# Minimal Plus One Design Points to Test Lack of Fit for Second-degree Mixture Model

Yanyan Li *

Teva Branded Pharmaceutical Products R&D, Inc., Frazer, PA 19355, USA

Damaraju Raghavarao

Department of Statistics, Temple University, Philadelphia, PA 19122, USA

**Abstract**

D-optimal minimal designs have been obtained for different mixture models with minimal support design points, i.e., the number of design points is equal to the number of parameters in the model. To test Lack of Fit, we need to add at least one additional distinct design point. Those distinct design points are within the design space with all factor levels greater than zero for practical reasons. We will discuss second-degree mixture model and second-degree mixture model with main effects and two factor interactions including one common factor. The new design will be compared with other possible designs for testing Lack of Fit.

*Keywords:* D-optimal Minimal Design, Lack of Fit, Second-degree Mixture Model

## 1 Introduction

Mixture experiments, where the predictor variables are proportions of ingredients adding to 1, are increasingly used in industry, such as food processing, chemical formulations, textile fibers, pharmaceutical drugs, etc. Consider a product formed by q non-negative factor levels $(x_1, x_2, \ldots, x_q)$ such that $\sum_{i=1}^{q} x_i = 1$, $x_i \geq 0$ for all i. The q-proportions can be expressed as a column vector $\mathbf{x} = (x_1, \ldots, x_q)'$ in the (q-1)-dimensional simplex space. In this context, $(x_1, x_2, \ldots, x_q)$ are called design points.

Scheffé (1958,1961,1963) presented the canonical polynomial mixture models such as linear, second-degree, special cubic mixture model, etc. D-optimal designs have been obtained for different types of mixture models. For example, Kiefer (1961) has found the D-optimal design for second-degree model. Lim (1990) has determined the D-optimal design for special cubic model. Chan (2000) summarized analytic and numeric solutions of optimal designs for various regression models for experiments with mixtures, which include polynomial models, log contrast models, models containing inverse terms, models with quantitative variables, etc. Those designs usually contain minimal support design points, i.e., the number of design points is equal to the number of parameters in the model. For a detailed discussion on mixture designs, the reader is referred

*Correspondence to Yanyan Li, Global Biostatistics, Teva Branded Pharmaceutical Products R&D, Inc. 18 Moores Rd, Frazer, PA 19355, USA. Email: yanyan.li@tevapharm.com

to Cornell (2002). We need to add at least one additional design point to test Lack of Fit. Those additional design points are not any of the minimal support design points.

In this paper, we start with D-optimal minimal design and aim to find one additional design point to test Lack of Fit. The location of the additional design point is inside the design space and away from the boundary points (such as vertices, edges, faces etc) for practical reasons. The new design will be compared with other possible designs for testing Lack of Fit.

# 2 Second-degree Mixture Model

Scheffé's second-degree mixture model fits data well in many cases and has been used extensively. It is defined as

$$y = \sum_{i=1}^{q} \beta_i x_i + \sum_{j=1}^{q} \sum_{i<j}^{q} \beta_{ij} x_i x_j + \epsilon. \tag{1}$$

There are a total of $\frac{q(q+1)}{2}$ parameters in the model and hence at least $\frac{q(q+1)}{2}$ design points are needed to estimate all parameters. Without loss of generality, we consider second-degree mixture model with 3 or more factors.

**D-optimal Minimal Design**

Kiefer (1961) proved that the $(q, 2)$ simplex-centroid design is D-optimal. Galil and Kiefer (1977) showed that this design performs well in terms of variance and bias, for fitting the second-degree mixture model. The design assigns equal weight to each of the extreme vertices $\leftrightarrow$ $(1, 0, \dots, 0)$ and the edge midpoints $\leftrightarrow$ $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$. Consider the design matrix

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \dots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 1/2 & 1/2 & \dots & 0 & 1/4 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \dots & \vdots \\ 0 & \dots & 1/2 & 1/2 & 0 & \dots & 1/4 \end{bmatrix} = \begin{bmatrix} \mathbf{X_{11}} & \mathbf{X_{12}} \\ \mathbf{X_{21}} & \mathbf{X_{22}} \end{bmatrix},$$

where $\mathbf{X_{11}} = \mathbf{I}_q$, where $\mathbf{I}_q$ is an identity matrix of order $q$, $\mathbf{X_{12}}$ is a zero matrix of $q \times \frac{q(q-1)}{2}$, $\mathbf{X_{22}} = \frac{1}{4}\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}}$, and

$$\mathbf{X_{21}} = (x_{ij,k}) = \begin{cases} \frac{1}{2} & \text{when } i = k \;\; or \;\; j = k, \\ 0 & \text{otherwise} \end{cases}$$

with $ij$ representing all interaction terms of factors $i$ and $j$, $i, j, k = 1, 2, \dots, q$ and $i < j$.

