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.
- When the number of significant main effects and interactions is small, the selected main effects and interactions are estimable.
- 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.
- This new perspective of designs with complex aliasing allows their known deficiency to be turned into an advantage.

❖ **Reading**: textbook, 9.3

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).
- Two other possible strategies are
  - Best subset regression using Mallows' C_p statistic.
  - Stepwise regression.

Best Subset Regression

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

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

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

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

- The Bayes information criterion (BIC) is defined by

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

- In choosing a model with the AIC/ BIC criterion, we choose the model that minimizes the criterion value.
- 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

- This method involves adding or dropping one variable at a time from a given model based on a *partial F*-statistic.

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

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

  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

- ***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.

- ***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

- The analysis of experiments with complex aliasing can be viewed as a *variable selection problem*. But,
  - 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
- The frequentist analysis strategy is based on:
  - Effect sparsity principle
  - Effect heredity principle

Frequentist Analysis Strategy for Complex Aliasing

- Method I

  Step 1.
  - For each factor X, entertain X and all its two-factor interactions XY with other factors.
  - Use a stepwise regression procedure to identify significant effects from the candidate variables and denote the selected model by M_X.
  - Repeat this for each of the factors and then choose the best model.
  - Go to Step 2.

  Step 2.
  - Use a stepwise regression procedure to identify significant effects among the effects identified in the previous step as well as all the main effects.
  - Go to Step 3.

Frequentist Analysis Strategy for Complex Aliasing

- Method I (cont.)

  Step 3.

  - Using effect heredity, entertain
    - (i) the effects identified in Step 2 and
    - (ii) the two-factor interactions that have at least one component factor appearing among the main effects in (i).
    - (iii) interactions suggested by the experimenter.
  - Use a stepwise regression procedure to identify significant effects among the effects in (i)-(iii).
  - Go to Step 2.

  Step 4. Iterate between Steps 2 and 3 until the selected model stops changing.

Frequentist Analysis Strategy for Complex Aliasing

- Some properties of Method I
  - Since the model search in the procedure is guided by effect heredity, the problem of obtaining uninterpretable models is lessened.
  - However, this problem cannot be completely avoided because effect heredity is not enforced throughout the procedure.
  - Effect sparsity suggests that only a few iterations will be required.
  - If all two factor interactions are entertained indiscriminately in Step 3, it is possible to get a good fitting model consisting only of interaction terms and no main effects; hence, nonsensical models may be obtained without assuming effect heredity.
  - Step 2 is motivated by the possibility of missing main effects in Step 1 because of the existence of interactions and complex aliasing.
  - If a more extensive search of models is desired, the final model obtained by Method I can be viewed as a good starting model.

Frequentist Analysis Strategy for Complex Aliasing

- Method II
  - The iterative search in Method I can be easily implemented computationally but does not provide a very extensive search for models.
  - Suppose effect sparsity suggests that no meaningful model can have more than h effects.
    - ∗ Box and Meyer (1986) have found that the proportion of active effects is typically between 0.13 and 0.27 of the design's run size, so that a choice for h about 0.30 of the run size seems reasonable.
  - Search procedure:
    - ∗ Search over all models that have no more than h effects (plus an intercept term) and satisfy the effect heredity requirement.
    - ∗ Choose the best model (or models) according to a sensible model selection criterion (e.g., the C_p or AIC, BIC criteria).

Frequentist Analysis Strategy for Complex Aliasing

- Some properties of Method II
  - Method II provides a more extensive search than Method I if h is not smaller than the number of effects in the model selected by Method I.
  - Rather than relying entirely on a model selection criterion, the analyst might inspect the best fitting models with $1,\ldots,h$ effect(s), respectively, to identify which terms are the most important and to assess the gains achieved as larger models are considered
  - When there are quite a few good fitting models for a given model size, this would suggest that the experimental data do not contain sufficient information to distinguish between them.
    - ∗ Methods I and II work well when there are only a few significant interactions that are partially aliased with the main effects. Otherwise, several incompatiable models may be identified (see an example in Section 9.4.1).