Workshop « Expériences et Simulations : comment les planifier et les exploiter au mieux ? » , Avignon 6 et 7 novembre 2025

Space-filling designs for mixture experiments

Astrid Jourdan

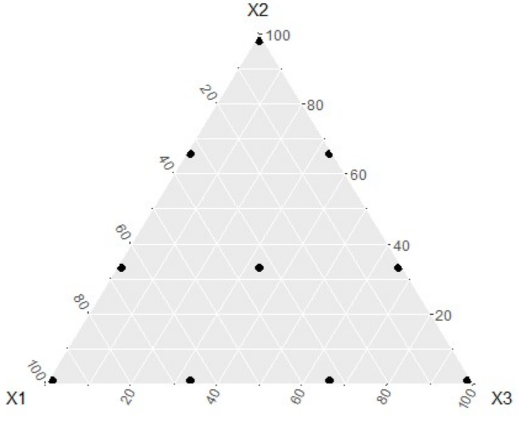
Université de Pau et des Pays de l'Adour

astrid.jourdan@univ-pau.fr

IVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

Mixture experiments

Mixture experiments consist in varying the proportions of some components involved in a physico-chemical phenomenon, and observe the resulting change on the response.



Simplex Lattice design

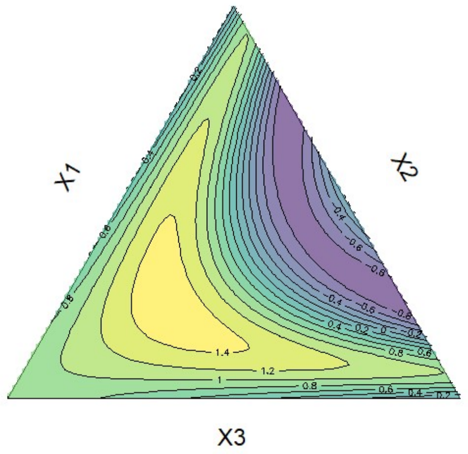
Denote x_k is the proportion of the *k*th component, $k = 1, \dots, d$.

- $x_k \in [0,1]$
- $x_1 + \cdots + x_d = 1$

The experimental region is reduced to a (d-1)-dimensional simplex S^{d-1}.

Usual approach

True response surface



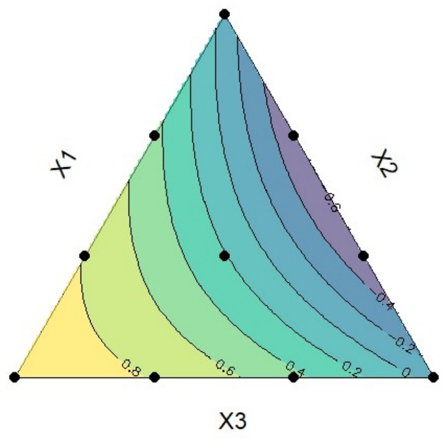
- Scheffé polynomial model:

$$Y(x) = \sum_i a_i x_i + \sum_{i<j} a_{ij} x_i x_j + \cdots + \varepsilon(x)$$

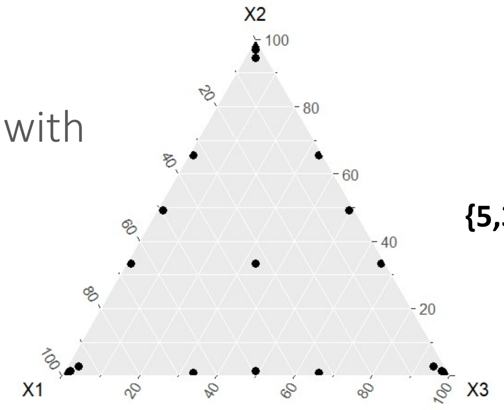
where $\varepsilon(x)$ are i.i.d. with $N(0, \sigma^2)$

- Simplex Centroid, Simplex Lattice designs, D_Optimal designs

Second order polynomial model



- The polynomial model is unable to represent a response surface with non-linearities
- Points on the edges do not detect non-linearities



{5,3} Simplex Lattice
n=31

Contents

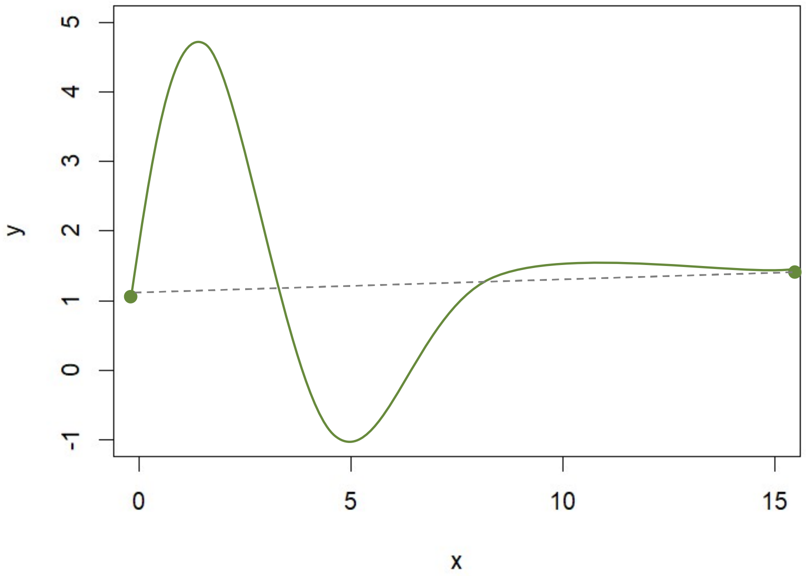
- Space-filling design in the hypercube
- Space-filling design for mixture experiments
- Constrained experiments

