Selection of Variables for Response Surface Experiments with Mixtures

Sung H. Park*

ABSTRACT

A strategy for selecting subsets of variables from a given linear model in a mixture system is discussed. The purpose is to achieve better fitting surfaces for estimation of the response in an experimental region of interest. A criterion is proposed for screening variables and illustrated with an example.

1. Introduction

This paper considers the problem of screening mixture components arising in empirical investigations of the response relationship between a response and a number of mixture variables. The functional relationship is assumed to be approximated by the linear mixture model in the Scheffe form [9], i.e.,

$$\eta(\boldsymbol{x})=\boldsymbol{x}'\boldsymbol{\beta} \tag{1}$$

where $\boldsymbol{x}$ is a q-vector of input mixture variables, $\boldsymbol{x}'=(x_1,x_2.\dots x_q)$, and $\boldsymbol{\beta}'=(\beta_1, \beta_2,\dots,\beta_q)$ is a q-vector of unknown regression parameters. Note that the possible mixtures are restricted to the regular $(q-1)$-dimensional simplex,

$$x_i\geq 0, \quad (i=1,2,\dots,q); \ \sum_{i=1}^{q} x_i=1. \tag{2}$$

It is assumed that at the start of experimental investigation, data on a large number of potentially important components are available and the ob-

*Department of Computer Science and Statistics, Seoul National University.

jective is to reduce the number of important components for which the response is fitted.

Suppose there are $n \geq q$ observations and each observed response, y, is determined by

$$y(x) = x'\beta + \varepsilon. \tag{3}$$

The error, ε, is assumed to be identically and independently distributed with mean zero and unknown variance, σ^2. For all observations the model (3) may be conveniently expressed as

$$Y = X\beta + e$$

where Y is the n-vector of observed responses, X is the $n \times q$ design matrix, assumed to have full rank, and e is the n-vector of errors.

The problem of selecting a subset of variables in a linear multiple regression model has been of interest to many applied statisticians. Draper and Smith [2] discuss several criteria and procedures. Allen [1], Mallows [5], Helms [3], Park [6] and others propose criteria, while Hocking [4] reviews several criteria for selection of variables. Part of the philosophy of selecting ordinary independent variables is applicable to the problem in mixture components. However, because of the inherent constraints described in (2), the screening philosophy of mixture variables would warrant special attention.

In a recent paper, Snee and Marquardt [11] suggest a strategy of determining 'essential' components in mixtures. The following section contains a brief review and discussion of their strategy, which will in turn provide a motivation for development of this paper.

2. Review of Screening Strategy

The idea by Snee and Marquardt [11] basically consists of hypothesis testing of types of linear contrasts among the regression coefficients, β. One type is that, if a coefficient β_i is equal to the average of the remaining coefficients, then the component x_i has no effect on the response and it may be dropped from the model. The null hypothesis may be written as

Ho: $\beta_i=(q-1)^{-1}\sum_{j\neq i}\beta_j$.

Another type is that, if two or more coefficients are equal, then the associated components have equal effects and their sum can be considered as one component, thus reducing the number of essential components. The null hypothesis is, for example,

Ho: $\beta_i=\beta_j=\beta_k$

Snee and Marquardt further note that the null hypotheses may be written as $C\beta=O$ where C is a $m\times q$ contrast matrix and each row of C is a linear contrast of β. The significance can be tested by F-statistic,

$$F=\frac{\hat{\beta}'C'[C(X'X)^{-1}C']^{-1}C\hat{\beta}}{m\hat{\sigma}^2}$$

where $\hat{\beta}=(X'X)^{-1}X'Y$ and $\hat{\sigma}^2=Y'[I-X(X'X)^{-1}X']Y/(n-q)$.

We observe that some combination of the above two types of null hypotheses would be also meaningful for screening purposes. For instance,

Ho: $\beta_i=\beta_j$

$\beta_i+\beta_j=2(q-2)^{-1}\sum_{k\neq i,j}\beta_k$.

