Complex Aliasing

from alias matrix $(X_1^T X_1)^{-1} X_1^T X_2$, may not be consistent with $X_1^T X_2$

one in X_1, the other in X_2

- Two effects are said to be **partially aliased** if the absolute value of their aliasing coefficient is strictly between 0 and 1.
- The **complex aliasing** refers to the enormous number of effects in $\mathbf{X}_2$ (e.g., 2-factor interactions) that are partially aliased with the effects in $\mathbf{X}_1$ (e.g., main effects).
- By comparison, for 2^{k-p} or 3^{k-p} regular designs,
  - the factorial effects are either orthogonal (i.e., uncorrelated) or fully aliased (i.e., with aliasing coefficient 1 or -1) with each other, and
  - the number of (full) aliases is much smaller.

2^{k-p}, each alias set contain 2^p effects

only effects in same alias set are fully aliased.

❖ **Reading**: textbook, 9.1

Traditional Analysis Strategy

∵ cannot fit $[X_1\ X_2]$ simultaneously.

- For designs with complex aliasing, it is difficult to disentangle the large number of aliased effects and to interpret their significance.
- For this reason, such designs were traditionally used for *screening factor effects* only, i.e., to estimate factor main effects but not their interactions.
- These designs are therefore referred to as **screening designs**.

purpose: identify important "factors" using ME-only model.

- The validity of the main effect estimates for screening designs depends on the assumption that the interaction effects are negligible, which is often questionable in many practical situations.
- It suggests the need for strategies that allow interactions to be entertained.

❖ **Reading**: textbook, 9.2

Simplification of Complex Aliasing

- The complex aliasing pattern can be greatly simplified under the assumption of *effect sparsity*.
  - Effect sparsity: number of relatively important effects (e.g., main effects and two-factor interactions) is small. larger contribution to R^2
  - By employing the effect sparsity and effect hierarchy principles, one can assume that only few main effects and even fewer two-factor interactions are relatively important.
- With this assumption, the complex aliasing pattern of a design can be greatly simplified. This simplification and the partial aliasing of effects make it possible to estimate some interactions.

Note: all the effect vectors are different in LNp.4-26 example. partial aliasing can be an advantage under the situation

Simplification of Complex Aliasing

- Consider the hypothetical example.
  - Suppose that in the cast fatigue experiment only $\underbrace{A, F, G}_{x_1}$, $\underbrace{AB, \text{ and } FG}_{x_2}$ are significant. i.e. the β's of other effects are zero. ∵ effect sparsity
  - If the seven main effect estimates were obtained under the main-effect-only model, their expectations in LNp.4-26 are simplified as follows:

$$\begin{bmatrix} \hat\beta_A \\ \hat\beta_B \\ \vdots \\ \hat\beta_G \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & -\frac{1}{3} \\ 0 & 0 & 0 & -\frac{1}{3} & \frac{1}{3} \\ & & \vdots & & \\ 0 & 0 & 1 & -\frac{1}{3} & 0 \end{bmatrix} \begin{bmatrix} \beta_A \\ \beta_F \\ \beta_G \\ \beta_{AB} \\ \beta_{FG} \end{bmatrix}$$

$$\begin{aligned}
E(\hat{A}) &= A - \tfrac{1}{3}FG \\
E(\hat{B}) &= -\tfrac{1}{3}FG \\
E(\hat{C}) &= -\tfrac{1}{3}AB + \tfrac{1}{3}FG \\
E(\hat{D}) &= -\tfrac{1}{3}AB + \tfrac{1}{3}FG \\
E(\hat{E}) &= -\tfrac{1}{3}AB - \tfrac{1}{3}FG \\
E(\hat{F}) &= F + \tfrac{1}{3}AB \\
E(\hat{G}) &= G - \tfrac{1}{3}AB
\end{aligned} \tag{1}$$

Simplification of Complex Aliasing

- Effect sparsity reduces the 16 terms in the original equation to one or two terms on the right side of equation (1).
- The model including A, F, G, AB, and FG is identifiable since the corresponding model matrix has full rank. need to put these effects in X_1
- When the number of significant main effects and interactions is small, the selected main effects and interactions are estimable. Note: design with complex aliasing has more estimable models
- In contrast with designs having complex aliasing, the 2^{k-p} designs have straightforward aliasing patterns; i.e., any two effects are either orthogonal or completely aliased.
  - If an 8-run 2^{7-4}_{III} design had been used instead with $A = FG$ as one of its defining relations, it's impossible to disentangle FG from its alias A.
  - The problem may not disappear by doubling the run size from 8 to 16. if the resolution is still III or IV
- This new perspective of designs with complex aliasing allows their known deficiency to be turned into an advantage. But → model selection becomes important

❖ **Reading**: textbook, 9.3

overfitting → p large, RSS small
underfitting → p small, RSS large
model complexity
fitting model

