

Theory of Screening Procedures to Identify Robust Product Designs Using Fractional Factorial Experiments

Guohua Pan
Biostatistics and Statistical Reporting
Novartis Pharmaceuticals Corporation
East Hanover, NJ 07936

Thomas J. Santner
Department of Statistics
Ohio State University
Columbus, OH 43210

Abstract

In many quality improvement experiments, there are one or more “control factors” that can be used to determine a final product design (or manufacturing process), and one or more “environmental factors” that vary under field (or manufacturing) conditions. This paper considers applications where the product design or process design is considered to be seriously flawed if its performance is inferior at any level of the environmental factor; for example, if a particular prosthetic heart valve design has poor fluid flow characteristics at certain pulse rates, then a manufacturer will not want to put this design into production. In such a situation, the appropriate measure of a product’s quality is its *worst* performance over the levels of the environmental factor. This paper develops theory for a class of subset selection procedures that identify product designs which maximize the worst-case performance over the environmental conditions for general Combined array experiments.

Keywords and phrases: Combined-array, Inner array, Minimax approach, Outer array, Product-array, Quality improvement, Response model, Robust process design, Screening, Simulation, Subset selection, Variance reduction.

1 Introduction

In some quality improvement experiments, two types of factors affect the performance of a product or process design. The first type are “control factors” which are variables that can be manipulated by the manufacturer. Control factors are sometimes called “engineering factors” or, in process improvement applications, “manufacturing factors”. The second type are “noise factors” which are variables that represent either different environmental conditions or, in process improvement applications, (uncontrollable) variability in component parts or raw materials. Because a process can be viewed as the process designer’s “product,” hereafter we refer only to the product design problem.

In some applications, the quality of a product design is measured by its worst performance under the different environments. This criterion is natural in situations where a low response at *any* level of the noise factors can have potentially serious consequences. Engineering designs for auto tires, seat belts, or heart valves that fail catastrophically under rare, though non-negligible, sets of operating conditions must be identified early in the product design cycle. The effect of the control and noise factors on the performance of a product design is commonly investigated at the product development stage using fractional factorial (combined-array) experiments involving both control and noise factors. Based on the data from such a (fractional factorial) experiment, the product developer wishes to screen out undesirable product designs whose worst-case performances are inferior. This screening goal is equivalent to selecting product designs whose worst-case performance is among the best. We term this selection goal the maximin criterion. Indeed, the best product design, as identified by the maximin criterion, can be viewed as a “robust” design in that it uses the larger-the-better approach advocated by Taguchi.

The subset selection methodology of Gupta (1956, 1965) was a screening methodology for balanced one-way layouts. In the simplest case, a screening procedure identifies a set of treatments (or treatment combinations) containing the treatment (treatment combination) with the greatest mean response based on a one-factor (multi-factor) experiment. Bechhofer, Santner, and Goldsman (1995) provide screening procedures for many goals/models in physical and simulation experiments. This paper extends the Gupta methodology to the area of quality improvement by proposing a method of selecting a subset of treatments containing the product design having the best performance, in the maximin sense described above, based on a fractional experiment. Previously, Pan and Santner (1998) considered the maximin criterion for complete experiments conducted under a variety of randomization restrictions. Santner and Pan (1997) discussed a pilot case study involving a 2^{5-1} experiment with three control factors and two noise factors. This paper provides a general screening theory based on the maximin criterion for *arbitrary*

combined-array experiments and any analysis-of-variance mean models.

Section 2 presents the basic model, screening goal, confidence requirement, and proposed subset selection procedure. Section 3 discusses the calculation of the critical values required to implement the proposed procedure using the “least favorable configuration” method. In those cases where the least favorable configuration method is not applicable, Section 4 provides an alternative method for determining conservative critical values based on the calculation of a certain lower bound for the confidence requirement. The section also provides a method for checking whether the lower bound used by this method is sharp (and hence the corresponding critical value is sharp). Several examples are given throughout the paper to illustrate the proposed methods. The proofs of all results are given in the Appendix.

2 Model, Goal, and Procedure

Suppose that the *mean* of a response that characterizes the output quality depends on $p + q$ control factors and $r + s$ noise factors. In our notation, p and r denote the number of control and noise factors, respectively, that interact with each other in the model for the mean, while q is the number of control factors having *no interactions* with noise factors, and s denotes the number of noise factors having *no interactions* with control factors. Of special practical importance is the frequently occurring case when all factors are at two levels; however, nothing in the development below requires this assumption. We introduce the following notation to distinguish these two types of control and noise factors.

Notation	Definition
$C_1^{\mathcal{I}}, \dots, C_p^{\mathcal{I}}$	Control Factors that <i>interact</i> with Noise Factors
$C_1^{\sim}, \dots, C_q^{\sim}$	Control Factors that do <i>not interact</i> with Noise Factors
$N_1^{\mathcal{I}}, \dots, N_r^{\mathcal{I}}$	Noise Factors that <i>interact</i> with Control Factors
$N_1^{\sim}, \dots, N_s^{\sim}$	Noise Factors that do <i>not interact</i> with Control Factors

The superscripts \mathcal{I} and \sim are mnemonic; the \mathcal{I} in $C_i^{\mathcal{I}}$ indicates that this control factor *interacts* with one or more noise factors while the \sim symbol is meant to suggest the negation, i.e., C_i^{\sim} does *not interact* with any noise factors. A parallel interpretation holds for the notations $N_i^{\mathcal{I}}$ and N_i^{\sim} .

Example 1 Box and Jones (1992) discuss a taste-testing experiment that is typical of those used in the food industry to evaluate recipes. The experiment involves five factors that (potentially) affect the taste of a cake recipe; each factor is used at two levels. Three of the factors (S = shortening, F = flour and E = egg powder) are “control factors” that are varied by the cake

manufacturer. The remaining two factors (T = baking temperature and Z = baking time) are noise factors because they represent deviations from the cake baking directions; the temperature controls in ovens can be biased and consumers both under and overbake cake mixes.

As an illustration of the general notation that is introduced below, let $\mu_{sfe,tz}$ denote the mean taste level when S , E , F , T , and Z are at levels s , f , e , t , and z , respectively, where each lower case index is either 0 or 1. The factor level is set equal to zero (unity) when the factor is at its low (high) level. *Suppose* the model

$$\mu_{sfe,tz} = m_0 + S_s + F_f + E_e + T_t + Z_z + (ST)_{st} \quad (2.1)$$

holds for all $2^5 = 32$ (sfe, tz) factor combinations. A comma is used to separate the list of indices for control factors (sfe) from the indices for noise factors (tz). As usual, the terms S_s , F_f , E_e , T_t , and Z_z are the shortening, flour, egg, temperature, and time main effects, respectively. For Model (2.1), there is one control factor, S , that interacts with one noise factor, T ; thus $p = 1 = r$, $C_1^{\mathcal{I}} = S$, and $N_1^{\mathcal{I}} = T$. The two control factors F and E do *not* interact with any noise factors while the single noise factor Z does *not* interact with any control factors; thus $q = 2$, $s = 1$, $C_1^{\sim} = F$, $C_2^{\sim} = E$, and $N_1^{\sim} = Z$. \square

Returning to the discussion of the general case, let $\mathbf{i}^{\mathcal{I}} = (i_1^{\mathcal{I}}, \dots, i_p^{\mathcal{I}})$ denote the $1 \times p$ vector of indices of the levels of those control factors each of which *interacts* with one or more noise factors, $\mathbf{j}^{\mathcal{I}} = (j_1^{\mathcal{I}}, \dots, j_r^{\mathcal{I}})$ denote the $1 \times r$ vector of indices of the levels of those noise factors each of which *interacts* with one or more control factors, $\mathbf{i}^{\sim} = (i_1^{\sim}, \dots, i_q^{\sim})$ denote the $1 \times q$ vector of indices of the levels of the non-interacting control factors, and $\mathbf{j}^{\sim} = (j_1^{\sim}, \dots, j_s^{\sim})$ denote the $1 \times s$ vector of indices of the levels of the non-interacting noise factors. Let $\mathcal{I}^{\mathcal{I}}$, \mathcal{I}^{\sim} , $\mathcal{J}^{\mathcal{I}}$, and \mathcal{J}^{\sim} denote the (cross product) range sets for the indices $\mathbf{i}^{\mathcal{I}}$, \mathbf{i}^{\sim} , $\mathbf{j}^{\mathcal{I}}$, and \mathbf{j}^{\sim} , respectively. For example, in the case of an experiment with each factor at two levels, $\mathcal{I}^{\mathcal{I}} = \{0, 1\}^p$, $\mathcal{I}^{\sim} = \{0, 1\}^q$, $\mathcal{J}^{\mathcal{I}} = \{0, 1\}^r$, and $\mathcal{J}^{\sim} = \{0, 1\}^s$. In general we assume that level 0 represents the *lowest* level of each factor. The highest level can depend on the factor although in many cases we assume two-level experiments in which case $(\mathbf{1}_p, \mathbf{1}_q)$ denotes the vector of highest levels of the control factors.

The selection procedure described below assumes that only a fractional factorial experiment need be conducted (subject to the restriction that all means are estimable). Therefore the indices of the observed treatment combinations are a subset of the Cartesian product $\mathcal{I}^{\mathcal{I}} \times \mathcal{I}^{\sim} \times \mathcal{J}^{\mathcal{I}} \times \mathcal{J}^{\sim}$. Throughout, this paper will use the notation $|\mathcal{S}|$ to denote the cardinality of the generic set \mathcal{S} . Also, it assumes that the lowest level of any factor is designated as 0 and that, for two-level factor experiments, the higher level is designated by 1.

Let $Y_{\mathbf{i}, \mathbf{j}}$ and $\mu_{\mathbf{i}, \mathbf{j}}$ denote the response and its mean when the control factors are at level \mathbf{i} and the noise factors are at level \mathbf{j} ; here $\mathbf{i} = (i^{\mathcal{I}}, i^{\sim}) \in \mathcal{I}$, $\mathbf{j} = (j^{\mathcal{J}}, j^{\sim}) \in \mathcal{J}$, where \mathcal{I} is $\mathcal{I}^{\mathcal{I}} \times \mathcal{I}^{\sim}$ and \mathcal{J} is $\mathcal{J}^{\mathcal{J}} \times \mathcal{J}^{\sim}$. We first describe the goal of the screening problem in terms of the means $\{\mu_{\mathbf{i}, \mathbf{j}}\}$ and then introduce the ANOVA model that describes the $\{\mu_{\mathbf{i}, \mathbf{j}}\}$ in terms of main effects and interactions.

For each combination \mathbf{i} of control factors, we determine the *worst mean performance* of the response over the levels of the noise factors, i.e.,

$$\xi_{\mathbf{i}} = \min_{\mathbf{j}} \mu_{\mathbf{i}, \mathbf{j}},$$

where, for notational simplicity, the minimum is taken over the range $\mathbf{j} \in \mathcal{J}$. In words, $\xi_{\mathbf{i}}$ is the smallest mean response for the product design described by control factor combination \mathbf{i} . We denote the ordered $\xi_{\mathbf{i}}$ corresponding to the $|\mathcal{I}|$ product designs by

$$\xi_{[1]} \leq \cdots \leq \xi_{[|\mathcal{I}|]}. \quad (2.2)$$

The design associated with $\xi_{[|\mathcal{I}|]} = \max_{\mathbf{i}} \xi_{\mathbf{i}}$ is defined to be the *best product design*; this design is robust to the unknown levels of the noise factors in a maximin sense.

Our *goal* is to use the data from a suitable (complete or fractional factorial) experiment to select a subset of the product designs so that the following confidence requirement is satisfied.

Confidence Requirement: Given α with $0 < \alpha < 1$, we desire that

$$P_{\boldsymbol{\mu}} \{\text{CS}\} \geq 1 - \alpha \quad (2.3)$$

for all $\boldsymbol{\mu}$ satisfying the mean Model (2.4)–(2.5) defined below, where CS denotes the event that the selected subset includes the control factor combination associated with $\xi_{[|\mathcal{I}|]}$, the best product design.

