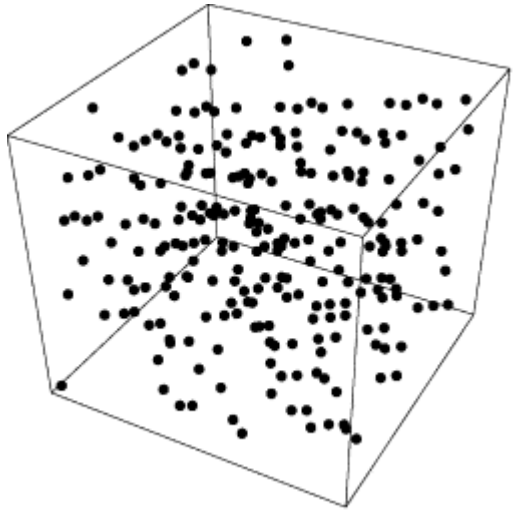
EISTI

SPACE-FILLING DESIGNS BASED ON RÉNYI ENTROPY

"how to put dots in cubes"



Astrid Jourdan
EISTI, Quartz Lab.
aj@eisti.eu

Thomas Alauzet
EISTI, master student
alauzettho@eisti.eu

GENESIS

The Annals of Statistics
2008, Vol. 36, No. 5, 2153–2182
DOI: 10.1214/07-AOS539
© Institute of Mathematical Statistics, 2008

A CLASS OF RÉNYI INFORMATION ESTIMATORS FOR MULTIDIMENSIONAL DENSITIES

BY NIKOLAI LEONENKO,[1] LUC PRONZATO [2] AND VIPPAL SAVANI

Journal de la Société Française de Statistique
Vol. 158 No. 1 (2017)

Minimax and maximin space-filling designs: some properties and methods for construction

Titre: Plans d'expériences à remplissage d'espace minimax et maximin : quelques propriétés et méthodes de construction

Luc Pronzato[1]

"but to the best of our knowledge the sum of power-weighted edge lengths for the MST or any other graph among those mentioned above, has scarcely been used as a criterion for space-filling design"

Design and Analysis of Computer Experiments
Jerome Sacks, William J. Welch, Toby J. Mitchell and Henry P. Wynn
Statistical Science 1989, Vol. 4, No. 4, 409–435

Jessica Franco

Oscar — 2004
Adélie — 2007
Violette — 2010

2000 2008 2009 2012 2017 2019

Journal de la Société Française de Statistique
Volume 150, numéro 2, 2009

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

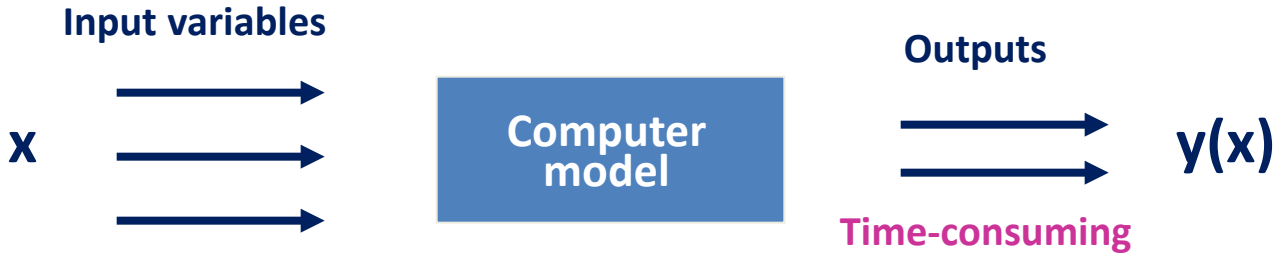
Astrid Jourdan[1], Jessica Franco[2]

Stat Comput (2012) 22:681–701
DOI 10.1007/s11222-011-9242-3

Design of computer experiments: space filling and beyond

Luc Pronzato · Werner G. Müller

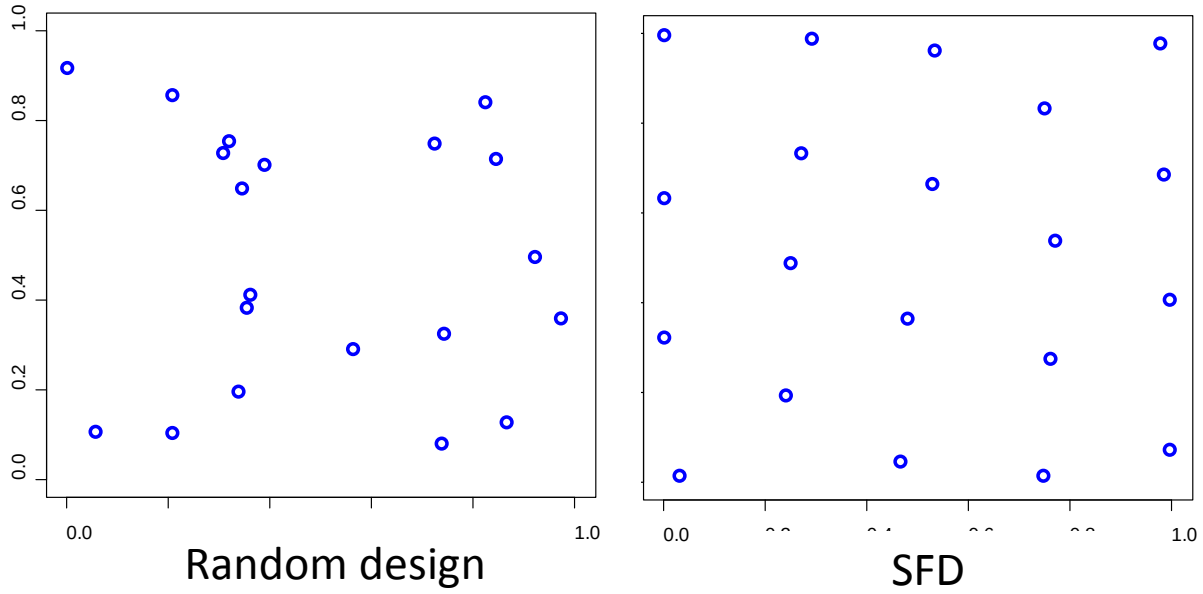
SPACE-FILLING DESIGNS (SFD)



How to determine the inputs for which the simulator is run (computer experiments)?

