais au détriment d'un biais (surestimation).

Exemple $Y \sim$ Loi Log Normale

Soit $Y = e^X$ avec $X \sim \mathcal{N}(0, 1) \Rightarrow y_{0.95} \simeq 5.18$. Echantillon de taille 200 \Rightarrow estimateur du 0.95-quantile empirique est la 191^{ème} valeur, celui de Wilks au niveau 95% est la 196^{ème} valeur. Histogrammes via $10^6 \times 200$ -échantillons.



Le quantile de Wilks donne une valeur plus sûre puisque la probabilité de sous-estimation du quantile n'est que de 5% au détriment évidemment d'un biais et d'une plus grande variance.

Plan

- ▶ Contexte
- ▶ Plan d'Expériences
- ▶ Analyse de la quantité da la variable d'intérêt
- ▶ Méta-Modèles
 - ▶ Réseaux de neurones
 - ▶ Krigeage
 - ▶ Polynômes de Chaos
- ▶ Applications avec Uranie

Méta-Modèles

Obtenir une fonction analytique qui simule le comportement d'un phénomène à partir d'un nombre réduit d'expériences numériques

- ▶ Exemples :
 - Régressions Linéaires, Polynômes, Réseaux de neurones,
 - Modèles Linéaires Généralisés ("GLM"), Krigeage, etc
- ▶ Objectifs :
 - Obtention d'un modèle de substitution au code,
 - Propagation d'incertitude,
 - Analyse de sensibilité globale,
 - Optimisation/Calibration,
 - Analyse fiabiliste
- ▶ Qualités attendues :
 - Représentativité du code,
 - Bonnes capacités de prédiction,
 - Temps de calcul négligeable
- ▶ Développement de méthodes visant à réduire le nombre d'appels au code numérique en général coûteux en temps de calcul

Neurone Formel (Mc Culloch et Pitts - 1943)

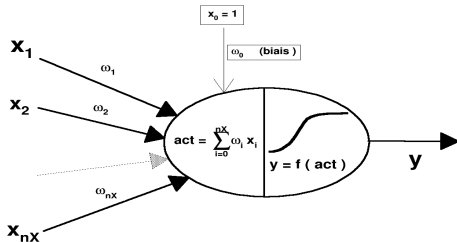
- ▶ Analogie au modèle biologique
- ▶ Ce Modèle n'est qu'une approximation des fonctions remplies par le neurone réel
- ▶ Le neurone formel est un automate reproduisant les deux fonctions suivantes :
 1. sommation pondérée des entrées. les ω_i sont appelés les **poids synaptiques** du neurone et la valeur obtenue est appelée **activité**.
 2. émission d'un signal dans l'axone si son activité est supérieure à une valeur seuil (**fonction de transfert**)
- ▶ fonction signe : $f(x) = \text{signe}(x) = +1$ si $x > 0$, -1 si $x < 0$
- ▶ fonction tangente hyperbolique :

$$f(x) = \frac{e^{2x} - 1}{e^{2x} + 1} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \tanh(x)$$

- ▶ fonction logistique

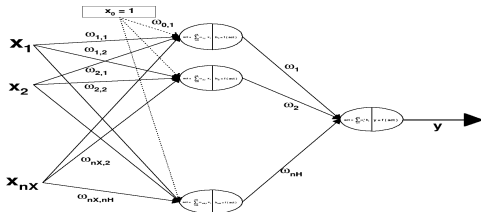
$$f(x) = \frac{1}{1 + e^{-x}}$$

- ▶ La **convention du perceptron** permet de simplifier l'écriture du signal en introduisant une entrée fictive $x_0 = +1$ de poids synaptique ω_0 appelé **biais**



Réseaux de Neurones Artificiel ("ANN")

- ▶ Premier ANN proposé par Rosenblatt (1962) afin d'étudier le mécanisme de la perception visuelle
- ▶ L'architecture d'un ANN concerne l'organisation des neurones. Elle caractérise la topologie des connexions entre les neurones. Il existe différentes topologies :
 - ▶ complète : tous les neurones sont connectés entre eux.
 - ▶ par couche : les neurones d'une couche sont connectés aux neurones de la couche précédente et de la couche suivante, ou bien, de façon plus générale représentée par un graphe acyclique (**feedforward**).
- ▶ Un ANN qui comporte au moins une couche cachée est appelé **perceptron multicouches** ("MLP")
- ▶ Les ANN sont des outils statistiques très puissants, permettant entre autres l'approximation de fonctions, la reconnaissance de formes
- ▶ Ils ont la qualité d'être des approximateurs universels du fait qu'il forme un espace dense sur tout compact Θ de \mathcal{C}^n , ensemble des fonctions continues de \mathbb{R}^n dans \mathbb{R} .



