



HAL
open science

Plans d'expériences pour simulations numériques

Charlie Sire

► **To cite this version:**

Charlie Sire. Plans d'expériences pour simulations numériques. École d'ingénieur. France. 2023.
hal-04527340

HAL Id: hal-04527340

<https://hal.science/hal-04527340v1>

Submitted on 2 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

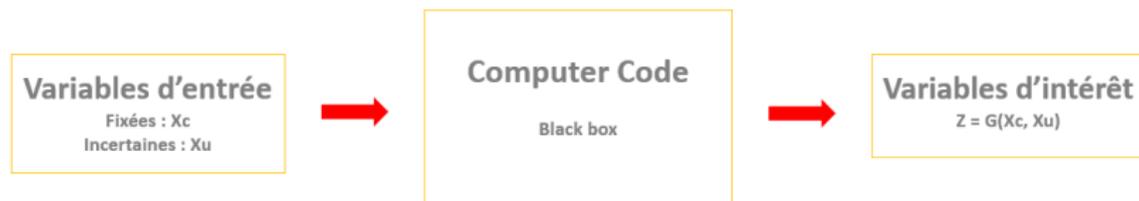
Plans d'expériences

Partie 2 : Plans d'expériences pour simulations numériques

Charlie SIRE

7 Décembre 2023

- 1 Simulation numériques
- 2 Introduction aux plans d'expériences pour simulations numériques
- 3 Critères de remplissage d'espace
- 4 Suites à faible discrédance
- 5 Hypercubes latins (LHS)
- 6 Conclusion

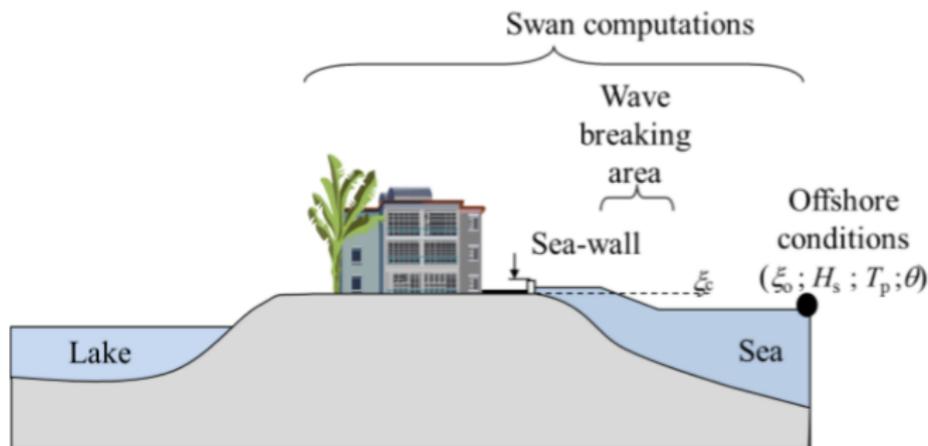


Variables d'entrée et d'intérêt:

- Scalaires
- Vectorielles
- Fonctionnelles (spatiales, temporelles, spatio-temporelles)

Exemple

Etude d'une inondation de zone côtière



- Variables fixées x_c : Dimensionnement des protections
- Variables incertaines X_u : Conditions maritimes au large
- Variable d'intérêt $G(x_c, X_u)$: Niveau d'eau à un endroit donné
- Code de calcul : Modélisateur hydraulique

Objectifs possibles

- Analyse de sensibilité : Trouver les variables les plus influentes, quantifier leur impact..
- Optimisation : Trouver le minimum ou le maximum de G (possiblement à X_c fixé)
- Inversion : Déterminer $\Gamma_{X_u} = \{x_c, G(x_c, X_u) > \text{seuil}\}$ ou $\Gamma = \{x_c, P(G(x_c, X_u) > \text{seuil}) > \alpha\}$