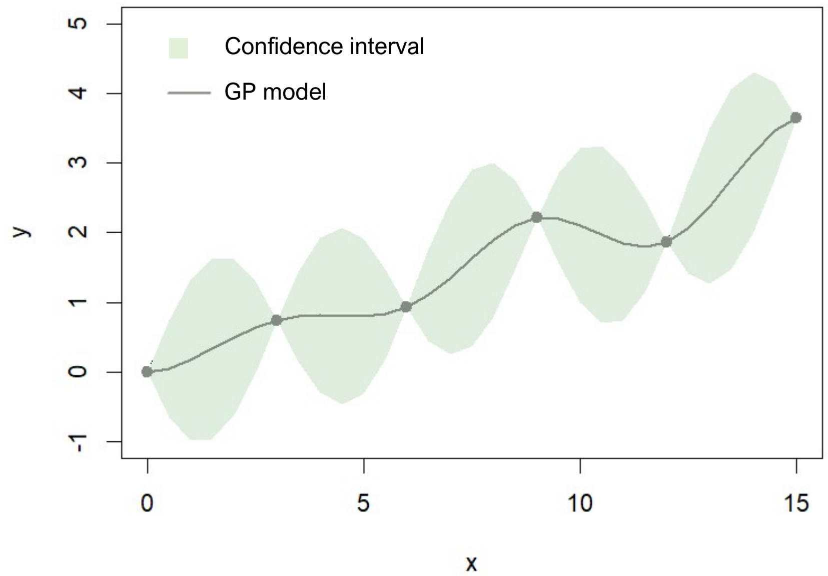
space-filling design in the hypercube

Why a space-filling design?

The aim of an experimental design is to detect any irregularities in the true response surface and reduce the uncertainty.

- Fill the experimental domain uniformly
- Test many levels for each variable

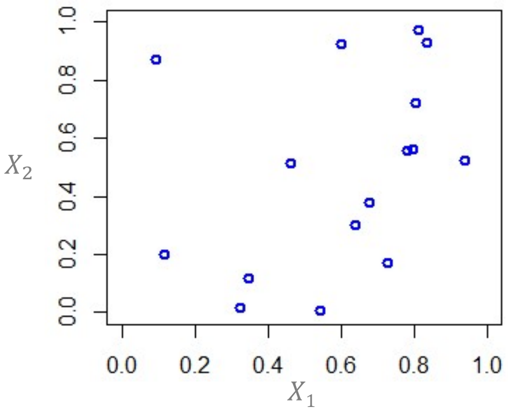




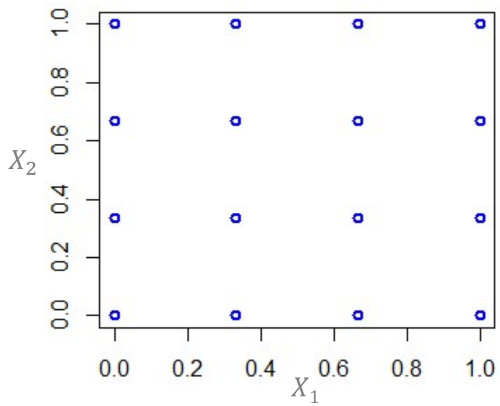
Why a space-filling design?

Two naïve sampling strategies:

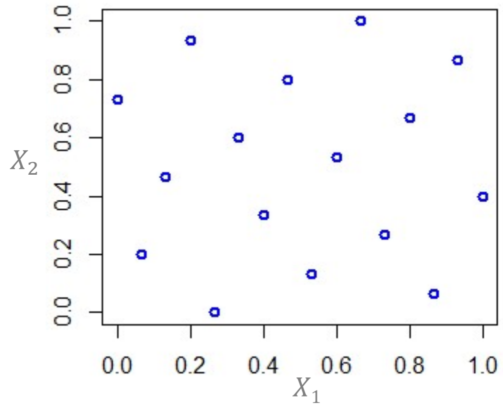
A simple random sample does not distribute points evenly



A grid requires a lot of experience for few tested levels ($n = p^d$)



The solution : a space-filling design



Examples with $n = 16$

Space-filling criteria

Criteria to evenly spread the design points over the experimental domain

- Distance-based criteria :

Goal: move the points away from each other

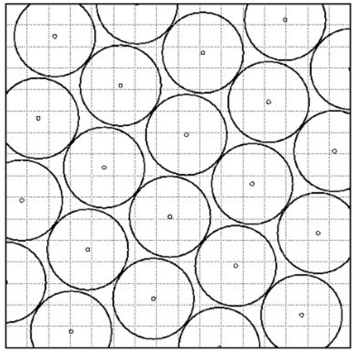
Maximin distance, Phi-p, Audze-Eglais, …

- Distribution-based criteria :

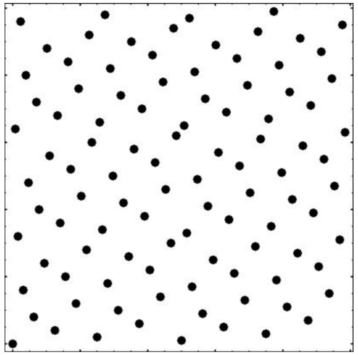
Goal: obtain an empirical distribution of the design points as close as possible to the uniform distribution

*.discrepancy : measure the difference between the proportion of points falling into an arbitrary set B and the measure of B – based on the cumulative distribution - centered L2-discrepancy, wrap-around L2-discrepancy,…

*.entropy : measure the difference between the empirical distribution of the points and the uniform distribution – based on the density function – Shannon entropy, Rényi entropy,…



Maximin distance design



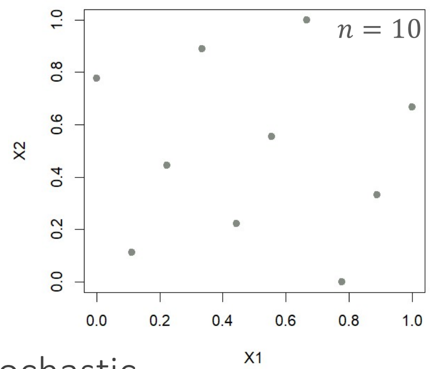
Halton sequence

Pronzato L., Muller W. (2012). Design of computer experiments : space filling and beyond. Statistics and Computing, 22, 681-701.