If this set of hypotheses is accepted, it means that x_i and x_j may compose a new component due to equal effects, which may be dropped from the model because of no effect. In fact, there are a large variety of possible contrasts to be tested.

The screening method described above reveals the following observations.

(i) The normality of $y(\boldsymbol{x})$ in (3) should be assumed for hypothesis testing.

(ii) It is possible that there exist different sets of contrasts that are accepted by F-test, but it may be hard to choose which set should be adopted to reduce the mixture components.

(iii) If the performance of the fitted equation with selected variables over the entire region of simplex is of main interest, it is not clear which set of contrasts should be chosen. For instance, suppose the null hypotheses $C_1\beta=O$ and $C_2\beta=O$, where C_1 is a $m_1\times q$ and C_2 is a $m_2\times q$ contrast matrix, are both accepted. It is not unlikely that there may be no clear-cut choice between the two hypotheses.

The objective of this paper is to propose a criterion to decide which set of contrasts should be adopted to provide a better performance of the fitted response surface over some region of experimental interest in the sense of mean square error(MSE). For the proposed criterion, the normality of $y(x)$ is not necessarily assumed.

3. Formulation of a criterion

Suppose a set of linear contrasts (or generally linear constraints) are imposed on the parameter space, i.e.,

$$C\beta=O \tag{4}$$

in which C is a $m\times q$ matrix of rank m. Let $\tilde{\beta}$ be the least squares estimator of β under the restriction (4). It is well known(see, for instance, Searle [10]) that $\tilde{\beta}$ has the form of

$$\tilde{\beta}=B\hat{\beta} \tag{5}$$

where $B=I-(X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C$,and $\hat{\beta}=(X'X)^{-1}\ X'Y$

Since the primary interest is in the precision of response estimation, consider the estimated response, $\tilde{y}(x)=x'\tilde{\beta}$, using the restricted estimator $\tilde{\beta}$ in (5). At an arbitrary point X on the region of interest R, the mean square error of $\tilde{y}(x)$ is

$$\begin{aligned}\text{MSE}[\tilde{y}(x)]&=E(x'\tilde{\beta}-x'\tilde{\beta})^2\\&=\sigma^2x'B(X'X)^{-1}B'x+x'LC\beta\beta'C'L'x\end{aligned}$$

where $L=(X'X)^{-1}C'[C(X'X)^{-1}]C'$

To Compare the precision of $\tilde{y}(x)$ with that of $\tilde{y}(x)=x'\hat{\beta}$, we examine the difference,

$$D(x)=\text{MSE}[\hat{y}(x)]-\text{MSE}[\tilde{y}(x)]$$

$$= \{\text{Var}[\hat{y}(\boldsymbol{x})] - \text{Var}[\tilde{y}(\boldsymbol{x})]\} - \{\text{squared bias of } \tilde{y}(\boldsymbol{x})\}$$
$$= \sigma^2 \boldsymbol{x}'[\boldsymbol{L}\boldsymbol{C}(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{C}'\boldsymbol{L}']\boldsymbol{x}$$
$$- \boldsymbol{x}'\boldsymbol{L}\boldsymbol{C}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{C}'\boldsymbol{L}'\boldsymbol{x}. \quad (7)$$

The first term in (7) is non-negative, which implies that the response can be estimated with smaller variance using $\tilde{y}(\boldsymbol{x})$. However, the penalty is in bias, the second term. So if one can be assured that bias does not become too much of problem, then justification is provided for adopting the linear contrasts in (4). This also implies that, if one is willing to accept some bias in trade for a reduction in variance, then even if the linear contrasts are not true, one might still prefer using them.

