Kullback-Leibler Designs

Astrid JOURDAN

EISTI
Ecole Internationale des Sciences du Traitement de l'Information

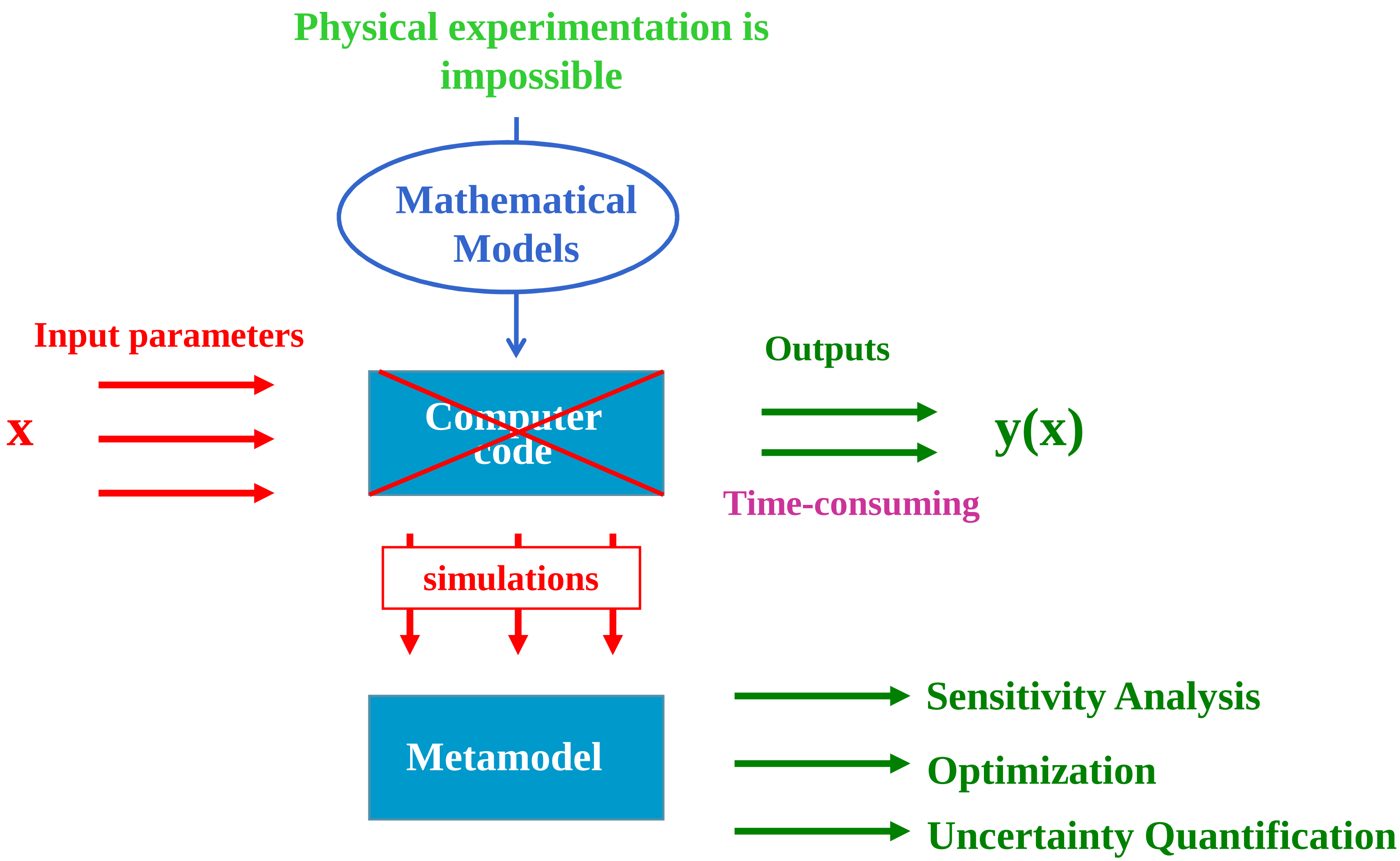
Jessica FRANCO

TOTAL

Contents

- Introduction
- Kullback-Leibler divergence
- Estimation by a Monte-Carlo method
- Design comparison
- Conclusion

Computer experiments



Design constraints

- No replication, in particular when projecting the design on to a subset of parameters (non-collapsing)
- Provide information about all parts of the experimental region

→ Space filling designs

- Allow one to adapt a variety of statistical models

→ Exploratory designs

Goal : fill up the space in uniform fashion with the design points

Kullback-Leibler Divergence

- Introduction
- Kullback-Leibler divergence
- Estimation by a Monte Carlo method
- Design comparison
- Conclusion

Goal

Suppose that the design points $X_1,\ldots,X_n$, are *n* independent observations of the random vector $X=(X^1,\ldots,X^d)$ with absolutely continuous density function *f*

select the design points in such a way as to have the density function "close" to the uniform density function.

The Kullback-Leibler (KL) divergence measures the difference between two density functions *f* and *g* (with $f \ll g$)

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

KL divergence properties

- The KL divergence is not a metric

  (it is not symmetric, it does not satisfy the triangle inequality)

- The KL divergence is always non-negative and

$$D(f, g) = 0 \quad \Rightarrow \quad f = g \quad \text{p.p.}$$

If $\{P_1,\ldots,P_n\}$ is a sequence of distributions then

$$P_n \xrightarrow[n \to +\infty]{\textit{KL divergence}} P \;\Rightarrow\; P_n \xrightarrow[n \to +\infty]{\textit{Total variation}} P$$

Minimizing the KL divergence

- The KL divergence is invariant under parameter transformations.

Design space = unit cube

The KL divergence and the Shannon entropy

If g is the uniform density function then

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

where H(f) is the Shannon entropy

Minimizing the KL divergence $\Leftrightarrow$ Maximizing the entropy

If f is supported by $[0,1]^d$, one always has $H(f) \le 0$ and the maximum value of $H(f)$, zero, being uniquely attained by the uniform density.

Using an exchange algorithm to build an "optimal" design

→

Entropy estimation

Estimation by a Monte Carlo method

- Introduction
- Kullback Leibler divergence
- Estimation by a Monte Carlo method
- Design comparison
- Conclusion

Estimation by a Monte Carlo method

The entropy can be written as an expectation

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

The Monte Carlo method (MC) provides a unbiased and consistent estimate of the entropy

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

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

the unknown density function f is replaced by its kernel density estimate (Ahmad and Lin, 1976)

Estimation by a Monte Carlo method

Joe (1989) obtained asymptotic bias and variance terms for the estimator

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

where $\hat{f}$ is the kernel estimate,

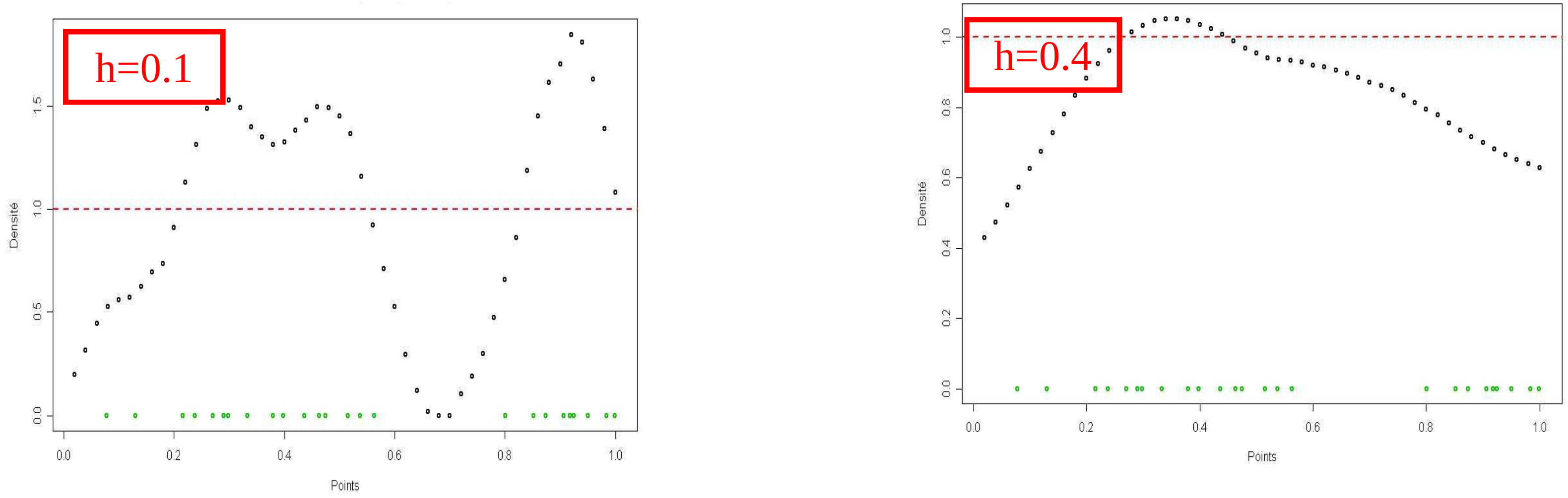
$$\forall x \in [0,1]^d,\ \hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