- Exploratory step
- Spot possible irregularities of the computer response
- Fit a variety of surrogate models

$\Rightarrow$ Fill up the experimental region in a uniform fashion



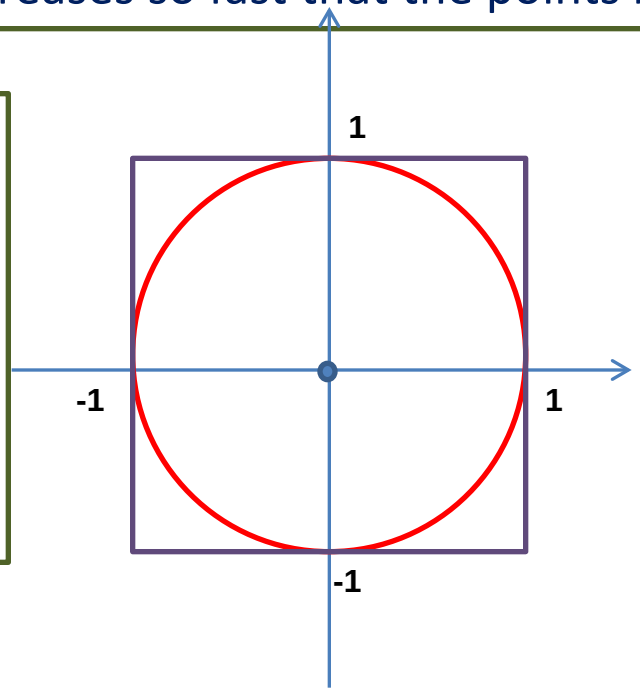
Many space-filling criteria : entropy (Shewry & Wynn, 1987) or maximin distance (generalized with ϕ_p) (Johnson et al., 1990) or Audze-Eglais criterion (ϕ_p with p=1/2) (Audze & Eglais, 1977), …

THE CURSE OF DIMENSIONALITY

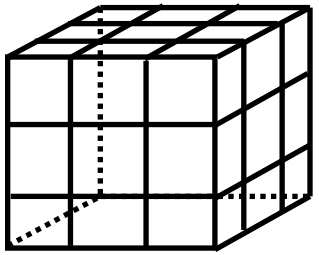
Dimension = number of variables

When the dimensionality increases, the volume of the space increases so fast that the points become sparse.

- Difficult to organize sparse data
- An enormous amount of training data are required to ensure a good exploration of a high dimensional space.



d	Vol. hypercube	Vol. sphere	% of Covering
2	4	3,1	78,5%
4	16	4,9	30,8%
6	64	5,2	8,1%
8	256	4,1	1,6%
10	1024	2,6	0,2%



Grid with 2 levels = 2^d points (e.g. d=20 $\Rightarrow$ 33554432 points)

k levels = k^d points (e.g. k=5 and d=20 $\Rightarrow$ 95367431640625 points)

UNIFORM DESIGNS

Suppose that the design points $x_1,\dots,x_n$, are n independent observations of the random vector $X=(X^1,\dots,X^d)$ with absolutely continuous density function f_D supported by $E=[0,1]^d$.

The aim is to select the design points in such a way that the empirical distribution of the points is "close" to the uniform distribution.

Many ways to measure the closeness of the two distributions:

- Discrepancy compares the cumulative distributions (Niederreiter, 1987, Fang et al. , 2006)
- Entropies compare the density functions
  - Shannon entropy (Jourdan & Franco, 2009, 2010)
  - Rényi entropy
  - Tsallis entropy

  } (Pronzato & Muller, 2012)
  - …

(RÉNYI, 1961)

THE RÉNYI ENTROPY

The Rényi entropy of the design D is

$$H_q(f_D) = \frac{1}{1-q} \ln \int_E f_D(x)^q dx \quad \text{with } q \in [0,1[$$

when q tends to 1, H_q tends to the Shannon entropy

- Properties : If f is supported by $[0,1]^d$, one always has $H(f) \leq 0$ and the maximum value of $H_q(f)$, zero, being uniquely attained by the uniform density.
  $H_q(f)$ is a concave function.
- Goal : The aim is to build a design which maximizes the Rényi entropy or more simply the integral

$$I_q(f_D) = \int_E f_D(x)^q dx$$

Estimation of the Rényi entropy

Many methods exist to estimate the entropy of a distribution from i.i.d. samples. We used the three most widely known :

- Monte Carlo estimation (*plug-in* method)
- Nearest neighbor distance
- Minimal Spanning tree

Pronzato and Muller (2012) linked these methods with other space-filling criteria (Maximin distance/ Φ_p criterion)

Jourdan and Franco (2009, 2010) used MC and NN estimation of Shannon entropy.

(BEIRLANT *ET AL.* 1997)

MONTE CARLO ESTIMATION (1/2)

The integral can be written as an expectation

$$I_q[f_D] = \int f_D(x)(f_D(x))^{q-1}\,dx = E_{P_f}\left[(f_D(X))^{q-1}\right]$$

The Monte Carlo method provides a unbiased and consistent estimate of the integral

$$\hat{I}_q(f_D) = \frac{1}{n}\sum_{i=1}^{n} f_D^{\,q-1}(X_i)$$

where $X_1,\ldots,X_n$ are the design points.