It is of interest to find when $D(\boldsymbol{x}) \geq 0$ at any point $\boldsymbol{x}$. It can be readily shown that

$$D(\boldsymbol{x}) = \boldsymbol{x}'L[\sigma^2 C(X'X)^{-1}C' - C\beta\beta'C']L'\boldsymbol{x}. \quad (8)$$

Thus, if the matrix $\sigma^2 C(X'X)^{-1}C' - C\beta\beta'C'$ is positive semi-definite, it is possible to estimate the true response $\eta(\boldsymbol{x})$ in (1) with smaller MSE using $\tilde{y}(\boldsymbol{x})$ at any arbitrary point $\boldsymbol{x}$ in the region R. It can be shown that

$$\beta'C'[C(C'C)^{-1}C']^{-1}C\beta \leq \sigma^2 \quad (9)$$

is a necessary and sufficient condition that the matrix $\sigma^2 C(X'X)^{-1}C' - C\beta\beta'C'$ is positive semi-definite. Therefore, if (9) is satisfied, $\tilde{y}(\boldsymbol{x})$ is a better estimator of $\eta(\boldsymbol{x})$ than $\hat{y}(\boldsymbol{x})$ since $D(\boldsymbol{x}) \geq 0$ at point $\boldsymbol{x}$.

In order to detect how well the estimator $\tilde{y}(\boldsymbol{x})$ performs over the whole region R, cosider the difference between the integrated MSE's of $\hat{y}(\boldsymbol{x})$ and $\tilde{y}(\boldsymbol{x})$ over the region R under a weighting function $W(\boldsymbol{x})$,

$$J = \frac{K}{\sigma^2} \int_R D(\boldsymbol{x}) dW(x)$$

where $K^{-1} = \int_R dW(\boldsymbol{x})$. The $W(\boldsymbol{x})$ may be treated as a probability distribution function on R, and it allows for differential importance of the difference $D(\boldsymbol{x})$ at different points in the region.

Let M be the region moments defined by

$$M = K\int_R \boldsymbol{x}\boldsymbol{x}' \, dW(\boldsymbol{x}). \quad (19)$$

It can be shown from (7) that

$$J=\frac{K}{\sigma^2}\int_R\{\text{MSE}[\hat{y}(x)]-\text{MSE}[\tilde{y}(x)]\}dW(x)$$

$$=IV-IB \qquad (11)$$

where IV is the difference between the integrated variances of $\hat{y}(x)$ an $\tilde{y}(x)$,

$$IV=K\int_R x'LC(X'X)^{-1}C'L'xdW(x)$$
$$=\text{Tr}\{LC(X'X)^{-1}C'L'M]$$

and IB is the integrated bias of $\tilde{y}(x)$,

$$IB=\frac{1}{\sigma_2}\int_R x'LC\beta\beta'C'L'xdW(x)$$

$$=\frac{1}{\sigma_2}\beta'C'L'MLC\beta,$$

and Tr denotes trace. Note that IV is the gain in precision from integrated variance and IB is the loss in precision from integrated squared bias over the region R. Therefore, a positive J tells that the drop in variance is not outweighted by the gain in bias, i.e., there is less error associated with the reduced model using the linear contrasts $C\beta=O$.

Now we are looking for a set of linear contrasts $C\beta=O$ for which J becomes large. For evaluation of J in practice, we need to know the unknown parameters, β and σ^2. Suppose that they may be estimated by $\hat{\beta}$ and $\hat{\sigma}^2$,respectively. Then the *proposed criterion* for screening mixture variables is "find the linear contrast matrix C" which maximizes the quantity,

$$\hat{J}=\text{Tr}[LC(X'X)^{-1}C'L'M]-\frac{1}{\hat{\sigma}^2}\hat{\beta}'C'L'MLC\hat{\beta} \qquad (12)$$

where $L=(X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}$ as defined before.

We observe that the hypothesis testing for $C\beta=O$ is not necessary for this criterion. However, it is easy to see that if the null hypothesis $C\beta=O$ is rejected, then the second term in (12) would be large, which would in turn makes the overall $\hat{J}$ small. Therefore, in general, if $C\beta=O$ is accepted, $\hat{J}$ tends to be large; otherwise, it becomes small.

We also observe that this criterion is characterized by the fact that the normality assumption of $y(\boldsymbol{x})$ is not necessary, and the overall performance of the fitted response surface resulted from the linear contrasts is taken into account. In the following section an example is illustrated to demonstrate the use of this criterion.

