GENERATORS FOR NONREGULAR 2^{K-P} DESIGNS

2015 Q&P RESEARCH CONFERENCE

ROBERT MEE

REGULAR FRACTIONAL FACTORIAL DESIGNS

- Generators => Defining relation => Alias sets

- Example: a regular 2^{7-2}
  - Generators: F = CDE, G = ABCD
  - Defining relation: I = CDEF = ABCDG = ABEFG
  - Aliasing: CD = EF = ABG = ABCDEFG, etc.

- We use:
  - Generators for convenient construction
  - Defining relation for characterizing the design
  - Aliasing to guide what model one can estimate and what confusion exists among possible effects

NONREGULAR FRACTIONAL FACTORIAL DESIGNS

- These designs have orthogonal main effects, but correlations other than 0, 1, and -1 between main effects and some interactions.

- **Pro**: These designs are more flexible wrt:
  - Run size n (any multiple of 4)
  - Model estimation - since the designs often have less full aliasing among lower-order effects
- **Con**: Data from these designs are more difficult to analyze because:
  - The aliasing is complex
  - Variation in the response can be "explained" many ways.

GENERATORS FOR NONREGULAR FRACTIONS

- When the run size is a power of 2, one can create the design using generators.

- Johnson and Jones (QREI 2011) have presented generators for all 16-run regular <u>and</u> nonregular designs with 6, 7 or 8 factors

- The generators:
  - Are convenient for construction <u>and</u> comparison
  - Help us compute the generalized word length pattern
  - May provide insights about the aliasing structure of the design

EXAMPLE 1: MIN G-ABERRATION 2^{7-2}

- Begin with a full factorial in A, B, C, D, E
- Add two columns:

  F = 0.5E(AB + AC + BD – CD)

  G = ABCD
- Each generator creates a "word" or string of words that equal I:

  I = 0.5EF(AB + AC + BD – CD)

  I = ABCDG

EXAMPLE 1: MIN G-ABERRATION 2^{7-2}

- Including the products of these words produces a defining relation:

  $I = 0.5EF(AB + AC + BD - CD)$

  $= ABCDG$

  $= 0.5EFG(CD + BD + AC - AB)$

- A defining relation contains all the interaction columns that do not sum to 0
- <u>Extended word length pattern</u>:

  $f_{4.5} = 4;\ f_5 = 1;\ f_{5.5} = 4$

EXAMPLE 1 WITH NEW BASIC COLUMNS

- Create a 2^5 for {B, C, E, F, G} and add

  $A=0.5EF(B + C - BG + CG)$

  $D=0.5EF(B - C + BG + CG)$

- Now the defining relation is

$I = 0.5AEF(B + C - BG + CG)$

$= 0.5DEF(B - C + BG + CG)$

$= ABCDG$

since $(B + C - BG + CG)\times(B - C + BG + CG) = 4BCG$

- The same 9 columns appear, but the columns that sum to ±16 are arranged in different strings

ALIASING FOR EXAMPLE 1

- The first version of the defining relation

$$I = 0.5EF(AB + AC + BD - CD) = \ldots$$

shows the linear dependency

$$2EF = (AB + AC + BD - CD)$$

- The second version

$$I = 0.5AEF(B + C - BG + CG) = 0.5DEF(B - C + BG + CG)$$
$$= ABCDG$$

obscures this; it puts length 5 words in every string.

EXAMPLE 2: RESOLUTION III* 2^{7-2}

- Generators:
  F = ABCDE
  G = 0.5A(B + C + D – BCD)
- Defining Relation:
  I = ABCDEF
  = 0.5AG(B + C + D – BCD)
  = 0.5EFG(BC + BD + CD – 1)
  Extended word length pattern:
  $f_{3.5} = 4$; $f_{5.5} = 4$; $f_6 = 1$
- Aliasing (since ABCD=EF):
  G = 0.5(AB + AC + AD – EF)

EXAMPLE 3: FITTING THE 2FI MODEL

- The two-factor-interaction (2FI) model for k = 7 factors has 29 terms, including the intercept.
- All three 2^{7-2} designs shown have rank(X) = 28, i.e., one linear dependency.
- Consider the design with generators

  F = 0.5ABC(1 + D + E – DE)

  G = 0.5DE(A + C + AB – AC)
- The two generators create words of length ≥4.5

  $f_{4.5} = 3$; $f_{5.5} = 4$; $f_{6.5} = 1$
- Compute the generalized interaction using symbolic computation in matlab

MATLAB CODE

```
syms   A B C D E F G W1 W2 W12
W1 = 0.5*A*B*C*F*(1 + D + E - D*E)
W2 = 0.5*D*E*G*(A + C + A*B - B*C)
W12=expand(W1*W2);
W12=subs(W12,A^2,1);
W12=subs(W12,B^2,1);
W12=subs(W12,C^2,1);
W12=subs(W12,D^2,1);
W12=subs(W12,E^2,1)
```

