New experimental design for mixture problems

Aurélie Béal[1], Charles Gomes[2], Magalie Claeys-Bruno[3] & Michelle Sergent[4]

1 Aix-Marseille Université, LISA EA4672
13397 Marseille Cedex 20, France
e-mail : aurelie.beal@univ-amu.fr

2 L'ORÉAL Research & Innovation
1, Av Eugène Schueller, 93601, Aulnay-sous-Bois, France
e-mail : cgomes@rd.loreal.com

3 Aix-Marseille Université, LISA EA4672
13397 Marseille Cedex 20, France
e-mail : m.claeys-bruno@univ-amu.fr

4 Aix-Marseille Université, LISA EA4672
13397 Marseille Cedex 20, France
e-mail : michelle.sergent@univ-amu.fr

Abstract
In the case of complex phenomenon, it is interesting to spread the points throughout the experimental domain. These designs are called Space-Filling Designs and are well-known in the case of independent variables.
In mixture problems, the variables which are the proportions of the various components are dependent. Therefore, classical uniforms designs could not be used. We propose two construction methods allowing the generation of uniform designs taking into account mixture constraints.

Keywords: uniform design, mixture, Space Filling Design
Mots-clés : plan d'expériences uniforme, mélange

1. Introduction

In many fields, such as cosmetic, pharmacy, food-processing industry, properties are often dependent on the proportion of various components in a mixture. In these studies, the input variables are not independent (their sum is equal to 1) since they are the proportions of the various components (their values are dimensionless numbers between 0 and 1). In order to represent the behavior of the responses (properties), mixture models and classical designs, as Scheffé designs [Scheffé], can be used to estimate the coefficients of the models. In the particular case of mixture with constraints, independent or relational constraints, whose consequence is a reduced domain of interest, D- or G-optimal designs can be used. These both designs spread experimental points on the periphery of the experimental domain leaving "hollow" center which could be harmful to the modelling step. In the case of complex phenomenon, it seems to be interesting to explore the whole experimental domain with Space-Filling Designs (SFD), whose points are distributed as uniformly as possible throughout

the experimental space. The most known SFD [Fang] are Latin hypercubes, low discrepancy sequences, but are not well adapted to mixture problems.
In this work, we propose two construction methods allowing the generation of uniform designs taking into account mixture constraints.

2. Construction of Space-Filling Design for mixture

2.1 Measures of uniformity

When the experimental space dimension is higher than 2, the uniformity of the space filling cannot be visually evaluated. It is thus necessary to use measures in order to know if the distribution is uniform and if the space of the variables is well filled. Only the most used criteria proposed in the literature are presented.

2.1.1 The Euclidean distance, *MinDist* [Johnson]

$$\boldsymbol{MinDist} = \min_{x_i \in X} \min_{\substack{x_k \in X \\ k \neq i}} dist\left(x^i, x^k\right)$$

with, $X = \{x^1, x^2, \ldots, x^n\} \subset [0,1]^D$, a set of n points in D dimensions.

A higher value of *MinDist* should correspond to a more regular scattering of design points and ensures that a point is never too close to another point.

2.1.2 The cover measure, *Cov* [Gunzburger]

$$\boldsymbol{Cov} = \frac{1}{\bar{\gamma}}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\gamma_i - \bar{\gamma}\right)^2\right)^{1/2}$$

with, $\gamma_i = \min_{k \neq i} dist\left(x^i, x^k\right)$ and $\bar{\gamma} = \frac{1}{n}\sum_{i=1}^{n}\gamma_i$

A low value of *Cov* corresponds to a distribution close to a regular grid and ensures that the points fill up the space.
We can plot these values, *MinDist* and *Cov*, to characterize different points distributions (random, cluster, ordered ...). In the (*MinDist, Cov*) plane, the best space-filling designs correspond to a quasi-periodical distribution which presents the best compromise between a regular grid (space filling) and a random distribution (uniformity). These designs are characterized by a low value of *Cov* and a high value of *Mindist*, which means that the desirable area is at the bottom, on the right.

2.2 Algorithms

There are many algorithms to build uniform experimental designs when variables are independent but to apply these algorithms in the case of mixtures we must adapt them.

2.2.1 WSP

The Wootton, Sergent, Phan-Tan-Luu's algorithm (WSP) [Santiago] was initially developed to build SFD in independent variables space and generate designs with good uniformity criteria whatever the number of dimensions.
In this algorithm, points are selected from a set of candidate points so as to be at a fixed minimal distance (d_{min}) from each point already in the independent variables space. The number of points in the subset depends on the value of d_{min}. If the d_{min} value increases, the number of points in the final subset decreases.