Problème

Ces objectifs requièrent des milliers d'appels au simulateurs

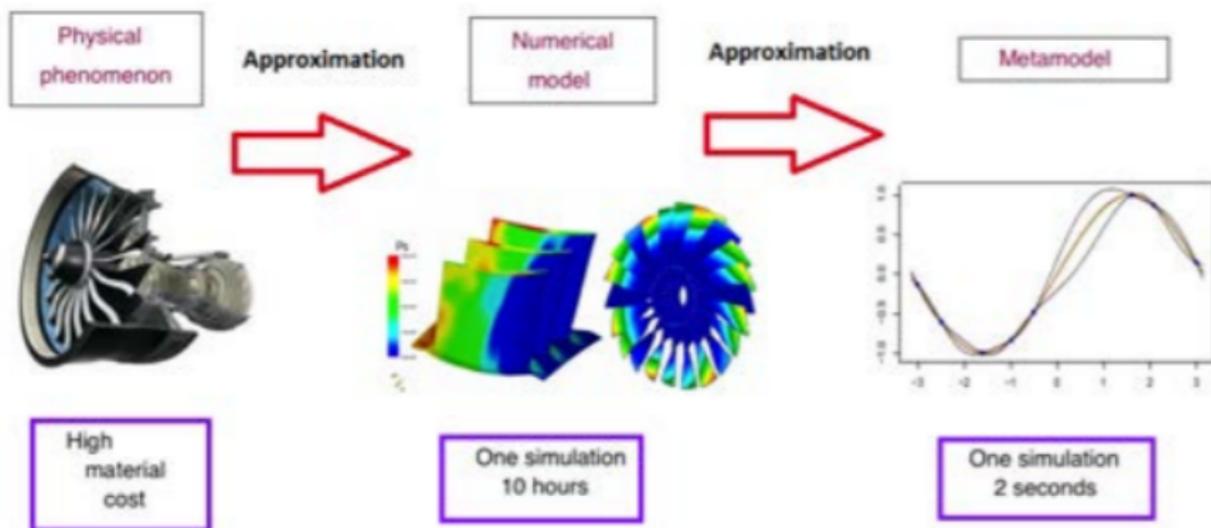
Or, temps de calcul du modélisateur très élevé : plusieurs jours

Approche statistique :

- Plans d'expériences
- Métamodèle

⇒ **Où faire les quelques simulations envisageables et quel métamodèle choisir pour atteindre un des objectifs précédents ?**

Intérêt du métamodèle



Content

- 1 Simulation numériques
- 2 Introduction aux plans d'expériences pour simulations numériques**
- 3 Critères de remplissage d'espace
- 4 Suites à faible discrédance
- 5 Hypercubes latins (LHS)
- 6 Conclusion

Objectifs

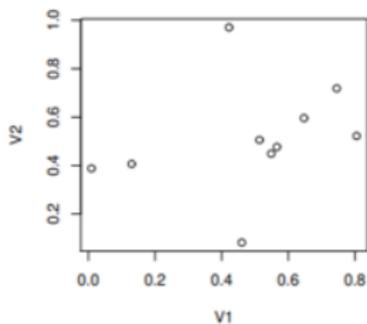
- Explorer le comportement des réponses d'un code à l'aide d'un nombre limité de calculs
- Propager les incertitudes (calcul des moments des réponses)
- Fournir des plans plus efficaces que les plans aléatoires purs pour estimer des indices de sensibilité
- Fournir un plan initial pour construire un métamodèle

Spécificité des plans pour simulations numériques

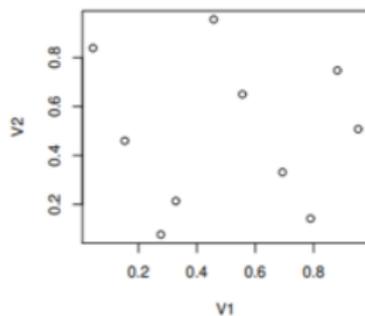
- Expériences déterministes (erreur=0)
- Grand nombre de facteurs
- Larges domaines de variation
- Variables d'intérêt multiples
- Modèles fortement non linéaires

⇒ **space filling designs**

Monte
Carlo



Space
Filling



Content

- 1 Simulation numériques
- 2 Introduction aux plans d'expériences pour simulations numériques
- 3 Critères de remplissage d'espace**
- 4 Suites à faible discrédance
- 5 Hypercubes latins (LHS)
- 6 Conclusion

Critère géométrique Maximin

Le design Maximin maximise la distance minimale entre les points du plan

On pose $\mathbb{X} = (x_1, x_2, \dots, x_n)$ avec $\forall 1 \leq i \leq n, x_i \in \mathbb{D}$ (ex $\mathbb{D} = [0, 1]^p$)