GENERALIZED INTERACTION VIA MATLAB

W12 = (A*F*G)/4 - (C*F*G)/4
- (A*B*F*G)/4 - (B*C*F*G)/4 - (A*D*F*G)/4
- (A*E*F*G)/4 + (C*D*F*G)/4 + (C*E*F*G)/4
+ (A*B*D*F*G)/4 + (A*B*E*F*G)/4
+ (B*C*D*F*G)/4 + (B*C*E*F*G)/4
- (A*D*E*F*G)/4 + (C*D*E*F*G)/4
+ (A*B*D*E*F*G)/4 + (B*C*D*E*F*G)/4

Extended word length pattern from this design

- 8 columns sum to ±n/2; 16 columns sum to ±n/4
- $f_{3.75} = 2$; $f_{4.5} = 3$; $f_{4.75} = 6$; $f_{5.5} = 4$; $f_{5.75} = 6$; $f_{6.5} = 1$; $f_{6.75} = 2$

No basic column set creates a string without words of length 5+

GENERATORS PROVIDE:

- A compact means of producing the design
- A basis for understanding the design's aliasing structure

Three more questions to answer:

1) What linear combinations can serve as generators?
2) Do all nonregular 2^{k-p} designs have generators and a defining relation?
3) How can we determine all the relevant singularities from the generators?

Q1: WHAT LINEAR COMBINATIONS ARE ELIGIBLE GENERATORS?

- This question has been answered by complete enumeration of solutions to the equation **Gc**=**e**, where:
  - **G** is an nxp matrix of all interaction contrasts,
  - **c** is a px1 vector of coefficients defining a generator
  - **e** is any nx1 vector of ±1 that sums to zero
- For n = 16, all generators involving either 1 or 4 interactions
- For n = 32, there are six types of generators:
  - A single ±1
  - Four ±0.5
  - Seven ±0.25 and one ±0.75
  - Eight ±0.25 and two ±0.5
  - Twelve ±0.25 and one ±0.5
  - Sixteen ±0.25
- The ones in red always include at least one 2fi and so create strength 2 arrays.

Q2: DO ALL NONREGULAR 2^{K-P} DESIGNS HAVE GENERATORS?

Three Types of 2^{k-p} Orthogonal Arrays

1. There exists at least one subset of b factors forming a full, unreplicated 2^b factorial
   - Type 1 designs have generators
2. There is partial replication, so we don't have 2^b distinct treatment combinations even in the full set of k factors
3. No partial replication but also no projection forms a full 2^b

Follow-up question:

- Can we still find generators for Types 2 and 3?

TYPE 2: 2^{K-P} DESIGNS WITH PARTIAL REPLICATION

- **Good News**:
  - Partial replication provides a pure error estimate of the error variance.
  - For many cases, this improves the power of tests more than having additional treatment combinations

- **Bad News**:
  - Orthogonal 2^{k-p} designs with partial replication have a generalized word length that sums to more than 2^{p}-1 (Xu and Wu 2001)

GENERALIZED WORD LENGTH PATTERN FOR 2^{7-2} EXAMPLES

- Minimum Aberration Design
  - I = CDEF = ABCDG = ABEFG
  - wlp = (0,1,2,0,0)
- Example 1: Mininum G-Aberration Design
  - ewlp: $f_{4.5} = 4$; $f_5 = 1$; $f_{5.5} = 4$
  - gwlp: $A_3 = 0$; $A_4 = 4(0.5)^2 = 1$; $A_5 = 1(1.0)^2 + 4(0.5)^2 = 2$
- Example 2: Resolution III* nonregular design
  - gwlp: (1,0,1,1,0)
- Example 3:
  - ewlp: $f_{3.75} = 2$; $f_{4.5} = 3$; $f_{4.75} = 6$; $f_{5.5} = 4$; $f_{5.75} = 6$; $f_{6.5} = 1$; $f_{6.75} = 2$
  - gwlp: (0.125, 1.125, 1.375, 0.375, 0)
- All these have no replication and so sum(gwlp) = 3

2^{7-2} EXAMPLE WITH PARTIAL REPL.

- 30 distinct treatment combinations
  - 28 appear once, 2 appear twice
- gwlp sums to $2^p[28(1)^2 + 2(2)^2]/n - 1 = 3.5$
  - 48 interaction columns sum to ±n/4, 2 sum to ±n/2
  - gwlp = (13/16, 22/16, 12/16 + 1/4,1/4, 1/16)

- How to construct generators for this example with maximum frequency = 2:
  - Find b-1 = 4 factors that form a strength 2 array
  - Add a 5th factor that sums to zero for each pair, which is orthogonal to the 7 factors of the design
  - The result is a 2^{8-3} design with a 2^5 projection

2^{7-2} EXAMPLE WITH PARTIAL REPL.

- {B, E, F, G} form an equally-replicated 2^4
- There are 2^{16} possible vectors for the 8th factor N; enumerate vectors until one is found that is orthogonal to A – G
- Resulting generators:
  - A and C have generators with 12 ±1 and one ±2
  - D has a generator with 16 ±1

$$
\begin{aligned}
A &= 0.25(BEG + EFG - BG - FG) \\
&\quad +0.25N(BEG + BFG + 2BE + EG + FG - BF - EF + B - E), \\
C &= 0.25(2BEFG + BEF + EFG + BE + BF - BG - EF) \\
&\quad +0.25N(BEFG - BEG - BFG - EFG + EF + B), \\
D &= 0.25(BEFG + BEF + BFG + EFG - BEG + BE + EF - BF) \\
&\quad +0.25N(BEG + EFG - BEF - BF - EG - B + F + G).
\end{aligned}
$$

2^{7-2} EXAMPLE WITH PARTIAL REPL.