Example 1 (Continued) The quality of the recipe $(S, F, E) = (s, f, e)$ is specified by

$$\xi_{sfe} = \min\{\mu_{sfe00}, \mu_{sfe01}, \mu_{sfe10}, \mu_{sfe11}\}$$

where $\mu_{sfe,tz}$ satisfies (2.1). Our goal is to identify a subset of the eight possible recipes that contains the best (S, F, E) combination, i.e., that associated with $\xi_{[8]} = \max \xi_{sfe}$. \square

Because we focus on the behavior of the $\mu_{\mathbf{i}, \mathbf{j}}$ for each fixed \mathbf{i} , we regard $\boldsymbol{\mu} = (\mu_{\mathbf{i}, \mathbf{j}})$ as an $|\mathcal{I}|$ by $|\mathcal{J}|$ matrix with each row of $\boldsymbol{\mu}$ corresponding to a single setting of the $p + q$ control factors and each column corresponding to a single setting of the $r + s$ noise factors. Several of

the proofs of the least favorable configuration results in the companion technical report, Pan and Santner (2003), use Kronecker product formulas for μ that assume the rows and columns of μ are arranged as follows. First, arrange the components of i corresponding to $i^{\mathcal{I}}$ in lexicographic order; then cross each fixed $i^{\mathcal{I}}$ with the set of i^{\sim} where the latter are also arranged in lexicographic order. The columns are ordered similarly as the cross product of the $j^{\mathcal{I}}$ and j^{\sim} , each arranged lexicographically. We illustrate the notation using our prototype Example 1.

Example 1 (Continued) Recall that this experiment had a total of three control factors, with one interacting control factor, S , and two noise factors, with but one interacting noise factor, T . Thus, the first four rows of μ have $S = 0$ and the remaining four rows have $S = 1$. The four rows with $S = 0$ are ordered *lexicographically* in the values of the non-interacting control factors F and E and similarly for the rows with $S = 1$ (see the right margin below)

$$\begin{array}{cccc}
 \left[\begin{array}{cccc}
 \mu_{00000} & \mu_{00001} & \mu_{00010} & \mu_{00011} \\
 \mu_{00100} & \mu_{00101} & \mu_{00110} & \mu_{00111} \\
 \mu_{01000} & \mu_{01001} & \mu_{01010} & \mu_{01011} \\
 \mu_{01100} & \mu_{01101} & \mu_{01110} & \mu_{01111} \\
 \mu_{10000} & \mu_{10001} & \mu_{10010} & \mu_{10011} \\
 \mu_{10100} & \mu_{10101} & \mu_{10110} & \mu_{10111} \\
 \mu_{11000} & \mu_{11001} & \mu_{11010} & \mu_{11011} \\
 \mu_{11100} & \mu_{11101} & \mu_{11110} & \mu_{11111}
 \end{array} \right] & & & \begin{array}{l}
 (S, F, E) \\
 (0, 0, 0) \\
 (0, 0, 1) \\
 (0, 1, 0) \\
 (0, 1, 1) \\
 (1, 0, 0) \\
 (1, 0, 1) \\
 (1, 1, 0) \\
 (1, 1, 1)
 \end{array} \\
 \begin{array}{cccc}
 (0, 0) & (0, 1) & (1, 0) & (1, 1) \\
 & & (T, Z) &
 \end{array}
 \end{array}$$

The first two columns of μ have $T = 0$ and the second two columns have $T = 1$. The pair of columns corresponding to each fixed T value are ordered so that the first has $Z = 0$ and the second has $Z = 1$. The entire μ matrix is $8 \times 4 = 2^{1+2} \times 2^{1+1}$. \square

The procedure below assumes that an ANOVA-type model

$$\mu_{i,j} = m_{CN}(i^{\mathcal{I}}, j^{\mathcal{I}}) + m_C(i^{\sim}) + m_N(j^{\sim}) \tag{2.4}$$

for the means is *known* and that this model holds for *all* $(i, j) \in \mathcal{I} \times \mathcal{J}$. The three components of the model

$$\begin{aligned}
m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{J}}) &= m_0 + \sum_{Q^{CN}} (C_{i_1^*}^{\mathcal{I}} C_{i_2^*}^{\mathcal{I}} \cdots C_{i_{p^*}^{\mathcal{I}}}^{\mathcal{I}} N_{j_1^*}^{\mathcal{I}} N_{j_2^*}^{\mathcal{I}} \cdots N_{j_{r^*}^{\mathcal{I}}}^{\mathcal{I}}) \mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{J}}. \\
m_C(\mathbf{i}^{\sim}) &= \sum_{Q^C} (C_{i_1^*}^{\sim} C_{i_2^*}^{\sim} \cdots C_{i_{q^*}^{\sim}}^{\sim}) \mathbf{i}^{\sim}, \text{ and} \\
m_N(\mathbf{j}^{\sim}) &= \sum_{Q^N} (N_{j_1^*}^{\sim} N_{j_2^*}^{\sim} \cdots N_{j_{s^*}^{\sim}}^{\sim}) \mathbf{j}^{\sim}
\end{aligned} \tag{2.5}$$

consist of the main effects and interactions that correspond to the interacting control and noise factors, to the non-interacting control factors, and to the non-interacting noise factors. The index set Q^{CN} identifies these quantities among the interacting control and noise factors while Q^C and Q^N are defined similarly for the non-interacting control factors and noise factors, respectively. We place the overall mean, m_0 , with the control \times noise interaction terms. Also, each term $(C_{i_1^*}^{\sim} \cdots C_{i_{q^*}^{\sim}}^{\sim}) \mathbf{i}^{\sim}$ of $m_C(\mathbf{i}^{\sim})$, for example, depends *only* on the q^* indices $1 \leq i_1^* < i_2^* < \cdots < i_{q^*}^* \leq q$ of \mathbf{i}^{\sim} . A similar convention applies to the terms $m_N(\mathbf{j}^{\sim})$ and $m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{J}})$. For example, in a model with $p = 3$ interacting control factors and $r = 2$ interacting noise factors, suppose $m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{J}})$ contains the interaction term $(C_1^{\mathcal{I}} C_2^{\mathcal{I}} N_1^{\mathcal{I}}) \mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{J}}$. Then $\mathbf{i}^{\mathcal{I}} = (i_1^{\mathcal{I}}, i_2^{\mathcal{I}}, i_3^{\mathcal{I}})$ and $\mathbf{j}^{\mathcal{J}} = (j_1^{\mathcal{J}}, j_2^{\mathcal{J}})$. The term $(C_1^{\mathcal{I}} C_2^{\mathcal{I}} N_1^{\mathcal{I}}) \mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{J}}$ depends only on the indices $i_1^{\mathcal{I}}$ and $i_2^{\mathcal{I}}$ in $\mathbf{i}^{\mathcal{I}}$ and on $j_1^{\mathcal{J}}$ in $\mathbf{j}^{\mathcal{J}}$. Lastly, we again emphasize that the screening procedure stated below assumes that the Model (2.4)—(2.5) is known. If the model is identified from the same data that is used to implement the screening procedure then the confidence level of the procedure will be affected, presumably negatively. Of course the same phenomenon is true in other settings where data are used for multiple purposes. This paper does not analyze the multiple stage procedure that first identifies the model and then chooses the critical value, although such a study would be of interest.

Example 1 (Continued) From (2.1), S and T are the interacting control and noise factors, respectively. The indices are $\mathbf{i}^{\mathcal{I}} = (s)$, $\mathbf{i}^{\sim} = (f, e)$, $\mathbf{j}^{\mathcal{J}} = (t)$, and $\mathbf{j}^{\sim} = (z)$. Thus, in the general notation of (2.4) applied to Model (2.1), we have $m_{CN}(s, t) = m_0 + S_s + T_t + (ST)_{st}$, $m_C(f, e) = F_f + E_e$, and $m_N(z) = Z_z$. \square

We assume that the experimental data satisfies

$$Y_{\mathbf{i}, \mathbf{j}} = \mu_{\mathbf{i}, \mathbf{j}} + \epsilon_{\mathbf{i}, \mathbf{j}}, \quad (\mathbf{i}, \mathbf{j}) \in \mathcal{D} \tag{2.6}$$

for each $(\mathbf{i}, \mathbf{j}) \in \mathcal{D} \subseteq \mathcal{I} \times \mathcal{J}$, where \mathcal{D} is the set of factor combinations at which data were observed. The $|\mathcal{D}|$ measurement errors $\epsilon_{\mathbf{i}, \mathbf{j}}$ are independent $N(0, \sigma_\epsilon^2)$ random variables. Terms

estimator of $\mu_{i,j}$. For each product design, $i \in \mathcal{I}$, estimate ξ_i by

$$\widehat{\xi}_i = \min_j \{\widehat{\mu}_{i,j}\}.$$

Let

$$\widehat{\xi}_{[1]} \leq \cdots \leq \widehat{\xi}_{[|\mathcal{I}|]}$$

denote the ordered $\widehat{\xi}_i$. In addition, we assume that there is an estimator S^2 of σ_ϵ^2 , for which $\nu S^2 / \sigma_\epsilon^2 \sim \chi_\nu^2$ and that S^2 is independent of the OLS estimator of $\mu_{\mathcal{D}}$. Ordinarily, such a chi-square estimator would be available when the number of observations, $|\mathcal{D}|$, is larger than the number of parameters estimated in the Model (2.4). We propose the following procedure to select a subset of the levels of the control factors containing the most robust control factor combination.

Procedure \mathcal{G} : Select control factor combination i if and only if

$$\widehat{\xi}_i \geq \widehat{\xi}_{[|\mathcal{I}|]} - hS,$$

where h is a cutoff value to be determined in the next section.

(s, f, e)	(t, z)				$\widehat{\xi}_{sfe}$
	(0, 0)	(0, 1)	(1, 0)	(1, 1)	
(0, 0, 0)	0.0	0.9	2.2	3.1	0.0
(0, 0, 1)	1.4	2.3	3.6	4.5	1.4
(0, 1, 0)	2.6	3.6	4.8	5.7	2.6
(0, 1, 1)	4.1	5.0	6.3	7.2	4.1
(1, 0, 0)	0.7	1.6	0.9	1.8	0.7
(1, 0, 1)	2.2	3.1	2.3	3.2	2.2
(1, 1, 0)	3.4	4.3	3.6	4.5	3.4
(1, 1, 1)	4.8	5.7	5.0	5.9	4.8

Table 2: OLS estimators $\widehat{\mu}_{sfe,tz}$ of the 32 treatment means for all recipe by baking conditions in Example 1 under Model (2.1) using the data in Table 1, and the corresponding estimated row minima $\widehat{\xi}_{sfe}$.

Example 1 (Continued) As described in the next section, we calculate $h = 2.29$ for Procedure \mathcal{G} to attain confidence level $100(1 - \alpha)\% = 95\%$. Table 2 lists the fitted means $\widehat{\mu}_{sfe,tz}$ and the associated $\widehat{\xi}_{sfe}$ based on Model (2.1) and the data in Table 1; the estimate of σ is $S = 0.4989$. Thus for Procedure \mathcal{G} , $\widehat{\xi}_{[8]} - hS = 4.8 - 1.14 = 3.66$ and the selected confidence subset is $\{(0, 1, 1), (1, 1, 1)\}$. Therefore, the data from this experiment shows that, with at least 95% confidence, the best recipe uses F and E at their high level and S is not determinable. Equivalently, the results of applying the statistical screening procedure can be interpreted as stating that recipes $\{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 1, 0), (1, 0, 1)\}$ need not be investigated further. \square

For simplicity, we assume below that there is one replicate of the design \mathcal{D} . This is the most common case in quality improvement experimentation. However, the theory and methods developed in this paper extend straightforwardly to situations where replicates of the design \mathcal{D} are observed—one simply views the sample means of the replicates as unreplicated observations from design \mathcal{D} and applies the results in this paper.

