

Plans d'expériences numériques d'information de Kullback-Leibler minimale

Astrid Jourdan¹, Jessica Franco²

Title

Space-filling designs based on the KL divergence

Résumé

Les utilisateurs de codes numériques onéreux en temps de calcul souhaitent réduire le coût en limitant le nombre de simulations suivant un choix judicieux fondé sur l'utilisation de plans d'expériences adaptés au contexte numérique et appelés "space-filling designs". Afin de remplir au mieux l'espace des variables d'entrée du simulateur, nous proposons une méthode de construction de plans dont les points sont le plus uniformément répartis dans l'hypercube unité. Pour 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. Celleci est estimée par une méthode de Monte Carlo dans laquelle la fonction de densité est remplacée par son estimation par noyaux gaussiens.

Mots-clés : space filling designs, estimation de l'entropie, méthodes à noyaux

Abstract

Experimental designs are tools which can dramatically reduce the number of simulations required by time-consuming computer codes. Because we don't know the true relation between the response and inputs, designs should allow one to fit a variety of models and should provide information about all portions of the experimental region. One strategy for selecting the values of the inputs at which to observe the response is to choose these values so they are spread evenly throughout the experimental region, according to "space-filling designs". In this article, 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 of 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 kernel density estimator.

Keywords : space filling designs, entropy estimation, kernel density estimation

Mathematics Subject Classification: (62K99)

astrid.jourdan@eisti.fr

¹E.I.S.T.I. 26, avenue des lilas, 64062 PAU CEDEX 9 FRANCE

²Total-DGEP/GSR/TG/G&I, Avenue Larribau, 64018 PAU CEDEX FRANCE

1 Introduction

Depuis quelques années, la simulation numérique modélise des phénomènes toujours plus complexes. De tels problèmes, généralement de très grande dimension, exigent des codes de simulation sophistiqués et très coûteux en temps de calcul (parfois plusieurs jours). Dans ce contexte, le recours systématique au simulateur devient illusoire. L'approche actuellement privilégiée consiste à définir un nombre réduit de simulations organisées selon un plan d'expériences numériques. Les réponses du code sont alors utilisées pour ajuster un métamodèle, c'est-à-dire un modèle mathématique fonction des variables d'entrée et rapide d'exécution, qui remplacera le simulateur pour diverses applications.

Dans le cadre de cet article, nous nous sommes intéressés à la construction des plans d'expériences en phase exploratoire, i.e. lorsque la dépendance entre les variables d'entrée et les sorties est a priori inconnue. Il est alors difficile de prévoir quel type de métamodèle va convenir. Les plans élaborés lors de cette phase exploratoire doivent donc s'affranchir de toute contrainte par rapport à un type de métamodèle (régression linéaire, krigeage, réseaux de neurones, ...) puisque celui-ci est choisi à l'issue de ces premières simulations. Par la suite, le plan de la phase exploratoire pourra être enrichi de façon optimale en fonction du métamodèle choisi.

Ainsi, les points de ces plans doivent remplir au mieux l'espace des variables afin d'obtenir des informations dans tout le domaine expérimental, et ainsi détecter les éventuelles irrégularités à l'intérieur du domaine de simulation. On cherche donc un plan dont les points seraient le plus uniformément répartis dans l'hypercube unité. Similairement à la discrépance, l'information de Kullback-Leibler (notée KL par la suite) permet de mesurer l'écart entre la fonction de densité associée aux points du plan et celle 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 (cf. §4.1).

Cet article est organisé comme suit. Dans un premier temps, nous présentons l'information de Kullback-Leibler ainsi que les propriétés qui lui sont associées, notamment son application dans le cas de la loi uniforme et son lien avec l'entropie de Shannon. Nous proposons ensuite une technique de construction de plans basée sur une estimation par méthode de Monte Carlo de l'entropie, dans laquelle la fonction de densité associée aux points du plan est estimée par une méthode à noyaux. Enfin, nous terminons par une étude comparative avec les plans usuels à l'aide des critères couramment utilisés dans la littérature, comme par exemple la distance Maximin ([11]) ou la discrépance ([16]).

