Followup Experiments Via Semi-foldover

Xuan Lu

Department of Mathematical Science

Tsinghua University

Dennis Lin

Department of Supply Chain and Information Systems

Penn State University

Fractional factorial design (FFD)

2^{k-p} FFD's of resolution IV or III are frequently used in industrial experiments for run-size economy, when only main effects and 2fi's (two-factor interactions) are considered important, while three- or more-factor interactions are negligible.

- In a 2_{III}^{k-p} FFD, some main effects are aliased with 2fi's;
- In a 2_{IV}^{k-p} FFD, some 2fi's (two-factor interactions) are aliased.

Foldover of FFD

Foldover is a technique in augmenting 2^{k-p} FFD to de-alias confounded main effects and 2fi's.

Some important work on foldover:

Box *et al* (1978),

Wu and Hamada (2000),

Montgomery (2001),

Li and Lin (2003),

etc.

The shortcoming of foldover

Mee & Peralta (2000) pointed out:

For most commonly used 2_{IV}^{k-p} FFD's, a foldover plan can only offer no more than half of the degree of freedoms in de-aliasing 2fi's.

Example 1

A 2_{IV}^{6-2} FFD with defining relations I=1235=2346= 1456 has 7 alias relations:

12=35, 13=25, 15=23=46, 24=36, 26=34, 14=56, 16=45.

By folding the factors 5 and 6 (Fo=56), 9 2fi's are dealiased; but there are still three pairs are aliased:

15=46, 14=56, 16=45.

Hence, for the foldover plan, only 5 more 2fi's can be estimated by adding 16 runs.

Semifolding FFD (Mee & Peralta, 2000)

Semifolding FFD:

1) Take a foldover plan;

2) Take a subset of the foldover plan, which is a regular half.

A regular half of a foldover plan is consisted of the runs in it corresponding to "+" (or "−") of a “branching column”.

Example 2 In the unique 2_{IV}^{4-1} FFD, the foldover plan Fo=4:

	column						
run	1	2	12=−34	3	13=−24	23=−14	4
1	−	−	+	−	+	+	+
2	−	−	+	+	−	−	−
3	−	+	−	−	+	−	−
4	−	+	−	+	−	+	+
5	+	−	−	−	−	+	−
6	+	−	−	+	+	−	+
7	+	+	+	−	−	−	+
8	+	+	+	+	+	+	−

The red 4 rows is the subset with 1=−. There are totally 14 regular half subsets.

The gain and lost of semifolding

Mee and Peltra (2000) pointed out:

- For most FFD's of resolution IV and a foldover plan, a carefully selected semifolding plan can estimate as many 2fi's as the foldover plan does.
- Hence, by semifolding, one can get the same estimation capacity and more run-size economy, compared with foldover plan.

- The lost (of semifolding instead of foldover): In the later the estimated effects are independent and of high estimation efficiency, but in the former, they are correlated and lower estimation efficiency.

Example 2 (Continued)

For the 2_{IV}^{4-1} FFD with the foldover plan Fo=4, all the 4 main effects (ME's) and 6 2fi's can be estimated independently in the total of 16 runs.

By taking the subset ss=4+ (the 4 runs with column "4" entries being "+", i.e., the runs 1, 4, 6, and 7), all the ME's and 2fi's can also be estimated (with correlations 0 or $\pm\frac{1}{3}$) in only 12 runs.

Mee & Peralta's suggestion

If the consideration of run-size economy is more serious than the consideration of estimation independency and efficiency for the same estimation capacity, semifolding is recommended.

They gave a detailed discussion of semifolding strategies for 16-run FFD's of resolution IV.

However, they did not offer general rules in semifolding FFD's of resolution IV or III and different sizes.

The pitfall of multicollinearity in semifolding FFD's

In trying to find some general rules in semifolding FFD's, we found that there may be pitfalls in semifolding a FFD:

Among the effects to be estimated in a semifolding plan, there can be hiding multicollinearity between them!

By multi-collinearity we mean that some effects seem to be dealiased, but are in fact strictly linearly dependent.

If a group of effects are strictly linearly dependent, they are not estimable.

Example 3

A 2_{III}^{5-2} FFD with defining relations I=124=135=2345 has aliasing relations among the ME's and 2fi's:

(1=24=35), (2=14), (3=15), (4=12), (5=13), (23=45), (25=34)

In the foldover plan Fo=45, the defining relations become I=−124=−135=2345, and the aliasing relations are:

[1=−(24=35)], [2=−14], [3=−15], [−4=12], [−5=13], (23=45), (25=34)

In the whole design, all the MEs are separated from 2fi's.