3 Determination of h Based on the Least Favorable Configuration

Sections 3 and 4 discuss two methods that can be applied to determine the critical value h of Procedure \mathcal{G} . The least favorable configuration (LFC) method discussed in this section searches for a configuration of means $\boldsymbol{\mu}$ at which the probability of correct selection is minimal (a least favorable configuration). The h value is computed by setting the probability of correct selection at the LFC equal to $1 - \alpha$. The LFC method provides the smallest h that guarantees $\min_{\boldsymbol{\mu}} P_{\boldsymbol{\mu}}\{\text{CS}\}$ attains $1 - \alpha$. The method described in Section 4 finds a *lower bound* (LB) for the probability of correct selection that depends on h and sets the lower bound equal to $1 - \alpha$. The LB method is more widely applicable than the LFC method but provides a conservative h ; in some situations, the h produced by the LB method is identical to that given by the LFC method.

To achieve (2.3), we must have $\inf_{\boldsymbol{\mu}} P_{\boldsymbol{\mu}}\{\text{CS}\} \geq 1 - \alpha$ where the infimum is over $\boldsymbol{\mu}$ satisfying Model (2.4)–(2.5). In some screening problems, it is possible to find a configuration of means $\boldsymbol{\mu}_{LFC}$ that satisfies the model and for which $P_{\boldsymbol{\mu}_{LFC}}\{\text{CS}\} = \inf_{\boldsymbol{\mu}} P_{\boldsymbol{\mu}}\{\text{CS}\}$; such a $\boldsymbol{\mu}_{LFC}$ is called a *least favorable configuration* of means. In this problem we will find a sequence of parameter configurations $\{\boldsymbol{\mu}_k\}$, each satisfying the Model (2.4)–(2.5), for which $\lim_{k \rightarrow \infty} P_{\boldsymbol{\mu}_k}\{\text{CS}\} = \inf_{\boldsymbol{\mu}} P_{\boldsymbol{\mu}}\{\text{CS}\}$. Informally, we refer to $\lim_{k \rightarrow \infty} \boldsymbol{\mu}_k$, when this limit exists, as a “least favorable configuration of means.”

As will be shown below, when the LFC exists, it is characterized by the model for the interacting control and noise factors,

$$m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{J}}) = m_0 + \sum_{Q_{\mathcal{I}}^{CN}} (C_{i_1^*}^{\mathcal{I}} C_{i_2^*}^{\mathcal{I}} \cdots C_{i_p^*}^{\mathcal{I}} N_{j_1^*}^{\mathcal{J}} N_{j_2^*}^{\mathcal{J}} \cdots N_{j_r^*}^{\mathcal{J}})_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{J}}}. \quad (3.1)$$

Following the convention used to arrange the rows and columns of the $\boldsymbol{\mu}$ matrix, we let $\mathbf{M} = (m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{J}}))$ be an $|\mathcal{I}^{\mathcal{I}}| \times |\mathcal{J}^{\mathcal{J}}|$ matrix with each row of \mathbf{M} corresponding to a single setting of the p interacting control factors and each column corresponding to a single setting of the r interacting noise factors with the elements of $\mathbf{i}^{\mathcal{I}}$ and $\mathbf{j}^{\mathcal{J}}$ ordered *lexicographically*.

Example 1 (Continued) For the taste-testing experiment, $\mu_{sfe,tz} = m_0 + S_s + F_f + E_e + T_t + Z_z + (ST)_{st}$, and so

$$m_{CN}(s, t) = m_0 + S_s + T_t + ST_{st},$$

for $s, t \in \{0, 1\}$ is the linear combination of the main effects and interactions between the interacting control factor S at level s and the interacting noise factor T at level t . The \mathbf{M} matrix in this example is

$$\begin{bmatrix} m_{CN}(0, 0) & m_{CN}(0, 1) \\ m_{CN}(1, 0) & m_{CN}(1, 1) \end{bmatrix}$$

with the rows corresponding to the levels of the control factor S and the columns to the levels of the noise factor T . \square

Theorem 3.1 provides the LFC of means for two-level experiments where exactly one control factor interacts with the noise factors. It (and later results) use the notation: $\mathbf{J}_{m \times n}$ to denote an $m \times n$ matrix of 1's, $\mathbf{1}_v$ to denote a row vector of v ones, $\mathbf{0}_v$ to denote a row vector of v zeros, and \otimes to denote Kronecker product.

Theorem 3.1 Assume the Model (2.4)—(2.5) has exactly one control factor that interacts with the noise factor(s) and the data $\mathbf{Y}_{\mathcal{D}}$ come from an orthogonal 2-level (fractional) factorial experiment for which all non-zero main effects and interactions are estimable, and free of non-zero confounding factors. If there exists a sequence of 2×2^r matrices \mathbf{M}_k for which

$$\lim_{k \rightarrow \infty} \mathbf{M}_k = \begin{bmatrix} 0 & +\infty & \cdots & +\infty \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (3.2)$$

then the $2^{1+q} \times 2^{r+s}$ matrix

$$\boldsymbol{\mu}_{LFC} = \lim_{k \rightarrow \infty} \mathbf{M}_k \otimes J_{2^q \times 2^s} = \begin{bmatrix} 0 & +\infty & \cdots & +\infty \\ 0 & 0 & \cdots & 0 \end{bmatrix} \otimes J_{2^q \times 2^s}, \quad (3.3)$$

is a least favorable configuration. Let h_k be the solution to the equation

$$P_{\boldsymbol{\mu}_k} \left\{ \widehat{\xi}_{1,1_q} \geq \max_{\mathbf{i}} \widehat{\xi}_{\mathbf{i}} - h_k S \right\} = 1 - \alpha, \quad (3.4)$$

where $\boldsymbol{\mu}_k = \mathbf{M}_k \otimes J_{2^q \times 2^s}$ then $h = \lim_{k \rightarrow \infty} h_k$ is the critical value for Procedure \mathcal{G} .

In practice, simulation is the simplest method to obtain h . Draw independent random variables of the form $\mathbf{Y}_{\mathcal{D}}$ having means given by the components of $\boldsymbol{\mu}_k = \mathbf{M}_k \otimes J_{2^q \times 2^s}$ and unit variances. Here \mathbf{M}_k is as in Theorem 3.1 with entries k that are at least 100 (100 is effectively ∞). Also draw $V \sim \chi_{\nu}^2$ and construct the estimators $\widehat{\xi}_{\mathbf{i}}$, for $\mathbf{i} \in \mathcal{I}$; form

$$T \equiv \sqrt{\nu} \left(\max_{\mathbf{i}} \widehat{\xi}_{\mathbf{i}} - \widehat{\xi}_{1,1_q} \right) / \sqrt{V}.$$

Then h_k (and thus h) can be estimated as the $100 \times (1 - \alpha)\%$ sample percentile of many such draws of T . Note that one can always take $\sigma_e^2 = 1$ in the simulation because the estimator S^2 of σ_e^2 has the property that νS^2 follows a chi-square distribution with ν degrees of freedom and the random variable $(\max_{\mathbf{i}} \widehat{\xi}_{\mathbf{i}} - \widehat{\xi}_{1,1,1})/S$ is equivalent to T .

Example 1 (Continued) We can take $\mathbf{M}_k = \begin{bmatrix} 0 & k \\ 0 & 0 \end{bmatrix}$ in Theorem 3.1 because the model $m_{CN}(s, t) = m_0 + S_s + T_t + (ST)_{st}$ spans $\mathbb{R}^{2 \times 2}$. Thus

$$\boldsymbol{\mu}_k = \mathbf{M}_k \otimes J_{2^2 \times 2} = \left[\begin{array}{cc|cc} 0 & 0 & k & k \\ 0 & 0 & k & k \\ 0 & 0 & k & k \\ 0 & 0 & k & k \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right] \rightarrow \left[\begin{array}{cc|cc} 0 & 0 & +\infty & +\infty \\ 0 & 0 & +\infty & +\infty \\ 0 & 0 & +\infty & +\infty \\ 0 & 0 & +\infty & +\infty \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

as $k \rightarrow \infty$ which is the LFC.

The configuration identified above can be explained intuitively as follows. In this problem, as in many subset selection problems, the LFC for $\boldsymbol{\mu}$ makes all associated ξ_i parameters *equal* (to zero). The LFC associates the last row, with levels $(1, 1, 1)$, as best control combination. The $\boldsymbol{\mu}$ configuration that minimizes the probability of the event

$$\left[\widehat{\xi}_{111} \geq \widehat{\xi}_{sfe}, (s, f, e) \neq (1, 1, 1) \right]$$

occurs when $\widehat{\xi}_{111}$ is stochastically *small* and the remaining $\widehat{\xi}_{sfe}$ are stochastically *large*. Recall that each $\widehat{\xi}_{sfe}$ is a minimum of four estimated means. The estimator $\widehat{\xi}_{111}$ is stochastically small because it is the minimum of four $\widehat{\mu}_{111,tz}$ each of which has zero mean. The LFC makes four of the remaining $\widehat{\xi}_{sfe}$ stochastically large by effectively eliminating two of the $\widehat{\mu}_{sfe,tz}$ from each minimum calculation by making their means large ($+\infty$). However, the $\boldsymbol{\mu}$ model restrictions allow only four of the seven rows corresponding to inferior control factor combinations to be based on two $\widehat{\mu}_{sfe,tz}$. The model constraints force the remaining three $\widehat{\xi}_{sfe}$ to be the minimum of four means, as is $\widehat{\xi}_{111}$.

To compute h , generate independent and identically distributed (iid) standard normal random

and each of $i_1, i_2, i_3, j_1, j_2,$ and j_3 is 0 or 1. The slight change in the subscript notation is used to emphasize the identity of the control and noise indices. Clearly (3.6) does not span $\mathbb{R}^{2 \times 2^2}$ but it can be shown that the 2×4 matrix $\mathbf{M} = (m_{CN}(i_1, j_1 j_2))$ satisfies Model (3.6) if and only if \mathbf{M} satisfies

$$\begin{bmatrix} \beta_0 + \beta_1^{\mathcal{I}} & \beta_0 + \beta_1^{\mathcal{I}} + \gamma_2^{\mathcal{I}} + \delta_{12}^{\mathcal{I}} & \beta_0 + \beta_1^{\mathcal{I}} + \gamma_1^{\mathcal{I}} + \delta_{11}^{\mathcal{I}} & \beta_0 + \beta_1^{\mathcal{I}} + \gamma_1^{\mathcal{I}} + \gamma_2^{\mathcal{I}} + \delta_{11}^{\mathcal{I}} + \delta_{12}^{\mathcal{I}} \\ \beta_0 & \beta_0 + \gamma_2^{\mathcal{I}} & \beta_0 + \gamma_1^{\mathcal{I}} & \beta_0 + \gamma_1^{\mathcal{I}} + \gamma_2^{\mathcal{I}} \end{bmatrix}$$

where each coefficient is an arbitrary real-valued number. Using this representation, it is easy to identify the required sequence \mathbf{M}_k in Theorem 3.1. Force the first column and last row of \mathbf{M}_k to be zero by setting $0 = \beta_1^{\mathcal{I}} = \beta_0 = \gamma_1^{\mathcal{I}} = \gamma_2^{\mathcal{I}}$. Then set $\delta_{11}^{\mathcal{I}} = k = \delta_{12}^{\mathcal{I}}$. The resulting matrix is

$$\mathbf{M}_k = \begin{bmatrix} 0 & k & k & 2k \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus

$$\boldsymbol{\mu}_k = \mathbf{M}_k \otimes J_{2^2 \times 2} = \begin{bmatrix} 0 & 0 & k & k & k & k & 2k & 2k \\ 0 & 0 & k & k & k & k & 2k & 2k \\ 0 & 0 & k & k & k & k & 2k & 2k \\ 0 & 0 & k & k & k & k & 2k & 2k \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & +\infty & +\infty & +\infty & +\infty & +\infty & +\infty \\ 0 & 0 & +\infty & +\infty & +\infty & +\infty & +\infty & +\infty \\ 0 & 0 & +\infty & +\infty & +\infty & +\infty & +\infty & +\infty \\ 0 & 0 & +\infty & +\infty & +\infty & +\infty & +\infty & +\infty \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

as $k \rightarrow \infty$ which yields a LFC. As above, an estimate of h can be obtained by solving Equation (3.4) via Monte Carlo simulation. \square

The following example shows a model where the condition (3.2)–(3.3) of Theorem 3.1 is not satisfied. For such situations, the lower bound method discussed in the next section can be used.

