On the Construction of Approximate Multi-Factor Designs from Given Marginals Using the Iterative Proportional Fitting Procedure

By F. Pukelsheim[1] and D. M. Titterington[2]

Summary: For multifactor designs based on linear models, the information matrix generally depends on a certain set of marginal tables created from the design itself. This note considers the problems of whether a set of marginal tables is consistent, in that a design exists that can yield them, and of calculating such a design when at least one does exist. The results are obtained by direct analogy with the problem of maximum likelihood estimation in loglinear models for categorical data.

1 Introduction

Suppose an experiment involves three factors at levels $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$, respectively. An approximate design is defined by a probability measure, ξ, on the design space

$$X = \{1, \ldots, a\} \times \{1, \ldots, b\} \times \{1, \ldots, c\},$$

so that $\xi(i, j, k)$ denotes the proportion of all observations to be taken at the combination (i, j, k) of factor levels. A particular manifestation of this is the class of multiway block designs in which one factor is singled out to be a set of varieties or treatments, and the remaining factors are blocking factors.

Any design ξ is to be thought of as a vector of probabilities of dimension abc with elements $\xi(i, j, k)$ in lexicographic order.

[1] Friedrich Pukelsheim, Institut für Mathematik, Universität Augsburg, D-8900 Augsburg, Federal Republic of Germany.

[2] D. Michael Titterington, Department of Statistics, University of Glasgow, Glasgow, G12 8QW, Great Britain.

We shall consider only linear models, with the assumptions that errors are uncorrelated, of zero means and of constant variances. Thus an observation generated from a combination (i, j, k) of factor levels will have mean

$$f(i, j, k)'\beta, \tag{1}$$

where $f(i, j, k)$ is a design vector, β is the vector of parameters in the linear model, and a prime denotes transposition. A design ξ therefore has information matrix

$$M(\xi) = \sum_{i,j,k} f(i, j, k)f(i, j, k)'\xi(i, j, k). \tag{2}$$

Here we do not discuss optimality questions for the particular case of multiway block designs, but take up the following rather more basic questions about the information matrix $M(\xi)$. Suppose we are given a matrix M that is alleged to be an information matrix as defined in (2) above.

(i) Is there indeed a design ξ such that $M(\xi) = M$?

(ii) Is it possible to compute a compatible design ξ?

Note that we are not going so far as to demand from (ii) a design that is necessarily optimal in any familiar sense. We may of course be lucky, but any design we come up with might at least be used as the starting point for some algorithm for computing an optimal design.

Section 2 of this note is devoted to the special model with additive main effects. The crucial link is forged with the topic of loglinear models for contingency tables and the Iterative Proportional Fitting Procedure for obtaining maximum likelihood estimates of the parameters therein. Section 3 looks beyond the main effects model, and indicates the straightforward extension to an arbitrary number of factors. Questions (i) and (ii) have been touched upon in the context of multiway block designs in Pukelsheim (1986).

2 Main Effects Model

Suppose the model contains only main effects and denote these effects by $\{\alpha_i, i = 1, \ldots, a\}$, $\{\gamma_j, j = 1, \ldots, b\}$, $\{\delta_k, k = 1, \ldots, c\}$. This model is described by (1) provided

$$\beta = \begin{bmatrix} \alpha \\ \gamma \\ \delta \end{bmatrix}, \quad \text{and} \quad f(i, j, k) = \begin{bmatrix} e_i^a \\ e_j^b \\ e_k^c \end{bmatrix},$$

where $\alpha = (\alpha_1, \ldots, \alpha_a)'$ and e_i^a is the a-dimensional vector with unity in the i-th position and zeroes elsewhere, etc. Since $f(i, j, k)'\beta = \alpha_i + \gamma_j + \delta_k$ this is the model with additive main effects as desired.