To test Lack of Fit, we need to add at least one additional distinct design point. The additional point extends the design matrix $\mathbf{X}$ with the additional row

$$\mathbf{z}' = \begin{bmatrix} x_1 & x_2 & \dots & x_q & x_1x_2 & \dots & x_{q-1}x_q \end{bmatrix},$$

repeated $t$ times to provide error sum of squares to test Lack of Fit. We take t responses and take its average as the required response in the analysis. Assume the responses to be independent and the variance-covariance matrix for the response vector is

$$\mathbf{V} = \begin{bmatrix} \mathbf{I}_{\frac{\mathbf{q(q+1)}}{\mathbf{2}}} & 0 \\ 0 & 1/t \end{bmatrix},$$

where $t$ is the number of replicates of additional design point consisting of the first q components of $\mathbf{z}'$.
The determinant of the new information matrix is

$$|\mathbf{X}^{*'}\mathbf{V}^{-\mathbf{1}}\mathbf{X}^{*}| = |\mathbf{X}^{'}\mathbf{X}|[1 + t\mathbf{z}^{'}(\mathbf{X}^{'}\mathbf{X})^{-\mathbf{1}}\mathbf{z}],$$

where $\mathbf{X^*}$ is the new design matrix augmenting $\mathbf{z}'$ as the last row to the design matrix $\mathbf{X}$. It is a linear increasing function of $t$. We choose $t$ as the value constrained by the fixed total budget. Let $C_i$ be the cost of collecting data at the $i$th point and $C_0$ be the total cost. We have

$$t = (C_0 - \sum_{j=1}^{\frac{q(q+1)}{2}} C_j)/C_{(\frac{q(q+1)}{2}+1)}.$$

Under such constraint, maximizing the determinant $|\mathbf{X}^{*'}\mathbf{V}^{-\mathbf{1}}\mathbf{X}^{*}|$ is equivalent to maximizing $\mathbf{z}^{'}(\mathbf{X}^{'}\mathbf{X})^{-\mathbf{1}}\mathbf{z}$.

**Inverse of the X′X Matrix**

Let

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{bmatrix},$$

where $\mathbf{A_{11}} = \frac{q+2}{4}\mathbf{I_q} + \frac{1}{4}\mathbf{J_q}$, $\mathbf{A_{22}} = \frac{1}{16}\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}}$, and $\mathbf{J_q}$ is the matrix of ones of order $q$. $\mathbf{A_{12}} = (a_{k,ij})$ is a $q \times \frac{q(q-1)}{2}$ matrix,

$$\mathbf{A_{12}} = (a_{k,ij}) = \begin{cases} \frac{1}{8} & \text{when } k = i \;\; or \;\; k = j, \\ 0 & \text{otherwise.} \end{cases}$$

where $i, j, k = 1, 2, \ldots, q$ and $i < j$ and $\mathbf{A_{21}} = \mathbf{A_{12}}^{'}$.
Since $\mathbf{X}'\mathbf{X}$ is non-singular, let

$$(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} = \begin{bmatrix} \mathbf{D_{11}} & \mathbf{D_{12}} \\ \mathbf{D_{21}} & \mathbf{D_{22}} \end{bmatrix} = \begin{bmatrix} \mathbf{A_{11}}^{-1}(\mathbf{I} + \mathbf{A_{12}}\mathbf{F}^{-1}\mathbf{A_{21}}\mathbf{A_{11}}^{-1}) & -\mathbf{A_{11}}^{-1}\mathbf{A_{12}}\mathbf{F}^{-\mathbf{1}} \\ -\mathbf{F}^{-1}\mathbf{A_{21}}\mathbf{A_{11}}^{-1} & \mathbf{F}^{-1} \end{bmatrix},$$

where $\mathbf{F} = \mathbf{A_{22}} - \mathbf{A_{21}}\mathbf{A_{11}}^{-1}\mathbf{A_{12}}$ and is non-singular. It can be verified that

$$\mathbf{A_{11}}^{-1} = \frac{4}{q+2}\{\mathbf{I_q} - \frac{1}{2(q+1)}\mathbf{J_q}\},$$

$$\mathbf{A_{12}}\mathbf{A_{21}} = \frac{q-2}{64}\mathbf{I_q} + \frac{1}{64}\mathbf{J_q},$$

and

$$\mathbf{A_{21}A_{12}} = (a_{jl,j'l'}) = \begin{cases} \frac{1}{32} & \text{when } j = j' \quad \text{and} \quad l = l' \\ \frac{1}{64} & \text{when } j = j' \quad \text{or} \quad j = l' \quad \text{or} \quad l = j' \quad \text{or} \quad l = l' \\ & \text{and } j \neq j' \quad \text{and} \quad l \neq l' \\ 0 & \text{otherwise} \end{cases}$$

where $j, j', l, l' = 1, 2, \ldots, q, \quad j < l$ and $j' < l'$.
Note that

$$\mathbf{A_{21}A_{12}} = \frac{1}{32}\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}} + \frac{1}{64}\mathbf{B_1},$$

where $\mathbf{B_1}$ is the association matrix of the first associates in a triangular association scheme of order $\frac{q(q-1)}{2}$ (See Raghavarao, 1971). Let $\mathbf{B_0}$, $\mathbf{B_1}$ and $\mathbf{B_2}$ be the association matrices of a triangular association scheme. We know the following results from Raghavarao (1971):

$$\mathbf{B_0} = \mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}}, \quad \sum_{i=0}^{2} \mathbf{B_i} = \mathbf{J}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}}, \tag{2}$$

$$\mathbf{B_1B_2} = (q-3)\mathbf{B_1} + (2q-8)\mathbf{B_2}, \tag{3}$$

$$\mathbf{B_1^2} = 2(q-2)\mathbf{B_0} + (q-2)\mathbf{B_1} + 4\mathbf{B_2}, \tag{4}$$

and

$$\mathbf{B_2^2} = \frac{(q-2)(q-3)}{2}\mathbf{B_0} + \frac{(q-3)(q-4)}{2}\mathbf{B_1} + \frac{(q-4)(q-5)}{2}\mathbf{B_2}. \tag{5}$$