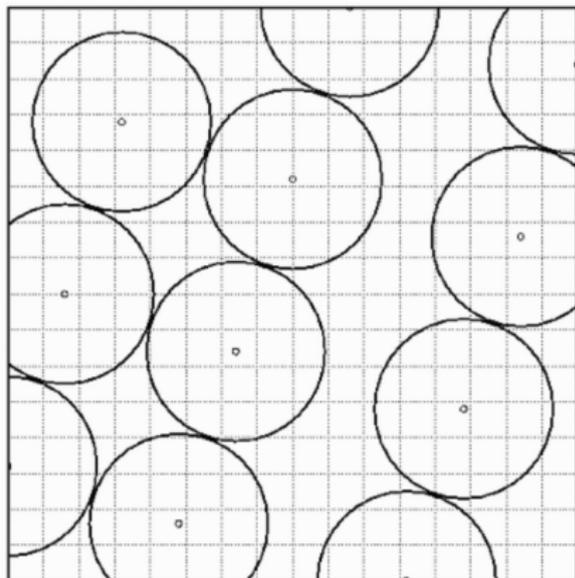
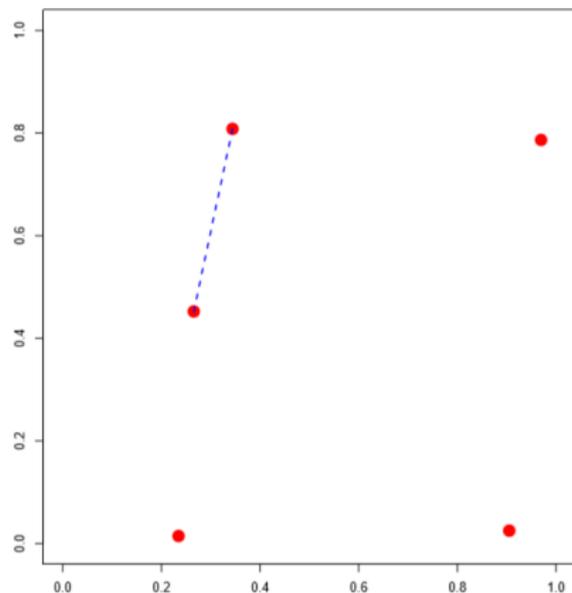
On définit :

$$\Phi_{Mm}(\mathbb{X}) = \min_{(x_i, x_k) \in \mathbb{X}^2, i \neq k} \|x_i - x_k\|$$

Φ_{Mm} est une mesure du remplissage de l'espace \mathbb{D} par le plan \mathbb{X}

$\mathbb{X}_{Mm}^* = \operatorname{argmax}_{\mathbb{X}} \Phi_{Mm}(\mathbb{X})$ est appelé plan d'expériences maximin

Visualisation du critère Maximin



$\Phi_{Mm}(\mathbb{X}) = 2\rho$ où ρ est le plus petit rayon r tel que l'ensemble des cercles de rayon r ayant pour centre les points du plan ne sont pas tous disjoints 2 à 2.

Critère géométrique : miniMax

Le design miniMax minimise la distance maximale entre un point du domaine et le plan

On définit la distance entre un point $x \in \mathbb{D}$ et un design \mathbb{X} :

$$d(x, \mathbb{X}) = \min_{x_i \in \mathbb{X}} \|x_i - x\|$$

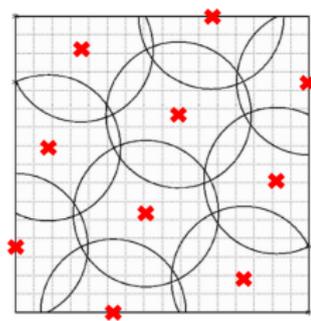
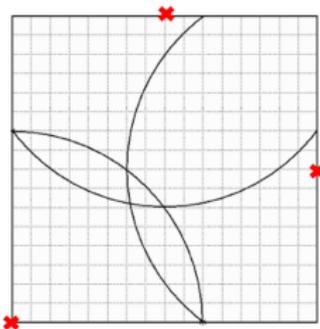
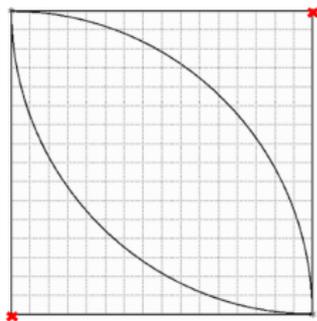
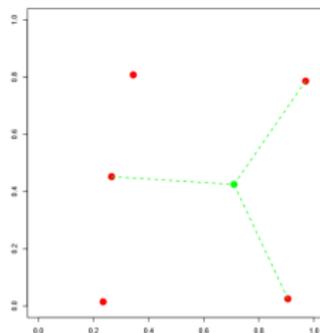
On définit maintenant :

$$\Phi_{mM}(\mathbb{X}) = \max_{x \in \mathbb{D}} d(x, \mathbb{X})$$

Φ_{mM} est également une mesure du remplissage de l'espace \mathbb{D} par le plan \mathbb{X}

$\mathbb{X}_{mM}^* = \operatorname{argmin}_{\mathbb{X}} \Phi_{mM}(\mathbb{X})$ est appelé plan d'expériences minimax

Visualisation du critère miniMax



$\Phi_{mM}(\mathbb{X}) = \rho$ où ρ est le plus petit rayon r tel que la réunion des disques de rayon r ayant pour centre les points du plan couvre l'ensemble du plan.