Algorithme de Rétropropagation du Gradient

- Initialiser les poids aléatoirement
- Pour chaque exemple (X, Y) de la base d'apprentissage
 - Présenter X aux entrées
 - Propager, en sens direct, les activations
 - Présenter Y aux sorties
 - Calculer les erreurs en sortie

$$E(x) = \frac{1}{2} \sum_{k=1}^{nY} (y_k - y_k(x; \omega))^2$$

- Calculer, en sens rétrograde, les gradients
- Mettre à jour les poids des connexions

modebatch Les gradients sont cumulés sur la totalité de la base d'apprentissage avant la mise à jour des paramètres

$$\omega(t+1) = \omega(t) - \eta \frac{\partial E_{moy}}{\partial \omega}$$

mode stochastique Les paramètres sont mis à jour après chaque présentation d'un exemple de la base d'apprentissage

$$\omega(t+1) = \omega(t) - \eta \frac{\partial E}{\partial \omega}$$

- Répéter la présentation de la base d'apprentissage plusieurs fois (cycle)

Calcul des gradients

Mise à jour des poids $\omega_{n,m}$ par une descente de gradient :

$$\frac{\partial e}{\partial \omega_{n,m}} = \frac{\partial e}{\partial a_n} \frac{\partial a_n}{\partial \omega_{n,m}} := \delta_n s_m$$

Le calcul explicite de δ_n dépend de la position du neurone :

- ① Le neurone n est situé sur la couche de sortie

$$\delta_n = \frac{\partial e}{\partial a_n} = \frac{\partial e}{\partial s_n} \frac{\partial s_n}{\partial a_n} = (s_n - y_n) f'(a_n)$$

- ② Le neurone n est un neurone de la couche cachée ; l'astuce est de décomposer $\frac{\partial e}{\partial s_n}$ en une somme de dérivation par rapport à l'activité des neurones fils du neurone n .

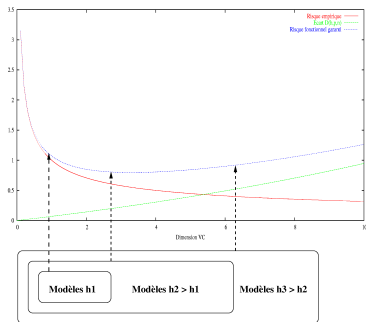
$$\frac{\partial e}{\partial s_n} = \sum_{k \in R(n)} \frac{\partial e}{\partial a_k} \frac{\partial a_k}{\partial s_n}$$

Nous reconnaissons d'une part, $\frac{\partial e}{\partial a_k} = \delta_k$ et d'autre part, $\frac{\partial a_k}{\partial s_n} = \omega_{k,n}$

$$\delta_n = \frac{\partial e}{\partial a_n} = \frac{\partial s_n}{\partial a_n} \frac{\partial e}{\partial s_n} = f'(a_n) \sum_{k \in R(n)} \delta_k \omega_{k,n}$$

Cette relation exprime le fait que l'erreur se calcule par récurrence en redescendant le long des connexions. L'erreur est ainsi propagée de la sortie vers les couches internes du réseau.

Sélection de l'architecture



- ▶ La dimension VC ("Vapnik-Chervonenkis") est un paramètre *difficile* à calculer
 - ▶ Elle est reliée au nombre de neurones cachés
- Estimation de l'erreur fonctionnelle par méthodes statistiques tout en utilisant le cadre de l'apprentissage statistique : adapter la *complexité* du modèle aux données.

Polynômes de Chaos

- ▶ Polynômes de chaos **approches fonctionnelles** permettant de construire le méta-modèle sur une base de polynômes orthogonaux particulièrement adaptés aux analyses de sensibilité par décomposition de la variance
- ▶ **Chaos Polynomial** introduit par Wiener (1938), popularisé par Ghanem (1999) : toute variable aléatoire de moyenne et variance finies, s'exprime par un développement en série :

$$X(\xi) = \sum_{\alpha} x_{\alpha} \Psi_{\alpha}(\xi)$$

- ▶ Variables aléatoire gaussiennes indépendantes $\xi_{i=1,2,\dots}$ centrées réduites et x_{α} coefficients du développement
- ▶ Polynômes $\Psi_{\alpha}(\xi)$ obtenus par tensorisation de polynômes de Hermite $H_{\alpha_i}(\xi_i)$ de degré α_i :