- 2^{8-3} design has a defining relation with 7 strings
- If we drop interactions involving N from each string:
  - We no longer have a defining relation
  - We obtain the columns of the indicator function
- Resulting indicator function:

$$
\begin{aligned}
&0.25AG(BE+EF-B-F)\\
&+0.25C(2BEFG+BE+BF+BEF+EFG-BG-EF)\\
&+0.25AC(-BE+BF-EF+FG-BEF+BEG)\\
&+0.25D(BE+EF+BEF+BFG+EFG+BEFG-BF-BEG)\\
&+0.25AD(G+BE+BF-EF-EG+BEF-BEG-EFG)\\
&+0.25CD(B+E+G-BF+BG-EF+FG+BFG)\\
&+0.25ACD(E-B-BF-BG+EF+EG+FG-2EFG+BEFG).
\end{aligned}
$$

TYPE 2: 2^{K-P} WITH PARTIAL REPL.

- If the maximum frequency is 3 or 4, we would need to find b-2 factors that project to an equally replicated 2^{b-2} factorial.
- Then we would need to add two factors N_1 and N_2 to create a $2^{(k+2)-(p+2)}$ Type 1 array.
- Gladly, the best designs with partial replication have frequencies of just 0 and 1, where we only need to find N to create a $2^{(k+1)-(p+1)}$ Type 1 array.

A THIRD CASE

- Consider Paley's Hadamard design of order 32
  - N=32, k=31
- This 2^{31-26} design is not type 1, since no 5-factor projection has more than 28 runs.
- This 2^{31-26} design is not type 2, since the 32 rows are distinct
- It is impossible to add a 32nd factor orthogonal to the original 31 factors.

GENERATORS FOR TYPE 3 DESIGNS

- Find b-1 factors that form an equally replicated 2^{b-1} factorial
- Create a new column that has:
  - Correlation of 0 with the b-1 basic factors
  - Correlation of 0, ±0.25, ±0.5, or ±0.75 with the remaining columns.
- The new column plus the original b-1 serve as basic factors
- The remaining k-b+1 factors can be defined in terms of the full factorial model of these b columns
  - The new column must be correlated with some of the factors.
  - Illustrate the computations in JMP

2^{7-3} EXAMPLE FROM DAVID'S TALK

- A-D can't serve as basic factors: only 12 distinct rows
- A, C, D, F do form a 2^4

Term	B	E	G
A*C	-0.5	-0.5	0
A*C*D	0	0.5	-0.5
A*C*D*F	0	-0.5	-0.5
A*C*F	0.5	-0.5	0
A*D	-0.5	0	-0.5
A*D*F	-0.5	0	0.5

Defining relation

- I = -ABEFG
- = 0.5ADG(F – C – 1 – CF) = 0.5BDE(C – 1 + F + CF)
- = 0.5ACE(D – 1 – F – DF) = 0.5BCG(1 + D + F – DF)
- = 0.5EG(D – C + CF + DF) = 0.5AB(CF – D – C – DF)

HOW TO FIND LINEAR DEPENDENCIES?

- This last question is not yet answered!
- For example, the 2^{7-3} design has the linear dependencies
  - G = - A*F + B*C - C*E
  - D = - A*G - B*E + D*F
- How can these be recognized from the defining relation?

I = -ABEFG

= 0.5ADG(F – C – 1 – CF) = 0.5BDE(C – 1 + F + CF)

= 0.5ACE(D – 1 – F – DF) = 0.5BCG(1 + D + F – DF)

= 0.5EG(D – C + CF + DF) = 0.5AB(CF – D – C – DF)

2^{7-2} EXAMPLE FROM DAVID'S TALK

Generators, with {A, C, D, F, G} basic:

- B=0.5AD(C + F + FG – CG)
- E = 0.5CF(A + D + DG – AG)

Defining relation

- I = 0.5ABD(C+F-CG+FG)
- = 0.5CEF(A+D-AG+DG)
- = 0.5BE(AC+DF+ACG-DFG)

Singularity Details

- A*D = - A*E + B*C + B*F + C*F - D*E
- A*C = - A*F + B*D + B*E + C*E - D*F
- A*B = - A*F + B*D + C*D + C*E - E*F

All 3 strings of the defining relation involve G, but singularities don't! Can I find these linear dependencies in the defining relation? Help!

CONCLUSION

- Progress has been made in understanding the structure of nonregular 2^{k-p}
- As n increases, the number of designs explodes
  - There are only 5 Hadamard matrices of order 16
  - There are over 13 million Hadamard matrices of order 32!
- Since the number of non-isomorphic designs is so numerous at n = 32, understanding the structure will be a help in the search for best designs.
- Having generators for designs should help us determine all the aliasing, but this is a work in progress.