4. An Example

The data for this example are the gasoline data which appeared in Snee and Marquardt [11] as a ten-component simple screening design example. If we take the weight function, $W(\boldsymbol{x})$, to be uniform over the entire region of simplex defined in (2), it can be shown (see Appendix) that the region moments in (10) will be

$$M=K\int_R \boldsymbol{x}\boldsymbol{x}'d\boldsymbol{x}=(m_{ij})$$

where $K^{-1}=\int_R d\boldsymbol{x}$ and $m_{ii}=0.01808$ for all i and $m_{ij}=0.00909$ for all $i\neq j$. For this particular example, the matrix $\boldsymbol{X}'\boldsymbol{X}$ takes the similar form as $\boldsymbol{M}$. The diagonal elements of $\boldsymbol{X}'\boldsymbol{X}$ are all 1.44611 and the off-diagonal elements are all equal to

Table 1: Single Linear Contrast ($m=1$)

Linear contrast	F	$\mathrm{Tr}[LC(X'X)^{-1}C'L'M]$	$\hat{\beta}'C'L'MLC\hat{\beta}/\hat{\sigma}^2$	$\hat{J}$
$\beta_5-\beta_6=0$	0.0114	0.0072	0.0001	0.0071
$\beta_8-\beta_9=0$	0.1025	0.0072	0.0007	0.0065
$\beta_4-\beta_7=0$	0.2307	0.0072	0.0017	0.0055
$\beta_2-\beta_7=0$	0.6220	0.0072	0.0045	0.0027
$\beta_2-\beta_{10}=0$	1.3234	0.0072	0.0095	−0.0023
$\beta_9-\sum_{j\neq 9}\beta_j/9=0$	1.3482	0.0072	0.0099	−0.0027
$\beta_2-\beta_4=0$	1.6104	0.0072	0.0116	−0.0044
$\beta_8-\sum_{j\neq 8}\beta_j/9=0$	2.6854	0.0072	0.0197	−0.0125

0.18376. First of all, consider a single linear contrast ($m=1$), i.e., C is a 1×10 matrix. In Table 1, the first eight best contrasts for F and $\hat{J}$ are listed and their F and $\hat{J}$ values are shown. Since the critical F value with (1, 21) degrees of freedom for a 5% Type I error (notation: $F_{0.05}(1, 21)$) is 4.32, each one of the contrasts in Table 1 is accepted by the F-test criterion. However, by the proposed $\hat{J}$ criterion in (12), the last four contrasts are not recommended, because $\hat{J}$ takes negative values. Note that a negative $\hat{J}$ implies that the drop in variance by imposing a linear contrast is not enough to offset the gain in precision by the squared bias. Table 1 also shows that $\mathrm{Tr}[\boldsymbol{LC}(\boldsymbol{X'X})^{-1}\boldsymbol{M}]=0.0072$ for any contrast. This is because of the special structure of the matrices of $\boldsymbol{M}$ and $\boldsymbol{X'X}$ for this particular example.

Now consider a set of linear contrasts ($m\geq 2$) where $\boldsymbol{C}$ contains more than one linear contrast. There are a large number of possible combinations. The result obtained by a computer program written by the author is summarized in Table 2. For each number of rows of the matrix $\boldsymbol{C}$, the best set in terms of $\boldsymbol{F}$ happens to be the best one for $\hat{\boldsymbol{J}}$, and they are given in Table 2. However,

