Space-filling designs with a Dirichlet distribution for mixture experiments

Astrid Jourdan

Space-filling designs with a Dirichlet distribution for mixture experiments

Astrid JOURDAN

ETIS UMR 8051, CY Paris University, 95000 Cergy, France

astrid.jourdan@cyu.fr

ABSTRACT: Uniform designs are widely used for experiments with mixtures. The uniformity of the design points is usually evaluated with a discrepancy criterion. In this paper, we propose a new criterion to measure the deviation between the design point distribution and a Dirichlet distribution. The support of the Dirichlet distribution, is defined by the set of d-dimensional vectors whose entries are real numbers in the interval [0,1] such that the sum of the coordinates is equal to 1. This support is suitable for mixture experiments. Depending on its parameters, the Dirichlet distribution allows symmetric or asymmetric, uniform or more concentrated point distribution. The difference between the empirical and the target distributions is evaluated with the Kullback-Leibler divergence. We use two methods to estimate the divergence: the plug-in estimate and the nearest-neighbor estimate. The resulting two criteria are used to build space-filling designs for mixture experiments. In the particular case of the flat Dirichlet distribution, both criteria lead to uniform designs. They are compared to an existing uniformity criterion. The advantage of the new criteria is that they allow other distributions than uniformity.

Keywords: Space-filling design, mixture experiments, Kullback-leibler divergence, nearest neighbor density estimation, kernel density estimation.

1. Introduction

Mixture experiments consist in varying the proportions of some components involved in a physico-chemical phenomenon, and observe the resulting change on the response. The proportions of the mixture components vary between 0 and 1 and they must sum to 1 for each run in the experiment. The experimental region is reduced to a (d-1)-dimensional simplex,

$$S^{d-1} = \{(x_1, \dots, x_d) | x_1 + \dots + x_d = 1, x_k \geq 0\},$$

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

The purpose of design for mixture experiments is to define a set of points in the simplex to catch as much information about the response as possible. Since Scheffé (1958) many authors have investigated designs for mixture experiments. The pioneers (Scheffé 1958, Kiefer 1961, Cornell 1981), defined optimal designs for linear and quadratic mixture models. An alternative approach of model-free designs is proposed by Wang and Fang (1990) and Fang and Wang (1994). The goal is to uniformly cover the experimental region. The main idea is to generate a uniform design on the (d-1) dimensional unit cube as explained in Hickernell (1998) or in Fang *et al.* (2005). Then they apply a mapping function to put the points in the simplex S^{d-1}. Following this principle, many articles suggested improvements specially to take into account complex constraints on the components, Fang and Yang (2000), Prescott (2008), Borkowski and Piepe (2009), Ning et al. (2011), Liu and Liu (2016).

The former design in the unit cube is uniform in the sense that the points minimize a discrepancy criterion. The discrepancy measures the distance between uniform distribution and the empirical distribution of the design points. It is not guarantee to conserve the uniformity after the mapping function. Some authors defined criteria to assess the uniformity of design for mixture experiments. Fang and Wang (1994) proposed to use the mean square distance (MSD), Borkowski and Piepe (2009) suggested the root mean squared distance, the maximum distance and the average distance, Chuang and Hung (2010) defined the central composite discrepancy. All these criteria require to compute the distance between the design points and the points of a much larger uniform set of points. The computational cost limits their usefulness in practice. To avoid this problem, Ning *et al.* (2011)

generalized the star discrepancy and proposed a new discrepancy, DM2_Discrepancy, to measure the uniformity of designs for mixtures. They also gave a computational formula of the DM2_discrepancy only based on the design points, which is useful in practice, specially to use it in an optimization algorithm to build a uniform design for mixture experiments.

In the same way, we defined in this paper a new criterion to measure the distribution of the design points in the simplex S^{d-1}. The purpose is to obtain uniform designs, and more generally designs with a Dirichlet distribution. Depending on its parameters, the Dirichlet distribution allows to obtain symmetric and asymmetric distributions, designs with points uniformly spread in the simplex or more concentrated in the center. We used the Kullback-leibler (KL) divergence to measure the difference between the design point distribution and the Dirichlet distribution. The resulting criterion is an estimator of the KL divergence computed with the design points. The KL divergence has already been used to define space-filling criteria but for a hypercube experimental domain (Jourdan and Franco 2009, 2010). The target distribution was the uniform distribution on the unit hypercube and the criterion was reduced to the estimation of the Shannon entropy. In this paper, we adapt the criterion to the Dirichlet distribution.

In section 2, we define the criterion from the Kullback-leibler divergence and the Dirichlet distribution. In section 3, we propose two methods to estimate the criterion. In section 4, we give some numerical examples.

2. Design points with a Dirichlet distribution

Suppose that the design points $\boldsymbol{x_1}, \dots, \boldsymbol{x_n}$, are n independent observations of the random vector $\boldsymbol{X} = (X_1, \dots, X_d)$ with absolutely continuous density function f concentrated on the simplex S^{d-1}. The aim is to select the design points in such a way as to have the corresponding empirical distribution "close" to the Dirichlet distribution.

Dirichlet distribution is a family of continuous multivariate probability distributions parameterized by a vector $\boldsymbol{\alpha}$ of positive reals. The support of the Dirichlet distribution is the (d-1)-simplex S^{d-1}. Its probability density function is

$$g(\boldsymbol{x}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{d} (x_k)^{\alpha_k - 1},$$

where $\boldsymbol{x}$ belongs to the (d-1)-simplex S^{d-1}, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ with $\alpha_i > 0$, and $B(\boldsymbol{\alpha})$ is the normalizing constant,

$$B(\boldsymbol{\alpha}) = \frac{\prod_{k=1}^{d} \Gamma(\alpha_k)}{\Gamma(\alpha_0)}$$

with $\alpha_0 = \sum_{k=1}^{d} \alpha_k$ and Γ the Gamma function.

