VARIANCE FUNCTIONS FOR COMPARING MIXTURE DESIGNS

A. K. NIGAM

(Received Dec. 20, 1971)

1. Introduction

The efforts of Scheffé [4], [5] and Murty and Das [2] led to a wide range of designs of experiments with mixtures. In the present paper we have considered the problem of comparing these designs when they are used to fit a quadratic surface of the form

$$y_u = \sum \beta_i x_{iu} + \sum \beta_{ij} x_{iu} x_{ju} \qquad u=1, 2, \cdots, N \tag{1.1}$$

where y_u is the observed response at the uth experimental point and x_{iu} is the proportion of ith component in the uth mixture combination.

Though certain problems of interest may arise when the model (1.1) fails to represent the true response surface, we have throughout assumed that the surface (1.1) represents quite reasonably the true relationship and any bias due to the inadequacy of fit is negligible.

Before any two designs can be compared, a requirement, as suggested by Box and Wilson [1], is to make the spread of items of the coded levels for each of the factors equal in both designs. While comparing two mixture designs in n components, the interesting feature is that the levels (the proportions) are always taken on real basis and the spread in the two designs is always the same for each of the components (because of $0 \leqq x_{iu} \leqq 1$).

In the next section we have considered different variance functions for comparing mixture designs.

2. Variance of the estimated response

Let $\mathrm{V}(\hat{y})$ be the variance of the estimated response $\hat{y}$. Then, while comparing different mixture designs, it would be worthwhile to compare the quantity $N\mathrm{V}(\hat{y})$ as the total number of mixtures N may not be equal for the designs under comparison.

We now proceed to describe three alternative variance functions for comparing mixture designs:

2.1. *Variances at the optimum*

One can compare the expressions $N_1 \mathrm{V}(\hat{y})_{\text{opt.}}$ and $N_2 \mathrm{V}(\hat{y}')_{\text{opt.}}$ where N_1 and N_2 are the number of mixtures in the two designs and $\mathrm{V}(\hat{y})_{\text{opt.}}$ and $\mathrm{V}(\hat{y}')_{\text{opt.}}$ are the variances at the optimum of the surfaces of the two designs. The more or less obvious limitation of this approach is that the optimum is never known in advance. The procedure is useful when there is some a priori information of the estimated optimum through previous experimentation.

2.2. *Variances at any other point* $(x_{1u}, x_{2u}, \cdots, x_{nu})$

The two variances can be compared at other specific points such as at the centre, vertices or the edges of the simplex depending upon the region of immediate interest.

2.3. *Over all variances*

When the optimum of the surface is not known or specific location of the region of interest is not available then it may be appropriate to consider the entire simplex as the region of interest. It is then proper to compare the 'over all' variances of the two designs. 'Over all' variance indicates the measure of dispersion over a region bounded by all points of the simplex. Such a comparison would yield the efficiency of one design on an average over the other.

Let there be n components in all the designs which are desired to be compared. If x_{iu} is the proportion of the ith component in the uth point of an n-component and N-point mixture design, then we know that

$$x_{iu} \geqq 0 \tag{2.1}$$

and

$$\sum_{1 \leqq i \leqq n} x_{iu} = 1 \; . \tag{2.2}$$

If we are to find 'over all' variance of the estimated response under the region (2.2), we must integrate $\mathrm{V}(\hat{y})$ subject to $x_{iu} \geqq 0$ and $\sum_{1 \leqq i \leqq n} x_{iu} = 1$. Thus, the 'over all' variance is given

$$U = \int_R \mathrm{V}(\hat{y}) \tag{2.3}$$

where R is the region bounded by the $(n-1)$-dimensional space given by (2.1) and (2.2).

