Mixture models based on homogeneous polynomials

Norman R. Draper [a,*], Friedrich Pukelsheim [b,1]

[a] *Department of Statistics, University of Wisconsin-Madison, Madison WI 53706-1685, USA*
[b] *Department of Statistics, University of Washington, Seattle WA 98195-4322, USA*

Received 27 May 1997; received in revised form 5 January 1998; accepted 6 January 1998

Abstract

Models for mixtures of ingredients are typically fitted by Scheffé's canonical model forms. An alternative representation is discussed which offers attractive symmetries, compact notation and homogeneous model functions. It is based on the Kronecker algebra of vectors and matrices, used successfully in previous response surface work. These alternative polynomials are contrasted with those of Scheffé, and ideas of synergism and model reduction are connected together in both algebras. Scheffé's 'special cubic' is shown to be sensible in both algebras. © 1998 Elsevier Science B.V. All rights reserved.

AMS classification: 62K15; 62J05

Keywords: Designs for mixtures; First-order models; Kronecker product; Mixture amount models; Scheffé canonical polynomial; Second-order models; Synergism; Third-order models

1. Introduction

Many practical problems are associated with the investigation of mixture ingredients $x_1, x_2, \ldots, x_q$ of q factors, with $x_i \geqslant 0$ and further restricted by

$$\sum x_i = 1 \tag{1.1}$$

or by some linear restriction which reduces to Eq. (1.1).

The definitive text Cornell (1990) lists numerous examples and provides a thorough discussion of both theory and practice. Early seminal work was done by Scheffé (1958, 1963) in which he suggested (1958) (p. 347) and analyzed the following canonical model forms of orders (degrees) one, two and three for the expected response η:

$$\eta = \sum_{1 \leqslant i \leqslant q} \beta_i x_i, \tag{1.2}$$

* Corresponding author. E-mail: draper@stat.wisc.edu.
[1] On leave from the Institut für Mathematik, Universität Augsburg, D-86135 Augsburg, Germany.

$$\eta = \sum_{1 \le i \le q} \beta_i x_i + \sum_{1 \le i < j \le q} \beta_{ij} x_i x_j, \tag{1.3}$$

$$\eta = \sum_{1 \le i \le q} \beta_i x_i + \sum_{1 \le i < j \le q} \beta_{ij} x_i x_j + \sum_{1 \le i < j \le q} \gamma_{ij} x_i x_j (x_i - x_j)$$
$$+ \sum_{1 \le i < j < k \le q} \beta_{ijk} x_i x_j x_k. \tag{1.4}$$

As stated by Cornell (1990) (p. 26) there is 'an infinite number of regression functions' derivable by resubstituting Eq. (1.1) in various ways. Scheffé (1958) (p. 346) remarks that Eqs. (1.2)–(1.4) constitute 'an appropriate form of polynomial regression'. We shall refer to Eqs. (1.1)–(1.4) as the *S-models*, or *S-polynomials*.

In the present paper, we propose an alternative representation of mixture models which appears to have certain advantages to be described. Our versions, to appear in Eqs. (2.3)–(2.5), are based on the Kronecker algebra of vectors and matrices, and give rise to homogeneous model functions. We shall refer to the corresponding expressions as the *K-models*, or *K-polynomials*.

A similar approach to non-mixture response surface models was used successfully in Draper et al. (1991), Draper and Pukelsheim (1994) and Draper et al. (1996); see also Ch. 15 in Pukelsheim (1993).

An outline of the present paper is as follows. In Section 2 we introduce the K-models; their expected response η is homogeneous in the ingredients x_i. By way of example, Section 3 illustrates the inhomogeneity of the S-models. Section 4 initiates the discussion of reducing the order of K-models through testable hypotheses, which is then carried through for reducing second order to first (Section 5), and third order to second (Section 6). In Section 7 we compare the second-order coefficients in a K-model with those in an S-model and in Section 8 we do the same for third order. The transition from S-models to K-models has consequences for the design choice for mixture experiments, and for the analysis of data. These aspects will be addressed in subsequent work.

