Space-filling designs based on Rényi entropy

Astrid Jourdan

Space-filling designs based on Rényi entropy

Astrid JOURDAN

ETIS UMR 8051, CY Paris University, 95000 Cergy, France
astrid.jourdan@cyu.fr

Abstract
Simulations using computationally intensive computer models need to be organized according to a design of experiments. Space-filling designs spread out the training examples in the experimental domain with the aim to catch the irregularities of the computer response. Among the existing space-filling designs, the uniform designs have the characteristic of having a distribution of their points close to the uniform distribution. In this paper we propose three uniformity criteria to build space-filling designs, defined from three methods of estimating the Rényi entropy: a plug-in estimation, a nearest neighbor estimation and a method based on the minimum spanning tree of the design points. An optimization algorithm is used to build optimal Latin hypercube designs. The space-filling properties of the resulting designs are studied with numerical tests.

Keywords: Uniform designs, plug-in entropy estimate, nearest neighbor estimate, minimal spanning tree estimate.

Subject classification codes: 62K05, 62G07

1. Introduction

Space-filling designs are commonly used for selecting the input values of time-consuming computer codes. Since the true relation between the computer response and inputs is not known, the design points should explore the entire experimental region, and should allow one to fit a variety of models. One strategy is to select the input values so that they are evenly spread throughout the experimental region, according to a "space-filling criterion". Many space-filling criteria have been investigated in the literature. Some of them quantify how the points fill up the space using the distance between points, such as the Maximin distance (Johnson *et al*., 1990) or the PHI-P criterion (Morris and Mitchell, 1995). Others measure the difference between the empirical distribution of the design points and the uniform distribution, such as the discrepancy (Niederreiter, 1987, Fang *et al*., 2006) or Kullback-Leibler criterion (Jourdan and Franco, 2010).

In this paper, we use the second strategy. The divergence between the empirical distribution of the design points and the uniform distribution is measured with the Rényi entropy. We suppose that the points $x_1, \ldots, x_n$ of the design D, are n independent observations of a random vector $X = (X_1, \ldots, X_d)$ with absolutely continuous density function f concentrated on the unit cube $[0,1]^d$ (we reduce the experimental space to the unit cube). Rényi entropy,

$$H_q(D) = \frac{1}{1-q} log \int_E f(x)^q dx \text{ , with } q \neq 1$$

measures the difference between f and the uniform density function in so far as, one always has $H_q(D) \leq 0$ and the maximum value of $H_q(D)$, zero, being uniquely attained by the uniform density. This latter property confirms that maximizing Rényi entropy makes f converge toward the uniform density.

The objective is to construct a design that maximizes the Rényi entropy or, more simply, that maximizes the integral,

$$I_q(f) = \int_E f(x)^q dx,$$

with $q \in]0,1[$ (or minimizes if $q > 1$).

We will say that design D_1 is better than design D_2 if $I_q(f_1) > I_q(f_2)$, where f_1 and f_2 are the density functions associated with D_1 and D_2 respectively. And we use an optimization algorithm to find the design

that maximizes the criterion. The main question is how to estimate the criterion with the design points. In the following section, we investigate three ways for estimating the entropy. Our goal is not to find a precise estimate of the entropy but to define criteria to compare the designs in the optimization algorithm in order to get closer to the uniform distribution.

When q tends to 1, $H_q(D)$ tends to the Shannon entropy. Jourdan and Franco (2010) defined a space-filling criterion based on the Shannon entropy. They used the plug-in estimator and the nearest neighbor estimator in order to obtain two criteria that can be calculated from the design points and used in an optimization process. In the same way, Pronzato and Muller (2011) suggested to use these two methods of estimation with the Rényi and Tsallis entropies. Using the Rényi entropy instead of the Shannon entropy allows to use a third estimation method based on minimum spanning tree as explained in Pronzato (2017). Next, we use the plug-in estimate, the nearest neighbor estimate and the minimum spanning tree estimate, to derive three criteria from of the Rényi entropy.

To guarantee the consistency of the following estimators, we assume that $q \in]0,1[$. The goal of this paper is not to study the properties of these estimators. For that, one can refer to the papers Pronzato (2017) and Pronzato and Muller (2011). The goal is to deduce computable criteria, to build the designs and to compare their performance with simulations.