2 Information de Kullback-Leibler

Supposons que les points du plan sont les n réalisations indépendantes d'un vecteur aléatoire $X = (X^1, ..., X^d)$ de fonction de densité inconnue f de support le cube unité $E = [0, 1]^d$. 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. Afin de mesurer l'écart entre les fonctions de densité, nous avons choisi d'utiliser l'information de Kullback-Leibler ([14]).

De façon générale, l'information de Kullback-Leibler permettant de mesurer l'écart entre

les deux fonctions de densité f et g de support E, est définie par

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

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, c'est-à-dire, si on considère g la fonction de densité de la loi uniforme sur E, cela revient à imposer que le support de f est dans E.

Notons que l'information KL n'est pas une distance au sens topologique du terme puisqu'elle ne vérifie pas l'inégalité triangulaire 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. En revanche, l'information de Kullback-Leibler est définie positive, $I(f,g) \ge 0$, 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 : plus l'information est proche de 0, et plus f est « proche » de g.

En appliquant l'information KL à notre cas, c'est-à-dire en considérant g comme la fonction de densité de la loi uniforme sur E et en supposant que f est à support dans E, on peut l'écrire sous la forme d'une espérance et faire le lien avec l'entropie de Shannon,

$$I(f) = \int_{E} f(x) \ln(f(x)) dx = E_{P_f} \left[\ln(f(x)) \right] = -H[f]$$
(1)

où H[f] désigne l'entropie. Ainsi, minimiser l'information KL revient à maximiser l'entropie.

On retrouve la définition des plans à entropie maximale couramment utilisés en planification numérique ([20], [5]) ou plans bayésiens optimaux ([4], [19]). 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 connu que la loi uniforme maximise l'entropie des lois à support dans E. 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 en dimension un ont été développés à partir de l'estimation de l'entropie ([6]). Afin de ne pas confondre avec les plans à entropie maximale, les plans construits dans cet article seront dits d'information KL minimale.

Finalement, les plans sont construits suivant un algorithme d'échange visant à maximiser l'entropie. Le point essentiel dans la méthode de construction reste l'estimation nonparamétrique de l'entropie d'un ensemble de points. Il existe différentes techniques d'estimation de l'entropie qui, pour certaines d'entre elles, sont exposées dans l'article de Beirlant et al. ([2]). Nous avons retenu ici une méthode de Monte Carlo dans laquelle la fonction de densité est estimée par une méthode à noyaux.

3 Estimation de l'entropie par méthode de Monte Carlo

Notons $X_1, ..., X_n$, les *n* réalisations indépendantes de *X* constituant les points du plan *D*. Etant donné que l'entropie s'écrit sous la forme d'une espérance (1), la méthode de Monte Carlo fournit un estimateur sans biais et convergent de l'entropie,

$$\hat{H}(X) = -\frac{1}{n} \sum_{i=1}^{n} \ln f(X_i)$$
(2)

Cette estimation fait intervenir la fonction de densité f inconnue mais pouvant être estimée à partir de $X_1, ..., X_n$. La solution consistant à remplacer f par une estimation dans l'expression (2), n'assure pas que l'estimateur reste sans biais. Cependant, avoir une estimation biaisée ne pose pas de problème dans notre application, pourvu que le biais reste fixe au cours de l'algorithme d'échange. En effet, le but n'est pas d'obtenir une estimation précise de l'entropie mais un critère de comparaison entre les plans. On dira qu'un plan D_1 est meilleur qu'un plan D_2 si

$$\hat{H}(D_1) \ge \hat{H}(D_2).$$

Nous avons choisi ici d'estimer la fonction de densité par une méthode à noyaux ([21], [18]),

$$\forall x \in [0,1]^d, \ \hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1 \dots h_d} K\left(\frac{x^1 - X_i^1}{h_1}, \dots, \frac{x^d - X_i^d}{h_d}\right)$$

où $h = (h_1, \ldots, h_d)$ défini la taille de la fenêtre et K est un noyau multidimensionnel, i.e une fonction positive définie dans \mathbb{R}^d telle que $\int_{\mathbb{R}^d} K(x) dx = 1$.

