

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**UNIVERSITÉ DE BLIDA 1**



**Faculté des Sciences  
Département de Mathématique**



# **Mémoire**

En vue de l'obtention du diplôme de

**MASTER EN MATHÉMATIQUES**

**Spécialité : Modélisation Stochastique et Statistique**

Thème

**PLANIFICATION DES EXPERIENCES NUMERIQUES A  
L'AIDE DE L'INFORMATION DE KULLBACK-  
LEIBLER MINIMALE**

Par

**M<sub>r</sub> GOUACEM TAREK**

Membres du jury :

Promoteur : M<sub>r</sub> ELMOSSAOUI HICHAM

Président : Mme RABER DALILA

Examineur : Mme MESSAOUDI NADIA

Année universitaire 2018/2019

ملخص : يقترح هذا العمل بناء تصميمات التجارب الرقمية وهو طريقة جديدة تقوم على مقارنة التوزيع التجريبي للنقاط في التصميم بتوزيع الموحد باستعمال معلومات Kullback-Leibler تعمل هذه المقاربة على تقدير هذا الاختلاف بالمرور على Shannon entropy, يتم تقدير الإنتروبي بواسطة طريقة Monte Carlo حيث يتم استبدال الكثافة بواسطة مقدر كثافة النواة.

**Résumé:** Ce travail propose une méthode de construction de plans d'expériences numériques, dont les points sont le plus uniformément répartis dans l'hypercube unité. Afin de mesurer l'écart entre la fonction de densité associée aux points du plan et celle de la loi uniforme, nous utilisons l'information de Kullback-Leibler, ce qui revient par ailleurs à utiliser l'entropie de Shannon. Celle-ci est estimée par une méthode de Monte Carlo dans laquelle la fonction de densité est remplacée par son estimé qui est aussi remplacé par noyaux gaussiens.

**Abstract:** this work suggests a method of constructing numerical experience designs, we suggest a new method based on comparing the empirical distribution of the points in a design to the uniform distribution with the Kullback-Leibler information. The considered approach consists in estimating this difference or, reciprocally, the Shannon entropy. The entropy is estimated by a Monte Carlo method where the density function is replaced by its gaussian's kernel density estimator.

## Remerciements

Avant toute chose, je tiens à remercier Dieu pour toutes ces choses dont on oublie la source, et quand elles nous quittent tout devient difficile et peut être même impossible à accomplir. Durant la période où je préparais mon mémoire, je n'ai manqué de rien « El Hamdoulilah », que ça soit matériellement, financièrement où moralement, c'est pour cette raison que je tiens à remercier tous ces acteurs qui m'ont porté vers ma réussite et qui m'ont accompagné tout le long de la route.

Je tiens à remercier mon encadreur Dr.Moussaoui Hicham pour ses efforts et ses innombrables conseils riches en expérience, ce qui m'a aidé à apprendre la rigueur dans le travail.

Mes sincères remerciements s'adressent à tous mes camarades, aux professeurs ainsi qu'aux autorités académiques de l'Université Saad Dahleb de Blida.

Le soutien indéfectible de mes amis et de ma famille m'a permis de ne pas m'écarter dans les moments difficiles de l'objectif c'est pour cela que je tiens à les remercier amplement et à leur exprimer ma gratitude.

Que tous ceux qui d'une manière ou d'une autre, ont rendu possible la réalisation de ce travail trouvent ici l'expression de ma profonde gratitude.

## Dédicace

Je dédie ce mémoire de fin d'étude :

A mes très chers parents,

A mon père, qui m'a aidé à construire l'homme et la personne que je suis aujourd'hui, sans lui toute ces années n'auraient jamais pu aboutir, grâce à son soutien moral et l'éducation qu'il m'a donné, il m'a appris de ne jamais abandonner et de me battre coute que coute.

A ma mère, l'incarnation de la douceur et de la bienveillance, celle qui m'a appris à aimer le travail, aimer les autres et aimer la vie, dans les moments difficiles tu as été la lumière qui m'a guidé et que j'ai suivie sans me poser de questions, je te dédis toute ces années de dur labeur et ce mémoire que j'ai façonné à ton image.

A mon frère et à ma sœur, pour leurs encouragements permanents, et leur soutien moral.

Mes amis, qui m'ont accompagné à chaque pas que je marchais, leurs réconfort inconditionnel et leur bonnes idées.

A tous mes camarades de l'université et du département de Mathématique ainsi qu'aux professeurs qui m'ont enseigné tout au long de mon parcours.

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infaillible,

Merci d'être toujours là pour moi.

## Table des matières

Introduction.....	1
Chapitre 1 : Les plans d'expériences numériques usuels.....	3
1.1 Plan d'expériences classiques utilisées à l'expérimentation numérique.....	3
1.1.1 Rappels sur le modèle polynomial.....	3
1.1.2 Quelques éléments sur les plans d'expériences usuels.....	4
1.1.3 Plan factoriel complet.....	6
1.1.4 Plans factoriels fractionnaires.....	7
1.1.5 Plans composites.....	7
1.1.6 Plans de Box-Behnken.....	8
1.1.7 Conclusion.....	9
1.2 Les plans uniformes ( <i>SpaceFillingDesign</i> « <i>SFD</i> » ).....	10
1.2.1 Les plans (S, F, D) basés sur des contraintes marginales(Hyper cube latin (LHS).....	11
1.2.2 Les plans (S, F, D) basés sur des lois de distribution des points (Suites à faible discrédance).....	13
1.2.2.1 Les suites de Van Der Corput.....	13
1.2.2.2 Suites de Halton.....	14
1.2.2.3 Les suites finies de Hammersley.....	15
1.2.2.4 Les suites de Faure.....	16
1.2.2.5 Les suites de Sobol.....	17
1.2.2.6 Les (t, s)-suites et (t, m, s)-réseaux.....	19
1.3 Discussion.....	21
Chapitre 2 : Critères et plans optimaux.....	23
2.1 Critères d'uniformité basés sur la discrédance et plans à faible discrédance.....	23
2.1.1 La discrédance locale.....	23
2.1.2 La discrédance infinie à l'origine.....	25
2.1.3 La discrédance extrême.....	26
2.1.4 La discrédance en norme $L_2$ -étoile.....	26
2.1.5 La Discrédance $L_2$ modifiée.....	28
2.1.6 La discrédance centrée $DL_2, Cx$ .....	28
2.1.7 La discrédance en norme $L_2$ -étoile pondérée $DL_2, *, \gamma$ .....	29
2.1.8 Plans à faible discrédance.....	31
2.2 Critères de distance et plans optimaux.....	33
2.2.1 Mesure de recouvrement.....	34
2.2.2 Le rapport de distances.....	35
2.2.3 Distances Maximin et Minimax.....	35
2.3 Critère d'entropie et plans à entropie maximale.....	37
2.3.1 Définition de l'entropie.....	37

2.3.2 Entropie et quantité d'information .....	38
2.3.3 Gain d'une expérience.....	39
2.3.4 Maximisation de l'entropie.....	40
2.3.5 Plans à entropie maximale.....	41
Chapitre 3 : Plans d'expérience numériques d'information de Kullback-Leibler minimale .....	44
3.1 Information de Kullback Leibler .....	44
3.2 Maximisation de l'entropie .....	46
3.3 Construction des plans.....	47
3.4 Estimation de l'entropie par Monte Carlo .....	48
3.4.1 La méthode de Monte Carlo .....	48
3.4.1.1 Méthode à noyaux .....	50
3.4.1.2 Estimation de l'entropie par plus proche voisin .....	55
Chapitre 4 : Implémentation, discussions et comparaison .....	58
4.1 Matlab comme outil de programmation .....	58
4.2 Etude de comparaison avec les plans usuels .....	59
4.3 Résultats et discussions .....	60
4.3.1 Résultat pour 10 points et 2 dimensions.....	60
4.3.2 Résultat pour 20 points et 3 dimensions.....	62
4.3.3 Résultat pour 40 points et 5 dimensions.....	64
4.3.4 Résultat pour 50 points et 7 dimensions.....	66
4.3.5 Résultat pour 80 points et 10 dimensions.....	67
Conclusion générale .....	70
Bibliographie.....	71

## Liste des figures

Figure 1-1: Un plan factoriel complet à 6 niveaux.....	6
Figure 1-2 : montrant quelques plans standards.....	10
Figure 1-3: Plans LHS avec 5 et 10 points. ....	12
Figure 1-4 : plans LHS pour 9 points. ....	12
Figure 1-5: Les 12 premiers points de la suite de Van Der Corput en base 3. ....	14
Figure 1-6 : Les 100, 1000 et 10000 premiers points de la suite de Halton en bases 2 et 3.....	14
Figure 1-7 : Des mauvaises projections en 2D. A gauche : les 80 premiers points d'une suite de Halton en bases 11 et 13. A droite : les 100 premiers points d'une suite de Halton en bases 17 et 19.....	15
Figure 1-8: (de la gauche vers la droite) Les 50, 250 et 500 premiers points d'une suite de Hammersley en base 2. ....	16
Figure 1-9: Les 100, 1000, 10000 premiers points d'une suite de Faure en dimension $s = 2$ et en base $b = 3$ . ....	17
Figure 1-10: Les 80 premiers points d'une suite de Faure en dimension 6, les 100 et 200 premiers points d'une suite de Faure en dimension 8, projetés sur le carré unité défini par les deux dernières dimensions.....	17
Figure 1-11 : Les 100, 1000 et 10000 premiers points d'une suite de Sobol. ....	18
Figure 1-12: A gauche : les 80 premiers points d'une suite de Sobol' en dimension 6 en projection sur le sous-espace des dimensions 5 et 6. A droite : les 100 et 150 premiers points d'une suite de Sobol' en dimension 8 sur le sous-espace des dimensions 7. ....	19
Figure 1-13 : Un $(0, 4, 2)$ -réseau en base $b = 2$ et des pavés de volume $b^{t-m} = 2^{-4}$ . ....	20
Figure 1-14: un plan Dmax de 300 points. ....	20
Figure 1-15: à droite Plans de Strauss de 40 points.....	21
Figure 2-1: une suite finie de $n = 30$ points en dimension 2 et un pavé $P = [0, t_1] \times [0, t_2]$ .....	24
Figure 2-2 :mantant le pavé $P'$ . ....	26

Figure 2-3 : Un sous-ensemble X pour le calcul de la discr�pance centr�e. ....	28
Figure 2-4: De la gauche vers la droite. Un plan al�atoire, une suite de Halton, et un plan factoriel complet � 49 points en dimension 2. ....	30
Figure 2-5: A gauche un plan � faible discr�pance et � droite une suite de Halton � 20 points en 2D. 32	
Discr�pance centr�e $L2 = 0.12$	discr�pance centr�e $L2 = 0.15$
Figure 2-6: Repr�sentation des projections sur les facteurs $X7, X8$ d'un plan � faible discr�pance sur la figure de gauche et d'une suite de Halton sur la figure de droite � 80 points en 8D. ....	33
Figure 2-7 : les 256 points d'une grille r�guli�re. ....	34
Figure 3-1 : figure illustre quelques noyaux. ....	52
Figure 3-2 Repr�sentation d'une fonction de densit� estim�e sur un �chantillon de taille 20 en dimension 2. ....	53
Figure 3-3: $N = 10$ observations dans $R^2$ , noyau gaussien de variance indiqu�e. ....	53
Figure 3-4: $N = 10$ observations dans $R^2$ , noyau gaussien de variance indiqu�e. ....	53
Figure 3-5: Evolution de l'entropie (MCGauss) en fonction de la taille de l'�chantillon pour des tirages al�atoires en dimension 3. ....	54
Figure 3-6: Evolution de l'entropie (MCGauss) au cours de l'algorithme d'�change pour des plans de taille 30 en dimension 3. ....	55
Figure 3-7: Evolution de l'entropie (PPV) en fonction de la taille de l'�chantillon pour des tirages al�atoires en dimension 3. ....	56
Figure 3-8: Evolution de l'entropie (PPV) au cours de l'algorithme d'�change pour des plans de taille 30 en dimension 3. ....	57
Figure 4-1 : boxplots des crit�res de qualit� pour deux dimensions. ....	61
Figure 4-2 : boxplots des crit�res de qualit� pour 3 dimensions. ....	63
Figure 4-3 : boxplots des crit�res de qualit� pour 5 dimensions. ....	65
Figure 4-4 : boxplots des crit�res de qualit� pour 7 dimensions. ....	67
Figure 4-5 : boxplots des crit�res de qualit� pour 10 dimensions. ....	69

## Liste des tableaux

Tableau 1-1: résumant les propriétés des plans usuels.....	22
Tableau 2-1: Discrépances en norme infinie pour $n=16, 25, 36$ et $49$ pour les 3 types de plans suivants aléatoires, suite de Halton, plan factoriel en dimension 2.....	30
Tableau 2-2 : Valeurs des critères de distance pour les 3 types de plans suivants : aléatoires, suite de Halton, plan factoriel où $n = 16, 25, 36$ et $49$ .....	37

## Liste des tables

Table 4-1: Listes des différents plans selon les quatre critères avec $n=10, d=2$ .....	60
Table 4-2 : Listes des différents plans selon les quatre critères avec $n=20, d=3$ .....	62
Table 4-3: Listes des différents plans selon les quatre critères avec $n=40, d=5$ .....	64
Table 4-4: Listes des différents plans selon les quatre critères avec $n=50, d=7$ .....	66
Table 4-5: Listes des différents plans selon les quatre critères avec $n=80, d=10$ .....	68

# Introduction

Pour des raisons de coût, de nombreux phénomènes scientifiques sont étudiés, non plus via l'expérimentation physique, mais à l'aide de modèles numériques. En effet, les progrès de ces dernières années concernant la modélisation mathématique de ces phénomènes ainsi que leur traitement numérique, permettent une étude extrêmement réaliste des phénomènes en question.

A titre d'exemples, nous pouvons citer l'étude de la production de pétrole brut en industrie pétrolière ([8], [18]), ou de la performance de circuits intégrés [34], ou encore la propagation d'un feu en milieu fermé [29]. De façon schématique, un modèle numérique peut être considéré comme un programme. Le simulateur prend des variables en entrée et donne une ou plusieurs réponses en sortie.

Une expérience numérique consiste alors à fixer un vecteur  $X$  de valeurs pour les variables d'entrée du simulateur puis à récolter la ou les réponses de celui-ci. On étudie ensuite le comportement de cette réponse en fonction des variations des variables d'entrée. Cependant, malgré les progrès des outils informatiques, le temps de calcul pour une expérience numérique reste très long.

L'objectif est donc de remplacer la réponse du simulateur par un modèle statistique simple représentant au mieux cette réponse.

## **Nous sommes alors confrontés à deux questions :**

- Comment choisir les valeurs des variables d'entrée,  $x_1, \dots, x_n$ , de façon à récolter un maximum d'informations sur le comportement de la réponse en un minimum de simulations ?
- Quel modèle statistique est approprié aux réponses du simulateur  $y(x_1), \dots, y(x_n)$  ?

Ces problématiques sont proches de celles rencontrées en expérimentation physique mais certaines particularités des expériences numériques doivent être prises en compte pour le traitement :

- les expériences sont déterministes, c'est-à-dire que deux simulations avec deux jeux de variables d'entrée identiques donnent la même réponse.
- les variables d'entrée sont très nombreuses. En effet, aux variables liées au phénomène physique, viennent s'ajouter des variables dues au modèle numérique (par exemple une taille de maillage).

Ce travail présente essentiellement une méthode de construction de plans d'expériences numériques basée sur **l'information de Kullback-Leibler (KL)**, nous introduisons l'information de KL comme mesure d'écart entre une distribution de points empirique et celle de la loi uniforme, ceci revient réciproquement à mesurer l'entropie de Shannon, nous détaillerons dans le chapitre 03 cette méthode.

Afin de comparer les résultats apportés par les plans d'expériences numériques d'information KL minimale nous avons besoin d'avoir connaissance sur les plans d'expériences usuellement utilisés en expérimentation numérique et les critères de comparaisons, ce que nous le verrons dans les deux premiers chapitres.

Enfin, nous comparons les plans **d'information de Kullback-Leibler** avec les plans usuels puis nous concluons les bonnes propriétés de ces plans dans le chapitre 04.

# Chapitre 1

## Les plans d'expériences numériques usuels.

### 1.1 Plan d'expériences classiques utilisées à l'expérimentation numérique.

#### 1.1.1 Rappels sur le modèle polynomial.

Les surfaces de réponse sont couramment utilisées pour les expériences numériques, nous pouvons par exemple citer [14] ou [8].

Cette technique bien connue consiste à utiliser un plan d'expérience classique tel qu'un plan Box-Behnken ou composite (voir par exemple [4]) pour sélectionner les valeurs des variables d'entrée.

Les réponses du simulateur en ces points permettent d'ajuster le modèle polynomial d'ordre deux,

$$y = a_0 + \sum a_i x_i + \sum_{i < j} a_{ij} x_i x_j + \sum a_i x_i^2 + \varepsilon \quad (1)$$

Où  $\varepsilon \sim N(0, \sigma^2)$  est un terme d'erreur.

Le modèle inclut certaines interactions  $x_i x_j$  et termes quadratiques  $x_i^2$  selon les choix spécifiés par l'utilisateur. La principale différence entre expériences physique et numérique est la réponse déterministe du simulateur. Le terme d'erreur  $\varepsilon$ , prend alors en compte un biais dû au modèle et non plus une erreur expérimentale, et on peut cependant considérer  $\varepsilon$  comme une erreur de mesure. En effet, l'objet de l'étude n'est pas la réponse du simulateur mais le phénomène simulé qui peut être considéré comme la réponse du simulateur plus une erreur aléatoire due aux simplifications du modèle mathématique ou aux erreurs systématiques du schéma numérique.

Considérons un plan d'expériences  $D = \{x_1, \dots, x_n\}$ . Les réponses du simulateur aux points du plan,  $Y_T = [y(x_1), \dots, y(x_n)]$ , sont utilisées pour estimer les paramètres inconnus du modèle (1),

$$\hat{y}(x) = X(x)\hat{A},$$

Où  $\hat{A} = (X^t X)^{-1} X^t Y$  est l'estimateur des moindres carrés de  $\beta$  et  $X = [X(x_1), \dots, X(x_n)]^t$  est la matrice du modèle.

Les outils statistiques traditionnels (étude des résidus, sélection de variables, ...) sont utilisés pour déterminer le meilleur modèle statistique.

**Remarque :** Les outils statistiques tels que la validation croisée ne semblent pas adaptés aux plans d'expériences classiques.

En effet, supprimer un point à un plan, par exemple factoriel, peut détruire sa structure, et il n'a alors plus les propriétés requises pour donner une bonne estimation des paramètres du modèle polynomial, on préférera alors valider le modèle par des expériences supplémentaires.

Les résultats, la méthodologie ainsi que les propriétés des plans d'expériences classiques sont exposés dans de multiples ouvrages tels que [2] ou [9].

### 1.1.2 Quelques éléments sur les plans d'expériences usuels

Pour ajuster le modèle (1) on utilise, soit des plans classiques tabulés tels que les plans composites, Box-Behnken...ect, soit des plans optimaux construits numériquement.

Il est cependant nécessaire de prendre quelques précautions dues au caractère déterministe des expériences :

- pas de répétition d'expériences puisque celles-ci donnent la même réponse.
- pas de randomisation des expériences car les conditions expérimentales n'influent pas sur la réponse.

On peut envisager des plans en blocs si certaines des variables d'entrée discrètes prennent en compte un contexte ou différents scénarios [18].

Quel que soit le plan utilisé, il peut présenter des inconvénients importants pour les expériences numériques:

- le nombre de niveaux testés pour chaque variable d'entrée est en général faible.

- les points se situent souvent sur les bords du domaine expérimental, Cette répartition des points, qui est optimale pour ajuster un modèle polynomial, ne permet pas de capter des éventuelles irrégularités à l'intérieur du domaine expérimental, Or la complexité des modèles mathématiques laisse supposer que la réponse du simulateur sera irrégulière et que le modèle polynomial ne sera pas adapté à cette réponse.

**Remarque** : Nous pouvons noter que dans un cas particulier, [5] ont établi un critère d'optimalité sélectionnant des points à l'intérieur du domaine expérimental (cas où les variables d'entrée sont incertaines mais connues avec une loi de probabilité).

### **Pourquoi utilise-t-on encore des plans d'expériences classiques en simulation?**

Les plans d'expériences classiques ne sont pas remis en cause, Si on les utilise dans le domaine où ils sont applicables, ils sont de très bonne qualité, Mais si les hypothèses initiales ne sont pas vérifiées, comme cela semble être souvent le cas en simulation, leur utilisation n'est pas appropriée.

Cependant, malgré la complexité des modèles numériques, un simple modèle polynomial peut être suffisant pour approcher la réponse du simulateur pour diverses raisons :

- La réponse du simulateur peut-être lisse même si le modèle mathématique décrivant le phénomène simulé est très complexe.
- Dans de nombreuses applications, le domaine d'étude peut être plus petit que le domaine de variation des variables d'entrée
- Le modèle polynomial peut être très utile pour faire une analyse de sensibilité.

Enfin une dernière raison pour laquelle la méthode des plans d'expériences classique est souvent utilisée vient du fait qu'elle est bien connue chez les ingénieurs et l'interprétation physique du modèle polynomial est simple, de plus elle est disponible en logiciel, Cependant, « l'automatisation » de cette méthode amène quelquefois à une mauvaise utilisation, le logiciel étant alors utilisé en boîte noire sans s'assurer que l'approche est bien adaptée au problème à traiter.

### **Avantages et inconvénients :**

Les deux principaux avantages de cette méthode sont sa simplicité et le faible coût de simulation.

En effet, si une surface de réponse polynomiale est adaptée à la réponse étudiée, alors cette approche est celle qui nécessite le moins de simulations.

Cependant, il se peut que ce modèle statistique soit trop simple au regard de la complexité de la réponse du simulateur.

Dans de nombreux cas il est nécessaire de faire appel à des modèles statistiques plus sophistiqués tels que le krigeage.

