A General Criterion for Factorial Designs Under Model Uncertainty

Steven Gilmour
Queen Mary University of London
`http://www.maths.qmul.ac.uk/~sgg`
and
Pi-Wen Tsai
National Taiwan Normal University

Fall Technical Conference Indianapolis 8-9 October 2009

1 Case studies

Case Study: Plasma Etching
Experiment on a plasma etching process: 6 factors studied in an 18-run three-level orthogonal main effects plan [Logothetis (1990)].

"Levels for the factors were chosen by the engineers so that the possibility of existence of strong interaction effects was minimized. This was based on previous evidence."

Fearn (1992), Tsai, Gilmour and Mead (1996) and Cox and Reid (2000) reanalyzed the data and all found clear evidence of strong interactions, even though the design used had limited power to detect these.

Case Study: Fuel Tank Welding
Experiment to improve quality of welding two halves of fuel tank: 6 quantitative factors studied in a regular 16-run two-level fractional factorial, along with three specific interactions [Grove and Davis (1992)].

Data analysis showed that the second largest effect on the mean and the largest on the dispersion were from contrasts measuring interactions other than those expected.

Theme of this talk: we should use engineers' prior knowledge in designing factorial experiments, but also recognize its limits. Otherwise factorial designs "just tell us what we already know".

2 Multifactor designs

Classification of types of multifactor designs:

- regular/irregular fractional factorials;
- two-, three- or mixed-level designs;
- qualitative/quantitative factor levels;
- order of terms which can be estimated - e.g. up to two-factor interactions, second order polynomial, main effects only, or less (supersaturated designs).

Distinctions are defined mainly by the models expected to be useful and the resources available.
Two different approaches (and almost non-intersecting sets of researchers):

- optimal design - choose a design to minimize some function of the variance matrix under an assumed model;
- classical design - choose a design to ensure estimation of as many effects as possible, with as little nonorthogonality as possible.

Advantages / disadvantages of optimal design:

- Optimal design criteria have natural statistical interpretations, e.g. for a model $E(Y) = \beta_0 + \sum_{i=1}^{v} \beta_i x_i$, minimizing $\sum_{i=1}^{v} V(\hat{\beta}_i)$ is A_s-optimality.
- The same optimal design criteria can be used for the many different types of multifactor designs.

- Heavily depend on an assumed model - we should be able to change the model if it doesn't fit. Chemistry professor: "If your student's data don't fit your model, change the student."

Classical approach is "model-free", or more correctly allows consideration of a very large number of candidate models, all of which are submodels of a *maximal model*.

3 The Q_B criterion

We present a criterion which builds a bridge between the optimal and classical design approaches, having a simple statistical interpretation, being applicable to all different types of design and being "model-free".

Define a maximal model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{\beta}' = [\beta_0 \cdots \beta_v]$, which need not be estimable from the design.

Let a_{ij} be the elements of $\mathbf{X}'\mathbf{X}$.

Ignoring higher order terms in the diagonal expansion of $|\mathbf{X}'\mathbf{X}|$ and in the Taylor series expansion of the inverse, we get

$$V\left(\hat{\beta}_i\right) \approx \sum_{j=0}^{v} \frac{a_{ij}^2}{a_{ii}^2 a_{jj}}.$$

This can be a poor approximation, but is still useful for comparing designs. The weighted-average approximate A_s-efficiency over all candidate models is

$$Q_B = \sum_{i=1}^{v} \sum_{j=0}^{v} \frac{a_{ij}^2}{a_{ii}^2 a_{jj}} p_{ij},$$

where p_{ij} is the sum, over models which include both β_i and β_j, of the prior probabilities of a model being the best (in this class of models) [Tsai, Gilmour & Mead (2007)].

Typically we will declare, for example, π_1, the probability of a linear effect being in the best model, π_2, the probability of a quadratic effect, given the corresponding linear effect, and π_3, the probability of an interaction, given the marginal linear effects. Then we just have to count the models.

When all possible submodels have equal prior probability, Q_B reduces to the Q criterion of Tsai, Gilmour & Mead (2000).

Like the A_s criterion, Q_B is scale dependent.

For a factor with p levels we define $p-1$ scaled contrast functions from the full factorial design, e.g. for a three-level factor, $\sqrt{3/2}(-1, 0, 1)$ and $\sqrt{1/2}(1, -2, 1)$.