2. K-polynomials for mixtures models

The mixture ingredients, x_i, can conveniently be written as a $q \times 1$ vector $x = (x_1, x_2, \ldots, x_q)'$. The Kronecker square $x \otimes x$ consists of a $q^2 \times 1$ vector of the q^2 cross products $x_i x_j$, in lexicographic order with subscripts $11, 12, \ldots, 1q; 21, 22, \ldots, 2q; \ldots; q1, q2, \ldots, qq$,

$$x \otimes x = (x_1^2, x_1 x_2, \ldots, x_1 x_q; x_2 x_1, x_2^2, \ldots, x_2 x_q; \ldots; x_q x_1, x_q x_2, \ldots, x_q^2)'. \tag{2.1}$$

In Eq. (2.1) individual mixed second-order terms appear twice, for example, we have $x_1 x_2$ and $x_2 x_1$. Although this may at first appear disadvantageous, the symmetry attained

more than compensates for the duplications, as will become apparent. The very same point is familiar from treating dispersion matrices as matrices, and *not* as arrays of a minimal number of functionally independent terms.

Similarly, the Kronecker cube $x \otimes x \otimes x$ is a $q^3 \times 1$ vector of all terms of the form $x_i x_j x_k$, in lexicographic order, and repeats third-order terms either six or three times depending on the number of different subscripts, ijk or iij. It has the form

$$x \otimes x \otimes x = (x_1 x_1 x_1, x_1 x_1 x_2, x_1 x_1 x_3, \ldots, x_1 x_1 x_q; x_1 x_2 x_1, x_1 x_2 x_2, x_1 x_2 x_3, \ldots, x_1 x_2 x_q;$$
$$\ldots; x_q x_q x_1, x_q x_q x_2, x_q x_q x_3, \ldots, x_q x_q x_q)' \tag{2.2}$$

for $q \ge 3$ factors. For $q = 2$, no products with three distinct subscripts occur, of course. The K-models that we propose to replace Eqs. (1.2)–(1.4) are the following:

$$\eta = x'\theta = \sum_{1 \le i \le q} \theta_i x_i, \tag{2.3}$$

$$\eta = (x \otimes x)'\theta = \sum_{1 \le i,j \le q} \theta_{ij} x_i x_j, \tag{2.4}$$

$$\eta = (x \otimes x \otimes x)'\theta = \sum_{1 \le i,j,k \le q} \theta_{ijk} x_i x_j x_k. \tag{2.5}$$

Since the regressors $x_i x_j$ and $x_j x_i$ are identical, we assume $\theta_{ij} = \theta_{ji}$. For the same reason, θ_{ijk} is assumed to be the same for all permutations of the subscripts i, j, k.

The first order K-model (2.3) and the first-order S-model (1.2) are of the same homogeneous form in the x_i's, of course. The second-order K-model from Eq. (2.4) is

$$\eta = \sum_{1 \le i \le q} \theta_{ii} x_i^2 + 2 \sum_{1 \le i < j \le q} \theta_{ij} x_i x_j, \tag{2.6}$$

and is fully homogeneous in second-order terms; the x_i terms of the S-model (1.3) are replaced by x_i^2 terms and, assuming that $\theta_{ij} = \theta_{ji}$, the multiplicity of mixed terms $x_i x_j$ for $i \ne j$ has been doubled. The third-order K-model is homogeneous of order three, and will be discussed in Section 6. Extension to higher-order models is evident.

The homogeneous representation of K-models should *not* be mistaken to mean that we 'lose' linear terms in Eq. (2.4), nor linear and quadratic terms in Eq. (2.5). The second-order S-model (1.3) and K-model (2.4) both feature $\binom{q+1}{2}$ parameters for the response function; for third order Eqs. (1.4) and (2.5) both involve $\binom{q+2}{3}$ parameters.