$$\Psi_{\alpha}(\xi) = \prod_i H_{\alpha_i}(\xi_i)$$

Polynômes de chaos généralisés

- ▶ Extension aux **Polynômes de Chaos Généralisés** : Xiu, Karniadakis (2000)
- ▶ Tensorisation de polynômes orthogonaux différents : Hermite (densité gaussienne), Legendre (densité uniforme), Laguerre (densité exponentielle), ...
- ▶ Ensemble de variables aléatoires $\{\xi_1, \xi_2, \dots\}$ indépendantes et polynômes orthogonaux $\varphi_{\alpha_i}^i$ associés aux densités des ξ_i de degré α_i
- ▶ Développement en polynômes de chaos généralisés :

$$X(\xi) = \sum_{\alpha} x_{\alpha} \Psi_{\alpha}(\xi) \quad \text{où} \quad \Psi_{\alpha}(\xi) = \prod_i \varphi_{\alpha_i}^i(\xi_i)$$

- ▶ Exemple d'une tensorisation Hermite \times Legendre

$$\begin{aligned} X(\xi_1, \xi_2) &= x_{00} + x_{10}H_1(\xi_1) + x_{01}L_1(\xi_2) \\ &+ x_{20}H_2(\xi_1) + x_{02}L_2(\xi_2) + x_{11}H_1(\xi_1)L_1(\xi_2) \\ &+ \dots \end{aligned}$$

Formalisation

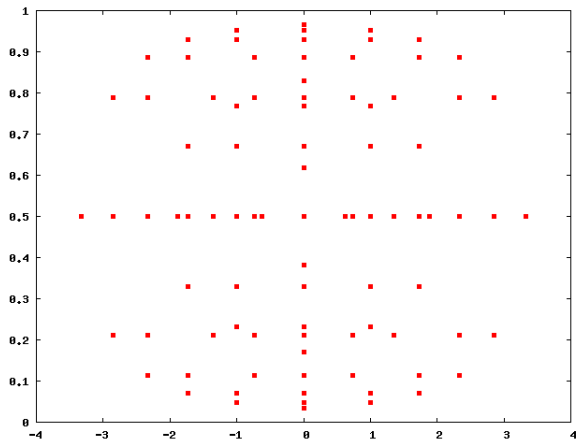
- ▶ Ensemble des variables aléatoires $X(\xi)$ de moyennes et variance finies est un espace de Hilbert munit du produit scalaire : $\langle X, Y \rangle = \mathbb{E}(XY)$
- ▶ Les polynômes Ψ_α sont orthogonaux
- ▶ Normalisés, ils constituent une base hilbertienne permettant de modéliser une variable aléatoire en chaos polynomial
- ▶ Coefficients du développement obtenus par projection

$$x_\alpha = \langle X, \Psi_\alpha \rangle$$

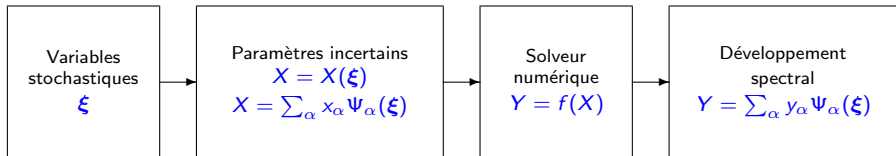
- ▶ Les grandeurs d'intérêt telles que moyenne, variance, indices de sensibilité de Sobol sont définies explicitement à partir des coefficients du chaos polynomial
- ▶ ANOVA fonctionnelle est fournie *naturellement* par le chaos polynomial
- ▶ Les coefficients peuvent être calculés par moindres carrés, ou par intégration numérique
 - ▶ Monte Carlo, Quasi Monte Carlo
 - ▶ Quadratures de Gauss, formules 1D tensorisées, construction de Smolyak (Sparse Grid)