La taille de la fenêtre h 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 la règle de SCOTT ([18])

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

où $\hat{\sigma}_j$ est l'estimation de l'écart-type de la jème composante X^j . Cependant, Joe ([10]) montre que dans le cas où f est estimée par une méthode à noyaux, l'estimateur de l'entropie (2) est asymptotiquement sans biais. Le biais dépend naturellement de la taille de l'échantillon n, de la dimension d, mais aussi de la taille de la fenêtre h. Lors de la construction d'un plan optimal, la taille n et la dimension d 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 l'écart-type estimé par l'écart-type de la loi uniforme sur [0, 1], d'où un choix de fenêtre anisotropique

$$h = \frac{1}{\sqrt{12}} \frac{1}{n^{1/(d+4)}}.$$

Finalement, l'estimateur de l'entropie s'écrit

$$\forall x \in [0,1]^d, \ \hat{H}(D) = -\frac{1}{n} \sum_{i=1}^n \ln\left(\frac{1}{nh^d} \sum_{k=1}^n K\left(\frac{X_k - X_i}{h}\right)\right)$$

Il est connu que la forme du noyau a peu d'influence sur l'estimation ([21]). Nous avons choisi d'utiliser un noyau gaussien multidimensionnel,

$$K(z) = \frac{(2\pi)^{-d/2}}{s^d} \exp\left(-\frac{1}{2s^2} \|z\|^2\right),$$

où s^2 est choisi égal à la dimension multipliée par la variance de la loi uniforme sur [0, 1],

$$s^2 = \frac{d}{12}.$$

Un noyau de support la boule unité (Epanechnikov, uniforme,...) n'est pas envisageable du fait de la grande dimension et de la petite taille de l'échantillon. En effet, dans notre application, la fonction noyau est appelée avec des valeurs de z variant dans l'intervalle $[0, \frac{d}{h^2}]$. Cet intervalle devient très grand quand le taille et la dimension augmentent (par exemple, [0; 231, 7] pour d = 10 et n = 100), et la probabilité pour que z soit à l'intérieur du support du noyau devient alors très faible. L'estimation de l'entropie des plans de même taille est alors quasiment constante, et le critère ne permet donc pas de comparer les plans.

Avec une estimation de f par un noyau gaussien, la propriété d'absolue continuité n'est pas vérifiée et n'est pas à support dans E. Cette méthode introduit donc un biais supplémentaire dans l'estimation de l'entropie. Une méthode d'estimation par histogrammes modifiées ([1]) aurait été plus satisfaisante concernant ce point. Cependant, cette estimation pose le problème de la partition de E en dimension élevée. Par exemple, une partition en rectangles d'un plan de dimension 10 avec seulement deux segments sur chaque axe, nécessite déjà l'évaluation de 210 cellules.

Les propriétés établies par Joe ([10]) 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 (cf. figure 1(a)). 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. Il est intéressant de voir que, malgré cette imprécision, l'algorithme d'échange converge assez rapidement (cf figure 1(b)) et que les plans ainsi construits ont les propriétés attendues, à savoir un remplissage uniforme de l'espace des variables, comme nous allons le voir dans le paragraphe suivant.

4 Comparaison des plans

Afin de juger de l'intérêt de la méthode de construction, nous avons construit des plans en dimension 2, 3 et 10 pour une taille standard fixée à $n = 10 \times d$, à l'aide de l'algorithme d'échange suivant.



FIGURE 1 – Illustration de convergences pour des plans en dimension 3

Algorithme d'échange³

- 1 Initialiser avec un plan aléatoire, i.e n points choisis au hasard dans $[0,1]^d$
- 2 Construire un nouveau plan en remplaçant un point du plan actuel avec un point candidat choisi aléatoirement dans $[0,1]^d$
- 3 Calculer le critère pour le nouveau plan et décider de garder ce nouveau plan s'il améliore le critère
- 4 Répéter l'étape 2 jusqu'à ce que les conditions d'arrêt soient vérifiées

Le plan résultant de l'algorithme d'échange est dit « optimal » ou « presque optimal » au regard du critère utilisé. Ce plan « optimal » n'est pas unique et dépend plus ou moins du plan initial. Afin de réduire le risque de convergence de l'algorithme vers un optimum local, plusieurs initialisations sont testées et le meilleur (au sens du critère utilisé) plan est sélectionné. Toutefois, dans le paragraphe 4.2, plusieurs plans sont construits mais à partir d'une seule initialisation de façon à étudier l'influence du plan initial sur les performances du plan résultant.