Exercice

- 1 En 1D, avec $\mathbb{D} = [0, 1]$ donner le plan Maximin à 2 points, 3 points.. Puis en déduire celui à n points. Calculer à chaque fois $\Phi_{Mm}(\mathbb{X})$ et $\Phi_{mM}(\mathbb{X})$.
- 2 Faire la même chose avec les plans minimax.

Correction exercice

- 1 $\mathbb{X} = \{x_i = \frac{i-1}{n-1}, 1 \leq i \leq n\}$; $\Phi_{Mm}(\mathbb{X}) = \frac{1}{n-1}$; ; $\Phi_{mM}(\mathbb{X}) = \frac{1}{2(n-1)}$
- 2 $\mathbb{X} = \{x_i = \frac{2i-1}{2n}, 1 \leq i \leq n\}$; $\Phi_{Mm}(\mathbb{X}) = \frac{1}{n}$; ; $\Phi_{mM}(\mathbb{X}) = \frac{1}{2n}$

Critère statistique : la discrédance

La discrédance mesure la déviation maximale entre la répartition des points du plan et une répartition uniforme

On pose $\mathbb{D} = [0, 1]^p$, $Q(t) = [0, t_1] \times [0, t_2] \times \dots \times [0, t_p]$ et $\mathbb{X} = (x_1, x_2, \dots, x_n)$

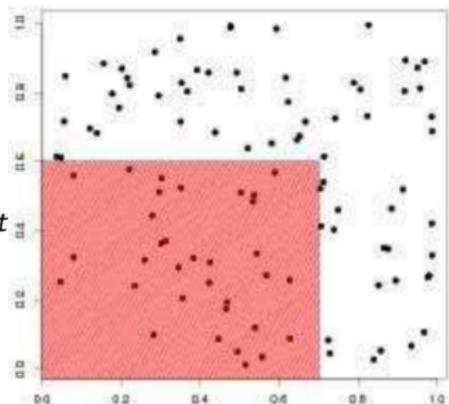
On définit la discrédance comme suit :

$$D_n(\mathbb{X}) = \sup_{t \in [0, 1]^p} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_n \in Q(t)} - \text{vol}(Q(t)) \right|$$

avec $\text{vol}(Q(t)) = \prod_{i=1}^p t_i$ et $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_n \in Q(t)}$ la proportion de points de points du plan dans $Q(t)$

D'autres formulations existent :

$$D_{nL_2}(\mathbb{X}) = \left[\int_{[0, 1]^p} \left[\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_n \in Q(t)} - \text{vol}(Q(t)) \right]^2 dt \right]^{1/2}$$



Suite uniformément distribuées

Définition : On appelle suite uniformément distribuée une suite u qui vérifie :

$\forall \mathcal{J}$ pavé de $[0, 1]^p$ (i.e. de la forme $[a_1, b_1] \times \dots \times [a_p, b_p]$),

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{u_n \in \mathcal{J}} = \text{vol}(\mathcal{J})$$

où $\mathbb{1}_{u_n \in \mathcal{J}}$ est la fonction indicatrice de \mathcal{J} appliquée en u_n

Dit autrement, $\forall \mathcal{J}$, la proportion de points dans \mathcal{J} tend vers le volume de \mathcal{J}

Propriété : Une suite (u_n) est uniformément distribuée si et seulement si $\lim_{n \rightarrow \infty} D_n(\mathbb{U}_n = (u_1, \dots, u_n)) = 0$

Lien avec le problème de l'intégration d'une fonction

$$I = \int_{[0,1]^p} F(x) dx$$

Propriété générale (Koksma Hlawka) :

$$\epsilon = \left| I - \frac{1}{n} \sum_{i=1}^n F(x_i) \right| \leq V(F) \times D_n(\mathbb{X})$$

où $V(F)$ est la variation de F et ne reflète que l'irrégularité de la fonction F

En dimension 1, pour une fonction F dérivable de dérivée intégrable sur $[0, 1]$:

$$V(F) = \int_{[0,1]} |F'(x)| dx$$

Monte Carlo basique : $\epsilon = O\left(\frac{1}{\sqrt{n}}\right)$ (en moyenne)

Suite à faible discrédance : $\epsilon = O\left(\frac{\ln(n)^p}{n}\right)$

Content

- 1 Simulation numériques
- 2 Introduction aux plans d'expériences pour simulations numériques
- 3 Critères de remplissage d'espace
- 4 Suites à faible discrédance**
- 5 Hypercubes latins (LHS)
- 6 Conclusion

Problématique en 1D

On va vouloir construire une suite (u_n) de discrédance minimale.