Construction of a space-filling design

Optimization algorithm

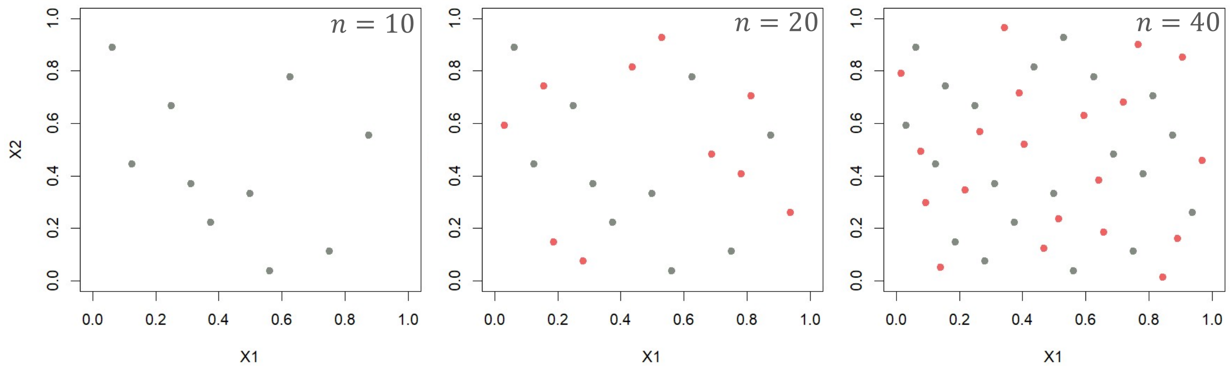
- Start with a set of n points randomly sampled in $[0,1]^d$
- Move the points in an optimization loop to minimize (or maximize) one of the criteria



- ✓ Stochastic
- ✓ Early stopping or local otpimum
- ✓ CPU time
- ✓ Tend to push the points on the boundaries

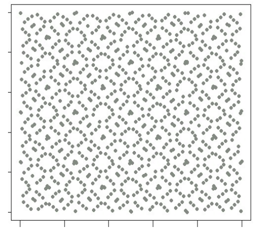
Low discrepancy sequence

is a sequence with the property that for all values of n, its subsequence $x_1, \dots, x_n$ has a low discrepancy.



Halton sequence

- ✓ Deterministic
- ✓ low discrepancy achieved if n is large enough
- ✓ Patterns



Sobol’ sequence

Space-filling designs in the simplex

UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

How to uniformly sample in the unit simplex?

Naïve approaches :

- Normalization

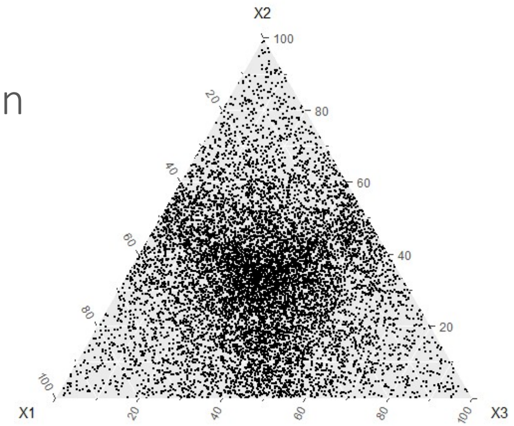
$U = (u_1, \ldots, u_d) \leftarrow \mathcal{U}[0,1]^d$
$S \leftarrow u_1 + \cdots + u_d$
$X \leftarrow U/S$

- Rejection

Repeat
  Randomly select $(u_1, \ldots, u_d)$ *in* $[0,1]^d$
until $u_1 + \cdots + u_d < 1$

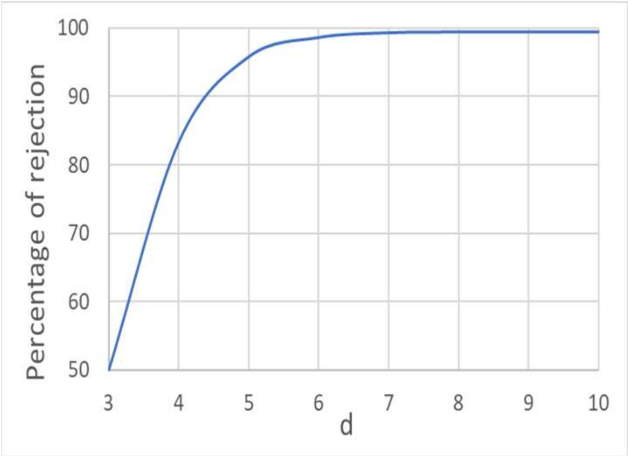
Favors points in the center

n=10000, d=3



Curse of dimensionality

Less than 1% of points accepted when d>5



How to uniformly sample in the unit simplex?

3 methods in the literature :

- Conditional distribution method

$(u_1, \dots, u_{d-1}) \leftarrow \mathcal{U}[0,1]^{d-1}$

$x_1 \leftarrow 1 - u_1^{1/(d-1)}$

$x_i \leftarrow \left(1 - u_i^{1/(d-i)}\right)\left(1 - \sum_{j=1}^{i-1} x_j\right)$