Now we have

$$\begin{aligned} \mathbf{F} &= \mathbf{A_{22}} - \mathbf{A_{21}}\mathbf{A_{11}}^{-1}\mathbf{A_{12}} \\ &= \frac{1}{16}\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}} - \frac{4}{q+2}(\frac{1}{32}\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}} + \frac{1}{64}\mathbf{B_1}) + \frac{1}{8(q+1)(q+2)}\mathbf{J}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}}. \end{aligned} \tag{6}$$

$\mathbf{F}$ can be rewritten as $\mathbf{F} = a_0\mathbf{B_0} + a_1\mathbf{B_1} + a_2\mathbf{B_2}$, where $a_0 = \frac{q^2+q+2}{16(q+1)(q+2)}$, $a_1 = -\frac{q-1}{16(q+1)(q+2)}$, and $a_2 = \frac{1}{8(q+1)(q+2)}$.
Let

$$(a_0\mathbf{B_0} + a_1\mathbf{B_1} + a_2\mathbf{B_2})^{-1} = b_0\mathbf{B_0} + b_1\mathbf{B_1} + b_2\mathbf{B_2}.$$

We solve for $b_0, b_1, b_2$ in terms of $a_0, a_1$ and $a_2$, using (2) - (5), and get

$$b_0 = 24 \quad , \quad b_1 = 4 \quad , \quad b_2 = 0.$$

Hence

$$\mathbf{D_{22}} = \mathbf{F^{-1}} = 24\mathbf{B_0} + 4\mathbf{B_1} = 24\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}} + 4\mathbf{B_1}.$$

Since

$$\mathbf{A_{12}B_1} = (q-4)\mathbf{A_{12}} + \frac{1}{4}\mathbf{J}_{\mathbf{q}\times\frac{\mathbf{q(q-1)}}{\mathbf{2}}},$$

we have

$$\begin{aligned}
\mathbf{D_{12}} &= -\mathbf{A_{11}}^{-1}\mathbf{A_{12}}\mathbf{F^{-1}} \\
&= \frac{-4}{q+2}(\mathbf{I_q} - \frac{1}{2(q+1)}\mathbf{J_q})\mathbf{A_{12}}(24\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}} + 4\mathbf{B_1}) \\
&= -16\mathbf{A_{12}} + \{\frac{2}{(q+1)} - \frac{4}{(q+2)} + \frac{2q}{(q+1)(q+2)}\}\mathbf{J}_{\mathbf{q}\times\frac{\mathbf{q(q-1)}}{\mathbf{2}}} \\
&= -16\mathbf{A_{12}},
\end{aligned}$$

$$\mathbf{D_{21}} = \mathbf{D'_{12}} = -\mathbf{16A_{21}},$$

and

$$\begin{aligned}
\mathbf{D_{11}} &= \mathbf{A_{11}}^{-1}(\mathbf{I_q} + \mathbf{A_{12}}\mathbf{F}^{-1}\mathbf{A_{21}}\mathbf{A_{11}}^{-1}) \\
&= \mathbf{A_{11}}^{-1}(\frac{q+2}{4}\mathbf{I_q} + \frac{1}{4}\mathbf{J_q}) \\
&= \mathbf{I_q}.
\end{aligned}$$

Thus, we have

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} \mathbf{I_q} & -\mathbf{16A_{12}} \\ -\mathbf{16A_{21}} & 24\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}} + \mathbf{4B_1} \end{bmatrix}. \tag{7}$$

**Selecting One Additional Design Point**

We want to take one additional design point $(x_1, x_2, \ldots, x_q)$ to test Lack of Fit. The added design point is displayed in the design matrix $\mathbf{X}^*$ as

$$\mathbf{z}' = \begin{bmatrix} x_1 & x_2 & \ldots, & x_q, & x_1x_2 & \ldots, & x_{q-1}x_q \end{bmatrix} = \begin{bmatrix} \mathbf{v}' & \mathbf{u}' \end{bmatrix},$$

where $\mathbf{v}'$=$(x_1, x_2, \ldots, x_q)$ and $\mathbf{u}'$=$(x_1x_2, x_1x_3, \ldots, x_{q-1}x_q)$.
To maximize $\mathbf{z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}$, such that $\mathbf{v}'\mathbf{1} = 1$, where $\mathbf{1}$ is a column vector of ones, consider

$$f(\mathbf{x}) = \begin{bmatrix} \mathbf{v}' & \mathbf{u}' \end{bmatrix} (\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{v} \\ \mathbf{u} \end{bmatrix} - \mathbf{2}\lambda(\mathbf{v}'\mathbf{1} - \mathbf{1}), \tag{8}$$

where $\lambda$ is a Lagrange multiplier.
Differentiating (8) w.r.t $\mathbf{v}$ and equating to zero, we get

$$\frac{\partial}{\partial \mathbf{v}}\{\begin{bmatrix} \mathbf{v}' & \mathbf{u}' \end{bmatrix}\} \begin{bmatrix} \mathbf{I_q} & -\mathbf{16A_{12}} \\ -\mathbf{16A_{21}} & 24\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}} + \mathbf{4B_1} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{u} \end{bmatrix} = \lambda\mathbf{1}, \tag{9}$$

Let

$$\frac{\partial}{\partial \mathbf{v}}\{\begin{bmatrix} \mathbf{v}' & \mathbf{u}' \end{bmatrix}\} = \begin{bmatrix} \mathbf{I_q} & \mathbf{K} \end{bmatrix}$$

where

$$\mathbf{K} = \begin{bmatrix} x_2 & x_3 & \ldots & x_q & 0 & 0 & \ldots & 0 \\ x_1 & 0 & \ldots & 0 & x_3 & x_4 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \vdots \\ 0 & 0 & \ldots & x_1 & 0 & 0 & \ldots & x_{q-1} \end{bmatrix}$$

Let $\mathbf{L}$ be a $(q-1)\times q$ matrix, such that

$$\mathbf{L}=\begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

Multiplying (9) by $\mathbf{L}$ on both sides, we get

$$\mathbf{L}\begin{bmatrix} \mathbf{I_q} & \mathbf{K} \end{bmatrix}\begin{bmatrix} \mathbf{I_q} & -\mathbf{16A_{12}} \\ -\mathbf{16A_{21}} & 24\mathbf{I}_{\frac{\mathbf{q(q-1)}}{\mathbf{2}}}+\mathbf{4B_1} \end{bmatrix}\begin{bmatrix} \mathbf{v} \\ \mathbf{u} \end{bmatrix}=\mathbf{0}, \tag{10}$$

or equivalently,

$$\mathbf{L}(\mathbf{v}-16\mathbf{KA_{21}v}-16\mathbf{A_{12}u}+24\mathbf{Ku}+4\mathbf{KB_1u})=\mathbf{0}.$$

By solving the above equations, we get $(2q+1)$ stationary points grouped as three solution groups below:

Solution I: $x=(\frac{1}{q},\dots,\frac{1}{q})$,

Solution II: q points of $x\leftrightarrow(1-(q-1)\delta,\delta,\dots,\delta)$, where $\delta=\frac{(5q+2+\sqrt{q^2-4q+76})}{8(q^2+q-3)}$,

Solution III: q points of $x\leftrightarrow(1-(q-1)\delta,\delta,\dots,\delta)$, where $\delta=\frac{(5q+2-\sqrt{q^2-4q+76})}{8(q^2+q-3)}$.

The three corresponding values of $\mathbf{z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}$ are listed below:

Ia. $\frac{q^2+4q-4}{q^3}$,

IIa. $\frac{1}{128(-3+q+q^2)^3}[q^6+115q^5+712q^4-772q^3-4648q^2+6088q-1664+\sqrt{76-4q+q^2}(q^5-3q^4+62q^3+124q^2-792q+608)]$,

IIIa. $\frac{1}{128(-3+q+q^2)^3}[q^6+115q^5+712q^4-772q^3-4648q^2+6088q-1664-\sqrt{76-4q+q^2}(q^5-3q^4+62q^3+124q^2-792q+608)]$.

Among them, $Ia$ is the maximum when $q=3$, $IIa$ is the maximum when $q\geq 4$, and $IIIa$ is always the minimum one.

In addition, the Jacobian matrix for the stationary points is

$$\frac{\partial^2 f(\mathbf{x})}{\partial\mathbf{v}\partial\mathbf{v}'}=2\begin{bmatrix} \mathbf{I_q} & \mathbf{K} \end{bmatrix}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\frac{\partial}{\partial\mathbf{v}'}\{\begin{bmatrix} \mathbf{v} \\ \mathbf{u} \end{bmatrix}\}+\mathbf{2}\frac{\partial}{\partial\mathbf{v}'}\{\begin{bmatrix} \mathbf{I_q} & \mathbf{K} \end{bmatrix}\}\begin{bmatrix} \mathbf{a_1'w} \\ \mathbf{a_2'w} \\ \dots \\ \mathbf{a'_{\frac{q(q+1)}{2}}w} \end{bmatrix}$$

$$=2\begin{bmatrix} \mathbf{I_q} & \mathbf{K} \end{bmatrix}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\begin{bmatrix} \mathbf{I_q} \\ \mathbf{K}' \end{bmatrix}+$$

$$2\begin{bmatrix} 0 & \mathbf{a}'_{\mathbf{q+1}}\mathbf{w} & \mathbf{a}'_{\mathbf{q+2}}\mathbf{w} & \dots & \mathbf{a}'_{\mathbf{2q-1}}\mathbf{w} \\ \mathbf{a}'_{\mathbf{q+1}}\mathbf{w} & 0 & \mathbf{a}'_{\mathbf{2q}}\mathbf{w} & \dots & \mathbf{a}'_{\mathbf{3q-3}}\mathbf{w} \\ \mathbf{a}'_{\mathbf{q+2}}\mathbf{w} & \mathbf{a}'_{\mathbf{2q}}\mathbf{w} & 0 & \dots & \mathbf{a}'_{\mathbf{4q-6}}\mathbf{w} \\ \vdots & \dots & \dots & \dots & \vdots \\ \mathbf{a}'_{\mathbf{2q-2}}\mathbf{w} & \mathbf{a}'_{\mathbf{3q-4}}\mathbf{w} & \mathbf{a}'_{\mathbf{4q-7}}\mathbf{w} & \dots & \mathbf{a}'_{\frac{\mathbf{q(q+1)}}{\mathbf{2}}}\mathbf{w} \\ \mathbf{a}'_{\mathbf{2q-1}}\mathbf{w} & \mathbf{a}'_{\mathbf{3q-3}}\mathbf{w} & \mathbf{a}'_{\mathbf{4q-6}}\mathbf{w} & \dots & 0 \end{bmatrix}. \tag{11}$$

The determinant of the Jacobian matrix is nonnegative for all three solutions and hence none of the stationary points maximizes the function $f(\mathbf{x})$.
Hence, we take the optimal design point for $q = 3$ as $x = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and for $q \geq 4$, as $x \leftrightarrow (1-(q-1)\delta, \delta, \dots, \delta)$, where $\delta = \frac{(5q+2+\sqrt{q^2-4q+76})}{8(q^2+q-3)}$, to maximize $\mathbf{z}'(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{z}$.