Construction de Smolyak

Formule de Gauss-Hermite \times Gauss-Legendre



Non Intrusive Spectral Projection



1. Choix d'un groupe de variables stochastiques $\xi_i, i = 1, \dots, nx$
2. Plan d'expériences stochastiques en spécifiant niveau/taille
 - ▶ MC, LHS, QMC, Quadrature/Smolyak $\xi^k, k = 1, \dots, np$
3. Génération du plan d'expériences physiques $\xi^k \rightarrow \mathbf{x}^k = X(\xi^k)$
4. Réalisation du plan $\mathbf{x}^k \rightarrow \mathbf{y}^k = f(\mathbf{x}^k)$
5. Choix d'un degré (compatible niveau/taille) du chaos polynomial et calcul des coefficients par intégration ou approximation

$$Y(\xi) = \sum_{\alpha} y_{\alpha} \Psi_{\alpha}(\xi)$$

Introduction au Krigeage

- ▶ Krigeage : modèle prédictif développé initialement en géostatistique par D. G. Krige (thèse, 1951) puis formalisé par G. Matheron (1970)
- ▶ Interprétation probabiliste : les données observées vues comme réalisations d'un processus stochastique
- ▶ On postule des dépendances spatiales \Rightarrow structure de covariance
- ▶ Une mise en oeuvre *pratique* lorsque le processus est supposé gaussien

Les utilisations du krigeage

- ▶ méthode utilisée en **Machine Learning**, régression, classification supervisée (reconnaitances de formes), (C. E. Rasmussen),
- ▶ en simulation numérique, **Design and Analysis of Computer Experiments** (T. J. Santer),
- ▶ en **optimisation** pour la recherche d'extremum, Algorithme EGO, Efficient Global Optimization (D. R. Jones)
- ▶ en **validation** en complétant la modélisation physique à partir de données expérimentales (thèse CEA F. Bachoc, Dir. J. Garnier),
- ▶ **GdR Mascot-Num** cadre alimentant et diffusant la R&D,
- ▶ Outils développés dans le consortium **ReDice** (package R)
- ▶ Librairie **gpLib** (CEA) → Scilab, Uranie

Processus gaussiens

- ▶ Processus gaussien $Y(\mathbf{x})$ indexé sur R^d est complètement spécifié par sa moyenne $\mu(\mathbf{x})$ et sa fonction de covariance $k(\mathbf{x}_1, \mathbf{x}_2)$

$$Y(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \bullet))$$

- ▶ Processus stationnaire lorsque $k(\mathbf{x}_1, \mathbf{x}_2) = k(\mathbf{x}_1 - \mathbf{x}_2)$
- ▶ Exemples : corrélations de Gauss et Matérn

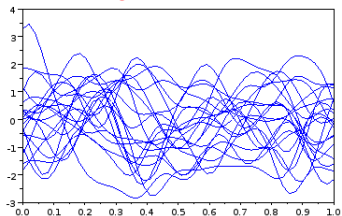
$$k(\mathbf{x}_1, \mathbf{x}_2) = \sigma^2 \exp - \sum_{i=1}^d \left(\frac{\mathbf{x}_{1i} - \mathbf{x}_{2i}}{l_{c_i}} \right)^2$$

$$k(\mathbf{x}_1, \mathbf{x}_2) = \sigma^2 \prod_{i=1}^d \frac{2}{\Gamma(\nu_i)} \left(\sqrt{\nu_i} \frac{\mathbf{x}_{1i} - \mathbf{x}_{2i}}{l_{c_i}} \right)^{\nu_i} K_{\nu_i} \left(2 \sqrt{\nu_i} \frac{\mathbf{x}_{1i} - \mathbf{x}_{2i}}{l_{c_i}} \right)$$

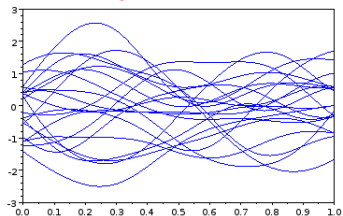
- ▶ Paramètres : la variance du processus σ^2 , les longueurs de corrélation l_{c_i} et les régularités ν_i pour la corrélation de Matérn

Echantillons de 4 processus gaussiens indexés sur R

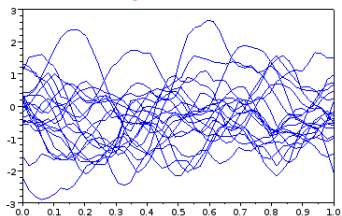
Gauss - longueur de corrélation = 0.1



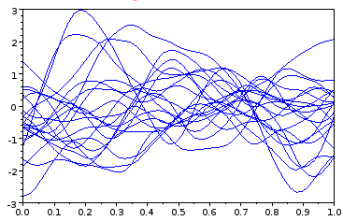
Gauss - longueur de corrélation = 0.3



Matérn 3/2 - longueur de corrélation = 0.2



Matérn 7/2 - longueur de corrélation = 0.2



Modèle de prédiction

- ▶ Vecteur gaussien pour le modèle des observations

$$Y_n = (Y(\mathbf{x}_1), Y(\mathbf{x}_2), \dots, Y(\mathbf{x}_n))^T \sim \mathcal{N}(\boldsymbol{\mu}, K)$$
$$\mu_i = \mu(\mathbf{x}_i) \quad , \quad K(i, j) = k(\mathbf{x}_i - \mathbf{x}_j)$$