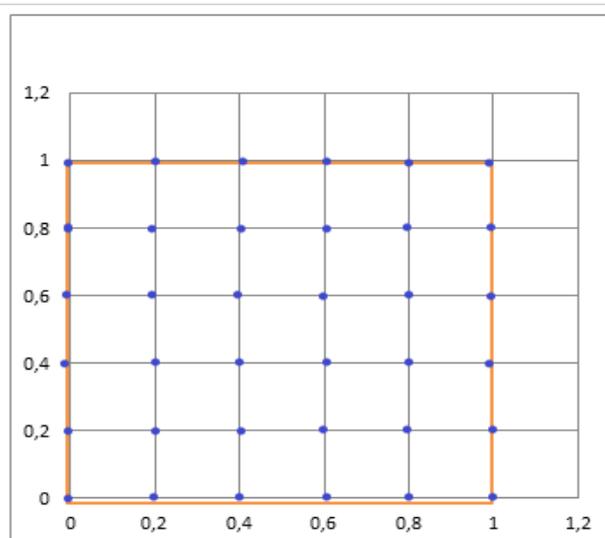
De plus, comme nous l'avons signalé précédemment, les points des plans classiques sont positionnés sur les bords du domaine expérimental, ils ne sont donc pas adaptés pour détecter les irrégularités à l'intérieur du domaine), ni d'ailleurs pour ajuster un modèle statistique autre que polynomial.

### 1.1.3 Plan factoriel complet

La méthode la plus simple pour répondre à la problématique du remplissage de l'espace consiste à choisir les points sur une grille régulière dans le domaine expérimental [15].

**Description :** Pour obtenir une grille régulière à  $k$  niveaux, il suffit de choisir  $k$  valeurs réparties à différents intervalles réguliers pour chaque facteur, Par exemple, dans le carré unité  $[0,1] * [0,1]$  nous choisissons 6 niveaux, on peut considérer la grille de points suivante (voir figure 1-1).

$$\{0, 0.2, 0.4, 0.6, 0.8, 1\} \times \{0, 0.2, 0.4, 0.6, 0.8, 1\}$$



**Figure 1-1 : Un plan factoriel complet à 6 niveaux.**

Il est évident que plus le nombre de niveaux sera important et plus le remplissage de l'espace sera satisfaisant mais cela entraîne une « explosion » du nombre de simulations, il est

donc essentiel de trouver un bon compromis et de choisir les niveaux paraissant les plus adaptés au problème.

**Discussion :** Cette méthode reste intéressante si la dimension du problème est petite, c'est-à-dire 2, voire 3, dès que la dimension est plus élevée, le nombre de simulations  $n = k^d$  augmente considérablement, et l'idée d'utiliser une grille devient alors inenvisageable, de plus, seulement  $k$  valeurs différentes apparaissent pour chaque dimension, dans le cas où la réponse dépend très fortement de seulement quelques variables (par exemple, une ou deux en dimension 5), le plan factoriel conduit à beaucoup de réplifications inutiles, à titre d'exemple cela pourrait nous conduire à étudier uniquement une minorité de points parmi tous les points du plan.

Ce type de plan devient donc inutilisable lorsque la dimension augmente car on perd la plupart des points en projection sur les axes factoriels.

**Remarque :** Si le nombre de coefficients du modèle à estimer est proche du nombre d'expériences, il est conseillé d'ajouter au plan factoriel quelques points répartis uniformément dans le domaine.

### 1.1.4 Plans factoriels fractionnaires [15]

Nous sommes dans l'obligation de réduire le mieux possible le nombre de simulations en récoltant le maximum d'information possible sans aucune perte, ce qui est n'est plus permis en utilisant des plan factoriel complet.

Par contre les plans fractionnaires paraissent plus intéressants, en nous permettant de réduire le nombre de simulations possibles à  $n = k^{d-p}$ , cependant ce type de plan présente les mêmes inconvénients que les plans factoriels, de plus il n'est plus conseillé de les utilisés si nous n'avons pas une connaissance minimum a priori sur les interactions.

### 1.1.5 Plans composites

Les plans composites ont été conçus pour assurer une précision à peu près uniforme des estimations de réponse  $\hat{y}$  dans la totalité du domaine expérimenté, avec le plus petit nombre d'essais possible, ces plans comprennent 3 catégories d'essais établis de telle sorte qu'à chaque variable  $x$  correspondent 3 niveaux :

- des essais d'un plan factoriel.

- des essais « en étoile » par rapport aux essais précédents (au total  $2n$ , soit 4 pour 2 facteurs).
- des essais au centre du domaine, 3<sup>ème</sup> niveau de chaque facteur.

D'autres types de plan peuvent être générés en faisant varier la distance entre le point central et les points aux bords du domaine, comme les plans composites centrés ordinaires, les plans composites à faces centrées, les plans composites centrés inscrits, etc.

**Discussion :** le plan composite présente les mêmes problèmes d'augmentation rapide de nombre simulations vus précédemment.

Ces plans ne répondent pas convenablement aux objectifs de remplissage de l'espace et notamment aux critères de bonne répartition de points en projection.

En effet, ces plans ne testent que 3 ou 5 niveaux (selon la valeur de  $x$ ) par paramètres et cela avec n'importe quelle taille du plan.

### 1.1.6 Plans de Box-Behnken

Dans le cas des plans  $2^{(k-p)}$ , Plackett et Burman (1946) ont développé des plans hautement fractionnés permettant d'obtenir le nombre maximum d'effets (principaux) avec le minimum d'essais.

L'équivalent pour un plan  $3^{(k-p)}$  est un plan dit de Box-Behnken [3], voir aussi [10].

Ces plans ne possèdent pas de générateurs de plan simple (ils sont construits en combinant des plans factoriels à deux niveaux avec des plans incomplets de blocs), et possèdent des interactions confondues complexes.

Ces plans ont toutefois l'avantage d'être économiques et donc particulièrement utiles lorsque les essais expérimentaux à réaliser sont coûteux.

**Discussion :** En termes de nombre de points, un plan de Box-Behnken est comparable à un plan composite en dimension 3 et 4, à noter qu'il n'existe pas de Box-Behnken en dimension 2.

Ces plans ayant leurs points sur les axes factoriels du domaine, ils ne remplissent pas correctement l'espace.

### 1.1.7 Conclusion

Les plans étudiés dans la première section du chapitre sont généralement de bonne qualité concernant le remplissage de l'espace.

Cependant, la plupart de ces méthodes sont soit coûteuses en nombre de simulations ou/et soit une cause de perte d'information dans le contexte des plans d'expériences numériques.

C'est essentiellement pour ces raisons que nous avons choisi de n'étudier qu'une infime partie de tous les plans d'expériences classiques connus, cela ne signifie pas qu'il faut les écarter pour autant des plans d'expériences numériques.

En effet les propriétés recherchées sur un plan construit en but de la simulation numérique sont les suivantes :

- expériences déterministes (*erreur* = 0).
- grand nombre de facteurs.
- larges domaines de variation.
- variables d'intérêt multiples.
- modèles fortement non linéaires.

La figure ci-dessous a pour objectif de montrer le remplissage des points dans les plans standards.

## Quelques plans d'expériences « classiques »

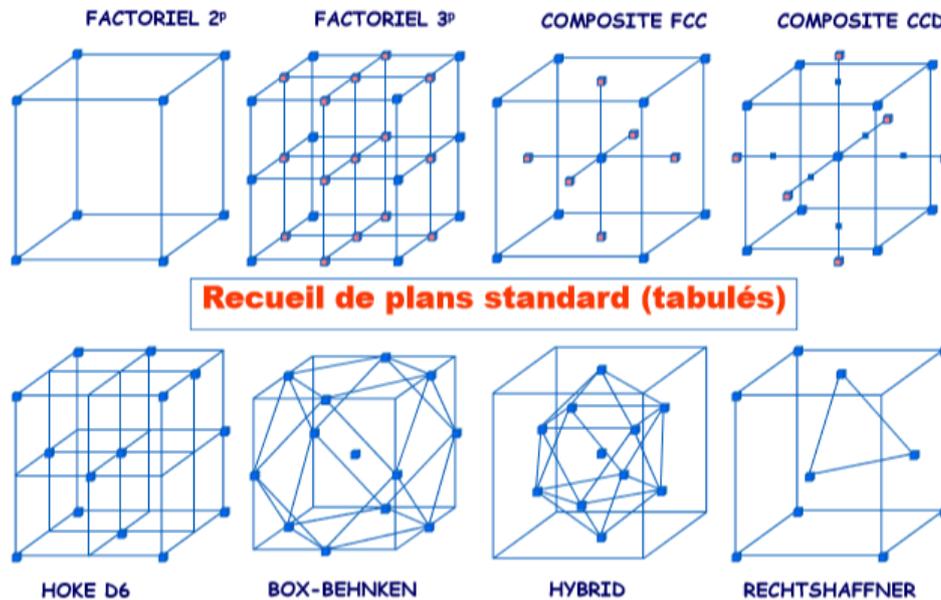


Figure 1-2 : montrant quelques plans standards.

### 1.2 Les plans uniformes ( *SpaceFillingDesign* «SFD» )

Le développement de codes qui modélisent ou simulent des phénomènes complexes sont de plus en plus réalistes et même si la puissance des ordinateurs augmente sans cesse, les temps de calcul demeurent importants et limitent ainsi le recours aux techniques de Monte-Carlo.

Le développement de méta modèle permet alors de remplacer le simulateur par un outil plus « simple » construit à partir du simulateur complexe.

Ces méta modèles ou surfaces de réponse sont en général des fonctions obtenues à l'aide de méthodes d'interpolation ou d'approximation à partir d'un nombre limité d'exécutions du simulateur sur des jeux de paramètres constituant le plan d'expériences numériques.

Du fait des caractéristiques non linéaires et/ou non paramétriques des codes de modélisation ou simulation, il est nécessaire de répartir les points dans l'espace le plus uniformément possible de façon à capter au mieux le comportement du simulateur, C'est le

mode de répartition des points que cherchent à proposer les plans d'expériences « Space Filling Designs ».

La particularité qu'un plan uniforme possède nous impose la nécessité de définir des critères intrinsèques de qualité en termes d'uniformité.

Ces plans sont basés sur deux types de critères que nous détaillerons dans le chapitre suivant :

- critère de distances entre les points du plan.
- critère de répartition uniforme des points (discrépance).

### 1.2.1 Les plans (S, F, D) basés sur des contraintes marginales (Hyper cube latin (LHS))

La méthode d'échantillonnage par hyper cube latin (Mac Kay, Conover, Beckman, 1979 [25] ) a été introduite pour évaluer numériquement les intégrales multiples.

Elle permet d'assurer la non redondance de l'information au travers d'une bonne répartition des projections sur les axes factoriels.

En pratique, les hyper cubes latins sont très utilisés en planification d'expériences numériques notamment pour leur simplicité d'usage et de construction.

**Définition :** Un hyper cube latin à  $n$  points et  $d$  variables est une matrice  $n \times d$  dont chacune des colonnes est une permutation de l'ensemble  $\{1, 2, \dots, n\}$ , on le note  $LHD(n, d)$ .

Ainsi, pour un plan LHS à  $n$  observations, chaque dimension de l'espace sera découpée en  $N$  intervalles de la façon suivante  $\left\{ \left[0, \frac{1}{n}\right], \left[\frac{1}{n}, \frac{2}{n}\right], \dots, \left[\frac{n-1}{n}, 1\right] \right\}$ , et un point par intervalle sera choisi (et ce pour chaque dimension).

Par exemple, sur  $R^2$ , un LHS à 5 points et un LHS à 10 points peuvent donner les plans suivants :

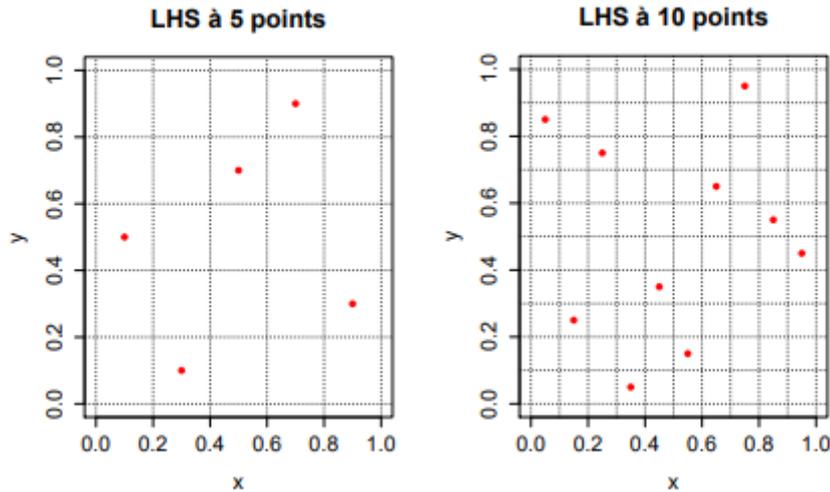


Figure 1-3 : Plans LHS avec 5 et 10 points.

La permutation des colonnes sur les  $d - 1$  dimensions génère  $n!^{d-1}$  LHS possibles, dans ce cas, différents critères nous imposent le meilleur plan à choisir comme (le remplissage, l'uniformité, l'indépendance).

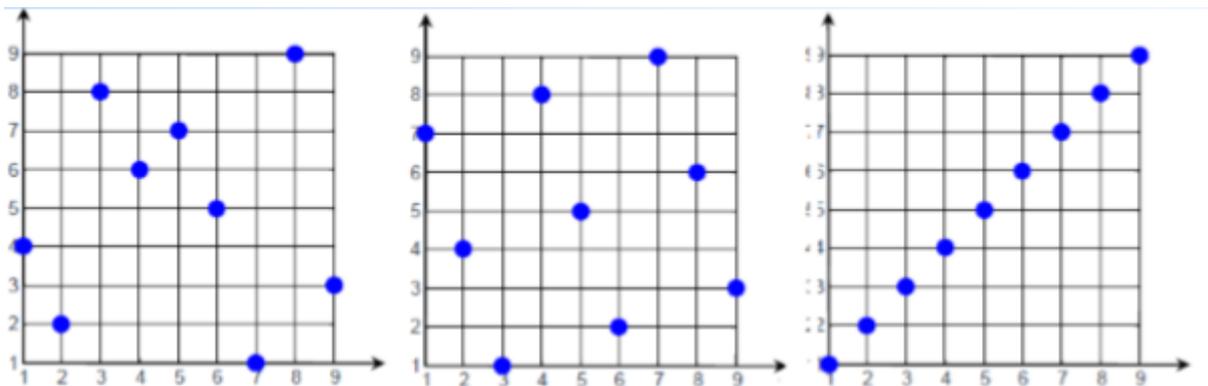


Figure 1-4 : plans LHS pour 9 points.

**Discussion :** les LHS assurent la propriété de la projection uniforme sur les marges, ce qui est n'est pas le cas dans tout l'espace, de plus le contexte dans lequel nous ne nous autorise pas à prendre un grand nombre de points.

## 1.2.2 Les plans (S, F, D) basés sur des lois de distribution des points (Suites à faible discrédance) [15], [33]

La discrédance mesure la dérivation de la répartition dans l'espace d'un échantillon de points par rapport à celle d'un échantillon de même taille tiré par loi uniforme, Plusieurs définitions de la discrédance existent.

Pour une suite  $(x_{i=1\dots N}^{(i)})$  elle peut notamment s'écrire :

$$D(x) = \sup_{y \in [0,1]^k} \left| \frac{1}{N} \sum_{k=1}^N 1_{(x^{(k)} \in [0,y])} - \text{volume}([0,y]) \right|$$

Les suites à faible discrédance sont des suites quasi-aléatoires ayant une faible valeur de discrédance.

Ce sont potentiellement des plans d'expérience très satisfaisants, car elles permettent un bon remplissage de l'espace.

### 1.2.2.1 Les suites de Van Der Corput

Sont le modèle le plus simple des suite à faible discrédance construite sur un espace unidimensionnel.

Soit un entier  $b \geq 2$ , tout entier  $i \in N$  peut se décomposer dans la base  $b$  d'une manière unique :

$$i = \sum_{k=0}^{\infty} c_k(i) b^k, \text{ avec } c_k(i) \in \{0, \dots, b-1\}, \forall k \in N.$$

On remarque que les coefficients  $c_k(i)$  sont nuls pour tout  $k > 1 + [\log_b(i)]$ , où  $[\cdot]$  désigne la partie entière inférieure, par conséquent, la somme précédente est finie.

A partir de cette décomposition, on définit la fonction radicale inverse par :

$$\varphi_b(i) = \sum_{k=0}^{\infty} c_k(i) b^{-k-1}, \forall i \in N, \varphi_b(i) \in [0,1[.$$

**Définition :** soit un entier  $b \geq 2$ , La suite  $VDC_d = (x_0, x_1, \dots) \in [0,1[$  donnée par :

$$x_i = \varphi_b(i)$$

est appelée suite de Van Der Corput en base  $b$ .



Figure 1-5 : Les 12 premiers points de la suite de Van Der Corput en base 3.

### 1.2.2.2 Suites de Halton

Les suites de Halton sont une généralisation en dimension  $s \geq 1$  des suites de Van Der Corput.

**Définition :** soient  $b_1, \dots, b_s$ ,  $s$  entiers positifs premiers entre eux, la suite  $H_{b_1, \dots, b_s} = (x_1, x_2, \dots)$  donnée par :

$$x_i = \left( \varphi_{b_1}(i), \dots, \varphi_{b_s}(i) \right) \in [0,1]^s.$$

Est appelée la suite de Halton en bases  $b_1, \dots, b_s$ .

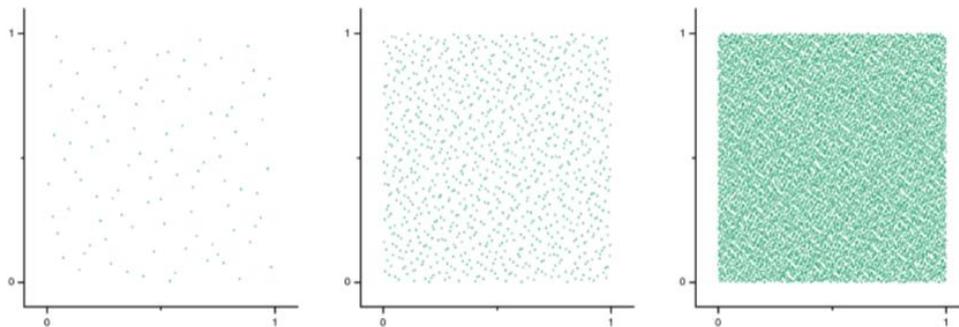


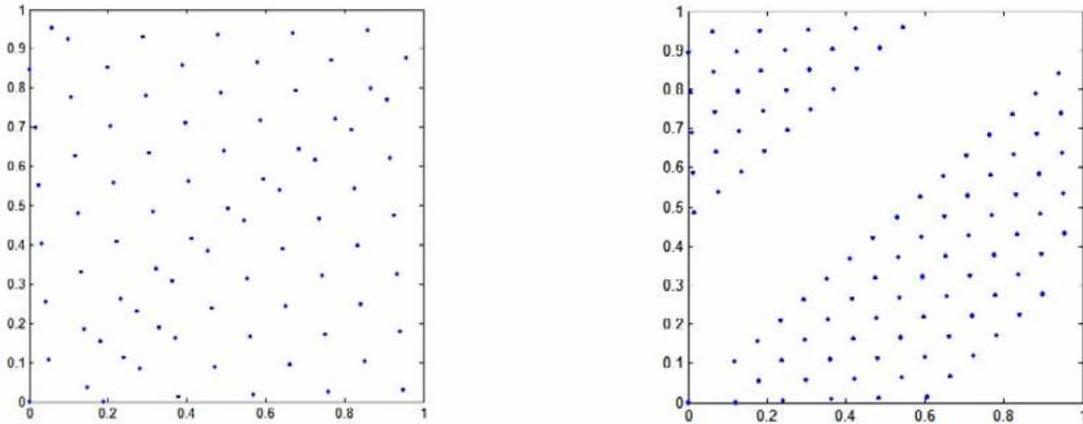
Figure 1-6 : Les 100, 1000 et 10000 premiers points de la suite de Halton en bases 2 et 3.

**Discussion :** les suites de Halton assurent le caractère de l'uniformité, cependant, des pathologies se présentent en dimension élevée.

En effet, lorsque l'on projette les points d'une telle suite dans le carré unité, l'allure de la structure est parfois décevante (figure 1-7), Nous pouvons observer des diagonales décalées à chaque itération, ce qui laisse place à de nombreuses zones lacunaires dans le carré unité.

Ce phénomène résulte de la régularité des sous-suites de longueur  $b$  des suites de Van der Corput.

A noter cependant que cette pathologie ne se manifeste pas pour n'importe quelle paire de bases, en général, il est plus fréquent d'observer ce phénomène en dimension élevée et lorsque la différence entre les deux bases est très petite par rapport aux bases utilisées.



**Figure 1-7 : Des mauvaises projections en 2D. A gauche : les 80 premiers points d'une suite de Halton en bases 11 et 13. A droite : les 100 premiers points d'une suite de Halton en bases 17 et 19.**

### 1.2.2.3 Les suites finies de Hammersley

Une suite finie de  $n$  points de Hammersley en dimension  $s$ , est construite à partir d'une suite finie de  $n$  points d'une suite de Halton en dimension  $s - 1$ .