Interaction contrasts are defined by multiplying the corresponding main effects.

For a factor with qualitative levels, giving equal weight to all polynomial contrast functions is the same as giving equal weight to all pairwise comparisons of treatments.
That's it! All we need to use Q_B is:

- a maximal model (usually bigger than we expect to eventually fit);
- a prior probability of each model being the best (usually obtained from probabilities of each type of effect being in the best model);
- a method of searching for an optimal design (the same as in any other optimal design problem).

We might look for a Q_B-optimal design within a restricted class, e.g. level-balanced designs, or we might use an exchange algorithm.

Now things get technical, as we show how Q_B is related to other criteria, but this is not needed for practical use.

4 Relationship with classical criteria

Regular Two-Level Designs
Regular fractions can be assessed by the lengths of the *words* in their defining contrasts,
e.g. quarter replicates of 2^7 (factors labeled $A, B, \ldots$):

- $I \equiv ABC \equiv DEFG \equiv ABCDEFG$
- $I \equiv ABCD \equiv AEFG \equiv BCDEFG$
- $I \equiv ABCD \equiv ABEFG \equiv CDEFG$

have *word-length patterns*

- 3,4,7
- 4,4,6
- 4,5,5

The *resolution* is the length of the shortest word in the defining contrast [Box & Hunter (1961)]. Designs are chosen to have maximum resolution - here the last two designs have resolution-IV, the maximum possible.
The *word-length counts* are:

- $(0, 0, 1, 1, 0, 0, 1)$
- $(0, 0, 0, 2, 0, 1, 0)$
- $(0, 0, 0, 1, 2, 0, 0)$

The *aberration* of a resolution-R design is the number of words of length R [Fries & Hunter (1980)]. Maximum resolution designs are chosen to have minimum aberration. The last design is the minimum aberration design.

Resolution and aberration can be adapted to three-level and mixed-level designs.

Irregular Fractional Factorials
Different *generalized aberration* criteria have been defined for factors of different types and with different numbers of levels.

We can rewrite them by defining the *generalized word-length count (GWC).* In an m-factor design, the GWC for words with k factors, i_q factors appearing as qth order effects, is

$$b_k(i_1, \cdots, i_p) =$$
$$\sum_{\substack{\{R_1, \ldots, R_p :\\ |R_1| = i_1, \cdots, |R_p| = i_p\}}} \left\{ \sum_{g=1}^{N} \left(\prod_{\substack{\{h_1, \ldots, h_p :\\ h_1 \in R_1, \cdots, h_p \in R_p\}}} d_{1gh_1} \cdots d_{pgh_p} \right) \right\}^2 / N^2 ,$$

where the Rs are subsets of $\{1, \ldots, m\}$ and $\mathbf{D}_q$ contains columns of order q from the design matrix. This is based on a generalization of J-characteristics of Tang and Deng (1999).

$b_k(i_1, \cdots, i_p)$ is an overall measure of the aliasing of factorial effects of these orders.

First Order Models
To choose a supersaturated design, typically minimize $E(s^2)$, the sum of squared off-diagonal terms of $\mathbf{X}'\mathbf{X}$ for the main effects model [Booth & Cox (1962)].

$E(s^2)$ has also been used for irregular saturated main effects designs, e.g. 5 factors in 6 runs [Lin (1993)].

For level-balanced designs with two-level factors, it can be shown that Q minimizes $b_2(2)$, which is equivalent to minimizing $E\left(s^2\right)$.

Q is also defined when there is no level-balance.
The Q_B criterion is equal to

$$\xi_1 b_1(1) + 2\xi_2 b_2(2),$$

where ξ_1 and ξ_2 are the sums of prior probabilities, over models including a specific factor and over models including a specific pair of factors, respectively, of models being the best model.

Q_B allows some factors to be given more importance, e.g. if they are expected to have large effects.

Example: 5 two-level factors in 6 runs
First, restrict ourselves to level-balanced designs and obtain all possible designs.

The Q_B-optimal level-balanced design, which is also the p-efficient design obtained by Lin (1993), has $b_1(1) = 0$ and $b_2(2) = 1.11$.