To apply the WSP algorithm in mixture studies, the set of candidate points must be generated in the mixture space, thus the selected points will be in the same space. Then, the minimal distance value, d_{min} must be fixed to obtain N points.
For example, if we consider a set of 5000 candidate points in the three dependent variables space, the WSP algorithm select N=10 points when d_{min}=0.34 and N=20 points when d_{min}=0.215.
These solutions are represented Figure 1.

a. b.

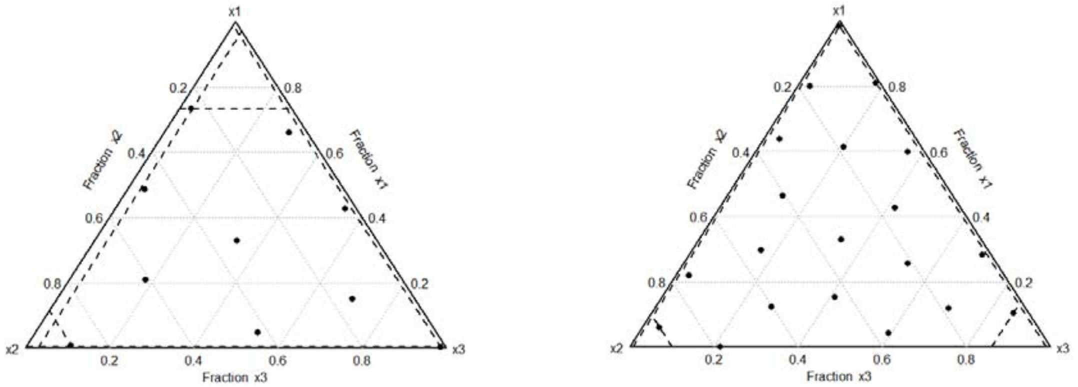


Figure 1: WSP algorithm applied to mixture design. a) Solution with N=10 points. b) Solution with N=20 points

2.2.2 SbS

The Step by Step algorithm (SbS) [Franco] is an *a priori* algorithm which builds a set of N points by iteration. The principle of this algorithm is to randomly choose one or several initial points then to add points at a distance R of the points already in the design.
The particularity of this algorithm is that all points are at the same distance of each other, whatever the number N of points, which guarantees a uniform filling all over the domain.

As the first points are randomly chosen, for a same value R we can obtain several solutions; consequently, to determine the R value which will allow to obtain a number N of points, it is necessary to repeat the algorithm several times (Figure 2).

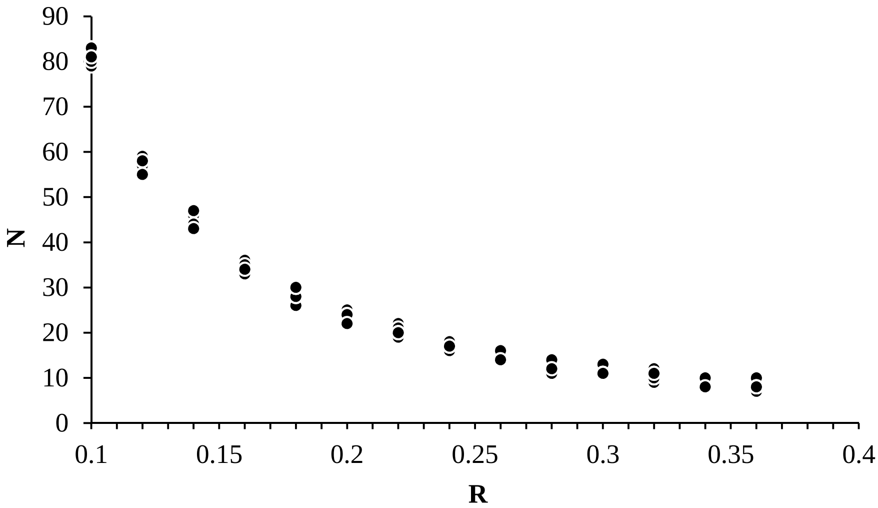
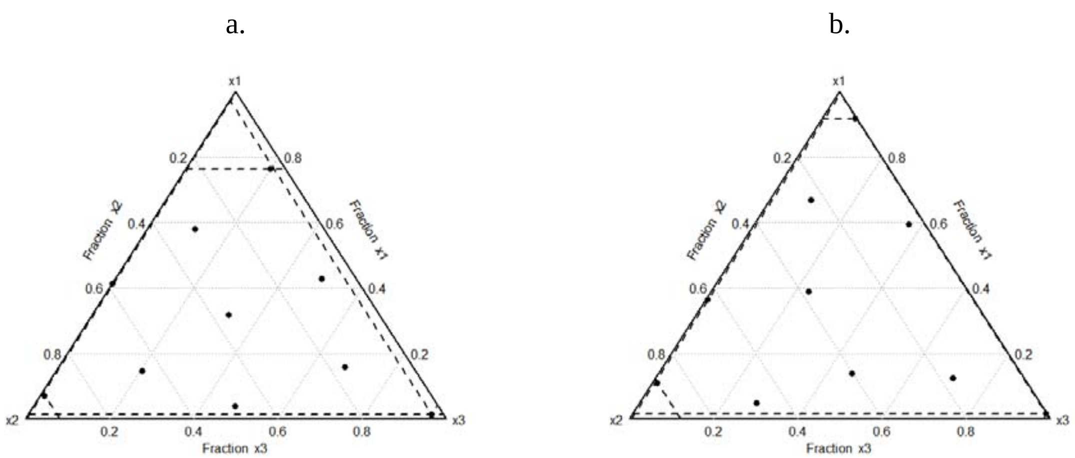