Example 3 Consider a 3-factor experiment with each factor at level 0 or 1. Suppose that there is one control factor $C_1^{\mathcal{I}}$ and two noise factors $N_1^{\mathcal{I}}$ and $N_2^{\mathcal{I}}$. For simplicity, suppose further that there are no non-interacting control or noise factors and the model for the mean of $Y_{i_1, j_1 j_2}$ is

$$\mu_{i_1, j_1 j_2} = m_0 + (C_1^{\mathcal{I}})_{i_1} + (N_1^{\mathcal{I}})_{j_1} + (C_1^{\mathcal{I}} N_1^{\mathcal{I}})_{i_1, j_2} + (C_1^{\mathcal{I}} N_1^{\mathcal{I}} N_2^{\mathcal{I}})_{i_1, j_2 j_3}$$

where each of i_1, j_1 and j_3 is 0 or 1. Using the technique of Example 2, it can be shown that the \mathbf{M}_k required in Theorem 3.1 does not exist. \square

When there is more than one interacting control factor, the lower bound method in Subsection 4.1 can be used to determine a (possibly conservative) cutoff value and then the method of Subsection 4.2 to determine whether the resulting critical value is exact or not.

We conclude this section by describing a systematic technique for determining whether the condition (3.2) of Theorem 3.1 is satisfied in any particular application. The technique involves solving a linear programming (LP) problem determined by the model $m_{CN}(\mathbf{i}^I, \mathbf{j}^I)$. Recall that the canonical form of an LP in the n decision variables $\mathbf{x} = (x_1, \dots, x_n)$ is

$$\begin{aligned} & \max \mathbf{c}^\top \mathbf{x} \\ & \text{s.t.} \\ & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq 0 \end{aligned}$$

where \mathbf{c} , \mathbf{A} , and \mathbf{b} are given $n \times 1$, $m \times n$, and $m \times 1$ arrays.

The LP problem for determining if (3.2) is satisfied is formulated as follows. Let \mathcal{Z} (“Zero”) denote those $(\mathbf{i}^I, \mathbf{j}^I)$ combinations corresponding to elements in the first column *or* last row of the matrix $\mathbf{M} = (m_{CN}(\mathbf{i}^I, \mathbf{j}^I))$ introduced previously. We wish to determine whether there exists a *sequence* of model matrices each satisfying (3.1) whose elements $m_{CN}(\mathbf{i}^I, \mathbf{j}^I) \rightarrow +\infty$ for all $(\mathbf{i}^I, \mathbf{j}^I) \notin \mathcal{Z}$ and $m_{CN}(\mathbf{i}^I, \mathbf{j}^I) = 0$ for all $(\mathbf{i}^I, \mathbf{j}^I) \in \mathcal{Z}$. To solve this problem, we introduce the auxiliary scalar variable w whose role is to be a lower bound on the elements we wish to simultaneously drive to $+\infty$. Then we solve

$$\begin{aligned} & \max w & (3.7) \\ & \text{s.t.} \\ & m_{CN}(\mathbf{i}^I, \mathbf{j}^I) = 0 \text{ for all } (\mathbf{i}^I, \mathbf{j}^I) \in \mathcal{Z} \\ & w \leq m_{CN}(\mathbf{i}^I, \mathbf{j}^I) \text{ for all } (\mathbf{i}^I, \mathbf{j}^I) \notin \mathcal{Z} \\ & \sum_{\mathbf{i}^I \in \mathcal{I}^I \text{ and } \mathbf{j}^I \in \mathcal{J}^I} (C_{i_1^*}^I C_{i_2^*}^I \cdots C_{i_p^*}^I N_{j_1^*}^I N_{j_2^*}^I \cdots N_{j_r^*}^I)_{\mathbf{i}^I, \mathbf{j}^I} = 0 \text{ for all } (\mathbf{i}^*, \mathbf{j}^*) \in Q_{\mathcal{I}}^{CN} \end{aligned}$$

where $m_{CN}(\mathbf{i}^I, \mathbf{j}^I)$ is a short hand for the for the sum

$$m_0 + \sum_{Q^{CN}} (C_{i_1^*}^I C_{i_2^*}^I \cdots C_{i_p^*}^I N_{j_1^*}^I N_{j_2^*}^I \cdots N_{j_r^*}^I)_{\mathbf{i}^I, \mathbf{j}^I}.$$

The decision variables for the LP are w , the sets of $\{(C_{i_1^*}^I C_{i_2^*}^I \cdots C_{i_p^*}^I N_{j_1^*}^I N_{j_2^*}^I \cdots N_{j_r^*}^I)_{\mathbf{i}^I, \mathbf{j}^I}\}$ for each $(\mathbf{i}^*, \mathbf{j}^*) \in Q^{CN}$, and m_0 . The first constraint of constraints forces the first column and last row of \mathbf{M} to be zero while the second insures w is the minimum of the $m_{CN}(\mathbf{i}^I, \mathbf{j}^I)$ for $(\mathbf{i}^I, \mathbf{j}^I) \notin \mathcal{Z}$. The third set of constraints are the usual main effect and interaction identifiability constraints. The sequence required in Theorem 3.1 is obtained by letting $w \rightarrow +\infty$.

4 Determination of h Based on a Lower Bound for the Probability of Correct Selection

Procedure \mathcal{G} requires the computation of the smallest h satisfying $\min_{\boldsymbol{\mu}} P_{\boldsymbol{\mu}}\{\text{CS}\} = 1 - \alpha$ where the minimum is over all $\boldsymbol{\mu}$ satisfying the mean Model (2.4)–(2.5). An exact determination of this h is possible when one knows an $\boldsymbol{\mu}_{LFC}$ that guarantees $P_{\boldsymbol{\mu}_{LFC}}\{\text{CS}\} = \min_{\boldsymbol{\mu}} P_{\boldsymbol{\mu}}\{\text{CS}\}$. Section 3 provides one set of sufficient conditions under which such an LFC can be found. This section provides a lower bound (LB) for $\min_{\boldsymbol{\mu}} P_{\boldsymbol{\mu}}\{\text{CS}\}$ that can be applied in more general settings when the sufficient condition is not satisfied. Setting the lower bound equal to $1 - \alpha$ gives a conservative estimate of h . Of course, if the lower bound is sharp, as is often the case for two-level experiments, then the resulting h is the exact critical value required. In this case, Theorem 3.1 and the LB method yield the same critical value.

4.1 A Lower Bound for the Probability of Correct Selection

The lower bound developed below is based on the fact that $P_{\boldsymbol{\mu}_{LFC}}\{\text{CS}\}$ is equal to the probability of an event *related* to CS that is calculated under the *homogeneity configuration* where *all* treatment means are equal to *zero*. We use Example 1 to illustrate this viewpoint and then generalize the idea.

Example 1 (Continued) In Section 3, we showed that h is the limit of the sequence $\{h_k\}$ where

$$P_{\boldsymbol{\mu}_k} \left\{ \widehat{\xi}_{111} \geq \widehat{\xi}_{sfe} - h_k S, \quad s, f, e \in \{0, 1\} \right\} = 1 - \alpha,$$

implicitly defines h_k where $\boldsymbol{\mu}_k = \mathbf{M}_k \otimes J_{2 \times 2}$ and $\mathbf{M}_k = \begin{bmatrix} 0 & k \\ 0 & 0 \end{bmatrix}$. Note that the product design $(S, F, E) = (1, 1, 1)$ is labeled as best in the [CS] event. As $k \rightarrow \infty$, $P_{\boldsymbol{\mu}_k}\{\text{CS}\} \rightarrow P_{\boldsymbol{\mu}_{LFC}}\{\text{CS}\}$ where $\boldsymbol{\mu}_{LFC}$ is the Kronecker product (3.3).

Suppose that $\boldsymbol{\mu}_0$ is a 8×4 matrix of zero means. Consider 16 observations with indices shown in the matrix equation (2.7) but having zero means and unit variances. Compute the 12 estimated means $\widehat{\mu}_{sfe,tz}$ shown in (4.2) based on these observations using the linear model (2.1). It can be shown that

$$P_{\boldsymbol{\mu}_{LFC}}\{\text{CS}\} = P_{\boldsymbol{\mu}_0} \left\{ \begin{array}{l} \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \widehat{\mu}_{00000} - h S \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \widehat{\mu}_{00100} - h S \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \widehat{\mu}_{01000} - h S \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \widehat{\mu}_{01100} - h S \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \min\{\widehat{\mu}_{10000}, \widehat{\mu}_{10010}\} - h S \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \min\{\widehat{\mu}_{10100}, \widehat{\mu}_{10110}\} - h S \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \min\{\widehat{\mu}_{11000}, \widehat{\mu}_{11010}\} - h S \end{array} \right\} \quad (4.1)$$

where the right-hand side of (4.1) involves the row minimums of the *reduced* matrix of means

$$\widehat{\boldsymbol{\mu}}^r \triangleq \begin{array}{c} \left[\begin{array}{cccc} \widehat{\mu}_{00000} & - & - & - \\ \widehat{\mu}_{00100} & - & - & - \\ \widehat{\mu}_{01000} & - & - & - \\ \widehat{\mu}_{01100} & - & - & - \\ \widehat{\mu}_{10000} & - & \widehat{\mu}_{10010} & - \\ \widehat{\mu}_{10100} & - & \widehat{\mu}_{10110} & - \\ \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{array} \right] \begin{array}{l} (S, F, E) \\ (0, 0, 0) \\ (0, 0, 1) \\ (0, 1, 0) \\ (0, 1, 1) \\ (1, 0, 0) \\ (1, 0, 1) \\ (1, 1, 0) \\ (1, 1, 1) \end{array} \end{array} \quad (4.2)$$

$$\begin{array}{c} (0, 0) (0, 1) (1, 0) (1, 1) \\ (T, Z) \end{array}$$

and row (1, 1, 1) designates the best recipe. Note that the event [CS] requires the entire 8×4 matrix of $\widehat{\boldsymbol{\mu}}_{sfe,tz}$.

To provide some intuition about the equality (4.1), we note first that $\widehat{\mu}_{000tz} \xrightarrow{a.s.} \infty$ as $k \rightarrow \infty$ under $\boldsymbol{\mu}_k$; hence it suffices to ignore the upper right 4×2 submatrix of estimated means when calculating row means. That the remaining means can be ignored follows from the representation

$$\widehat{\mu}_{sfe,tz} = \widehat{m}_{CN}(st) + \widehat{m}_C(fe) + \widehat{m}_N(z)$$

for the estimated mean where the three terms are the OLS estimators of the model parts. Using this representation it is straightforward to show that, for any $(f, e) \neq (1, 1)$, the following are equivalent

$$\begin{aligned} \widehat{\xi}_{111} &\geq \widehat{\xi}_{1fe} \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}, \widehat{\mu}_{11101}, \widehat{\mu}_{11111}\} &\geq \min\{\widehat{\mu}_{1fe00}, \widehat{\mu}_{1fe10}, \widehat{\mu}_{1fe01}, \widehat{\mu}_{1fe11}\} \\ \widehat{m}_C(11) + \widehat{m}_N(0) + \min\{\widehat{m}_{CN}(10), \widehat{m}_{CN}(11)\} &\geq \widehat{m}_C(fe) + \widehat{m}_N(0) + \min\{\widehat{m}_{CN}(10), \widehat{m}_{CN}(11)\} \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} &\geq \min\{\widehat{\mu}_{1fe00}, \widehat{\mu}_{1fe10}\}. \end{aligned}$$

A similar argument shows that for any (j, k) ,

$$\widehat{\xi}_{111} \geq \widehat{\xi}_{0fe} \text{ if and only if } \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \widehat{\mu}_{0fe00}.$$

The estimators $\widehat{\mu}_{sfe,tz}$ involved in the matrix $\widehat{\boldsymbol{\mu}}^r$ can be viewed as the result of applying the following (general) process to the complete 8×4 matrix of $\widehat{\boldsymbol{\mu}}_{sfe,tz}$. First, $\widehat{\mu}_{sfe,tz}$ at the *high* level of the noninteracting noise factors are eliminated; in this case, $z = 1$ is the high level of the noninteracting noise factor and so both columns with $z = 1$ are eliminated. Second, among the

remaining $\widehat{\mu}_{sfe,tz}$, those estimated means having the *same* levels of the *interacting* control factor as the best product design are *retained* in the reduced $\widehat{\boldsymbol{\mu}}^r$ matrix. In this example, the best product design is at control level $S = 1$ so all eight $\widehat{\mu}_{sfe,tz}$ with $s = 1$, i.e., the last four rows of $\widehat{\boldsymbol{\mu}}^r$, are retained. Third, for each of the interacting control factor levels that are *different* from those of the best product design, i.e., $s = 0$ in this example, we choose *one* level of the interacting noise factors. In this case T is the interacting control factor and for each $f, e \in \{0, 1\}$, $\widehat{\boldsymbol{\mu}}^r$ has retained $\widehat{\mu}_{0fe,00}$ in the choice between $\widehat{\mu}_{0fe,00}$ and $\widehat{\mu}_{0fe,10}$; thus the first four rows of $\widehat{\boldsymbol{\mu}}^r$ contain only the first element.