In the next section, space-filling criteria are derived from the three estimation methods of the Rényi entropy. Section 3 is devoted to numerical to study the space-filling performances of the designs built with the new criteria. A conclusion and some future work are given in Section 4.

2. Three methods for estimating the Rényi entropy

2.1. The plug-in estimate (KERN)

The integral is the expected value,

$$I_q(f) = \int_E f(x)^{q-1} f(x) dx = E_{P_f}[f(X)^{q-1}].$$

The Monte-Carlo method gives an unbiased estimation,

$$\hat{I}_q(f) = \frac{1}{n} \sum_{i=1}^{n} f^{q-1}(x_i).$$

The unknown density function f is estimated with the design points $D = \{x_1, \dots, x_n\}$ by a kernel method with a Gaussian kernel

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^{n} (2\pi)^{-d/2} |H|^{-1/2} e^{-\frac{1}{2}(x-x_j)^T H^{-1} (x-x_j)},$$

where H is the bandwidth matrix (symmetric and positive definite matrix).

The choice of the bandwidth matrix has a great influence on the accuracy of the estimation. Joe (1989) shows that in the case where f is estimated by a kernel method, the bias in the estimation of $I_q(f)$ depends on the sample size n, the dimension d, and the bandwidth matrix H. When constructing an optimal design, the size n and the dimension d are fixed. We need to fix the bandwidth so that the bias does not vary during the optimization algorithm.

Usually the bandwidth matrix is simplified into a diagonal matrix with the Scott's rule [(1992), $H = diag(h_1^2, \dots, h_d^2)$ with $h_k = n^{-1/(d+4)} \hat{\sigma}_k$ where $\hat{\sigma}_k$ is the estimation of X_k standard deviation. The estimation $\hat{\sigma}_k$ changes at each iteration of the optimization algorithm. In order to fix the bias during the algorithm, we replace it with the standard deviation of the target (uniform) distribution,

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

Finally,

$$\hat{I}_q(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n(2\pi)^{d/2}h^d}\sum_{j=1}^{n} e^{-\frac{1}{2h^2}\|x_j - x_i\|^2}\right)^{q-1}$$

We remove the terms independent of the design points and the symmetric terms in the double sum, and we obtain the simplified criterion,

$$C_{KERN}(D) = \sum_{i=1}^{n-1}\left(\sum_{j=i}^{n} e^{-\frac{1}{2h^2}\|x_j - x_i\|^2}\right)^{q-1}.$$

2.2. The nearest neighbor estimate (NN)

Wang *et al.* (2006) and Leonenko *et al.* (2008) proposed to estimate the entropy with the k-nearest neighbor density estimation.

Let $\rho(x,y)$ denote the Euclidian distance between two points x and y of IR^d. We note $\rho^{(1)}(x,S) \leq \rho^{(2)}(x,S) \leq \cdots \leq \rho^{(m)}(x,S)$, the order distances between $x \in \mathrm{IR}^d$ and $S = \{y_1, \ldots, y_m\}$ a set of points of IR^d such that $x \notin S$. $\rho^{(k)}(x,S)$ is the k-nearest-neighbor distance from x to points of S. The previous authors demonstrated that the following estimate of $I_q(f)$ with the design points $D = \{x_1, \ldots, x_n\}$ is asymptotically unbiased and consistent,

$$\hat{I}_q(f) = \frac{1}{n}\sum_{i=1}^{n}\left((n-1)C_k V_d\left(\rho^{(k)}(x_i, D_{-i})\right)^d\right)^{1-q}$$

with $C_k = (\Gamma(k)/\Gamma(k+1-q))^{1/(1-q)}$ where Γ is the Gamma function, V_d the volume of the unit ball in IR^d and $D_{-i} = D \backslash \{x_i\}$.

The bias depends on n, d and k. We need to fix the value of k so that the bias does not vary during the optimization algorithm. Pronzato (2017) justified to restrict the estimation to $k = 1$. We remove the terms independent of the design points and we obtain the simplified criterion,

$$C_{nn}(D) = \sum_{i=1}^{n}\left(\rho^{(1)}(x_i, D_{-i})\right)^{d(1-q)}.$$

2.3. The minimum spanning tree estimate (MST)

Another way to estimate the Rényi entropy is to use the minimum spanning tree of the design points.