Take a subset ss=1+ in the foldover plan Fo=45. Since we have I=B=1 in the subset where "B" represents the blocking effect, the effects are grouped as:

Group	Effects
1	{I=B=1=(24=35)}
2	{[2=−14]=[−4=12]}
3	{[3=−15]=[−5=13]}
4	(23=45)
5	(25=34)

Note: The effects in ( ) are always aliased; the effects in [ ] are aliased in the foldover; the effects in { } are aliased in the semifolding subset

It can be verified that in the total of 12 runs, the effects in each of Groups 1, 2, and 3 are strictly linearly dependent.

The N-S condition for multicollinearity

Theorem 1 For a group of inner-correlated effects in a semifolding plan, multicollinearity exists if and only if the group is formed by two pairs of effects, in each of which the two effects are aliased in the original design, and dealiased in the foldover plan.

Proof of the sufficiency

Let a, b, c, and d are four effects (ME or 2fi),

1) $a=b$ and $c=d$ in the original design,

2) $[a=-b]$ and $[c=-d]$ in the foldover plan, and

3) $\{[a=-b]=[c=-d]\}$ in the semifolding subset.

Then, in the whole design, $a-b-c+d=0$, i.e., the four effects are strictly linearly dependent.

Semifolding strategy (resolution IV)

In semifolding a 2_{IV}^{k-p} FFD, Mee & Peralta propose:

The semifolding subset is determined by a single factor, i.e., formed by the runs where a single factor at a fixed level ("+", or "–").

They, however, only considered the case of 16-run FFD's and did not explain the reason as clearly as we do.

We proved the following theorem.

Theorem 2 For a 2_{IV}^{k-p} FFD, taking a semifolding subset determined by a single factor branching column can avoid the occurrence of multicollinearity among the main effects and 2fi's.

Optimal single-semifold designs

We considered the problem of optimal single-semifold designs for FFD's of resolution IV.

Criterion:

- Take a optimal foldover plan in Li and Lin (2003);
- Select a subset determined by a single factor so that the whole design is D-optimal among all the possible choices.

For 16-run FFD's of resolution IV, optimal semifolding plans can be found in Mee & Peralta (2000).

We have found the optimal semifolding plan for each 32-run FFD of resolution IV.

Optimal single-semifold designs for 32-run FFD's of resolution IV

(The initial design coding is based on Chen, *et al* (1993).)

Initial design	Generators	Foldover design	Possible factors as Branching columns
7-2.1	6=1234,7=1245	6	1, 2, 4
7-2.2	6=123,7=145	67	2, 3, 4, 5, 6, 7
7-2.3	6=123,7=124	67	5
8-3.1	6=123,7=124,8=2345	678	5, 8
8-3.2	6=123,7=124,8=135	78	4, 5, 7, 8
8-3.3	6=123,7=124,8=125	67	2, 3, 4, 5, 6, 7, 8
8-3.4	6=123,7=124,8=134	6	5

9-4.1	6=2345,7=1345,8=1245, 9=1235	67	5
9-4.2	6=123,7=124,8=134, 9=2345	67	5, 9
9-4.3	6=123,7=124,8=135, 9=145	78	1, 2, 3, 4, 5, 6, 7, 8, 9
9-4.4	6=123,7=124,8=134, 9=125	89	8
9-4.5	6=123,7=124,8=134, 9=234	67	5
10-5.1	6=1234,7=1235,8=1245, 9=1345,10=2345	67	1, 2, 3, 4, 5, 6, 7, 8, 9 10
10-5.2	6=123,7=124,8=135, 9=145,10=12345	678	1, 2, 3, 4, 5, 6, 7, 8, 9 10
10-5.3	6=123,7=124,8=134,	89	1, 2, 3, 4, 5, 6, 7, 8, 9

	9=125,10=135		10
10-5.4	6=123,7=124,8=135,	8910	1, 2, 3, 4, 5, 6, 7, 8, 9
	9=234,10=125		10
11-6.1	6=123,7=124,8=134,	689	1, 2, 3, 4, 5, 6, 7, 8, 9
	9=125,10=135,11=145		10, 11
11-6.2	6=123,7=124,8=134,	8910	1, 2, 3, 4, 5, 6, 7, 8, 9
	9=234,10=125,11=135		10, 11

Example 4

- The FFD 7-2.1 has defining relations I=12346=12457=3567. There are three pairs aliased 2fi's: 35=67, 36=57, 37=56.