A design ξ then has information matrix

$$M(\xi) = \begin{bmatrix} \Delta_0 & W_{01} & W_{02} \\ W_{01}' & \Delta_1 & W_{12} \\ W_{02}' & W_{12}' & \Delta_2 \end{bmatrix} \tag{3}$$

in which the W's are matrices of two-dimensional marginals of ξ, and the Δ's are diagonal matrices whose diagonal elements are a set of one-dimensional marginals of ξ. For instance, the (i, j)-th element of W_{01} is obtained from $\xi(i, j, k)$ by summing over k, and the i-th diagonal element of Δ_0 is obtained by then summing over j as well.

Another way of describing the "weight matrices" W_{pq} is obtained with the help of a further notation. Let a dot denote summation. Then, for example,

$$(W_{01})_{ij} = \xi(i, j, \cdot), \qquad \text{for all } i, j, \tag{4}$$

$$(W_{02})_{ik} = \xi(i, \cdot, k), \qquad \text{for all } i, k, \tag{5}$$

$$(W_{12})_{jk} = \xi(\cdot, j, k), \qquad \text{for all } j, k, \tag{6}$$

and

$$(\Delta_0)_{ii} = \xi(i, \cdot, \cdot) \quad \text{etc.}$$

The numbering of the factors is motivated by the context of multiway block designs where factor 0 comprises varieties or treatments and plays a distinctive role, while the blocking factors 1 and 2 give rise to nuisance parameters.

The grand information matrix $M(\xi)$ in (3) is determined by the two-dimensional marginals of ξ. These are sufficient because they clearly imply the one-dimensional marginals.

Question (i) of Section 1 becomes the following. Given matrices $\{W_{pq}\}$, is there a design ξ for which the matrices $\{W_{pq}\}$ are indeed the two-dimensional marginals?

The main problem is to discover whether the marginals $\{W_{pq}\}$ are mutually consistent. Kellerer (1964) derives a necessary and sufficient condition for consistency that applies to any of the problems considered in this note. However, the condition appears to be very difficult to apply in practice and here we present a more practical, if algorithmic and inexact, check. To describe the check we must introduce the procedure by which we intend to answer question (ii) of Section 1: if there is at least one design with which the two-dimensional marginals are compatible, can we compute one of them?

The proposed procedure for finding a design, if one exists, is as follows.

First choose a design measure ξ_0 with two mild provisos, that ξ_0 is of a factorized form consonant with (12) below and that if $(W_{pq})_{ij} > 0$, say, then the appropriate margin of ξ_0 is also positive. Clearly these conditions are achieved by taking $\xi_0(i, j, k) \equiv 1/(abc)$.

Then we compute

$$\xi_1(i, j, k) = \frac{\xi_0(i, j, k)}{\xi_0(i, j, \cdot)} (W_{01})_{ij}, \quad \text{for all } i, j, k, \tag{7}$$

$$\xi_2(i, j, k) = \frac{\xi_1(i, j, k)}{\xi_1(i, \cdot, k)} (W_{02})_{ik}, \quad \text{for all } i, j, k, \tag{8}$$

$$\xi_3(i, j, k) = \frac{\xi_2(i, j, k)}{\xi_2(\cdot, j, k)} (W_{12})_{jk}, \quad \text{for all } i, j, k. \tag{9}$$

For any case for which $(W_{pq})_{ij} = 0$, the corresponding elements in the next ξ_n are to be automatically set to zero.

Note that ξ_1 satisfies (4), ξ_2 satisfies (5), and ξ_3 satisfies (6).

Equations (7), (8) and (9) represent one cycle of the algorithm. It continues by returning to (7) with ξ_3 in place of ξ_0 and recycling until "convergence" occurs.

We may state the following result.

Theorem 1:

(a) There is at least one design ξ with which W_{01}, W_{02} and W_{12} are compatible if and only if the above algorithm converges.

(b) If the algorithm converges then its limit $\hat{\xi}$ is a design compatible with W_{01}, W_{02} and W_{12}. Furthermore, the elements of $\hat{\xi}$ are of a special form, namely

$$\hat{\xi}(i, j, k) = \hat{\rho}_{ij}\hat{\sigma}_{ik}\hat{\tau}_{jk}, \tag{10}$$

for all i, j, k, and for certain $\hat{\rho}, \hat{\sigma}, \hat{\tau}$.