Run	A	B	C	D	E
1	−	−	−	−	−
2	−	−	+	+	+
3	−	+	+	−	+
4	+	−	−	−	+
5	+	+	−	+	−
6	+	+	+	+	−

Alternatively, consider the 3/16 fractional replicate of John (1971), obtained by deleting two points from a regular 2^{5-2} design.

Run	A	B	C	D	E
1	−	+	−	+	+
2	−	−	+	−	+
3	+	+	−	−	−
4	+	−	+	+	−
5	−	+	+	−	−
6	−	−	−	+	−

Main effects of factors A and E are partially aliased with the intercept and $b_1(1) = 0.22$.
Each of the main effects of A and E is orthogonal to each of the main effects of B, C and D, the remaining pairs, {(A,E), (B,C), (B,D), (C,D)}, are partially aliased together, and $b_2(2) = 0.44$.

In terms of Q, this design is better than the best level-balanced design.

If $\xi_1 > 6.11\xi_2$, then Lin's design is optimal with respect to Q_B.

Whereas Lin contrasted his design with A-optimal designs, Q_B shows that there is a smooth transition.

First Order Models
Traditionally, extensions to three-level and mixed-level factors require new criteria, $Ave\,(\chi^2)$ and $E(f_{NOD})$ [Yamada & Lin (1999; 2002)].

For level-balanced designs with three-level qualitative factors, Q minimizes $b_2(2,0) + b_2(1,1) + b_2(0,2)$.

$Ave\,(\chi^2)$-optimal designs are Q-optimal only if they are orthogonal.

Second Order Models
For two-level factors, we consider only linear effects and (for level-balanced designs) the G_2 *generalized aberration* criterion sequentially minimizes $b_2(2)$, $b_3(3)$, $b_4(4)$, etc. [Tang & Deng (1999)].

In this case Q_B minimizes

$$\{2\xi_{20} + \xi_{21} + 2(m-2)\xi_{32}\}b_2(2) + 6\xi_{31}b_3(3) + 6\xi_{42}b_4(4),$$

where ξ_{ab} is the sum, over all models with a main effects and b two-factor interactions, of the prior probabilities of the model being best.

Hence, if $2\xi_{20} + \xi_{21} + 2(m-2)\xi_{32} \gg 6\xi_{31} \gg 6\xi_{42}$, the Q_B criterion converges to the G_2-aberration criterion.

For three-level qualitative factors (level-balanced orthogonal main-effects designs), the *generalized minimum aberration (GMA)* criterion sequentially minimizes $A_3 = b_3(3,0) + b_3(2,1) + b_3(1,2) + b_3(0,3)$,
$A_4 = b_4(4,0) + b_4(3,1) + b_4(2,2) + b_4(1,3) + b_4(0,4)$, etc. [Xu & Wu (2001); Ma & Fang (2001)].

Here Q_B minimizes $2\xi_3 A_3 + \xi_4 A_4$, where ξ_s is the sum of prior probabilities for models in which at least effects of particular sets of s factors are included.

Hence, if $2\xi_3 \gg \xi_4$, the Q_B criterion converges to the GMA criterion.

Polynomial Models

For three-level quantitative factors (in level-balanced orthogonal main-effects designs), the *β-aberration* criterion sequentially minimizes $b_3(3,0)$, $b_3(2,1) + b_4(4,0)$, etc. [Cheng & Ye (2004)].

In this case Q_B minimizes

$$6\xi_{301} b_3(3,0) + \{2\xi_{311} + \xi_{302}\} b_3(2,1) + 6\xi_{402} b_4(4,0),$$

where ξ_{abc} denotes the sum, over models which include at least linear effects of a factors, b out of the a factors' quadratic effects and c linear×linear interactions of these a factors, of the prior probabilities of a model being the best.

Hence, if $6\xi_{301} \gg \{2\xi_{311} + \xi_{302}\}$ and $\{2\xi_{311} + \xi_{302}\} = 6\xi_{402}$, the Q_B criterion converges to the β-aberration criterion.

Regular Fractional Factorials
Here the approximation to A_s-efficiency is perfect for any estimable model, so the Q_B criterion is an exact weighted average, over all models, of the A_s-criterion.

The generalized aberration criteria reduce to the aberration criterion.

Thus Q_B provides an interesting link between the classical and optimal design approaches.