The unknown density function f_D is replaced by its kernel density estimate.

$$\hat{f}_D(x) = \frac{1}{nh^d}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)$$

where K is the kernel and h the bandwidth.

Joe (1989) obtained asymptotic bias and variance terms for the estimator of I_q. The bias depends on the size n, the dimension d and the bandwidth h $\Rightarrow$ fix the bias (h) during the optimization algorithm

Monte Carlo Estimation (2/2)

The bandwidth is determined by the Scott's rule with the variance of the uniform distribution,

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

The kernel is a multidimensional Gaussian function,

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

$\hat{f}$ *is no more supported by* $[0,1]^d$

After simplification, the criteria to maximize is

$$MC_q(D) = \sum_{i=1}^{n} \left(\sum_{i<j\leq n} e^{-\frac{72n^{2/(d+4)}}{d} \|x_i - x_j\|^2} \right)^{q-1}$$

LEONENKO. ET AL. (2008)

NEAREST NEIGHBOR DISTANCE

Leonenko, Pronzato and Savani (2008) suggested to estimate the Renyi entropy by using the nearest neighbor distances,

$$\hat{I}_q(D) = \frac{[(n-1)C_k V_d]^{1-q}}{n} \sum_{i=1}^{n} \rho_{k,i}^{1-q}$$

where V_d is the volume of the unit ball in R^d, $C_q=(\Gamma(k)/\Gamma(k+1-q))^{(1/q-1)}$ and $\rho_{k,i}$ is the k-th nearest-neighbor distance of X_i to some other sample. If k=1

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

Under conditions between q and k, they proved the asymptotic unbiasedness and the consistency of this estimator.

After simplification, the criteria to maximize is

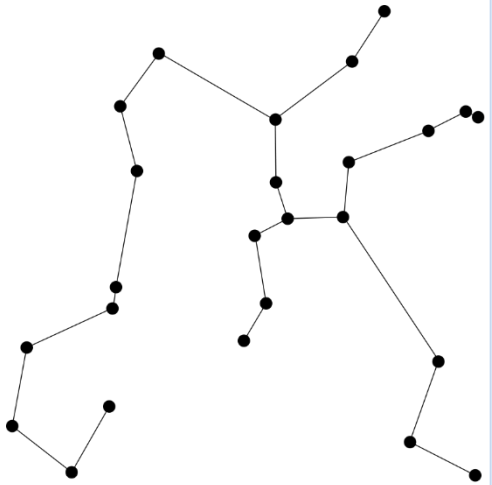
$$NN_q(D) = \sum_{i=1}^{n} \rho_{k,i}^{d(1-q)}$$

Note that this method is also a MC estimation where the density is replaced by its estimation with nearest neighbor

Redmond&Yukich (1996)
Hero & Mitchell (1999)

Minimal spanning tree

https://en.wikipedia.org/wiki/File:Euclidean_minimum_spanning_tree.svg



A minimum spanning tree (MST) is a subset of the edges, $e_{i,j}$, of a connected, edge-weighted undirected graph that connects all the vertices together, without any cycles and with the minimum possible total edge weight. That is, it is a spanning tree whose sum of edge weights

$$L_\gamma(D) = \sum_{e_{i,j}} \left\| e_{i,j} \right\|^\gamma$$

is as small as possible.

Redmond & Yurich (1996) or Hero & Mitchell (1999, k-MST) demonstrated the convergence of the estimator

$$\hat{H}_q(D) = \frac{1}{1-q} \ln\left(n^{-q} L_{d(1-q)}(D)\right) + \beta(q,d)$$

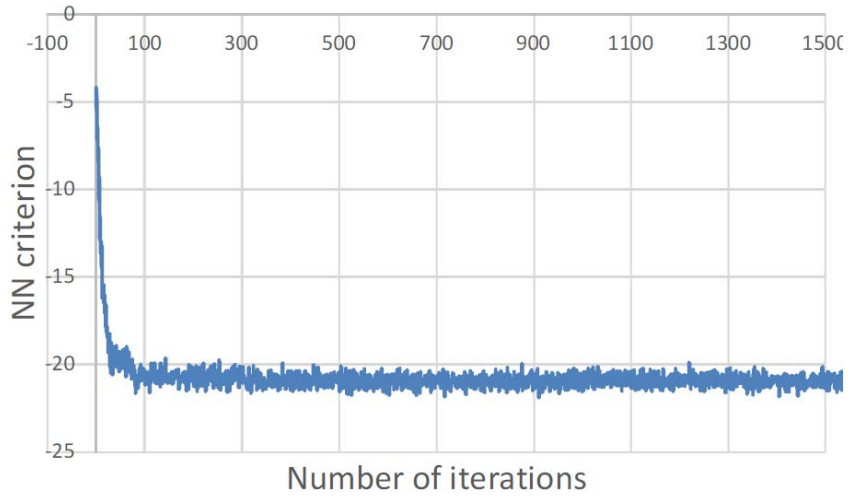
After simplification, the criteria to maximize is

$$MST_q(D) = L_{d(1-q)}(D)$$

Franco et al. (2009) used the mean and standard deviation of the edge lengths of the MST for design classification.

OPTIMIZATION ALGORITHM

The simulated annealing algorithm



Finding a neighborhood is decisive. In our case, we move each point from its nearest neighbor by a factor proportional to n^{-1}.

