A RADAR-SHAPED STATISTIC FOR TESTING AND VISUALIZING UNIFORMITY PROPERTIES IN COMPUTER EXPERIMENTS

Working Paper, March 2007

Jessica Franco[1], Laurent Carraro[1], Olivier Roustant[1] et Astrid Jourdan[2].

[1]Département 3MI Ecole Nationale Supérieure des Mines, Saint-Etienne, nom@emse.fr.
[2]Ecole Internationale des Sciences du Traitement de l'Information, Pau, astrid.jourdan@eisti.fr.

In the study of computer codes, filling space as uniformly as possible is important to describe the complexity of the investigated phenomenon. However, this property is not conserved by reducing the dimension. Some numeric experiment designs are conceived in this sense as Latin hypercubes or orthogonal arrays, but they consider only the projections onto the axes or the coordinate planes. In this article we introduce a statistic which allows studying the good distribution of points according to all 1-dimensional projections. By angularly scanning the domain, we obtain a radar type representation, allowing the uniformity defects of a design to be identified with respect to its projections onto straight lines. The advantages of this new tool are demonstrated on usual examples of space-filling designs (SFD) and a global statistic independent of the angle of rotation is studied.

KEYWORDS: Computer experiments ; Space-Filling Designs ; Dimension Reduction ; Discrepancy ; Kolmogorov-Smirnov Statistic, Cramer-Von Mises Statistic

1. INTRODUCTION

For over 15 years the theory of experiment designs initiated by Fisher (1926) has experienced a revival with its use for the investigation of large-scale industrial computer codes. This change in setting has led to at least two major changes. First, the large-scale computer codes describe phenomena of an increasing complexity, which implies that the corresponding models are often nonlinear and/or nonparametric. Second, the experiment itself is different. Numeric experiments are simulations and, except for stochastic codes implementing a Monte Carlo-based method, they give the same response for identical experimental conditions (including algorithmic and computer-based parameters). Therefore, repeating an experiment under the same conditions doesn't make sense since it doesn't help in acquiring new information.

In this new setting, the experiment planning methods are therefore different, For example when the code is to be remodeled in the scanning phase (before any simulation has been realized), one often tries to satisfy the following two requirements. On the one hand, distribute the points in the space as uniformly as possible to catch non-linearities; this also allows avoiding repetitions. On the other hand, try to make this space filling last by reducing the dimension. The first requirement was the starting point of researches in space filling designs (SFD). The quality of the spatial distribution is measured either by using deterministic criteria like minimax or maximin distances (Johnson, Moore and Ylvisaker, 1990), or by using statistical criteria like discrepancy (Niederreiter, 1987, Hickernell, 1998, Fang, Li, Sudjianto, 2006). The second requirement stems from the fact that it is frequently observed that the code depends only on a few influential variables, which may be either direct factors or "principal components" made up of linear combinations of these variables. It is interesting to note that dimension reducing techniques like SIR (Li, 1991) or KDR (Fukumizu, Bach, Jordan, 2004) allow effectively identifying the space generated by these main directions. It is therefore desirable that the space filling property be also satisfied *in the projection onto subspaces*. This idea has motivated the use of Latin hypercubes (LH) and orthogonal arrays (OA) in numeric design experiments. With a LH design of the SFD type, one can guarantee a very good distribution in the projection onto margins, so that there is no loss of information if the code depends on only one variable. SFD type OAs extend this property to the projection onto marginal subspaces of higher dimension (see, for example, Koehler and Owen, 1996, and more generally, Owen, 1992 or Santner, Williams, Notz, 2003). However, considering only the projections onto margins is not sufficient if, for example, the code is a function of a linear combination of 2 variables.

In this article we introduce a statistic built around 1-dimensional projections to test the uniformity of an experiment design. The choice of 1 dimension is linked to the difficulty of obtaining the theoretical distribution of the projections onto a higher dimensional space. However, the advantage of the restriction to 1 dimension is that it offers a simple viewing tool based on the principle of a radar. By representing the statistic's value in all directions, one obtains a parameterized curve (or a surface), allowing the uniformity defects of a design to be identified with respect to its projections onto straight lines. The article is structured as follows. In section 2 we define the new statistic and the associated viewing tool, called uniformity radar, and give a few properties. In section 3 we show one application of the radar to SFD designs. Section 4 is devoted to an enhancement, where we define a global statistic which doesn't depend on a particular axis of rotation. In section 5 we give some conclusions.