Il peut être démontré que pour tout ensemble de points $\{u_1, \dots, u_n\}$,

$$D_n(u_1, \dots, u_n) \geq \frac{1}{2n}$$

avec égalité quand $u_i = \frac{2i-1}{2n}$

Problème : La position des n points optimaux dépend de n , ce qui complique les approches séquentielles

Il a même été démontré que :

$$\exists c \in \mathbb{R}, \forall u \in [0, 1]^{\mathbb{N}}, \forall n \in \mathbb{N}, D_n(u_1, \dots, u_n) \geq c \frac{\log(n)}{n}$$

\Rightarrow Construction de suites dont la discrédance est de l'ordre de la borne ci-dessus

Suites de Van der Corput : base b

Les suites de Van der Corput sont basés sont l'écriture en base b d'un entier n.

En effet, chaque entier n a une unique écriture en base b :

$$n = \sum_{j=0}^{\infty} a_j(n, b)b^j$$

où $\forall j, a_j(n, b) \in \{0, 1, \dots, b-1\}$ et $a_j(n, b) = 0$ pour j suffisamment grand

Exemple :

$$\begin{aligned}(10433.2)_5 &= 1 \times 5^4 + 0 \times 5^3 + 4 \times 5^2 + 3 \times 5^1 + 3 \times 5^0 + 2 \times 5^{-1} \\ &= (765.4)_{10}\end{aligned}$$

Suites de Van der Corput : exercice base b

Compléter les valeurs de n en base b pour les b et n ci-dessous :

n	$b = 2$	$b = 3$	$b = 4$
0			
$(1)_{10}$			
$(2)_{10}$			
$(3)_{10}$			
$(4)_{10}$			

n	$b = 2$	$b = 3$	$b = 4$
$(5)_{10}$			
$(6)_{10}$			
$(7)_{10}$			
$(8)_{10}$			
$(9)_{10}$			

Suites de Van der Corput : correction exercice base b

n	$b = 2$	$b = 3$	$b = 4$	n	$b = 2$	$b = 3$	$b = 4$
0	$(0)_2$	$(0)_3$	$(0)_4$	(5) ₁₀	$(101)_2$	$(12)_3$	$(11)_4$
$(1)_{10}$	$(1)_2$	$(1)_3$	$(1)_4$	$(6)_{10}$	$(110)_2$	$(20)_3$	$(12)_4$
$(2)_{10}$	$(10)_2$	$(2)_3$	$(2)_4$	$(7)_{10}$	$(111)_2$	$(21)_3$	$(13)_4$
$(3)_{10}$	$(11)_2$	$(10)_3$	$(3)_4$	$(8)_{10}$	$(1000)_2$	$(22)_3$	$(20)_4$
$(4)_{10}$	$(100)_2$	$(11)_3$	$(10)_4$	$(9)_{10}$	$(1001)_2$	$(100)_3$	$(21)_4$

Suites de Van der Corput : Définition

Pour tout entier $b \geq 2$, on pose :

$$\Phi_b(n) = \sum_{j=0}^{\infty} a_j(n, b) b^{-j-1}$$

La suite de Van der Corput en base b est ainsi :

$$C_b = \{\Phi_b(n), n \in \mathbb{N}\}$$

n	$b = 2$	$\phi_b(n)$
0	$(0)_2$	$0 \times 2^{-1} = 0,0000$
1	$(1)_2$	$1 \times 2^{-1} = 0,5000$
2	$(10)_2$	$1 \times 2^{-2} + 0 \times 2^{-1} = 0,2500$
3	$(11)_2$	$1 \times 2^{-2} + 0 \times 2^{-1} = 0,7500$
4	$(100)_2$	$1 \times 2^{-3} + 0 \times 2^{-2} + 0 \times 2^{-1} = 0,1250$
5	$(101)_2$	$1 \times 2^{-3} + 0 \times 2^{-2} + 1 \times 2^{-1} = 0,6250$
6	$(110)_2$	$1 \times 2^{-3} + 1 \times 2^{-2} + 0 \times 2^{-1} = 0,3750$
7	$(111)_2$	$1 \times 2^{-3} + 1 \times 2^{-2} + 1 \times 2^{-1} = 0,8750$
8	$(1000)_2$	$1 \times 2^{-4} + 0 \times 2^{-3} + 0 \times 2^{-2} + 0 \times 2^{-1} = 0,0625$