- Optimal foldover plans are Fo=6 or 7.

- Taking Fo=6, there are three pairs aliased 2fi's in the foldover design: [35=−67], [−36=57], [37=−56].

- Semifold designs produced from a single factor branching column are divided into two equivalence class: {3+, 5+, 6+, and 7+ (or −) } and {1+, 2+, and 4+ (or −)}.

- Taking the subset 6+ in the 1st class, the inner-correlated effect groups in the whole design are: {[35=−67]=−7}, {[36=−57]=3}, {37=−56}=−5}, {1=16}, {2=26}, and {4=46}. It has the determinant of the correlation matrix equal to $\frac{8^7}{27^5}$.

- Taking the subset 1+ in the 2nd class, the inner-correlated effect groups in the whole design are: {[35=−67]}, {[36=−57]}, {37=−56}}, and {1*x*=*x*}, *x*=2 to 6. It has the determinant of the correlation matrix equal to $\frac{8^9}{27^6}$.

- Since $\frac{8^9}{27^6}$ is larger than $\frac{8^7}{27^5}$, the subset 1+ is better than the subset 6+ based on D-optimality criterion, and hence, is optimal.

Optimal double-semifold designs

It is natural to consider such a question:

Instead of a single foldover design, if we use two semifold designs as a followup design for a FFD of resolution IV, whether the whole design can estimate more 2fi's?

The answer is Yes.

Thus further considerations are:

1. Select a double-semifold design (two semifold designs) for a FFD of resolution IV such that the whole design can estimate as many 2fi's as possible.

2. If there are more than one double-semifold designs satisfying condition (1), select the one with the highest D-efficiency.

A double-semifold design satisfying conditions (1) and (2) is called optimal double-semifold design.

Optimal double-semifold designs for 16- and 32-run FFD's of resolution IV

Initial design	1st semifold design	2nd semifold design	# of 2fi's estimated in the DS design	# of 2fi's estimated in a single foldover
6-2.1	fo=5, bc=1	fo=6, bc=2	All the 15	12
7-3.1	fo=5, bc=1	fo=6, bc=2	18	13
8-4.1	fo=56, bc=1	fo=7, bc=2	20	13
7-2.1				All
7-2.2				All
7-2.3	fo=6, bc=5	fo=7, bc=5	All the 21	18
8-3.1	fo=6, bc=5	fo=7, bc=8	All the 28	25
8-3.2	fo=78, bc=4	fo=6, bc=5	All the 28	25
8-3.3	fo=67, bc=1	fo=8, bc=2	25	22
8-3.4	fo=6, bc=5	fo=7, bc=1	25	20

9-4.1	fo=67, bc=5	fo=8, bc=1	33	30
9-4.2	fo=67, bc=5	fo=8, bc=9	33	28
9-4.3	fo=78, bc=1	fo=679, bc=2	All the 36	27
9-4.4	fo=89, bc=8	fo=6, bc=5	33	28
9-4.5	fo=67, bc=5	fo=8, bc=1	28	21
10-5.1	fo=67, bc=1	fo=8, bc=2	39	34
10-5.2	fo=678, bc=1	fo=69, bc=2	42	29
10-5.3	fo=89, bc=1	fo=7910, bc=4	42	29
10-5.4	fo=8910, bc=1	fo=6, bc=5	39	30
11-6.1	fo=689, bc=1	fo=10, bc=2	39	30
11-6.2	fo=7810, bc=1		39	30

References

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, Wiley, New York.

Chen, J., Sun. D. X., and Wu, C. F. J. (1993), "A Catalogue of Two-Level and Three-Level Fractional Factorial Designs with Small Runs," *International Statistical Review*, 61, 131-145.

Li, W. and Lin, D. K. J. (2003), "Optimal Foldover Plans for Two-Level Fractional Factorial Designs," *Technometrics*, 45, 142-149.

Mee, R. W. and Peralta, M. (2000), "Semifolding 2^{k-p} Designs," *Technometrics*,42, 122-134.

Montgomery, D. C. (2001), *Design and Analysis of Experiments*(5th ed), Wiley, New

York.

Wu, C. F. J. and Hamada, M. (2000), “Experiments: Planning, Analysis, and Parameter Design Optimization, Wiley, New York.

Thank You!