In general, the final step can be viewed as the specification of a function $\boldsymbol{j}^{\mathcal{I}}(\cdot)$ with values in $\mathcal{J}^{\mathcal{I}}$ which is defined as follows. Fix a vector $\boldsymbol{I} = (\boldsymbol{I}^{\mathcal{I}}, \boldsymbol{I}^{\sim\mathcal{I}}) \in \mathcal{I}^{\mathcal{I}} \times \mathcal{I}^{\sim\mathcal{I}}$ which is designated as the set of control factor levels corresponding to the best product design. For each vector $\boldsymbol{i}^{\mathcal{I}}$ of non-best interacting control factor levels, i.e., $\boldsymbol{i}^{\mathcal{I}} \in \mathcal{I}^{\mathcal{I}}$ with $\boldsymbol{i}^{\mathcal{I}} \neq \boldsymbol{I}^{\mathcal{I}}$, let $\boldsymbol{j}^{\mathcal{I}}(\boldsymbol{i}^{\mathcal{I}}) \in \mathcal{J}^{\mathcal{I}}$ specify a set of levels of the interacting noise factors. In the taste test case of Example 1, $\{\boldsymbol{i}^{\mathcal{I}} \in \mathcal{I}^{\mathcal{I}} : \boldsymbol{i}^{\mathcal{I}} \neq \boldsymbol{I}^{\mathcal{I}}\} = \{0\}$ is the only non-best level of the control factor. The possible levels of the interacting noise variables are $\mathcal{J}^{\mathcal{I}} = \{0, 1\}$; the matrix $\widehat{\boldsymbol{\mu}}^r$ in (4.2) corresponds to $\boldsymbol{j}^{\mathcal{I}}(0) = 0$. With this notation, we denote the right-hand side probability (4.1) by $P\{1, \mathbf{1}_2, \boldsymbol{j}^{\mathcal{I}}(0) = 0\}$. \square

The probability $P\{1, \mathbf{1}_2, \boldsymbol{j}^{\mathcal{I}}(0) = 0\}$ is determined by the levels of *both* the interacting *and* non-interacting control factors for the product design $(1, \mathbf{1}_2)$ and by the function $\boldsymbol{j}^{\mathcal{I}}(0) = 0$. This observation is now extended to a general setting. Recall that $\boldsymbol{I} = (\boldsymbol{I}^{\mathcal{I}}, \boldsymbol{I}^{\sim\mathcal{I}})$ denotes an arbitrary specification of control factor levels (that should be thought of as the “best product design” in the probability calculation of (4.3)). Assume that for every *interacting* control factor combination $\boldsymbol{i}^{\mathcal{I}} \neq \boldsymbol{I}^{\mathcal{I}}$, that an interacting noise factor combination $\boldsymbol{j}^{\mathcal{I}}(= \boldsymbol{j}^{\mathcal{I}}(\boldsymbol{i}^{\mathcal{I}})) \in \mathcal{J}^{\mathcal{I}}$ has been selected. Define

$$P\{\boldsymbol{I}^{\mathcal{I}}, \boldsymbol{I}^{\sim\mathcal{I}}, \boldsymbol{j}^{\mathcal{I}}(\cdot)\} = P_{\boldsymbol{\mu}_0} \left\{ \begin{array}{l} \min_{\boldsymbol{j}^{\mathcal{I}}} \left\{ \widehat{\boldsymbol{\mu}}_{\boldsymbol{I}^{\mathcal{I}}, \boldsymbol{I}^{\sim\mathcal{I}}, \boldsymbol{j}^{\mathcal{I}}, \mathbf{0}_s} \right\} \geq \widehat{\boldsymbol{\mu}}_{\boldsymbol{i}^{\mathcal{I}}, \boldsymbol{i}^{\sim\mathcal{I}}, \boldsymbol{j}^{\mathcal{I}}(\boldsymbol{i}^{\mathcal{I}}), \mathbf{0}_s} - h S, \text{ for } \boldsymbol{i}^{\mathcal{I}} \neq \boldsymbol{I}^{\mathcal{I}}, \boldsymbol{i}^{\sim\mathcal{I}} \in \mathcal{I}^{\sim\mathcal{I}}; \\ \min_{\boldsymbol{j}^{\mathcal{I}}} \left\{ \widehat{\boldsymbol{\mu}}_{\boldsymbol{I}^{\mathcal{I}}, \boldsymbol{I}^{\sim\mathcal{I}}, \boldsymbol{j}^{\mathcal{I}}, \mathbf{0}_s} \right\} \geq \min_{\boldsymbol{j}^{\mathcal{I}}} \left\{ \widehat{\boldsymbol{\mu}}_{\boldsymbol{I}^{\mathcal{I}}, \boldsymbol{i}^{\sim\mathcal{I}}, \boldsymbol{j}^{\mathcal{I}}, \mathbf{0}_s} \right\} - h S, \text{ for } \boldsymbol{i}^{\sim\mathcal{I}} \in \mathcal{I}^{\sim\mathcal{I}} \end{array} \right\}, \quad (4.3)$$

where all three minima are over $\boldsymbol{j}^{\mathcal{I}} \in \mathcal{J}^{\mathcal{I}}$. As seen in Example 1, such a probability is a lower bound for $P_{\boldsymbol{\mu}}\{\text{CS}\}$. We emphasize that the probability $P\{\boldsymbol{I}^{\mathcal{I}}, \boldsymbol{I}^{\sim\mathcal{I}}, \boldsymbol{j}^{\mathcal{I}}(\cdot)\}$ involves only $\widehat{\boldsymbol{\mu}}_{\boldsymbol{i}, \boldsymbol{j}}$ where each of the s non-interacting control factors are set at their low level, i.e., $\boldsymbol{j}^{\sim\mathcal{I}} = \mathbf{0}_s$. The next theorem provides a lower bound for the probability of correct selection in terms of the set of $P\{\boldsymbol{I}^{\mathcal{I}}, \boldsymbol{I}^{\sim\mathcal{I}}, \boldsymbol{j}^{\mathcal{I}}(\cdot)\}$; the bound is valid for *any* mean model using *any* complete or fractional factorial design.

Theorem 4.1 Consider an experiment involving factors with arbitrary numbers of levels and having mean $\boldsymbol{\mu} = (\mu_{i,j})$ that satisfies (2.4)—(2.6). Suppose that each element of $\boldsymbol{\mu}$ is estimable, independent of non-zero confounding factors. Then

$$P_{\boldsymbol{\mu}}\{\text{CS}\} \geq \min_{(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim})} \min_{\mathbf{j}^{\mathcal{I}}(\cdot)} P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\} \quad (4.4)$$

where the first minimum is over $(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}) \in \mathcal{I}^{\mathcal{I}} \times \mathcal{I}^{\sim}$ and the second is over all functions $\mathbf{j}^{\mathcal{I}}(\cdot)$.

In principle, one can set the lower bound on the right-hand side of (4.4) equal to $1 - \alpha$ and solve for the (best possible) constant, call it h_{LB} . The value h_{LB} is a conservative critical value for Procedure \mathcal{G} when the lower bound is not sharp and is exact when the lower bound is sharp. In general, there are a large number of different combinations of $\mathbf{I}^{\mathcal{I}} \in \mathcal{I}^{\mathcal{I}}, \mathbf{I}^{\sim} \in \mathcal{I}^{\sim}$ and functions $\mathbf{j}^{\mathcal{I}}(\cdot)$ to be considered in the minima calculations of (4.4).

However, for the important case of orthogonal two-level fractional factorial or two-level full factorial experiments, both common in screening studies, the right-hand side of Equation (4.4) is particularly simple.

Corollary 4.2 For an orthogonal 2-level (full or fractional) factorial experiment,

$$\min_{\mathbf{j}^{\mathcal{I}}(\cdot)} P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\} \quad (4.5)$$

is constant over $\mathbf{I}^{\mathcal{I}} \in \{0, 1\}^p$ and $\mathbf{I}^{\sim} \in \{0, 1\}^q$. Thus $\min_{\mathbf{j}^{\mathcal{I}}(\cdot)} P\{\mathbf{1}_p, \mathbf{1}_q, \mathbf{j}^{\mathcal{I}}(\cdot)\}$ is a lower bound for $P_{\boldsymbol{\mu}}\{\text{CS}\}$.

Under the hypotheses of Corollary 4.2 one must solve

$$P\{\mathbf{1}_p, \mathbf{1}_q, \mathbf{j}^{\mathcal{I}}(\cdot)\} = 1 - \alpha$$

for all $\mathbf{j}^{\mathcal{I}}(\cdot)$ to obtain the best possible solution (although possibly conservative) h_{LB} to Equation (4.5). In practice, often there are only a small number of functions $\mathbf{j}^{\mathcal{I}}(\cdot)$ involved in the left-hand minimum of Equation (4.5) and this brute force approach is feasible.

Example 1 (Continued) Section 3 presented a lower bound based on the function $\mathbf{j}^{\mathcal{I}}(0) = 0$ for obtaining the critical value h for this example. Now we examine the alternative lower bound

given by Corollary 4.2 corresponding to $\mathbf{j}^{\mathcal{I}}(0) = 1$. The corresponding reduced matrix is

$$\widehat{\boldsymbol{\mu}}^r \triangleq \begin{array}{c} \left[\begin{array}{ccc} - & \widehat{\mu}_{00010} & - \\ - & \widehat{\mu}_{00110} & - \\ - & \widehat{\mu}_{01010} & - \\ - & \widehat{\mu}_{01110} & - \\ \widehat{\mu}_{10000} & - & \widehat{\mu}_{10010} \\ \widehat{\mu}_{10100} & - & \widehat{\mu}_{10110} \\ \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} \end{array} \right] \begin{array}{l} (S, F, E) \\ (0, 0, 0) \\ (0, 0, 1) \\ (0, 1, 0) \\ (0, 1, 1) \\ (1, 0, 0) \\ (1, 0, 1) \\ (1, 1, 0) \\ (1, 1, 1) \end{array} \\ (0, 0)(0, 1)(1, 0)(1, 1) \\ (T, Z) \end{array} \quad (4.6)$$

The probability of correctly selecting the last row of this matrix based on the row minimums and using a yardstick of length hS is

$$P\{1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 1\} = P_{\boldsymbol{\mu}_0} \left\{ \begin{array}{l} \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \widehat{\mu}_{00010} - hS \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \widehat{\mu}_{00110} - hS \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \widehat{\mu}_{01010} - hS \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \widehat{\mu}_{01110} - hS \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \min\{\widehat{\mu}_{10000}, \widehat{\mu}_{10010}\} - hS \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \min\{\widehat{\mu}_{10100}, \widehat{\mu}_{10110}\} - hS \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \min\{\widehat{\mu}_{11000}, \widehat{\mu}_{11010}\} - hS \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \geq \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} - hS \end{array} \right\}.$$

By Corollary 4.2, the smaller of $P\{1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 0\}$ and $P\{1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 1\}$ is a lower bound for the probability of correct selection. However, the next corollary shows that these two probabilities are equal and thus both give the same lower bound. That the lower bound is sharp follows from Theorem 4.4 in Subsection 4.2 and will be discussed there. \square

The following example illustrates a case where Corollary 4.2 represents an effective way to find a critical value.