Suites de Halton : Définition

Les suites de Halton sont une extension naturelle en dimension d des suites de Van der Corput

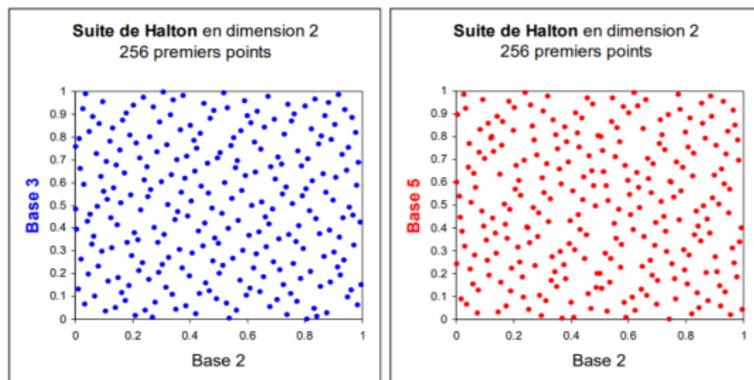
Soient b_1, \dots, b_p p entiers supérieurs ou égaux à 2. La suite de Halton en dimension d est :

$$H_{b_1, \dots, b_p} = \{u_n, n \in \mathbb{N}\}$$

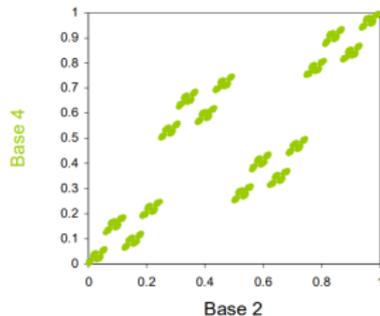
où $u_n = (\Phi_{b_1}(n), \dots, \Phi_{b_p}(n))$

Attention : Les b_i doivent être premiers entre eux

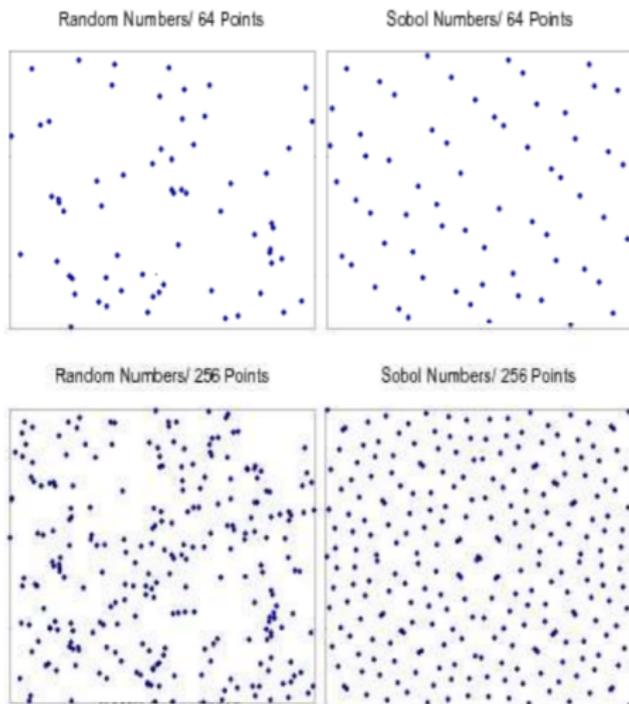
Suites de Halton : Visualisation



Suite de Halton en dimension 2
256 premiers points

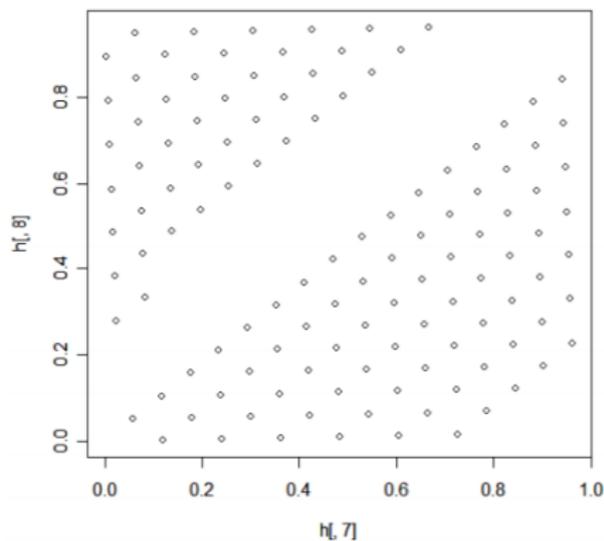
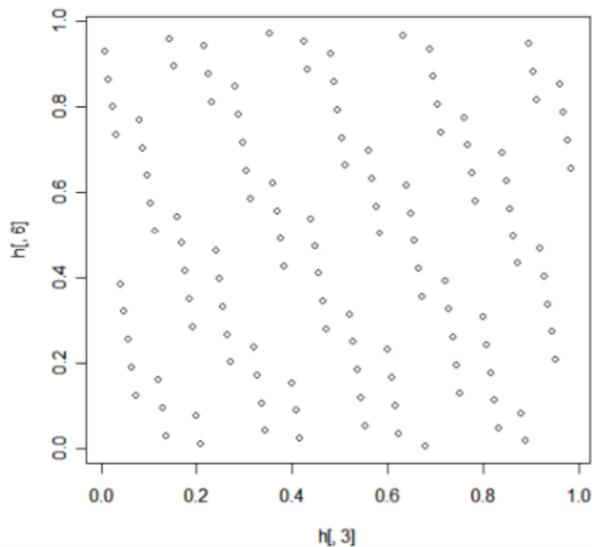