4.1 Caractéristiques des plans d'information KL minimale

L'objectif de la méthode de construction est un remplissage uniforme de l'espace des paramètres. On constate visuellement sur la figure 2 que cet objectif est atteint et ce quelle que soit l'initialisation du plan. On remarque que les deux plans finaux présentent les mêmes caractéristiques :

- Les points du plan sont disposés au bord du domaine mais aussi à l'intérieur suivant

³Le programme R peut être mis à disponibilité.

la disposition d'une grille régulière légèrement perturbée (plans dits quasi-périodiques).
Ainsi le remplissage de l'espace est garanti tout en conservant une distribution aléatoire.
Chaque paramètre est testé sur un grand nombre de niveaux.

Ces caractéristiques sont conservées en dimension $d \ge 2$.

Quel que soit le plan initial, l'algorithme converge vers le même type de plan décrit cidessus. Cette remarque, confirmée par les tests ci-après, permet de limiter le nombre d'initialisations aléatoires dans l'algorithme et ainsi réduire le temps de calcul.





Sur la figure 3, on remarque que les plans présentent une mauvaise répartition des points en projection sur des sous-ensembles de variables. En effet, le critère semble privilégier les bords du domaine lorsque le nombre de points est faible. Ce phénomène n'est pas spécifique à notre critère mais a souvent été remarqué dans la littérature à propos de divers space-filling designs (Koehler[12]). Il est cependant peut-être amplifié du fait d'avoir choisi un noyau à support infini. Lorsqu'une projection uniforme des points sur chaque axe factoriel est nécessaire, la stratégie communément adoptée consiste alors à construire des hypercubes latins optimaux ([17]).



FIGURE 3 – Projection d'un plan de dimension 10 et de taille 100

4.2 Comparaison avec les critères et plans usuels

Etudier l'uniformité d'une distribution de points lorsque $d \ge 2$, ne peut se faire visuellement. Il est donc utile de se référer à des critères afin de pouvoir décider si une distribution est uniforme ou non ou/et si elle vérifie également le bon remplissage de l'espace. Il existe pour cela des critères intrinsèques aux plans d'expériences - c'est-à-dire des critères qui ne préjugent en rien de la qualité de la surface de réponse déterminée par la suite à partir de ces plans - qui se prêtent particulièrement bien à cet objectif. Nous avons choisi de tester uniquement les critères les plus courants, auxquels nous ajoutons notre critère noté EntGauss.

La mesure de recouvrement (Cov), définie par

$$\frac{1}{\overline{\gamma}} \left(\frac{1}{n} \sum_{i=1}^{n} (\gamma_i - \overline{\gamma})^2 \right)^{1/2} \operatorname{avec} \gamma_i = \min_{j \neq i} \|x_i - x_j\| \text{ et } \overline{\gamma} = \frac{1}{n} \sum_{i=1}^{n} \gamma_i$$

est nulle aux points d'une grille régulière ([8]). L'objectif est donc de minimiser ce critère pour se rapprocher d'une grille régulière, et ainsi assurer le remplissage de l'espace, sans toutefois l'atteindre pour respecter une distribution uniforme

 La distance maximin (Mindist), définie par Johnson et al. ([11]), maximise la distance minimale entre deux points du plan,

$$\min_{i\neq j} d(x_i, x_j)$$

 La discrépance permet de mesurer l'écart entre la fonction de répartition empirique des points du plan et celle de la loi uniforme,

$$\sup_{e \in [0,1]^d} |F_n(t) - U(t)| \quad \text{avec} \quad F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i \le t\}}$$

Il existe différentes mesures de discrépance ([16], , [23]). Nous retenons la discrépance en norme L2 (DL2).

Les plans construits ici sont comparés avec les plans usuellement utilisés en expériences numériques. Nous conseillons la lecture des travaux de Koehler & Owen ([12]) et Franco ([7]) pour un état de l'art concernant les « space filling designs ». On notera que certains de ces plans sont optimaux pour les critères cités ci-dessus.