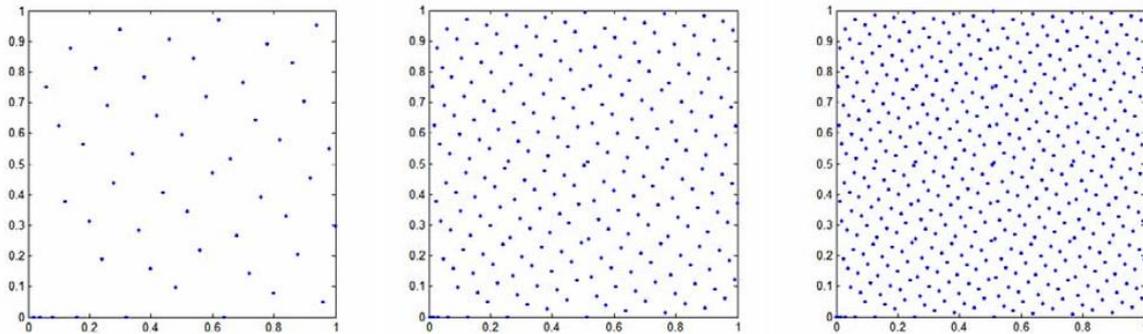
**Définition :** soient  $b_1, \dots, b_{s-1}$ ,  $s - 1$  entiers premiers positifs distincts, La suite finie de  $n \in \mathbb{N}^*$  points  $H_{n, b_1, \dots, b_s} = (x_1, \dots, x_n)$  donnée par :

$$x_i = \left( \frac{i}{n}, \varphi_{b_1}(i), \dots, \varphi_{b_{s-1}}(i) \right) \in [0, 1]^s.$$

Est appelée suite finie de Hämmerli en bases  $b_1, \dots, b_{s-1}$ .

**Discussion :** ces suites étant construites à partir de suites de Halton, elles présentent le même phénomène de diagonales successives, de plus, il est impossible de rajouter des points supplémentaires à ces suites sans pour autant perturber la discrèpance.

Dans le cas où le nombre de points à générer n'est pas connu à l'avance, il est déconseillé d'utiliser une suite de Hammersley, aussi, ces suites ont perdu l'aspect itératif des suites de Halton qui permettait de rajouter des points facilement.



**Figure 1-8 : (de la gauche vers la droite) Les 50, 250 et 500 premiers points d'une suite de Hammersley en base 2.**

### 1.2.2.4 Les suites de Faure

Les suites de Faure sont définies à partir d'une base unique  $b$ , où  $b \geq s$ , est un entier premier.

**Définition :** soient  $b \geq s$  un entier premier, et  $B = (b_1, \dots, b_s)$ ,  $s$ , éléments de  $\{0, \dots, b - 1\}$ .

La suite de Faure  $F_{b,B} = (x_1, x_2, \dots)$  est définie par :

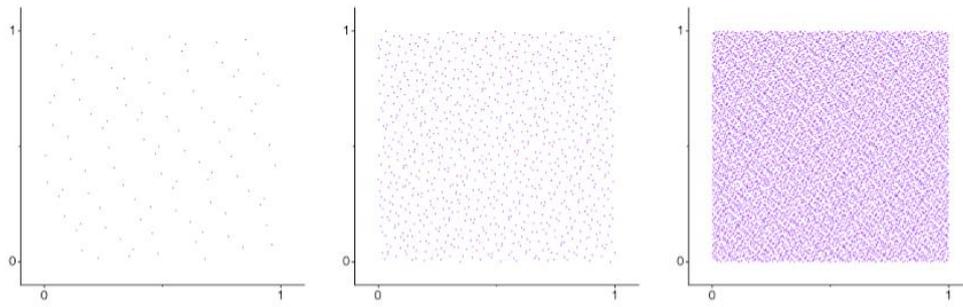
$$x_i^j = \sum_{k=1}^{\infty} x_i^{j,k} b^{-k}, \forall (i \geq 1 \text{ et } j = 1, \dots, s).$$

Où

$$x_i^{j,k} = \sum_{p=k}^{\infty} C_p^{k-1} b_j^{p-k+1} c_k(i) \pmod{b}$$

Avec  $c_k(i)$  pour  $k \geq 1$ , les coefficients de la décomposition de  $i$  dans la base  $b$ , et  $C_p^k = \frac{p!}{(p-k)!k!}$ , est un coefficient binomial.

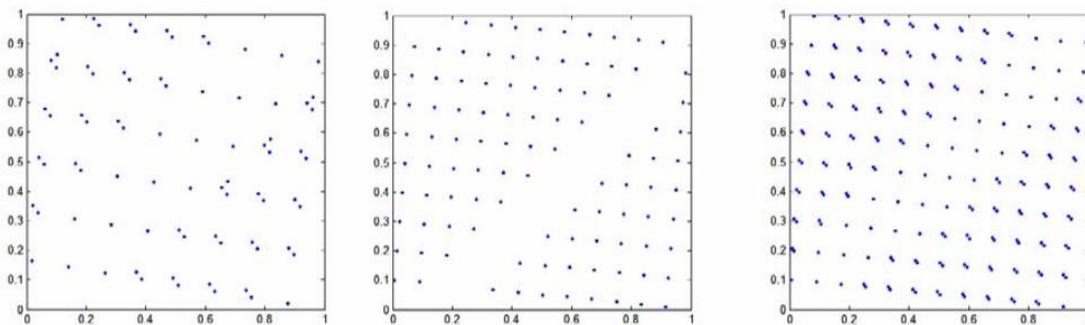
Les coefficients  $c_k(i)$  étant nuls pour  $k > 1 + \lfloor \log_b i \rfloor$ , les sommes de la définition ci-dessus sont en fait finies.



**Figure 1-9 : Les 100, 1000, 10000 premiers points d'une suite de Faure en dimension  $s = 2$  et en base  $b = 3$ .**

**Discussion :** Les suites de Faure permettent de générer une distribution uniforme localement (figure 1-9), Cependant, les projections sur le carré unitaire des points de l'hyper cube sont échantillonnées en bandes (figure 1-10).

Nous pouvons remarquer que les nouveaux points sont placés dans le voisinage de ceux produits précédemment.

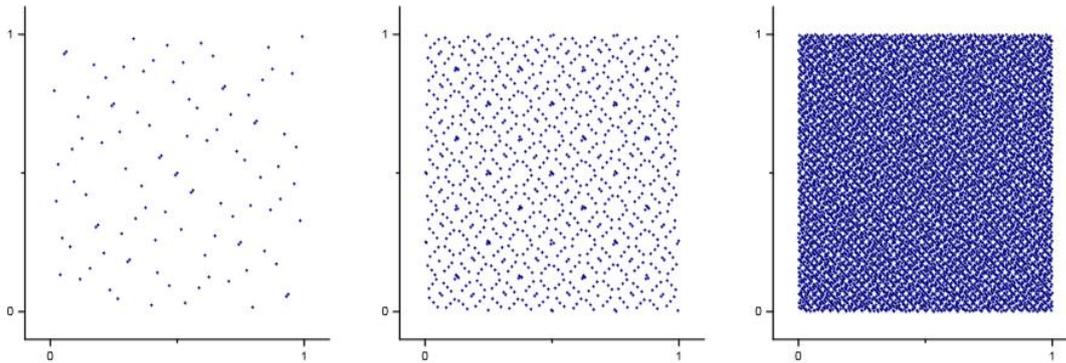


**Figure 1-10 : Les 80 premiers points d'une suite de Faure en dimension 6, les 100 et 200 premiers points d'une suite de Faure en dimension 8, projetés sur le carré unité défini par les deux dernières dimensions.**

### 1.2.2.5 Les suites de Sobol

Les suites de Sobol sont construites en utilisant des propriétés arithmétiques des polynômes sur  $Z_2$ , à noter que nous ne détaillerons pas la construction de ces suites. Pour les

applications, les suites de Sobol utilisées sont générées à l'aide du logiciel «R» qui reprend l'algorithme de Bratley et Fox.



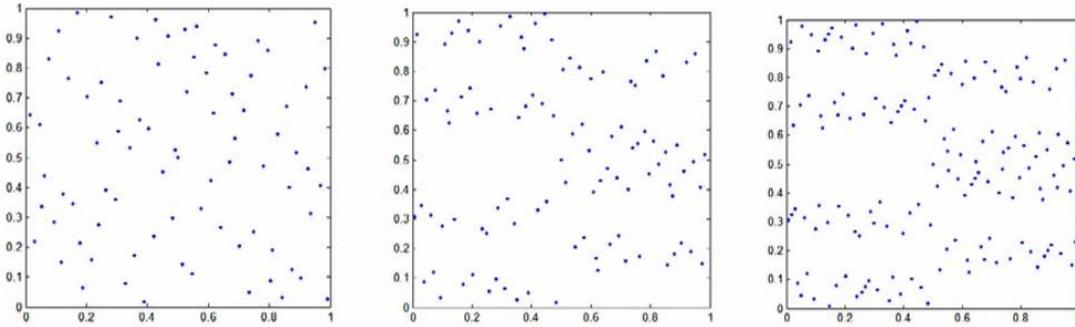
**Figure 1-11 : Les 100, 1000 et 10000 premiers points d'une suite de Sobol.**

**Discussion :** ces suites ont l'avantage d'être rapides à construire car la nature binaire des ordinateurs permet de réduire le temps de calcul, elles ont aussi l'avantage de préserver généralement l'uniformité de la distribution lorsque la dimension augmente.

Cependant, en raison de la nature arbitraire des états initiaux  $I_1, \dots, I_s$  de ces récurrences, les points sont souvent mal répartis dans l'hypercube unité (voir figure 1-12).

Seulement, on ne peut difficilement prédire les états initiaux qui donnent des distributions non uniformes.

Un autre inconvénient apparaît si l'on souhaite rajouter de nouveaux points à une distribution, dans ce cas, on observe qu'ils tombent dans le voisinage de ceux produits précédemment augmenter le nombre de points n'est donc pas une solution envisageable.



**Figure 1-12 : A gauche : les 80 premiers points d'une suite de Sobol' en dimension 6 en projection sur le sous-espace des dimensions 5 et 6. A droite : les 100 et 150 premiers points d'une suite de Sobol' en dimension 8 sur le sous-espace des dimensions 7.**

### 1.2.2.6 Les (t, s)-suites et (t, m, s)-réseaux

Les (t, s)-suites qui ont été introduite par H.Niederreiter, sont une généralisation des suites de Van Der Corput, Sobol et Faure.

Tout d'abord il est nécessaire de définir un pavé élémentaire en base  $b, \forall b \in \mathbb{N}^*, P_b$  est un pavé de la forme

$$: P_b = \prod_{j=1}^s \left[ \frac{a_j}{b^{d_j}}, \frac{a_j + 1}{b^{d_j}} \right]$$

Où  $a_j, d_j \in \mathbb{N}$  et  $a_j < b^{d_j}$ , pour tout  $j = 1, \dots, s$ .

**Définition :** soient  $t$  et  $m$  deux entiers tels que  $0 \leq t \leq m$ .

Un  $(t, m, s)$ -réseau en base  $b$  est une suite finie  $x = (x_1, \dots, x_{b^m})$  de  $b^m$  points de  $[0,1]^s$  telle que pour tout pavé élémentaire  $P_b$  en base  $b$  de volume  $\lambda(P_b) = b^{t-m}$ , avec  $A(P_b, x) = b^t$ , est le nombre de points de  $x$  qui appartiennent au pavé  $P_b$ .

Une illustration du concept de  $(t, m, s)$ -réseau est donné par la Figure 1-1, et sur laquelle on peut vérifier que pour les  $b^m = 16$  points du  $(0,4,2)$ -réseau, tout pavé de volume  $b^{t-m} = 2^{-4}$  ne contient que  $b^t = 1$  point.

**Définition :** soit un entier  $t \geq 0$ . Une  $(t, s)$ -suite en base  $b$  est une suite de point  $x = (x_1, x_2, \dots)$  telle que pour toute paire d'entier  $k \geq 0$  et  $m \geq t$ , la suite finie  $(x_{kb^m}, \dots, x_{(k+1)b^m-1})$  est un  $(t, m, s)$ -réseau en base  $b$ .

Il est à noter que les suites de Sobol sont des  $(t, s)$ -suites en base  $b = 2$  où le paramètre  $t$  dépend de  $s$ .

Les suites de Faure sont des  $(0, s)$ -suites, enfin, les suites de Van Der Corput, sont des  $(0, 1)$ -suites.

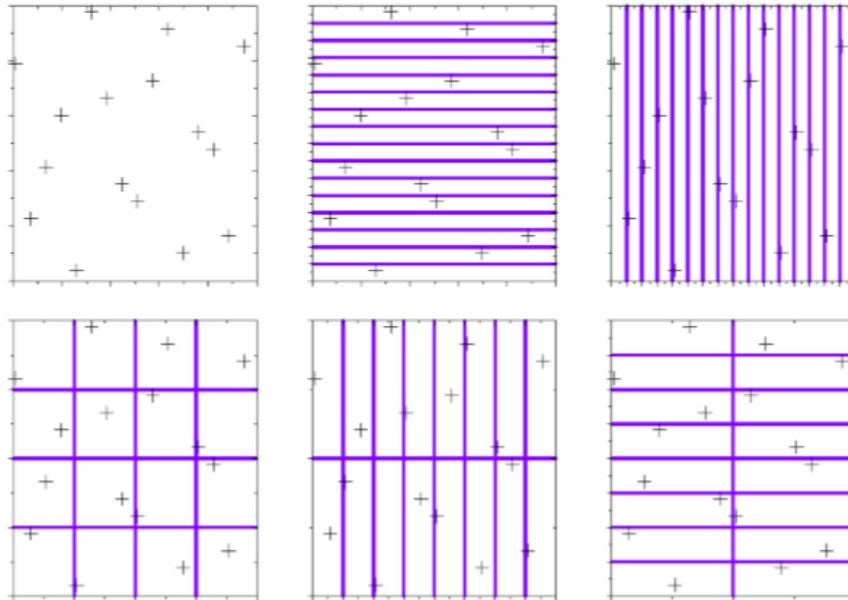


Figure 1-13 : Un  $(0, 4, 2)$ -réseau en base  $b = 2$  et des pavés de volume  $b^{t-m} = 2^{-4}$ .

**Remarque :** Il existe d'autres types plans uniformes nous les citons sans les expliciter.

- Plans à entropie maximale : Plans Dmax.

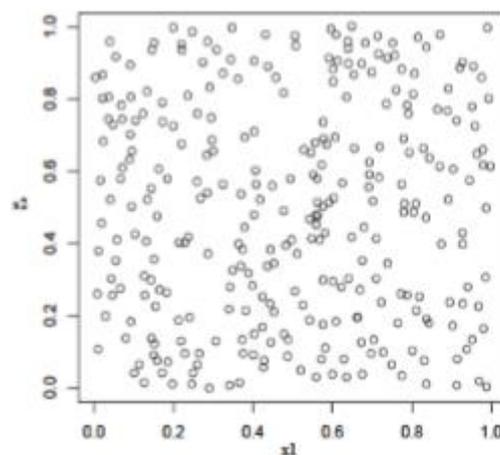
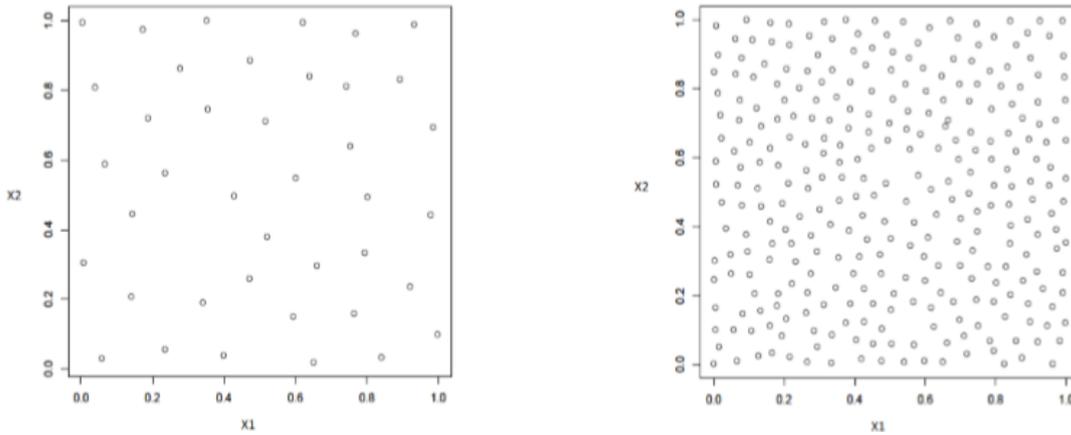


Figure 1-14 : un plan Dmax de 300 points.

- Plans de Strauss: inspiré du principe de répulsion entre particules.



**Figure 1-15 : à droite Plans de Strauss de 40 points.**

**À gauche Plans de Strauss de 400 points.**

### 1.3 Discussion

Nous présentons sur le tableau ci-dessous, les propriétés vérifiées par les plans étudiés au sein de ce chapitre.

Chaque type de plan est critiqué en regardant si :

- Les distributions obtenues permettent ou non de remplir l'hypercube uniformément.
- les distributions, notamment en projection dans le carré unité, suivent des alignements.
- les distributions sont séquentielles c'est-à-dire s'il est possible de rajouter des points à cette distribution sans pour autant détruire l'uniformité initiale.
- à partir des distributions obtenues, les distributions en dimension inférieure ont une meilleure uniformité.

	<b>A</b>	<b>b</b>	<b>c</b>	<b>d</b>
<b>Aléatoires</b>	<b>Non : lacune.</b>	<b>non</b>	<b>oui</b>	<b>Oui</b>
<b>Doehlert</b>	<b>Oui</b>	<b>oui</b>	<b>oui</b>	<b>Non</b>
<b>Grille régulière</b>	<b>Oui</b>	<b>oui</b>	<b>non</b>	<b>Non</b>
<b>Hypercube latin</b>	<b>Oui, en version centrée, sinon, lacune</b>	<b>Oui en version centrée</b>	<b>non</b>	<b>Oui en dimension 1</b>

<b>Tableaux orthogonaux</b>	<b>Oui, en version centrée, sinon, lacune</b>	<b>oui</b>	<b>non</b>	<b>Oui, en dimension égale a la force du tableau</b>
<b>Good lattice points</b>	<b>Oui, si les paramètres sont bien choisis</b>	<b>oui</b>	<b>oui</b>	<b>Non</b>
<b>Halton</b>	<b>Oui</b>	<b>Oui, en dimension élevée</b>	<b>oui</b>	<b>Non</b>
<b>Hammersley</b>	<b>Oui</b>	<b>Oui, en dimension élevée</b>	<b>oui</b>	<b>Non</b>
<b>Faure</b>	<b>Oui</b>	<b>oui</b>	<b>oui</b>	<b>Non</b>
<b>Nederreiter</b>	<b>Oui</b>	<b>non</b>	<b>oui</b>	<b>Non</b>
<b>Sobol</b>	<b>Oui</b>	<b>Non, si les états initiaux sont biens choisis</b>	<b>oui</b>	<b>non</b>

Tableau 1-1: résumant les propriétés des plans usuels.

## Chapitre 2

### Critères et plans optimaux.

La difficulté de détecter ou de mesurer l'uniformité d'un plan nous impose la nécessité de définir des critères intrinsèques de qualité en termes d'uniformité. Nous verrons dans ce chapitre deux types de critères intrinsèques mesurant l'uniformité d'un plan :

- Critère basés sur la notion de discrédance :

Ces critères sont fondés sur la conception de comparer une distribution empirique de points à une distribution de points uniforme.

- Critères basés sur la notion de la distance :

Ces critères sont fondés sur la conception de comparer une distribution de points données à une grille régulière.

#### 2.1 Critères d'uniformité basés sur la discrédance et plans à faible discrédance [15], [34]

##### 2.1.1 La discrédance locale

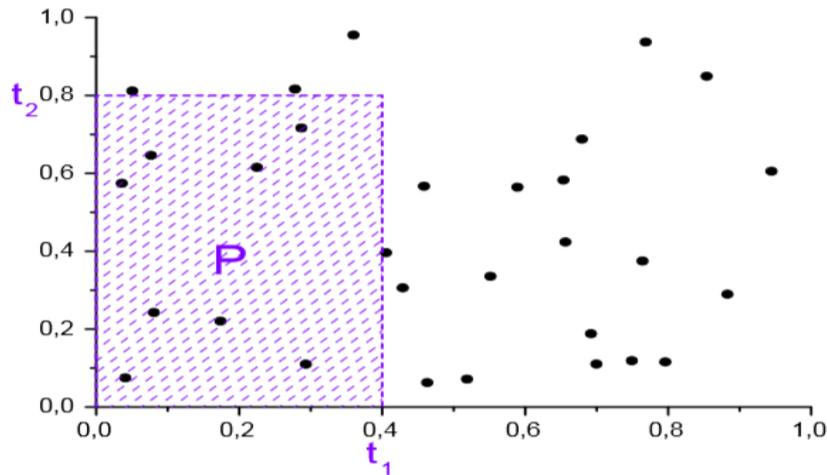
**Définition :** la discrédance est une mesure de l'écart existant entre une situation de référence (l'uniformité théorique) et une configuration donnée.

Soient  $n$  points,  $X = (x_1, \dots, x_n)$  dans  $[0,1]^s$ ,  $x_i = (x_i^1, \dots, x_i^s)$  pour  $i = 1, \dots, n$  et  $t = (t^1, \dots, t^s) \in [0,1]^s$ .

La discrédance locale « disc » est définie par :

$$disc(t) = \frac{A(t, x)}{n} - \prod_{j=1}^s t^j$$

Où  $A(t, x)$  est le nombre de points de  $X$  qui appartiennent au pavé  $P = [0, t^1[ \times \dots \times [0, t^s[$ .



**Figure 2-1 :** une suite finie de  $n = 30$  points en dimension 2 et un pavé  $P = [0, t^1] \times [0, t^2]$ .

Comme nous l'avons dit précédemment, la discrédance mesure l'écart entre une distribution de points empirique et une distribution uniforme, à la lumière de cette conception nous remarquons que si la proportion de points étant situés dans le pavé  $P$ , est bien proche de la quantité  $\prod_{j=1}^n t^j$ , alors la distribution empirique se rapproche de l'uniformité.

### Équipartition :

**Définition :** une suite de points  $X = (x_1, \dots, x_n)$  est dite équirépartie si :

$$\lim_{n \rightarrow \infty} \left( \sup_{P=[0, t^1] \times \dots \times [0, t^s] \subset [0, 1]^s} \left| \frac{A(t, x)}{n} - \prod_{j=1}^s t^j \right| \right) = 0$$

L'égalité peut s'écrire de la forme suivante :

$$\lim_{n \rightarrow \infty} \left( \sup_{P=[0, t^1] \times \dots \times [0, t^s] \subset [0, 1]^s} \left| \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^s 1_{x_i^j \leq t^j} - \prod_{j=1}^s t^j \right| \right) = 0$$

Ce qui signifie qu'une suite est équirépartie lorsque sa fonction de répartition empirique converge vers la fonction de répartition uniforme.

On voit ainsi qu'une suite de variables aléatoires uniformes indépendantes est équirépartie.

Après avoir connaissance sur l'équipartition, nous pouvons maintenant aborder la notion de discrédance dans ces divers types.

### 2.1.2 La discrédance infinie à l'origine

**Définition :** soit  $X = (x_1, \dots, x_n)$  dans  $[0,1]^s$ , une suite finie de  $n$  points données, la discrédance infinie à l'origine notée  $D_{\infty,*}$  est la norme  $L_{\infty}$  de la discrédance locale étant définie ci-dessus. [13], [28]

$$D_{\infty,*}(X) = \sup_{t \in [0,1]^s} |disc(t)|.$$

Le qualificatif « à l'origine » signifie que les pavés sont ancrés en 0, en dimension  $s = 1$ , cette discrédance est facile à calculer, de même, il existe des résultats sur le calcul explicite de la discrédance infinie à l'origine en dimension  $s = 2$ . [23]

En dimension  $s > 2$ , il a été montré que le problème était discrétisable et pouvait être résolu en un nombre fini d'étapes, cependant, la complexité des algorithmes croît de manière exponentielle avec la dimension et ils sont donc inutilisables en dimension élevée, c'est pourquoi en pratique on se contente de bornes. [13]

Le théorème suivant donne une borne inférieure pour la discrédance classique à l'origine. [19]

**Théorème :** Il existe une constante  $B_s > 0$  ne dépendant que de la dimension  $s$ , telle que pour toute suite  $X = (x_1, \dots, x_n)$  dans  $[0,1]^s$ , on a :

$$D_{\infty,*}(X) \geq B_s \frac{(\log(n))^{s/2}}{n}$$

Pour une infinité de valeurs de  $n$ .

Le théorème ci-dessus montre que les suites à faible discrédance, c'est-à-dire dont la discrédance est en  $O\left(\frac{(\log n)^s}{n}\right)$ , sont proches de l'optimalité, de plus, la discrédance d'une suite aléatoire uniforme est en  $O\left(\frac{1}{\sqrt{n}}\right)$ , par conséquent les suites à discrédance faible permettent de gagner approximativement, un facteur  $\sqrt{n}$  au dénominateur.

### 2.1.3 La discrédance extrême

**Définition :** Pour tout  $X \in P' = [t^{1'}, t^1[ \times \dots \times [t^{s'}, t^s[$ , où  $(t^j, t^{j'}) \in [0,1]^2$ , et  $j = [1, \dots, s]$ .

Nous appelons une discrédance extrême la formule suivante :

$$D_{\infty}(X) = \sup_{t \in [0,1]^s} |disc'(t)|.$$

avec

$$disc'(t) = \frac{A(t, x)}{n} - \prod_{j=1}^n (t^j - t^{j'})$$

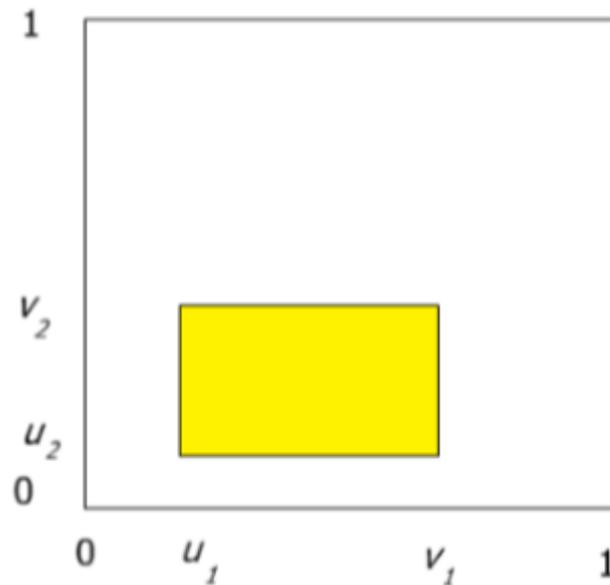


Figure 2-2 : mantant le pavé  $P'$ .

### 2.1.4 La discrédance en norme $L_2$ -étoile [34]

**Définition :** la discrédance en norme  $L_2$ -étoile, appelée aussi la discrédance carrée moyenne à l'origine  $D_{L_2,*}$  d'une suite finie de  $n$  points,  $X = (x_1, \dots, x_n)$  dans  $[0,1]^s$  est donnée par :

$$D_{L_2,*}(x) = \left[ \int_{[0,1]^s} disc(t)^2 dt \right]^{\frac{1}{2}}$$

On a clairement  $0 < D_{L_2,*}(x) \leq D_{\infty,*}(X) \leq 1$ .

**Remarque :** la discrédance carrée moyenne à l'origine  $D_{L_2,*}$  d'une suite finie de  $n$  points,  $X = (x_1, \dots, x_n)$  dans  $[0,1]^s$  peut-être reformuler de la manière suivante :

$$D^2_{L_2,*}(x) = \left(\frac{1}{3}\right)^s - \frac{2}{n} \sum_{i=1}^n \prod_{j=1}^s \left(\frac{1}{2} - \frac{(x_i^j)^2}{2}\right) + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^s [1 - \max(x_i^j, x_k^j)].$$

Cette formulation permet d'implémenter facilement le calcul de la discrédance carrée moyenne.

Toutefois, Matousek (voir [16]) met en évidence l'incapacité de la discrédance en norme  $L_2$ -étoile à quantifier la non-uniformité d'une suite de points en grande dimension lorsque le nombre de points est restreint.

Plus précisément, on peut montrer que dans le cas pathologique extrême de la suite de  $n$  points constituée de  $n$  répliques du point  $(1, \dots, n)$ , si  $n$  n'est pas suffisamment grand pour la dimension  $s$ .

La discrédance en norme  $L_2$ -étoile de cette suite est proche de la borne inférieure de la discrédance en norme  $L_2$ -étoile de  $n$  points quelconques en dimension  $s$ .

Le concept de discrédance peut se généraliser à n'importe quelle norme, mais à l'heure actuelle, on ne dispose de formules explicites que pour la discrédance carrée moyenne.

Cependant, cette discrédance, comme la discrédance infinie à l'origine, ne tient pas compte d'éventuelles spécificités de la fonction à intégrer.

En particulier, la plupart des suites à discrédance faible, que nous avons vues au cours du chapitre 01, présentent des irrégularités de répartition lorsqu'on les projette de l'espace  $[0,1]^s$  sur  $[0,1]^d$  avec  $1 < d \ll s$ , or toutes les variables et interactions n'ont pas toujours le même poids pour la fonction, on imagine alors aisément que si une projection mal distribuée correspond à une interaction importante, le résultat de l'estimation de l'intégrale risque d'être mauvais.

Par conséquent, l'écart à l'uniformité dans l'hyper cube  $[0,1]^s$  n'est pas forcément la mesure la plus adaptée au regard de l'intégrale.

### 2.1.5 La Discrépance $L_2$ modifiée

**Définition :** La notion de la discrèpance modifiée est fonder sur le calcul de  $D_{L_2,*}$  dans une dimension inferieure.

$$D_{L_2}(x) = \left[ \sum_{\emptyset \neq \mathcal{U} \subseteq \{1, \dots, s\}} D_{L_2,*}(x^{\mathcal{U}}) \right]^{\frac{1}{2}}$$

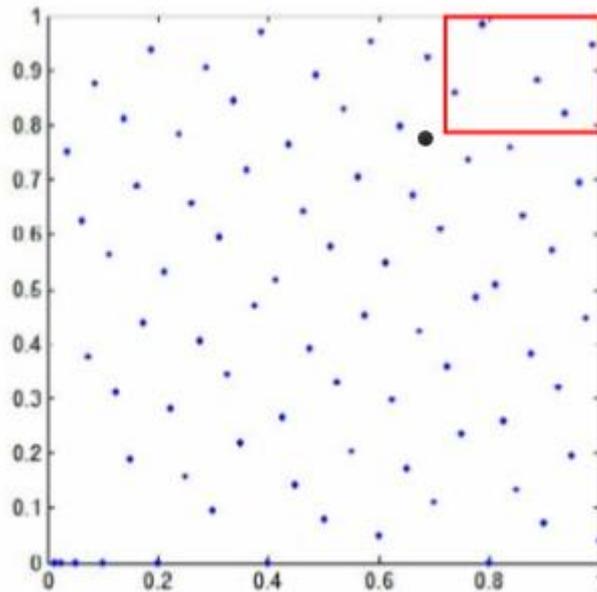
$x^{\mathcal{U}}$  est la projection de  $X$  sur le sous-espace des variables indicées dans  $\mathcal{U}$ .

### 2.2.6 La discrèpance centrée $D_{L_2,C}(x)$

**Définition :**  $x^{\mathcal{U}}$  est la projection du sous-ensemble construit à partir du point considéré dans  $X$  et de son sommet le plus proche.

$$D_{L_2,C}(x) = \left(\frac{13}{12}\right)^s - \frac{2}{n} \sum_{i=1}^n \prod_{k=1}^s \left(1 + \frac{1}{2} \left|x_k^{(i)} - \frac{1}{2}\right| - \frac{1}{2} \left|x_k^{(i)} - \frac{1}{2}\right|^2\right) + \frac{1}{n^2} \sum_{i,j=1}^n \prod_{k=1}^s \left(1 + \frac{1}{2} \left|x_k^{(i)} - \frac{1}{2}\right| + \frac{1}{2} \left|x_k^{(j)} - \frac{1}{2}\right| - \frac{1}{2} \left|x_k^{(i)} - x_k^{(j)}\right|\right).$$

**Remarque :** en dimension 2, l'ensemble  $x^{\mathcal{U}}$  peut-être de 4 formes différentes dont un est représenté sur la Figure 2-3 , Plus généralement, en dimension d , il existe  $2^d$  cas.



**Figure 2-3 :** Un sous-ensemble  $X$  pour le calcul de la discrèpance centrée.

### 2.1.7 La discrédance en norme $L_2$ -étoile pondérée $D_{L_2,*,\gamma}$

**Définition:** pour toute suite  $X = (x_1, \dots, x_n)$  de  $n$  points dans  $[0,1]^s$ , la discrédance en norme  $L_2$ -étoile pondérée  $D_{L_2,*,\gamma}$  est définie par :

$$D_{L_2,*,\gamma}(x) = \left[ \sum_{\emptyset \neq \mathcal{U} \subseteq \{1, \dots, s\}} \gamma_{\mathcal{U}} D_{L_2,*}(x^{\mathcal{U}}) \right]^{\frac{1}{2}}$$

Où  $\gamma = (\gamma_{\mathcal{U}})_{\emptyset \neq \mathcal{U} \subseteq \{1, \dots, s\}}$  est l'ensemble des poids réels positifs et  $x^{\mathcal{U}}$  est la projection de  $X$  sur le sous-espace des variables indicées dans  $\mathcal{U}$ .

La formule de la discrédance carrée moyenne  $D_{L_2,*}$  permet de calculer facilement la discrédance  $L_2$ -étoile pondérée.

Cependant, lorsque la dimension est grande, le nombre de termes dans la somme ci-dessus est très élevé et ceci peut rendre le calcul de  $D_{L_2,*,\gamma}$  relativement long.

Ainsi, une formule explicite de la discrédance  $L_2$ -étoile pondérée a été montrée lorsque les poids sont définis par :

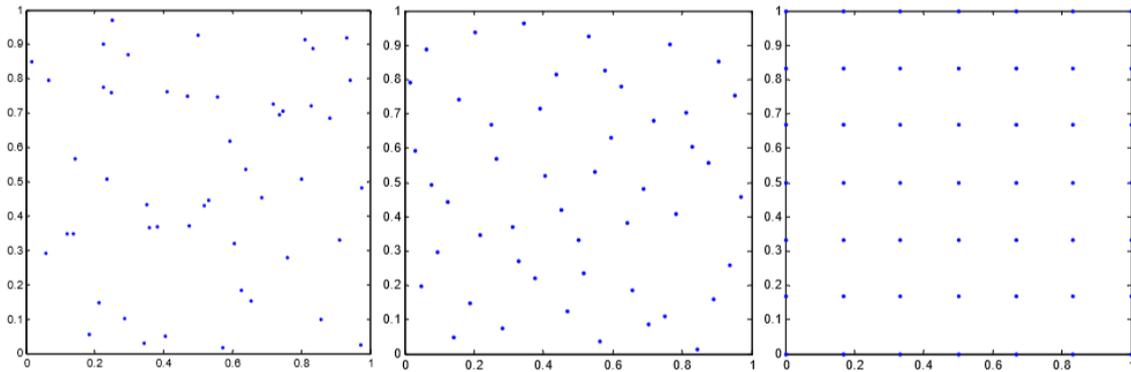
$$\begin{cases} 0 \leq \gamma_s \leq \dots \leq \gamma_1 \\ \gamma_{\mathcal{U}} = \prod_{j \in \mathcal{U}} \gamma_j. \end{cases}$$

A l'aide des poids produits définis ci-dessus La discrédance en norme  $L_2$ -étoile pondérée  $D_{L_2,*,\gamma}$  s'écrit de la manière suivante :

$$D_{L_2,*,\gamma}(x) = \prod_{j=1}^s \left( 1 + \frac{\gamma_j}{3} \right) - \frac{2}{n} \sum_{i=1}^n \prod_{j=1}^s \left( 1 + \frac{\gamma_j}{2} \left[ 1 - (x_i^j)^2 \right] \right) + \frac{1}{n^2} \sum_{i=1}^n \prod_{j=1}^s \left( 1 + \gamma_j \left[ 1 - x_i^j \right] \right) + \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{k=i+1}^n \prod_{j=1}^s \left( 1 + \gamma_j \left[ 1 - \max(x_i^j, x_k^j) \right] \right).$$

Le concept de discrédance  $L_2$ -étoile pondérée a été étendu à la norme  $L_p$ , et le même résultat a été obtenu par F. J. Hickernell (voir [12]) à travers la théorie des espaces à noyau reproduisant.

**Remarque :** quelle que soit la discrédance, plus la valeur calculée pour un plan donné est faible et plus les points du plan seront distribués uniformément.



**Figure 2-4 : De la gauche vers la droite. Un plan aléatoire, une suite de Halton, et un plan factoriel complet à 49 points en dimension 2.**

Le tableau ci-dessous montre que pour le plan factoriel, on obtient les discrédances les plus mauvaises, ceci implique que la discrédance seule ne suffit pas à mesurer la qualité d'un plan car nous ne souhaitons pas uniquement mesurer l'uniformité des points mais aussi le remplissage de l'espace.

A noter que cette remarque est valable pour n'importe quel nombre de points, ainsi, il est nécessaire de choisir plusieurs critères afin de mesurer la bonne répartition des points dans l'hypercube unité, ce que permettent entre autre les mesures basées sur les critères de distances présentés dans la section suivante.

	<b>Aléatoire</b>	<b>halton</b>	<b>factoriel</b>
<b>n=16</b>	0.26	0.17	0.44
<b>n=25</b>	0.21	0.14	0.36
<b>n=36</b>	0.26	0.1	0.31
<b>n=49</b>	0.12	0.08	0.27

**Tableau 2-1 : Discrédances en norme infinie pour n=16, 25, 36 et 49 pour les 3 types de plans suivants : aléatoires, suite de Halton, plan factoriel en dimension 2.**

### 2.1.8 Plans à faible discrédance [15]

La notion de plans à faible discrédance décrits ici est différente de celle des suites détaillée dans le chapitre précédent pour les raisons mentionnées ci-dessous.

Les plans à faible discrédance sont des plans stochastiques obtenus à l'aide d'un simple algorithme d'échange qui privilégie les configurations minimisant la discrédance.

Il existe de nombreux résultats de majoration de la discrédance, comme l'inégalité de Koksma-Hlawka, il existe aussi des majorations pour chacune des suites discutées précédemment, par exemple, Faure montre que la discrédance à l'origine d'une suite de Halton à  $n$  points en dimension  $d$  générée à partir des bases  $b_1, \dots, b_d$ . Est majorée par :

$$\frac{d}{n} + \frac{1}{n} \prod_{j=1}^d \frac{b_j - 1}{2 \log(b_j)} \log(n) + \frac{b_j + 1}{2} .$$

L'idée d'obtenir des plans à faible discrédance à l'aide d'un simple algorithme d'échange ne paraît pas être complètement irréaliste.

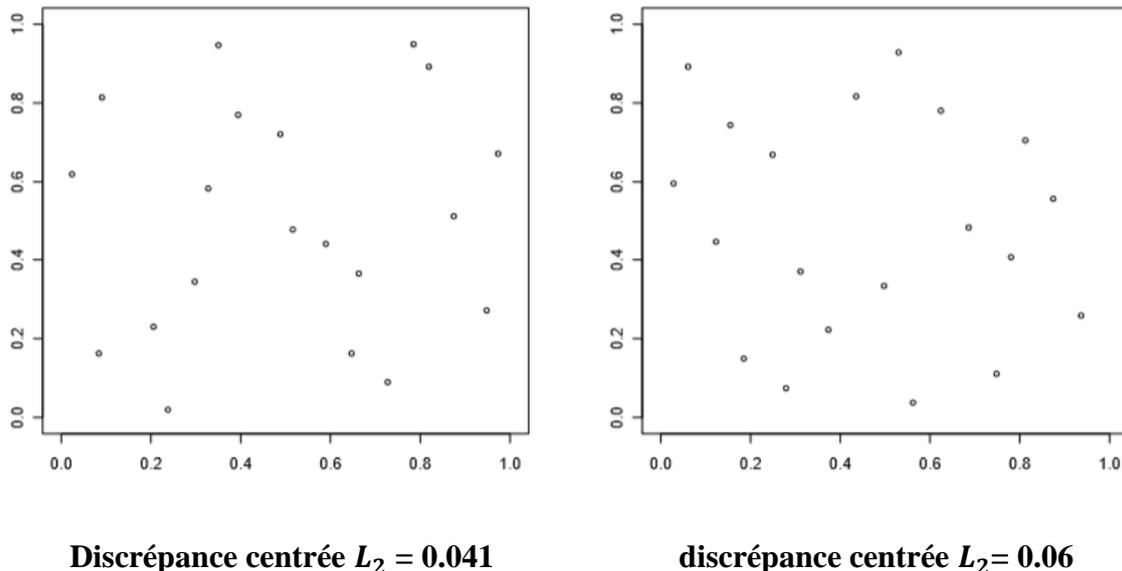
L'intérêt de discuter de ces plans à faible discrédance est double, d'une part, nous avons le souhait de montrer qu'il est possible d'obtenir des plans à plus faible discrédance que celle obtenue pour la majorité des suites présentées dans le chapitre précédent, d'autre part, ces plans non déterministes, par nature, ont le mérite de ne pas hériter des mauvaises propriétés (en grande dimension) de projections des suites causées par leur régularité.

La génération de ces plans nécessitant le calcul de la discrédance, ils sont construits uniquement à partir de la discrédance en norme  $L_2$ , beaucoup plus simple à calculer que la discrédance en norme  $L_\infty$ , il devient alors possible d'envisager des plans basés sur n'importe laquelle des discrédances définies au cours du chapitre ce qui signifie également la possibilité de construire des plans où l'on considère les points en projection sur les marges.

L'inconvénient de ces plans, outre le fait qu'ils ne peuvent être basés uniquement sur des discrédances en norme  $L_2$ , est leur temps d'exécution, en effet, même si le calcul des discrédances en norme  $L_2$  est obtenu à l'aide de formules analytiques simples, les suites à faible discrédance sont très rapides à mettre en Œuvre.

Par exemple, générer une suite de Halton à 600 points en dimension 60 est immédiat alors qu'un plan à faible discrédance de cet ordre de grandeur est inenvisageable.

Maintenant, vérifions sur quelques exemples, que ces plans permettent de réduire considérablement la discrédance annoncée des suites à faible discrédance, que l'exemple ci-dessous, nous avons généré un plan à 20 points en dimension 2 à l'aide de l'algorithme d'échange qui nous donne une discrédance de 0.041 alors que la discrédance pour la suite de Halton de même caractéristique est égale à 0.06.



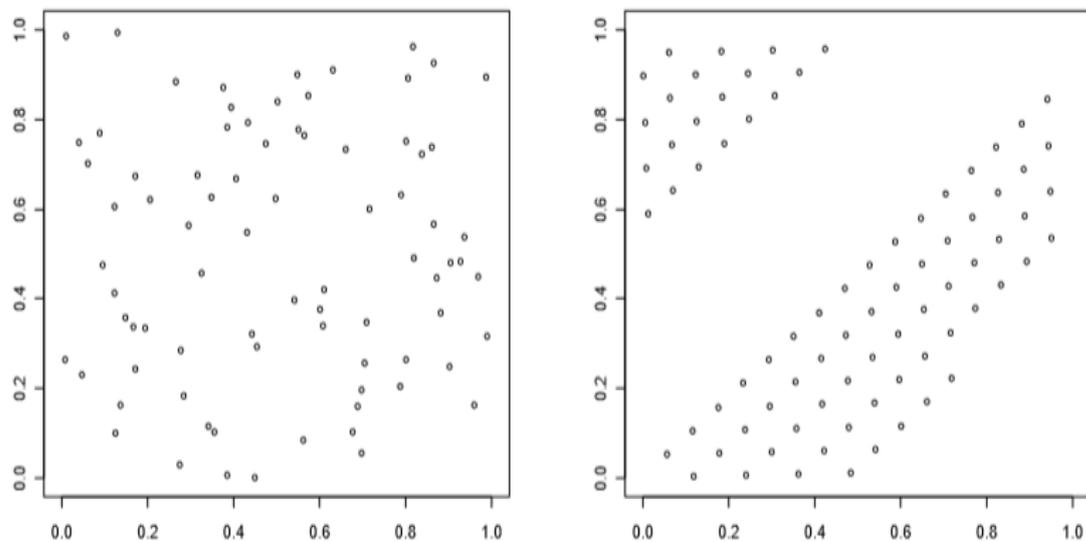
**Figure 2-5 : A gauche un plan à faible discrédance et à droite une suite de Halton à 20 points en 2D.**

Une autre possibilité est d'initialiser l'algorithme d'échange avec une suite à faible discrédance, au lieu de choisir une distribution aléatoire, afin de diminuer plus rapidement la discrédance.

Par exemple, si l'initialisation est faite à partir d'une suite de Halton, il est possible de diviser par 2 la discrédance en moitié moins d'itérations.

Regardons ce qu'il en est maintenant en grande dimension et plus particulièrement au regard des projections.

L'exemple ci-dessous montre les projections d'un plan à faible discrédance et d'une suite de Halton sur les facteurs  $(X_7, X_8)$ .



Discrédance centrée  $L_2 = 0.12$

discrédance centrée  $L_2 = 0.15$

**Figure 2-6 : Représentation des projections sur les facteurs  $(X_7, X_8)$  d'un plan à faible discrédance sur la figure de gauche et d'une suite de Halton sur la figure de droite à 80 points en 8D.**

A titre de comparaison, la discrédance centrée  $L_2$  d'une distribution aléatoire est en moyenne (sur 1000 plans à 80 points en dimension 8) égale à 0.22, nous avons donc l'assurance ici d'obtenir des distributions de meilleure qualité qu'un plan aléatoire.

Le caractère stochastique de ces plans est le défaut de leur qualité, en effet, ces plans permettent d'éviter les éventuels problèmes numériques que l'on peut avoir avec des plans à structure régulière.

Cependant, générer un plan en très grande dimension (par exemple en dimension 60) devient vite inenvisageable à cause du temps de calcul.

## 2.2 Critères de distance et plans optimaux [15]

L'objectif des critères basés sur la distance entre les points est de quantifier la proximité entre une distribution de points donnée et celle d'une grille régulière.

L'idée ici est de générer des plans dont les points sont proches d'une grille régulière mais pas exactement ceux d'une grille à cause des mauvais alignements que cela peut engendrer.

On rappelle que nous recherchons des plans avec des distributions quasi-périodiques c'est-à-dire qui sont un bon compromis entre une grille régulière et une bonne uniformité (dictée par exemple avec le critère de discrédance).

On note,  $dist(x^i, x^k)$  la distance euclidienne entre deux points  $x^i$  et  $x^k$  telle que :

$$dist(x^i, x^k) = \left[ \sum_{j=1}^s (x_j^i - x_j^k)^2 \right]^{\frac{1}{2}}$$

avec

$$x^i = (x_j^i)_{j \in \{1, \dots, s\}} \text{ Et } x^k = (x_j^k)_{j \in \{1, \dots, s\}}$$

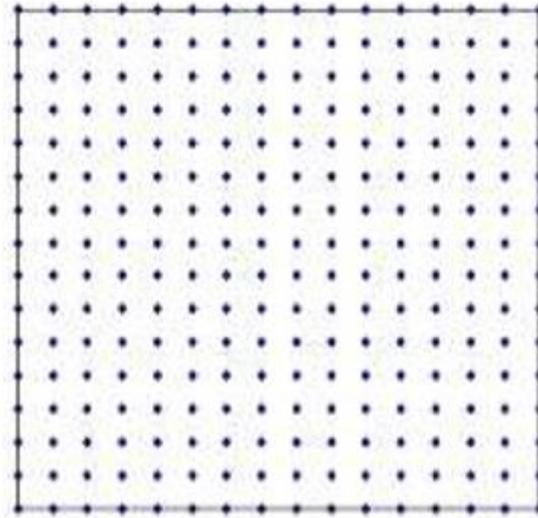


Figure 2-7 : les 256 points d'une grille régulière.

### 2.2.1 Mesure de recouvrement

**Définition :** Soit  $X = \{x^1, \dots, x^n\} \subset [0,1]^s$ , une suite de  $n$  points en «  $s$  » dimensions.

On définit la mesure de recouvrement  $\lambda$  par :

$$\lambda = \frac{1}{\bar{y}} \left( \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \right)^{\frac{1}{2}}$$

avec  $y_i = \min_{k \neq i} \text{dist}(x^i, x^k)$ , représentant la distance minimale entre le point  $x^i$  et les autres points de la suite  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ .

Si les points sont sur une grille régulière, on a  $y_1 = \dots = y_n = \bar{y}$ , donc  $\lambda = 0$ , ainsi plus  $\lambda$  est faible, et plus les points sont proches de ceux d'une grille régulière.

## 2.2.2 Le rapport de distances

**Définition :** Soit  $X = \{x^1, \dots, x^n\} \subset [0,1]^s$  une distribution de  $n$  points en «s» dimensions.

On définit le rapport de distance par :

$$R = \frac{\max_{i=1, \dots, n} y_i}{\min_{i=1, \dots, n} y_i}$$

avec  $y_i = \min_{k \neq i} \text{dist}(x^i, x^k)$ .

Dans le cas où les points sont sur une grille régulière, on a  $y_1 = \dots = y_n$ , donc  $R = 1$ , ainsi plus  $R$  est proche de 1, et plus les points sont proches de ceux d'une grille régulière.

## 2.2.3 Distances Maximin et Minimax

Johnson et al (1990) (voir[15]) ont introduit les distances Maximin et minimax afin de construire des plans répondant à la question de remplissage de l'espace.

**Définition (Chen, 2003) [6] :** On définit les critères suivants au sens de la distance euclidienne

$$\text{MinDist} = \min_{x_i \in X} \min_{x_k \in X, k \neq i} \text{dist}(x^i, x^k) = \min_{x_i \in X} y_i$$

et

$$\text{AvgDist} = \frac{1}{n} \sum_{x_i \in X} \min_{x_k \in X, k \neq i} \text{dist}(x^i, x^k) = \bar{y}$$

Où  $X = \{x^1, \dots, x^n\} \subset [0,1]^s$  est le plan d'expériences à  $n$  points en «s» dimensions.

Maximiser le critère «MinDist» revient à considérer le critère de distance «Maximin», si la valeur de MinDist est grande, alors il n'y a aucun point qui soit proche au sens de cette grandeur.

Ainsi, plus les critères MinDist et AvgDist seront grands et plus cela traduira une distribution uniforme.

**Définition :** on définit les critères suivants au sens de la distance euclidienne :

$$MaxDist(D, x) = \sup_{x_i \in D} \min_{x_k \in X} dist(x^i, x^k) = \sup_{x_i \in D} y_i$$

et

$$EDist(D, x) = \int_D \min_{x_k \in X} dist(x^i, x^k) dx^i$$

Où  $D$  est un ensemble de points du domaine expérimental et  $X = \{x^1, \dots, x^n\}$ , le plan d'expériences à  $n$  points en dimension  $s$ .

**Remarque :** l'ensemble  $D$  considéré dans la définition est généralement  $D = [0,1]^s$ , ainsi le supremum et l'intégrale ci-avant sont définis sur des ensembles infinis de points.

Cependant, afin de calculer ces 2 expressions simplement, nous considérons un échantillonnage fini de points dans  $D$ , à savoir un plan factoriel ou un échantillonnage de ce type.

Le principe du critère minimax est de mesurer la distance maximale entre un point du domaine expérimental n'appartenant pas au plan, qui varie, et les points du plan d'expérience.

Afin de calculer les critères, on prend généralement pour les points de  $D$  une grille régulière.

Aussi, en grande dimension il n'est plus envisageable de considérer ce critère sans prendre pour l'ensemble  $D$  un plan aléatoire, Le critère EDist correspond au critère de dispersion introduit par Niederreiter (1992). [28]

### Exemple

$n$	Aléatoire				Halton				Factoriel			
	$\lambda$	R	MinDist	MaxDist	$\lambda$	R	MinDist	MaxDist	$\lambda$	R	MinDist	MaxDist
16	0.5	7.56	0.049	0.43	0.17	1.81	0.16	0.35	$3 \cdot 10^{-8}$	1	0.33	0.23
25	0.43	4.03	0.049	0.32	0.17	1.80	0.12	0.27	0	1	0.25	0.17
36	0.68	19.8	0.012	0.29	0.19	2.06	0.097	0.18	$3 \cdot 10^{-8}$	1	0.2	0.14
49	0.58	11.14	0.016	0.19	0.22	2.81	0.053	0.17	$2 \cdot 10^{-6}$	1	0.17	0.11

**Tableau 2-2: Valeurs des critères de distance pour les 3 types de plans suivants : aléatoires, suite de Halton, plan factoriel où  $n = 16, 25, 36$  et  $49$ .**

### Conclusion :

Afin de répondre au problème du remplissage de l'espace, il est nécessaire de considérer les deux catégories de critères – discrédance et distance.

## 2.3 Critère d'entropie et plans à entropie maximale [15]

Ce critère est différent de ceux présentés précédemment dans le sens où il ne permet pas à proprement parler d'étudier l'uniformité et/ou le remplissage de l'espace d'un plan en phase exploratoire, en effet, le calcul de l'entropie n'est bien souvent envisageable que si la distribution sous-jacente est connue ; hypothèse que nous ne possédons généralement pas en phase exploratoire.

Le souhait de définir la notion d'entropie, ici, est motivé par la présentation d'une méthode de plans optimaux au sens de ce critère ,en effet, même si le critère n'est pas intrinsèque, les plans générés présentent de bonnes propriétés au regard du remplissage de l'espace, de plus, ce critère permet de prendre en compte le caractère anisotropique des variables qui lui peut être pressenties dès la phase exploratoire grâce à la connaissance du phénomène physique.

### 2.3.1 Définition de l'entropie

Historiquement, La notion d'entropie a été introduite officiellement en thermodynamique au milieu du *XIX*<sup>ème</sup> siècle par Clausius pour compléter le principe de conservation de l'énergie.

Boltzmann a été le premier à formaliser de manière probabiliste l'entropie qui fut utilisée par Shannon en 1948 [31] dans le contexte de la théorie de l'information, celui-là même a défini, dans un cadre probabiliste, une notion d'entropie, qui a des interprétations diverses en informatique, en thermodynamique et en probabilités, pourtant toutes liées par des intuitions communes.

Shewry et Wynn (1987) (voir [32]), l'ont décrit comme étant « la quantité d'information dans une expérience », plus généralement, l'entropie mesure la quantité d'information contenue dans une distribution.

Voilà dans le cadre probabiliste, la définition qui a été retenue

**Définition :** l'entropie d'une variable aléatoire continue  $X$  à densité de probabilité  $f$  est définie par :

$$H(X) = - \int_{x \in R} f(x) \log(f(x)) dx = -E_X(\log(f(x)))$$

Avec la convention  $0, \ln(0) = 0$ .

Pour une étude mathématique simplifiée, on choisit le logarithme népérien, ceci n'ayant aucune influence sur les résultats énoncés car l'entropie est alors translatée d'une constante.

Par ailleurs, l'entropie d'un vecteur aléatoire  $X = (X_1, \dots, X_n)$  à valeurs dans  $R^s$  et de même densité  $f$  se définit de façon analogue à la dimension un :

$$H(X) = - \int_{x \in R^s} f(x) \ln(f(x)) d\mu(x)$$

Où  $f$  est la densité de probabilité et  $\mu$  la mesure de Lebesgue.

**Remarque :** la valeur de l'entropie ne dépend pas des valeurs prises par  $X$ , mais uniquement de la densité de probabilité d'où l'impossibilité de la calculer directement à partir d'un plan d'expériences.

### 2.3.2 Entropie et quantité d'information

Shannon (1948) [31] décrit l'entropie comme une quantité d'information, afin de comprendre cette notion, nous allons considérer une variable aléatoire discrète  $X$  dont la distribution est donnée par  $(X = x_i) = p_i$ .

L'entropie de  $X$  est alors définie par :

$$H(X) = - \sum_{i=1}^n p_i \ln(p_i)$$

D'une part, on remarque que si toutes les valeurs de  $X$  sont distribuées de manière régulière (c'est-à-dire où  $p_1 = \dots = p_n = \frac{1}{n}$ ), alors  $X$  est la plus imprévisible possible et  $H(X) = - \sum_{i=1}^n \frac{1}{n} \ln\left(\frac{1}{n}\right) = \ln(n)$  atteint sa valeur maximale, cette propriété est connue comme le principe d'entropie maximale, si au contraire  $X$  prend une seule valeur avec la probabilité 1 alors  $X$  est complètement prévisible et l'entropie associée sera minimale avec  $H(X) = 0$ .

D'autre part, on note que l'entropie doit augmenter avec le nombre d'états possibles ce qui signifie que plus il y a de choix et plus l'incertitude est grande.

Cet exemple montre que la quantité d'information disponible sur la distribution peut se mesurer par  $I = -H$ , Shannon (1948) montre que cette quantité d'information  $I$  apportée par un paramètre  $\theta$  est la seule fonction ayant la propriété d'additivité à une constante près et la propriété de continuité.

### 2.3.3 Gain d'une expérience

Lindley (1956) (voir [24]) utilise la notion de quantité d'information au sens de Shannon afin de mesurer l'information apportée par une expérience.

Notons  $e$  une expérience d'un système aléatoire  $X$  de densité  $p(x|\theta)$  et  $\theta$  un paramètre.

Généralement  $\theta$  est le paramètre du modèle statistique que nous souhaitons estimer.

Selon le modèle choisi,  $\theta$  est défini différemment, par exemple, en régression,  $\theta$  peut être le vecteur des paramètres de la régression ; alors que si on considère un processus gaussien,  $\theta$  correspond au vecteur des différents paramètres (espérance, variance, corrélations).

Si la distribution a priori de  $\theta$  est notée  $\pi(\theta)$  et si  $x$  est la donnée obtenue après l'expérience  $e$  alors la densité a posteriori de  $\theta$  est proportionnelle à

$$\pi(\theta|x) \propto p(x|\theta)\pi(\theta).$$

L'information globale sur, avant  $e$  est

$$I(\theta) = \int \pi(\theta) \log(\pi(\theta)) d\theta$$

Et après  $e$ ,

$$I(\theta|x) = \int \pi(\theta|x) \log(\pi(\theta|x)) d\theta$$

L'accroissement d'information est donc donné par

$$I(\theta|x) - I(\theta)$$

Le gain d'information est défini en tenant compte de la loi conjointe de  $X$  et de  $\theta$ ,

$$g(e) = E[I(\theta|x) - I(\theta)] = E_x E_\theta \log \left( \frac{\pi(\theta|x)}{\pi(\theta)} \right)$$

### 2.3.4 Maximisation de l'entropie

Avec la définition ci-dessus, l'idée est de sélectionner une expérience  $e$  dans un ensemble  $E$  qui maximise le gain.

Shewry et Wynn (1987) [32] soulignent plusieurs problèmes liés à cette stratégie et notamment sur la définition de l'ensemble  $E$ , ils proposent de considérer  $E$  comme un ensemble fini d'expériences possibles et établissent ainsi le lien entre gain et l'entropie.

Considérons le domaine expérimental  $E$  comme discret et composé de  $N$  points auxquels on associe une réponse,  $Y_i$ ,  $i = 1, \dots, N$ .

Partitionnons le domaine expérimental en deux ensembles,  $D$  le plan d'expériences et  $\bar{D}$  les points restants, Alors, la décomposition classique de l'entropie s'écrit

$$H(Y_E) = H(Y_D) + E_{Y_D} [H(Y_{\bar{D}}|Y_D)]$$

L'espérance dans le membre de droite correspond à  $-g(e)$  dans le cas où  $Y_D$  joue le rôle de  $\theta$  et  $e$  au plan  $D$ .

On cherche donc à minimiser  $(Y_E) - H(Y_D)$ . L'entropie totale du système,  $H(Y_E)$ , étant fixe, cela revient à maximiser l'entropie aux points du plan  $H(Y_D)$ .

Construire un plan à entropie maximale a donc pour but de maximiser la quantité d'information apportée par les expérimentations relativement à un paramètre.

On trouve au sujet des plans à entropie maximale de nombreuses discussions, nous retiendrons principalement celles de Koelher et Owen (1996) [20] et Santner et al (2003) [29].

Le critère d'entropie a été, par la suite, appliqué aux plans d'expériences pour l'approximation d'un modèle numérique complexe déterministe par Mitchell et Scott (1987) [26] puis aussi par Currin et al (1991) [7] et dernièrement par Sebastiani et Wynn (2000) [30].

La section suivante propose une méthode simple, basée sur un algorithme d'échanges, pour générer des plans à entropie maximale où l'on présuppose un modèle sous-jacent.

### 2.3.5 Plans à entropie maximale

La définition générale des plans à entropie maximale nécessite d'avoir la réponse aux points du plan d'expériences, ce qui ne fait pas a priori de l'entropie un critère intrinsèque.

Shewry et Wynn (1987) [32] appliquent cette définition de façon à ne pas avoir à utiliser les réponses aux points du plan d'expériences pour construire celui-ci.

La méthode qui suit propose un remplissage de l'espace, en distribuant des points selon une matrice de corrélation spatiale, dans le cas particulier d'un processus gaussien centré  $f$ , Shewry et Wynn (1987) démontrent que  $H(Y_X)$  dépend simplement de  $\ln(\det(C(X)))$ , où  $C(X)$  dénote la matrice de covariance, ainsi, générer un plan à entropie maximale revient à, sous l'hypothèse de stationnarité, maximiser le déterminant de la matrice de corrélation.

**Remarque :** si le modèle est linéaire, alors le déterminant s'exprime en fonction de la matrice du plan, aussi, le plan obtenu à l'aide d'une méthode classique telle qu'un algorithme d'échange est, dans ce cas, D-optimal.

Cette équivalence existe uniquement si les réponses aux points suivent une loi normale multidimensionnelle, mais sans aucune condition sur la structure de covariance.

Soit  $X = (X^1, \dots, X^n)^t$  un vecteur de variables aléatoires, la matrice de variance – covariance de  $X$  s'écrit :

$$C(X) = \begin{bmatrix} \sigma_1^2 & \text{cov}(X^1, X^2) & \dots & \text{cov}(X^1, X^n) \\ & & \dots & \\ & & & \sigma_n^2 \end{bmatrix}$$

Où  $\sigma_i$  est l'écart-type de  $X^i$  ET  $\text{cov}(X^i, X^j) = \sigma_i \sigma_j \rho_{ij}$  est la covariance de  $X^i$  et  $X^j$ . Si les variables  $X^i$  sont réduites,  $C(X)$  s'identifie avec la matrice de corrélation :

$$C(X) = \sigma^2 \begin{bmatrix} 1 & \dots & \rho_{ij} \\ \dots & \dots & \\ \rho_{ij} & \dots & 1 \end{bmatrix}$$

Définissons à présent une matrice de corrélation spatiale  $C = [\rho_{ij}]$  avec :

$$\begin{cases} \rho_{ii} = 1 \\ \rho_{ij} = 1 - \gamma(h_{ij}) \\ \rho_{ij} = 0 \text{ si } h_{ij} > a \end{cases} \quad (1).$$

On reconnaît en  $\gamma(h)$  le variogramme où  $h_{ij}$  est la distance entre les points  $i$  et  $j$  et  $a$  la portée dudit variogramme.

On peut alors calculer pour  $X = (X^1, \dots, X^n)^t$ , vecteur de points dans un espace à  $s$  dimensions, une matrice de corrélation spatiale par :

$$C(X) = \begin{bmatrix} 1 & \dots & \rho_{ij} \\ \dots & \dots & \\ \rho_{ij} & \dots & 1 \end{bmatrix}$$

Où  $\rho_{ij}$  est une fonction de la distance entre les points  $i$  et  $j$ , calculée selon un modèle de corrélation spatiale que l'on attribue a priori à l'espace étudié (voir (1) ci-dessus).

Le déterminant de  $C(X)$  est maximum pour  $\rho_{ij} = 0$ , c'est-à-dire lorsque les points pris deux à deux sont éloignés l'un de l'autre d'une distance supérieure à la portée  $a$  de la fonction de corrélation spatiale.

On cherche donc à maximiser le déterminant de  $C(X)$  par un algorithme d'échange du type de Fedorov ou de Mitchell comme ci-après :

---

**Algorithme pour générer un plan à entropie maximale.**

---

**Initialisation** : choisir un vecteur de  $n$  points  $X_{(0)}$  au hasard et fixer un variogramme

**Calcul du déterminant de la matrice de corrélation spatiale associée** :  $\det(C(X_{(0)}))$

**POUR**  $k = 1$  à  $N_{\text{Max}}$

**Choisir une expérience**  $i$  au hasard dans  $\{1, \dots, n\}$

**Simuler**  $z^i$  uniformément dans  $[0, 1]^s$

**Accepter de remplacer**  $x^i$  par  $z^i$  si  $\det(C(X_{(k)})) > \det(C(X_{(k-1)}))$

**Fin pour**

## Chapitre 3

### Plans d'expérience numériques d'information de Kullback-Leibler minimale

Similairement à la discrédance, l'information de Kullback-Leibler (KL) (1951) (voir [22]) permet de mesurer l'écart entre la distribution empirique et la loi uniforme. L'idée est alors de construire, de façon empirique, des plans d'information KL minimale à l'aide d'un simple algorithme d'échange.

#### 3.1 Information de Kullback Leibler [1]

Supposons que les points du plan sont les  $n$  réalisations d'un vecteur aléatoire  $X = (X_1, \dots, X_s)$  de fonction de densité inconnue  $f$  de support le cube unité  $E = [0,1]^s$ .

L'objectif est de construire un plan dont la fonction de densité associée est la plus proche possible de celle de la loi uniforme sur  $E$ .

Pour cela, nous avons choisi d'utiliser l'information de Kullback-Leibler (1951) afin de mesurer l'écart entre les fonctions de densité.

**Définition :** l'information de Kullback-Leibler permettant de mesurer l'écart entre les deux fonctions de densité  $f$  et  $g$  est définie par

$$I(f, g) = \int f(x) \ln \left( \frac{f(x)}{g(x)} \right) dx$$

L'intégrale ci-dessus n'est pas toujours définie.

Une condition nécessaire pour que l'intégrale converge est que  $P_f$ , la mesure de probabilité sous-jacente à la fonction de densité  $f$ , est absolument continue par rapport à  $P_g$  la mesure de probabilité induite par  $g$ .

Dans notre cas, si on considère  $f$  continue et  $g$  la fonction de densité de la loi uniforme sur  $E$ , cela revient à supposer que pour tout  $A \subset R^s \setminus E$ ,  $P_f(A) = 0$ .

### 3.1.1 Quelques propriétés de l'information de Kullback-Leibler

Nous présentons ici les propriétés de l'information de Kullback-Leibler nécessaires à la bonne compréhension de nos propos.

- **Propriétés de distance.**

Même si l'information de Kullback-Leibler est utilisée pour mesurer un écart entre deux fonctions de densité, elle n'est pas une distance au sens topologique du terme.

En effet, l'inégalité triangulaire n'est pas vérifiée ainsi que la propriété de symétrie. Ceci explique par ailleurs l'importance du choix entre les fonctions  $f$  et  $g$ .

Par convention,  $f$  est choisie comme la fonction de densité associée au plan et  $g$  comme la fonction de densité théorique, à savoir ici celle de la loi uniforme.

Néanmoins, il existe des modifications pour symétriser l'information KL, par exemple en utilisant,

$$d(f, g) = I(f, g) + I(g, f)$$

On parle alors de divergence de Kullback-Leibler.

En revanche, l'information de Kullback-Leibler est définie positive  $I(f, g) \geq 0$  avec égalité si et seulement si  $f = g$  presque partout.

Cette propriété justifie en partie l'utilisation de l'information de Kullback-Leibler pour mesurer l'écart entre deux fonctions de densité.

Ainsi, plus l'information est proche de 0, et plus  $f$  est « proche » de  $g$ .

**Remarque 1 :** l'inégalité triangulaire exprime l'idée que la distance est une mesure minimale.

**Rappel :** soient  $f$  et  $g$  deux fonctions mesurables.

On dit que  $f$  et  $g$  sont égales presque partout si  $\{x \in \Omega \mid f(x) \neq g(x)\}$  est un ensemble négligeable avec  $\Omega$  un ouvert de  $R^S$ .

- **Formulation sous forme d'espérance**

En développant l'intégrale, on obtient

$$I(f, g) = \int f(x) \ln(f(x)) dx - \int f(x) \ln(g(x)) dx$$

Ce qui peut encore s'exprimer sous la forme d'une espérance relativement à la mesure  $P_f$ ,

$$I(f, g) = E_{P_f}[\ln(f(x))] - E_{P_f}[\ln(g(x))]$$

- **Propriété d'additivité**

L'information K-L est additive, c'est-à-dire que l'information entre deux lois conjointes de variables i.i.d. est la somme de l'information des lois marginales.

Si on suppose que les coordonnées des points du plan sont variables i.i.d. alors la propriété d'additivité de l'information KL revient à tester la distribution uniforme des points du plan sur les axes uniquement.

Or une répartition uniforme sur les marges n'entraîne pas nécessairement une répartition uniforme dans le cube unité.

**Rappel :** Soit  $(X, Y)$  un vecteur aléatoire réel, on appelle loi conjointe de  $(X, Y)$  la probabilité définie sur  $R^2$  par

$$P_{(X,Y)}(I * J) = P(X \in I \text{ et } Y \in J)$$

Les lois de probabilité de  $X$  et  $Y$  sont appelées lois marginales de  $(X, Y)$ .

- **Changement d'échelle**

On note de plus que l'information KL est invariante par changement d'échelle.

Cette propriété permet de travailler dans le cube unité pour construire des plans d'expériences génériques.

### 3.2 Maximisation de l'entropie [1], [11]

Le lien entre l'information de Kullback-Leibler et l'entropie se fait immédiatement en appliquant la définition à notre cas, c'est-à-dire en considérant  $g$  comme la fonction de densité de la loi uniforme sur  $E$ ,

$$I(f) = \int f(x) \ln(f(x)) dx = E_{P_f}[\ln(f(x))] = -H[f]$$

Où  $H[f]$  est l'entropie.

Ainsi, minimiser l'information KL revient à maximiser l'entropie.

**Rappel :** mesurer la quantité d'information contenue dans une distribution  $f$  donnée revient à calculer l'entropie de la variable aléatoire continue, qui est définie par :

$$H[X] = - \int f(x) \ln(f(x)) dx$$

L'originalité de l'idée proposée ici vient du fait que nous sommes en phase exploratoire et que nous n'avons aucun modèle sous-jacent.

La maximisation de l'entropie n'a donc pas pour objectif d'augmenter la quantité d'information (au sens de Shannon) contenue dans l'échantillon relativement à des paramètres du modèle.

En revanche, il est bien connu que la loi uniforme maximise l'entropie des lois à support dans  $[0,1]$ .

L'entropie du plan est donc négative et faire tendre cette entropie vers 0, revient à s'approcher d'une distribution uniforme.

On notera d'ailleurs que des tests d'uniformité ont été développés à partir de l'estimation de l'entropie mais uniquement dans le cas pour la dimension un et ne peuvent se généraliser en dimension supérieure.

Afin de ne pas confondre avec les plans à entropie maximale, les plans construits dans ce rapport seront dits d'information KL minimale.

**Remarque :** on note, dans ce cas, que la divergence de Kullback-Leibler s'écrit

$$\begin{aligned} d(f, g) &= I(f, g) + I(g, f) = \int f(x) \ln(f(x)) dx - \int \ln(f(x)) dx \\ &= \int f(x) [\ln(f(x)) - 1] dx = E_{P_f} [\ln(f(x)) - 1] \end{aligned}$$

### 3.3 Construction des plans

Au vu des propriétés énoncées précédemment, les plans sont construits suivant un simple algorithme d'échange visant à maximiser l'entropie.

---

**Algorithme d'échange**

---

**Initialisation** : choisir un plan d'expériences  $x_1, \dots, x_n$  selon une loi de probabilité donnée

**debut**

**TANT QUE** (condition d'arrêt n'est pas satisfaite)

Choisir une expérience  $x_i$  au hasard parmi les points du plan

Remplacer  $x_i$  par  $y_i$  simuler uniformément dans  $[0, 1]^s$

SI l'entropie obtenue à l'action précédente est plus grande que celle du plan d'origine

ALORS accepter le changement  $x_i = y_i$

**FIN TANT QUE**

**fin**

La condition d'arrêt n'est pas spécifiée.

Pratiquement, on considère soit un nombre maximum d'échanges, soit la stabilisation de l'entropie qui se traduit par un nombre arbitraire d'échanges consécutifs testés et refusés.

Suite à l'étude de la convergence de l'algorithme sur plusieurs essais, le nombre maximal d'échanges est fixé à  $1000 \times s$  et le nombre d'échanges consécutifs sans amélioration à  $100 \times s$ .

Le plan final dépend plus ou moins du plan initial choisi aléatoirement.

Afin de réduire cette dépendance, plusieurs initialisations sont testées et le meilleur plan est sélectionné.

Il est évident que dans l'algorithme ci-dessus le point essentiel concernant le calcul de l'entropie d'un ensemble de points reste à définir.

### 3.4 Estimation de l'entropie par Monte Carlo

#### 3.4.1 La méthode de Monte Carlo

**Rappel** : a l'aide des logiciels informatique come « R, Matlab » il est possible de simuler les  $n$  réalisations d'une variable aléatoire uniforme sur l'ordinateur.

Les chercheurs ont développé des techniques mathématiques basées sur la simulation uniforme afin de pouvoir simuler les  $n$  réalisations d'une variable aléatoire suivant une loi de probabilité quelconque.

La méthode Monte-Carlo est appliquée afin de mesurer une quantité du type

$$I = E[\varphi(X)] = \int \varphi(x) f(x),$$

Où  $\varphi : R^s \rightarrow R$  est une fonction donnée et  $X$  un vecteur aléatoire de densité  $f$  suivant laquelle on sait simuler.

Dans ce contexte, l'estimateur Monte-Carlo de base est défini par

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \varphi(X_i),$$

Où les  $X_i$  sont générées de façon i.i.d. selon  $f$ .

Si  $E[\varphi(X)] < \infty$ , alors d'après la Loi des grands nombres,  $\hat{I}_n \xrightarrow[n \rightarrow \infty]{p.s.} I$ .

Notons  $X_1, \dots, X_n$ , les  $n$  réalisations de  $X$  constituant les points du plan, et étant donné que l'entropie s'écrit sous la forme d'une espérance,

$$H[X] = - \int f(X) \ln(f(X)) dX = -E[\ln(f(X))]$$

La méthode de Monte Carlo fournit un estimateur sans biais et convergent de l'entropie,

$$\widehat{H[X]} = -\frac{1}{n} \sum_{i=1}^n \ln(f(X_i))$$

Cette estimation fait intervenir la fonction de densité  $f$  inconnue mais pouvant être estimée à partir de  $X_1, \dots, X_n$ .

La solution consistant à remplacer  $f$  par une estimation dans l'expression ci-dessus, à l'aide d'une méthode à noyaux.

### 3.4.1.1 Méthode à noyaux

#### Principe de la méthode à noyaux

L'originalité de la méthode vient du fait que  $f(x)$  peut s'écrire de la forme suivante lorsque  $h$  devient de plus en plus petit :

$$f(x) \cong \frac{F(x+h) - F(x-h)}{2h}$$

D'où l'estimateur plug-in (non naïf)

$$\begin{aligned} \hat{f}_{n,h}(x) &= \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{2h} 1_{\{X_i \in ]x-h, x+h]\}} = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \Phi_0\left(\frac{X_i - x}{h}\right) \end{aligned}$$

Où  $\Phi_0(u) = 1_{\{u \in ]-1, 1]\}}(u)/2$  est le noyau de Rosenblatt (1956).

Ce principe ci-dessus peut être généralisé dans le cas multidimensionnel, ce que résumons les étapes suivantes :

- 1) Choisir un type de noyau  $\Phi$  et un paramètre (de lissage)  $h$ .
- 2) Centrer un noyau  $\Phi_h$  sur chaque observation  $X_i$  appartenant au  $[0,1]^s$ .
- 3) Pour tout  $x \in [X_j - h_j, X_j + h_j]^s$ ,  $j = 1, \dots, s$ , La densité en un point  $x$  est estimée par  $\widehat{f}(x) = \frac{1}{nh^s} \sum_{i=1}^n \Phi_h\left(\frac{x - X_i}{h}\right)$ .

Nous détaillerons dans ce qui suit le bon choix du noyau et le paramètre  $h$ .

#### Taille de la fenêtre ou paramètre de lissage $h$ .

La taille de la fenêtre a une grande influence sur la qualité de l'estimation.

En supposant que les variables sont non corrélées, il est assez usuel d'estimer la taille de la fenêtre par (Règle de Scott, 1992).

$$\hat{h}_j = n^{-1/(s+4)} \hat{\sigma}_j, j = 1, \dots, s.$$

Où  $\hat{\sigma}_j$  est l'écart-type de la  $j$ -ème variable aléatoire  $X_j$ .

Cependant, Joe (1989) montre que dans le cas où  $f$  est estimée par noyaux, l'estimateur de l'entropie par la méthode de Monte Carlo est biaisé.

Le biais dépend naturellement de la taille de l'échantillon  $n$ , et aussi de la dimension, mais aussi de la taille de la fenêtre  $h$ .

Lors de la construction d'un plan optimal, la taille  $n$  et la dimension  $s$  sont fixées.

Il reste donc à fixer la taille de la fenêtre de façon à ce que le biais ne varie pas au cours de l'algorithme d'échange.

Pour cela, nous remplaçons  $\hat{\sigma}_j$  par l'écart-type de la loi uniforme sur  $[0,1]$ , d'où

$$\hat{h} = \frac{n^{-1/(s+4)}}{\sqrt{12}}$$

### Choix du noyau :

Il est connu que la forme du noyau a peu d'influence sur l'estimation, nous évoquons ci-dessous, différents noyaux qui ont été testés.

- **Noyau d'Epanechnikov :**

Le noyau d'Epanechnikov multidimensionnel est défini par

$$\Phi(z) = \begin{cases} \alpha(1 - \|z\|^2) & \text{si } \|z\| \leq 1 \\ 0 & \text{sinon} \end{cases}$$

Avec  $\alpha$  la constante de normalisation, et  $\| \cdot \|$  représente la norme euclidienne.

- **Noyau gaussien :**

$$\Phi(z) = \frac{(2\pi)^{-\frac{s}{2}}}{\sigma^s} \exp\left[-\frac{1}{2\sigma^2}\|z\|^2\right]$$

Où  $\sigma = \sqrt{\frac{s}{12}}$ , avec  $s$  désigne la dimension et,  $\frac{1}{12}$  désigne la variance d'une variable aléatoire uniforme sur  $[0,1]$ .

La forme du noyau n'a effectivement que peu d'influence sur le résultat.

Le noyau d'Epanechnikov est à support borné et convient ainsi à notre problématique. Cependant, cette caractéristique pose problème en dimension élevée car la probabilité que la norme de  $z$  soit inférieure à 1 devient très faible.

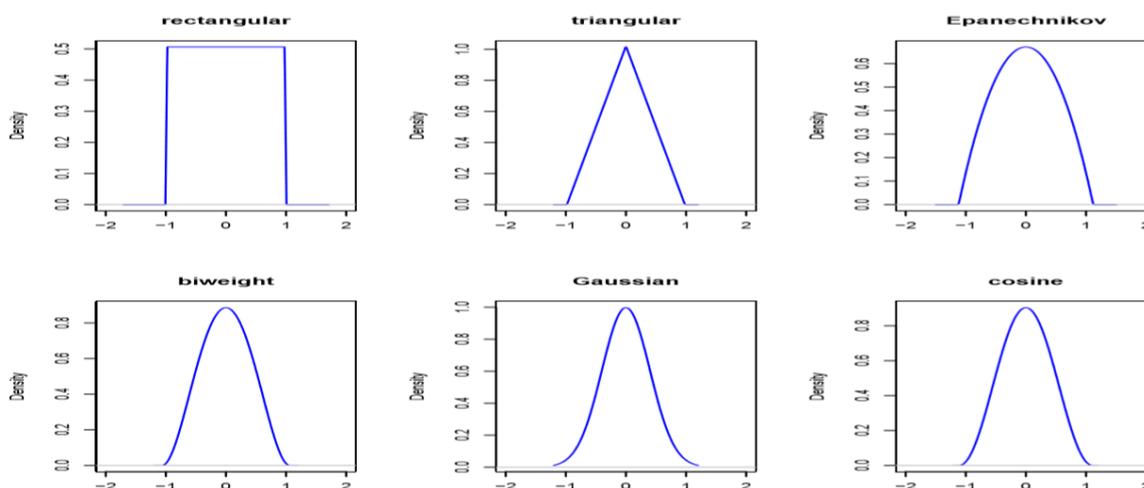
En effet, si on suppose que la taille de la fenêtre est fixée comme expliqué ci-dessus alors

$$P(\|z\| < 1) = \frac{n^{-2/(s+4)}}{12} = \hat{h}^2$$

Par exemple pour  $s = 3$  et  $n = 30$ , cette probabilité est de  $10^{-1}$ , et pour  $s = 10$  et  $n = 100$ , elle vaut  $4 \times 10^{-3}$ .

Cela signifie que le noyau est toujours nul et ainsi la fonction de densité estimée est identique quel que soit l'échantillon utilisé, ce qui n'apporte pas d'intérêt à l'utilisation de l'algorithme d'échange.

Etant à support infini, le noyau gaussien ne présente pas ce type de problème, et est finalement retenu pour l'estimation de la fonction de densité.

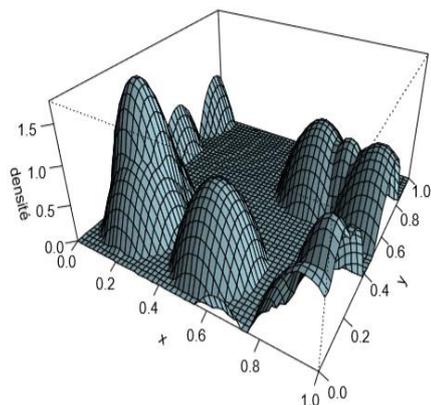


**Figure 3-1 : figure illustre quelques noyaux.**

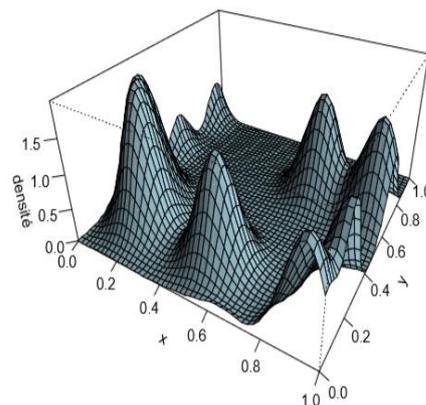
## Conclusion

Le choix du noyau gaussien, et une taille de fenêtre  $\hat{h}$  assure une bonne estimation de la fonction de densité  $f$ .

Les figures ci-dessous illustrent l'influence du choix des noyaux et la taille de fenêtre sur l'estimation de la densité  $f$ .

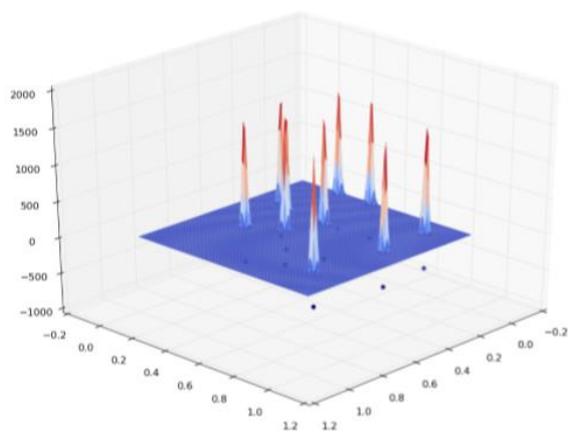


Noyau d'Epanechnikov

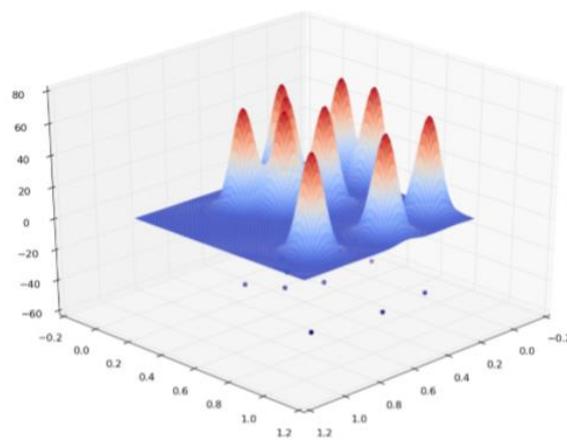


Noyau gaussien

**Figure 3-2 : Représentation d'une fonction de densité estimée sur un échantillon de taille 20 en dimension 2.**

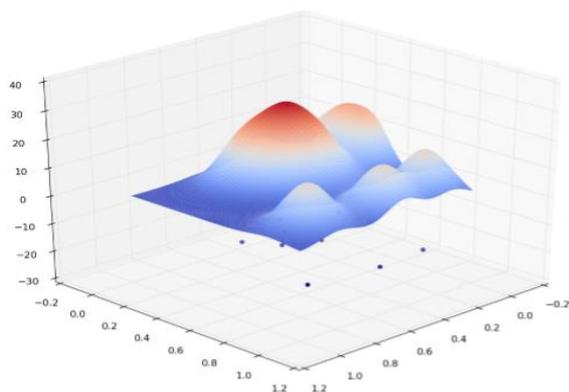


$h^2 = 0,01$

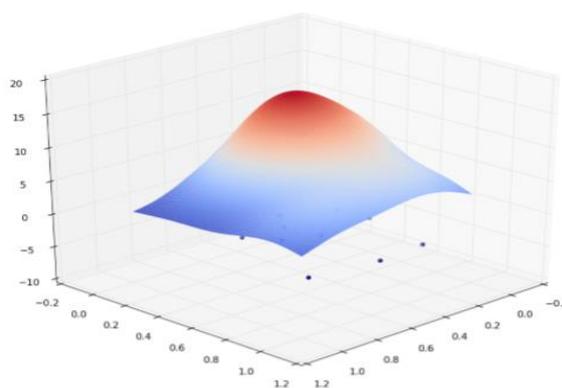


$h^2 = 0,05$

**Figure 3-3 : N = 10 observations dans  $\mathbb{R}^2$ , noyau gaussien de variance indiquée.**



$h^2 = 0,1$



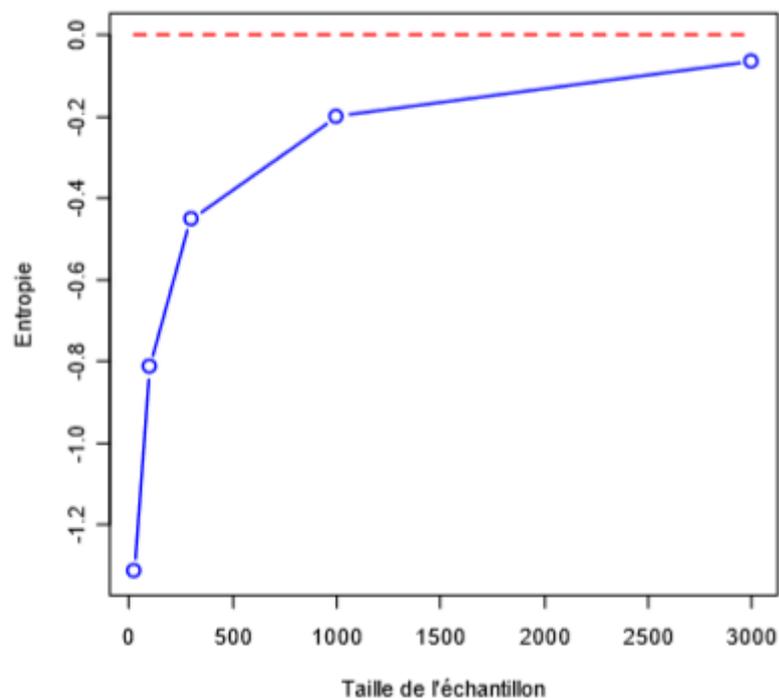
$h^2 = 0,2$

**Figure 3-4 : N = 10 observations dans  $\mathbb{R}^2$ , noyau gaussien de variance indiquée.**

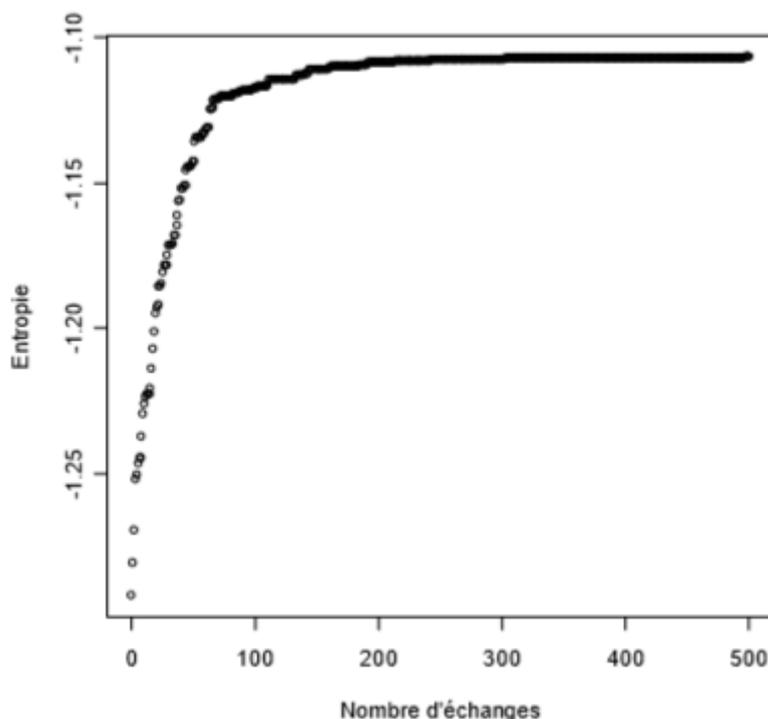
Les propriétés établies par Joe (1989) [17] concernant cet estimateur sont asymptotiques. De plus, il montre que la taille de l'échantillon augmente rapidement avec la dimension pour avoir une estimation correcte de l'entropie.