(c) The design $\hat{\xi}$ given in (10) is the only design of form (10) compatible with W_{01}, W_{02} and W_{12}.

(d) The design $\hat{\xi}$ given in (10) is the design ξ, of the factorized form indicated by (10), which maximizes

$$\sum_{i,j} (W_{01})_{ij} \log \xi(i, j, \cdot) + \sum_{i,k} (W_{02})_{ik} \log \xi(i, \cdot, k) + \sum_{j,k} (W_{12})_{jk} \log \xi(\cdot, j, k), \tag{11}$$

over ρ, σ, τ provided that W_{01}, W_{02}, W_{12} are consistent. □

It should be emphasized that there may be many designs other than (10) that are compatible with W_{01}, W_{02} and W_{12}. It should also be acknowledged that the design given by (10) may not be an exact design, thus limiting the immediate practical importance of the result.

From the point of view of question (i) of Section 1 the important part of the theorem is part (a), in that convergence of the algorithm clearly leads to a feasible design.

We are relieved of the need to prove Theorem 1 ourselves, provided we notice the parallel between the structure of our problem and that of maximum likelihood estimation of parameters in loglinear models for contingency tables. Suppose, in a three-way contingency table, $\xi(i, j, k)$ is the probability of obtaining an observation in row i, column j, and layer k. Suppose also that a loglinear model is specified in which

$$\log \xi(i, j, k) = \rho_{ij} + \sigma_{ik} + \tau_{jk}, \tag{12}$$

for all i, j and k. Note that, for instance, ρ_{ij} in (12) corresponds to $\log \hat{\rho}_{ij}$ in (10). Suppose also that, under a multinomial model, W_{01}, W_{02} and W_{12} represent the row/column, row/layer and column/layer marginal sets of relative frequencies. Then (11)

is the loglikelihood, (4), (5) and (6) are the likelihood equations, and the following theorem can be compiled.

Theorem 2:

(a) The algorithm converges if and only if W_{01}, W_{02} and W_{12} are consistent.

(b) If it does converge, the algorithm converges to the unique set of probabilities of the form (12) that satisfy (4), (5) and (6).

(c) The limit $\hat{\xi}$ is the unique distribution ξ of the form (12) that maximizes the loglikelihood. □

Proofs of the more general versions of these statements, which we shall state in Section 3, appear in Darroch and Ratcliff (1972) and, in part, in Darroch (1962), as remarked by Bishop et al. (1975, p. 101); see also Csiszár (1975).

The algorithm described earlier is called the *Iterative Proportional Fitting Procedure* (IPFP) or *iterative scaling,* in the context of loglinear models. It is also called the *raking method* when used as a device for constructing a full table of frequencies from sets of marginal tables in the context of incomplete data from sample surveys: see, for instance, Oh and Scheuren (1983). Section 3.5 of Bishop et al. (1975) provides a detailed discussion of the practice and properties of the IPFP.

Theorem 1 clearly follows from Theorem 2.

In numerical examples we used total variation distance to measure the deviations between the current and the last round, i.e. between ξ_n and ξ_{n-3}, ξ_{n+1} and ξ_{n-2}, and ξ_{n+2} and ξ_{n-1}. Within each round we computed the sum of the total variation distances between ξ_n and ξ_{n+1}, ξ_{n+1} and ξ_{n+2}, and ξ_{n+2} and ξ_n. Consider for instance the two sets of marginals from Krafft (1978, pp. 186–188) both of which are exact for 20 observations on 3 treatments in 4 x 5 blocks with uniform marginals W_{12}. In the case

$$W_{01} = \frac{1}{20}\begin{pmatrix} 4 & 0 & 1 & 0 \\ 0 & 3 & 3 & 3 \\ 1 & 2 & 1 & 2 \end{pmatrix}, \quad W_{02} = \frac{1}{20}\begin{pmatrix} 2 & 1 & 0 & 0 & 2 \\ 1 & 2 & 3 & 3 & 0 \\ 1 & 1 & 1 & 1 & 2 \end{pmatrix}$$

it is known that no common joint distribution ξ exists for W_{01}, W_{02}, and W_{12}. Indeed it becomes evident after a first dozen of rounds or so that the IPFP gets into a cycle with one round as period, while within each round the sum of the total variation distances remains constant 0.39.