- Aléatoire : plans construits par tirage aléatoire dans le cube unité.
- LH : plans construits à partir d'hypercubes latins (sans critère d'optimisation) ([22])
- Discrépance : suites à faible discrépance (Sobol,Niederreiter, Hammersley, Halton) ([16], [23]).
- Dmax : plans à entropie maximale ([20], [5]) construits de façon à maximiser le déterminant d'une matrice de covariance. Ces plans sont ainsi très utilisés lorsque la surface de réponse est ajustée par krigeage. Ils supposent cependant un modèle sous-jacent.
- Strauss : plans élaborés à partir d'un processus de Strauss qui considère de la répulsion entre les points de manière à remplir au mieux l'espace des paramètres ([7])
- Maximin : plans optimaux pour le critère maximin ([11]).
- MCGauss : plans d'information KL minimale construits par la méthode décrite précédemment.

Pour chacune de ces catégories, 20 plans ont été construits (exceptées les suites de faible discrépance). Les boxplot des figures 4 et 5 représentent les scores obtenus par les 20 plans pour les quatre critères utilisés.

En dimension 3, les plans d'information KL minimale (MCGauss) donnent les meilleures valeurs excepté pour le critère de discrépance. Ce résultat est inattendu, car le critère KL et la discrépance visent tous les deux à minimiser l'écart entre la distribution empirique et la loi uniforme alors que les critères de recouvrement et de distance maximin sont basés sur la distance entre les points. Ce résultat peut s'expliquer par la difficulté à évaluer la discrépance.

Bien souvent, les space-filling designs perdent de leurs qualités en grande dimension. La figure 5 illustre les performances des plans en dimension 10. Seules les quatre meilleures classes de plan ont été conservées. On constante que les plans d'information KL minimale gardent de bonnes propriétés en grande dimension. Ils semblent concurrencer les plans maximin traditionnellement utilisés en phase exploratoire, et ceci même pour le critère de distance maximin.

On note sur les figures 4 et 5 que les résultats des plans MCGauss sont très peu dispersés. Cela confirme la remarque du paragraphe 4.1, à savoir que les plans dépendent peu de l'initialisation de l'algorithme d'échange. Le temps de calcul peut ainsi être réduit en diminuant le nombre d'initialisations dans l'algorithme.

5 Conclusions

L'objectif du critère présenté dans cet article est de minimiser l'écart entre la distribution empirique des points du plan et la loi uniforme. Minimiser l'information de Kullback-Leibler pour optimiser ce critère revient à maximiser l'entropie. Ainsi, toute la technique de construction des plans repose sur l'estimation de l'entropie. Celle-ci se fait par une méthode de Monte Carlo dans laquelle la fonction de densité est estimée à l'aide de noyaux gaussiens. Les tests effectués montrent que les plans d'information KL minimale répondent à l'objectif principal, à



FIGURE 4 – Représentation des critères usuels calculés sur 20 plans à 30 points en dimension 3

savoir une bonne représentation du domaine de simulation par les points du plan, et ce, même en grande dimension. Quel que soit le critère choisi, ces plans concurrencent les space-filling designs usuels et notamment les plans maximin habituellement utilisés en phase exploratoire.



(e) zoom de EntGauss



Malgré des conditions éloignées de l'asymptotique, un critère basé sur l'estimation de l'entropie semble fournir de très bons résultats. Cependant, lorsque la dimension augmente, la méthode d'estimation par Monte Carlo devient moins performante lorsque l'échantillon est de petite taille et requière un temps de calcul prohibitif. De plus, nous avons vu que cette méthode pose le problème du support de la fonction de densité f. Il pourrait être intéressant de calculer le critère à l'aide d'une estimation de l'entropie qui évite, à la fois la méthode de Monte Carlo, et à la fois l'estimation de la fonction de densité. Par exemple, la méthode des plus proches voisins ([13]) se prête particulièrement bien aux grandes dimensions ([15]).Une autre perspective consiste à définir le critère à partir d'une entropie différente de celle de Shannon, par exemple, l'entropie de Rényi qui peut être estimée à partir des arbres de longueur minimale ([9]), ou bien l'entropie de Tsallis qui, à l'ordre deux et dans le cas d'une méthode d'estimation de la densité par noyaux gaussiens, s'écrit de façon analytique comme une somme des noyaux et évite ainsi l'erreur d'approximation de la méthode de Monte Carlo ([3]).