- ▶ Modèle intégrant la valeur à prédire en un nouveau point \mathbf{x}_{new}

$$\mathbf{k} = (k(\mathbf{x}_1, \mathbf{x}_{new}), k(\mathbf{x}_2, \mathbf{x}_{new}), \dots, k(\mathbf{x}_n, \mathbf{x}_{new}))^T$$
$$\begin{pmatrix} Y_n \\ Y(\mathbf{x}_{new}) \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \boldsymbol{\mu} \\ \mu(\mathbf{x}_{new}) \end{pmatrix}, \begin{pmatrix} K & \mathbf{k} \\ \mathbf{k}^T & k(0) \end{pmatrix} \right)$$

- ▶ Conditionnement gaussien \Rightarrow loi de $Y(\mathbf{x}_{new})$ conditionnellement à une réalisation du vecteur gaussien Y_n

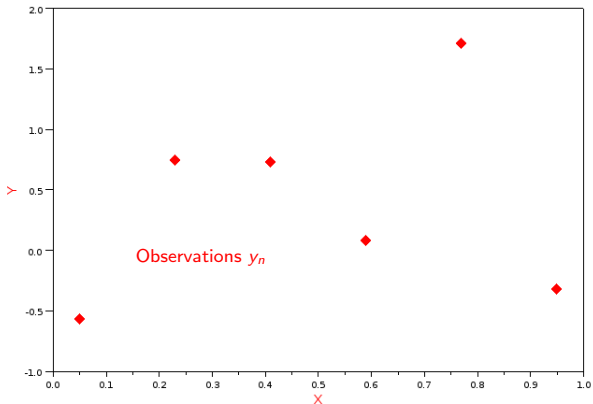
$$\mathbb{E}[Y(\mathbf{x}_{new}) | Y_n] = \mu(\mathbf{x}_{new}) + \mathbf{k}^T K^{-1} (Y_n - \boldsymbol{\mu})$$
$$\text{Var}[Y(\mathbf{x}_{new}) | Y_n] = k(0) - \mathbf{k}^T K^{-1} \mathbf{k}$$

- ▶ BLUP, Best Linear Unbiased Predictor $\hat{Y}(\mathbf{x}_{new}) = \mathbb{E}[Y(\mathbf{x}_{new}) | Y_n]$

Processus gaussien conditionnellement aux observations

Processus $Y(x) \sim (\mu = 0, \sigma^2 = 1 \times \text{Matérn}[lc = 0.3, \nu = 5/2])$

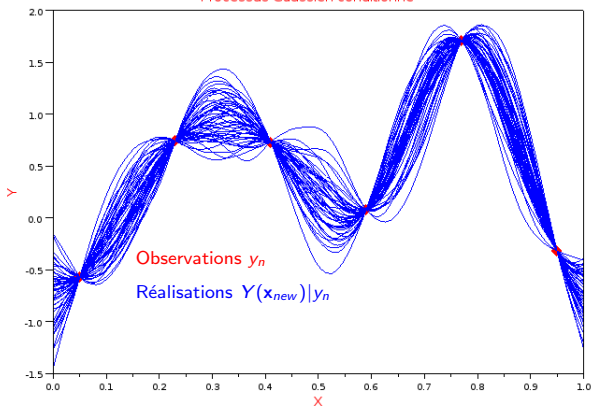
Processus Gaussien conditionné



Processus gaussien conditionnellement aux observations

Processus $Y(x) \sim (\mu = 0, \sigma^2 = 1 \times \text{Matérn}[lc = 0.3, \nu = 5/2])$

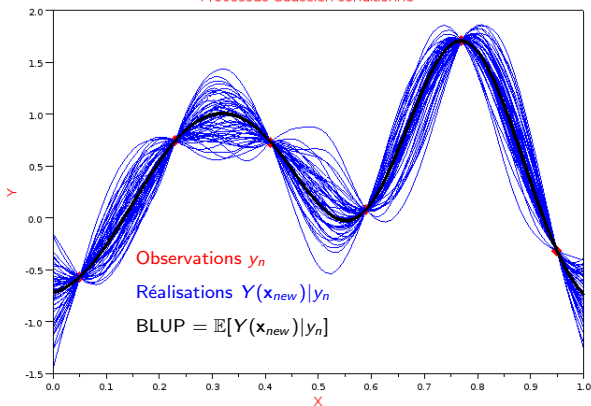
Processus Gaussien conditionné



Processus gaussiens conditionnellement aux observations

Processus $Y(x) \sim (\mu = 0, \sigma^2 = 1 \times \text{Matérn}[lc = 0.3, \nu = 5/2])$