$x_d \leftarrow 1 - \sum_{j=1}^{d-1} x_j$

- Uniform order statistics

$(u_1, \dots, u_{d-1}) \leftarrow \mathcal{U}[0,1]^{d-1}$

$u_{(1)} \le u_{(2)} \le \dots \le u_{(d-1)}$

$x_1 \leftarrow u_{(1)}$

$x_i \leftarrow u_{(i)} - u_{(i-1)}$

$x_d \leftarrow 1 - u_{(d-1)}$

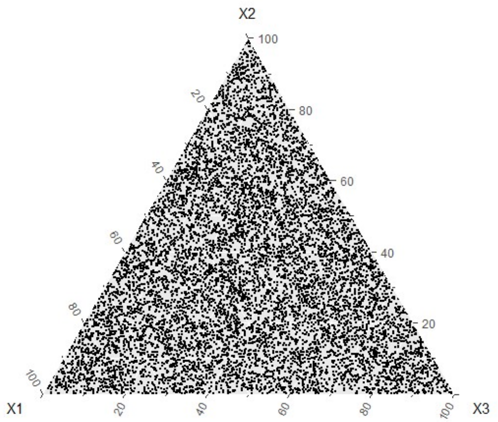
- Dirichlet distribution

$U = (u_1, \dots, u_d) \leftarrow \mathcal{U}[0,1]^{d}$

$E \leftarrow -\log(U)$

$S \leftarrow e_1 + \dots + e_d$

$X \leftarrow \mathrm{E}/S$



✓ Better distribution than the naive normalization method

✓ No rejection

But …

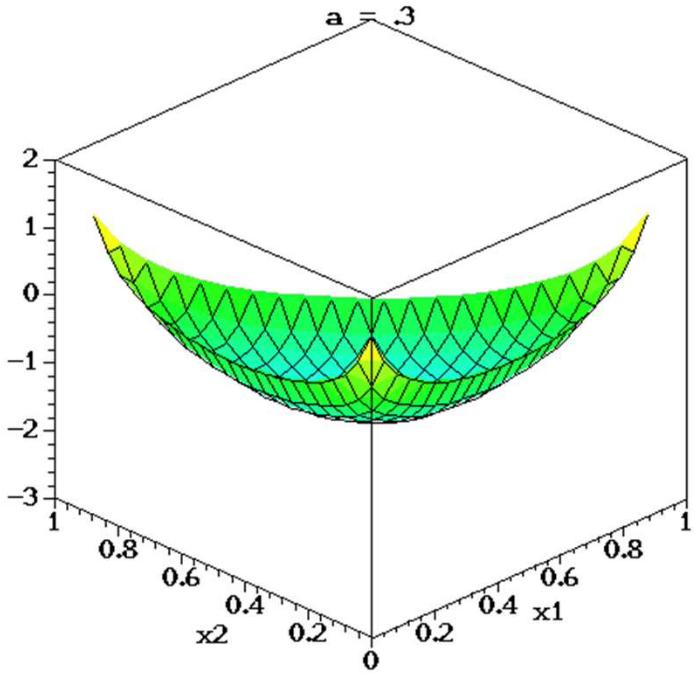
Dirichlet distribution

Dirichlet distribution is a family of continuous multivariate probability distributions parameterized by a vector $\boldsymbol{a}$ of positive reals,

$$g(x) = \frac{1}{B(\boldsymbol{a})} \prod_{k=1}^{d} (x_k)^{a_k - 1}$$

where x belongs to the (d-1)-simplex S^{d-1} and $B(\boldsymbol{\alpha})$ is the normalizing constant.

If $\boldsymbol{a}$=(1,…,1) the Dirichlet distribution is equivalent to a uniform distribution over the simplex.



Wikipedia

How to uniformly sample in the unit simplex?

3 methods in the literature :

- Conditional distribution method

$(u_1, \dots, u_{d-1}) \leftarrow \mathcal{U}[0,1]^{d-1}$

$x_1 \leftarrow 1 - u_1^{1/(d-1)}$

$x_i \leftarrow \left(1 - u_i^{1/(d-i)}\right)\left(1 - \sum_{j=1}^{i-1} x_j\right)$

$x_d \leftarrow 1 - \sum_{j=1}^{d-1} x_j$

- Uniform order statistics

$(u_1, \dots, u_{d-1}) \leftarrow \mathcal{U}[0,1]^{d-1}$

$u_{(1)} \leq u_{(2)} \leq \cdots \leq u_{(d-1)}$

$x_1 \leftarrow u_{(1)}$

$x_i \leftarrow u_{(i)} - u_{(i-1)}$

$x_d \leftarrow 1 - u_{(d-1)}$

- Dirichlet distribution

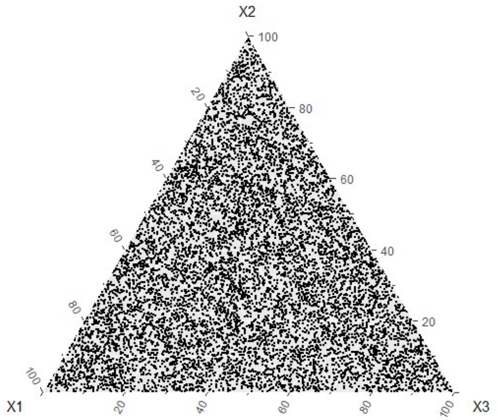
$U = (u_1, \dots, u_d) \leftarrow \mathcal{U}[0,1]^{d}$

$E \leftarrow -\log(U)$

$S \leftarrow e_1 + \cdots + e_d$

$X \leftarrow \mathrm{E}/S$

Very fast for simulating a large number of points (with R or Python NumPy)



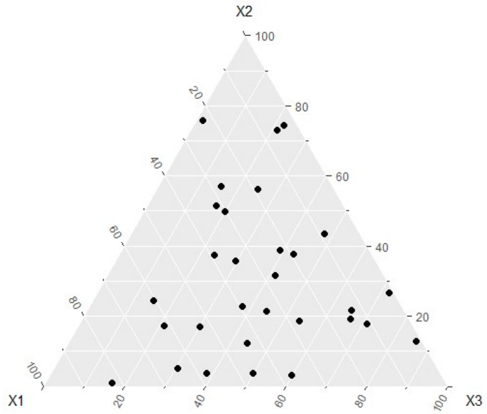
✓ Better distribution than the naive normalization method

✓ No rejection

But …

Space-filling design in the simplex

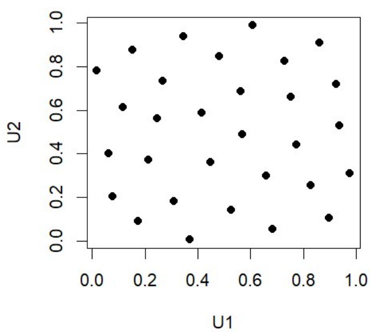
But the points are not enough evenly spread in the unit simplex.