The tree is constructed by connecting the points of the design by edges $(e_{i,j})$ such that:

- only one edge connects two points,
- only one path allows to go from one point to another,
- there is no cycle,
- all the points of the design are connected,

and such that the sum of the lengths of the edges (Euclidean norm),

$$L_\delta(D) = \sum_{e_{i,j}} \|e_{i,j}\|^\delta, \ \delta \in]0,d[,$$

is minimal.

Redmond and Yukich (1996) and Hero and Mitchel (1999) showed that

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

is an asymptotically unbiased and almost surely consistent estimator of the entropy, where β is a constant bias correction independent on f. Maximizing the above estimator is equivalent to maximize the simplified criterion

$$C_{MST}(D) = \sum_{i=1}^{n} L_{d(1-q)}(D).$$

The idea of using minimum length trees to assess the distribution of points in a multidimensional space is not new. In 1984, Smith and Jain defined a multivariate uniformity test based on this structure. This idea can also be found in the comparative study of different topographic analysis methods by Wallet and Dussert (1998). Franco *et al.* (2009) presented a nice graphic tool based on the empirical mean and standard deviation of the edge lengths of the MST to compare space-filling designs. Finally, as mentioned in Pronzato (2017), the sum of power-weighted edge lengths for the MST has never been used as a criterion for space-filling designs.

In the next section, we compare the new MST criterion with the two criteria defined above. We can already note that the complexity of the MST criterion ($O(d\times n\times(n+1)/2) + O(n^2\times\log(n))$) is greater than that of the KERN ($O(d\times n\times(n-1)/2)$) or NN criterion ($O(d\times n\times(n-1))$) because of the Kruskal algorithm used to build the minimum spanning tree.

3. Numerical tests

In this section we compare the three criteria described previously for q=0.1, 0.5 and 0.9. The idea is to use an optimization algorithm to build designs that maximize the criteria and compare the space-fillling performances of the resulting designs. It is well-known that space-filling criteria tend to push the design points on the boundaries of the unit cube as d increases (curse of dimensionality). A common strategy to overcome this problem is to use Latin hypercube designs (LHD). Each column of a d-dimensional LHD of n points is a random permutation of {1,2,…,n}. This property ensures that each variable is tested n times regularly between 0 and 1 (and not only at the edges), but does not ensure a good spatial distribution of the points in the unit cube. It is necessary to optimize an LHD with a space-filling criterion. Many algorithms have been developed to optimize LHD. We use the enhanced stochastic algorithm (ESE) defined by Jin *et al.* [17] with the same settings. In the numerical tests, we build 20 Latin hypercube designs with d=2 and n=20, d=4 and n=40, and d=10 and n=100. We compare both the uniformity of the distribution of the design points and their inter-site distance.

3.1. Uniformity of the distribution of the design points

A common method to assess the space-filling properties of a design is to use the discrepancy, that is a measurement of the difference between the cumulative function of the uniform distribution and the cumulative function of the distribution of the design points (Fang *et al.*, 2006). Many ways to calculate the discrepancy have been defined in the literature. Figure 1 give the boxplots of the centered L2-discrepancy,

$$DISCL2(D) = \left(\frac{13}{12}\right)^d - \frac{2}{n}\sum_{i=1}^{n}\prod_{k=1}^{d}\left(1+\frac{1}{2}|x_{ik}-0.5| - \frac{1}{2}|x_{ik}-0.5|^2\right) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\prod_{k=1}^{d}\left(1+\frac{1}{2}|x_{ik}-0.5| + \frac{1}{2}|x_{jk}-0.5| - \frac{1}{2}|x_{ik}-x_{jk}|\right)$$

but the results are quiet the same with the L2-discrepancy and the wrapped discrepancy. The aim is to minimize the discrepancy or maximize its opposite value as in Figure 1.

The choice of q value has a big impact on the discrepancy for NN and MST designs. The designs defined with q=0.1 have a high discrepancy value with a high variability. Thus, with q=0.1, the distribution of the design points is further from the uniform distribution (in the sense of discrepancy) than the distribution obtained with q=0.5 or q=0.9. The high variability implies that the discrepancy of the design strongly depends on the design initialization in the optimization algorithm.

The distribution of the points of the KERN designs seems to be less impacted by the value of q. Whatever q is, the points of the KERN planes are as uniformly distributed as the points of the NN and MST designs with q=0.5 and q=0.9.

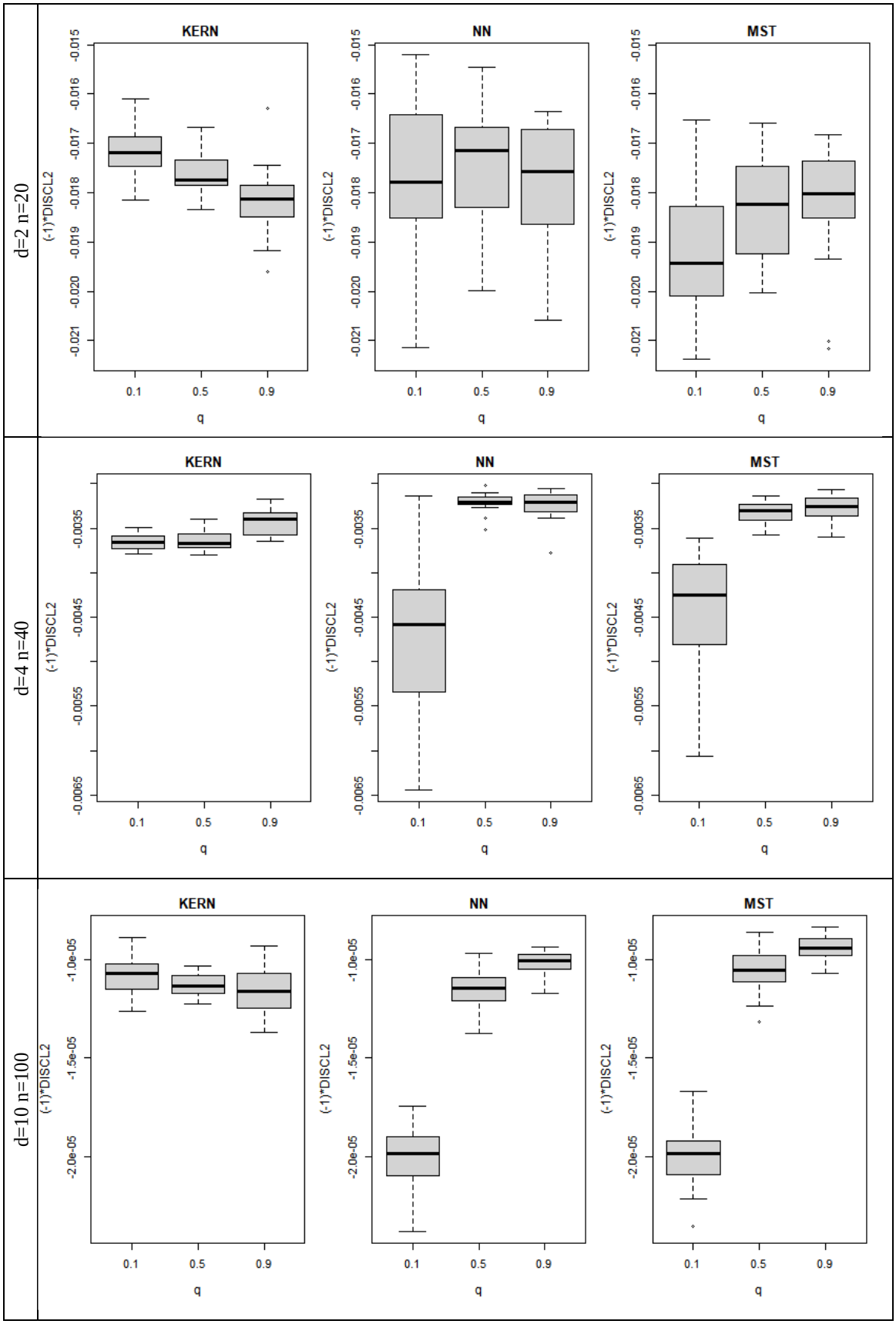


Fig. 1. Boxplot of (-1)*DISCL2 criterion for KERN, NN and MST designs.

3.2. Inter-site distance of the design points

Another way to assess the space-filling properties of a design is to study the distance between the design points (Johnson *et al.*, 1990). Here, we use the PHI-P criterion with p=50 (Figure 2) as proposed by Morris and Mitchell (1995),

$$\Phi_p(D) = \left(\sum_{i=1}^{n-1} \sum_{j>i}^{n} \left\| x_i - x_j \right\|^{-p} \right)^{1/p},$$

In order to compare the inter-site distance of the design points, we add a graphical tool (Figure 3). For a design, we compute the nearest neighbor distance of each point. The x-axis is the average of the nearest neighbor distances of the design points (μ) and the y-axis is the standard deviation (σ). A good coverage of the experimental region is obtained by a design with points far from each other (high average) and close to a regular grid (small standard deviation) like a scrambled grid. Then the target area is at the bottom right of this graphic.

As for discrepancy, we can see that the PHI-P distance criterion is rather little impacted by the value of q for KERN designs except for the variability in dimension d=2. However, we can see in Figure 3 (especially in dimension d=4 and d=10) that the average of the minimum distances between two points increases when q decreases.

The NN and MST designs obtained with q=0.1 also perform the worst in terms of PHI-P criterion or minimum distances. The NN and MST designs constructed with q=0.9 give the best results. The PHI-P criterion values are almost the same for KERN designs and NN and MST design with q=0.9 (Figure 2). However, the minimum distance between points (Figure 3) is greater for the NN and MST designs for q=0.5 and q=0.9 than for the KERN designs. The NN and MST designs with q=0.5 have the highest average of the nearest neighbor distances but some designs have a high standard deviation. This means that globally the points are far from each other in the design, but that some points are close. The NN and MST designs with q=0.9 have a slightly lower average but with a very small standard deviation. This means that the points are almost all equidistant.

4. Conclusion

In this paper we used the Rényi entropy to measure the uniformity of the point distribution in an experimental design. From three methods of estimating the entropy, we proposed three criteria that can be computed directly from a set of points. The criteria are used in an optimization algorithm to build space-filling Latin hypercubes.
Numerical tests allow to establish that the performance depends on the value of q for the criteria based on the minimal spanning tree and the nearest neighbor distance. The space-filling performances are better when q is close to 1, i.e. when the Rényi entropy is close to the Shannon entropy.

In order to avoid the curve of dimensionality (and thus points pushed to the edges of the experimental domain), we used a Latin hypercube structure to build the space-filling designs. This guarantees to have points inside the experimental domain and an equidistribution of the points on the factorial axes, but it constrains the optimization algorithm. An alternative idea is to use the additivity property satisfied by the Rényi entropy. The estimate of the entropy in dimension d can be reduced to the sum of the d estimates of the entropies of the projections of the points on factorial axis. Maximizing the sum of entropies should ensure a good spatial distribution and a good point distribution on each factorial axe. One idea would be to use a multi-objective algorithm to ensure that each entropy is maximized in the sum. The first objective function would be the sum of the entropies and the second objective function would be the uniformity of the entropy values on the axis. Another advantage of using the sum of entropies is that it allows to consider other methods for univariate density estimation like Fourier series (Sylverman, 1986) or histogram-based estimate Barron *et al.* (1992).