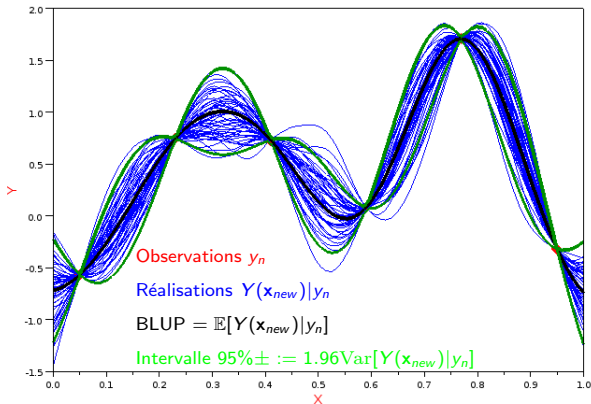
Processus Gaussien conditionné



Processus gaussien conditionnellement aux observations

Processus $Y(x) \sim (\mu = 0, \sigma^2 = 1 \times \text{Matérn}[lc = 0.3, \nu = 5/2])$

Processus Gaussien conditionné



Construction d'un krigeage

- ▶ Structure général du modèle du processus (sans erreurs de mesure)

$$Y_{obs}(\mathbf{x}) = \underbrace{\mathbf{h}^T(\mathbf{x})\boldsymbol{\beta}}_{\text{moyenne}:=\mu(\mathbf{x})} + \underbrace{Z(\mathbf{x}; \sigma^2, \theta)}_{\text{processus centré}}$$






- ▶ Estimation des paramètres $\boldsymbol{\beta}; \sigma^2, \theta$ par différentes méthodes : maximum de vraisemblance, maximum de la vraisemblance restreinte, par leave one out
- ▶ Dans le cas où il n'y a pas d'erreur de mesures, la matrice de covariance $K_{\theta, \sigma^2} = \sigma^2 K_{\theta}$
- ▶ Estimation des paramètres θ par maximum de vraisemblance à partir des données observées \mathbf{y}_{obs} . **Problème à résoudre :**

$$\hat{\theta} = \arg \min_{\theta} \ln |K_{\theta}| + n \ln [\mathbf{y}_{obs}^T K_{\theta}^{-1} (I - H(H^T K_{\theta}^{-1} H)^{-1} H^T K_{\theta}^{-1}) \mathbf{y}_{obs}]$$

- ▶ Puis estimation de la variance et des paramètres de la moyenne :

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n} \mathbf{y}_{obs}^T K_{\hat{\theta}}^{-1} (I - H(H^T K_{\hat{\theta}}^{-1} H)^{-1} H^T K_{\hat{\theta}}^{-1}) \mathbf{y}_{obs} \\ \hat{\boldsymbol{\beta}} &= (H^T K_{\hat{\theta}}^{-1} H)^{-1} H^T K_{\hat{\theta}}^{-1} \mathbf{y}_{obs}\end{aligned}$$

Bibliographie

-  Saltelli, A., Chan, K., and Scoot, M., editors (2000). *Sensitivity Analysis*, John Wiley and Sons.
-  Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. (2005). *Sensitivity Analysis in Practise*, John Wiley and Sons.
-  Fang, K.T., Li, R., and Sudjianto, A. (2006). *Design and Modeling for Computers Experiments*, Chapman.
-  Saltelli, and all (2008). *Global Sensitivity Analysis : the Primer*, John Wiley and Sons.
-  McKay, M. D. (1995). Evaluating Prediction Uncertainty, NUREG/CR-6311.
-  Franco, J. (Sept. 2008). Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes, Mines de Saint-Etienne.
-  Bishop, C. M. (1995). *Neural Networks for Patterns Recognition*, Oxford University Press.
-  Rasmussen, C. E. and Willians, C. (2006). *Gaussian Processes for Machine Learning*, MIT Press.
-  Helton, J. and Davis, F. (2000). *Sampling-Based Methods for Uncertainty and Sensitivity Analysis*, Sandia Report SAND99-2240.