2. UNIFORMITY RADAR

In the investigation of a computer code, let us consider a uniform experiment design on a cubic domain $\Omega = [-1,1]^d$. Note $x_1,...,x_N$ the experimental points, and (H_0) the hypothesis " $x_1,...,x_N$ were generated by independent random drawings according to the uniform distribution in Ω ". If the computer code depends only on one single main direction, it is important that the projections on this axis be distributed in the best possible way. More generally, let us denote L_a the straight line generated by the unitary vector $a = (a_1, ..., a_d)$ of Ω, and μ_a the probability distribution of the projections of $x_1, ..., x_N$ onto L_a. Ideally, we may therefore expect that in all the directions a the distribution μ_a is uniform. However, this is not realistic when L_a is not a coordinate axis. For example, in the case of the cubic domain $[-1,1]^2$, there will be a larger number of projected points in the central part of the axis of projection as can be seen in Figure 1.

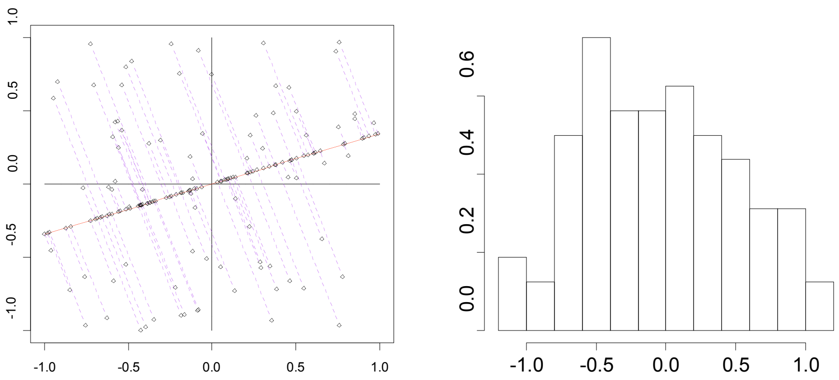
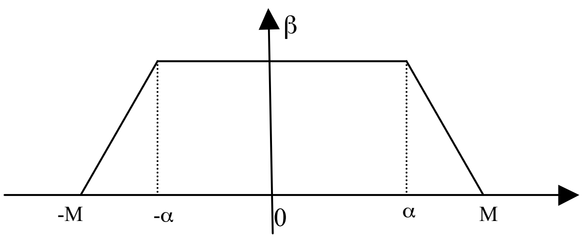


Figure 1. Left : Projections of points onto an axis L_a. Right : The histogram of the projections.

More specifically, the distribution of the projected points may be subdivided into 3 areas defined by the projection of the apexes of the domain. Actually, the distribution μ_a is continuous, with probability density represented below. The apexes of the trapezoid correspond to the apexes of the square Ω projected onto the axis L_a, where $a = (\cos\theta, \sin\theta)$.



where :

M = |cosθ| + |sinθ|
α = | |cosθ|-|sinθ| |
$\beta = \dfrac{1}{M+\alpha}$

Figure 2. The distribution of the projections for a 3 dimensional cubic domain.

In the general case, the projection onto L_a is a linear combination of independent random variables of a uniform distribution, which reduces the calculating of its distribution to a traditional problem of probabilities first investigated in the 18th century by Lagrange (see discussion in Elias and Shiu (1987)). If we denote $X = (X_1,...,X_d)$, where $X_1,...,X_d$ are independent random variables identically distributed following the uniform distribution over $[-1,1]$, and Z the projection of X onto L_a, where $a_j \neq 0\ \forall j \in \{1,...,d\}$, then the distribution function of Z is given by :

$$F_Z(z) = \left(\prod_{j=1}^{d} \frac{1}{2a_j} \right) \times \sum_{s \in \{-1,1\}^d} \varepsilon(s) \frac{(z + s.a)_+^d}{d!}$$

where $s = (s_1,...,s_d) \in \{-1,1\}^d$ are the apexes of the hypercube $\Omega = [-1,1]^d$, $\varepsilon(s) = \prod_{j=1}^{d} s_j$, $s.a$ is the scalar product of the vectors s and a, and $(y)_+ = \max(y,0)$ the positive part of y. As a result, for a given axis, Z accepts a continuous density in bits whose nodes correspond to projections onto the axis of the apexes of the domain.

It is interesting to note that the distribution of the projections is known in other situations, as in the case of a spherical domain : if Ω is the unit sphere of R^d, a direct calculation shows that μ_a accepts the density $f_a(x) = \frac{2}{\pi}\sqrt{1-x^2}\, 1_{[-1,1]}(x)$, the distribution function being equal to $F_a(x) = \frac{1}{2} + \frac{1}{\pi}(\text{Arc}\sin x + x\sqrt{1-x^2})$ for $x \in [-1,1]$.

In sum, for a uniform experiment design to have good distribution properties on the 1-dimensional projections, it will be necessary that in all the directions a, the empirical distribution of the projections onto L_a is close to their theoretical distribution under the hypothesis (H_0). There exist many distribution adequacy statistics (see D'agostino and Stephens, 1986), which allow for a large number of choices to define a criterion adapted for this purpose. However, possibilities are limited by special requirements. To start with, it is preferable that the statistic's distribution be known to avoid the approximate calculation of its distribution. Furthermore, one would also like the statistic to be distribrution-free, that is, its distribution doesn't depend on the projection direction to have a unique rejection threshold for all the angles. Also, for the sake of consistency, it would be desirable for the retained statistic to be interpretable in terms of discrepancy when projections are made onto a coordinate axis.

Finally, two statistics (at least) correspond to these requirements : the Kolmogorov-Smirnov statistic

$$D_N(a) = \sup\left|\hat{F}_{N,a}(z) - F_a(z)\right| \qquad (1)$$

and the Cramér-Von Mises statistic

$$N\omega_N^{\;2}(a) = \int (\hat{F}_{N,a}(z) - F_a(z))^2 dz \qquad (2)$$

where $\hat{F}_{N,a}$ is the empirical distribution function of the projections of $x_1,...,x_N$ onto L_a and F_a the distribution function of μ_a . When L_a is a coordinate axis, μ_a is the uniform distribution on [0,1] and these statistics correspond, respectively, to the discrepancies L_∞ and L_2 (Niederreiter, 1987). In what follows, we decided to work on the first because the conclusions seem equivalent, while the corresponding graphics are a little more readable (see section 5). By analogy with the case of coordinate axes, we will talk about a *discrepancy of projections* to designate the Kolmogorov-Smirnov statistic of the formula (1).

The discrepancy of projections provides a viewing tool of the uniformity defects based on the projections. This tool is built around the principle of a radar and, for this reason, we propose to call it *uniformity radar*. Its usage varies according to the dimension of the space Ω on which the uniformity of the points are to be verified. In 2 dimensions, the discrepancy of projections is calculated in all directions by making an angular scan in Ω . Thus, a parameterized curve will be obtained, called *2D radar*, with an equation in polar coordinates

$$\theta \mapsto D_N(\theta)$$

defined over [0,2π], allowing a good distribution of the points to be displayed in all directions and to decide whether the design is appropriate or not. In 3 dimensions, one calculates the discrepancy of the projected onto an axis $L_{\theta,\varphi}$ pivoting around the center of the domain and defined in spherical coordinates by an angle θ in longitude and φ in latitude. This time a parameterized surface is obtained, called *3D radar*,

$$(\theta,\varphi) \mapsto D_N(\theta,\varphi)$$

defined over $[0,2\pi]\times\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$. In higher dimensions, it is unthinkable to try to make an angular scan of the space Ω and especially because it becomes impossible to represent the result graphically (although the calculation is still possible). However, the hypothesis (H_0)

remains valid on 2- and 3-dimensional coordinate spaces. Therefore, uniformity radar may be applied to all pairs and/or triplets of possible dimensions.

In practice, the quality of the representation can be degraded by discretization. Here it can be shown that the 2D and 3D radars are continuous applications. But D_N is not differentiable on all the axes L_a so that at least two points of Ω are projected onto the same point, which explains why there are many visible breaks in the graphics of the next sections.

3. APPLICATIONS OF UNIFORMITY RADAR

In this section we present an example to focus on the interest of uniformity radar for testing the good distribution of experimental points in projection. As required by the scope of application, we refer to cases where the hypothesis (H_0) of a uniform distribution in the experimental domain is plausible. For each representation of the uniformity radar we have added the circle (or the sphere) of radius *ks* equal to the statistic of the Kolmogorov-Smirnov test associated with a level of confidence at 95%. Recall that since the statistic is distribution free, *ks* doesn't depend on *a*. This provides a decision-making element or, at least, a means of comparison with the random designs obtained by a uniform drawing. Should the studied design be stochastic (pseudorandom, Latin hypercubes or randomized orthogonal arrays), the graph displays the directions *a* along which the hypothesis (H_0) is rejected. If the studied design is deterministic (due to low discrepancy), we are no longer in the usual scope of application of the test. If one of the values of $D_N(a)$ is greater than *ks*, then we can only say that the investigated design is less good than a random design in the sense where the probability that a random design has a better discrepancy exceeds 95%. The following examples apply essentially to those cases because we preferred known SFD designs without transforming them. Nonetheless, in practice, it would suffice to apply randomization or scrambling (see, for example, Fang, Li, Sudjianto, 2006) to return to the usual application conditions of statistics tests.

Example 1 : Detection of holes in 15-dimensional Halton sequences using 2D radar. We consider the first 250 points of a 15-dimensional Halton sequence of low discrepancy (1960). Since the design is a high dimensional design, we apply the radar to all pairs of possible dimensions. Among the rejected pairs we have, for example, the pair (14,15), which may be viewed on Figure 3.

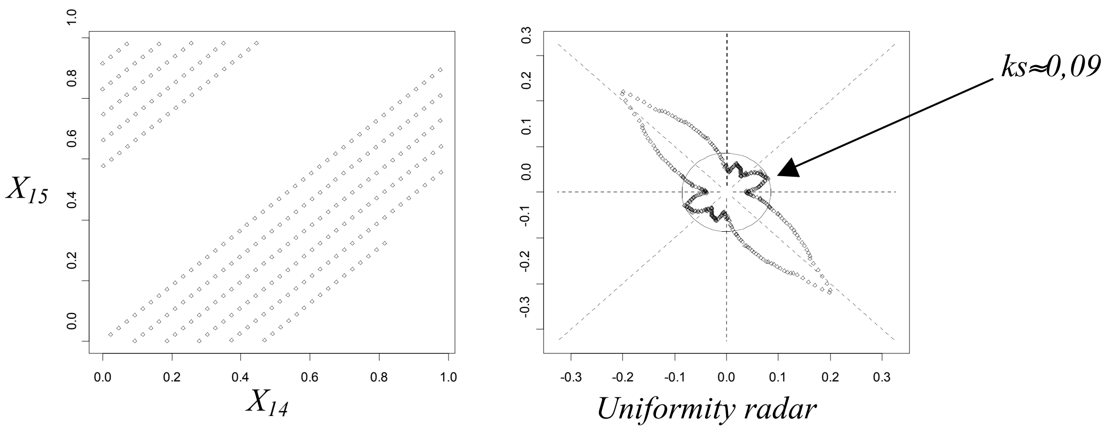


Figure 3. Left : The design (X_{14},X_{15}) of a Halton sequence. Right : The curve obtained using radar.

In this example, since there exist values of $D_N(a)$ outside the circle of radius *ks,* the uniformity radar detects the bad uniform distribution in the design $\left(X_{14}, X_{15}\right)$. The largest deviation in the uniformity is observed in direction a associated with an angle of approximately 135°, corresponding here to the direction orthogonal to the visible alignments on the figure to the left. Here we find ourselves faced with the well-known defect of high-dimensional Halton sequences, which do not preserve a low discrepancy in projection (Thiémard (2000), Morokoff, Caflisch (1994)). It is interesting to note, however, that the radar is not designed to systematically detect the directions of alignment.

4. A GLOBAL STATISTIC FOR 2D RADAR

Example 2. Toward an extension of the uniformity radar. Let us consider the first 100 points of an 8-dimensional Halton sequence projected onto the subspace formed by $\left(X_3, X_6\right)$.