Example 4 Consider a five-factor experiment with each factor at level 0 or 1. Suppose that there are three control factors and two noise factors. Two control factors interact with the same noise factor and the third control factor does not interact with any noise factors; the other noise factor does not interact with any control factors. Our notation describing this situation is

Index	Name	Type
i_1, i_2	$C_1^{\mathcal{I}}, C_2^{\mathcal{I}}$	Interacting
i_3	C_1^{\sim}	Non-interacting
j_1	$N_1^{\mathcal{I}}$	Interacting
j_2	N_1^{\sim}	Non-interacting

Suppose that the 2^{5-1} fractional factorial experiment is conducted that yields the data

$$\mathbf{Y}_{\mathcal{D}} = \begin{array}{c} \begin{bmatrix} & Y_{00001} & Y_{00010} & & \\ Y_{00100} & & & & Y_{00111} \\ Y_{01000} & & & & Y_{01011} \\ & Y_{01101} & Y_{01110} & & \\ Y_{10000} & & & & Y_{10011} \\ & Y_{10101} & Y_{10110} & & \\ & Y_{11001} & Y_{11010} & & \\ Y_{11100} & & & & Y_{11111} \end{bmatrix} & (C_1^{\mathcal{I}}, C_2^{\mathcal{I}}, C_1^{\sim}) \\ \begin{array}{l} (0, 0, 0) \\ (0, 0, 1) \\ (0, 1, 0) \\ (0, 1, 1) \\ (1, 0, 0) \\ (1, 0, 1) \\ (1, 1, 0) \\ (1, 1, 1) \end{array} \end{array} \cdot$$

$$(0, 0)(0, 1)(1, 0)(1, 1)$$

$$(N_1^{\mathcal{I}}, N_1^{\sim})$$

and a $10 = 16 - 6$ degrees of freedom estimator of σ_{ϵ}^2 . Assume the model

$$\mu_{i_1 i_2 i_3, j_1 j_2} = m_{CN}(i_1 i_2, j_1) + m_C(i_3) + m_N(j_2). \quad (4.7)$$

for the mean of $Y_{i_1 i_2 i_3, j_1 j_2}$ where $m_N(j_2) = (N_1^{\sim})_{j_2}$, $m_C(i_3) = (C_1^{\sim})_{i_3}$,

$$m_{CN}(i_1 i_2, j_1) = m_0 + (C_1^{\mathcal{I}})_{i_1} + (C_2^{\mathcal{I}})_{i_2} + (N_1^{\mathcal{I}})_{j_1} + (C_1^{\mathcal{I}} C_2^{\mathcal{I}})_{i_1 i_2} + (C_1^{\mathcal{I}} N_1^{\mathcal{I}})_{i_1, j_1} + (C_2^{\mathcal{I}} N_1^{\mathcal{I}})_{i_2, j_1}$$

and each of i_1, i_2, i_3, j_1 , and j_2 is 0 or 1. By Corollary 4.2, a lower bound for $\min_{\boldsymbol{\mu} \in (4.7)} P_{\boldsymbol{\mu}}\{\text{CS}\}$ is given by the minimum of $P\{\mathbf{1}_2, 1, \mathbf{j}^{\mathcal{I}}(\cdot)\}$ over the 8 possible functions $\{\mathbf{j}^{\mathcal{I}}(i_1, i_2)\}$ on $\{(0, 1), (1, 0), (0, 0)\}$. Here, the last row of $\boldsymbol{\mu}$, that with $i_1 = i_2 = i_3 = 1$, is labeled as the best. For example, in Step 3, the function $\mathbf{j}^{\mathcal{I}}(i_1, i_2) \equiv 0$ always uses the estimator $\widehat{\mu}_{i_1 i_2 i_3, 00}$, at the lower level $j_1 = 0$ of the interacting noise factor $N_1^{\mathcal{I}}$, rather than that at the high level, $\widehat{\mu}_{i_1 i_2 i_3, 10}$. The reduced matrix of $\widehat{\mu}_{i_1 i_2 i_3, j_1 j_2}$ associated with $\mathbf{j}^{\mathcal{I}}(\cdot) \equiv 0$ is

$$\widehat{\boldsymbol{\mu}}^r = \begin{array}{c} \begin{bmatrix} \widehat{\mu}_{00000} & - & - & - \\ \widehat{\mu}_{00100} & - & - & - \\ \widehat{\mu}_{01000} & - & - & - \\ \widehat{\mu}_{01100} & - & - & - \\ \widehat{\mu}_{10000} & - & - & - \\ \widehat{\mu}_{10100} & - & - & - \\ \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{bmatrix} & (C_1^{\mathcal{I}}, C_2^{\mathcal{I}}, C_1^{\sim}) \\ \begin{array}{l} (0, 0, 0) \\ (0, 0, 1) \\ (0, 1, 0) \\ (0, 1, 1) \\ (1, 0, 0) \\ (1, 0, 1) \\ (1, 1, 0) \\ (1, 1, 1) \end{array} \end{array} \cdot$$

$$(0, 0)(0, 1)(1, 0)(1, 1)$$

$$(N_1^{\mathcal{I}}, N_1^{\sim})$$

The complete list of $\widehat{\boldsymbol{\mu}}^r$ matrices corresponding to the 8 different $\{j^{\mathcal{I}}(i, j)\}$ is

$$\begin{array}{cccc}
\left[\begin{array}{c|ccc} \widehat{\mu}_{00000} & - & - & - \\ \widehat{\mu}_{00100} & - & - & - \\ \hline \widehat{\mu}_{01000} & - & - & - \\ \widehat{\mu}_{01100} & - & - & - \\ \hline \widehat{\mu}_{10000} & - & - & - \\ \widehat{\mu}_{10100} & - & - & - \\ \hline \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{array} \right], &
\left[\begin{array}{c|ccc} - & - & \widehat{\mu}_{00010} & - \\ - & - & \widehat{\mu}_{00110} & - \\ \hline \widehat{\mu}_{01000} & - & - & - \\ \widehat{\mu}_{01100} & - & - & - \\ \hline \widehat{\mu}_{10000} & - & - & - \\ \widehat{\mu}_{10100} & - & - & - \\ \hline \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{array} \right], &
\left[\begin{array}{c|ccc} \widehat{\mu}_{00000} & - & - & - \\ \widehat{\mu}_{00100} & - & - & - \\ \hline - & - & \widehat{\mu}_{01010} & - \\ - & - & \widehat{\mu}_{01100} & - \\ \hline \widehat{\mu}_{10000} & - & - & - \\ \widehat{\mu}_{10100} & - & - & - \\ \hline \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{array} \right], &
\left[\begin{array}{c|ccc} - & - & \widehat{\mu}_{00010} & - \\ - & - & \widehat{\mu}_{00110} & - \\ \hline - & - & \widehat{\mu}_{01010} & - \\ - & - & \widehat{\mu}_{01100} & - \\ \hline \widehat{\mu}_{10000} & - & - & - \\ \widehat{\mu}_{10100} & - & - & - \\ \hline \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{array} \right] \\
\left[\begin{array}{c|ccc} \widehat{\mu}_{00000} & - & - & - \\ \widehat{\mu}_{00100} & - & - & - \\ \hline \widehat{\mu}_{01000} & - & - & - \\ \widehat{\mu}_{01100} & - & - & - \\ \hline - & - & \widehat{\mu}_{10010} & - \\ - & - & \widehat{\mu}_{10110} & - \\ \hline \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{array} \right], &
\left[\begin{array}{c|ccc} - & - & \widehat{\mu}_{00010} & - \\ - & - & \widehat{\mu}_{00110} & - \\ \hline \widehat{\mu}_{01000} & - & - & - \\ \widehat{\mu}_{01100} & - & - & - \\ \hline - & - & \widehat{\mu}_{10010} & - \\ - & - & \widehat{\mu}_{10110} & - \\ \hline \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{array} \right], &
\left[\begin{array}{c|ccc} \widehat{\mu}_{00000} & - & - & - \\ \widehat{\mu}_{00100} & - & - & - \\ \hline - & - & \widehat{\mu}_{01010} & - \\ - & - & \widehat{\mu}_{01100} & - \\ \hline - & - & \widehat{\mu}_{10010} & - \\ - & - & \widehat{\mu}_{10110} & - \\ \hline \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{array} \right], &
\left[\begin{array}{c|ccc} - & - & \widehat{\mu}_{00010} & - \\ - & - & \widehat{\mu}_{00110} & - \\ \hline - & - & \widehat{\mu}_{01010} & - \\ - & - & \widehat{\mu}_{01100} & - \\ \hline - & - & \widehat{\mu}_{10010} & - \\ - & - & \widehat{\mu}_{10110} & - \\ \hline \widehat{\mu}_{11000} & - & \widehat{\mu}_{11010} & - \\ \widehat{\mu}_{11100} & - & \widehat{\mu}_{11110} & - \end{array} \right]
\end{array}$$

where the first $\widehat{\boldsymbol{\mu}}^r$ in this list can be recognized as corresponding to $j^{\mathcal{I}}(\cdot) \equiv 0$. To obtain a conservative critical value for Procedure \mathcal{G} , one can solve $P\{\mathbf{1}_2, 1, j^{\mathcal{I}}(\cdot)\} = 1 - \alpha$ for h_{LB} for each of the 8 functions $j^{\mathcal{I}}(\cdot)$; the maximum of these h_{LB} is a conservative critical value for Procedure \mathcal{G} .

As an example of solving $P\{\mathbf{1}_2, 1, j^{\mathcal{I}}(\cdot)\} = 1 - \alpha$, consider $j^{\mathcal{I}}(\cdot) \equiv 0$. The expression $P\{\mathbf{1}_2, 1, j^{\mathcal{I}}(\cdot) \equiv 0\}$ is the probability, under the zero mean homogeneity configuration $\boldsymbol{\mu} = (0)_{8 \times 8}$ with $\sigma_\epsilon = 1.0$, of selecting the last row as the best product design based on the reduced set of $\widehat{\mu}_{sfe,tz}$ corresponding to $j^{\mathcal{I}}(\cdot) \equiv 0$. The solution, h_{LB} , to $P\{\mathbf{1}_2, 1, j^{\mathcal{I}}(\cdot) \equiv 0\} = 1 - \alpha$ is the $100 \times (1 - \alpha)$ percentile of the random variable

$$T = \sqrt{10} \left(\max \left\{ \begin{array}{l} \widehat{\mu}_{00000}, \widehat{\mu}_{00100}, \widehat{\mu}_{01000}, \\ \widehat{\mu}_{01100}, \widehat{\mu}_{10000}, \widehat{\mu}_{10100}, \\ \min\{\widehat{\mu}_{11000}, \widehat{\mu}_{11010}\}, \\ \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \end{array} \right\} - \min\{\widehat{\mu}_{11100}, \widehat{\mu}_{11110}\} \right) / \sqrt{V},$$

where $V \sim \chi_{10}^2$. The critical points corresponding to the other three $\widehat{\boldsymbol{\mu}}^r$ matrices in the first row above are computed similarly.

Lastly, we mention that it is straightforward to show the second row of four matrices can be obtained from those in the first row by relabeling the high and low levels of the factor $N_1^{\mathcal{I}}$. In terms of T , the two sets are equivalent. Therefore, we need only compute h_{LB} corresponding to each of $\widehat{\boldsymbol{\mu}}^r$ matrices in the first row above. \square

The next corollary allows the determination of the minimum in (4.5) of Corollary 4.2 (and hence the lower bound in Theorem 4.1) for the special case of a *single interacting control factor*.