A common special case is the symmetric Dirichlet distribution, where all of the elements making up the parameter vector $\boldsymbol{\alpha}$ have the same value α, called the concentration parameter. When $\alpha = 1$, the symmetric Dirichlet distribution is equivalent to a uniform distribution over the (d-1)-simplex S^{d-1}. It is called the flat Dirichlet distribution.

The aim is to generate n points in the simplex with a distribution as close as possible of a Dirichlet distribution. On Figures 2 and 3 (starting design), we can see that a simple random generation of the Dirichlet distribution is not efficient to obtain a good point distribution. Especially in the case of the flat Dirichlet distribution (Fig. 2a and 3d), the points do not uniformly cover the simplex: some points are very close to each other while some areas are empty.

We defined a criterion to measure the "distance" between the point distribution and the Dirichlet distribution. The criterion is then used in an optimization algorithm to build a set of points with the expected distribution.

There are different ways to measure the difference between two distributions. In this paper, we use the Kullback-Leibler divergence to evaluate the deviation between two probability density functions f and g,

$$I(f,g) = \int_{S^{d-1}} f(\boldsymbol{x}) log\left(\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\right) d\boldsymbol{x}$$

where g is the density function of the Dirichlet distribution and f is the density function of the design points. This integral can be written as the expected value of a random vector $\boldsymbol{X}$ with Dirichlet distribution,

$$I(f,g) = E\left[log\left(\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\right)\right].$$

If we consider that the design points $D = \{\boldsymbol{x_1}, \dots, \boldsymbol{x_n}\}$ are n i.i.d. realizations of a Dirichlet distribution, the Monte Carlo method gives an unbiased and consistent estimator,

$$\hat{I}(f,g) = \frac{1}{n}\sum_{i=1}^{n} \log(f(\boldsymbol{x_i})) - \log(g(\boldsymbol{x_i})).$$

By replacing g with the Dirichlet density function, we obtain,

$$\hat{I}(f,g) = \frac{1}{n}\sum_{i=1}^{n}\left[\log(f(\boldsymbol{x_i})) - \log\left(\prod_{k=1}^{d}(x_{ik})^{\alpha_k - 1}\right) + \log\big(B(\boldsymbol{\alpha})\big)\right] \quad (1)$$

with $x_{ik} \neq 0$, the kth component of the ith design point, i=1,…,n and k=1,…,d.

If the Dirichlet distribution is symmetric, then

$$\hat{I}(f,g) = \frac{1}{n}\sum_{i=1}^{n}\log(f(\boldsymbol{x_i})) - \frac{\alpha - 1}{n}\sum_{i=1}^{n}\sum_{k=1}^{d}\log(x_{ik}) + \log\big(B(\boldsymbol{\alpha})\big) \quad (2)$$

This expression is not a computational formula since the density function f is unknown. There are two common ways to estimate integral $I(f,g)$: the plug-in estimate which consists in replacing the density function f by its kernel estimate, and the nearest-neighbor estimate.

The two estimations are no more unbiased. However, having a bias is not a problem in our application, if the bias is fixed for a given n and d. The goal is not to obtain an accurate estimate of the integral but a criterion to compare two set of points in the optimization algorithm. We say that a design D_1 is better than a design D_2 if

$$\hat{I}(f_1,g) \leq \hat{I}(f_2,g)$$

with f_1 and f_2 the density functions associated to D_1 and D_2 respectively.

The optimization algorithm is an adaptation of the exchange algorithm described in Jin *et al.* (2005)

```
Exchange Algorithm
     1. Simulate n points from a Dirichlet distribution[1]
     2. Randomly select a new point in the simplex S^(d-1)
     3. For i in 1 to n
             Build a new design by replacing the i-th point by the
             new point
             Compute the criterion value of the new design
             Replace the current design with the new one iff the
             exchange improves the criterion
        End for i
     4. Repeat steps 2 and 3 until terminating condition is met
```

[1] Let $y_1,\dots,y_d$ be i.i.d realizations of the Gamma distribution with $y_k \sim \Gamma(\alpha_k,1)$. The random vector $\mathbf{x}=\mathbf{y}/S$ where $S=y_1+\dots+y_d$ has a Dirichlet distribution with parameter $\boldsymbol{\alpha}=(\alpha_1,\dots,\alpha_d)$.

3. Estimation of the criterion

In this section we propose two methods to estimate the unknown density function f in (1) and (2). In each case we explain our choices (kernel, bandwidth, k in the k-nearest neighbor distance) and we give a computational formula for the criterion.

3.1. Plug-in estimate

The unknown density function f is estimated with the design points $D = \{\boldsymbol{x_1}, \dots, \boldsymbol{x_n}\}$ by a kernel method (Scott 1992)

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n|H|^{1/2}} \sum_{i=1}^{n} K\big(\boldsymbol{H}^{-1/2}(\boldsymbol{x} - \boldsymbol{x_i})\big),$$

where K is a multivariate kernel and $\boldsymbol{H}$ is the bandwidth matrix (symmetric and positive definite matrix).

Choice of the bandwidth

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 $I(f, g)$ depends on the sample size n, the dimension d, and the bandwidth matrix $\boldsymbol{H}$. When constructing an optimal design, the size n and the dimension d are fixed. The bandwidth still needs to be fixed so that the bias does not vary during the optimization algorithm.