For fixed values of n and d we can estimate the computation time for MC, NN and MST estimation of the entropy with the “O” common notation :

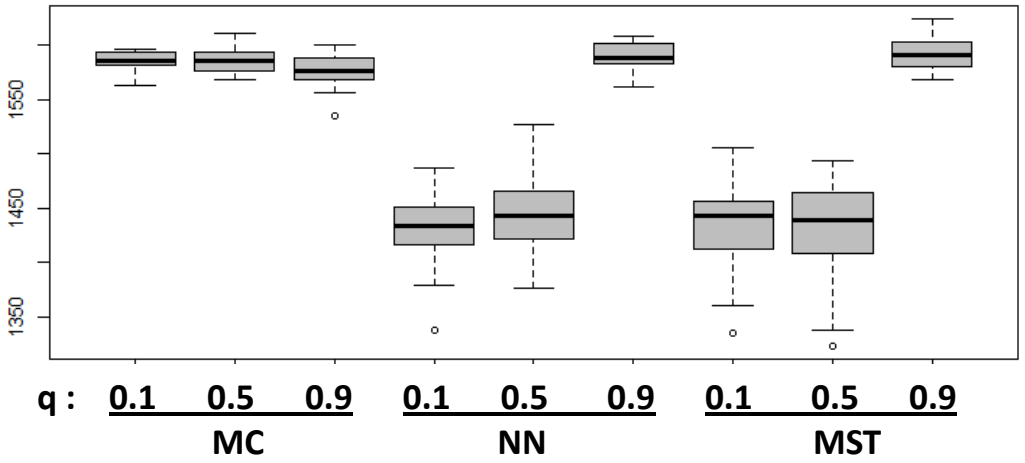
MC(D): $O(d*n*(n+1) / 2)$
NN(D): $O(d*n*(n-1))$
MST(D): $O(d*n*(n+1) / 2) + O(n^2*\log(n))$

For the MST estimation the first term is the graph initialization. The second one is the complexity for Kruskal’s algorithm.

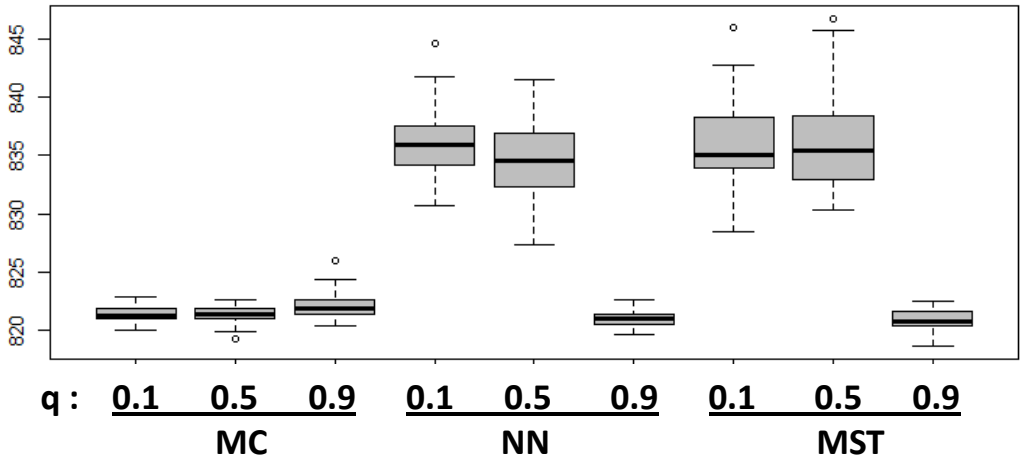
DESIGN COMPARISON (1/3)

Perform well under any criterion (Discrepancy, Maximin distance, AE, …)

Discrepancy (DC2) (to maximize)

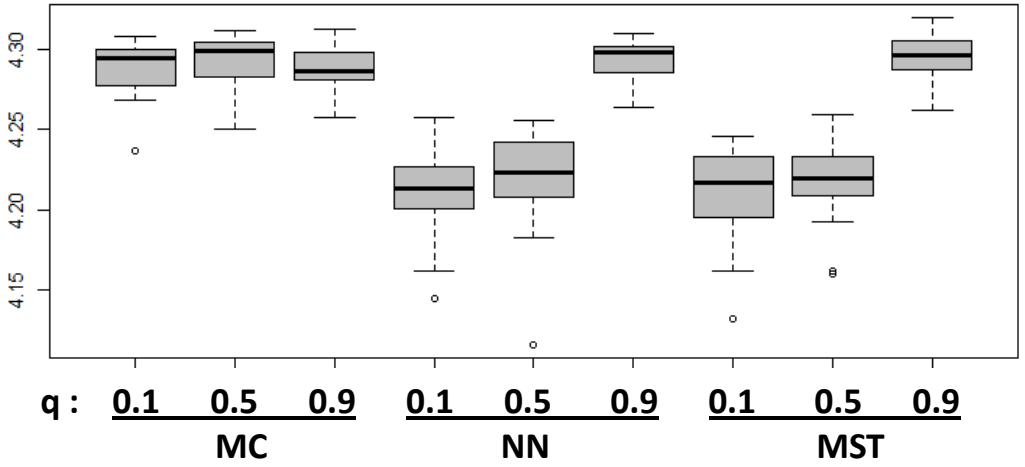


AE (to minimize)

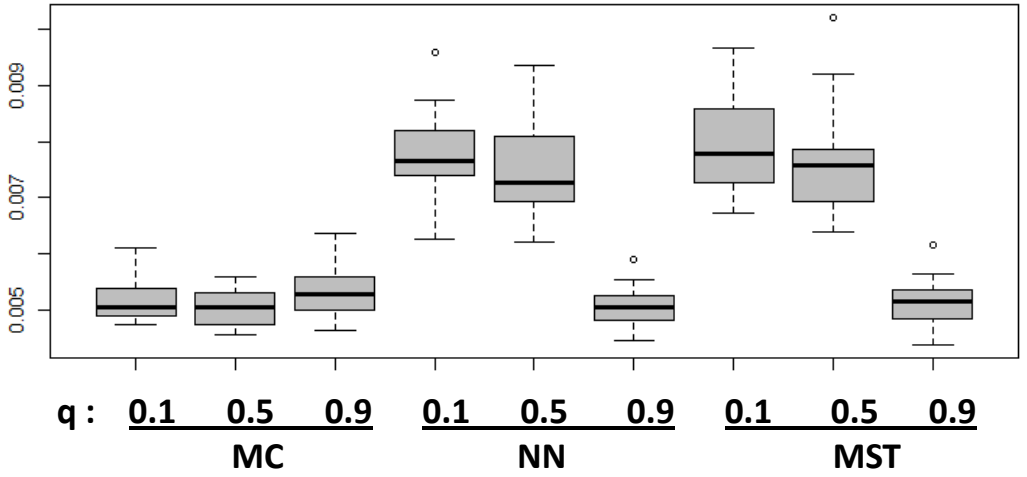


d=50, n=200

Maximin (to maximize)