For the second example, with

$$W_{01} = \frac{1}{20}\begin{pmatrix} 2 & 1 & 3 & 3 \\ 1 & 2 & 2 & 1 \\ 2 & 2 & 0 & 1 \end{pmatrix}, \quad W_{02} = \frac{1}{20}\begin{pmatrix} 2 & 3 & 0 & 2 & 2 \\ 0 & 1 & 3 & 1 & 1 \\ 2 & 0 & 1 & 1 & 1 \end{pmatrix},$$

the algorithm converges after a couple of dozen of rounds to a design $\hat{\xi}$ which is distinct from the design D of Krafft (1978, p. 186).

As discussed in Section 3.4 of Bishop et al. (1975) there are some models for which parameter estimates of the required form can be written down explicitly without the need for recourse to the IPFP, although in these cases the algorithm generally reaches the solution after one cycle anyway. For the three-way table one such model is of the form

$$\log \xi(i, j, k) = \rho_{ij} + \sigma_{ik},$$

for which the associated sufficient statistics are W_{01} and W_{02} and the maximum likelihood estimates of ξ are

$$\hat{\xi}(i, j, k) = (W_{01})_{ij}(W_{02})_{ik}/(\Delta_0)_{ii}. \tag{13}$$

In the design context, this is a special case of the conditional block-block product designs considered by Pukelsheim (1986). That (13) represents a valid design is clear. Each component of $\hat{\xi}$ is nonnegative, and summation over j, k and the i gives a total of unity. We note that the conditional block-block product designs often have good properties but may not be optimal. Design BAD of Pukelsheim (1983) is a nonoptimal design of this type.

3 More General Models

First we briefly indicate the extension to more than three factors, which follows the same pattern as for multiway block designs (Pukelsheim 1983, 1986). Suppose an experiment involves $m + 1$ factors and that factor k can appear at levels $j_k = 1, \ldots, b_k$, $k = 0, 1, \ldots, m$. A design ξ then is a probability measure on the design space

$$X = \{1, \ldots, b_0\} \times \ldots \times \{1, \ldots, b_m\}.$$

In the model with only additive main effects the information matrix of ξ turns out to be

$$M(\xi) = \begin{bmatrix} \Delta_0 & W_{01} & \dots & W_{0m} \\ W'_{01} & \Delta_1 & \dots & W_{1m} \\ \dots & \dots & \dots & \dots \\ W'_{0m} & W'_{1m} & \dots & \Delta_m \end{bmatrix}.$$

Hence the grand information matrix $M(\xi)$ again is solely determined by the two-dimensional marginals $\{W_{pq}\}$ of ξ. Application of the IPFP to this case is straightforward.

If the underlying model contains terms other than main effects, then the information matrix contains third-order or higher-order marginals. The following example illustrates inclusion of three-dimensional marginals.

Suppose the model includes all first-order interactions of factor 0 with factors $1, \dots, m$. Thus a single observation at levels $j_0, \dots, j_m$ has expected value

$$\alpha^{(01)}_{j_0 j_1} + \alpha^{(02)}_{j_0 j_2} + \dots + \alpha^{(0m)}_{j_0 j_m}.$$

If we define $\alpha^{(01)}$ etc. by lexicographic order, this corresponds to

$$\beta' = [\alpha^{(01)'}, \dots, \alpha^{(0m)'}],$$

$$f(j_0, \dots, j_m)' = [\{e^{b_0}_{j_0} \otimes e^{b_1}_{j_1}\}', \dots, \{e^{b_0}_{j_0} \otimes e^{b_m}_{j_m}\}'],$$

in which $\otimes$ denotes the Kronecker product.