Usually the bandwidth matrix is chosen to be proportional to the covariance matrix of the data. This solution implies that $\boldsymbol{H}$ varies during the optimization algorithm. An idea to fix it, is to replace the covariance matrix of the data by the target covariance matrix, *i.e* the covariance matrix of the Dirichlet distribution. Unfortunately, this matrix is singular. Then, even if the variables are correlated, we simplify the bandwidth matrix into a diagonal matrix with the Scott's rule (1992), $\boldsymbol{H} = diag\big(h_1^2, \dots, h_d^2\big)$ with $h_k = n^{-1/(d+4)}\hat{\sigma}_k$. As previously, the estimation of the standard deviation of the k^{th} component, $\hat{\sigma}_k$ changes at each iteration of the optimization algorithm. In order to fix the bias, we replace it with the standard deviation of the target distribution,

$$\hat{\sigma}_k = \frac{1}{\alpha_0}\sqrt{\frac{\alpha_k(\alpha_0 - \alpha_k)}{(\alpha_0 + 1)}}, \qquad k = 1, \dots, d.$$

Choice of the kernel

It is known that the shape of the kernel has a minor influence on the estimation (Silverman 1986). We have chosen to use a multidimensional Gaussian kernel,

$$K(\boldsymbol{z}) = (2\pi)^{-d/2} e^{-\frac{1}{2}\|\boldsymbol{z}\|^2}.$$

A kernel of finite support (Epanechnikov, uniform) might have seemed more appropriate for the Dirichlet distribution. But, in our application, the kernel function has input values $\boldsymbol{z}$ in the interval $\big[0, d/h_k^2\big]$. This interval becomes very large when the size and the dimension increase, and the probability for $\boldsymbol{z}$ to be inside the kernel support then becomes very low. The estimation of f is then almost constant during the optimization process, and the criterion therefore does not allow to compare the designs.

Finally,

$$\hat{I}(\hat{f}, g) = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left\{\frac{1}{n}\sum_{j=1}^{n}\frac{(2\pi)^{-d/2}}{h_1 \dots h_d} exp\left(-\frac{1}{2}\sum_{k=1}^{d}\left(\frac{x_{jk} - x_{ik}}{h_k}\right)^2\right)\right\} - \log\left(\prod_{k=1}^{d}(x_{ik})^{\alpha_k - 1}\right) + \log\big(B(\boldsymbol{\alpha})\big)\right]$$

where $h_k = n^{-1/(d+4)} \frac{1}{\alpha_0} \sqrt{\frac{\alpha_k(\alpha_0 - \alpha_k)}{(\alpha_0 + 1)}}$, $k = 1, \dots, d$.

By removing the terms independent of the design points, we obtain a simplified criterion, especially for the symmetric Dirichlet distribution,

$$C_{kern}(D) = \sum_{i=1}^{n} \left[log \left(\sum_{j=1}^{n} e^{-\frac{1}{2}\left\| \frac{x_j - x_i}{h} \right\|^2} \right) \right] - (\alpha - 1) \sum_{i=1}^{n} \sum_{k=1}^{d} log(x_{ik}) \tag{3}$$

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

3.2. Nearest-neighbor estimate

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

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

$$\hat{I}(\hat{f}, g) = \frac{1}{n} \sum_{i=1}^{n} \left[-\log\left\{ (n-1) e^{-\psi(k)} V_d \left(\rho^{(k)}(\boldsymbol{x_i}, D_{-i}) \right)^d \right\} - \log\left(\prod_{k=1}^{d} (x_{ik})^{\alpha_k - 1} \right) + \log\big(B(\boldsymbol{\alpha})\big) \right]$$

with ψ the digamma function, V_d the volume of the unit ball in IR^d and $D_{-i} = D \backslash \{\boldsymbol{x_i}\}$. By removing the terms independent of the design points, we obtain the following criterion for a symmetric Dirichlet distribution,

$$C_{nn}(D) = -\sum_{i=1}^{n} \left[log \left\{ \left(\rho^{(k)}(\boldsymbol{x_i}, D_{-i}) \right)^d \right\} \right] - (\alpha - 1) \sum_{i=1}^{n} \sum_{k=1}^{d} log(x_{ik}) \tag{4}$$

The choice of k is discussed in the next section with numerical examples.

Remark: Note that criteria C_{kern} and C_{nn} are reduced to their first term for the flat Dirichlet (uniform) distribution ($\alpha = 1$).

Remark: The complexity of the two criteria is $O(n^2)$.

4. Numerical tests

In this section we built designs with the two criteria C_{kern} and C_{nn} and the optimization algorithm given in section 2 for different values of d, n and α. For each configuration, we built 30 designs to take into account the randomness of the algorithm.

When the target is the uniform distribution, we use the uniform DM2 criterion defined by Ning *et al.* (2011) in order to give some conclusions about C_{kern} and C_{nn} criteria.

4.1. Impact of the k value

In order to study the impact of the k value on the resulting C_{nn} design, different values have been tested between 1 and $2n/3$. Figure 1 represents the mean value of the C_{nn} criterion computed with 30 designs with $d = 5, n = 50, \alpha = 1$ and different values of k. We note that the criterion decreases with k. It is therefore important to use the same value of k to compare the designs in the optimization process.

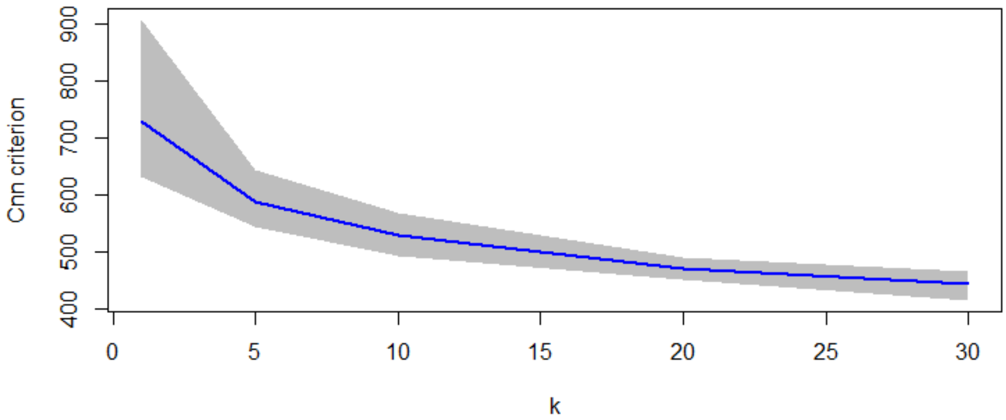


Fig.1. Average the C_{nn} criteria values computed on 30 C_{nn} designs with $d = 5$ and $n = 50$ and the flat Dirichlet distribution.

To determine the k value, we built 30 optimal designs with the C_{nn} criterion with $\alpha = 1$ and different values ok k. Table 1 gives the squared correlation between the DM2 criterion and the k value used to build the 30 designs. The square correlation is very small. It seems that the choice of k has no impact on the resulting design. Hereafter, we fix it to $k = 1$.

	$d = 3$ $n = 30$	$d = 4$ $n = 40$	$d = 5$ $n = 50$	$d = 6$ $n = 60$	$d = 7$ $n = 70$
k	$-0{,}04$	$-0{,}34$	$-0{,}36$	$0{,}14$	$-0{,}02$

Tab.1. Squared correlation between the DM2 criterion and the k values used to build the 30 designs with the C_{nn} criterion with $\alpha = 1$.

4.2. Design comparison

Figures 2 and 3 give some examples of designs obtained with C_{kern} and C_{nn} criteria from the same starting design and the same number of iterations in the optimization process. The criteria work well since the design points are more evenly spread in the simplex than the original ones.

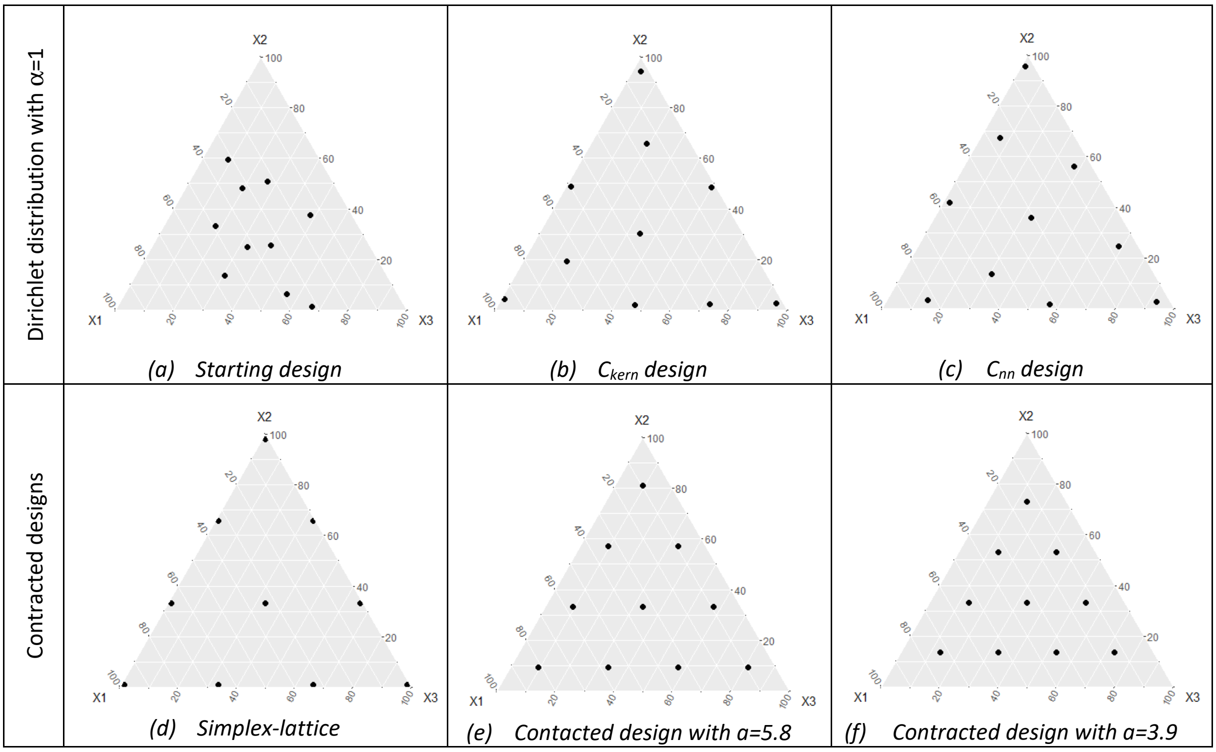


Fig.2. Uniform designs for mixture experiments with $d = 3$ and $n = 10$. The starting design points are n i.i.d random generation of the Dirichlet distribution (a). (b) and (c) are the resulting designs of the optimization algorithm with the same initialization (a). (e) and (f) are the contracted form of the simplex-lattice (d) ($4.2).

In the uniform distribution case $\alpha = 1$ (Fig. 2b&c, Fig. 3b&c), the design points explore in a uniform fashion the experimental domain. The designs obtained with C_{nn} are not so far from a regular grid (Fig. 2c and 3c). In the general case (Fig. 3), the point distributions of the resulting designs are better than those of the initial designs in the sense that there are no more points close to each other and no more empty area. In all cases, the C_{kern} criterion tends to put the points on the boundaries while the points obtained with C_{nn} are more concentrated in the middle. If we constraint the algorithm to put points near the boundaries, it degrades the C_{nn} criterion. Inversely, more points in the middle degrades the C_{kern} criterion.

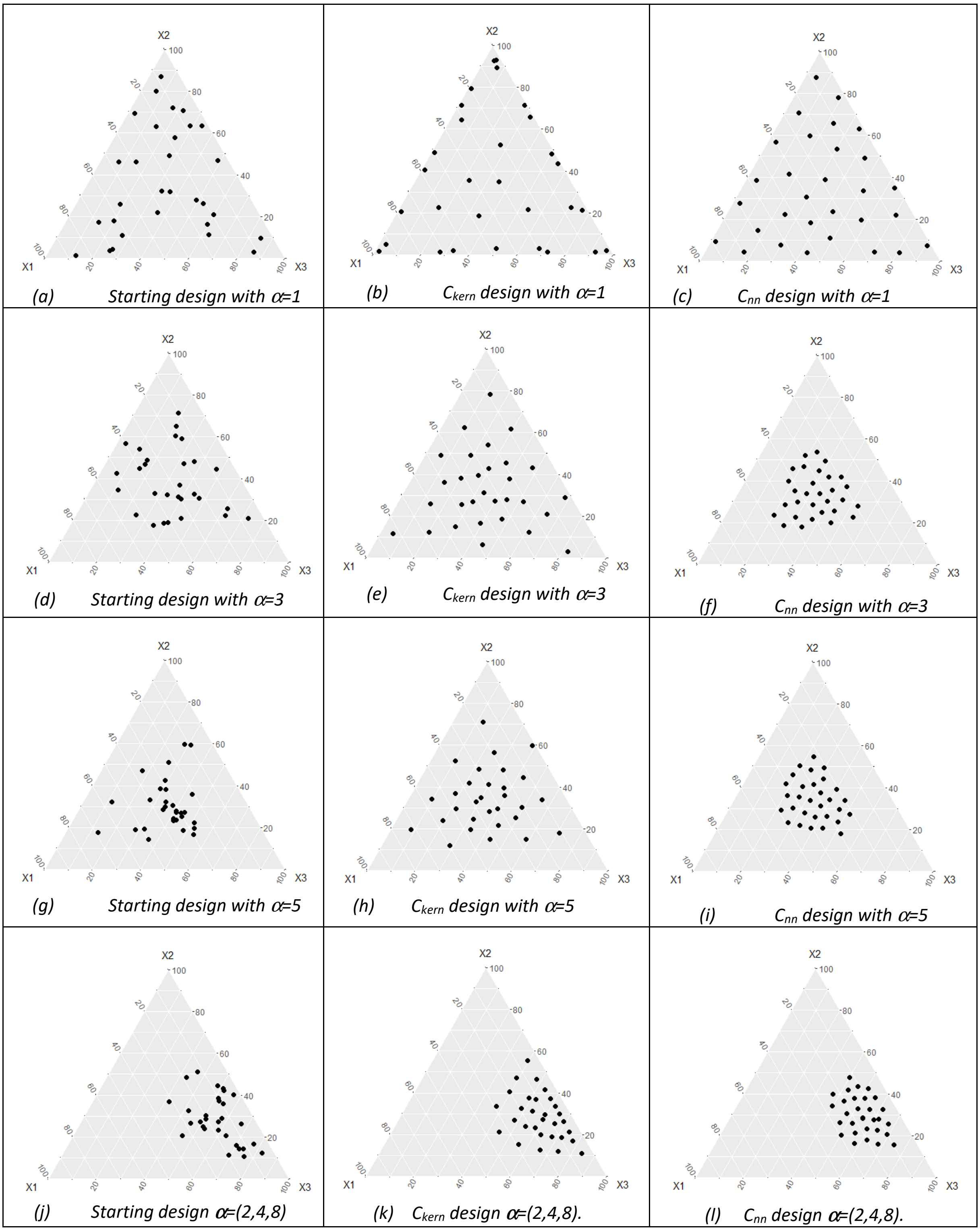


Fig.3. Designs with for $d = 3$ and $n = 30$ built with C_{kern} or C_{nn} criterion and different α values. The starting design points are n i.i.d random generation for the Dirichlet distribution (left column). The middle and right columns are the resulting designs of the optimization algorithm with the same initialization given in the left column.

Figure 4 represents the average of the DM2 criterion values for the sampling of 30 designs with $d = 3, n = 30$ and $\alpha = 0.5, 1, 3$ and 5. The DM2 criterion measures the uniformity of the designs. It should therefore reach its minimum value for $\alpha = 1$.

Indeed, we see that the minimum value is reached for $\alpha = 1$ for the designs built with criterion C_{nn}. On the other hand, for the designs built with criterion C_{kern}, the minimum value is reached for $\alpha = 3$. This confirms the previous conclusions, namely that criterion C_{kern} tends to push the points towards the edges of the simplex and that criterion C_{nn} with $\alpha = 1$ corresponds well to a uniformity criterion.

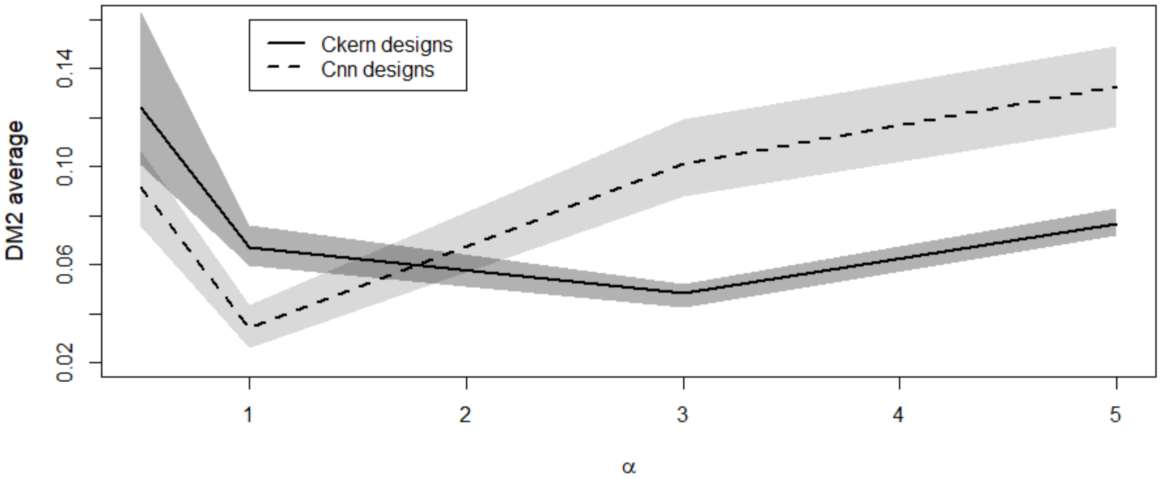


Fig.4. Average of the DM2 criterion values for the sampling of 30 designs with $d = 3$ and $n = 30$ with $\alpha = 0.5, 1, 3$ and 5.

In order to compare the three criteria (C_{kern}, C_{nn} with $\alpha = 1$ and DM2) in the case of the uniform distribution, we build 30 designs with $d = 3$ and $n = 30$ by optimizing criterion DM2 with the optimization process with the same starting designs and the same number of iterations. Figure 5 gives the boxplots of the three criteria calculated on the three type of designs. Of course, each type of design optimizes its own criteria. We notice that C_{nn} and DM2 designs have a behavior quite close. This reinforces the previous conclusion and allows us to conclude that criterion C_{nn} is more appropriate than criterion C_{kern} to construct uniform designs.
The interest of this criterion compared to the DM2 criterion, is that it also allows to obtain designs with distributions other than the uniform distribution. In particular, it is possible to construct designs with points concentrated in the middle of the simplex or asymmetric and which have a better distribution than a simple random sampling following the corresponding Dirichlet distribution (Fig. 3e,f,h,I,k and l).

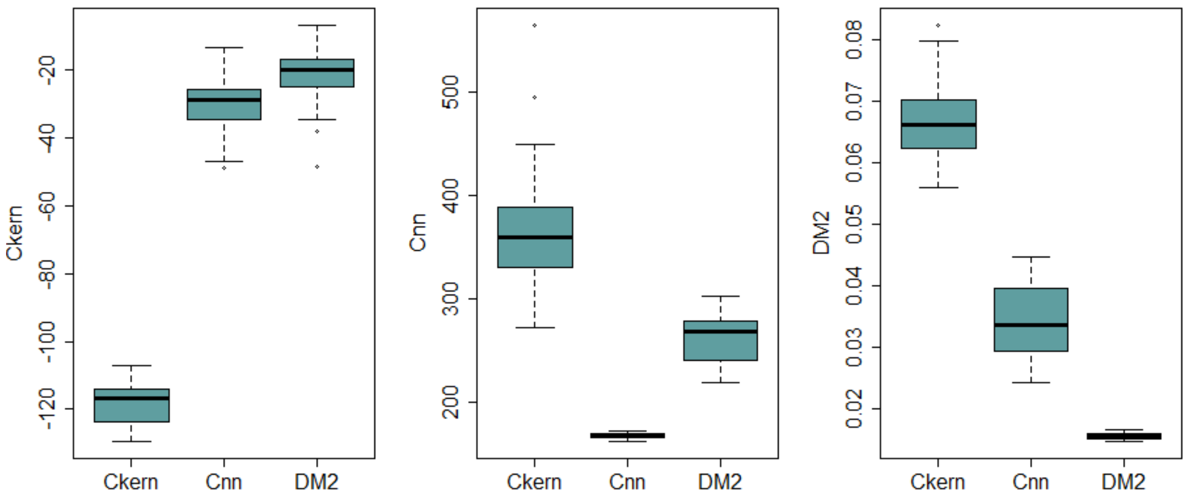


Fig.5. Boxplot of the three criteria calculated on the three type of designs with α=1.

4.3. Concentrated design

An alternative to build uniform design for mixture experiments is the contraction of a simplex-lattice (Scheffé 1958). The points of a simplex-lattice seem to be uniformly distributed on S^{d-1}but most of them lies on the boundaries (Fig. 2d). The mixture is then reduced to d-1 or d-2 ingredients and one or two ingredients are not in the mixture. Fang and Wang (1994) proposed to keep the simplex-lattice pattern while moving the points towards the centroid of the simplex. An example of a lattice-simplex and the contracted design is given in Table 2. The smaller the contraction constant *a*, the more the points are concentrated in the center (Fig. 2e and 2f). Ning *et al.* (2011) used DM2 criterion to find the best value of *a*. In the same way, we optimize the C_{kern} and C_{nn} criteria to determine *a* (Fig.6).

The C_{kern} criterion is optimal for a high value of *a* (*a*=13.8). This confirms that criterion C_{kern} tends to push the points on the edges of the simplex. The C_{nn} and DM2 criteria have the same behavior and give nearly the same value (*a*=5.8 with C_{nn} and *a*=5.26 with DM2).

Simplex-lattice design			Contracted design		
x_1	x_2	x_3	x_1	x_2	x_3
1	0	0	1-1/a	1/(2a)	1/(2a)
0	1	0	1/(2a)	1-1/a	1/(2a)
0	0	1	1/(2a)	1/(2a)	1-1/a
2/3	1/3	0	2/3-1/(2a)	1/3	1/(2a)
1/3	2/3	0	1/3	2/3-1/(2a)	1/(2a)
2/3	0	1/3	2/3-1/(2a)	1/(2a)	1/3
1/3	0	2/3	1/3	1/(2a)	2/3-1/(2a)
0	2/3	1/3	1/(2a)	2/3-1/(2a)	1/3
0	1/3	2/3	1/(2a)	1/3	2/3-1/(2a)
1/3	1/3	1/3	1/3	1/3	1/3

Tab.2. {3,3}-simplex lattice design and its contracted design.

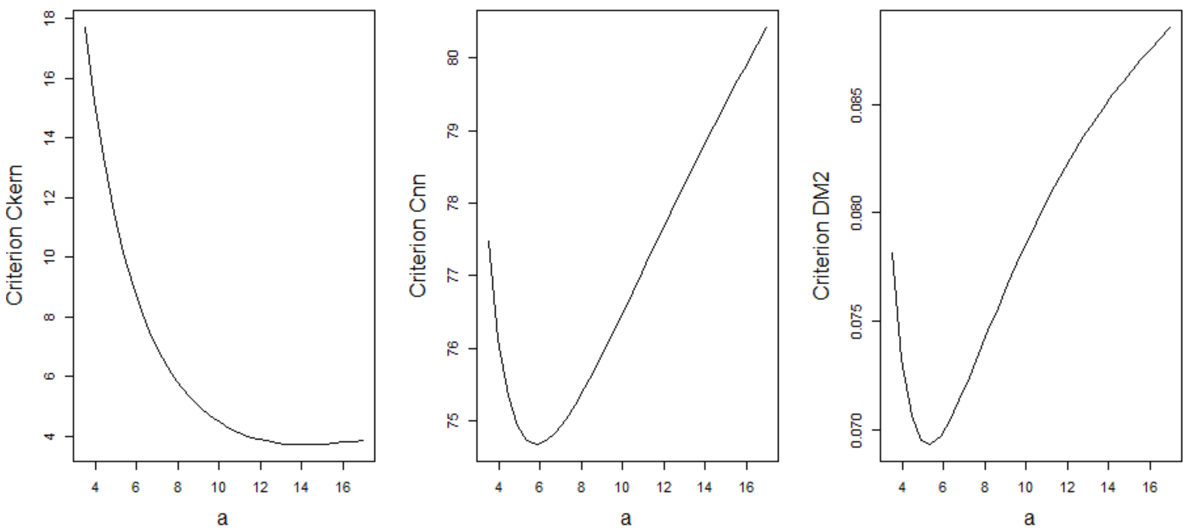
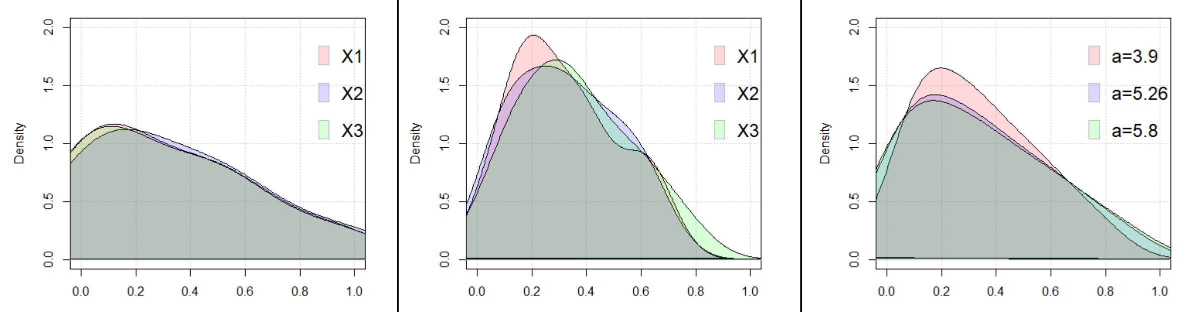


Fig.6. C_{kern}, C_{nn}, DM2 criteria against the contraction constant a. Best values of a are 13.8 with C_{kern}, 5.8 with C_{nn}, 5.26 with DM2.

4.4. Marginal distributions

The marginal distributions of the Dirichlet distribution are Beta distributions, $Beta(\alpha_i, \alpha_i - \alpha_0)$. In the special case of the uniform distribution ($\alpha = 1$), the distributions are $Beta(1, d-1)$. The asymmetric shape of the density function implies that the small values (proportions) of the components are over-represented while the larger values are under-represented. The same behavior is observed with the contracted design (Fig.7). There is no reason to make more experiments with small values. We would also like to have a uniform distribution for each of the components. But the two objectives, uniform distribution on the simplex and uniform distribution for each component, are conflicting.

Figure 7 illustrates the component distributions with the designs of Figure 2. Figures 7a and 7b show the three distributions of X_1, X_2 and X_3 for the C_{kern} and C_{nn} designs. Figure 7c displays only X_1 distribution of the contracted design but for different values of the contraction constant *a*. We can see that the marginal distributions are more uniform with the design computed with the C_{kern} criterion since they are very asymmetric for C_{nn} and contacted designs.