Random design in the simplex

2 strategies :

- Replace $U \sim \mathcal{U}[0,1]^{d-1}$ by a space-filling design in the unit cube in the previous **mapping** functions
- Defined a **space-filling criterion** on the unit simplex (uniformity criterion or distance-based criterion)

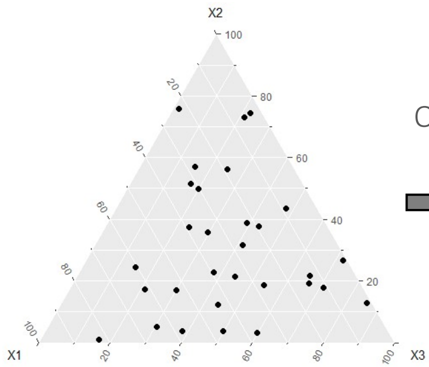


Space-filling Latin Hypercube

No guarantee to preserve the space-filling properties

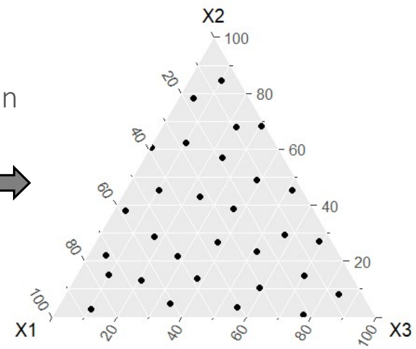
Mapping

« Space-filling » design in the simplex



Random design in the simplex

Optimization algorithm



Space-filling design in the simplex

Criteria for uniformity over the simplex

Discrepancy criterion : DM2 is an adaptation of the star discrepancy to the simplex

$$M2(D) = \left(\frac{\sqrt{d}}{(d-1)!}\right)^{\frac{1}{2}} \left\{ c_{n,s} - \frac{2(d-1)!}{n} \sum_{i=1}^{n} \sum_{(\tau_2,\ldots,\tau_d)\in\{0,1\}^{d-1}} a_\tau . (x_{i1})^{2(d-1)-\sum_{j=2}^{2}\tau_j} . \prod_{j=2}^{d} (x_{ij})^{\tau_j} + \frac{1}{n^2} \sum_{i=1,k=1}^{n} \left(\max\left(1 - \sum_{j=2}^{d} \max(z_{ij}, z_{kj}), 0\right)\right)^{d-1} \right\}^{\frac{1}{2}}$$

where $c_{n,s} = \left((d-1)!\right)^3 2^{d-1} / (2(d-1)! \prod_{k=0}^{d-2}(2d+k-1))$ and $a_\tau = (d-1)!/\left(2(d-1) - \sum_{i=2}^{d} \tau_i\right)!$.

Ning JH, Zhou YD, Fang KT (2011) Discrepancy for uniform design of experiments with mixtures, Journal of Statistical Planning and Inference.

Dirichlet criterion : measure the distance between the point distribution and the Dirichlet distribution with the Kullback-Leibler divergence

$$C_{kern}(D) = \sum_{i=1}^{n} \left[log \left(\sum_{j=1}^{n} e^{-\frac{1}{2}\left\| \frac{x_j - x_i}{h} \right\|^2} \right) \right]$$

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

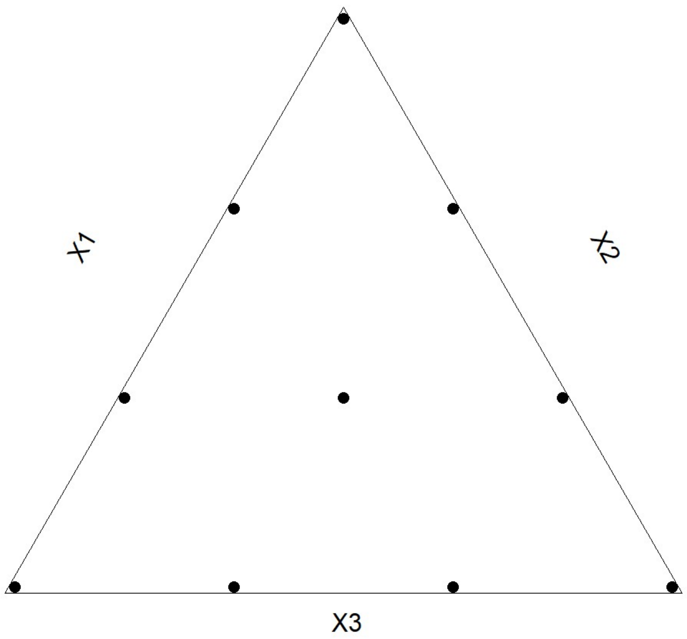
$$C_{nn}(D) = -\sum_{i=1}^{n} log \left\{ \left(\rho^{(k)}(\boldsymbol{x_i}, D \backslash \{\boldsymbol{x_i}\}) \right)^d \right\}$$

where $\rho^{(k)}$ is the k-nearest-neighbor distance.

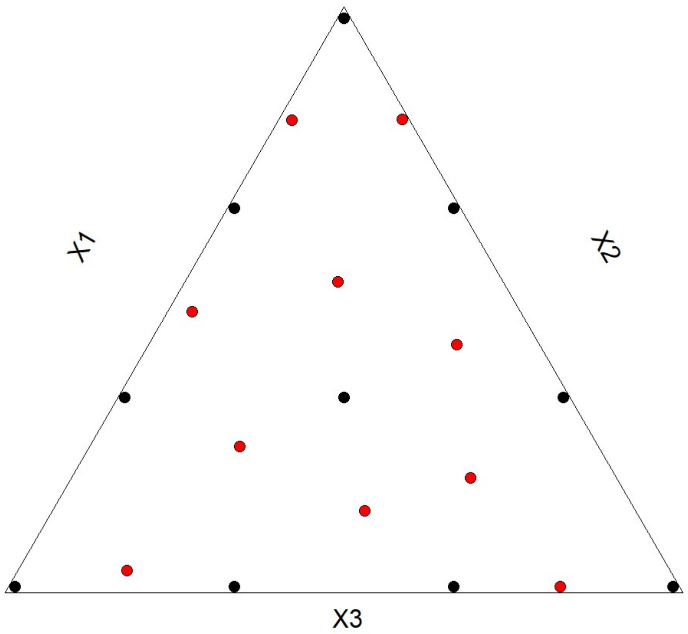
Jourdan A. (2023). Space-filling designs with a Dirichlet distribution for mixture experiments. *Statistical Papers*, 65, 2667–2686.