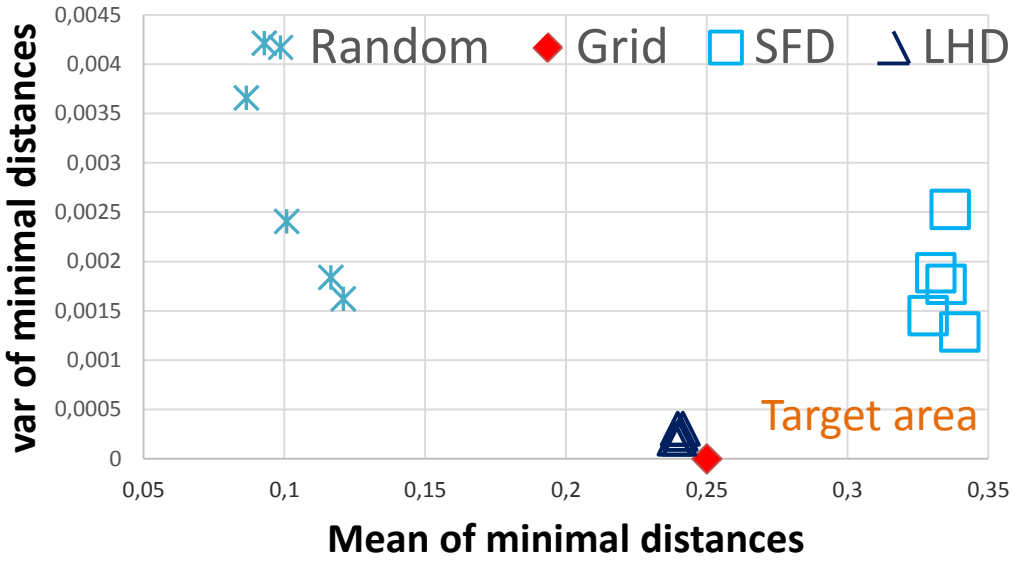


Coverage (to minimize)



DESIGN COMPARISON (2/3)

Scrambled regular grid

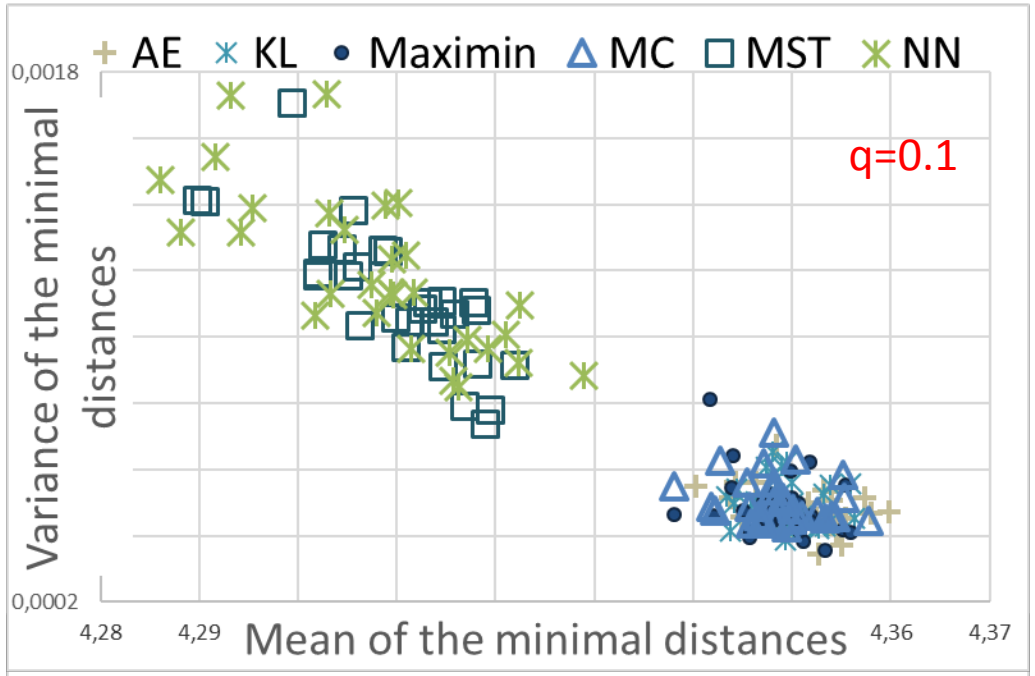


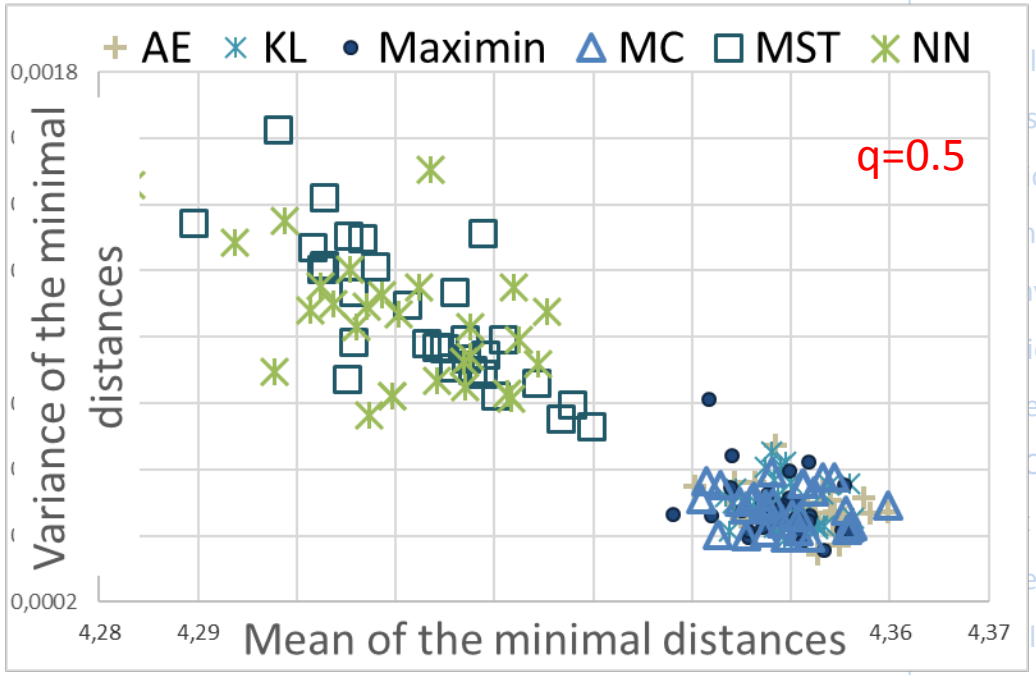
Compute the minimal distance of each of point with the others

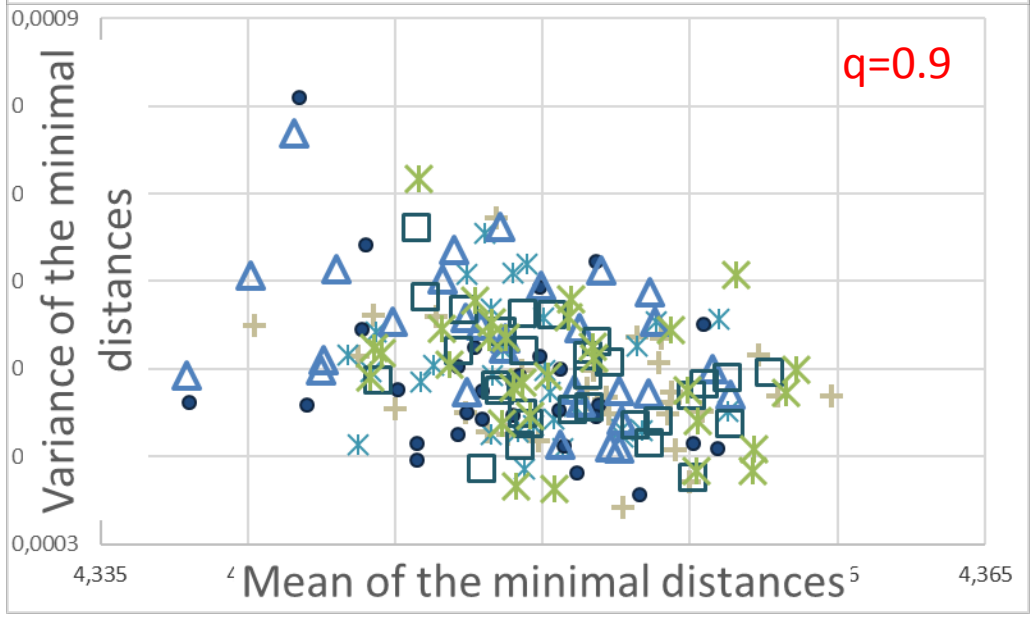
X-axis= average of the minimal distances
Y-axis = variance of the minimal distances

Goal = high value on x-axis and small value on y axis

DESIGN COMPARISON (3/3)







- Whatever the value of q, MC designs are in the target area
- When q=0.9, all points are in the target area
- High variability for NN and MST designs with q=0.1 and q=0.5
- Not better than existing designs

CONCLUSION

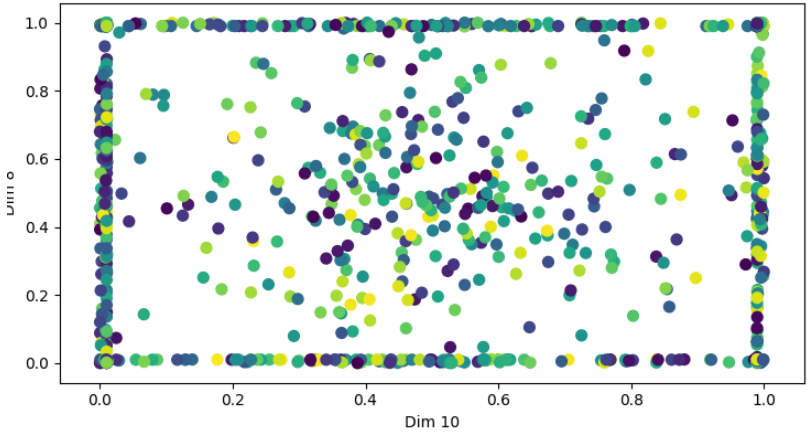
- Renyi entropy is not a so good criterion to build space-filling designs

- As the other criteria, Renyi entropy spreads the points on the borders of the experimental domain
  $\Rightarrow$ Latin hypercubes

- Key point is the optimization algorithm

➢ The late acceptance hill-climbing heuristic (Burke&Bykov, 2017)