We know that for a quadratic model (1.1), the variance of the estimated response $\mathrm{V}(\hat{y})$ is given by

$$\text{(2.4)} \qquad \mathrm{V}(\hat{y})=\mathrm{V}(b_i)\sum x_i^2+\mathrm{V}(b_{ij})\sum x_i^2x_j^2+2\operatorname{cov}(b_ib_j)\sum x_ix_j$$
$$+2\operatorname{cov}(b_ib_{ij})\sum x_i^2x_j+2\operatorname{cov}(b_ib_{jk})\sum x_ix_jx_k$$
$$+2\operatorname{cov}(b_{ij}b_{kl})\sum x_ix_jx_kx_l+2\operatorname{cov}(b_{ij}b_{ik})\sum x_i^2x_jx_k$$

where the summations range over 1 to n. We, thus, have

$$\text{(2.5)} \qquad U=\mathrm{V}(b_i)\int_R \sum x_i^2 \prod dx_i+\mathrm{V}(b_{ij})\int_R \sum x_i^2x_j^2 \prod dx_i+\cdots$$
$$+2\operatorname{cov}(b_{ij}b_{ik})\int_R \sum x_i^2x_jx_k \prod dx_i$$

$$\text{(2.6)} \qquad =\mathrm{V}(b_i)I_1+\mathrm{V}(b_{ij})I_2+\cdots+2\operatorname{cov}(b_{ij}b_{ik})I_7 \qquad \text{(say)}$$

where R is given by (2.1) and (2.2).

We now evaluate the integrals I_j; $j=1, 2, \cdots, 7$. We have

$$\text{(2.7)} \qquad I_1=\int \sum_{1\leq i\leq n} x_i^2 \prod_{i=1}^{n-1} dx_i\,; \qquad x_i\geqq 0,\ \sum_{1\leq i\leq n} x_{iu}=1,$$
$$\text{for all } i=1, 2, \cdots, n$$

I_1 may be written as

$$\text{(2.8)} \qquad I_1=n\int x_1^2 \prod_{i=1}^{n-1} dx_i$$

$$\text{(2.9)} \qquad =n\int x_1^{3-1}x_2^{1-1}\cdots x_{n-1}^{1-1}x_n^{1-1} \prod_{i=1}^{n-1} dx_i\,.$$

Now, I_1 may be rewritten, by virtue of (2.2), as

$$\text{(2.10)} \qquad I_1=n\int x_1^{3-1}x_2^{1-1}\cdots x_{n-1}^{1-1}(1-x_1-x_2-\cdots-x_{n-1})^{1-1} \prod_{i=1}^{n-1} dx_i$$

where $\sum_{1\leq i\leq n-1} x_i\leqq 1$. This can easily be evaluated by using Liouville's extension of Dirichlet integrals. Thus, I_1 is given by

$$\text{(2.11)} \qquad I_1=n\frac{\Gamma(3)\Gamma(1)\cdots\Gamma(1)}{\Gamma(3+1+\cdots+1)}=\frac{n\Gamma(3)}{\Gamma(n+2)}=\frac{2n}{\Gamma(n+2)}\,.$$

Similarly, we have the following

$$\text{(2.12)} \qquad I_2=\int \sum_{1\leq i<j\leq n} x_i^2x_j^2 \prod_{i=1}^{n-1} dx_i \qquad \text{for all } i, j=1, 2, \cdots, n$$

$$\text{(2.13)} \qquad =\frac{n(n-1)}{2}\,\frac{\Gamma(3)\Gamma(3)}{\Gamma(6+n-2)}=\frac{2n(n-1)}{\Gamma(n+4)}$$

$$\text{(2.14)} \qquad I_3=\int \sum_{1\leq i<j\leq n} x_ix_j \prod_{i=1}^{n-1} dx_i$$

$$\text{(2.15)} \qquad =\frac{n(n-1)}{2}\,\frac{1}{\Gamma(n+2)}=\frac{n(n-1)}{2\Gamma(n+2)}$$

$$(2.16)\quad I_4=\int \sum_{1\leq i<j\leq n} x_i^2 x_j \prod_{i=1}^{n-1} dx_i$$

$$(2.17)\quad =\frac{n(n-1)}{2}\,\frac{2}{\Gamma(n+3)}=\frac{n(n-1)}{\Gamma(n+3)}$$