Sequential sampling

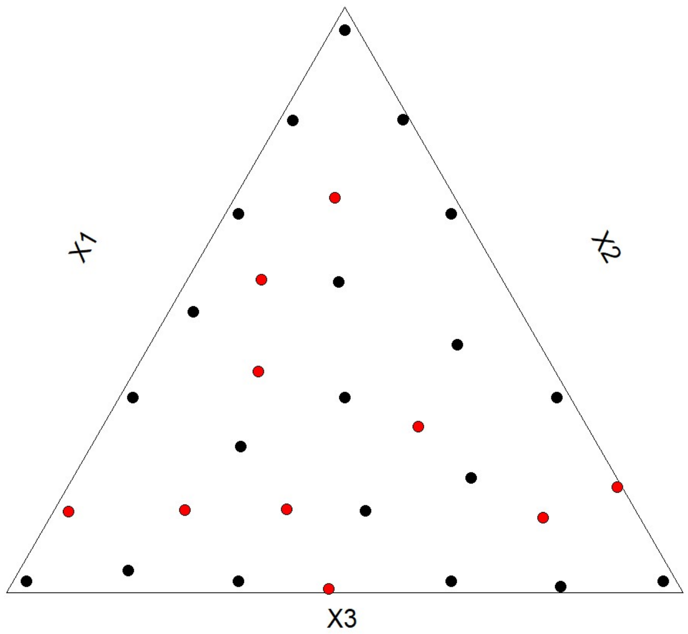
One advantage of the second method is the possibility of a sequential approach