5 Case studies revisited

π_1	π_2	π_3	D2	D25	$Df2.1$	$Df2.2$
0.9	0.1	0.5	0.8998	0.9064	**0.5500**	0.5672
0.9	0.1	0.8	0.9383	0.9604	0.6944	**0.6913**
0.9	0.2	0.5	0.9220	0.9324	**0.6478**	0.6653
0.9	0.2	0.8	0.9260	0.9501	0.7389	**0.7370**
0.9	0.3	0.2	0.6127	**0.6126**	0.6534	0.6742
0.9	0.5	0.2	**0.6850**	0.6887	1.0839	1.1047
0.9	0.5	0.5	**0.9425**	0.9615	1.0472	1.0647
0.8	0.3	0.2	0.5160	**0.5159**	0.5533	0.5681
0.8	0.3	0.5	0.7776	0.7881	**0.6533**	0.6661
0.8	0.5	0.5	**0.8117**	0.8273	0.9298	0.9425
0.7	0.2	0.5	0.6030	0.6083	**0.4446**	0.4535
0.7	0.2	0.8	0.7456	0.7630	0.5318	**0.5261**
0.7	0.4	0.1	0.3887	**0.3886**	0.5641	0.5701
1	0.4	0.4	**1.0355**	1.0496	1.0465	1.0767
1	1	0.2	**0.9713**	0.9868	3.2925	3.3195

6 Discussion

- This link between optimal and classical design criteria helps to unify the subject.
- The comparison with aberration-type criteria reflects favorably on Q_B.
- This makes the practical use of aberration-type criteria obsolete.
- However, the research in these areas is not obsolete; it just needs to be rewritten in terms of Q_B.
- Being able to fit many models is not the same as being able to discriminate between them - this might be a weakness of Q_B (and the classical criteria).
- However, since the "supremum" of the models can be fitted, this is indirectly taken into account.

References

Booth, K. H. V. and Cox, D. R. (1962) Some systematic supersaturated designs. *Technometrics*, **4**, 489–495.

Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} fractional factorial designs, Part I. *Technometrics*, **3**, 311-351.
Cheng, S. W. and Ye, K. Q. (2004). Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. *Annals of Statistics*, **32**, 2168–2185.
Cox, D. R. and Reid, N. (2000). *The Theory of the Design of Experiments.* London: Chapman & Hall.
Fearn, T. (1992). Box-Cox transformations and the Taguchi method: an alternative analysis of a Taguchi case study. *Applied Statistics*, **41**, 553–559.
Fries, A. and Hunter, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics*, **22**, 601–608.
Grove, D. M. and Davis, T. P. (1992). *Engineering, Quality and Experimental Design.* Harlow: Longman.
Lin, D. K. J. (1993). Another look at first-order saturated design: The p efficient designs. *Technometrics*, **35**, 284–292.
Logothetis, N., (1990). Box-Cox transformations and the Taguchi method. *Applied Statistics*, **39**, 31–38.
Ma, C. X. and Fang, K. T. (2001). A note on generalized aberration in factorial designs. *Metrika*, **53**, 85–93.
Tang, B. and Deng, L. Y. (1999). Minimum G_2-aberration for non-regular fractional factorial designs. *Annals of Statistics*, **27**, 1914–1926.
Tsai, P.-W. and Gilmour, S. G. (2010). A General Criterion for Factorial Designs Under Model Uncertainty. *Technometrics*, to appear.
Tsai, P.-W., Gilmour, S. G. and Mead, R. (1996). Letter to the Editors: An alternative analysis of Logothetis's plasma-etching data. *Applied Statistics*, **45**, 498–503.
Tsai, P.-W., Gilmour, S. G. and Mead, R. (2000). Projective three-level main effects designs robust to model uncertainty. *Biometrika*, **87**, 467–475.
Tsai, P. W., Gilmour, S. G. and Mead, R. (2007). Three-level main-effects designs exploiting prior information about model uncertainty. *Journal of Statistical Planning and Inference*, **137**, 619–627.
Xu, H. and Wu, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Annals of Statistics*, **29**, 1066–1077.
Yamada, S. and Lin, D. K. J. (1999). Three-level supersaturated designs. *Statistics and Probability Letters*, **45**, 31–39.
Yamada, S. and Lin, D. K. J. (2002). Construction of mixed-level supersaturated design. *Metrika*, **56**, 205–214.