Table 2: Linear Contrasts for $2\leq m\leq 6$

Number of contrast, m	Contrasts	F	$F_{0.05}(m,21)$	$\hat{J}=\mathrm{Tr}[\boldsymbol{LC}(\boldsymbol{X'X})^{-1}\boldsymbol{C'L'M}]-\hat{\boldsymbol{\beta}}'\boldsymbol{C'L'MLC}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2$
2	$\beta_5-\beta_6=0$ $\beta_8-\beta_9=0$	0.0570	3.47	0.0136=0.0144−0.0008
3	$\beta_5-\beta_6=0$ $\beta_8-\beta_9=0$ $\beta_4-\beta_7=0$	0.1149	3.07	0.0191=0.0216−0.0025
4	$\beta_5-\beta_6=0$ $\beta_8-\beta_9=0$ $\beta_4-\beta_7=0$ $\beta_2-\beta_{10}=0$	0.4170	2.84	0.0168=0.0288−0.0120
5	$\beta_5-\beta_6=0$ $\beta_2=\beta_4=\beta_7$ $\beta_8-\sum_{j\neq 8}\beta_j/9=0$ $\beta_9-\sum_{j\neq 9}\beta_j/9=0$	1.2332	2.68	−0.0090=0.0360−0.0450
6	$\beta_5-\beta_6=0$ $\beta_2=\beta_4=\beta_7=\beta_{10}$ $\beta_8-\sum_{j\neq 8}\beta_j/9=0$ $\beta_9-\sum_{j\neq 9}\beta_j/9=0$	1.8857	2.57	−0.0383=0.0432−0.0815

in general, this is not necessarily true.

We observe that every set of the contrasts listed in Table 2 is accepted by the F-test and it is not clear which set should be used to reduce the number of mixture variables. By the $\hat{J}$ criterion, it is easily observed that the set corresponding to $m=3$ is recommended since $\hat{J}$ reaches the maximum. If one intends to reduce more variables, the set for $m=4$ is a good choice. The sets for $m=5$ and 6 are not recommended because they produce negative $\hat{J}$ values. If one chooses the contrasts corresponding to $m=3$, then the gasoline data may be analyzed as a seven-component system, since the sums of components(5 and 6), (8 and 9) and (4 and 7) may be treated as three new components.

5. Discussion

When we search for the contrast matrix $\boldsymbol{C}$ which makes $\hat{J}$ large, we first note that $\hat{J}$ is dimensionless and unaffected by a change of scale of the elements of $\boldsymbol{C}$. This indeterminacy can be removed without loss of generality by imposing the constaint

$$\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'=\boldsymbol{I}. \tag{13}$$

For a given matrix $\boldsymbol{C}$ observe that, if $\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'=\boldsymbol{H}\boldsymbol{H}'$ where $\boldsymbol{H}$ is an $m\times m$ nonsingular matrix, there always exists a corresponding matrix $\tilde{\boldsymbol{C}}=\boldsymbol{H}^{-1}\boldsymbol{C}$ such that $\tilde{\boldsymbol{C}}(\boldsymbol{X}'\boldsymbol{X})^{-1}\tilde{\boldsymbol{C}}'=\boldsymbol{I}$.

With the restriction (13), the $\hat{J}$-statistics in (12) may be written as

$$\hat{J}=\mathrm{Tr}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{M}]-\frac{1}{\hat{\sigma}^2}\hat{\boldsymbol{\beta}}'\boldsymbol{C}'\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{M}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\boldsymbol{C}\hat{\boldsymbol{\beta}}. \tag{14}$$

Frequently one is interested in the estimated responses at the design points only. If the weight function $W(\boldsymbol{x})=1/n$ at the data points and $W(\boldsymbol{x})=0$ elsewhere, then

$$\begin{aligned}\boldsymbol{M}&=K\int_R \boldsymbol{x}\boldsymbol{x}'\, dW(\boldsymbol{x})\\&=\boldsymbol{X}'\boldsymbol{X}/n.\end{aligned}$$

Substituting this $\boldsymbol{M}$ into (14), we obtain

$$\hat{J}=\frac{1}{n}\mathrm{Tr}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\boldsymbol{C}]-\frac{1}{n\hat{\sigma}^2}\hat{\boldsymbol{\beta}}'\boldsymbol{C}'\boldsymbol{C}\hat{\boldsymbol{\beta}}$$

$$=(\text{rank of } \boldsymbol{C})/n - mF/n$$
$$= m(1-F)/n$$

where m is the rank of $\boldsymbol{C}$ and F is the F-statistic described in Section 2. Note that for this particular case of $\boldsymbol{M}=\boldsymbol{X}'\boldsymbol{X}/n$, linear contrasts $\boldsymbol{C}\boldsymbol{\beta}=\boldsymbol{O}$ are adopted for purposes of screening variables if $F<1$ This is a more restrictive condition than the usual F-test.