We may sketch the essential argument by rewriting Eq. (1.1) in succinct notation as $l'_q x = 1$, where $l_q = (1, 1, \ldots, 1)'$ is the unity vector in $\mathbb{R}^q$. Then the first-order part of the response surface (1.3) can be blended into the second-order part to produce a homogeneous second-order function of form (2.4) by noting that

$$x'\beta \cdot 1 = x'\beta \cdot x' l_q = (x'\beta) \otimes (x' l_q) = (x \otimes x)'(\beta \otimes l_q), \tag{2.7}$$

where the last equation uses a key property of Kronecker products, see Eq. (5.4) in Draper et al. (1991) (p. 140) or Eq. (1.4) in Pukelsheim (1993) (p. 392). In similar fashion, Eq. (1.4) can be converted into the homogeneous third-order form (2.5) by blending both the first- and second-order parts of Eq. (1.4) into the third-order part. Sections 7 and 8 elaborate the equivalences of the K-models with the corresponding S-models.

An immediate advantage of model homogeneity is apparent. In problems where the component sum in Eq. (1.1) is $A \neq 1$, the homogeneity of the K-models ensures that all model terms are affected by the same multiple A^d where d is the degree or order of the model. This is not true of the S-models. The possible effects of model dependence on the total amount A is illustrated by the following example, after which we continue our discussion of the K-models.

3. An example of inhomogeneity for the S-models

We consider the simplest case of two components, $x_1 + x_2 = A$, where A is the total amount. Suppose we consider the three point design $(A,0),(\tfrac{1}{2}A,\tfrac{1}{2}A),(0,A)$, with respective weightings $\alpha/2, 1-\alpha, \alpha/2$, with $\alpha \in [0,1]$.

The second-order S-model involves the terms $x_1, x_2, x_1 x_2$, and gives rise to the moment matrix

$$M = \frac{1-\alpha}{16}\begin{pmatrix} 4A^2\frac{1+\alpha}{1-\alpha} & 4A^2 & 2A^3 \\ 4A^2 & 4A^2\frac{1+\alpha}{1-\alpha} & 2A^3 \\ 2A^3 & 2A^3 & A^4 \end{pmatrix}. \quad (3.1)$$

The trace of the inverse is found to be

$$\operatorname{trace} M^{-1} = \frac{4}{A^2\alpha}\left(1 + \frac{4}{A^2(1-\alpha)}\right). \quad (3.2)$$

If we optimize the design weights with respect to the average-variance criterion which requires maximization of Eq. (2.8), the solution for α depends on the amount A,

$$\alpha(A) = 1 + \frac{4}{A^2} - \frac{2}{A}\sqrt{1 + \frac{4}{A^2}}.$$

Including the limiting values 0 and ∞ for A, the weight α ranges from $\tfrac{1}{2}$ to 1:

A	0	1/8	1/4	1/2	1	2	4	8	∞
$\alpha(A)$	0.5	0.5005	0.502	0.508	0.528	0.586	0.691	0.805	1

(3.3)

For $A = 0$, the distribution of the weights is $\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}$. For $A = 1$ we reproduce the entry in line 11 of Table 2 of Galil and Kiefer (1977) (p. 451). For $A = \infty$ the inhomogeneity in the S-model has the effect that the linear portion dominates, and the central weight is zero.

4. Conditions for reducing the order of K-models

A standard procedure of polynomial model building is not only to check whether the current model is suitable for representing the data, but also to determine whether a more parsimonious lower order model might be adequate.

A great advantage of the S-model hierarchy is that higher-order models visibly include the terms of lower-order models. Thus reduction of the order of an S-model is attained simply by setting certain coefficients to zero, and so appropriate hypotheses are easy to formulate.

This is not so obvious for K-models. Thus, we now investigate what conditions are necessary for reduction of a K-model to one of a lower order. The resulting hypotheses will be seen to permit a pleasing interpretation.

5. Reduction of second order to first order

We will work with the excess function $\operatorname{Exc}_{21}(x)$ obtained by subtracting the first-order model function (2.3) from the second-order model function (2.4). We multiply Eq. (2.3) by $x_1 + x_2 + \cdots + x_q$, which is equal to one by Eq. (1.1), to achieve second-order terms throughout.

$$\operatorname{Exc}_{21}(x) = \sum_{1\le i,j\le q}\theta_{ij}x_ix_j - \sum_{1\le i\le q}\theta_i x_i\left(\sum_{1\le j\le q}x_j\right) \quad (5.1)$$

$$= \sum_{1\le i\le q}(\theta_{ii}-\theta_i)x_i^2 + \sum_{1\le i<j\le q}(2\theta_{ij}-\theta_i-\theta_j)x_ix_j,$$

using the fact that $\theta_{ij} = \theta_{ji}$. It follows that the excess will vanish identically on the region (1.1) if and only if

$$\theta_{ii} = \theta_i \quad \forall i.$$

$$\theta_{ij} = \tfrac{1}{2}(\theta_i + \theta_j) \quad \forall i,j. \quad (5.2)$$

Hence, the appropriate hypothesis that a second order K-model reduces to a first-order K-model is

$$\theta_{ij} = \tfrac{1}{2}(\theta_{ii} + \theta_{jj}) \quad \forall i\neq j. \quad (5.3)$$

If the hypothesis (5.3) is true, then the first-order parameters are obtained from the second-order parameters via $\theta_i = \theta_{ii}$.

In the spirit of Scheffé's (1958) (pp. 347–348) synergism discussion, we call $2\theta_{ij} - \theta_i - \theta_j$ the *coefficient of binary synergism* of x_ix_j *for the second-order K-model relative to the first-order K-model*. With this terminology we see that the fulfillment of Eq. (5.3) is equivalent to the vanishing of all coefficients of binary synergism.

6. Reduction of third order to second order

In similar fashion, the excess of a third-order K-model over a second-order K-model is

$$\text{Exc}_{32}(x) = \sum_{1 \le i,j,k \le q} \theta_{ijk} x_i x_j x_k - \sum_{1 \le i,j \le q} \theta_{ij} x_i x_j \left(\sum_{1 \le k \le q} x_k \right)$$

$$= \sum_{1 \le i,j,k \le q} (\theta_{ijk} - \theta_{ij}) x_i x_j x_k$$

$$= \sum_{1 \le i \le q} (\theta_{iii} - \theta_{ii}) x_i^3 + \sum_{1 \le i \ne j \le q} (3\theta_{iij} - \theta_{ii}) x_i^2 x_j$$

$$+ \sum_{1 \le i < j < k \le q} \{6\theta_{ijk} - 2(\theta_{ij} + \theta_{ik} + \theta_{jk})\} x_i x_j x_k, \tag{6.1}$$

using the fact that $\theta_{iij} = \theta_{iji} = \theta_{jii}$, and that $\theta_{ijk} = \theta_{ikj} = \theta_{jik} = \theta_{jki} = \theta_{kij} = \theta_{kji}$. It follows that the excess will vanish identically on the region (1.1) if and only if

$$\theta_{ijk} = \frac{1}{3}(\theta_{ij} + \theta_{ik} + \theta_{jk}) \quad \forall i, j, k. \tag{6.2}$$

This condition specializes to $\theta_{iii} = \theta_{ii}$ when $i = j = k$, and so to

$$2\theta_{iij} = 3\theta_{iii}, \tag{6.3}$$

when $i = k \ne j$. By solving Eq. (6.3) for θ_{ij} and substituting into Eq. (6.2), we obtain the appropriate hypothesis that a third-order K-model reduces to a second-order K-model as

$$\theta_{ijk} = \frac{1}{12}\{(3\theta_{iij} - \theta_{iii}) + (3\theta_{ijj} - \theta_{jjj}) + (3\theta_{jjk} - \theta_{jjj})$$
$$+ (3\theta_{iik} - \theta_{kkk}) + (3\theta_{jkk} - \theta_{kkk})\} \quad \forall i,j,k. \tag{6.4}$$