Dans le contexte des plans d'expériences, le nombre de points est faible et donc le calcul de l'entropie sera entaché d'une erreur importante, cependant, notre objectif n'est pas d'obtenir une valeur précise pour l'entropie.

Il est intéressant de voir que, malgré cette imprécision, l'algorithme d'échange converge assez rapidement (Figure 3-6) et que les plans ainsi construits ont les propriétés attendues, à savoir un remplissage uniforme de l'espace des paramètres.



**Figure 3-5 : Evolution de l'entropie (MCGauss) en fonction de la taille de l'échantillon pour des tirages aléatoires en dimension 3.**



**Figure 3-6 : Evolution de l'entropie (MCGauss) au cours de l'algorithme d'échange pour des plans de taille 30 en dimension 3.**

### 3.4.1.2 Estimation de l'entropie par plus proche voisin

Lorsque la dimension  $s$  augmente, la méthode d'estimation par Monte Carlo devient moins performante lorsque l'échantillon est de petite taille, comme cela est le cas pour un plan d'expériences numériques.

Etant donné que la densité doit être ré-estimée à chaque échange, cette méthode est trop coûteuse en temps de calcul.

Kozachenko et Leonenko (1987) (voir [21]), proposent une estimation de l'entropie directement basée sur la répartition des points sans passer par l'estimation de la fonction de densité,

$$\widehat{H}[X] = \frac{s}{n} \sum_{i=1}^n \ln(\rho_i) + \ln \left[ \frac{\pi^{s/2}}{\Gamma\left(\frac{s}{2} + 1\right)} \right] + C_E + \ln(n-1)$$

Où  $C_E \approx 0.5772$  est la constante d'Euler,  $\Gamma$  est la fonction Gamma et  $\rho_i$  est la distance euclidienne entre  $X_i$  et son plus proche voisin dans l'échantillon,

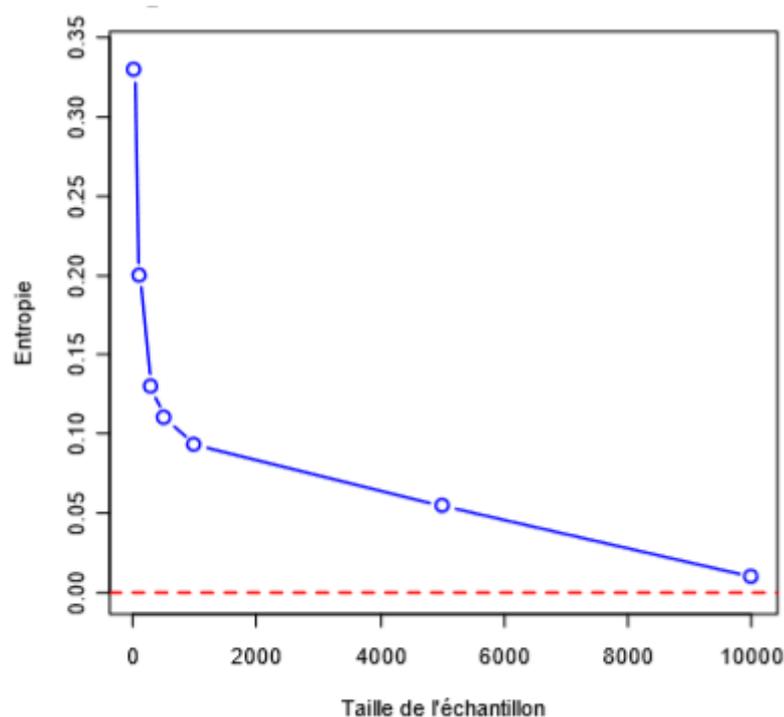
$$\rho_i = \min_{j \neq i, 1 \leq j \leq n} \|X_i - X_j\|$$

Ils montrent que cet estimateur est asymptotiquement sans biais et converge en moyenne quadratique pour des conditions faibles sur  $f$ .

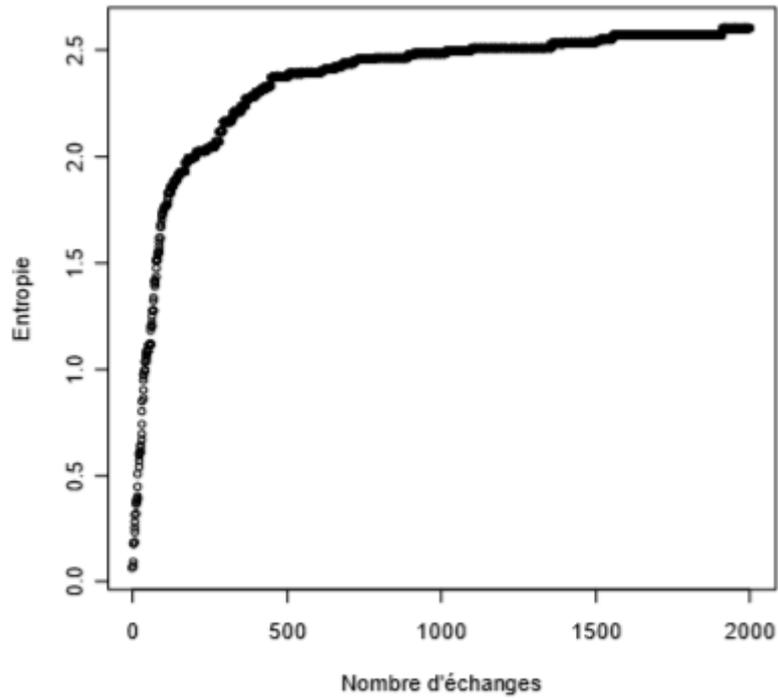
On constate effectivement sur la figure 3-7, cependant les valeurs sont positives alors qu'elles devraient être négatives en théorie, ce décalage doit certainement retranscrire le biais de l'estimation.

L'algorithme d'échange (figure 3-8) est lui aussi convergent mais semble plus lent que pour la méthode précédente.

Le temps de calcul gagné sur l'estimation de l'entropie se trouve en partie perdu par un nombre d'échanges plus important.



**Figure 3-7: Evolution de l'entropie (PPV) en fonction de la taille de l'échantillon pour des tirages aléatoires en dimension 3.**



**Figure 3-8 : Evolution de l'entropie (PPV) au cours de l'algorithme d'échange pour des plans de taille 30 en dimension 3.**

# Chapitre 4

## Implémentation, discussions et comparaison

Les expérimentations ont été réalisées sur un PC équipé d'un processeur Intel(R) Core (TM) i3-4005U CPU @ 1.70 GHz avec une mémoire de 4 GB sous Windows 7 Professionnel. Les algorithmes des différentes méthodes utilisées dans la comparaison sont implémentés en MATLAB. Les programmes prennent comme donnée d'entrée des matrices d'expériences contenant des points générés d'une manière aléatoire, selon une densité donnée (généralement, on utilise la loi uniforme).

### 4.1 Matlab comme outil de programmation

Matlab est un langage hautes performances pour le calcul scientifique et technique. Il intègre la possibilité de calculs, de visualisation et de programmation dans un environnement très simple d'emploi. Les résultats sont exprimés sous une forme mathématique standard. L'utilisation typique est :

- Calcul scientifique.
- Développement d'algorithmes.
- Acquisition de données.
- Modélisation et simulation.
- Analyse de données, exploration et visualisation.
- Graphisme scientifique.
- Développement d'applications, interface graphique.

Matlab est un système interactif dont la brique de base est un tableau dont la taille n'est pas nécessairement connue. Ceci permet de résoudre des problèmes, en particulier ceux qui ont une formulation matricielle, en un minimum de temps (contrairement aux langages de bas niveau comme le C ou le fortran). Matlab a été développé à l'origine pour avoir un accès simple et rapide aux projets EISPACK et LINPACK. Aujourd'hui, Matlab intègre les

librairies LAPACK et BLAS incorporant ainsi les dernières techniques pour le calcul matriciel.

Dans l'enseignement universitaire Matlab s'est imposé comme un standard pour l'apprentissage de l'algorithmique scientifique. Dans l'industrie, il est l'outil de choix pour une productivité accrue en recherche et développement. Matlab peut aussi être enrichi à l'aide de Toolbox (boîtes à outils) pour des problèmes spécifiques. Matlab fournit une documentation détaillée à la fois en ligne et sous forme de fichiers PDF.

## 4.2 Etude de comparaison avec les plans usuels

Avant de comparer les plans en dimension supérieure où égale à 2, il est important de faire appel à des critères usuels permettant de juger le bon remplissage de l'espace ainsi la distribution uniforme. (Pour plus de détail voir le chapitre 2)

- Mesure de recouvrement: permet de mesurer l'écart entre les points du plan et ceux d'une grille régulière. Ce critère est nul pour une grille régulière. L'objectif est donc de le minimiser pour se rapprocher d'une grille régulière, et ainsi assurer le remplissage de l'espace, sans toutefois l'atteindre pour respecter une distribution uniforme notamment en projection sur les axes factoriels.
- Critère de distance: Le critère maximin (Mindist) consiste à maximiser la distance minimale entre deux points du plan.
- Critère de discrédance : La discrédance permet de mesurer l'écart entre la fonction de répartition empirique des points du plan et celle de la loi uniforme. Contrairement aux deux critères précédents, la discrédance n'est pas basée sur la distance entre les points. Il existe différentes mesures de discrédance. Nous retenons la discrédance en norme L2.
- Le critère R : est le rapport entre la distance maximale et minimale entre deux point du plan.

Les plans proposés dans ce mémoire sont comparés avec les plans usuellement utilisés en expériences numériques:

- Plans aléatoires (RD).
- Hyper cubes latins (LHS).
- Plan maximal LHS (mLHS).
- Plans de Strauss (SD).
- Plans à entropie maximale (Dmax).

### 4.3 Résultats et discussions

Dans cette partie, nous présentons les tables regroupant les résultats d'implémentation des différents algorithmes pour les différents plans considérés. Ces tables sont divisées selon le nombre de points et la dimension de l'espace expérimental. Les plans proposés donnent incontestablement des bons résultats quel que soit le critère. Ils semblent les meilleurs en discrédance et en rapport, cependant il ne semble plus meilleurs pour les autres critères « mesure de recouvrement et mindist », ce que nous expliquons par l'influence de la dimension et le nombre de points dans le plan d'expériences.

#### 4.3.1 Résultat pour 10 points et 2 dimensions

	recouvrement	discrédance	distance	rapport
Aléatoire	0.5215326	0.03200207	0.07773079	5.553536
LHS	0.3359125	0.03009458	0.1340521	3.323186
mLHS	0.3161762	0.01432051	0.152288	3.11082
Sobol	0.1362939	0.003047299	0.1767767	1.118034
Halton	0.2810232	0.007570195	0.1607922	2.080255
Faure	0.5775323	0.005308041	0,08838835	6
Dmax	0.1547445	0.008652122	0.2925121	3.238632
Strauss	0.1642766	0.01307772	0.290778	4.458029
DKL	0.30306	0.007885	0.15154	1.67595375

**Table 4-1 : Listes des différents plans selon les quatre critères avec n=10, d=2.**

80 plans de chaque type seront générés afin de donner un sens aux résultats. Les figures ci-dessous représentent les boxplots obtenus à la suite de cette étude pour les plans :

- Plans aléatoires (RD).
- Hyper cubes latins (LHS).
- Plan maximal LHS (mLHS).
- Plans de Strauss (SD).
- Plans à entropie maximale (Dmax).

- Plans d'information KL minimale (DKL).

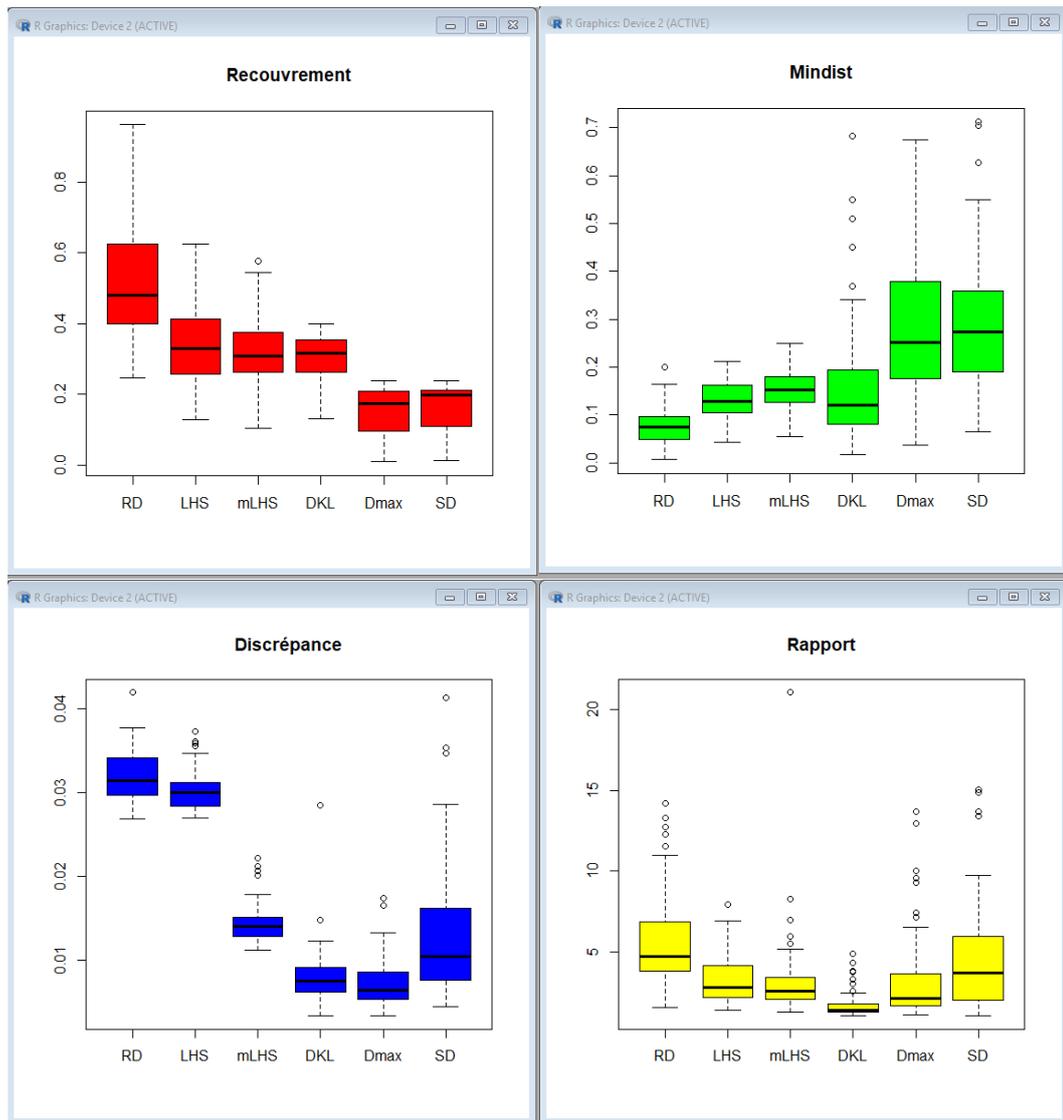


Figure 4-1 : boxplots des critères de qualité pour deux dimensions.

### 4.3.2 Résultat pour 20 points et 3 dimensions

	recouvrement	discrédance	distance	rapport
<b>Aléatoire</b>	0.4084524	0.01432051	0.09660525	6.229212
<b>LHS</b>	0.3229826	0.01133634	0.1353985	3.721318
<b>mLHS</b>	0.2592255	0.01072663	0.1744015	2.925884
<b>Sobol</b>	0.2864609	0.001382617	0.1623798	2.553139
<b>Halton</b>	0.2601684	0.001732852	0.1719613	2.562003
<b>Faure</b>	0.2082265	0.00113882	0.169725	2
<b>Dmax</b>	0.08613133	0.01254935	0.5148527	5.336851
<b>Strauss</b>	0.08336419	0.01235391	0.4957753	6.010683
<b>DKL</b>	0.3279325	0.00303625	0.526920892	1.80208875

**Table 4-2 : Listes des différents plans selon les quatre critères avec n=20, d=3.**

Les figures ci-dessous représentent les résultats de 80 plans générés pour chaque type de plans suivants :

- Plans aléatoires (RD).
- Hyper cubes latins (LHS).
- Plan maximal LHS (mLHS).
- Plans de Strauss (SD).
- Plans à entropie maximale (Dmax).
- Plans d'information KL minimale (DKL).

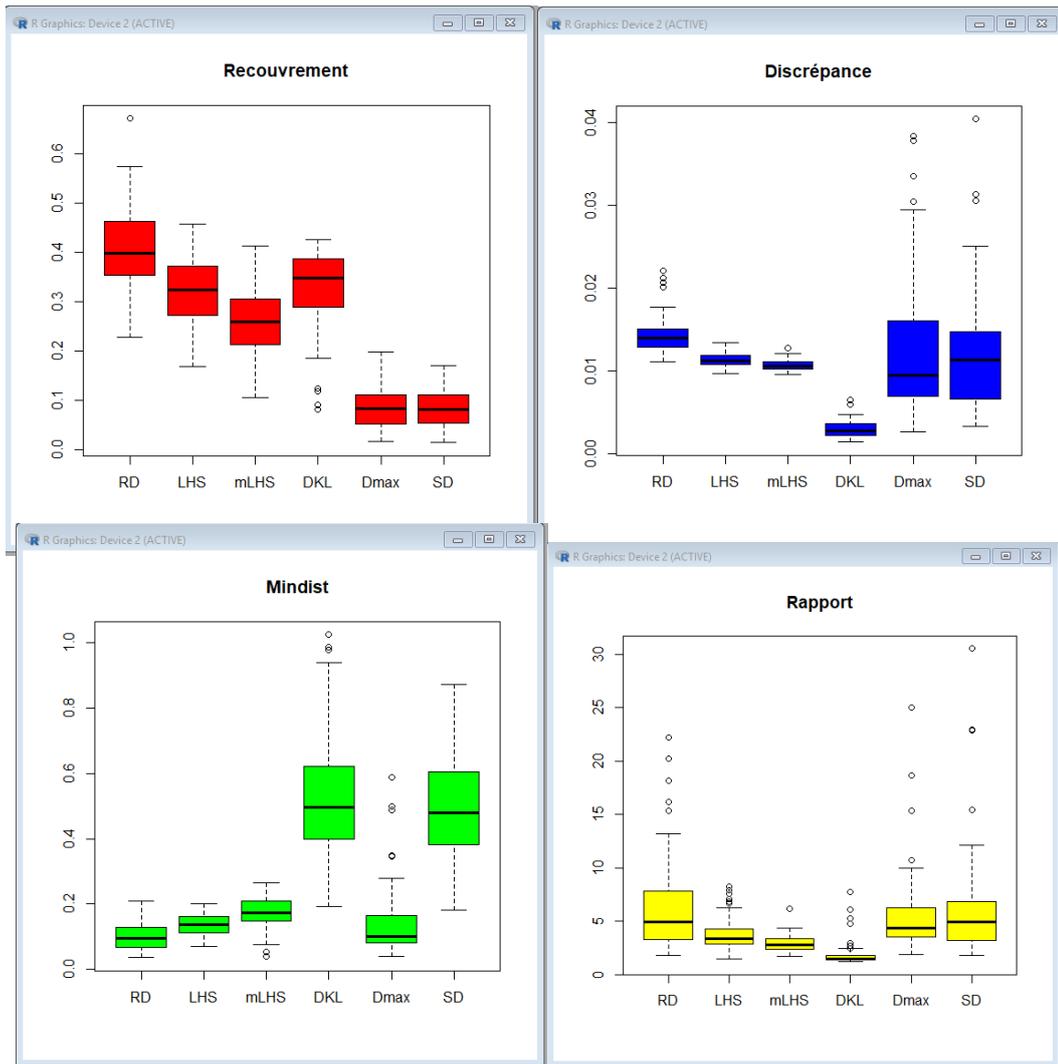


Figure 4-2 : boxplots des critères de qualité pour 3 dimensions.