Let the matrix $\boldsymbol{C}$ which maximizes the first term in (14) be called 'all variance matrix' and the $\boldsymbol{C}$ which minimizes the second term in (14) be called 'all bias matrix'. It is of theoretical interest to find the optimal matrices.

Let Λ denote the diagonal matrix of eigenvalues, λ_i, of $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{T}$ denote the corresponding orthogonal matrix of eigenvectors. That is,

$$\boldsymbol{T}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}=\Lambda \quad \text{and} \quad \boldsymbol{T}'\boldsymbol{T}=\boldsymbol{T}\boldsymbol{T}'=\boldsymbol{I}$$

where $\Lambda=\text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_q)$. We assume that the mixture components are arranged so that the magnitude of the eigenvalues is $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_q > 0$. Also let the *ith* column of matrix $\boldsymbol{T}$ be denoted by $\boldsymbol{t}_i$. Then one can write

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}=\boldsymbol{T}\Lambda^{-1}\boldsymbol{T}'=\boldsymbol{T}\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}\boldsymbol{T}'.$$

Now the first term in (14) may be written as

$$\begin{aligned}\text{Tr}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{M}] &= \text{Tr}[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{M}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'] \\ &= \sum_{i=1}^{m} \boldsymbol{r}'_i(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{M}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{r}_i \\ &= \sum_{i=1}^{m} \boldsymbol{r}'_i\boldsymbol{T}\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}\boldsymbol{T}'\boldsymbol{M}\boldsymbol{T}\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}\boldsymbol{T}'\boldsymbol{r}_i \\ &= \sum_{i=1}^{m} \boldsymbol{a}'_i\boldsymbol{V}\boldsymbol{a}_i\end{aligned}$$

where $\boldsymbol{r}'_i$ is the ith row of matrix $\boldsymbol{C}, \boldsymbol{a}_i'=\boldsymbol{r}'_i\boldsymbol{T}\Lambda^{-\frac{1}{2}}$ and $\boldsymbol{V}=\Lambda^{-\frac{1}{2}}\boldsymbol{T}'\boldsymbol{M}\boldsymbol{T}\Lambda^{-\frac{1}{2}}$. Note that $\boldsymbol{a}'_i\boldsymbol{a}_i=1$ from the constraint (13). It is known from the matrix theory (see, for instance, Rao [7]) that $\boldsymbol{a}'_1\boldsymbol{V}\boldsymbol{a}_1$ is maximized if $\boldsymbol{a}_1=\boldsymbol{v}_1$ where $\boldsymbol{v}_1$ is the eigenvector corresponding to the largest eigenvalue of matrix $\boldsymbol{V}$. Therefore, from $\boldsymbol{a}'_1=\boldsymbol{r}'_1\boldsymbol{T}\Lambda^{-\frac{1}{2}}=\boldsymbol{v}_1$ the first row of 'all variance matrix' will be

$$\boldsymbol{r}'_1=\boldsymbol{v}'_1\Lambda^{\frac{1}{2}}\boldsymbol{T}'$$

and similarly we can obtain that the 'all variance matrix' C is

$$C=\begin{pmatrix} v'_1 \\ v'_1 \\ \vdots \\ v'_m \end{pmatrix} \Lambda^{\frac{1}{2}} T'.$$

The second term in (14) becomes zero if $C\hat{\beta}=O$. Therefore, the 'all bias matrix' C has independent rows which are orthogonal to $\hat{\beta}$,

$$C=\{r'_1 : r'_i\beta=O \text{ for } i=1, 2, \ldots, m\}.$$

Lastly, it should be mentioned that the development so far was discussed in terms of the linear mixture model in (1); however, the screening strategy, in general, may be applicable to a higher order mixture model (quadratic, cubic, etc.). Note that the matrix C is not necessarily restricted to linear contrasts. It can be any type of linear constraints. In particular, for a higher order model, the zero restriction (for instance, dropping the variable $x_i x_j$ if $\beta_{ij}=0$ is adopted) may well be used. For screening polynomial terms for a high order mixture model, it would be interesting to compare the results obtained by this strategy with those obtained from the Park's method [6].