Figure 2: Evolution of the number of points as a function of R distance for the SbS algorithm solutions in a mixture design for 3 components. For each R value, the algorithm is repeated 5 times.

All solutions obtained by SbS algorithm present a set of points which are separated by the same R distance from each other. If we consider another initial point, the solution can be different but with the same properties and a close distance R (Figure 3).



c. d.

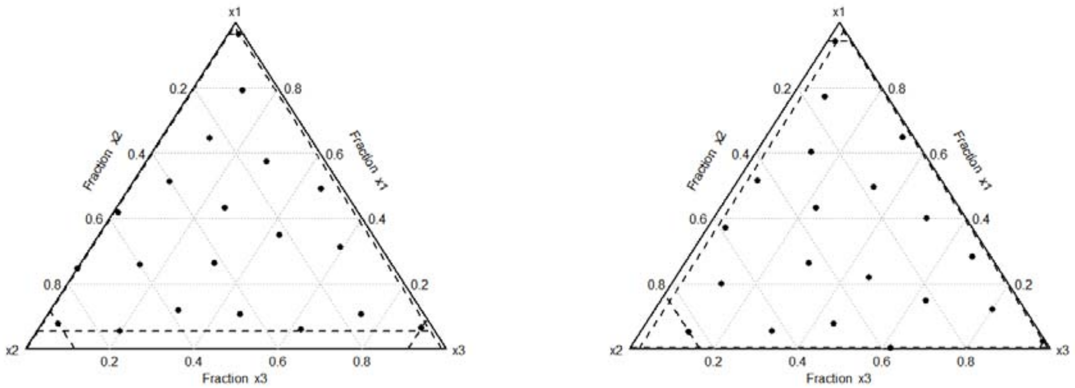


Figure 3: Solutions obtained by SbS algorithm with N=10 points for R=0.34 (a. and b.) and with N=20 points for R=0.21 (c. and d.)

3. Results and comparison

We can compare solutions obtained from these two algorithms by calculating the uniformity criteria such as *MinDist* and *Cov*.

		MinDist	***Cov***
N = 10 points	WSP (Fig 1 a.)	0.340	0.0097
	SbS (Fig 3 a.)	0.340	0
	SbS (Fig 3 b.)	0.340	0
	Random design	0.018	0.5661
N = 20 points	WSP (Fig 1 b.)	0.215	0.0319
	SbS (Fig 3 c.)	0.210	0
	SbS (Fig 3 d.)	0.210	0
	Random design	0.082	0.3624

Table 1: Comparison of uniformity criteria for mixture designs obtained by WSP and SbS algorithms

These values of uniformity criteria can be compared to those obtained with a random distribution with the same number of points. It is obvious that the two algorithms are equivalent and lead to uniform designs with good criteria.

4. Conclusion

When we work with dependent variables, several methods exist to build uniform experimental designs. In the case of mixtures, we propose two algorithms: the WSP algorithm which is adapted from the space-filling design for independent variables and a construction method which guarantees the uniform repartition of points in the dependent variables space. The solutions obtained by these algorithms present good values of criteria with a high value of *MinDist* and a low value of *Cov* measure that guarantee a good filling of the mixture space.

References

Scheffé, H. (1958). Experiments With Mixtures. Journal of the Royal Statistical Society. Series B (Methodological), 20(2), 344‑360

Fang, K.T., Li, R., & Sudjianto, A. (2006). Design and modeling for computer experiments. Chapman & Hall/CRC.

Johnson, M. E., Moore, L. M., & Ylvisaker, D. (1990). Minimax and maximin distance designs. Journal of Statistical Planning and Inference, 26(2), 131‑148.

Gunzburger, M., & Burkhardt, J. (2004). Uniformity measures for point samples in hypercubes, http://people.sc.fsu.edu/~jburkardt/pdf/ptmeas.pdf.

Santiago, J., Claeys-Bruno, M., & Sergent, M. (2012). Construction of space-filling designs using WSP algorithm for high dimensional spaces. Chemometrics and Intelligent Laboratory Systems, 133, 26-31.

Franco, J. (2008). Planification d’expériences numériques en phase exploratoire pour la simulation des phénomènes complexes. Thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne.