REGULAR ARTICLE

Space-filling designs with a Dirichlet distribution for mixture experiments

Astrid Jourdan[1]

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 existing uniformity criteria. The advantage of the new criteria is that they allow other distributions than uniformity and they are fast to compute.

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

---

✉ Astrid Jourdan
  astrid.jourdan@cyu.fr

[1] ETIS UMR 8051, CY Paris University, 95000 Cergy, France

to 1 for each run in the experiment. The experimental region is reduced to a (d-1)-dimensional simplex,

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

where x_k is the proportion of the kth component, $k = 1, \ldots, d..$

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 Piepel (2009), Ning et al. (2011), and 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 the cumulative function of the uniform distribution and the empirical cumulative function of the design points. It is not guaranteed 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 Piepel (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 drawback, 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 probability density function of the design point distribution and the probability density function of the Dirichlet distribution. 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. We propose two methods

to estimate the KL divergence, a plug-in estimation and a nearest neighbor estimation. This leads to two criteria for assessing the distribution of the design points.

Applied with the flat Dirichlet distribution, the new criteria lead to designs with a uniform distribution of their points but they are not uniform designs in the sense defined by Fang and Wang (1994). The new criteria are based on the density probability function whereas the discrepancy for uniform designs is based on the cumulative distribution function.

In Sect. 2, we define the criterion from the Kullback–Leibler divergence and the Dirichlet distribution. In Sect. 3, we propose two methods to estimate the criterion. In Sect. 4, we carry out a numerical comparison between the new and existing criteria in the case of the uniform distribution. In Sect. 5, we propose two applications, one concerning simplex-lattice designs and the other on the marginal distribution of components.

2 Design points with a Dirichlet distribution

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 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

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

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

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

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

Hereafter, we focus on the symmetric Dirichlet distribution, that is all of the elements making up the parameter vector $\boldsymbol{\alpha}$ have the same value α, called the concentration parameter, and we suppose that $\alpha \geq 1$. 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 Fig. 1a (starting design), we can see that a simple random generation of the Dirichlet distribution is not efficient to obtain a good point distribution. The points do not uniformly cover the simplex: some points are very close to each other while some areas are empty.
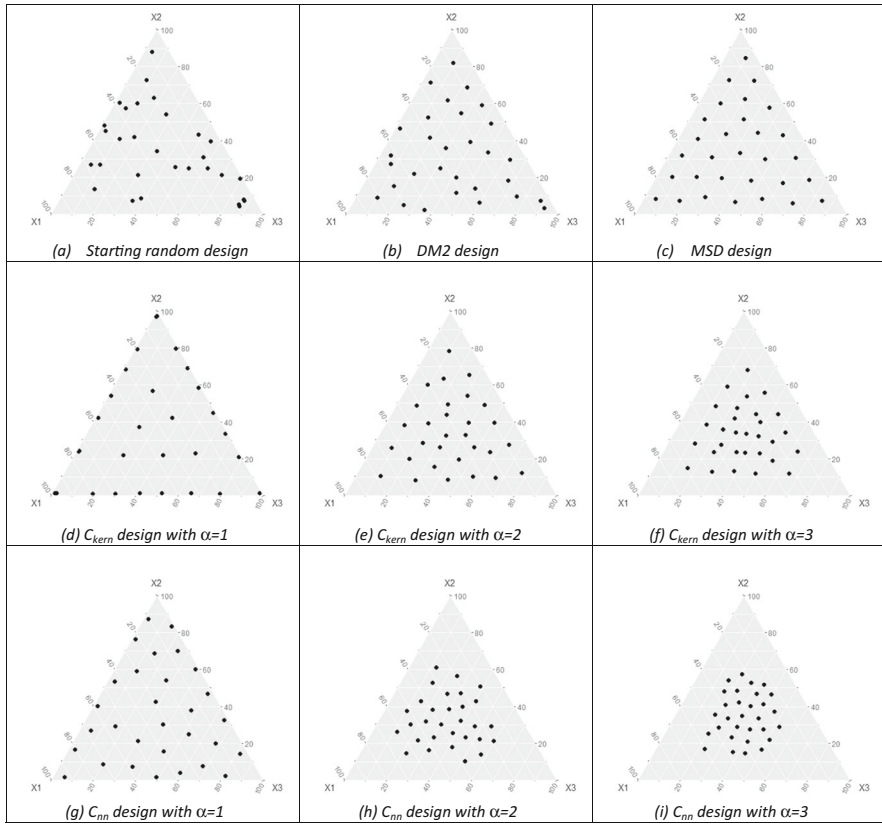
Fig. 1 Designs with for $d = 3$ and $n = 30$. The flat Dirichlet random sampling **a** is a simple random generation of 30 points with a Dirichlet distribution with $\alpha = 1$. DM2, MSD, C_{kern} and C_{nn} designs **b**–**f** are the resulting designs of the optimization algorithm

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 the case of uniform design, discrepancies are based on the cumulative distribution function (Fang et al. 2005). In this paper, we use the Kullback–Leibler divergence to evaluate the deviation between two probability density functions f and g,

$$I(f, g) = \int f(\boldsymbol{x}) log\left(\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\right) d\boldsymbol{x}.$$

This integral can be written as the expected value of a random vector X distributed according to f,

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

We denote

$$I(f, g) = I_f(f) - I_f(g)$$

where $I_f(f) = E\big[\log(f(X)\big]$ and $I_f(g) = E\big[\log(g(X)\big].$

Theorem 1. Let g be the probability density function of the Dirichlet distribution (1), then integral $I_f(g)$ exists.

The proof of this theorem is given in Appendix A.

Throughout we suppose that f is the unknown density function of the design points such that integral $I_f(f)$ exists. This assumption is feasible since the goal is to obtain a density function f close to the Dirichlet density function g. Then we can use the Kullback–Leibler divergence to evaluate the deviation between the design points distribution and the Dirichlet distribution.

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

$$\hat{I}(f, g) = \hat{I}_f(f) - \hat{I}_f(g) \tag{2}$$

where $\widehat{I}_f(f) = \frac{1}{n}\sum_{i=1}^{n}\log(f(x_i))$ and

$$\hat{I}_f(g) = \frac{(\alpha - 1)}{n}\sum_{i=1}^{n}\sum_{k=1}^{d}\log(x_{ik}) + \log(B(\boldsymbol{\alpha}))$$

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

The estimator $\widehat{I}_f(f)$ 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. We detail the two approaches in the next section.

The two estimations are not 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 sets 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 minimization algorithm is an adaptation of the exchange algorithm described in Jin et al. (2005)

```
Exchange Algorithm
    1. Simulate n points from a Dirichlet distribution
    2. Randomly select a new point in the simplex Sᵈ⁻¹
    3. For i in 1 to n
            Build a new design by replacing the iᵗʰ 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
```

3 Estimation of the criterion

In this section we propose two methods to estimate the unknown density function f in Eq. 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 $D = \{x_1, \ldots, x_n\}$ by a kernel method (Scott 1992)

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

where K is a multivariate kernel and H is the bandwidth matrix (symmetric and positive definite matrix). It is known that the shape of the kernel has a minor influence on the estimation (Silverman 1986). We use a multidimensional Gaussian kernel,

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

On the contrary, the choice of the bandwidth matrix has a great influence on the accuracy of the estimation. We use a diagonal matrix, $H = h^2 I_d$, where

$$h = n^{-1/(d+4)} \frac{1}{\alpha_0} \sqrt{\frac{\alpha(\alpha_0 - \alpha)}{(\alpha_0 + 1)}}.$$

This choice is motivated by Theorem 2.

Theorem 2. We consider the estimator $\widehat{I}_f(\widehat{f})$ of $\widehat{I}_f(f)$. Suppose that f has continuous first and second order derivatives and $\int f(\boldsymbol{x}) log^2(f(\boldsymbol{x}))d\boldsymbol{x}$ exists then the bias is.

$$E\left[\hat{I}_f(f) - \hat{I}_f\left(\hat{f}\right)\right] = O\left(n^{-1}h^{-d}\right) + O\left(h^2\right).$$

The proof of this theorem is given in Appendix B.

The bias depends on the sample size n, the dimension d, and the bandwidth 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), $\boldsymbol{H} = diag\left(h_1^2, \ldots, h_d^2\right)$ with $h_k = n^{-1/(d+4)}\widehat{\sigma}_k$, where $\widehat{\sigma}_k$ is the estimation of the standard deviation of the k th component. The estimate $\widehat{\sigma}_k$ depends on the design points, so h_k and thus the bias varies from one iteration to another in the algorithm. In order to fix the bias, we will replace the estimate h_k by a value independent of the design points. Since our goal is to get closer to a Dirichlet distribution, the most obvious value for $\widehat{\sigma}_k$ is the standard deviation of the target distribution,

$$\hat{\sigma}_k = \frac{1}{\alpha_0}\sqrt{\frac{\alpha_k(\alpha_0 - \alpha_k)}{(\alpha_0 + 1)}}.$$

Finally, by removing the terms independent of the design points and with $\alpha_k = \alpha$, $k = 1, \ldots, d$, we obtain a simplified criterion,

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

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 $\rho^{(1)}(\boldsymbol{x}, S) \leq \rho^{(2)}(\boldsymbol{x}, S) \leq \ldots \leq \rho^{(m)}(\boldsymbol{x}, S)$, the ordered distances between $\boldsymbol{x} \in \mathrm{IR}^d$ and $S = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y_m}\}$ 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 $I_f(f)$ with the design points $D = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x_n}\}$ is asymptotically unbiased and consistent,

$$\hat{I}_f\left(\hat{f}\right) = -\frac{1}{n} \sum_{i=1}^{n} \log\left\{(n-1)e^{-\psi(k)} V_d\left(\rho^{(k)}(\boldsymbol{x_i}, D_{-i})\right)^d\right\}$$

with ψ the digamma function, V_d the volume of the unit ball in IR^d and $D_{-i} = D\backslash\{\boldsymbol{x_i}\}$. Note that in this expression, we suppose that $\boldsymbol{x_i} \neq \boldsymbol{x_j}$. The bias depends on n, d and k. We need to fix the value of k so that the bias does not vary during the optimization algorithm. Pronzato (2017) justified to restrict the estimation to $k = 1$.

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

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

Remark 1. Note that criteria C_{kern} and C_{nn} are reduced to their first term for the flat Dirichlet (uniform) distribution ($\alpha = 1$), which are estimations of the Shannon entropy of the random vector X (except the coefficient $1/n$).

Remark 2. As the points get closer to the edges of the simplex, the second term increases. This means that the criteria will favor points inside the simplex. The higher the α coefficient, the more the points will be in the center of the simplex, which respects the behavior of the Dirichlet distribution.

Remark 3. Since the criteria are based on the Euclidian distance, they are invariant under permuting factors or runs, and invariant under rotation of the coordinates.

4 Numerical tests

There is no criterion in the literature (except C_{kern} and C_{nn}) to assess whether a sample follows a Dirichlet distribution in the general case. In the case of the flat distribution, one can use existing criteria defined to evaluate the uniform distribution of points in a simplex. Most of the criteria compute the distances between the design points and the points of a much larger number-theoretic set within the simplex. Among these distance-based criteria, we select the mean square distance (MSD) defined by Fang and Wang (1994),

$$MSD(D) = \frac{1}{N} \sum_{i=1}^{N} \min_{1 \leq j \leq n} d^2(x_j, z_i),$$

where d is the Euclidian distance between x_j and z_i, and $z_1, \ldots, z_N$ are N points of a *glp* sets[1] given in Fang and Wang (1994) with $N = 610, 597, 701, 1069$ and 2129 for $d = 3$ to 7 respectively, $N = 3997$ for $d = 8$ and 9, and $N = 4661$ for $d = 10$ (which is the smallest size found in Fang and Wang 1994). The computational cost of

---

[1] *Good Lattice Point* sets.

the distance-based criteria increases rapidly with the dimension d and the size n (see Sect. 4.1) and limits their usefulness in practice. As far as we know, only the DM2 criterion defined by Ning et al. (2011) does not involve calculation with a large set of points. It is an adaptation of the star discrepancy to the simplex and it is estimated with the design points only,

$$
DM2(D) = \left(\frac{\sqrt{d}}{(d-1)!} \right)^{\frac{1}{2}} \left\{ c_{n,s} - \frac{2(d-1)!}{n} \sum_{i=1}^{n} \sum_{(\tau_2, \ldots, \tau_d) \in \{0,1\}^{d-1}} a_\tau . \right.
$$

$$
\times (x_{i1})^{2(d-1) - \sum_{j=2}^{2} \tau_j} \cdot \prod_{j=2}^{d} (x_{ij})^{\tau_j} + \frac{1}{n^2}
$$

$$
\left. \times \sum_{i=1, k=1}^{n} \left(\max \left(1 - \sum_{j=2}^{d} \max(z_{ij}, z_{kj}), 0 \right) \right)^{d-1} \right\}^{\frac{1}{2}}
$$

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

In this section we use the optimization algorithm given in Sect. 2 (with 1000 iterations) to build designs with the four criteria C_{kern}, C_{nn}, DM2 and MSD and for different values of d, n and α. For each configuration, we built several designs to consider the randomness in the initialization of the algorithm. Table 1 shows the correlation between the four criteria computed with 1000 random designs with $n = 10d$. The correlation is fairly weak especially when the dimension increases. This means that the criteria do not operate in the same way to assess the uniform distribution of the points. In the following sections, we compare the performance and behavior of the criteria.

4.1 Design comparison in the case of the flat distribution

In Fig. 1, we use the same starting design (Fig. 1a) in the exchange algorithm in order to visually compare the resulting designs for $d = 3$ and $n = 30$. The starting design is a random set of points according the flat Dirichlet distribution. We observe that some points are very close together, providing redundant information, while some areas in the simplex are not explored by the points. The designs obtained the criteria DM2, MSD, C_{kern} and C_{nn} with $\alpha = 1$ (Fig. 1b, c, d, g) explore more uniformly the experimental domain. Some points are still close together with the DM2 criterion but there is no more empty area. The point distribution of the criteria MSD and C_{kern} designs with $\alpha = 1$ (Fig. 1c, d) is very regular like a grid distribution. The C_{kern} criterion tends to push the points on the edges of the simplex. Figures 1e, f, h and i illustrate that as α increases, the points are more concentrated inside the simplex. In this case, the designs are not space-filling since they don't explore the entire domain,

Table 1 Correlation between the four criteria computed with 1000 random designs with $n = 10d$

	$d = 3$			$d = 5$			$d = 7$			$d = 9$		
	C_{nn}	C_{kern}	DM2	C_{nn}	C_{kern}	DM2	C_{nn}	C_{kern}	DM2	C_{nn}	C_{kern}	DM2
C_{kern}	0.17	1		0.15	1		0.23	1		0.32	1	
DM2	0.12	0.37	1	0.07	0.02	1	0.00	− 0.04	1	0.00	− 0.05	1
MSD	0.15	0.61	0.41	0.02	0.29	0.05	0.02	0.17	− 0.06	− 0.11	− 0.11	− 0.05
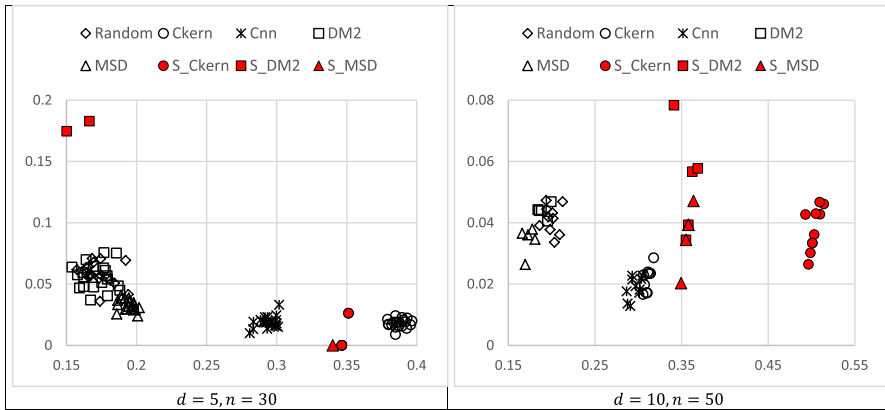
Fig. 2 Average (x-axis) and standard deviation (y-axis) the of the nearest neighbor distances of the design points with $\alpha = 1$ for C_{kern} and C_{nn} criteria

but the points are well distributed, i.e. they're not too close together and they explore the concentrated area evenly.

This first visual comparison illustrates the fact that a simple random draw according to the flat Dirichlet distribution is not enough, and an optimization algorithm with an appropriate criterion is necessary to construct a design of experiments for mixtures with points evenly spread in the simplex. We have drawn some conclusions about the behavior of the criteria in dimension 3, but a visual comparison is not sufficient to draw conclusions in dimensions greater than 3. That is why we introduce a graphical tool (Fig. 2) in order to compare the inter-site distance of the design points in any dimension. For a design, we compute the nearest neighbor distance of each point. The x-axis is the average of the nearest neighbor distances of the design points (μ) and the y-axis is the standard deviation (σ). A good coverage of the experimental region is obtained by a design with points far from each other (high average) and close to a regular grid (small standard deviation) like a scrambled grid. Then the target area is at the bottom right of this graphic. In Fig. 2, we have designs in dimension $d = 5$ and size $n = 30$ (left), and dimension $d = 10$ and $n = 50$ (right). In both cases C_{kern} criterion gives the best results since the points are on average far from their nearest neighbor. The C_{nn} criterion is not as good in dimension 5, but gives almost the same results in dimension 10. The DM2 and MSD criteria have the same results than simple random designs, the average of the inter-site distance is smaller with a high standard deviation. This means that some points are close to each other and will provide redundant information (the red points will be explained in Sect. 5.2). The new criteria are better than the existing ones in terms of inter-site distance.

The next comparison is about the computational time. The complexity is.

- $O(d \times n^2)$ for C_{kern} and C_{nn} criteria,
- $O((d-2) \times n^2 + n \times 2^{d-1} \times (d-2) + 2(d-2))$ for DM2 criterion,
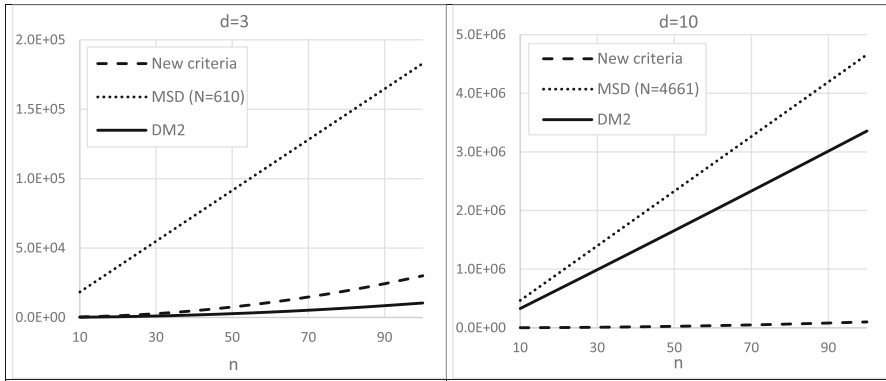- $O(d \times n \times N)$ for MSD criterion.

Fig. 3 Evolution of complexity as a function of size *n*

Figure 3 gives the evolution of the complexity as a function of *n* for $d = 3$ and $d = 10$. The calculation time for the MSD criterion is high, due to the calculation of the distance between the design points and the points of a larger *glp* set. Time increases dramatically with size, even if we choose the smallest size for the *glp* set in Fang and Wang (1994). The DM2 criterion has the lowest complexity for $d = 3$ but the complexity becomes very high for $d = 10$. The significant cost of the two existing criteria explains why they are only five designs in Fig. 2 for $d = 10$.

4.2 Design behavior according to α

Figure 4 represents the average of DM2 and MSD criterion values of 20 designs with $d = 3$, $n = 30$ according to α. DM2 and MSD criteria measure the uniformity of the designs. They should therefore reach their minimum value for $\alpha = 1$, and this is the case for C_{nn} designs. However, the C_{kern} designs have a minimal value for
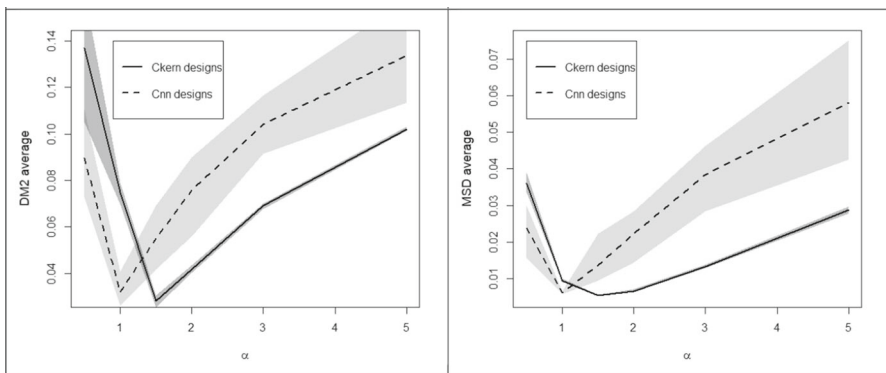

Fig. 4 Average of the DM2 and MSD criterion values for the sampling of 20 designs with $d = 3$ and $n = 30$ (95% confidence interval in grey)

$\alpha = 1.5$. This means that the C_{kern} criterion tends to push the points on the edges of the simplex, and that α must be increased to bring them back to the center. This confirms our conclusion from the previous paragraph when visually comparing the 3-dimensional design in Fig. 1. We also note that there is less variability in criterion values for C_{kern} designs.

5 Applications

5.1 Concentrated design

An alternative to build a 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 lie on the boundary (Fig. 5a) and some experiments are reduced one or two components in the mixture (e.g. the first experiment in the {3,3}-simplex lattice in Table 2 involves the first component X_1 only). 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 and Fig. 5.

The smaller the contraction constant a, the more the points are concentrated in the center. Fang and Wang (1994) and Ning et al. (2011) used MSD and DM2 criteria 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_{nn} criterion is optimal for a high value of a ($a = 9.6$). It means that C_{nn} criterion tends to push the points inside the simplex. The C_{kern} criterion find an optimal value very close to the values obtained by Fang and Wang (1994) and Ning et al. (2011) ($a \cong 5$).
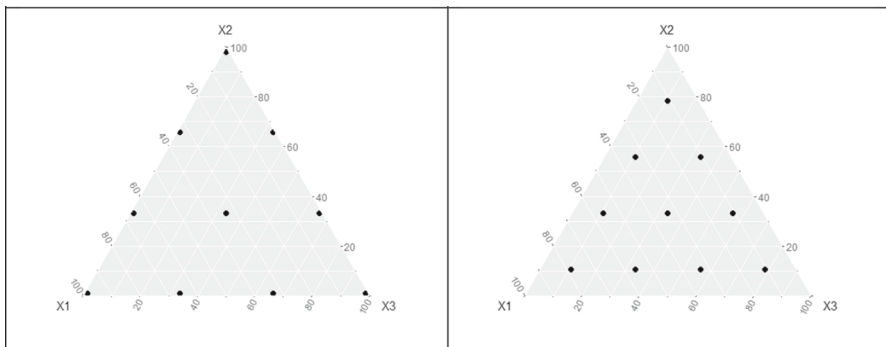

Fig. 5 Simplex-lattice (left) and contracted designs with $a = 5$ (right) with $d = 3$ and $n = 10$

Table 2 {3,3}-Simplex lattice design and its contracted design

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
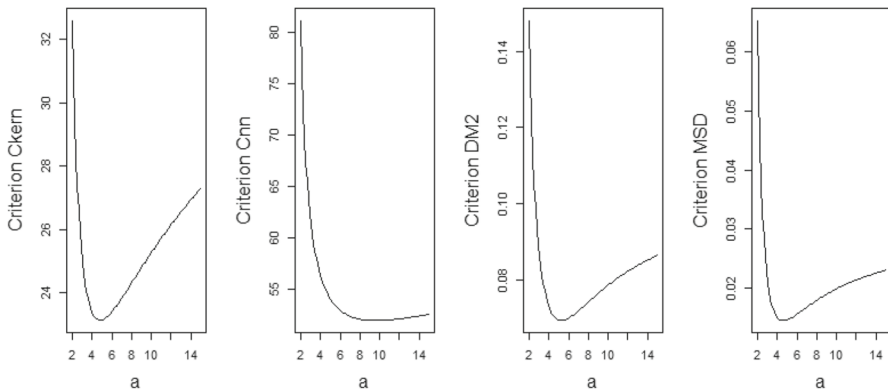
Fig. 6 C_{kern}, C_{nn}, DM2 and MSD criteria against the contraction constant a. Best values of a are 4.9 with C_{kern}, 9.6 with C_{nn}, 5.3 with DM2 and 4.6 with MSD

5.2 The curse of dimensionality and marginal distribution

As the dimension d increases, some phenomena appear that cannot be ignored.

The first one is the prohibitive size of the {d,q} simplex-lattice when d increases, $n = (d + q - 1)!/(d - 1)!\,q!$. Some examples are given in Table 3. In dimension 10, the simplex-lattice with $q = 2$ requires $n = 55$ experiments and tests only three levels {0, 0.5, 1} for each component. If we need to test more levels, {0, 1/3, 2/3, 1} with $q = 3$ for example, the size increases to $n = 220$ experiments. The idea is to use the optimization algorithm with the previous criteria to select a well-distributed subset of points in the simplex-lattice. The red points in Fig. 2 are the designs obtained by this method with a simplex-lattice with $q = 4$. The inter-site distance of designs

Table 3 Example of sizes for a $\{d,q\}$ simplex-lattice

d	5	5	5	5	10	10	10	10
q	2	3	4	5	2	3	4	5
n	15	35	70	126	55	220	715	2002
Levels	3	4	5	6	3	4	5	6

constructed with the MSD criterion increases considerably in dimension 5 and 10. In dimension 5, the DM2 and C_{kern} criteria perform less well than their original versions, with DM2 in particular showing strong variability in inter-site distance. In dimension 10, all criteria increase the inter-site distance, with still high variability for the DM2 criterion. There is no result with the C_{nn}. The nearest-neighbor distance is constant when we restrict the experimental domain to the points of the simplex-lattice, so the optimization algorithm does not converge.

The second phenomenon that arises as the dimension increases is that the optimization process tends to push the points to the edges of the experimental domain. This phenomenon is well known in the construction of space-filling designs in the unit cube. It is reinforced in the case of design for mixture experiments by the fact that the marginal distributions of the Dirichlet distribution are Beta distributions, $Beta(\alpha_i, \alpha_0 - \alpha_i)$. In the special case of the uniform distribution ($\alpha = 1$), the distributions are $Beta(1, d - 1)$. As shown in Fig. 7, the skewness of the density function increases with dimension. When the dimension is large, small proportions are over-represented in the experimental design. To avoid this inconvenience, we can build designs with a Dirichlet distribution, which allows us to control the distribution of small proportions by choosing an appropriate α value. For example, if X is a random variable with a
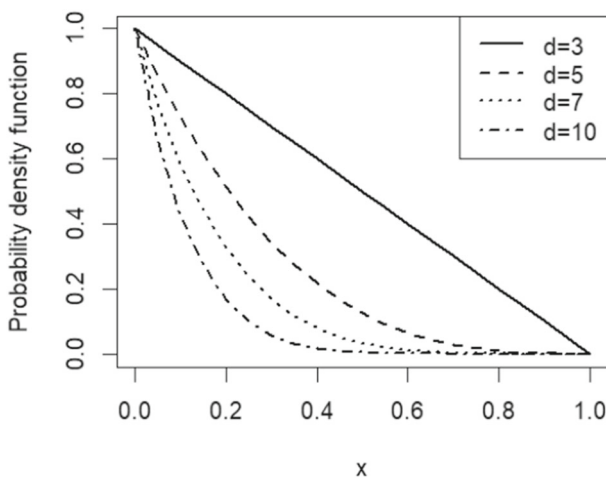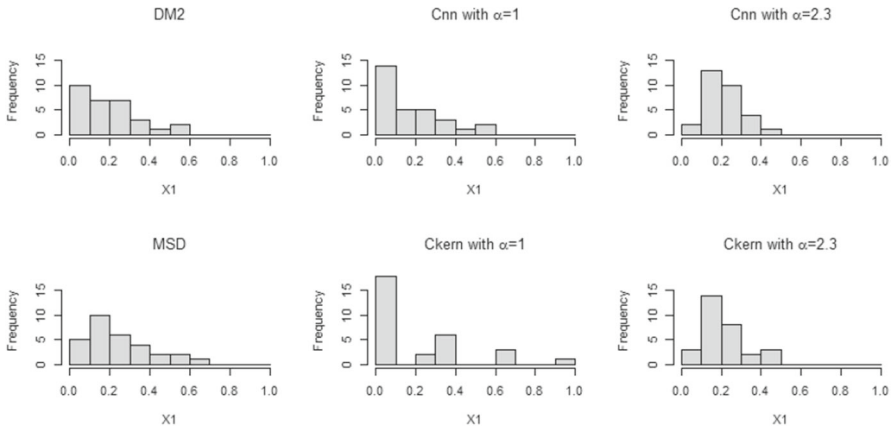

Fig. 7 Density function of Beta(1,d-1)

Fig. 8 Marginal distribution of X_1 for designs with $d = 5$ and $n = 30$

Dirichlet distribution, then $P(X < 0.1) = 20\%$ implies that $\alpha = 2.3$ for $d = 5$. Figure 8 provides the marginal distribution of X_1 for different designs with $d = 5$ and $n = 30$ build with the optimization algorithm. Note that with $\alpha = 2.3$, the frequency of component proportion below 0.1 decreases considerably. The objective is therefore achieved. On the other hand, increasing α contracts the design symmetrically. As a result, the frequency of large proportions also decreases. There is no longer any proportion greater than 0.5.

6 Conclusion

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

Numerical tests in dimension 3 show that the criteria allow to evenly spread the points in the simplex as well as the existing criteria. Tests in higher dimensions show that the new criteria give better results. The distance between points is higher with the new criteria than with the existing DM2 and MSD criteria, and with lower variability. On the other hand, calculation time for the existing criteria increases considerably as the dimension increases. This makes them difficult to use for mixtures with many components. The new criteria therefore seem to be the best choice in this case.

We have proposed two applications in the high-dimensional case. The first comes from the observation that the number of points in a simplex-lattice becomes excessive as the dimension increases, especially if we wish to test many levels for each
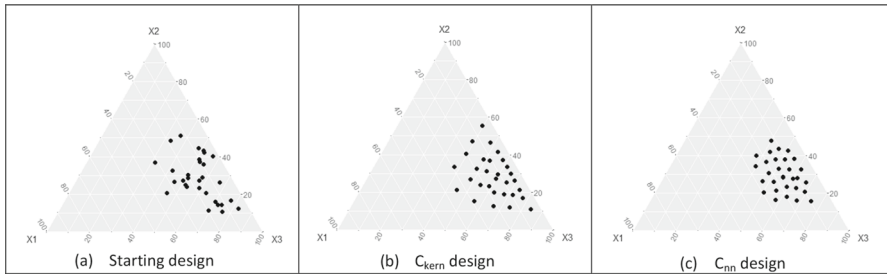
Fig. 9 C_{kern} and C_{nn} designs with $d = 3$ and $n = 10$ for an asymmetric Dirichlet distribution with $\boldsymbol{\alpha} = (2,4,8)$. The starting design points are n i.i.d random generation of the Dirichlet distribution

component. The criteria are then used to select a subset of simplex-lattice points that are well distributed over the experimental domain. The C_{kern} and MSD criteria give good results in terms of inter-site distance. The advantage of the C_{kern} criterion lies in its computational speed. The second application comes from the observation that low proportions are over-represented in the design, and that this phenomenon is amplified as the dimension increases. One of the advantages of the two criteria proposed in this paper is that they are based on Dirichlet distribution, which deals with more than just uniform distribution. We can determine the value of the α parameter to control the frequency of small proportions in the experimental design. We can also set different values of α depending on the component and obtain an asymmetrical design as shown in Fig. 9.

The second application is not entirely satisfactory, because by reducing the frequency of small proportions, we also reduce the large proportions that are already under-represented. Having a uniform distribution on the simplex S^{d-1} and a symmetric distribution on each axis seems 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 univariate 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 a criterion for the second objective function. The Pareto front could be used to find the best compromise between the two objectives.

Appendices

Appendix A. The proof of theorem 1

We apply the Jensen's inequality to the expected value $I_f(g) = E\left[\log(g(X))\right]$.

Let denote the function $\varphi(\boldsymbol{x}) = \log(g(\boldsymbol{x}))$,

$$\varphi(\boldsymbol{x}) = \log\left(\frac{1}{B(\boldsymbol{\alpha})}\prod_{k=1}^{d}(x_k)^{\alpha_k-1}\right) = \sum_{k=1}^{d}(\alpha_k - 1)\log(x_k) - \log(B(\boldsymbol{\alpha})).$$

φ is a concave function since the logarithmic function is concave and $(\alpha_k - 1)$ is positive for $\alpha_k \geq 1$. The Jensen's inequality implies that, $E[\varphi(X)] \leq \varphi(E[X])$. Let $E[X] = (\mu_1, \ldots, \mu_d)$, then

$$\varphi(E[X]) = \sum_{k=1}^{d}(\alpha_k - 1)\log(\mu_k) - \log(B(\boldsymbol{\alpha})) < \infty$$

since $0 < \mu_k < 1$ ($supp(X_k) = [0, 1]$ and we exclude the special case of a constant random variable equals to 0).

Appendix B. The proof of theorem 2

The prof of Theorem 2 is a direct application of a result demonstrated by Joe (1989). We just have to verify the assumptions.

The choice of a Gaussian kernel satisfies the conditions.

$K(-z) = K(z)$
The kernel is of the form $K(z) = K(z_1, \ldots, z_d) = \prod_{j=1}^{d} K_0(z_j)$ where K_0 is a symmetric univariate density satisfying $\int u^2 K_0(u)du = 1$.

The d components of X have approximately the same scale in $[0,1]$, the logarithmic function is thrice differentiable. Moreover, we suppose that $\int f(\boldsymbol{x})log(f(\boldsymbol{x}))d\boldsymbol{x}$ and $\int f(\boldsymbol{x})log^2(f(\boldsymbol{x}))d\boldsymbol{x}$ exists. Hence all conditions are satisfied to apply the results demonstrated by Joe (1989).

We have already noted that the existence hypothesis of $\int f(\boldsymbol{x})log(f(\boldsymbol{x}))d\boldsymbol{x}$ is feasible since f is close to g, and we proved the existence of this integral in Theorem 1. However, we have not demonstrated the existence of $\int f(\boldsymbol{x})log^2(f(\boldsymbol{x}))d\boldsymbol{x}$ when $f = g$. This is demonstrated below in the case of $d = 2$ to simplify notation. It remains true in the general case.

$$I = \int f(\boldsymbol{x})log^2(f(\boldsymbol{x}))d\boldsymbol{x} = \int_{S^1} f(x_1, x_2)log^2(f(x_1, x_2))dx_1dx_2.$$

The line $x_1 + x_2 = 1$ has the parametric representation,

$$\begin{cases} x_1(t) = -t \\ x_2(t) = 1 + t \end{cases}$$

where $t \in [-1, 0]$. We define the mapping function $M : [-1, 0] \rightarrow [0, 1] \times [0, 1], M(t) = (-t, 1 + t)$. Then,

$$I = \int_{-1}^{0} f(-t, 1+t) log^2(f(-t, 1+t))\left(x_1'(t), x_2'(t)\right)dt$$

$$= \sqrt{2} \int_{-1}^{0} f(-t, 1+t) log^2(f(-t, 1+t))dt$$

If $f = g$,

$$I = \frac{\sqrt{2}}{B(\alpha)} \int_{-1}^{0} (-t)^{\alpha_1-1}(1+t)^{\alpha_2-1} log^2\left(\frac{1}{B(\alpha)}(-t)^{\alpha_1-1}(1+t)^{\alpha_2-1}\right)dt = I_1 + I_2 + I_3$$

With

$$I_1 = \frac{\sqrt{2}}{B(\alpha)} \int_{-1}^{0} (-t)^{\alpha_1-1}(1+t)^{\alpha_2-1} log^2\left(\frac{1}{B(\alpha)}\right)dt$$

$$I_2 = (\alpha_1 - 1)\frac{\sqrt{2}}{B(\alpha)} \int_{-1}^{0} (-t)^{\alpha_1-1}(1+t)^{\alpha_2-1} log^2(-t)dt$$

$$I_3 = (\alpha_2 - 1)\frac{\sqrt{2}}{B(\alpha)} \int_{-1}^{0} (-t)^{\alpha_1-1}(1+t)^{\alpha_2-1} log^2(1+t)dt$$

$I_1 < +\infty$ since $\alpha_i \geq 1$.

I_2 is an improper integral in 0, but $(-t)^{\alpha_1-1}(1+t)^{\alpha_2-1}log^2(-t) \sim (-t)^{\alpha_1-1}log^2(-t)$ when t tends to 0 and $\int_{-1}^{0}(-t)^{\alpha_1-1}log^2(-t)dt$ is a convergent Bertrand's integral.

I_3 is an improper integral in -1, but $(-t)^{\alpha_1-1}(1+t)^{\alpha_2-1}log^2(1+t) \sim (1+t)^{\alpha_2-1}log^2(1+t)$ when t tends to -1 and $\int_{-1}^{0}(1+t)^{\alpha_2-1}log^2(1+t)dt$ is a convergent Bertrand's integral.

References

Borkowski JJ, Piepel GF (2009) Uniform designs for highly constrained mixture experiments. J Qual Technol. 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. Comput Stat Data Anal. 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, Wang Y (1994) Number-theoretic methods in statistics. Chapman & Hall, London. https://doi.org/10.1007/978-1-4899-3095-8

Fang KT, Li R, Sudjianto A (2005) Design modeling for computer experiments. Chapman & Hall, London. https://doi.org/10.1201/9781420034899

Hickernell FJ (1998) A generalized discrepancy and quadrature error bound. Math Comput. 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. J Stat Plan Inference. https://doi.org/10.1016/j.jspi.2004.02.014

Joe H (1989) Estimation of entropy and other functional of multivariate density. Ann Inst Stat Math. https://doi.org/10.1007/BF00057735

Jourdan A, Franco J (2009) Plans d'expériences numériques d'information de Kullback-Leibler minimale. J Soc Fr Stat 150(2):52–64

Jourdan A, Franco J (2010) Optimal Latin hypercube designs for the Kullback-Leibler criterion. Adv Stat Anal. https://doi.org/10.1007/s10182-010-0145-y

Kiefer J (1961) Optimum designs for regression model, II. Ann Math Stat. https://doi.org/10.1214/aoms/1177705160

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

Liu Y, Liu M (2016) Construction of uniform designs for mixture experiments with complex constraints. Commun Stat. https://doi.org/10.1080/03610926.2013.875576

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

Prescott P (2008) Nearly uniform designs for mixture experiments. Commun Stat. https://doi.org/10.1080/03610920701824257

Pronzato L (2017) Minimax and maximin space-filling designs: some properties and methods for construction. J Soc Fr Stat 158(1):7–36

Scheffé H (1958) Experiments with mixtures. J R Stat Soc Ser B. https://doi.org/10.1111/j.2517-6161.1958.tb00299.x

Scott DW (1992) Multivariate density estimation: theory, practice and visualization. Wiley, New York. https://doi.org/10.1002/9780470316849

Wang Y, Fang KT (1990) Number theoretic methods in applied statistics (II). Chin Ann Math. https://doi.org/10.1142/9789812701190_0039

Wang Q, Kulkarni R, Verdu S (2006) A nearest-neighbor approach to estimating divergence between continuous random vectors. In: 2006 IEEE international symposium on information theory. https://doi.org/10.1109/ISIT.2006.261842