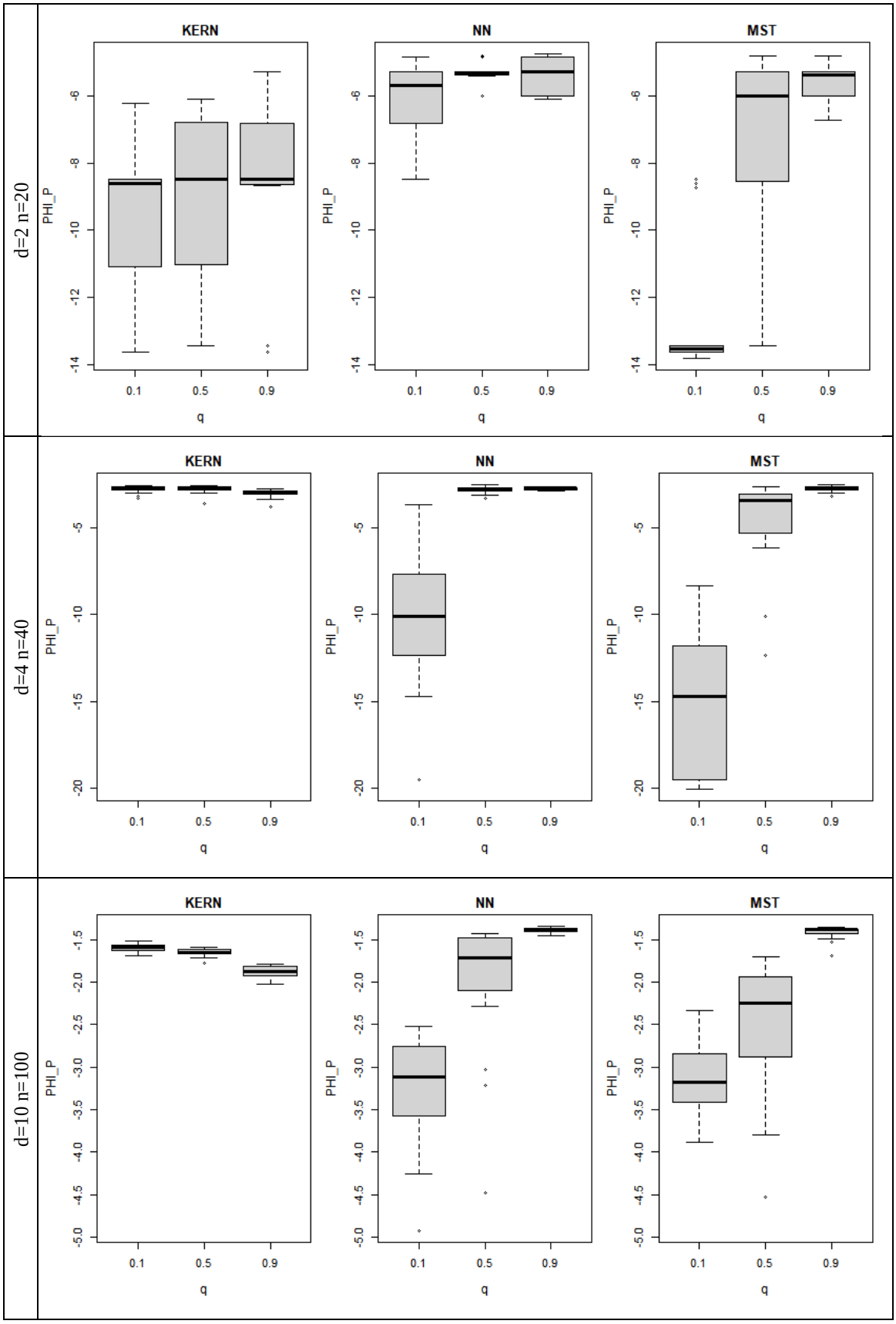
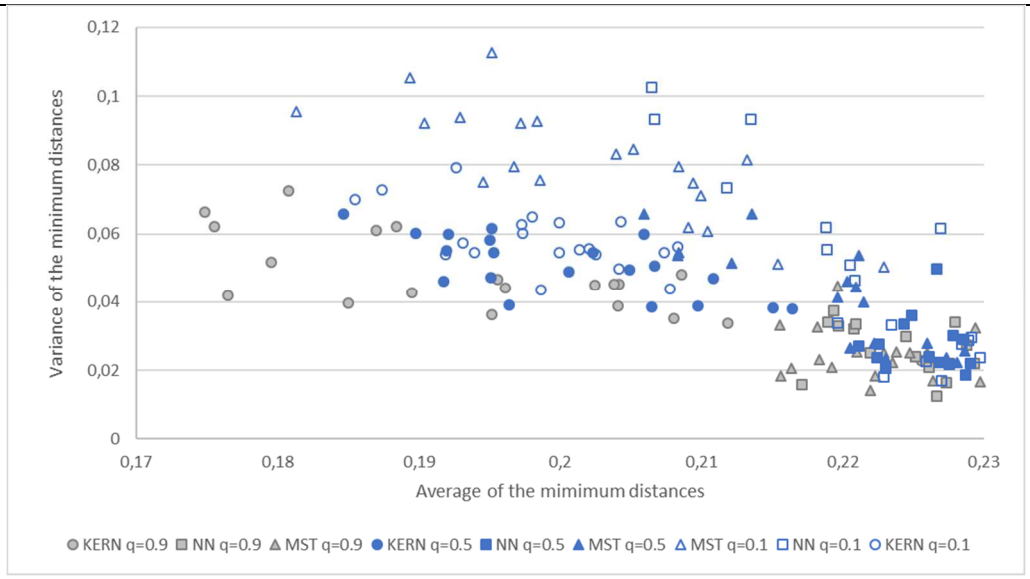
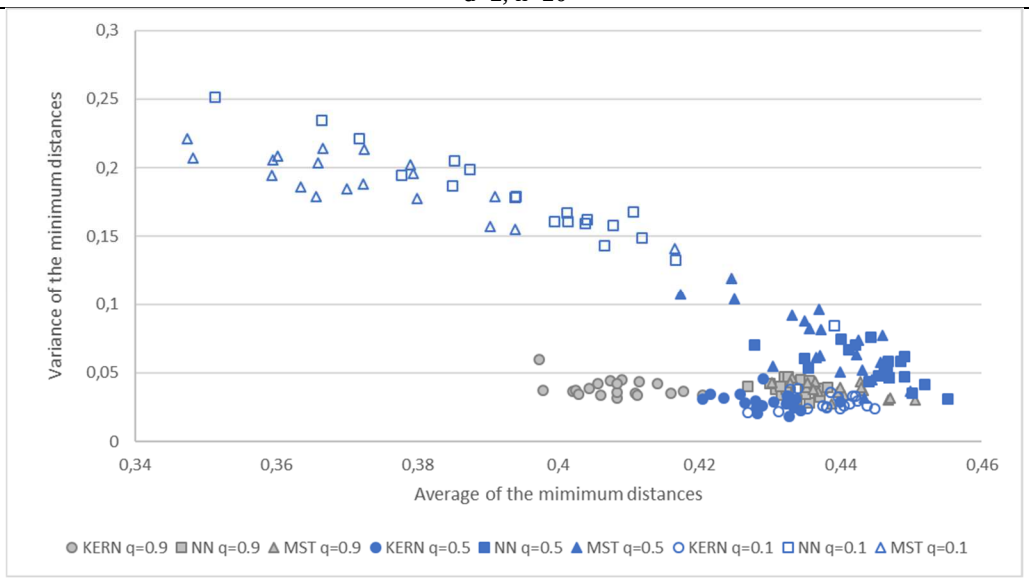


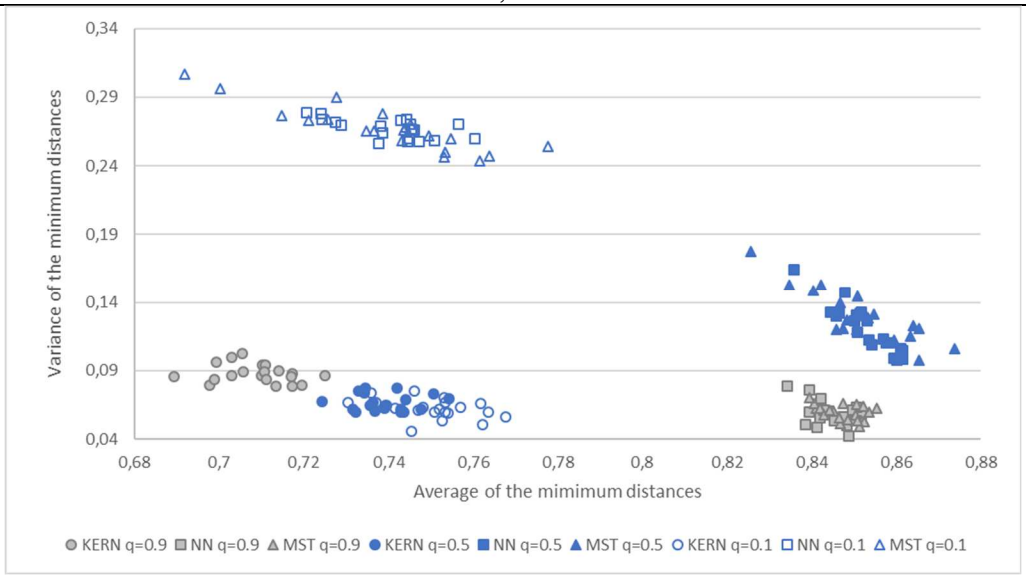
Fig. 2. Boxplot of PHI-P criterion for KERN, NN and MST designs.



d=2, n=20



d=4, n=40



d=10, n=100

Figure 3. Average (μ) and standard deviation (σ) the of the nearest neighbor distances of the design points.