Standard
initial design



10 additionnal
space-filling points

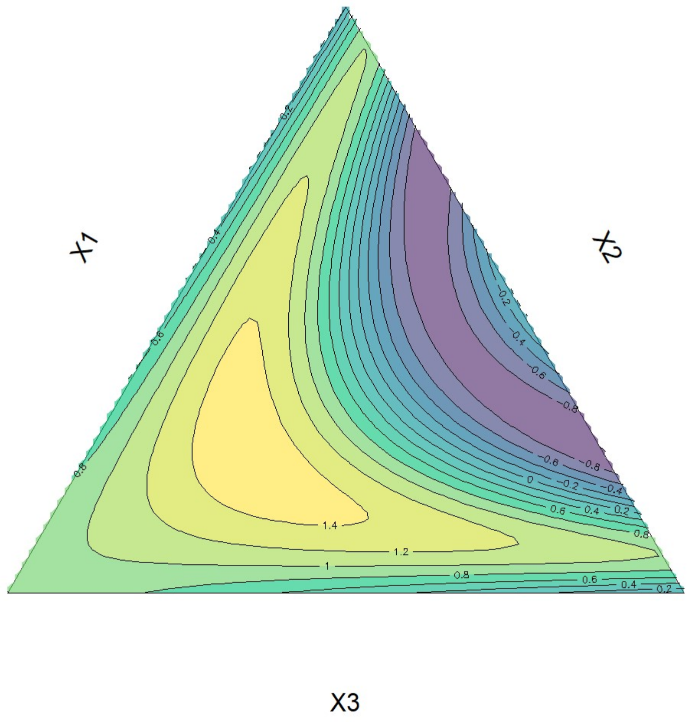


10 additionnal
points more

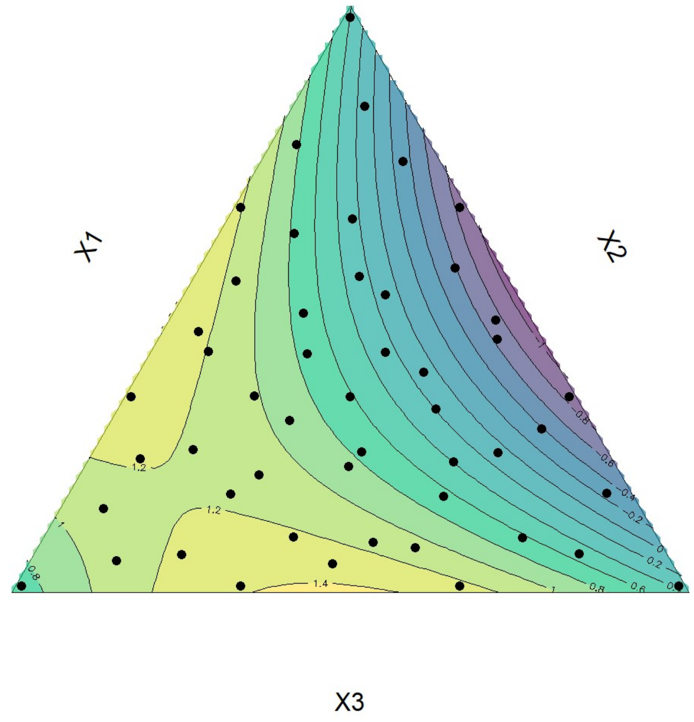
…

Sequential sampling

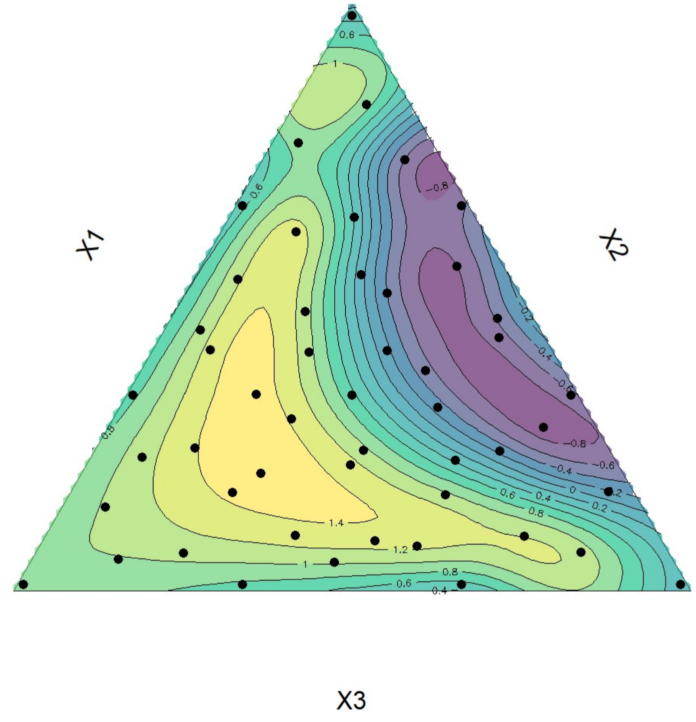
True response



2nd order polynomial model



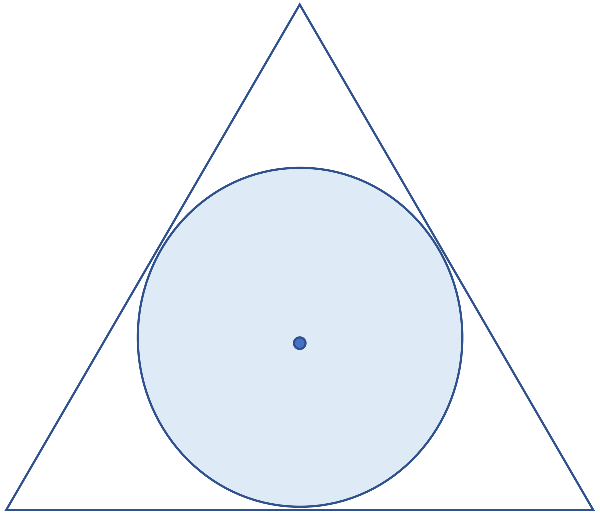
Gaussian process model

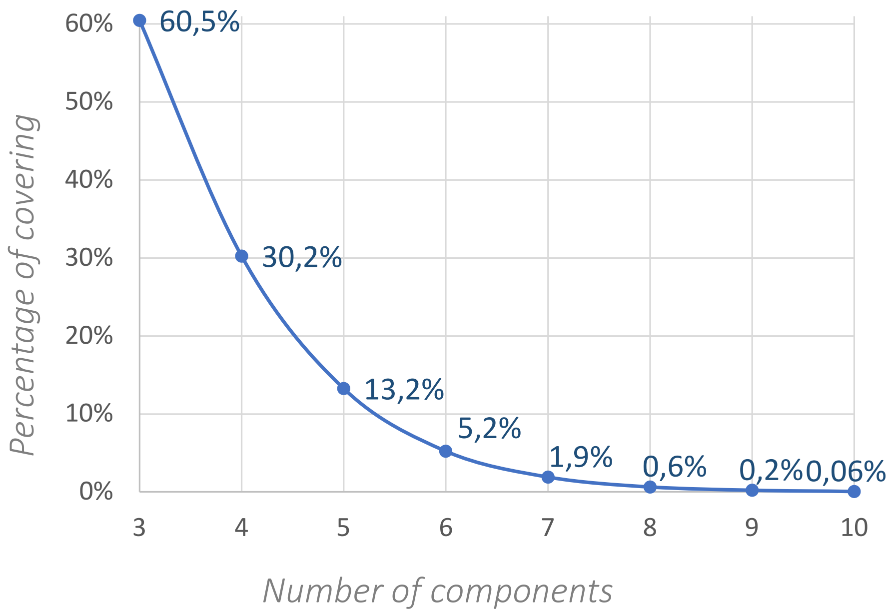


Curse of dimensionality

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

The biggest circle covers 60,5% of the triangle (3 components) and the biggest hypersphere covers less than 2% of the experimental domain for 7 or more components





Constrained experimental domain

Constrained mixture experiments

For certain components of the mixture, it is necessary to specify some constraints,

Simple constraints $0 \leq a \leq X_k \leq b \leq 1$

or

Complex constraints $0 \leq a \leq \alpha_1 X_1 + \cdots + \alpha_d X_d \leq b \leq 1$

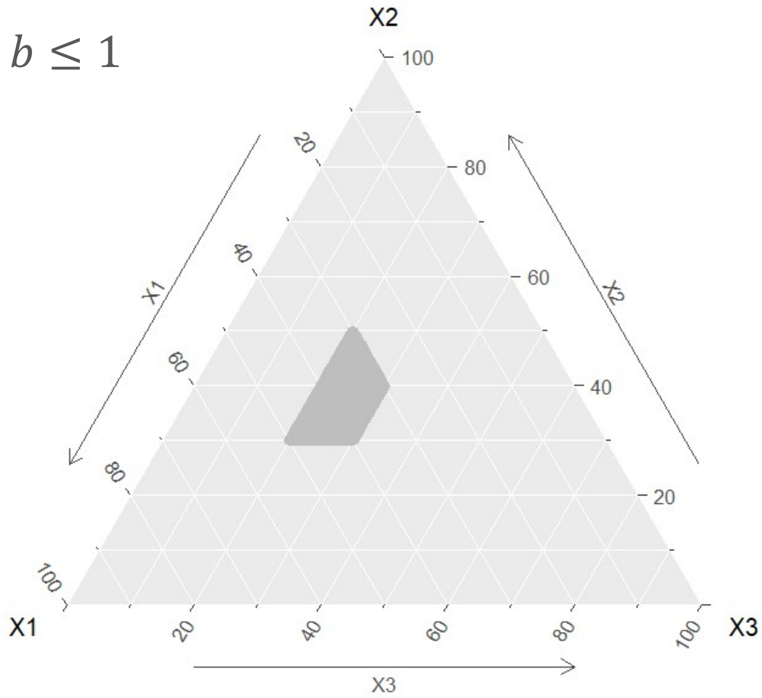
The experimental region can be considerably reduced.

Example:

$0,3 \leq X_1 \leq 0,6$ $\quad X_3 \geq 0,2$

$0,3 \leq X_2 \leq 0,6$ $\quad X_1 + X_2 \geq 0,7$

The volume of the constrained domain is reduced to 3% of the volume of the simplex.



Constrained mixture experiments

- Mapping function from [0,1]^d to constrained experimental domain (Fang and Wang).