Corollary 4.3 For an orthogonal 2-level (full or fractional) factorial experiment in which only one design factor, $C_1^{\mathcal{I}}$, interacts with r noise factors, i.e., $\mathcal{I}^{\mathcal{I}} = \{0, 1\}$, $\mathcal{I}^{\sim} = \{0, 1\}^q$, and $\mathcal{J}^{\mathcal{I}} = \{0, 1\}^r$, then

$$P \{ \mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot) \} = P \{ 1, \mathbf{1}_q, \mathbf{j}_{\star}^{\mathcal{I}}(\cdot) \}$$

for any $\mathbf{j}^{\mathcal{I}}(\cdot) : \{0\} \rightarrow \{0, 1\}^r$ where $\mathbf{j}_{\star}^{\mathcal{I}}(0) \equiv \mathbf{0}_r$.

Suppose that, in addition, there is a *single interacting noise factor* then there are but two possible $\mathbf{j}^{\mathcal{I}}(\cdot)$ functions, i.e., $\mathbf{j}^{\mathcal{I}}(0) = 0$ and $\mathbf{j}^{\mathcal{I}}(0) = 1$. Corollary 4.3 states they are equal. In particular, as applied to Example 1, Corollary 4.3 guarantees that the two probabilities $P \{ 1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 0 \}$ and $P \{ 1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 1 \}$ are equal and hence the lower bound of Theorem 4.1 is equal to $P \{ 1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 0 \}$, say. Furthermore, as the discussion following Equation (4.1) shows, this lower bound is sharp because it is equal to $\min_{\boldsymbol{\mu}} P_{\boldsymbol{\mu}} \{ \text{CS} \}$. Next we address the general topic of determining when the lower bound (4.1) is sharp.

4.2 Checking Whether the Lower Bound (4.4) is Sharp

This subsection provides a method to check whether the lower bound of Theorem 4.1 is sharp. First, we describe how to check whether a given $P \{ \mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot) \}$ corresponds to the probability of correct selection at a certain parameter configuration. The conclusion depends only on the identity of the interacting control and noise factors and on the model m_{CN} that relates these factors. Recall that $\mathbf{M} = (m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}))$ is an $|\mathcal{I}^{\mathcal{I}}| \times |\mathcal{J}^{\mathcal{I}}|$ matrix ordered *lexicographically* in both the elements of $\mathbf{i}^{\mathcal{I}}$ and $\mathbf{j}^{\mathcal{I}}$.

Fix a vector of indices $\mathbf{I} = (\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim})$ which is regarded as the “best” product design in the probability calculation below. For the given $\mathbf{I}^{\mathcal{I}}$ and a given function $\mathbf{j}^{\mathcal{I}}(\cdot)$ from $\{ \mathbf{i}^{\mathcal{I}} \in \mathcal{I}^{\mathcal{I}} : \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}} \}$ to $\mathcal{J}^{\mathcal{I}}$, let $\mathbf{M}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\cdot))$ denote any $|\mathcal{I}^{\mathcal{I}}| \times |\mathcal{J}^{\mathcal{I}}|$ matrix, if one exists, such that

$$\mathbf{M}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\cdot)) \text{ satisfies model } m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}), \quad (4.8)$$

$$\mathbf{M}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) = 0 \text{ for } \mathbf{j}^{\mathcal{I}} \in \mathcal{J}^{\mathcal{I}}, \quad (4.9)$$

$$\mathbf{M}_k(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})) = 0 \text{ for } \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \text{ and} \quad (4.10)$$

$$\mathbf{M}_k(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \geq k \text{ for } \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}} \text{ and } \mathbf{j}^{\mathcal{I}} \neq \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}}) \quad (4.11)$$

In practice, k is a large positive number, say $k \geq 100$. In words, row $\mathbf{I}^{\mathcal{I}}$ of $\mathbf{M}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\cdot))$ consists of zeroes. For any other row $\mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}$, the entry in the $\mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})$ th column of row $\mathbf{i}^{\mathcal{I}}$ is zero

and all other entries in that same row are at least k (while the entire matrix must satisfy model $m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}})$).

Given $M_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\cdot))$, let $\boldsymbol{\mu}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)) = M_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\cdot)) \otimes J_{|\mathcal{I}^{\sim}| \times |\mathcal{J}^{\sim}|}$ denote the corresponding vector of means for a (conceptual) full factorial experiment in the control and noise factors. It is straightforward to check that the mean matrix $\boldsymbol{\mu}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot))$ satisfies $m_C(\mathbf{i}^{\sim}) = 0$ and $m_N(\mathbf{j}^{\sim}) = 0$. Notice that $\xi_{\mathbf{i}} = 0$ for all rows of $\boldsymbol{\mu}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot))$ and hence the product designs corresponding to these rows have $\xi_{\mathbf{i}}$ that are *equal*. However, intuition suggests that the row $\mathbf{I}^{\mathcal{I}}$ will tend to have smaller estimated $\xi_{\mathbf{i}}$ than the other $\xi_{\mathbf{i}} = 0$ rows because all of its entries are (tied at) zero, whereas other rows have exactly one zero entry and its remaining components are large (magnitude k) components. Theorem 4.4 states that the probability that Procedure \mathcal{G} correctly identifies the row $(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim})$ as the best product design under $\boldsymbol{\mu}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot))$ approaches $P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\}$ as k increases.

Theorem 4.4 Suppose that there exists a sequence $\{M_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\cdot))\}_{k=1}^{\infty}$ satisfying (4.8)–(4.11). Let $\boldsymbol{\mu}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)) = M_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\cdot)) \otimes J_{|\mathcal{I}^{\sim}| \times |\mathcal{J}^{\sim}|}$. Then,

$$P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\} = \lim_{k \rightarrow \infty} P_{\boldsymbol{\mu}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot))} \left\{ \hat{\xi}_{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}} \geq \max_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}^{\sim}} \hat{\xi}_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}^{\sim}} - hS \right\}. \quad (4.12)$$

Roughly speaking, Theorem 4.4 states that if

$$\boldsymbol{\mu}(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)) \equiv \lim_{k \rightarrow \infty} \boldsymbol{\mu}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)) = \lim_{k \rightarrow \infty} \{M_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\cdot))\} \otimes J_{|\mathcal{I}^{\sim}| \times |\mathcal{J}^{\sim}|}$$

exists, it is a “parameter configuration” for which the probability of correctly identifying $(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim})$ as the best control factor combination is equal to $P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\}$. The following corollary provides a sufficient condition that the lower bound in Theorem 4.1 be sharp.

Corollary 4.5 Suppose that $\mathbf{I}_{\star}^{\mathcal{I}}, \mathbf{I}_{\star}^{\sim}$ and $\mathbf{j}_{\star}^{\mathcal{I}}(\cdot)$ yield the smallest $P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\}$, i.e.,

$$P\{\mathbf{I}_{\star}^{\mathcal{I}}, \mathbf{I}_{\star}^{\sim}, \mathbf{j}_{\star}^{\mathcal{I}}(\cdot)\} = \min_{(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim})} \min_{\mathbf{j}^{\mathcal{I}}(\cdot)} P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\}.$$

In addition, suppose that there exists a sequence $\{M_k\}_{k=1}^{\infty}$ corresponding to $(\mathbf{I}_{\star}^{\mathcal{I}}, \mathbf{j}_{\star}^{\mathcal{I}}(\cdot))$ satisfying (4.8)–(4.11). Then

$$\boldsymbol{\mu}(\mathbf{I}_{\star}^{\mathcal{I}}, \mathbf{I}_{\star}^{\sim}, \mathbf{j}_{\star}^{\mathcal{I}}(\cdot)) = \lim_{k \rightarrow \infty} M_k \otimes J_{|\mathcal{I}^{\sim}| \times |\mathcal{J}^{\sim}|}$$

is a least favorable configuration and

$$P\{\mathbf{I}_{\star}^{\mathcal{I}}, \mathbf{I}_{\star}^{\sim}, \mathbf{j}_{\star}^{\mathcal{I}}(\cdot)\} = \min_{\boldsymbol{\mu}} P\boldsymbol{\mu}\{\text{CS}\}.$$

In sum, to check if the lower bound in Theorem 4.1 is sharp, one must first find the control factor levels $\mathbf{I}_*^{\mathcal{I}}, \mathbf{I}_*^{\sim}$, and the function $\mathbf{j}_*^{\mathcal{I}}(\cdot)$ that correspond to the smallest $P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\}$. Next, construct a sequence $\{\mathbf{M}_k\}_{k=1}^{\infty}$ corresponding to $(\mathbf{I}_*^{\mathcal{I}}, \mathbf{I}_*^{\sim}, \mathbf{j}_*^{\mathcal{I}}(\cdot))$ and satisfying (4.8)–(4.11) using the technique in Example 2. If the construction is successful, the lower bound $P\{\mathbf{I}_*^{\mathcal{I}}, \mathbf{I}_*^{\sim}, \mathbf{j}_*^{\mathcal{I}}(\cdot)\}$ is sharp and $\boldsymbol{\mu} = \lim_{k \rightarrow \infty} \mathbf{M}_k \otimes J_{|\mathcal{I}^{\sim}| \times |\mathcal{J}^{\sim}|}$ is an LFC.

In particular, the condition of Corollary 4.5 is satisfied if for every $P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\}$ there exists the required $\{\mathbf{M}_k = \mathbf{M}_k(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\cdot))\}_{k=1}^{\infty}$ sequence. By Corollary 4.2, there are usually a small number of distinct values for $P\{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot)\}$ in orthogonal two-level experiments, because only probabilities $(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}) = (\mathbf{1}_p, \mathbf{1}_q)$ need to be considered. Therefore, orthogonal two-level experiments are cases where it is feasible to check whether it is possible to construct an $\{\mathbf{M}_k(\mathbf{1}_p, \mathbf{j}^{\mathcal{I}}(\cdot))\}_{k=1}^{\infty}$ sequence for each $\mathbf{j}^{\mathcal{I}}(\cdot)$.

Example 1 (Continued) We have already used Corollaries 4.2 and 4.3 to establish that the lower bound of Theorem 4.1 is sharp. Now we show how to use Corollary 4.5 to establish this fact. Recall that there are only two $\mathbf{j}^{\mathcal{I}}(\cdot)$ functions possible for this example: $\mathbf{j}^{\mathcal{I}}(0) = 0$ and $\mathbf{j}^{\mathcal{I}}(0) = 1$. As in Subsection 3.2, we choose $\mathbf{M}_k = \mathbf{M}_k(1, \mathbf{j}^{\mathcal{I}}(0) = 0) = \begin{bmatrix} 0 & k \\ 0 & 0 \end{bmatrix}$ corresponding to $\mathbf{j}^{\mathcal{I}}(0) = 0$. Similarly, we choose $\mathbf{M}_k = \mathbf{M}_k(1, \mathbf{j}^{\mathcal{I}}(0) = 1) = \begin{bmatrix} k & 0 \\ 0 & 0 \end{bmatrix}$ corresponding to $\mathbf{j}^{\mathcal{I}}(0) = 1$. The smaller of $P\{1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 0\}$ and $P\{1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 1\}$ is thus a sharp lower bound. In this example, one does not have to consider both possible $\mathbf{j}^{\mathcal{I}}(\cdot)$ because Corollary 4.3 implies that $P\{1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 0\}$ and $P\{1, \mathbf{1}_2, \mathbf{j}^{\mathcal{I}}(0) = 1\}$ are equal. \square

Example 4 provides another case where Theorem 4.4 can be applied to show that the (possibly conservative) critical value h_{LB} obtained by setting the right-hand side (4.4) in Theorem 4.1 equal to $1 - \alpha$ is, in fact, sharp.