Autres suites à faibles discrédances : suites de Sobol



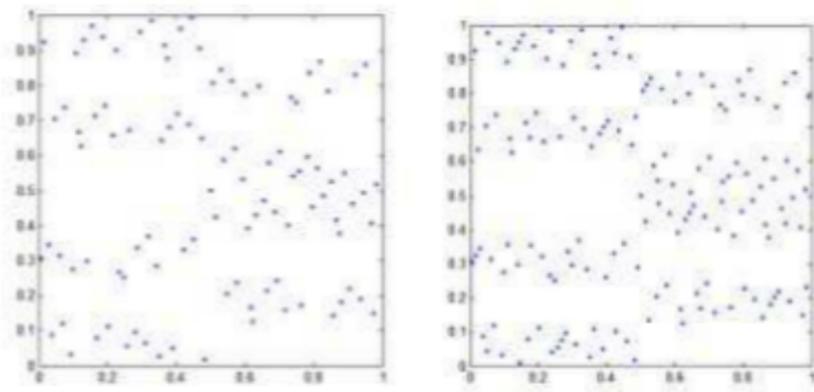
Pathologie sur les projections en dimensions inférieures - Halton

Exemple avec 150 points en 8 dimensions pour la suite de Halton



Pathologie sur les projections en dimensions inférieures - Sobol

Exemple avec 100 points (gauche) 150 points (droite) en 8 dimensions pour la suite de Sobol, avec projection sur les axes 7 et 8



Importance de la robustesse dans les dimensions inférieures : le code de calcul a souvent peu de dimensions effectives (nb de variables influentes $\ll p$)

Content

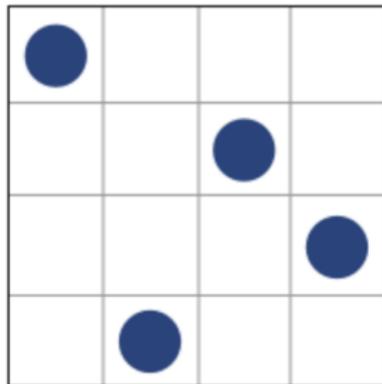
- 1 Simulation numériques
- 2 Introduction aux plans d'expériences pour simulations numériques
- 3 Critères de remplissage d'espace
- 4 Suites à faible discrédance
- 5 Hypercubes latins (LHS)**
- 6 Conclusion

LHS : Principe

On note $LHS(p, n)$ un LHS à p variables et n points.

L'idée est la suivante :

- On divise chaque dimension en n intervalles égaux : on obtient ainsi n^p cellules de même taille (produit cartésien des intervalles)
- On choisit n cellules parmi les n^p obtenues, en faisant en sorte que chaque intervalle de chaque dimension ne soit représentées qu'une fois
- On tire au hasard un point dans chacune des n cellules sélectionnées