$$(2.18)\quad I_5=\int \sum_{1\leq i<j<k\leq n} x_i x_j x_k \prod_{i=1}^{n-1} dx_i$$

$$(2.19)\quad =\frac{n(n-1)(n-2)}{3\cdot 2}\,\frac{1}{\Gamma(n+3)}=\frac{n(n-1)(n-2)}{6\Gamma(n+3)}$$

$$(2.20)\quad I_6=\int \sum_{1\leq i<j<k<l\leq n} x_i x_j x_k x_l \prod_{i=1}^{n-1} dx_i$$

$$(2.21)\quad =\frac{n(n-1)(n-2)(n-3)}{4\cdot 3\cdot 2}\,\frac{1}{\Gamma(n+4)}=\frac{n(n-1)(n-2)(n-3)}{24\Gamma(n+4)}$$

$$(2.22)\quad I_7=\int \sum_{1\leq i<j<k\leq n} x_i^2 x_j x_k \prod_{i=1}^{n-1} dx_i$$

$$(2.23)\quad =\frac{n(n-1)(n-2)}{3\cdot 2}\,\frac{2}{\Gamma(n+4)}=\frac{n(n-1)(n-2)}{3\Gamma(n+4)}\,.$$

We, thus, have

$$(2.24)\quad \begin{aligned} U=&\frac{2n}{\Gamma(n+2)}\,\mathrm{V}\,(b_i)+\frac{2n(n-1)}{\Gamma(n+4)}\,\mathrm{V}\,(b_{ij})+\frac{n(n-1)}{\Gamma(n+2)}\,\mathrm{cov}\,(b_ib_j)\\ &+\frac{2n(n-1)}{\Gamma(n+3)}\,\mathrm{cov}\,(b_ib_{ij})+\frac{n(n-1)(n-2)}{3\Gamma(n+3)}\,\mathrm{cov}\,(b_ib_{jk})\\ &+\frac{n(n-1)(n-2)(n-3)}{12\Gamma(n+4)}\,\mathrm{cov}\,(b_{ij}b_{kl})+\frac{2n(n-1)(n-2)}{3\Gamma(n+4)}\,\mathrm{cov}\,(b_{ij}b_{ik})\\ =&\frac{n}{\Gamma(n+2)}\Big[2\,\mathrm{V}\,(b_i)+(n-1)\Big\{\frac{2\,\mathrm{V}\,(b_{ij})}{(n+3)(n+2)}+\mathrm{cov}\,(b_ib_j)\\ &+\frac{2}{n+2}\,\mathrm{cov}\,(b_ib_{ij})+\frac{n-2}{3(n+2)}\,\mathrm{cov}\,(b_ib_{jk})\\ &+\frac{(n-2)(n-3)}{12(n+3)(n+2)}\,\mathrm{cov}\,(b_{ij}b_{kl})+\frac{2(n-2)}{3(n+3)(n+2)}\,\mathrm{cov}\,(b_{ij}b_{ik})\Big\}\Big]\,. \end{aligned}$$

The expression (2.24) may be used to compare 'over all' variances for different designs under comparison.

Acknowledgement

The author is grateful to Dr. M. N. Das, senior professor of statistics at IARS, New Delhi for his help and guidance.

BANARAS HINDU UNIVERSITY, VARANASI, INDIA*

REFERENCES

[1] Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimum conditions, *J. R. Statist. Soc.*, B, **13**, 1-45.
[2] Murty, J. S. and Das, M. N. (1968). Design and analysis of experiments with mixtures, *Ann. Math. Statist.*, **39**, 1517-1539.
[3] Nigam, A. K. (1969). Contributions to design and analysis of experiments with mixtures, Ph.D. dissertation submitted to Banaras Hindu University.
[4] Scheffé, H. (1958). Experiments with mixtures, *J. R. Statist. Soc.*, B, **20**, 344-360.
[5] Scheffé, H. (1963). Simplex-centroid designs for experiments with mixtures, *J. R. Statist. Soc.*, B, **25**, 235-263.

* Now at Institute of Agricultural Research Statistics, New Delhi.