➢ Multi-objective optimisation

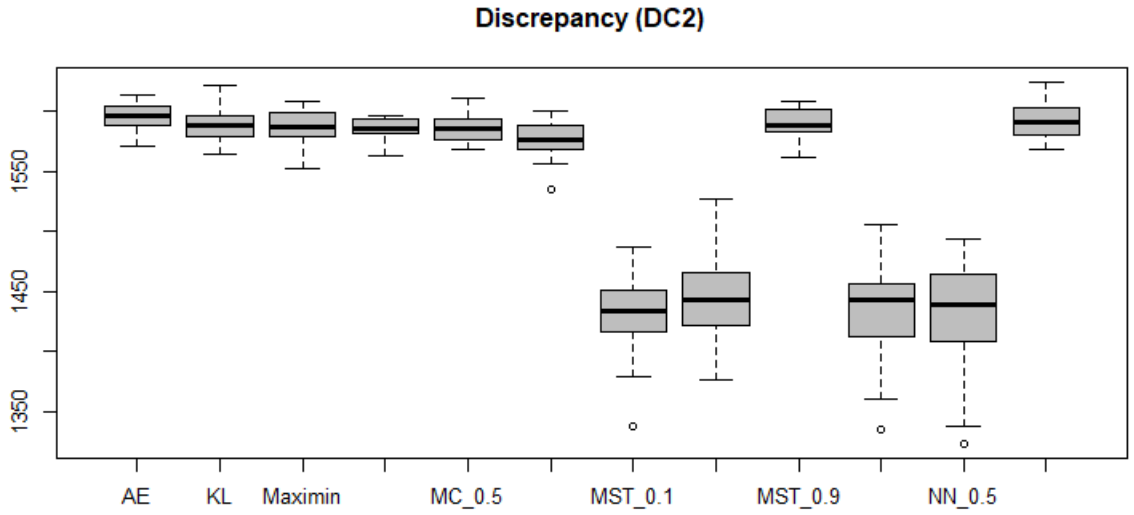


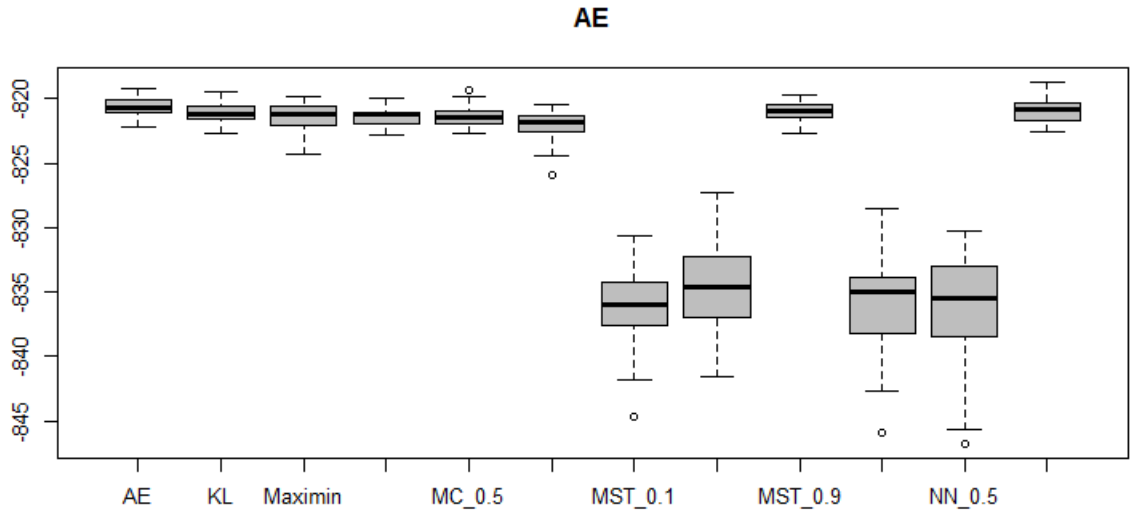
2D projection of 1000x10 design
Number of corners is
2^{10}=1024>1000