LHS : formulation

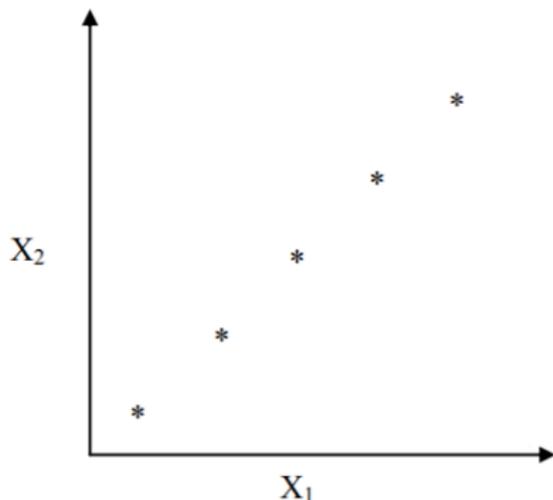
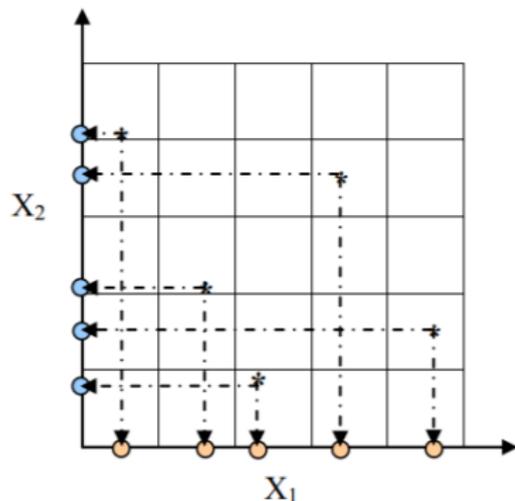
Un LHS n points sur $[0, 1]^d$ est défini par l'ensemble des points x^i tels que :

$$x_j^i = \frac{\pi_j(i) - u_j^i}{n}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq p$$

où π_j est une permutation de $\{1, \dots, n\}$ et u_j^i est une réalisation de la variable aléatoire U_j^i avec $U_j^i \sim \mathcal{U}(0, 1)$

Remarque : Les U_j^i peuvent suivre d'autres lois, et même par exemple placer les points au centre des cellules

LHS : exemples

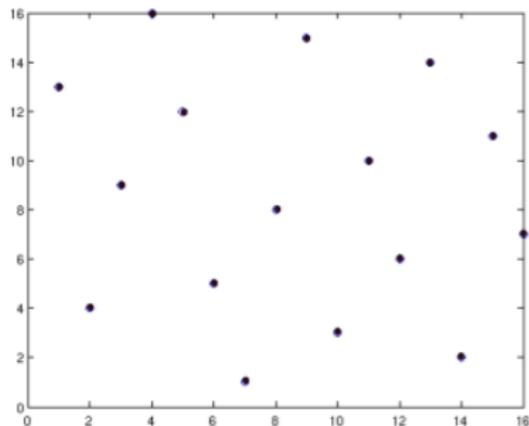


Sans prendre en compte l'aléa dû à la réalisation de la variable aléatoire uniforme, il y a $(n!)^{p-1}$ LHS possibles ! (et non $(n!)^p$ car on ne tient pas compte de l'ordre des points)

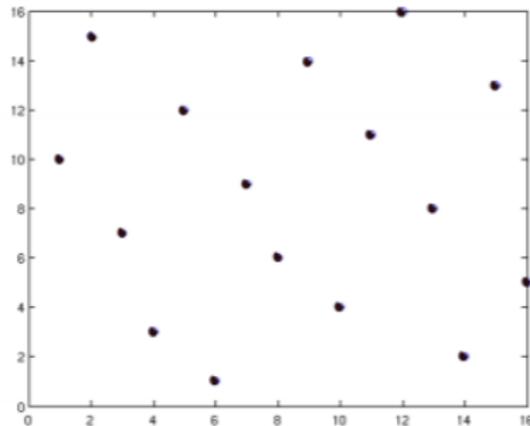
LHS : optimisation

Idée : Optimisation d'un des critère de remplissage d'espace

Comme le nombre de LHS est très grand, on peut faire appel à des algorithmes d'optimisation adaptés, pour ne pas parcourir l'ensemble des possibilités



Maximin LHS



LHS à faible discrédance

Content

- 1 Simulation numériques
- 2 Introduction aux plans d'expériences pour simulations numériques
- 3 Critères de remplissage d'espace
- 4 Suites à faible discrédance
- 5 Hypercubes latins (LHS)
- 6 Conclusion

Bilan sur les plans d'expériences numériques

Enjeu : Echantillonner un espace de grande dimension de manière "optimale" (obtenir le plus d'informations sur le comportement de la sortie)

Problème : Un échantillon aléatoire (Monte Carlo) remplit mal l'espace

- 1 **Plans "space filling"** sont de bons candidats pour bien remplir l'espace :
 - Basés sur un critère de remplissage d'espace (minimax, maximin)
 - Basés sur un critère de répartition uniforme des points (discrépance)
- 2 Ces plans peuvent avoir des problèmes de projections en dimensions inférieures. Les **hypercubes latins (LHS)** y remédient pour les projections en dimension 1.
- 3 Il est possible de coupler les 2 propriétés avec des **LHS optimisés**