When all three subscripts are equal, Eq. (6.4) is an identity. There are $\binom{q}{2}$ conditions when two subscripts are equal, in which case Eq. (6.4) simplifies to

$$3\theta_{iij} = 3\theta_{iii} = \theta_{jjj} \quad \forall i \ne j. \tag{6.5}$$

There are $\binom{q}{3}$ conditions in Eq. (6.4) with three distinct subscripts. When Eq. (6.5) holds they simplify to

$$\theta_{ijk} = \frac{1}{6}\{(3\theta_{iij} - \theta_{iii}) + (3\theta_{ijk} - \theta_{jjj}) + (3\theta_{ikk} - \theta_{kkk})\} \quad \forall i \ne j \ne k \ne i. \tag{6.6}$$

If the hypothesis (6.4) is true, then the second-order parameters are obtained from the third-order parameters via Eqs. (6.3) and (6.5) as

$$\theta_{ij} = \frac{1}{2}(3\theta_{iij} - \theta_{iii}) = \frac{1}{2}(3\theta_{iij} - \theta_{jjj}) = \theta_{ji}. \tag{6.7}$$

Again in the spirit of Scheffé's (1958) (pp. 347–348) synergism discussion, we call $\theta_{ijk} - \frac{1}{3}(\theta_{ij} + \theta_{ik} + \theta_{jk})$ the *coefficient of ternary synergism* of x_i, x_j, x_k for the third-order K-model relative to the second-order K-model. With this terminology we see that the fulfillment of Eq. (6.2) is equivalent to the vanishing of all coefficients of ternary synergism.

7. Connections between second-order coefficients in S-models and K-models

In order to determine the relationships between the coefficients of the second-order models (1.3) and (2.4), we must convert the first term in Eq. (1.3) to be homogeneous of second order, by multiplying by $x_1 + x_2 + \cdots + x_q$:

$$\sum_{1 \le i \le q} \beta_i x_i = \sum_{1 \le i \le q} \beta_i x_i^2 + \sum_{1 \le i < j \le q} (\beta_i + \beta_j) x_i x_j. \tag{7.1}$$

Thus, the difference between Eqs. (2.4) and (1.3) is seen to be

$$\sum_{1 \le i,j \le q} \theta_{ij} x_i x_j - \sum_{1 \le i \le q} \beta_i x_i$$

$$= \sum_{1 \le i \le q} (\theta_{ii} - \beta_i) x_i^2 + \sum_{1 \le i < j \le q} (2\theta_{ij} - \beta_i - \beta_j) x_i x_j. \tag{7.2}$$

This difference vanishes for all x from Eq. (1.1) if and only if

$$\beta_i = \theta_{ii} \quad \text{and} \quad \beta_{ij} = 2\theta_{ij} - \theta_{ii} - \theta_{jj}. \tag{7.3}$$

This connects to Eq. (5.3), in that a reduction to a first-order model takes place if and only if all the β_{ij} vanish.

8. Connections between third-order coefficients in S-models and K-models

In order to determine the relationships between the coefficients of the third-order models (1.4) and (2.5), we first convert the first two terms in Eq. (1.4) to be homogeneous of third order, by multiplying by $x_1 + x_2 + \cdots + x_q$ as needed to raise to third order:

$$\sum_{1 \le i \le q} \beta_i x_i = \sum_{1 \le i \le q} \beta_i x_i^3 + \sum_{1 \le i < j \le q} (2\beta_i + \beta_j) x_i^2 x_j + \sum_{1 \le i < j \le q} (\beta_i + 2\beta_j) x_i x_j^2$$

$$+ \sum_{1 \le i < j < k \le q} 2(\beta_i + \beta_j + \beta_k) x_i x_j x_k, \tag{8.1}$$

$$\sum_{1\le i<j\le q}\beta_{ij}x_ix_j=\sum_{1\le i<j\le q}\beta_{ij}x_i^2x_j+\sum_{1\le i<j\le q}\beta_{ij}x_ix_j^2$$
$$+\sum_{1\le i<j<k\le q}(\beta_{ij}+\beta_{ik}+\beta_{jk})x_ix_jx_k.$$