### 4.3.3 Résultat pour 40 points et 5 dimensions

Les résultats sont regroupés dans le tableau ci-dessous, montrons une amélioration sur les valeurs des différents critères retenus.

	Recouvrement	discrépance	distance	rapport
<b>Aléatoire</b>	0.2779045	0.01130093	0.1872624	6.893661
<b>LHS</b>	0.2537125	0.01067706	0.2654198	2.527428
<b>mLHS</b>	0.2030364	0.001754879	0.2606106	2.690024
<b>Sobol</b>	0.2110594	0.000260329	0.2952257	2.066133
<b>Halton</b>	0.2378431	0.000400047	0.2406925	2.483516
<b>Faure</b>	0,1952901	0.00028458	0.3179937	2.008684
<b>Dmax</b>	0.08142545	0.001628364	0.5382086	6.025024
<b>Strauss</b>	0.07330626	0.009685785	0.5521754	6.219888
<b>DKL</b>	0.113516057	0.000513009	0.130810472	3.546064918

**Table 4-3 : Listes des différents plans selon les quatre critères avec n=40, d=5.**

Les figures ci-dessous représentent les résultats de 80 plans générés pour chaque type de plans suivants :

- Plans aléatoires (RD).
- Hyper cubes latins (LHS).
- Plan maximal LHS (mLHS).
- Plans de Strauss (SD).
- Plans à entropie maximale (Dmax).
- Plans d'information KL minimale (DKL).

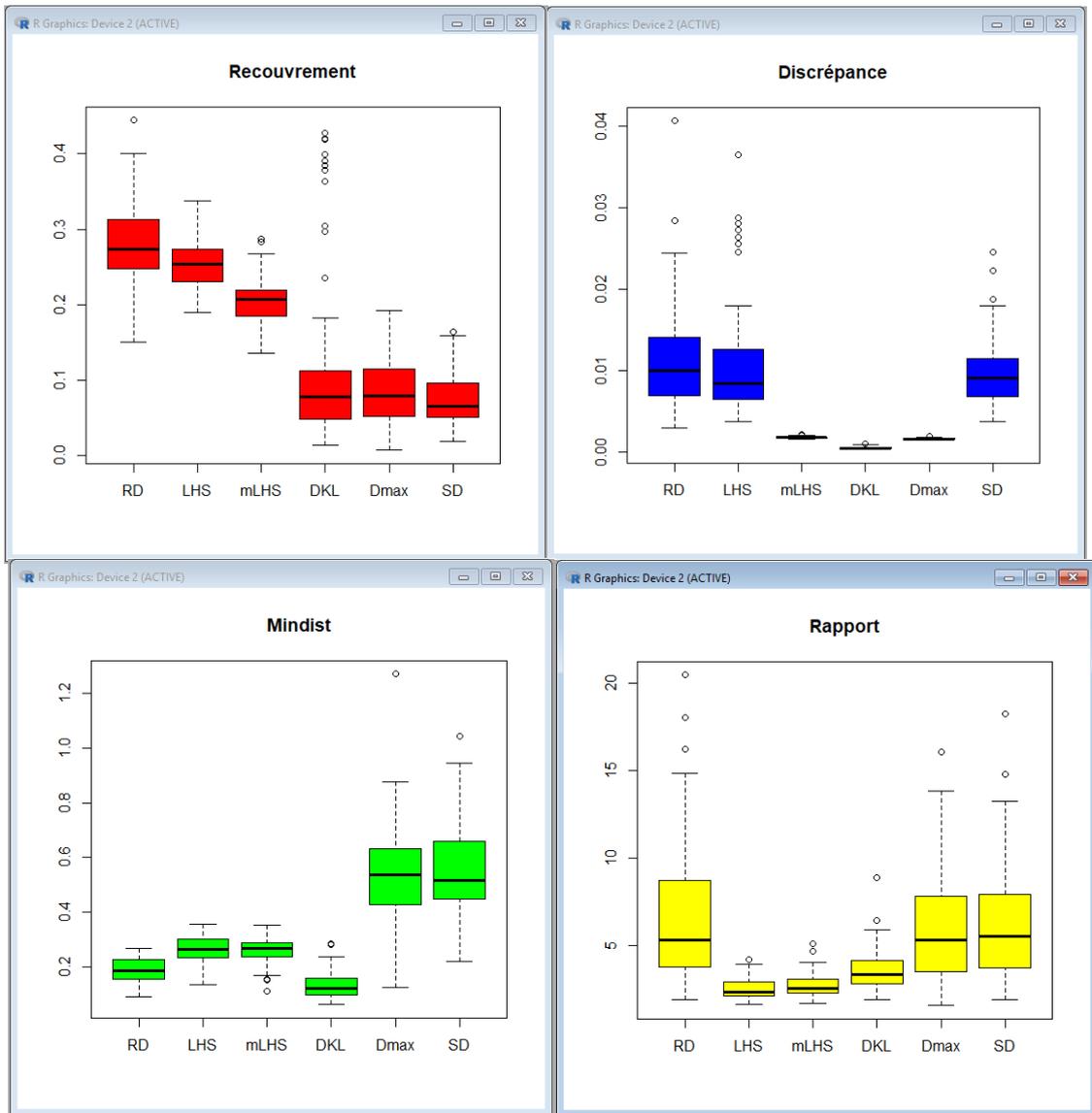


Figure 4-3 : boxplots des critères de qualité pour 5 dimensions.

#### 4.3.4 Résultat pour 50 points et 7 dimensions

Les résultats sont regroupés dans le tableau ci-dessous, montrons une amélioration sur les valeurs des différents critères retenus.

	recouvrement	discrépance	distance	rapport
<b>Aléatoire</b>	0.2132717	0.01130093	0.2936924	9.629222
<b>LHS</b>	0.1984752	0.01824426	0.3156578	2.571722
<b>mLHS</b>	0.1774884	0.0003156118	0.3628827	2.251442
<b>Sobol</b>	0.110397	0.00000743	0,5595626	2.250821
<b>Halton</b>	0.1349754	0.0000560	0.507455	1.828507
<b>Faure</b>	0.3664268	0.00013279	0.2874798	3.200336
<b>Dmax</b>	0.06925627	0.01462145	0.5151662	6.803585
<b>Strauss</b>	0.07542196	0.01183629	0.5071142	6.394307
<b>DKL</b>	0.091277764	0.00013411	0.437503693	1.988384876

**Table 4-4 : Listes des différents plans selon les quatre critères avec n=50, d=7.**

Les figures ci-dessous représentent les résultats de 80 plans générés pour chaque type de plans suivants:

- Plans aléatoires (RD).
- Hyper cubes latins (LHS).
- Plan maximal LHS (mLHS).
- Plans de Strauss (SD).
- Plans à entropie maximale (Dmax).
- Plans d'information KL minimale (DKL).

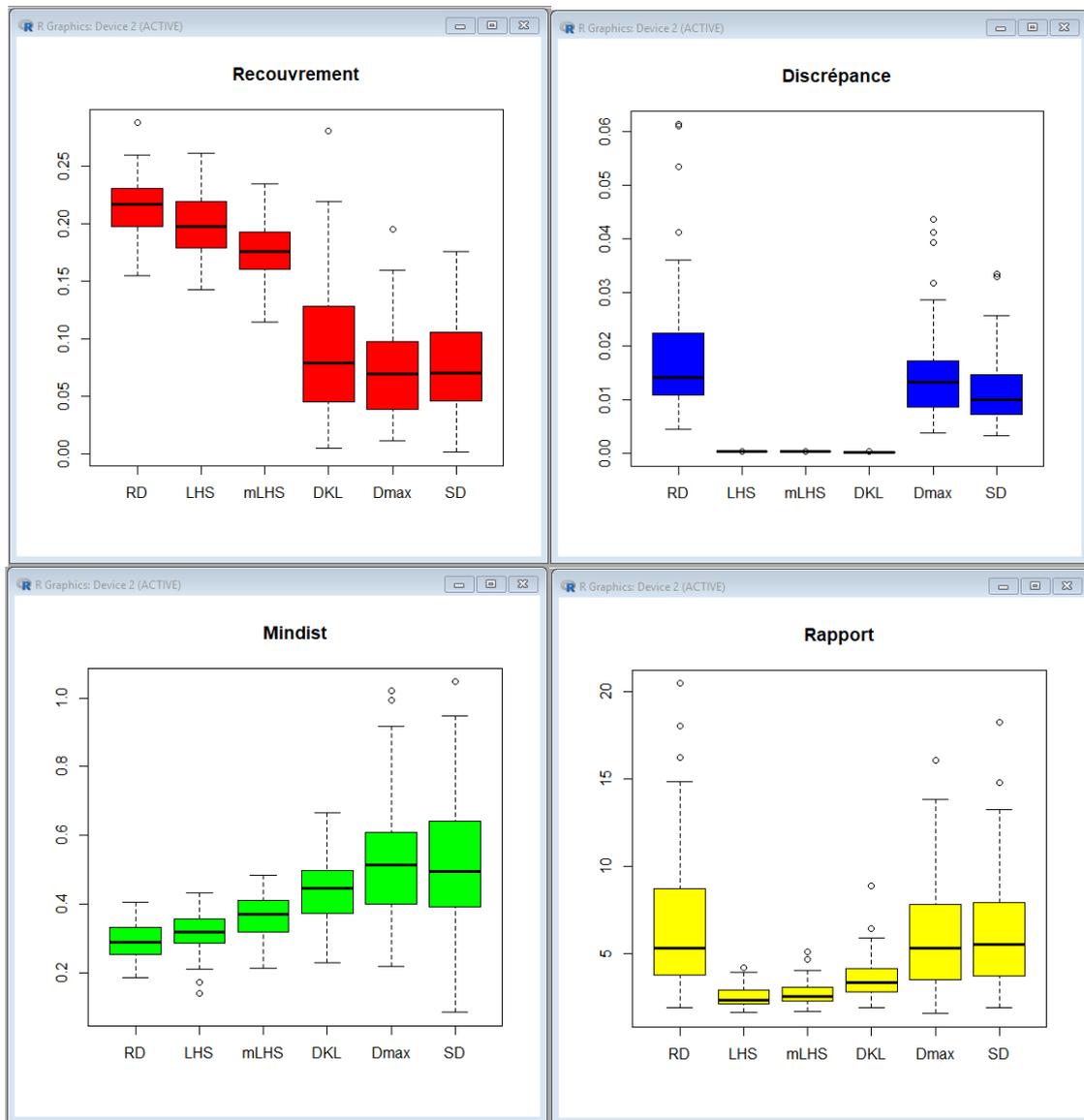


Figure 4-4 : boxplots des critères de qualité pour 7 dimensions.

### 4.3.5 Résultat pour 80 points et 10 dimensions

Les résultats sont regroupés dans le tableau ci-dessous, montrons une amélioration sur les valeurs des différents critères retenus.

	<b>recouvrement</b>	<b>discrédance</b>	<b>distance</b>	<b>rapport</b>
<b>Aléatoire</b>	0.1677978	0.01679324	0.4303464	6.000734
<b>LHS</b>	0.156489	0.000014	0.4518074	2.196265
<b>mLHS</b>	0.1570912	0.000023	0.4673578	2.216084
<b>Sobol</b>	0.110397	0.000007	0.5595626	1.766043
<b>Halton</b>	0.1349754	0.000056	0.507455	1.828507
<b>Faure</b>	0.3664268	0.00013279	0.2874798	3.200336
<b>Dmax</b>	0.07310013	0.01331964	0.5354068	4.958062
<b>Strauss</b>	0.07493931	0.01239033	0.521444	5.846509
<b>DKL</b>	0.123693511	1.08619E-05	0.553443977	1.855607891

**Table 4-5 : Listes des différents plans selon les quatre critères avec n=80, d=10.**

Les figures ci-dessous représentent les résultats de 80 plans générés pour chaque type de plans suivants:

- Plans aléatoires (RD).
- Hyper cubes latins (LHS).
- Plan maximal LHS (mLHS).
- Plans de Strauss (SD).
- Plans à entropie maximale (Dmax).
- Plans d'information KL minimale (DKL).

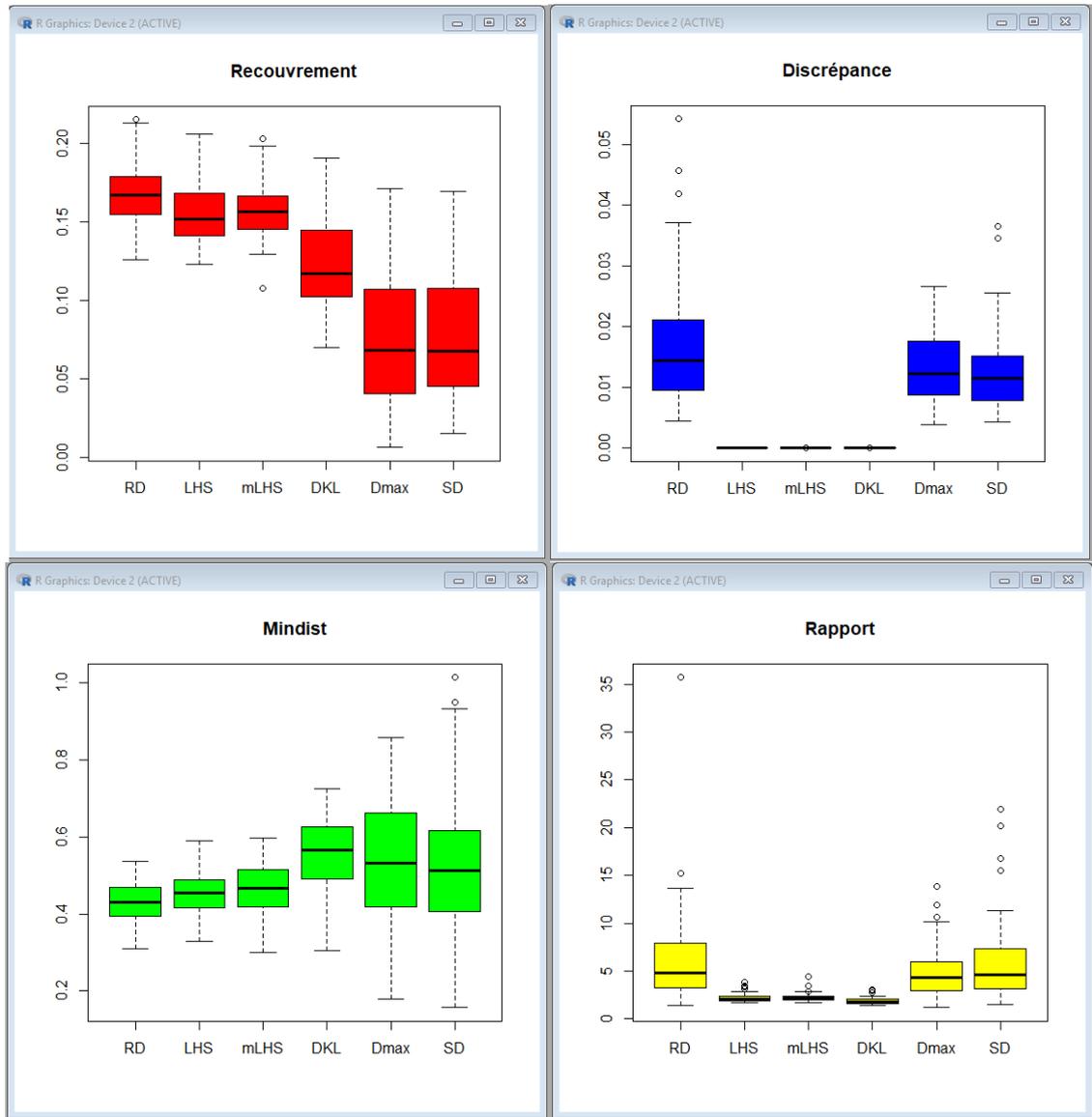


Figure 4-5 : boxplots des critères de qualité pour 10 dimensions.

## Conclusion générale

La méthode des plan d'expériences est un ensemble de techniques mathématiques et mode de raisonnement, fondés essentiellement afin d'obtenir la bonne planification des expérimentations sous forme de données numériques, en permettant sur lesquelles l'étude mathématique détaillée, qui cherche à minimiser le coût et le temps de l'expérimentation, dans ce contexte nous nous sommes intéressés à construire des plan d'expériences à partir de simulateurs déterministes et en phase exploratoire lorsque la dépendance entre les entrée et les sortie est a priori inconnue.

L'utilisation de l'information KL minimale dans le contexte des plans d'expériences numériques, permet de générer des meilleurs plans en discrédance et en rapport, ainsi les résultats apportés dans ce mémoire affirment que ces plans répondent parfaitement aux problématiques « du bon remplissage et de l'explosion du domaine expérimental ».

## BIBLIOGRAPHIE

1. Astrid Jourdan, Jessica Franco. Plans d'expériences numériques d'information de Kullback-Leibler minimale. 41<sup>èmes</sup> Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France.
2. Benoist D., Tourbier Y., Germain-Tourbier S. (1995). Plans d'expériences : construction et analyse. Tec et Doc, Paris.
3. Box G.E.P., Behnken D.W. (1960). Some New Three-Level Designs for the Study of Quantitative Variables, *Technometrics*, 2, 455-475.
4. Box G.E.P., Draper N.R. (1987). *Empirical Model Building and Responses Surfaces*. Wiley. New York.
5. Carraro L., Corre B., Helbert C., Roustant O. (2005). Construction d'un critère d'optimalité pour la quantification d'incertitudes d'une sortie de simulateur. A paraître dans la *Revue de Statistique Appliquée*.
6. Chen V.C.P., Tsui K.L., Barton R.R, Allen J.K. (2003). A review of design and modeling in computer experiments. *Handbook of Statistics* 22, 231-261.
7. Currin C., Mitchell T., Morris M. Ylvisaker D. (1991). Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments. *Journal of the American Statistical Association*, 86, 953-963.
8. Dejean J.-P., Blanc G. (1999). Managing Uncertainties on Production Prediction Using Integrated Statistical Methods. Paper SPE 56696 presented at the 1999 SPE Annual Technical Conference and Exhibition, Houston, U.S.A., October 3-6.
9. Drosbeke J.-J, Fine J., Saporta G. (1997) (Editeurs scientifiques). Plans d'expériences, application à l'entreprise. Technip, Paris.
10. Draper N . R. (1984) : The Box-Wetz criterion versus  $R^2$  . *J R Statist Soc A*, 147, Part 1 : 101-3 .
11. Éric Wolsztynski, thèse de doctorat, Critère d'entropie pour l'estimation semi-paramétrique, Université de Nice – Sophia Antipolis
12. F. J. Hickernell – « A generalized discrepancy and quadrature error bound. », *Mathematics of Computation* 67 (1998), no. 221, p. 299–322 (English).
13. H. Niederreiter – « Discrepancy and convex programming », *Annali di Matematica Pura ed Applicata* 93 (1972), no. 1, p. 89–97.
14. Iman R., Helton J.C. (1988). An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Analysis*, 8, 71-90.
15. Jessica FRANCO, thèse de doctorat, Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes, département Méthodes et Modèles Mathématiques pour l'Industrie (3MI) de l'ENSM-SE et au département Géomodélisation et Incertitudes (G&I) de TOTAL
16. J. Matoušek – « On the L2-discrepancy for anchored boxes. », *Journal of Complexity* 14 (1998), no. 4, p. 527–556 (English).
17. Joe H. (1989). Estimation of entropy and other functional of multivariate density. *Ann. Int. Statist. Math.*, 41, 683-697
18. Jourdan A., Zabalza-Mezghani I. (2004). Response surface designs for scenario management and uncertainty quantification in reservoir production. *Mathematical Geology* 36 (8), 965-985.
19. K. F. Roth – « On irregularities of distribution », *Mathematika* 1 (1954), p. 73–79.
20. Koehler J.R. et Owen A.B. (1996). Computer Experiments, *Handbook of Statistics*, 13, 261308.
21. Kosachenko L.F., Leonenko N.N. (1987). Sample estimate of entropy of a random vector. *Problem of Information Transmission*, 23, 95-101.

22. Kullback S., Leibler R.A. (1951). On information and sufficiency. *Ann. Math. Statist.*, 22 79-86.
23. L. D. Clerck – « A method for exact calculation of the stardiscrepancy of plane sets applied to the sequences of Hammersley », *Monatshefte für Mathematik* 101 (1986), p. 261–278.
24. Lindley D.V. (1956). On a Measure of the Information Provided by an Experiment. *The Annals of Mathematical Statistics*, 27, 986-1005
25. Mc Kay M.D., Conover W.J. Beckman R.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 21, 239-245.
26. Mitchell T.J., Scott D.S. (1987). A computer program for the design of group testing experiments. *Communications in Statistics – Theory and methods*, 16, 2943-2955.
27. Plackett, R.L. and Burman, J.P. (1946) *The Design of Optimum Multifactorial Experiments*. *Biometrika*, 33, 305-325.
28. Random number generation and quasi-Monte Carlo methods, *CBMS-NSF Regional Conference Series in Applied Mathematics*. 63. Philadelphia, PA : SIAM, Society for Industrial and Applied Mathematics. vi, 241 p., 1992.
29. Santner T.J., Williams B.J., Notz W. I. (2003) *The Design and Analysis of Computer Experiments*, Springer Series in Statistics, Springer-Verlag New York.
30. Sebastiani P., Wynn H.P. (2000) Maximum entropy sampling and optimal Bayesian experimental design. *J. of Royal Statist. Soc.*, 62, 145-157.
31. Shannon C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.
32. Shewry M.C., Wynn H.P. (1987). Maximum Entropy Sampling. *Journal of Applied Statistics*, 14, 165-170.
33. Suzanne Varet, thèse de doctorat, Développement de méthodes statistiques pour la prédiction d'un gabarit de signature infrarouge Suzanne Varet, Université Paul Sabatier - Toulouse III, 2010.
34. Welch W.J, Yu T.K., Kang S.M., Sacks J. (1988). Computer experiments for quality control by parameter design. Rapport technique 4, université de l'Illinois, dept. de statistics.