Remerciement

Merci à Bernard Corre de l'entreprise Total qui est à l'origine de ce travail.

Références

- A. R. BARRON, L. GYÖRFI et E. C. VAN DER MEULEN « Distribution estimation consistent in total variation and two types of information divergence », *IEEE Trans. Inform. Theory* 5 (1992), p. 1867–1883.
- [2] J. BEIRLANT, E. DUDEWICZ, L. GYÖRFI et E. C. VAN DER MEULEN « Nonparametric entropy estimation : an overview. », *Int. J. Math. Stat. Sci.* **6(1)** (1997), p. 17–36.
- [3] R. BETTINGER, P. DUCHÊNE, L. PRONZATO et E. THIERRY « Design of experiments for response diversity. », Proc. 6th International Conference on Inverse Problems in Engineering (ICIPE) (Dourdan (Paris)), vol. Journal of Physics : Conference Series, June, 2008.
- [4] K. CHALONER et I. VERDINELLI « Bayesian experimental design : A review. », *Statis. Sci.* **10** (1995), p. 237–304.
- [5] C. CURRIN, T. MITCHELL, M. MORRIS et D. YLVISAKER « A bayesian approach to the design and analysis of computer experiments. », Tech. Report 6498, ORNL, 1988, available from the national technical information service, Springfield, Va. 22161.
- [6] E. J. DUDEWICZ et E. C. VAN DER MEULEN « Entropy-based tests of uniformity. », J. Amer. Statist. Assoc. **76** (1981), p. 967–974.
- [7] J. FRANCO « Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes. », Thèse, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2008.
- [8] M. GUNZBURGER et J. BURKARDT « Uniformity measures for point sample in hypercubes. », https://people.scs.fsu.edu/~burkardt/pdf/ptmeas.pdf, 2004.
- [9] A. HERO, BING MA, O. MICHEL et J. GORMA « Applications of entropic spanning graphs. », *Signal Processing Magazine, IEEE* **19** (2002), p. 85–95.
- [10] H. JOE « Estimation of entropy and other functional of multivariate density. », Ann. Int. Statist. Math. 41 (1989), p. 683–697.

- [11] M. E. JOHNSON, L. M. MOORE et D. YLVISAKER « Minimax and maximin distance design. », J. Statist. Plann. Inf. 26 (1990), p. 131–148.
- [12] J. R. KOEHLER et A. B. OWEN « Computer experiments. », *Handbook of statistics*, vol. 13, 1996, p. 261–308.
- [13] L. F. KOSACHENKO et N. N. LEONENKO « Sample estimate of entropy of a random vector. », Problem of Information Transmission 23 (1987), p. 95–101.
- [14] S. KULLBACK et R. A. LEIBLER « On information and sufficiency. », Ann. Math. Statist. 22 (1951), p. 79–86.
- [15] N. N. LEONENKO, L. PRONZATO et V. SAVANI « A class of rényi information estimators for multidimensional densities. », Annals of Statistics (2009), to appear.
- [16] H. NIEDERREITER « Point sets and sequences with small discrepancy. », Monasth. Math. 104 (1987), p. 273–337.
- [17] J. S. PARK « Optimal latin hypercube designs for computer experiments. », J. Stat. Plan. Inf. **39** (1994), p. 95–111.
- [18] D. W. SCOTT Multivariate density estimation : Theory, practice and visualization, John Wiley & Sons, New York, Chichester, 1992.
- [19] P. SEBASTIANI et H. P. WYNN « Maximum entropy sampling and optimal bayesian experimental design. », J. Royal Statist. Soc. 62 (2000), p. 145–157.
- [20] M. C. SHEWRY et H. P. WYNN « Maximum entropy sampling. », J. Appl. Statist. 14 (1987), p. 165–170.
- [21] B. W. SILVERMAN Density estimation for statistics and data analysis., Chapman & Hall, London, 1986.
- [22] M. STEIN « Large sample properties of simulations using latin hypercube sampling. », *Technometrics* 29 (1987), p. 143–151.
- [23] E. THIÉMARD « Sur le calcul et la majoration de la discrépance à l'origine. », Thèse, École polytechnique fédérale de Lausanne, 2000.