The bias depends on the size *n*, the dimension *d* and the bandwidth *h*

fix the bias during the exchange algorithm

The kernel density estimation : the bandwidth

- The bandwidth *h* plays an important role in the estimation



⇒ Scott's rule

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

j=1,…,d

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

Standard deviation of the uniform distribution

The kernel density estimation : the kernel

- the choice of the kernel function κ is much less important

Multidimensional Gaussian function

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

where $z = \frac{X_i - X_j}{h}$ i,j=1,…,n

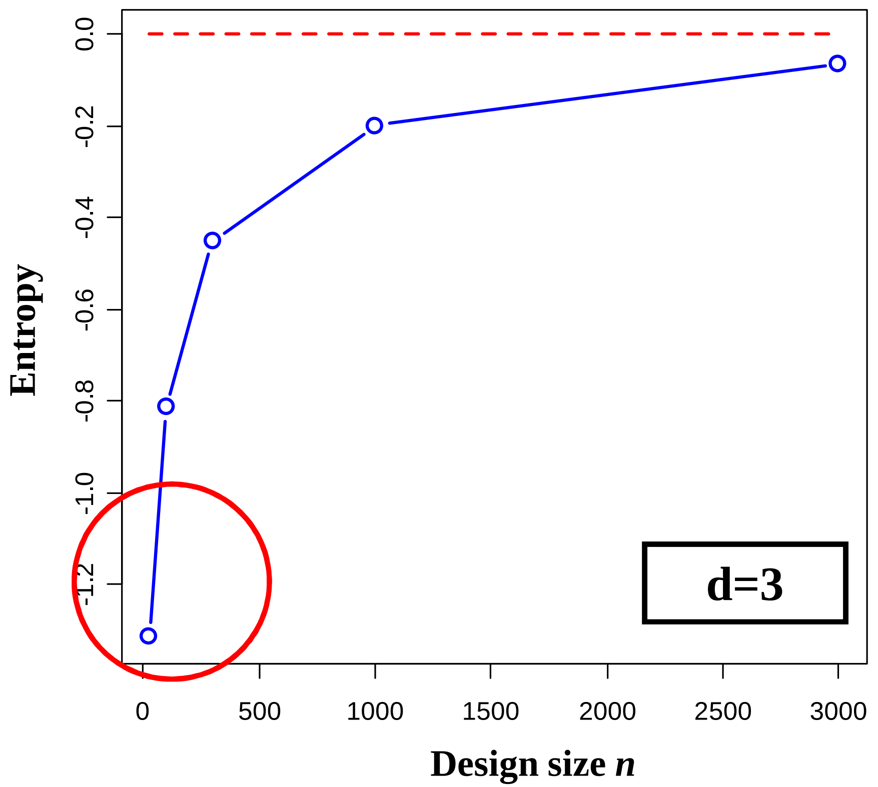
$\Longrightarrow$ $\|z\|^2 \in [0, d/h^2]$