(a) C_{kern} design | *(b) C_{nn} design* | *(c) Contracted design (X_1 distribution)*

Fig.7. Component distributions for designs with $d = 3$ and $n = 10$ given in Fig.1.

5. Conclusion

In this paper we proposed a new class of designs for mixture experiments. The Dirichlet distribution allows to build design points with symmetric or asymmetric distribution, uniform or contracted distribution. The Kullback-Leibler divergence is used to measure the difference between the Dirichlet and design points distributions. We used the plugin estimate with a Gaussian kernel and the nearest neighbor estimate of the Kullbeck-Leibler divergence to define two criteria to assess the design point distribution. The two criteria are simplified to be used in an optimization process to build designs for mixture experiments with a target Dirichlet distribution.

The numerical tests and the comparison with the existing criteria DM2 show that C_{nn} criterion performs better than C_{kern} criterion to build uniform designs. With the same α value, the C_{kern} criterion tends to push the points on the boundaries while the C_{nn} criterion concentrates the points in the middle of the simplex. The advantage of the criteria proposed in this article compare to the existing criteria, is that they allow the construction of designs with distributions other than uniform, symmetric or asymmetric.

However, our method presents a shortcoming. The optimization algorithm converges very slowly and requires many iterations until terminating condition is met (the same inconvenient is observed for criterion DM2). An early stopping of the optimization process may produce poor quality designs. One idea to speed up convergence is to choose the new point in a large NT-net instead of choosing it randomly in the simplex (step 2 in the exchange algorithm).

Another drawback of uniform design for mixture experiments (not only C_{kern} and C_{nn} designs) is the asymmetric distribution of each component. Having a uniform distribution on the simplex S^{d-1} and a symmetric distribution on each axis seem to be two conflicting objectives. A multi-objective optimization algorithm (instead of the exchange algorithm) would allow to manage this problem. The first objective function would be one of the two criteria defined in this paper. The second objective function could be defined in order to measure the difference between the distribution of each component and a symmetric distribution with support [0,1] (e.g. symmetric triangular or truncated normal distribution). As we did in this paper, the Kullback-Leibler divergence and its estimates could be used to define this second objective function.

References

Borkowski JJ, Piepel GF (2009) Uniform designs for highly constrained mixture experiments. Journal of Quality Technology. https://doi.org/10.1080/00224065.2009.11917758

Chuang SC, Hung YC (2010) Uniform design over general input domains with applications to target region estimation in computer experiments. Computational Statistics & Data Analysis. https://doi.org/10.1016/j.csda.2009.08.008

Cornell JA (1981) Experiments with Mixtures, designs, models, and the analysis of mixture data. Wiley, New York.

Fang KT, Li R, Sudjianto A (2005) Design Modeling for Computer Experiments. Chapman & Hall/CRC Press, London. https://doi.org/10.1201/9781420034899

Fang KT, Wang Y (1994) Number-theoretic Methods in Statistics. Chapman & Hall, London. https://doi.org/10.1007/978-1-4899-3095-8

Hickernell FJ (1998) A generalized discrepancy and quadrature error bound. Mathematics of Computation. https://doi.org/10.1090/S0025-5718-98-00894-1

Jin R, Chen W, Sudjianto A (2005) An efficient algorithm for constructing optimal design of computer experiments. Journal of Statistical Planning and Inference. https://doi.org/10.1016/j.jspi.2004.02.014

Joe H (1989) Estimation of entropy and other functional of multivariate density. Annals of the Institute of Statistical Mathematics. https://doi.org/10.1007/BF00057735

Jourdan A, Franco J (2009) Plans d'expériences numériques d'information de Kullback-Leibler minimale. Journal de la Société Française de Statistique, 150 (2), 52-64.

Jourdan A, Franco J (2010) Optimal Latin hypercube designs for the Kullback-Leibler criterion. AStA Advances in Statistical Analysis. https://doi.org/10.1007/s10182-010-0145-y

Kiefer J (1961) Optimum designs for regression model, II. The Annals of Mathematical Statistic. https://doi.org/10.1214/aoms/1177705160

Leonenko N, Pronzato L, Savani V (2008) A class of Rényi information estimators for multidimensional densities. The annals of Statistics. https://doi.org/10.1214/07-AOS539

Liu Y, Liu M (2016) Construction of uniform designs for mixture experiments with complex constraints. *Communications in Statistics - Theory and Methods*. https://doi.org/10.1080/03610926.2013.875576

Ning JH, Zhou YD, Fang KT (2011) Discrepancy for uniform design of experiments with mixtures, Journal of Statistical Planning and Inference. https://doi.org/10.1016/j.jspi.2010.10.015

Prescott P (2008) Nearly uniform designs for mixture experiments. Communication in Statistics - Theory and Methods. https://doi.org/10.1080/03610920701824257

Scheffé H (1958) Experiments with Mixtures. Journal of the Royal Statistical Society: Series B. https://doi.org/10.1111/j.2517-6161.1958.tb00299.x

Scott DW (1992) Multivariate Density Estimation: Theory, practice and visualization. John Wiley & Sons, New York, Chichester. https://doi.org/10.1002/9780470316849

Wang Q, Kulkarni R, Verdu S (2006) A Nearest-Neighbor Approach to Estimating Divergence between Continuous Random Vectors. 2006 IEEE International Symposium on Information Theory. https://doi.org/10.1109/ISIT.2006.261842

Wang Y, Fang KT (1990) Number theoretic methods in applied statistics (II). Chinese Annals of Mathematics - Series B. https://doi.org/10.1142/9789812701190_0039