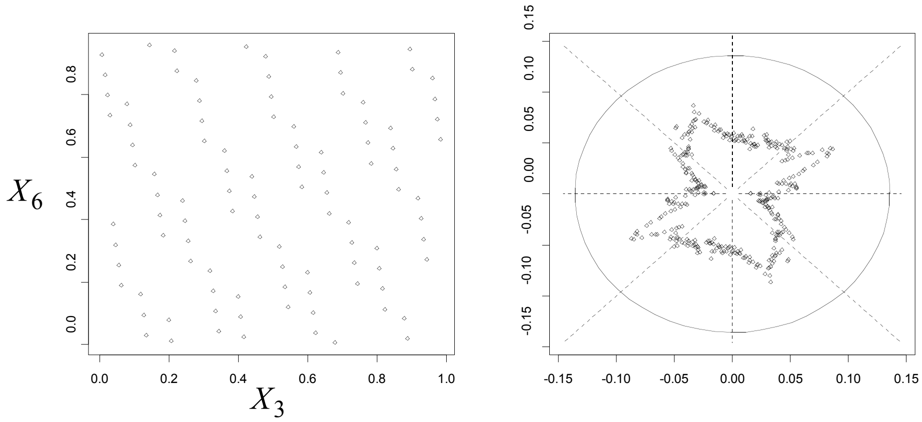


Figure 4. The first 100 terms of an 8-dimensional Halton sequence projected onto (X_3, X_6).

We note that all the points of the uniformity radar are inside the circle of radius *ks* and, as a result, the radar accepts the plane although we can see that the points of the plane (X_3, X_6) are not uniformly distributed. However, the discrepancy values are rather scattered with a low value for angle $\theta = 0°$, and rather high values, for example, for angle $\theta = 29°$, which seems to correspond to the orthogonal direction of alignments. The idea to reject this type of design amounts, therefore, to defining a new statistic which introduces minimum and maximum discrepancies. In order to avoid scale problems, we suggest taking the ratio of these quantities. We have:

$$G_N = \frac{\sup_{\theta \in [0, 2\pi]} D_N(\theta)}{\inf_{\theta \in [0, 2\pi]} D_N(\theta)} \quad (3)$$

This statistic should allow rejecting designs which have a bad distribution in one direction compared to a direction where the points will be uniformly distributed. This statistic has the advantage of being global, that is, only its value allows accepting or rejecting a design, while up to now a statistic test was available for each value of θ, which may be criticizable at the decision-making level. For a fixed value of N, the distribution of G_N seems difficult to obtain other than by simulation. For a 100 points design, we obtain a threshold equal to 4.23 at a 95% level. In the case of example 2, the observed value of the statistic G_N is equal to 6.07, which allows very clearly deciding to reject this design.

5. CONCLUSION

Due to the complexity of the phenomena described by large-scale computer codes, it is preferable to distribute numeric experiments as uniformly as possible in the domain. In addition, the distribution should also be satisfactory in projections onto subspaces in case the code would depend only on a small number of factors or main components. In this article we introduced a statistical criterion to allow generalizing the use of the discrepancy L_∞ to projections onto all the 1-dimensional subspaces. This criterion is called uniformity radar because it allows graphically testing in 2 and 3 dimensions the uniformity hypothesis of the design plane by scanning in all directions. We also introduced a global statistic in 2 dimensions to help reject unsatisfactory designs which are still accepted by the uniformity radar.

The interest of these criteria was studied on usual SFD designs. In these examples (only one is presented here), the uniformity radar was able to detect the main defects of these designs, including some with a very bad behavior in projection, such as the 15-dimensional sequence with low discrepancy in example 1. It is obviously not enough to check the 1-dimensional projections for the detection of detects in higher dimensions. The radar can succeed when there is a rectangular shaped empty area in the domain, as in the aforementioned example. In such a case, the distribution of the points is bad when the points are projected onto the rectangle's width. It can also detect the alignments of points, but cannot detect them if the directions of alignments are well distributed such as with the factorial design. This underscores the lack of power of the Kolmogorov-Smirnov test when the sample is generated in reality from a continuous distribution supported by the union of small intervals regularly distributed. In practice, this situation is not very detrimental because the SFD experiment designs obtained by a deterministic process are often randomized or scrambled (see, for example, Fang, Li and Sudjianto, 2006).