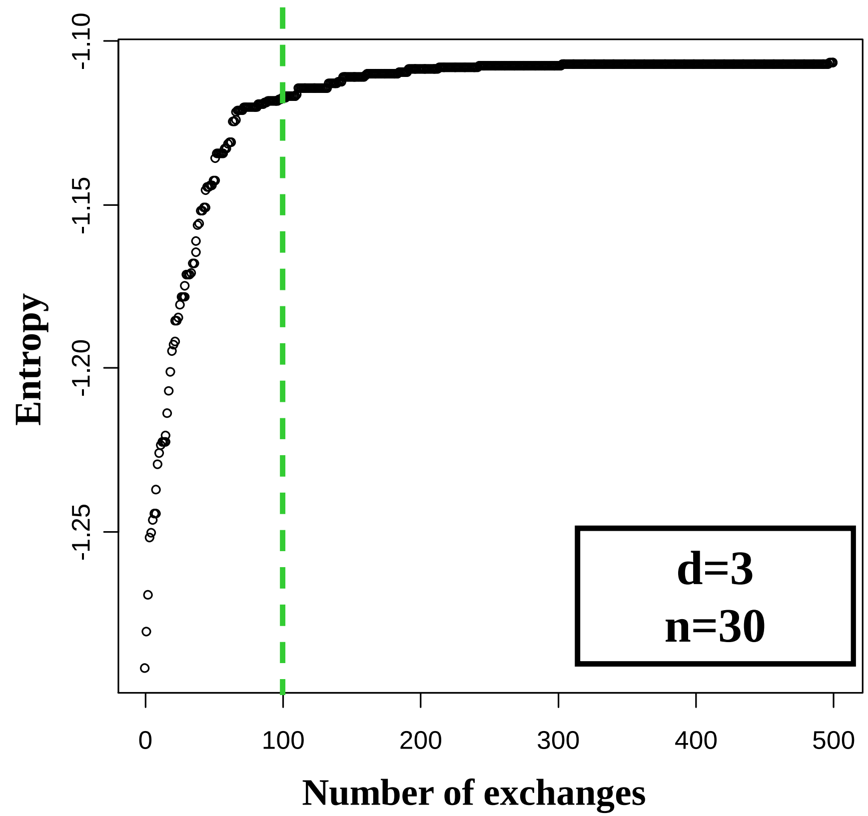
(d=10 and n=100 : $\|z\|^2 \in [0;231.7]$)

} Epanechnikov, uniform,… kernel functions are not desirable

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

Convergences





The entropy estimation converges slowly towards 0

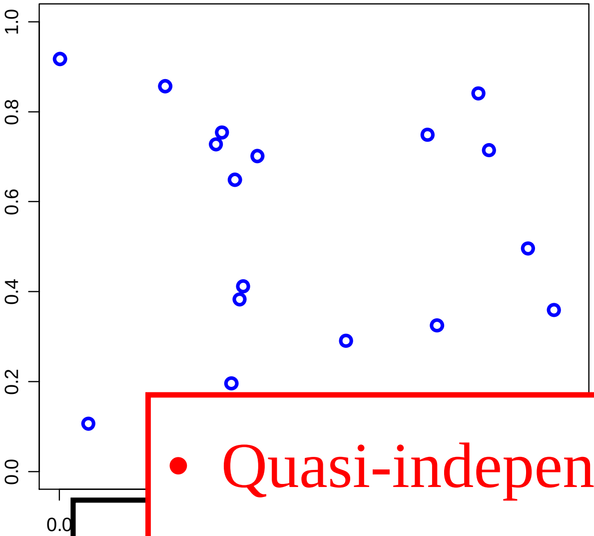
The exchange algorithm converges rapidly

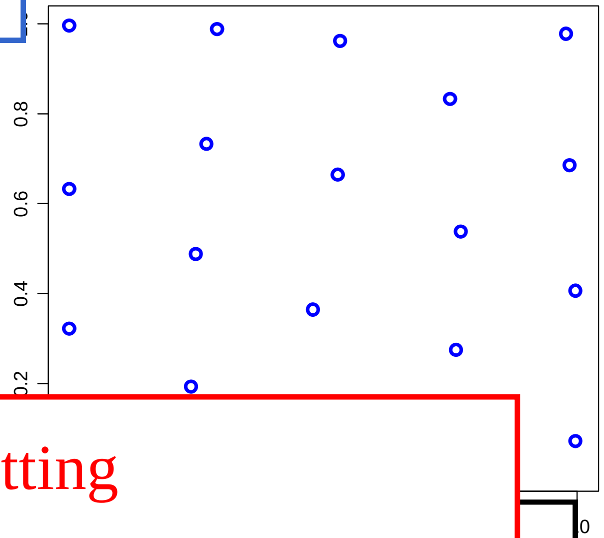
Design comparison

- Introduction
- Kullback-Leibler divergence
- Estimation by a Monte Carlo method
- Design comparison
- Conclusion