Then $M(\xi)$ contains all three-dimensional marginals $\{W_{0pq}, 0 < p < q\}$, as well as certain lower-order marginals which can be derived from $\{W_{0pq}\}$.

It is now clear that the set of marginals which determines the grand information matrix $M(\xi)$ depends on which effects enter into the regression of the expected value. The list of examples could be extended along this line.

In general, $M(\xi)$ can be defined by a minimal system of marginals $\{W_T, T \in \mathcal{T}\}$. In the contingency-table literature, these marginals represent the minimal sufficient statistics associated with the type of loglinear model under investigation. All members of $\mathcal{T}$ are subsets of the subscript set $\{0, 1, \dots, m\}$.

The general IPFP amounts to solving the equations

$$\hat{\xi}_T = W_T, \tag{14}$$

for all $T \in \mathcal{T}$. The algorithm can be defined concisely as follows. Let the members of $\mathcal{T}$ be denoted by $T_1, \ldots, T_t$, choose ξ_0 suitably (see later) and let n denote the number of cycles that have been carried out so far. Then

$$\xi_{nt+s}(j) = \xi_{nt+s-1}(j) W_{T_s}(j^{T_s}) / \xi^{T_s}_{nt+s-1}(j^{T_s}), \tag{15}$$

for all $j = (j_0, \ldots, j_m)$, for $s = 1, \ldots, t$ and for $n = 0, 1, \ldots$. The notations ξ^{T_s} and j^{T_s} indicate appropriate marginals from ξ and from j.

In terms of loglinear models, provided the W_T are consistent and provided ξ_0 is of the factorized form

$$\xi = \prod_{s=1}^{t} \rho_s, \tag{16}$$

then the IPFP converges to the maximum likelihood estimate of ξ which is of the form (16). In fact, a general version of Theorem 2 holds and from it we can deduce the following result, which is largely a generalization of Theorem 1.

Theorem 3:

(a) There is at least one design ξ with which $\{W_T: T \in \mathcal{T}\}$ are compatible if and only if the IPFP based on (15) converges.

(b) If the algorithm converges, then the resulting design $\hat{\xi}$ is the unique design satisfying (14) that is of the form (16).

(c) The design $\hat{\xi}$ in (b), if it exists, is the unique design ξ of the form (16) that maximizes

$$\sum_{T \in \mathcal{T}} \sum W_T \log \xi_T,$$

where the inner summation is over all values of subscripts within each T.

Acknowledgement: We would like to thank L. Rüschendorf for pointing out to us the work of Csiszár (1975). One of us (DMT) gratefully acknowledges support from a British Council Academic Travel Grant.

References

Bishop YMM, Fienberg SE, Holland PW (1975) Discrete multivariate analysis. MIT Press

Csiszár I (1975) *I*-divergence geometry of probability distributions and minimization problems. Ann Probab 3:146–158

Darroch JN (1962) Interaction in multi-factor contingency tables. J Roy Statist Soc Ser B 24: 251–263

Darroch JN, Ratcliff D (1972) Generalized iterative scaling for loglinear models. Ann Math Statist 43:1470–1480

Kellerer HG (1964) Verteilungsfunktionen mit gegebenen Marginalverteilungen. Z Wahrsch verw Gebiete 3:247–270

Krafft O (1978) Lineare statistische Modelle und optimale Versuchspläne. Vandenhoeck & Ruprecht, Göttingen

Oh HL, Scheuren FT (1983) Weighting adjustment for unit nonresponse. In: Madow WG, Olkin I, Rubin DB (eds) Incomplete data in sample survey, vol 2. Academic Press, New York, pp 143–184

Pukelsheim F (1983) Optimal designs for linear regression. In: Heiler S (ed) Recent trends in statistics. Allgemeines Statistisches Archiv, Sonderheft 21, pp 32–39

Pukelsheim F (1986) Approximate theory of multiway block designs. Canad J Statist 14, No 4