**Comparison of Designs for Testing Lack of Fit**

For each $q$-factor second-degree mixture model, we compare the following five different designs which are composed of $\frac{q(q+1)}{2}$ D-optimal minimal design points plus one extra design point:

Design I: One of the minimal design points, $x \leftrightarrow (1, 0, 0, \dots, 0)$ or $x \leftrightarrow (\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$.

Design II: One of the face centroids, $x \leftrightarrow (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \dots, 0)$.

Design III: Overall centroid $x = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$.

Design IV: One interior design point $x = (\frac{1}{2}, \frac{1}{2(n-1)}, \dots, \frac{1}{2(n-1)})$.

Design V: One design point, $x = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for $q = 3$, and one of the following points: $x \leftrightarrow (1-(q-1)\delta, \delta, \dots, \delta)$, where $\delta = \frac{(5q+2+\sqrt{q^2-4q+76})}{8(q^2+q-3)}$ for $q \geq 4$.

To compare the designs listed above, we calculate the D-efficiency by using the ratio of the determinant of any design to that of Design I, the D-optimal minimal design plus one of the replicate of the minimal design points. Table 1 presents the design points, the determinant of the information matrix $\mathbf{X}^{*'}\mathbf{X}^*$ taking $t = 1$ and D-efficiency for three, four and five-factor second-degree mixture models. As expected, the designs having the boundary points such as vertices, edges, faces, etc (Design I and II) as the extra design point, has larger determinant than those with only interior point (Design III-V). For practical reasons, the interior design point having the highest efficiency given at Design V is optimal and recommended.

# 3 Second-degree Mixture Model With Main Effects and Two Factor Interactions Including One Common Factor

We want to discuss one special mixture model, second-degree mixture model with main effects and two factor interactions including one common factor (WLOG, $x_1$). The model could be expressed as:

$$y = \sum_{i=1}^{q} \beta_i x_i + \sum_{j=2}^{q} \beta_{1j} x_1 x_j + \epsilon. \tag{12}$$

There are (2q-1) parameters in the model and at least (2q-1) minimal design points are needed to estimate all factors.

**D-optimal Minimal Design**

Table 1: Comparisons of designs for Three-, Four- and Five-factor Second-degree Mixture Model

| Factors | Designs | Additional Design Point | $\|X^{*'}X^*\|$ | D-efficiency |
|---|---|---|---|---|
| 3 | I | One of the minimal design points | $4.8828 * 10^{-4}$ | 100% |
| | II,III,V | $(\frac{1}{3}\mathbf{1}_3)$ | $3.9786 * 10^{-4}$ | 81.48% |
| | IV | One interior point $(\frac{1}{2}, \frac{1}{4}\mathbf{1}_2)$ | $3.8910 * 10^{-4}$ | 79.69% |
| 4 | I | One of the minimal design points | $1.1921 * 10^{-7}$ | 100% |
| | II | One face centroid | $0.9713 * 10^{-7}$ | 81.48% |
| | III | Overall centroid $(\frac{1}{4}\mathbf{1_4})$ | $0.8568 * 10^{-7}$ | 71.88% |
| | IV | One interior point $(\frac{1}{2}, \frac{1}{6}\mathbf{1_3})$ | $0.8389 * 10^{-7}$ | 70.37% |
| | V | One of the following design point $x \leftrightarrow (\frac{35-3\sqrt{19}}{68}, \frac{11+\sqrt{19}}{68}, \frac{11+\sqrt{19}}{68}, \frac{11+\sqrt{19}}{68})$ | $0.8575 * 10^{-7}$ | 71.93% |
| 5 | I | One of the minimal design points | $1.8190 * 10^{-12}$ | 100% |
| | II | One face centroid | $1.4821 * 10^{-12}$ | 81.48% |
| | III | Overall centroid $(\frac{1}{5}\mathbf{1_5})$ | $1.2078 * 10^{-12}$ | 66.40% |
| | IV | One interior point $(\frac{1}{2}, \frac{1}{8}\mathbf{1_4})$ | $1.1902 * 10^{-12}$ | 65.43% |
| | V | One of the following design point $x \leftrightarrow (\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}), (\frac{1}{6}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ | $1.2127 * 10^{-12}$ | 66.67% |

The minimal design matrix is:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \dots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 1/2 & 1/2 & \dots & 0 & 1/4 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \dots & \vdots \\ 1/2 & 0 & \dots & 1/2 & 0 & \dots & 1/4 \end{bmatrix}$$

**Inverse of the X′X Matrix**

The inverse matrix $(\mathbf{X'X})^{-1}$ is

$$(\mathbf{X'X})^{-\mathbf{1}} = \begin{bmatrix} \mathbf{I_q} & \mathbf{M} \\ \mathbf{M}' & 20\mathbf{I_{q-1}} + 4\mathbf{J_{q-1}} \end{bmatrix}, \tag{13}$$

where $\mathbf{M}$ is a $q$ by $(q-1)$ matrix with $\mathbf{M} = \begin{bmatrix} -2\mathbf{1}'_{q-1} \\ -2\mathbf{I_{q-1}} \end{bmatrix}$.