→ Defined only for simple constraints:

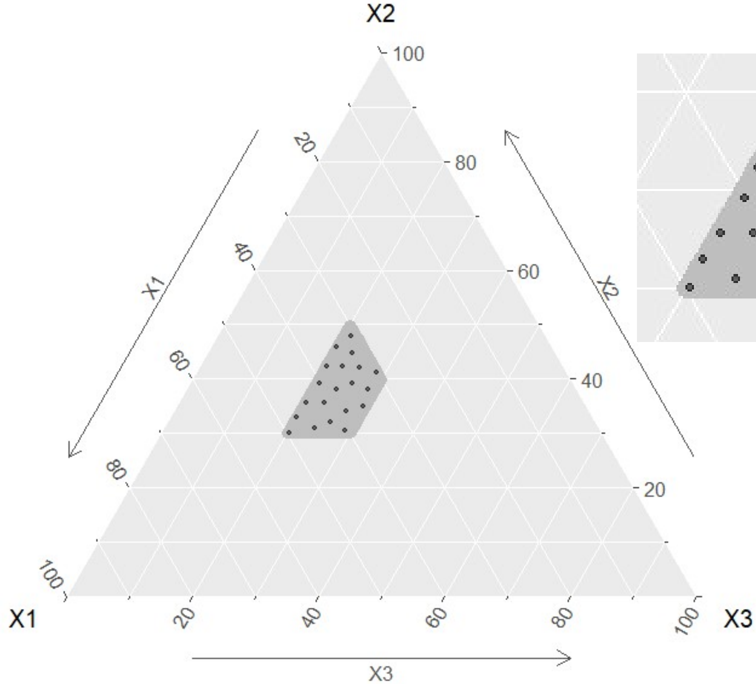
$$0 \leq a \leq X_k \leq b \leq 1$$

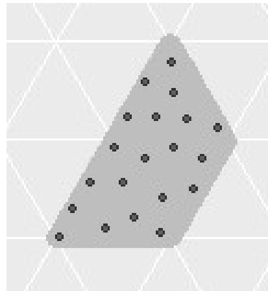
- Optimization algorithm

Requires generating a large set of points in the simplex and selecting only those that lie within the constrained experimental domain.

→ Curse of dimensionality

Previous example with d=5 : Volume ↘ 0,5%





Thank you
I welcome your questions

Google DeepMind 2025-5-14

AlphaEvolve: A coding agent for scientific and algorithmic discovery

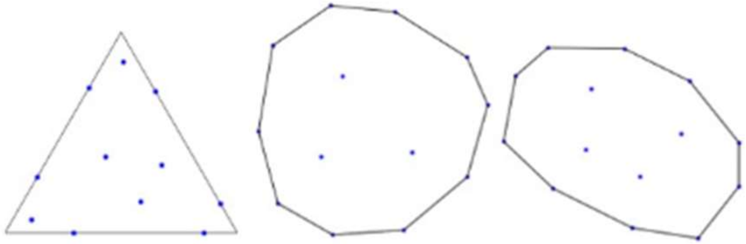


Figure 13 | New constructions found by *AlphaEvolve* improving the best known bounds on two variants of the Heilbronn problem. Left: 11 points in a unit-area triangle with all formed triangles having area ≥ 0.0365. Middle: 13 points inside a convex region with unit area with all formed triangles having area ≥ 0.0309. Right: 14 points inside a unit convex region with minimum area ≥ 0.0278.

References

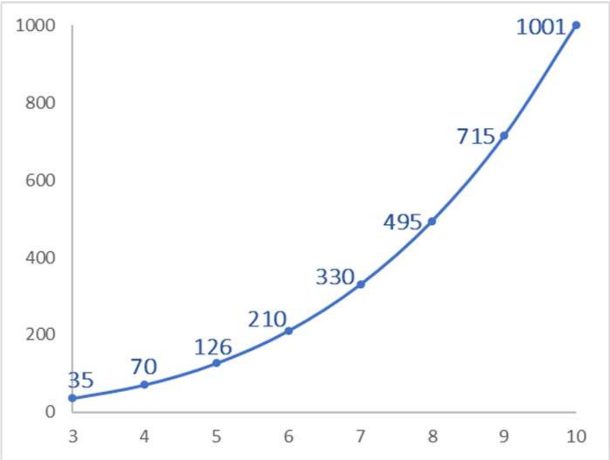
- Borkowski JJ, Piepel GF (2009). Uniform designs for highly constrained mixture experiments. Journal of Quality Technology.
- Fang KT, Li R, Sudjianto A (2005). Design Modeling for Computer Experiments. Chapman & Hall/CRC Press, London.
- Fang KT, Wang Y (1994). Number-theoretic Methods in Statistics. Chapman & Hall, London.
- Jourdan A. (2023). Space-filling designs with a Dirichlet distribution for mixture experiments. Statistical papers.
- Gomes C., Claeys-Bruno M., Sergent M. (2018), Space-filling designs for mixtures. Chemometrics and Intelligent Laboratory Systems.
- Menchaca-Mendez A., Zapotecas-Martinez S., Garcia-Velazquez L.M., Coello-Coello C.A. (2021). Uniform mixture design via evolutionary Multi-objective optimization. Swarm and Evolutionnary Computation.
- Ning JH, Zhou YD, Fang KT (2011). Discrepancy for uniform design of experiments with mixtures. Journal of Statistical Planning and Inference.
- Verret E. (2024). Optimisation des systèmes de protection incendie par machine learning Application aux systèmes intumescents. Phd manuscript, université de Lille.

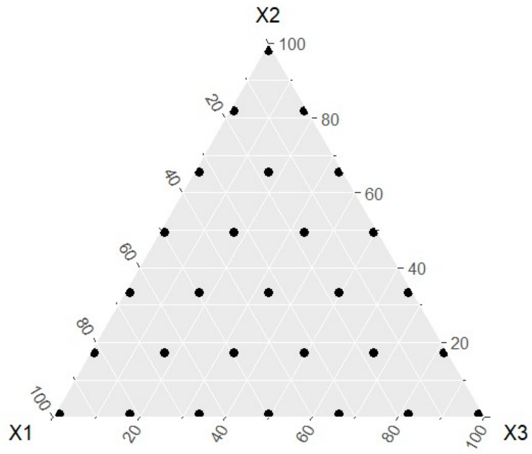
Curse of dimensionality 2

The number of points in a $\{d, q\}$ simplex lattice design is

$$n = (d + q - 1)!/(q!\,(d - 1)!)$$

Example for $d = 5$





$\{3, 6\}$ **Simplex Lattice**
$\boldsymbol{n = 28}$