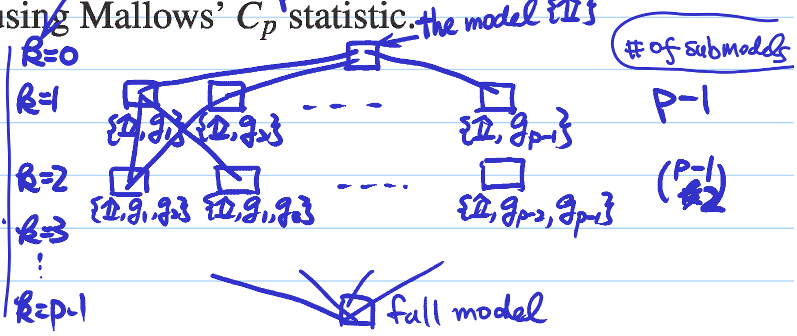
Variable Selection Methods

- **Principle of Parsimony (Occam's razor):** Choose fewer variables with sufficient explanatory power. This is a desirable modeling strategy.
- The goal is thus to identify the smallest subset of covariates that provides good fit. One way of achieving this is to retain the significant predictors in the fitted multiple regression. This may not work well if some variables are strongly correlated among themselves or if there are too many variables (e.g., exceeding the sample size).

computationally more intensive

- Two other possible strategies are
  - Best subset regression using Mallows' C_p statistic. AIC, BIC. # of effects in the submodels
  - Stepwise regression.

full model: $\{1, g_1, \cdots, g_{p-1}\}$.
$\{1\} \subset$ any submodel $\subset$ full model.
* full model may be unidentifiable.
# of submodels $= 2^{p-1}$



Best Subset Regression

sub # of parameters in the sub model.

- For a sub model with p regression coefficients, (i.e., $p-1$ covariates plus the intercept β_0), define its C_p value as

fitting model complexity

$$C_p = \frac{RSS}{s^2} - (N - 2p),$$

$\hat{\sigma}^2$ under full model.

where RSS = residual sum of squares for the given model, s^2 = mean square error = $\frac{RSS \text{ (for the complete model)}}{\text{df (for the complete model)}}$, N = number of observations.

- If the model is true, then $E(C_p) \approx p$. Thus one should choose p by picking models whose C_p values are low and close to p. For the same p, *choose a model with the smallest C_p value* (i.e., the smallest RSS value).

AIC and BIC Information Criteria

- The Akaike information criterion (AIC) is defined by

fitting $e^2 = 7.38$ model complexity

$$AIC = N ln(\frac{RSS}{N}) + 2p$$

- The Bayes information criterion (BIC) is defined by

different weights

fitting

$$BIC = N ln(\frac{RSS}{N}) + p\,ln(N)$$

model complexity

- In choosing a model with the AIC/ BIC criterion, we choose the model that $\boxed{\text{minimizes}}$ the criterion value.

But, only compare p for unidentifiable models.

- Unlike the C_p criterion, the AIC criterion is applicable even if the number of observations do not allow the complete model to be fitted.
- The BIC criterion favors smaller models more than the AIC criterion.

Stepwise Regression

Alternative, use AIC or BIC

- This method involves adding or dropping one variable at a time from a given model based on a *partial F*-statistic. ① Model I nested Model II ② Model I obtained by dropping one effect from model II

  Let the smaller and bigger models be Model I and Model II, respectively. The partial F-statistic is defined as

$$\frac{RSS(Model\ I) - RSS(Model\ II)/1}{RSS(Model\ II)/\nu}$$

$\leftarrow \hat{\sigma}^2$ under Model II.

  where ν is the degrees of freedom of the RSS (residual sum of squares) for Model II.

- There are three possible ways
  1. **Backward elimination :** starting with the full model and removing covariates.
  2. **Forward selection :** starting with the intercept and adding one variable at a time.
  3. **Stepwise selection :** alternate backward elimination and forward selection.

  Usually stepwise selection is recommended.

Criteria for Inclusion and Exclusion of Variables

critical value

- F**-to-remove :** At each step of backward elimination, compute the partial F value for each covariate being considered for removal. The one with the lowest partial F, provided it is smaller than a preselected value, is dropped. The procedure continues until no more covariates can be dropped. The preselected value is often chosen to be $F_{1,\nu,\alpha}$, the upper α critical value of the F distribution with 1 and ν degrees of freedom. Typical α values range from 0.1 to 0.2. ← multiple testing → choose a larger α than 0.05.
- F**-to-enter :** At each step of forward selection, the covariate with the largest partial F is added, provided it is larger than a preselected F critical value, which is referred to as an F*-to-enter* value.
- For stepwise selection, the F-to-remove and F-to-enter values should be chosen to be the same.

❖ **Reading**: textbook, 1.7

Frequentist Analysis Strategy for Complex Aliasing

$[X_1 \,.\, X_2]$ ← # of columns > # of rows (# of effects > # of runs)

- The analysis of experiments with complex aliasing can be viewed as a *variable selection problem*. But, ∵ cannot fit $\hat{\sigma}^2$ for full model.
  - backward selection and C_p criterion cannot be used here because the full model is not estimable
  - all-subsets regression may be computationally infeasible because of the large number of effects

    ① $OA(12, 2^{11})$ ② → ME, 11; 2fi, $\binom{11}{2}=55$] 66.

    all submodels 2^{66}.
- The frequentist analysis strategy is based on:
  - Effect sparsity principle → prefer model with smaller R (LN p. 4-32)
  - Effect heredity principle → apply on submodels

$y \sim A+B+AB$

$y \sim A+B+AC$

$y \sim A+B+CD$ ← lesser prefer