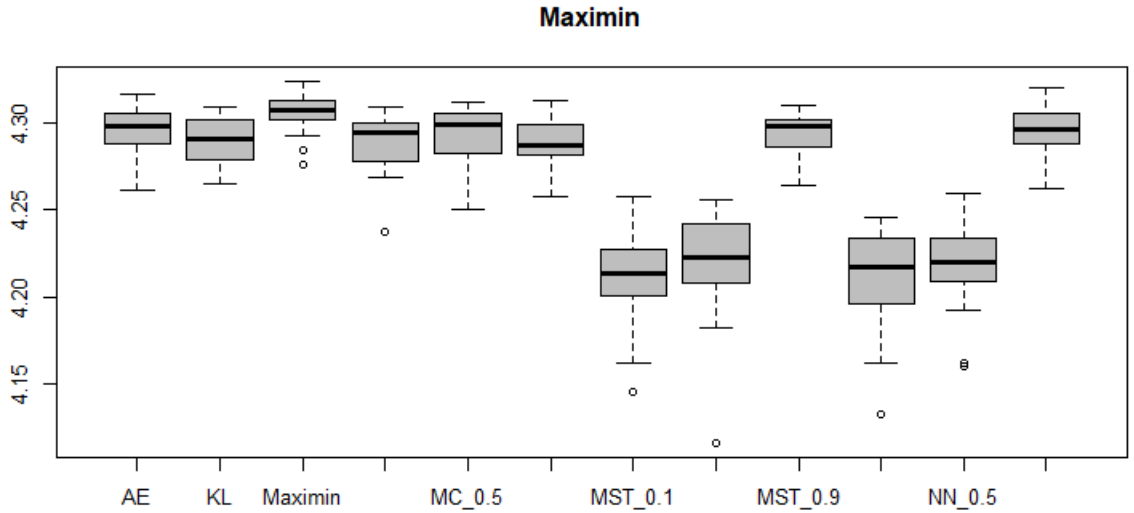
Thank you for your attention

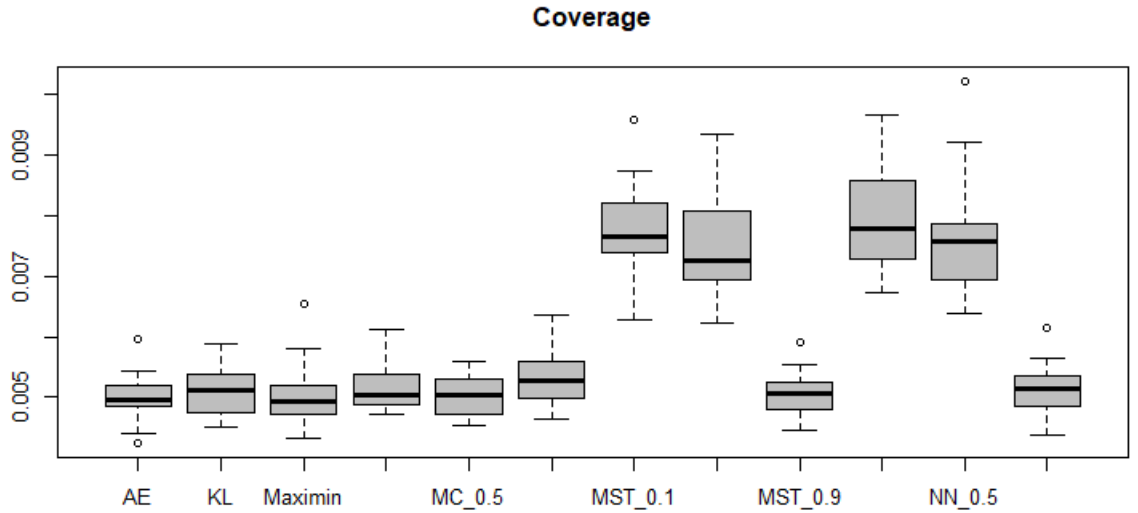
REFERENCES

- Audze P, Eglais V. (1977). New approach for planning out of experiments. *Problems of dynamics and strengths*, 35. Zinatne Publishing House; 104-107.
- Beirlant J., Dudewicz E.J., Györfi L., Van Der Meulen E.C. (1997). Nonparametric entropy estimation : an overview. *International Journal of Mathematical and Statistical Sciences*, 6(1), 17-39.
- Burke E., Bykov Y. (2017). The late acceptance hill-climbing heuristic. European Journal of Operational Research, 258(1), Pages 70-78.
- Fang K.T., Li R., Sudjianto A. (2006). *Design and modeling for computer experiments*. Chapman&Hall, London.
- Franco J., Vasseur O., Corre B., Sergent M. (2009). Minimum spanning tree : a new approach to assess the quality of the design computer. *Chemiometrics and Intelligence Laboratory Systems*, 97, 164-169.
- Hero A., Ma B., Michel O., Gorman J. (2002). Applications of entropic spanning graphs. *IEEE Signal Processing Magazine*, 19(5), 85–95.
- Hero A., Michel O. (1999). Asymptotic theory of greedy approximations to minimal k-point random graphs. *IEEE Trans. on Inform. Theory* ,45(6), 1921–1939.
- Joe H. (1989). Estimation of entropy and other functional of multivariate density. *Ann. Int. Statist. Math*., 41, 683-697.
- Johnson M.E., Moore L.M., Ylvisaker D. (1990). Minimax and maximin distance design. *Journal of Statistical Planning and Inference*, 26,131-148.
- Jourdan A. et Franco J. (2009). Plans d'expériences numériques d'information de Kullback-Leibler minimale. *Journal de la Société Française de Statistique*, 150 (2), 52-64.
- Jourdan A. et Franco J. (2010). Optimal Latin hypercube designs for the Kullback-Leibler criterion. *AStA Advances in Statistical Analysis*, 94 (4), 341-351.
- Leonenko N., Pronzato L., Savani V. (2008). A class of Rényi information estimators for multidimensional density. *The Annals of Statistics*. 36(5), 2153-2182.
- Niederreiter H. (1987). Point sets and sequences with small discrepancy. *Monatshefte fur Mathematik.*, 104, 273-337.
- Pronzato L. (2017). Minimax and maximin space-filling designs: some properties and methods for construction. *Journal de la Société Française de Statistique*, 158 (1), 7-36.
- Pronzato L., Muller W. (2012). Design of computer experiments : space filling and beyond. *Statistics and Computing*, 22, 681-701.
- Redmond C., Yukich J.E. (1995). Asymptotics for Euclidean functionals with power-weigthed edges. *Stochastic processes and their applications. 61, 289-304.*
- Rényi A. (1961). On measures entropy and information. In Proc. 4th Berkeley Symp. On Math. Statist. and Prob., 547-561.
- Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. (1989). Design and analysis of Computer Experiments. Statistical Science, 4, 409-435.
- Shewry M.C., Wynn H.P. (1987). Maximum Entropy Sampling. *Journal of Applied Statistics*, 14, 165-170.


Discrepancy (DC2)
1550
1450
1350
AE
KL
Maximin
MC_0.5
MST_0.1
MST_0.9
NN_0.5


AE
-820
-825
-830
-835
-840
-845
AE
KL
Maximin
MC_0.5
MST_0.1
MST_0.9
NN_0.5


Maximin
4.30
4.25
4.20
4.15
AE
KL
Maximin
MC_0.5
MST_0.1
MST_0.9
NN_0.5


Coverage
0.009
0.007
0.005
AE
KL
Maximin
MC_0.5
MST_0.1
MST_0.9
NN_0.5