Appendix

The elements of the M matrix defined in (10) are of the form

$$M=K\int_R x_1^{c_1} x_2^{c_2} \ldots x_q^{c_q} dx_1 dx_2 \ldots dx_q$$

and

$$K^{-1}=\int_R dx_1 dx_2 \ldots dx_q$$

if $W(x)$ is assumed to be uniform over some region of interest, R. Let R denote the $(q-1)$-dimensional simplex described in (2). For computational purposes R may be written as

$$R^*=\{(x_1, x_2, \ldots, x_{q-1}) : x_i \geqslant 0 \text{ and } \sum_{i=1}^{q-1} x_i \leqslant 1\}$$

since $x_q=1-\sum_{i=1}^{q-1} x_i$.

It is is known (see, for instance, Ryzhik and Gradshtein [8]) that

$$\int_{R^*} \Pi x_i^{p_i-1} dx_1 dx_2 \ldots dx_{q-1} = \frac{\Pi \Gamma(p_i)}{\Gamma(\sum p_i)} \int_0^1 y^{(\Sigma p_i)-1} dy$$

where the sums ($\sum$) and products (Π) are over $i=1, 2, \ldots, q-1$. For example, consider $\int_{R^*} x_i^2 \, dx^*$ where $dx^* = dx_1 dx_2 \ldots dx_9$. Since $p_i=3$ and $p_j=1$ for $j \neq i$, we can obtain that

$$\int_{R^*} x_i^2 dx^* = \frac{2!}{10!} \int_0^1 y^{10} dy = \frac{2}{11!}$$

Similarly, it can be readily shown that, for $j \neq i$,

$$\int_{R^*} x_i x_j dx^* = \frac{1}{11!} \text{ and} \int_R dx^* = \frac{1}{9!}.$$

Therefore, in the example in Section 4,

$$m_{ii} = K \int_R x_i^2 dx = \left(\frac{2}{11!}\right)(9!) = 0.01818$$

and

$$m_{ij} = K \int_R x_i x_j dx = \left(\frac{1}{11!}\right)(9!) = 0.00909.$$

REFERENCES

[1] Allen, D, "Mean Square Error of Prediction as Criterion for Selecting Variables," *Technometrics*, Vol. 13 (1971), pp. 469-476.

[2] Draper, N. and Smith, H., *Applied Regression Analysis*, Wiley Inc., New York, 1966.

[3] Helms, R., "The Average Estimated Variance Criterion for the Selection of Variables Problem in General Linear Models," *Technometrics*, Vol. 16 (1974), pp. 261-274.

[4] Hocking, R., "The Analysis and Selection of Variables in Linear Regression," *Biometrics*, Vol. 32(1976), pp. 1-49.

[5] Mallows, C., "Some Comments on C_p," *Technometrics*, Vol. 15(1973), pp. 661-675.

[6] Park, S., "Selection of Polynomial Terms for Response Surface Experiments," *Biometrics*, Vol. 33(1977), pp. 225-229.

[7] Rao, C. R., *Linear Statistical Inference and Its Applications*, Wiley Inc., New York, 1973.

[8] Ryzhik, I. and Gradshteyn, I., *Tables of Integrals, Series and Products*, Academy Press, New York, 1965.

[9] Scheffe, H., "Experiments with Mixtures," *Journal of Royal Statistical Society* B, Vol. 20(1958), pp. 344-360.

[10] Searle, S., *Linear Models*, Wiley Inc., New York, 1971.

[11] Snee, R. and Marquardt, D., "Screening Concepts and Designs for Experiments with Mixtures," *Technometrics*, Vol. 18(1976), pp. 19-30.