Thus, the difference between Eqs. (2.5) and (1.4) is seen to be

$$\sum_{1\le i,j,k\le q}\theta_{ijk}x_ix_jx_k-\sum_{1\le i\le q}\beta_ix_i-\sum_{1\le i<j\le q}\beta_{ij}x_ix_j$$
$$-\sum_{1\le i<j\le q}\gamma_{ij}x_ix_j(x_i-x_j)-\sum_{1\le i<j<k\le q}\beta_{ijk}x_ix_jx_k$$
$$=\sum_{1\le i\le q}(\theta_{iii}-\beta_i)x_i^3+\sum_{1\le i<j\le q}(3\theta_{iij}-2\beta_i-\beta_j-\beta_{ij}-\gamma_{ij})x_i^2x_j$$
$$+\sum_{1\le i<j\le q}(3\theta_{ijj}-2\beta_j-\beta_i-\beta_{ij}+\gamma_{ij})x_ix_j^2$$
$$+\sum_{1\le i<j<k\le q}(6\theta_{ijk}-2\beta_i-2\beta_j-2\beta_k-\beta_{ij}-\beta_{ik}-\beta_{jk}-\beta_{ijk})x_ix_jx_k.$$

This difference vanishes for all x if and only if

$$\begin{aligned}\beta_i&=\theta_{iii},\\\beta_{ij}&=\tfrac{3}{2}(\theta_{iij}-\theta_{iii}+\theta_{ijj}-\theta_{jjj}),\\\gamma_{ij}&=\tfrac{1}{2}((3\theta_{iij}-\theta_{iii})-(3\theta_{ijj}-\theta_{jjj})),\\\beta_{ijk}&=6\theta_{ijk}+\tfrac{3}{2}(\theta_{iij}+\theta_{ijj}+\theta_{iik}+\theta_{ikk}+\theta_{jjk}+\theta_{jkk})-(\theta_{iii}+\theta_{jjj}+\theta_{kkk}).\end{aligned}$$

$$(8.2)$$

Scheffé (1958) (p. 352) refers to the reduced model when all γ_{ij} are zero as the special cubic model. In the K-model this requires Eq. (6.5) to be true. We see that the cubic is 'special' in the sense that it satisfies not all the conditions (6.4) but a particular subset of them, namely Eq. (8.3) to

$$(8.3)$$

the last equation of Eq. (8.3) to

$$\beta_{ijk}=6\theta_{ijk}+(3\theta_{iij}-\theta_{iii})+(3\theta_{jjk}-\theta_{jjj})+(3\theta_{ikk}-\theta_{kkk}).$$

$$(8.4)$$

Note that Eq. (6.3) implies Eq. (6.5) and hence that all γ_{ij} in Eq. (8.3) are zero; the reverse implication is not true, however.

Acknowledgements

References

Cornell, J.A., 1990. Experiments with Mixtures, 2nd ed. Wiley, New York.

Draper, N.R., Pukelsheim, F., 1994. On third order rotatability. Metrika 41, 137–161.

Draper, N.R., Heiligers, B., Pukelsheim, F., 1996. On optimal third order rotatable designs. Ann. Institute Statist. Math. 48, 395–402.

Draper, N.R., Gaffke, N., Pukelsheim, F., 1991. First and second order rotatability of experimental designs, moment matrices, and information surfaces. Metrika 38, 129–161.

Galil Z., Kiefer, J., 1977. Comparison of simplex designs for quadratic mixture models. Technometrics 19, 445–453. Also in: J.C. Kiefer Collected Papers III, Springer, New York, pp. 417–425.

Pukelsheim, F., 1993. Optimal Design of Experiments. Wiley, New York.

Scheffé, H., 1958. Experiments with mixtures. J. Roy. Statist. Soc. B 20, 344–360.

Scheffé, H., 1963. The simplex centroid design for experiments with mixtures. J. Roy. Statist. Soc. B 25, 235–263.