The uniformity radar may be adapted to other distribution adequacy statistics, such as the Cramér-Von Mises statistic (see section 2), which corresponds to the discrepancy L_2, for a projection onto a coordinate axis. For example, we repeated the examples with the corresponding radar. As expected, the conclusions are exactly the same because the Kolmogorov-Smirnov and Cramér-Von Mises tests do not present any clear-cut difference in terms of power. The main difference is graphic: the curve of the radar defined with the Cramér-Von Mises statistic is smoother, which is due to the norm L_2, and introduces sometimes large scale variations from one plane to another, while these differences are attenuated by the norm L_∞ in the examples given here.

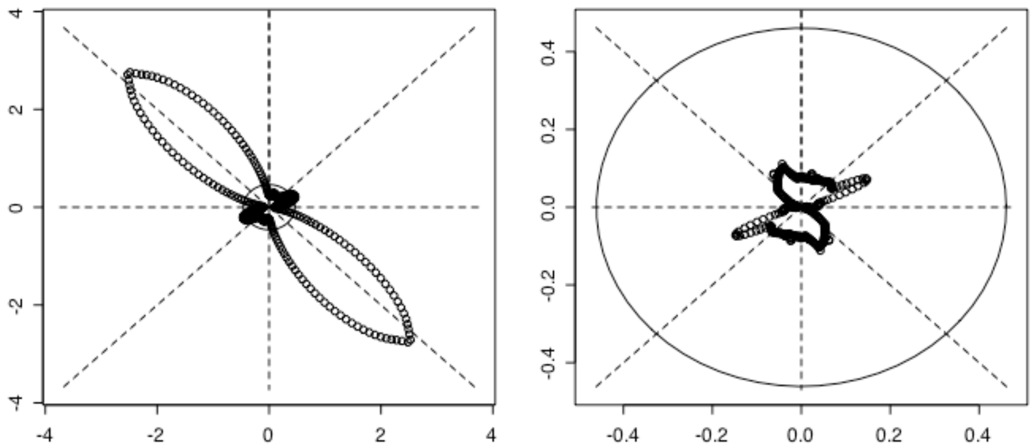


Figure 5. Uniformity radar with the Cramér-Von Mises statistic for examples 1 and 2.

ACKNOWLEDGMENTS

This work was sponsored by the DICE consortium (ARMINES STSIM 70106, 70108, 70111, 70113, 70115)

REFERENCES

D'agostino R.B., Stephens M.A. (1986). *Goodness-of-fit Techniques*. Marcel Dekker, New-York.

Elias S.W., Shiu W. (1987). Convolution of Uniform Distributions and Ruin Probability. *Scandinavian Actuarial*, 191-197.

Fang K.-T., Li R., Sudjianto A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall.

Fisher, R.A. (1926). The arrangement of field experiments. *J. Ministry Agric*. **33**, 503-513.

Fukumizu K., Bach F.R., Jordan M.I. (2004), Dimension Reduction for Supervised Learning with Reproducing Kernel Hilbert Spaces, *Journal of Machine Learning Research*, **5**, 73-99.

Halton, J.H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math,* **2**, 84-90.

Hickernell, F. (1998). A generalized discrepancy and quadrature error bound, *Mathematics of computation*, **67**, 299-322.

Koehler, J.R. et Owen, A.B. (1996). Computer Experiments, *Handbook of Statistics*, **13**, 261-308.

Jourdan A. (2000). Analyse statistique et échantillonnage d'expériences simulées, Université de Pau et des Pays de l'Adour.

Li K-C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association*, **86**, 316-342.

Morokoff, W.J., Caflisch R.E. (1994). Quasi-random sequences and their discrepancies. *S.I.A.M. Journal Scientific Computing,* **15**, 1251-1279.

Niederreiter, H. (1987). Low-Discrepancy and Low-Dispersion Sequences, *Journal of number theory,* **30**, 51-70.

Ostrowski A.M. (1952). Two Explicit Formulae for the Distribution Function of the Sums of n Uniformly Distributed Independent Variables, *Arch. Math*, **3**, 451-459.

Owen A.B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statisitica Sinica* **2**, 439-452.

Santner T.J., Williams B.J., Notz W.I. (2003). *The Design and Analysis of Computer Experiments*, Springer.

Thiémard E. (2000). Sur le calcul et la majoration de la discrépance à l'origine. Thèse No 2259, Département de mathématiques, école polytechnique fédérale de Lausanne.