**Selecting One Additional Design Point**

We need at least one design point to determine Lack of Fit. The design point provides the added row to the design matrix $\mathbf{X}$

$$\mathbf{z}' = \begin{bmatrix} x_1 & x_2 & \dots & x_q & x_1x_2 & \dots & x_1x_q \end{bmatrix} = \begin{bmatrix} \mathbf{v}' & x_1\mathbf{u}'_\mathbf{1} \end{bmatrix},$$

where $\mathbf{v}'=(x_1, x_2, \dots, x_q)$, $\mathbf{u}'_\mathbf{1}=(x_2, x_3, \dots, x_q)$.

To maximize $\mathbf{z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}$ such that $\mathbf{v}'\mathbf{1} = 1$, we consider the following function

$$f(\mathbf{x}) = \begin{bmatrix} \mathbf{v}' & x_1\mathbf{u_1}' \end{bmatrix} (\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{v} \\ x_1\mathbf{u_1} \end{bmatrix} - \mathbf{2}\lambda(\mathbf{v}'\mathbf{1} - \mathbf{1}), \tag{14}$$

where $\lambda$ is a Lagrange multiplier.
Differentiating $f(\mathbf{x})$ w.r.t $\mathbf{v}$ and equating to zero, we get

$$\frac{\partial}{\partial \mathbf{v}}\{\begin{bmatrix} \mathbf{v}' & \mathbf{x_1 u_1}' \end{bmatrix}\}(\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{v} \\ \mathbf{x_1 u_1} \end{bmatrix} = \lambda\mathbf{1}, \tag{15}$$

Let

$$\frac{\partial}{\partial \mathbf{v}}\{\begin{bmatrix} \mathbf{v}' & \mathbf{x_1 u_1}' \end{bmatrix}\} = \begin{bmatrix} \mathbf{I_q} & \mathbf{K_1} \end{bmatrix}$$

where $\mathbf{K}_1$ is a $q \times (q-1)$ matrix with

$$\mathbf{K}_1 = \begin{bmatrix} x_2 & x_3 & \dots & x_q \\ x_1 & 0 & \dots & 0 \\ \vdots & \dots & \dots & \vdots \\ 0 & 0 & \dots & x_1 \end{bmatrix}.$$

Let $\mathbf{L}_1$ be a $q \times (q-1)$ matrix, s.t.

$$\mathbf{L}_1 = \begin{bmatrix} \text{-(q-1)} & 1 & 1 & \dots & 1 & 1 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \dots & \dots & \dots & \dots & \vdots \\ 0 & 1 & 0 & \dots & 0 & -1 \end{bmatrix},$$

Note that

$$\mathbf{L}_1 \begin{bmatrix} \mathbf{I_q} & \mathbf{K}_1 \end{bmatrix} = \begin{bmatrix} \text{-(q-1)} & 1 & \dots & x_1-(q-1)x_2 & \dots & x_1-(q-1)x_q \\ 0 & 1 & \dots & x_1 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \vdots \\ 0 & 1 & \dots & x_1 & \dots & -x_1 \end{bmatrix}, \tag{16}$$

and

$$(\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{v} \\ x_1\mathbf{u_1} \end{bmatrix} = \begin{bmatrix} \mathbf{v} - 2x_1 \begin{pmatrix} 1-x_1 \\ \mathbf{u}_1 \end{pmatrix} \\ 2x_1(1-2x_1)\mathbf{1}_{q-1} - 2(1-10x_1)\mathbf{u_1} \end{bmatrix}. \tag{17}$$

Multiplying (15) by $\mathbf{L}_1$ on both sides and taking into account (16) and (17), we get

$$\begin{bmatrix} \text{-(q-1)} & 1 & 1 & \dots & x_1-(q-1)x_2 & \dots & x_1-(q-1)x_q \\ 0 & 1 & -1 & \dots & x_1 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \dots & \vdots \\ 0 & 1 & 0 & \dots & x_1 & \dots & -x_1 \end{bmatrix} (\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{v} \\ x_1\mathbf{u_1} \end{bmatrix} = \mathbf{0},$$

It can be further simplified as

$$\left[\begin{array}{c} g \\ (1-4x_1+20x_1^2)\begin{pmatrix} x_2-x_3 \\ x_2-x_4 \\ \dots \\ x_2-x_q \end{pmatrix} \end{array}\right] = \mathbf{0}, \tag{18}$$

where g is a suitable scalar. Since $1-4x_1+20x_1^2>0$ for all $x_1$, equation (18) gives that $x_2 = x_3 = \ldots = x_q$. Using the fact that $\mathbf{v}'\mathbf{1}=1$, we have $x_2=x_3=\ldots=x_q=(1-x_1)/(q-1)$ and subsequently

$$g = (q-1)(1-2x_1)[4(q+4)x_1^2-(q+22)x_1+3] = 0 \tag{19}$$

Therefore the stationary points are:

1. $(\frac{1}{2}, \frac{1}{2(q-1)}\mathbf{1}'_{q-1})$
2. $(\frac{q+22+\sqrt{q^2-4q+292}}{8(q+4)}, \frac{7q+10-\sqrt{q^2-4q+292}}{8(q+4)(q-1)}\mathbf{1}'_{q-1})$
3. $(\frac{q+22-\sqrt{q^2-4q+292}}{8(q+4)}, \frac{7q+10+\sqrt{q^2-4q+292}}{8(q+4)(q-1)}\mathbf{1}'_{q-1})$

With corresponding $\mathbf{z}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}$ :

1a. $\frac{1}{q-1}$

2a. $\frac{1}{128(q-1)(q+4)^3}(q^4+120q^3+840q^2+8992q-5040+(q^3-6q^2+300q-584)\sqrt{q^2-4q+292})$

3a. $\frac{1}{128(q-1)(q+4)^3}(q^4+120q^3+840q^2+8992q-5040-(q^3-6q^2+300q-584)\sqrt{q^2-4q+292})$

When $q \leq 8$, 1a is maximum and when $q \geq 9$, 2a is the maximum. The difference between 1a and 2a is small when $q \geq 9$. In addition, 3a is the minimal value among the stationary points and is not recommended.

The Jacobian matrix for stationary points is:

$$\frac{\partial^2 f(\mathbf{x})}{\partial\mathbf{v}\partial\mathbf{v}'} = 2\left[\begin{array}{cc} \mathbf{I_q} & \mathbf{K}_1 \end{array}\right](\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\frac{\partial}{\partial\mathbf{v}'}\{\left[\begin{array}{c} \mathbf{v} \\ x_1\mathbf{u_1} \end{array}\right]\} + \mathbf{2}\frac{\partial}{\partial\mathbf{v}'}\{\left[\begin{array}{cc} \mathbf{I_q} & \mathbf{K}_1 \end{array}\right]\}\left[\begin{array}{c} \mathbf{a_1}'\mathbf{w} \\ \mathbf{a_2}'\mathbf{w} \\ \dots \\ \mathbf{a_{2q-1}}'\mathbf{w} \end{array}\right]$$

$$= 2\left[\begin{array}{cc} \mathbf{I_q} & \mathbf{K}_1 \end{array}\right](\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\left[\begin{array}{c} \mathbf{I_q} \\ \mathbf{K_1}' \end{array}\right] +$$

$$2\left[\begin{array}{ccccc} 0 & \mathbf{a_{q+1}}'\mathbf{w} & \mathbf{a_{q+2}}'\mathbf{w} & \dots & \mathbf{a_{2q-1}}'\mathbf{w} \\ \mathbf{a_{q+1}}'\mathbf{w} & 0 & 0 & \dots & 0 \\ \mathbf{a_{q+2}}'\mathbf{w} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{a_{2q-1}}'\mathbf{w} & 0 & 0 & \dots & 0 \end{array}\right], \tag{20}$$

where we rewrite

$$(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} = \begin{bmatrix} \mathbf{I_q} & \mathbf{M} \\ \mathbf{M}' & 20\mathbf{I_{q-1}} + 4\mathbf{J_{q-1}} \end{bmatrix} = \begin{bmatrix} \mathbf{a_1}' \\ \mathbf{a_2}' \\ \cdots \\ \mathbf{a_{2q-1}}' \end{bmatrix},$$

and

$$\mathbf{w} = \begin{bmatrix} \mathbf{v} \\ x_1\mathbf{u_1} \end{bmatrix}.$$

Since $x_2 = x_3 = \ldots = x_q$, we substitute $x_3, \ldots, x_q$ to $x_2$, the Jacobian matrix becomes

$$2\begin{bmatrix} 1 + 4(1 - x_1)(5x_2 - x_1) & 2(2x_1 - 2x_2 + 20x_1x_2 - 4x_1^2)\mathbf{1_{q-1}}' \\ 2(2x_1 - 2x_2 + 20x_1x_2 - 4x_1^2)\mathbf{1_{q-1}} & (20x_1^2 - 4x_1 + 1)\mathbf{I_{q-1}} + 4x_1^2\mathbf{J_{q-1}} \end{bmatrix}.$$

As earlier, the determinant of the Jacobian matrix is not negative definite for all stationary points, therefore none of the stationary points maximize function $f(\mathbf{x})$. Thus, the stationary point which attains the maximum determinant is chosen as optimal design point. Considering the difference of determinant between stationary point 1 and 2 is very small when $q \geq 9$. For practical reason, stationary point 1 is recommended as the optimal design point, i.e. $(\frac{1}{2}, \frac{1}{2(q-1)}\mathbf{1}_{q-1}')$.

**Comparison of Designs for Testing Lack of Fit**

We compare new design with other existing designs to test Lack of Fit. The following six designs, which are composed of $(2q - 1)$ D-optimal minimal design points and one additional design point are discussed and compared.

Design I: One of the minimal design points.

Design II: One of the midpoints with main factor 0, such as $(0, \frac{1}{2}, \frac{1}{2}, \ldots, 0)$.

Design III: Face centroid with main factor of $\frac{1}{3}$, such as $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \ldots, 0)$.

Design IV: One of the face centroids with main factor 0, such as $(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \ldots, 0)$ etc.

Design V: Overall centroid $(\frac{1}{q}, \frac{1}{q}, \ldots, \frac{1}{q})$.

Design VI: One design point $(\frac{1}{2}, \frac{1}{2(q-1)}, \ldots, \frac{1}{2(q-1)})$.

Table 2 presents the determinant of information matrix $\mathbf{X}'\mathbf{X}$ and D-efficiency (defined as the ratio of the determinant to that of Design I) for the above designs. Similarly, the designs having the boundary points lead to larger determinant and higher efficiency than those with only interior points (Design V and VI). For practical reasons, the interior design point with higher efficiency given at Design VI is optimal and recommended.

Table 2: Comparisons for designs for Three-, Four- , Five- and Six-factor Second-degree mixture model with main effects and two factor interactions including one common factor

| Factors | Designs | Additional Design Point | $\lvert X^{*'} X^{*} \rvert$ | D-efficiency |
|---|---|---|---|---|
| 3 | I | One of the minimal design points | $7.8125 * 10^{-3}$ | 100% |
| | II | One midpoint with main factor 0 | $5.8594 * 10^{-3}$ | 75.00% |
| | III,IV,V | $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ | $5.594 * 10^{-3}$ | 71.62% |
| | VI | One design point $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ | $5.8594 * 10^{-3}$ | 75.00% |
| 4 | I | One of the minimal design points | $4.8828 * 10^{-4}$ | 100% |
| | II | One midpoint with main factor 0 | $3.6621 * 10^{-4}$ | 75.00% |
| | III | One face centroid with main factor $\frac{1}{3}$ | $3.4963 * 10^{-4}$ | 71.60% |
| | IV | One face centroid with main factor 0 | $3.2552 * 10^{-4}$ | 66.67% |
| | V | Overall centroid $(\frac{1}{4}\mathbf{1}_4)$ | $3.0518 * 10^{-4}$ | 62.50% |
| | VI | One design point $(\frac{1}{2}, \frac{1}{6}\mathbf{1}_3)$ | $3.2552 * 10^{-4}$ | 66.67% |
| 5 | I | One of the minimal design points | $3.0518 * 10^{-5}$ | 100% |
| | II | One midpoint with main factor 0 | $2.2888 * 10^{-5}$ | 75.00% |
| | III | One face centroid with main factor $\frac{1}{3}$ | $2.1852 * 10^{-5}$ | 71.60% |
| | IV | One face centroid with main factor 0 | $2.0345 * 10^{-5}$ | 66.67% |
| | V | Overall centroid $(\frac{1}{5}\mathbf{1}_5)$ | $1.7920 * 10^{-5}$ | 58.72% |
| | VI | One design point $(\frac{1}{2}, \frac{1}{8}\mathbf{1}_4)$ | $1.9074 * 10^{-5}$ | 62.50% |
| 6 | I | One of the minimal design points | $1.9074 * 10^{-6}$ | 100% |
| | II | One midpoint with main factor 0 | $1.4305 * 10^{-6}$ | 75.00% |
| | III | One face centroid with main factor $\frac{1}{3}$ | $1.3658 * 10^{-6}$ | 71.60% |
| | IV | One face centroid with main factor 0 | $1.2716 * 10^{-6}$ | 66.67% |
| | V | Overall centroid $(\frac{1}{6}\mathbf{1}_6)$ | $1.0832 * 10^{-6}$ | 56.79% |
| | VI | One design point $(\frac{1}{2}, \frac{1}{10}\mathbf{1}_5)$ | $1.1444 * 10^{-6}$ | 60.00% |

# 4 Conclusion

In this paper we developed a method of including one additional design point for second-degree mixture model to test Lack of Fit. This design point is inside the design space with all factor levels greater than zero for practical reasons. In summary, when number of factors equal to 3 , the optimal design point is $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and when the number of factors $\geq 4$, the optimal design point is $x \leftrightarrow (1-(q-1)\delta, \delta, \ldots, \delta)$, where $\delta = \frac{(5q+2+\sqrt{q^2-4q+76})}{8(q^2+q-3)}$.

In addition, we also discussed one special mixture model, second-degree mixture model with main effects and two factor interactions including one common factor. The optimal design point in this case is $(\frac{1}{2}, \frac{1}{2(q-1)}\mathbf{1}'_{q-1})$.

One may consider optimal design with minimal design plus one additional point by using computer-aid program. However this will give repeated point for design (Design I), which are not useful to test Lack of Fit. Also one may think of adding more than one point to the minimal design, this will also give design of repeated points. Hence it is preferable that one adds the points sequentially by extending the methods of this paper.

# References

[1] Chan, L.Y.(2000). "Optimal designs for experiments with mixtures: a survey". *Communications in Statistics - Theory and Methods*, Vol.29, Issue 9-10, pp. 2281-2312.

[2] Cornell, J. A. (2002). *Experiments with mixtures: Designs, Models, and Analysis of mixture data*, Third Edition. Wiley, New York.

[3] Kiefer, J. (1961). "Optimal designs in Regression problems, II". *Annals of Mathematical Statistics*, Vol.32, Issue 1, pp. 298-325.

[4] Lim, Y.B. (1990). "D-optimal designs for cubic polynomial regression on the q-simplex". *Journal of Statistical Planning and Inference*, 25, pp. 141-152.

[5] Raghavarao, D. (1971). *Constructions and combinatorial problems in design of experiments*, Wiley, New York.

[6] Scheffé, H. (1958). "Experiments with Mixtures". *Journal of Royal Statistical Society*, Series B, 20, pp. 344-366.

[7] Scheffé, H. (1961). "Reply to Mr. Quenouille's Comments About my Paper on Mixtures". *Journal of Royal Statistical Society*, Series B, 23, pp. 171-172.

[8] Scheffé, H. (1963). "Simplex-centroid designs for experiments with Mixtures". *Journal of Royal Statistical Society*, Series B, 25, pp. 235-263.