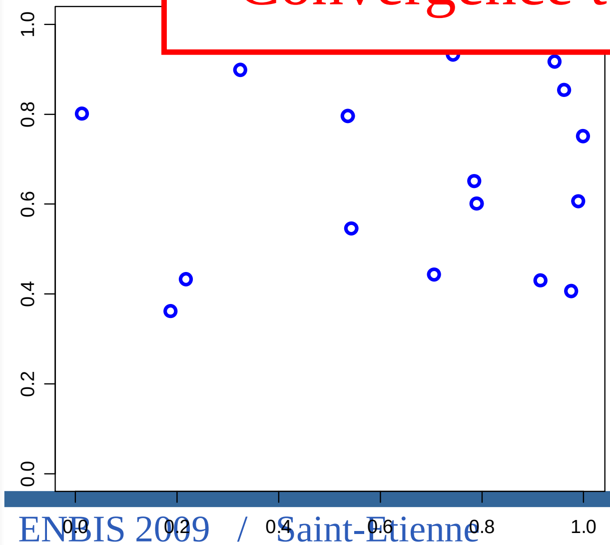
Improvement of the initial setting

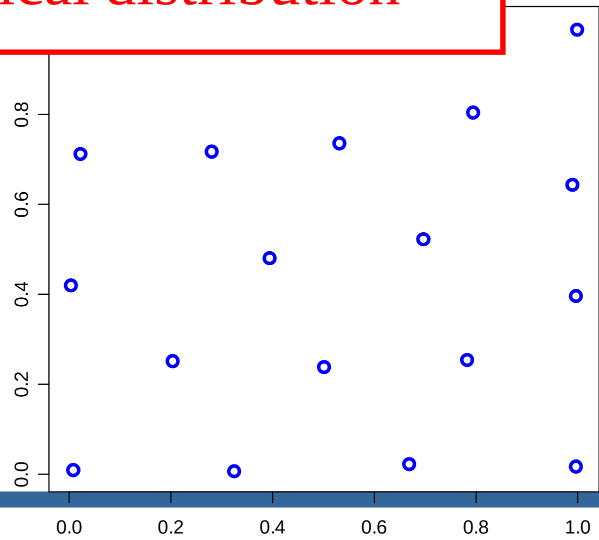
d=2 n=20





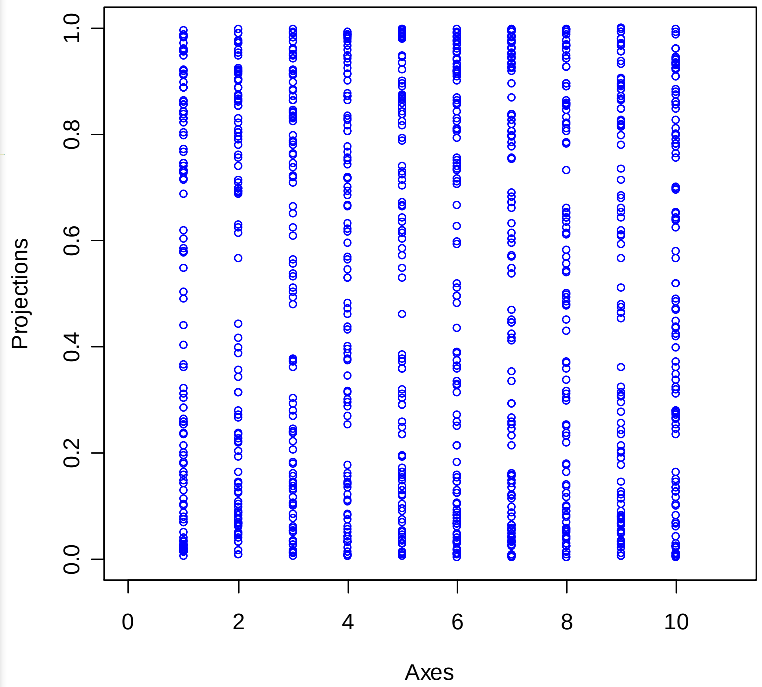
- Quasi-independent of the initial setting
- Convergence towards quasi-periodical distribution





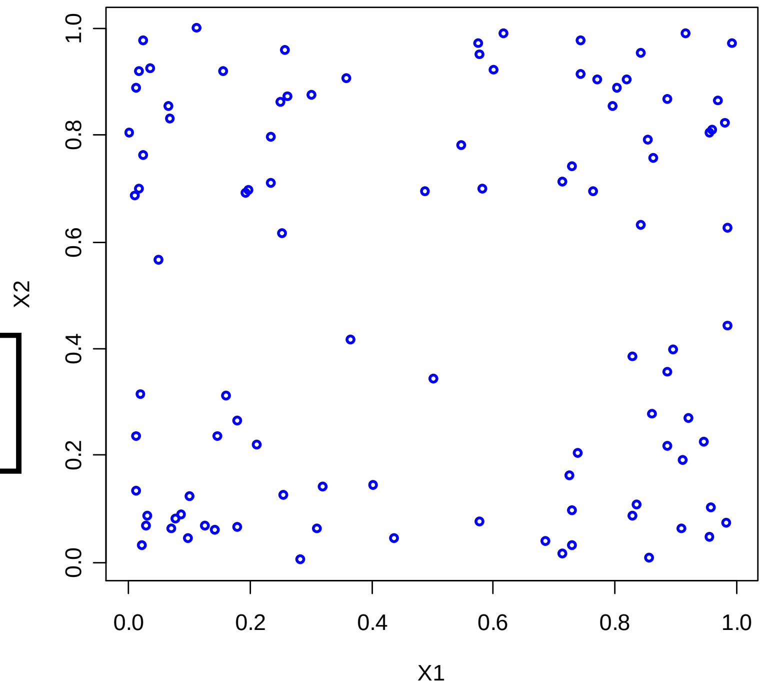
Projections

the design points will generally lie on the boundary of the design space, especially for small size *n*



Projections on each dimension

d=10
n=100



Projections on 2D plane X_1X_2

Usual space-filling designs

- The maximin criterion (**Maximin**) maximizes the minimal distance between the design points (Johnson *et al.*, 1990),

$$\min_{1 \le i < j \le n} d(x_i, x_j)$$

- The entropy criterion (**Dmax**) is the maximization of the determinant of a covariance matrix (Shewry & Wynn, 1987),

$$R(x_i, x_j) = \exp\left\{ - \sum_{k=1}^{d} \theta_k \left| x_i^k - x_j^k \right|^p \right\}$$

- Two kind of designs are based on the analogy of minimizing forces between charged particles

Strauss designs (**Strauss**) built with a MCMC method (Franco, 2008)

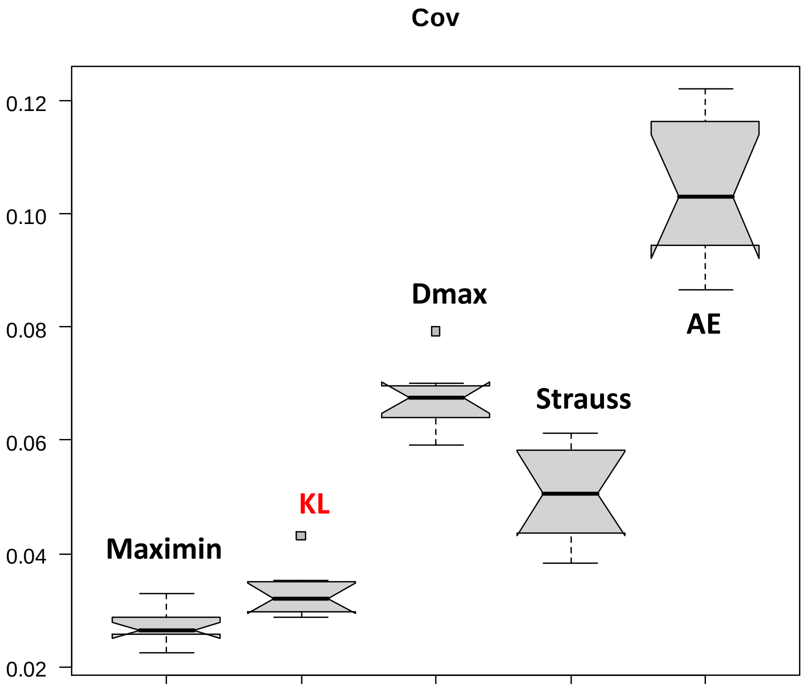
Audze-Eglais (1977) criterion (**AE**) minimizes

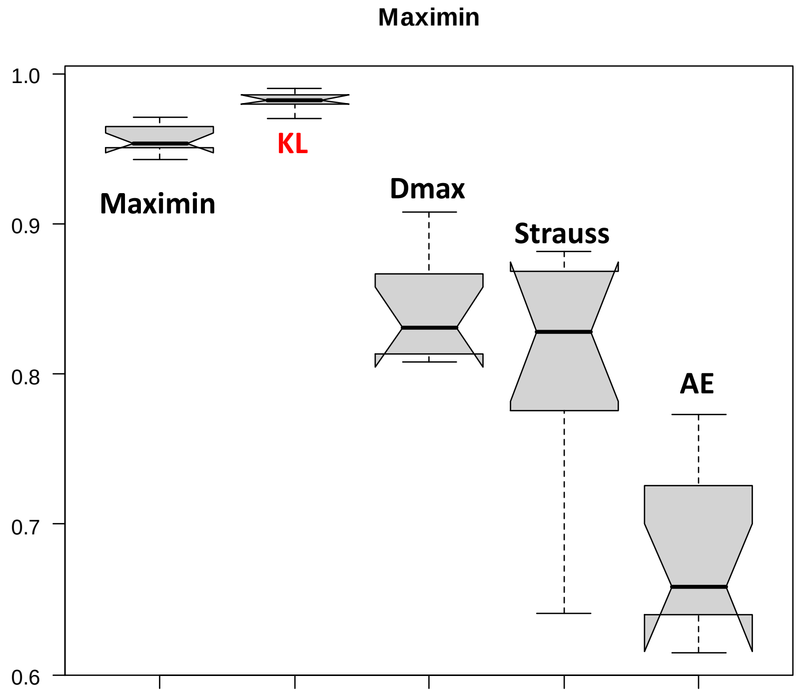
$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d(x_i, x_j)^2}$$

Usual criteria (d=10 and n=100)

Distance criteria

quantify how the points fill up the space





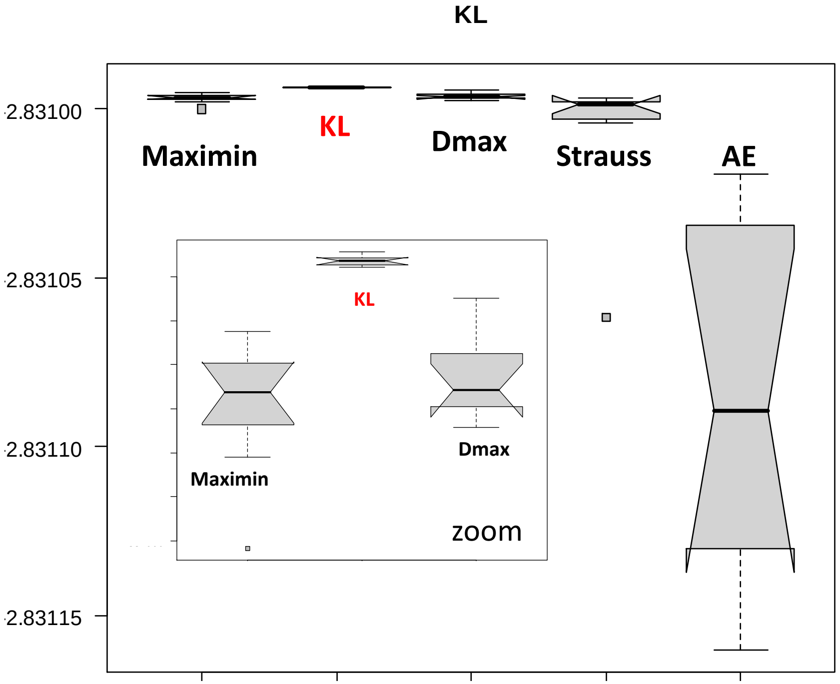
The cover measure calculates the difference between the design and a uniform mesh (min)

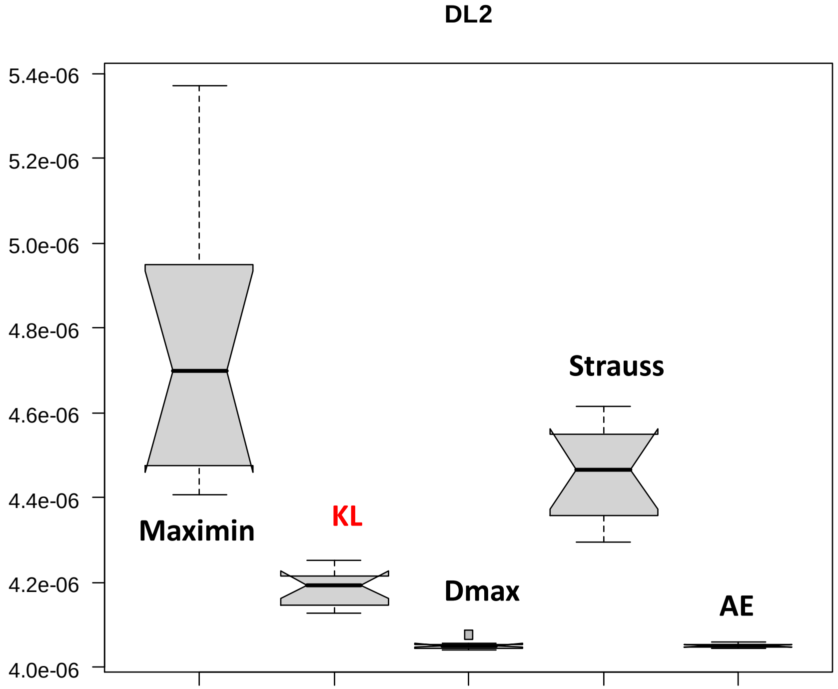
The Maximin criterion maximizes the minimal distance between the design points (max)

Usual criteria (d=10 and n=100)

Uniformity criteria

Measure how close points being uniformly distributed





KL divergence (max)

The discrepancy measures the difference between the empirical cumulative distribution of the design points and the uniform one (min)

Conclusion

- Introduction
- Kullback-Leibler divergence
- Estimation by a Monte Carlo method
- Design comparison
- Conclusion

Conclusion

Results

- The KL criterion spread points evenly throughout the unit cube
- The KL designs outperform the usual space-filling designs

Outlooks

- Estimation based on the nearest neighbor distances (CPU time + support of *f*)
- Construction of optimal Latin hypercube (projection)
- Tsallis entropy (analytic expression) , Rényi entropy (estimated by MST)

References

•Beirlant J., Dudewicz E.J., Györfi L., Van Der Meulen E.C. (1997). Nonparametric entropy estimation : an overview. Int. *J. Math. Stat. Sci*., 6(1) 17-39.

•Franco J (2008). *Planification d'expériences numériques en phase exploratoire pour des codes de calculs simulant des phénomènes complexes*. Thèse présentée à l'Ecole Nationale Supérieure des Mines de Saint-Etienne

•Gunzburger M., Burkardt J. (2004). Uniformity measures for point sample in hypercubes. https://people.scs.fsu.edu/~burkardt/pdf/ptmeas.pdf

•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. *J. Statist. Plann. Inf*., 26,131-148

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

•Scott D.W. (1992). *Multivariate Density Estimation : Theory, practice and visualization*, John Wiley & Sons, New York, Chichester

•Silverman B.W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall, London

•Shewry M.C., Wynn H.P. (1987). Maximum Entropy Sampling. *J. Appl. Statist*., 14, 165-170