References

[1] Barron A.R., Györfi L., Van Der Meulen E.C. (1992). Distribution estimation consistent in total variation and two types of information divergence. *IEEE Trans. Inform. Theory*, 5, 1867-1883. DOI: 10.1109/18.149496

[2] Fang K.T., Li R., Sudjianto A (2005). *Design and modeling for computer experiments*. Chapman&Hall, London. DOI: 10.1201/9781420034899

[3] J. Franco, O. Vasseur, B. Corre, M. Sergent. Minimum Spanning Tree (2009). A new approach to assess the quality of the design of computer experiments. *Chemometrics and Intelligent Laboratory Systems*, 97, 164-169. DOI: 10.1016/j.chemolab.2009.03.011

[4] Jin R., Chen W., Sudjianto A. (2005). An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*, 134, 268-287. DOI: 10.1016/j.jspi.2004.02.014

[5] Joe H. (1989). Estimation of entropy and other functional of multivariate density. *Annals of the Institute of Statistical Mathematics*, 41, 683-697. DOI: 10.1007/BF00057735

[6] Johnson M.E., Moore L.M., Ylvisaker D. (1990). Minimax and maximin distance design. *J. Statist. Plann. Inf.*, 26, 131-148. DOI: 10.1016/0378-3758(90)90122-B

[7] Jourdan A. et Franco J. (2010). Optimal Latin hypercube designs for the Kullback-Leibler criterion. *AStA Advances in Statistical Analysis*, 94(4), 341-351. DOI: 10.1007/s10182-010-0145-y

[8] Leonenko N., Pronzato L.,Savani V. (2008) A class of Rényi information estimators for multidimensional densities, *Ann. Statist.* 36, 2153-2182. Correction by Leonenko and Pronzato 2010, *Ann. Statist*., 38, 3837–3838, 2008. DOI: 10.1214/07-AOS539

[9] Morris, M., Mitchell, T. (1995). Exploratory designs for computational experiments. *J. Stat. Plan. Inference* 43, 381–402. DOI: 10.1016/0378-3758(94)00035-T

[10] Niederreiter H. Point sets and sequences with small discrepancy (1987). *Monasth. Math.*, 104, 273-337. DOI: 10.1007/BF01294651

[11] Hero O. and Michel O.J.J., (1999). Asymptotic theory of greedy approximations to minimal k-point random graphs, *IEEE Transactions on Information Theory*, 45(6), 1921-1938. DOI 10.1109/18.782114

[12] Pronzato L. (2017). Minimax and maximin space-filling designs: some properties and methods for construction. *Journal de la Société Française de Statistique*, 158(1).

[13] Pronzato L., Muller W. (2011). Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3), 681-701. DOI 10.1007/s11222-011-9242-3

[14] Redmond, C., Yukich, J. (1996). Asymptotics for Euclidian functionals with power-weighted edges. *Stoch. Process. Appl*., 61, 289–304 (1996) DOI: 10.1016/0304-4149(95)00075-5

[15] Scott D.W. (1992). *Multivariate Density Estimation : Theory, practice and visualization*, John Wiley & Sons, New York, Chichester. ISBN: 978-0-471-69755-8

[16] Smith SP, Jain AK. Testing for uniformity in multidimensional data. IEEE Trans Pattern Anal Mach Intell. 1984 Jan;6(1):73-81. doi: 10.1109 /tpami.1984.4767477

[17] Wallet F., Dussert C. (1998). Comparison of spatial point patterns and processes characterization methods. *Europhysics Lett.*, 42, 493-498. DOI: 10.1209/epl/i1998-00279-7

[18] Q. Wang, S. R. Kulkarni and S. Verdu (2006). A Nearest-Neighbor Approach to Estimating Divergence between Continuous Random Vectors, *IEEE International Symposium on Information Theory*, Seattle, WA, USA, 2006, 242-246. DOI: 10.1109/ISIT.2006.261842.