Example 4 (Continued) The critical value h_{LB} is exact if there exists a sequence $\{\mathbf{M}_k\}_{k=1}^{\infty}$ corresponding to each of the 8 probabilities $P\{\mathbf{1}_2, 1, \{\mathbf{j}^{\mathcal{I}}(\cdot)\}\}$ involved in the right-hand side of Theorem 4.1. As stated earlier, only four probabilities $P\{\mathbf{1}_2, 1, \{\mathbf{j}^{\mathcal{I}}(\cdot)\}\}$ can possibly be distinct. Using the technique in Example 2, it can be shown easily that

$$\mathbf{M}_k = \begin{bmatrix} 0 & 2k \\ 0 & k \\ 0 & k \\ 0 & 0 \\ (0, 1) \\ N_1^{\mathcal{I}} \end{bmatrix} \begin{matrix} (C_1^{\mathcal{I}}, C_2^{\mathcal{I}}) \\ (0, 0) \\ (0, 1) \\ (1, 0) \\ (1, 1) \end{matrix}$$

satisfies (4.8)–(4.11) and provides the required ν_k sequence for $(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}) = (\mathbf{1}_2, 1)$ and $\mathbf{j}^{\mathcal{I}}(i, j) \equiv 0$. It can be shown similarly that such a sequence exists for the other 3 choices of $\mathbf{j}^{\mathcal{I}}(\cdot)$ function. Therefore, h_{LB} is sharp and the mean sequence corresponding to $\{\mathbf{M}_k\}_{k=1}^{\infty}$ generates h_{LB} corresponding to a least favorable configuration. \square

5 Discussion

We conclude this paper by mentioning one simplification and several extensions of the results discussed above. There is a certain “robustness” in the calculation of the critical value h for this screening procedure in the sense described next. For 2-level experiments with a single control factor, Theorem 3.1 shows that the LFC of $\boldsymbol{\mu}$ depend only on the *numbers* of non-interacting control and noise variables and interacting part of the model (2.4)–(2.5), $m_{CN}(\cdot, \cdot)$. Thus the *same* critical value h is used for any experiment involving the same $m_{CN}(\cdot, \cdot)$, and the same numbers of non-interacting control and noise variables. This

In some applications, the response is bounded above by a *known* value. For example, in the taste test study in Example 1, the highest taste score is 8. The maximin criterion judges the quality of a product or process design by its lowest mean across the environmental conditions. If there is an upper bound on the means, say, U , then the design can be equivalently judged by the largest difference between the means across the environmental conditions and the upper bound U . The best product based on maximin criterion is *also* the product that minimizes the difference from its mean and U , across the environmental conditions.

If, in addition, L is a lower bound on the the means, then Theorem 3.1 can be improved by replacing each $+\infty$ by $U - L$. For example, in Theorem 3.1, the least favorable configuration is

$$\begin{bmatrix} 0 & U-L & \cdots & U-L \\ 0 & 0 & \cdots & 0 \end{bmatrix} \otimes J_{2^q \times 2^s}.$$

However, unless $U - L$ is “small” relative to the measurement error standard deviation σ_ϵ (see Equation 2.6), the resulting h will be close to the value determined directly from Theorem 3.1.

Finally we mention that this paper has not addressed the important problem of designing a complete or fractional factorial experiment when the objective is to select the (single) best control-factor combination. Rather we have focused on the problem of screening, a methodology that does not identify the best treatment but which attempts to find a subset containing the best treatment. See Bechhofer et al. (1995) for a discussion of design for selection in normal means and binomial success probability problems.

Acknowledgment

We would like to thank two referees and the Editor for comments that substantially improved the presentation of this paper.

References

- Bechhofer, R. E., Santner, T. J. and Goldsman, D. M. (1995). *Designing Experiments for Statistical Selection, Screening, and Multiple Comparisons*. New York: J. Wiley & Sons.
- Box, G. E. P. and Jones, S. (1992). Split-plot designs for robust product experimentation. *Journal of Applied Statistics*, **19**, 3-26.
- Gupta, S. S. (1956). *On A Decision Rule for a Problem in Ranking Mean*. Ph.D. Dissertation, Institute of Statistics, University of North Carolina, Chapel Hill, NC.
- Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics*, **7**, 225-245.
- Pan, G. and Santner, T. J. (1998). Selection and Screening Procedures to Determine Optimal Product Designs, *Journal of Statistical Planning and Inference*, **67** (2), 311-330.
- Santner, T. J. and Pan, G. (1997). The Use of Subset Selection in Combined-Array Experiments to Determine Optimal Product or Process Designs. *Advances in Statistical Decision Theory and Methodology*, Editors: Panchapakesan, S. and Balakrishnan, N., 417-430.

A Proofs

First, we establish that the least favorable configuration must occur when the effects of the non-interacting control and non-interacting noise factors are zero.

A.1 Location of the LFC

Lemma A.1 Assuming that the design \mathcal{D} is such that all non-zero main effects and interactions are estimable, independent of non-zero confounding factors,

$$\inf_{\Omega} P_{\boldsymbol{\mu}} \{CS\} = \inf_{\Omega_0} P_{\boldsymbol{\mu}} \{CS\} = \inf_{\Omega_{00}} P_{\boldsymbol{\mu}} \{CS\} \quad (\text{A.1})$$

where

$$\Omega = \{\boldsymbol{\mu} : (2.4) \text{ holds}\}, \quad \Omega_0 = \{\boldsymbol{\mu} \in \Omega : m_N(\mathbf{j}^\sim) \equiv 0\}, \quad \Omega_{00} = \{\boldsymbol{\mu} \in \Omega_0 : m_C(\mathbf{i}^\sim) \equiv 0\}.$$

Proof: Recall that we use lexicographic order to write any $\boldsymbol{\mu}$ satisfying the model (2.4) as an $|\mathcal{I}|$ by $|\mathcal{J}|$ ($= |\mathcal{I}^\mathcal{I}| \times |\mathcal{I}^\sim|$ by $|\mathcal{J}^\mathcal{I}| \times |\mathcal{J}^\sim|$) matrix (Section 2). Thus such a $\boldsymbol{\mu}$ can be uniquely expressed in the form

$$\boldsymbol{\mu} = \mathbf{m}_{CN} \otimes J_{|\mathcal{I}^\sim| \times |\mathcal{J}^\sim|} + J_{|\mathcal{I}^\mathcal{I}| \times 1} \otimes \mathbf{m}_C \otimes J_{1 \times |\mathcal{J}|} + J_{|\mathcal{I}| \times |\mathcal{J}^\mathcal{I}|} \otimes \mathbf{m}_N^\top \quad (\text{A.2})$$

where $\mathbf{m}_{CN} = (m_{CN}(\mathbf{i}^\mathcal{I}, \mathbf{j}^\mathcal{I}))$ is $|\mathcal{I}^\mathcal{I}| \times |\mathcal{J}^\mathcal{I}|$, $\mathbf{m}_C = (m_C(\mathbf{i}^\sim))$, is $|\mathcal{I}^\sim| \times 1$, $\mathbf{m}_N = (m_N(\mathbf{j}^\sim))$ is $|\mathcal{J}^\sim| \times 1$. Notice that model component (2.5) specifies that $\mathbf{m}_{CN} \in \mathbf{V}_{CN}$, where

$$\mathbf{V}_{CN} = \{\mathbf{v} \in \mathbb{R}^{|\mathcal{I}^\mathcal{I}| \times |\mathcal{J}^\mathcal{I}|} : (2.5) \text{ holds}\}.$$

To prove the left-hand equality of (A.1), pick $\boldsymbol{\mu} \in \Omega$ and let $(\mathbf{m}_{CN}, \mathbf{m}_C, \mathbf{m}_N)$ denote the corresponding elements in (A.2). Define

$$\boldsymbol{\mu}^* = \boldsymbol{\mu} - J_{|\mathcal{I}| \times |\mathcal{J}^\mathcal{I}|} \otimes \mathbf{m}_N^\top.$$

If $(\mathbf{m}_{CN}^*, \mathbf{m}_C^*, \mathbf{m}_N^*)$ correspond to $\boldsymbol{\mu}^*$, then it straightforward to show that $\mathbf{m}_{CN}^* = \mathbf{m}_{CN}$, $\mathbf{m}_C^* = \mathbf{m}_C$, and $\mathbf{m}_N^* \equiv 0$, which shows that $\boldsymbol{\mu}^* \in \Omega_0$. We claim that $P_{\boldsymbol{\mu}} \{CS\} = P_{\boldsymbol{\mu}^*} \{CS\}$ which will establish the left-hand equality in (A.1).

Consider the relationship between the row minimums for $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^*$; let ξ_i correspond to $\boldsymbol{\mu}$ and ξ_i^* correspond to $\boldsymbol{\mu}^*$. Fix $\mathbf{i} = (\mathbf{i}^\mathcal{I}, \mathbf{i}^\sim)$; then

$$\begin{aligned} \xi_i &= \min_{(\mathbf{j}^1, \mathbf{j}^2)} \{m_{CN}(\mathbf{i}^\mathcal{I}, \mathbf{j}^1) + m_C(\mathbf{i}^\sim) + m_N(\mathbf{j}^2)\} \\ &= \min_{\mathbf{j}^1} \{m_{CN}(\mathbf{i}^\mathcal{I}, \mathbf{j}^1) + m_C(\mathbf{i}^\sim)\} + \min_{\mathbf{j}^2} \{m_N(\mathbf{j}^2)\} \\ &= \xi_i^* + Q, \end{aligned} \quad (\text{A.3})$$

say, where Q is independent of i . Using this fact, it is easy to see that the index i that maximizes ξ_i is identical for both $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^*$ and the relationship between the row minimums is

$$\xi_{[I]} \equiv \max \{ \xi_i \} = \max \{ \xi_i^* + Q \} = \xi_{[I]}^* + Q$$

Let $\boldsymbol{\mu}_{\mathcal{D}}$ and $\boldsymbol{\mu}_{\mathcal{D}}^*$ consist of the components of $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^*$ corresponding to the design \mathcal{D} in (2.6); similarly let $\mathbf{Y}_{\mathcal{D}} = \boldsymbol{\mu}_{\mathcal{D}} + \boldsymbol{\epsilon}_{\mathcal{D}}$ and $\mathbf{Y}_{\mathcal{D}}^* = \boldsymbol{\mu}_{\mathcal{D}}^* + \boldsymbol{\epsilon}_{\mathcal{D}}$ where the elements of $\boldsymbol{\epsilon}_{\mathcal{D}}$ are independent $N(0, \sigma_{\epsilon}^2)$ random variables. We determine the relationship between the *estimated* row minimums for the two models. Let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\mu}}^*$ denote the estimated means using the data $\mathbf{Y}_{\mathcal{D}}$ and $\mathbf{Y}_{\mathcal{D}}^*$, respectively, and let $\hat{\xi}_i$ and $\hat{\xi}_i^*$ denote the corresponding row minimums. We have $\hat{\boldsymbol{\mu}}_{i,j} = \widehat{\mathbf{m}}_{CN}(\mathbf{i}^I, \mathbf{j}^I) + \widehat{\mathbf{m}}_C(\mathbf{i}^{\sim}) + \widehat{\mathbf{m}}_N(\mathbf{j}^{\sim})$ where each of the three terms on the right-hand side of $\hat{\boldsymbol{\mu}}_{i,j}$ is the sum of estimated main effects and interactions. In particular, $(\widehat{\mathbf{m}}_{CN}, \widehat{\mathbf{m}}_C, \widehat{\mathbf{m}}_N)$ minimizes

$$\sum_{\mathcal{D}} \left\{ Y_{i,j} - m_{CN}^a(\mathbf{i}^I, \mathbf{j}^I) - m_C^a(\mathbf{i}^{\sim}) - m_N^a(\mathbf{j}^{\sim}) \right\}^2$$

where the summation is over indices \mathcal{D} in the experimental region and the superscript a denotes that these are the variables over which the minimization is to take place. To derive $\hat{\boldsymbol{\mu}}^*$, we observe that

$$\begin{aligned} & \sum_{\mathcal{D}} \left\{ Y_{i,j}^* - m_{CN}^a(\mathbf{i}^I, \mathbf{j}^I) - m_C^a(\mathbf{i}^{\sim}) - m_N^a(\mathbf{j}^{\sim}) \right\}^2 \\ &= \sum_{\mathcal{D}} \left\{ Y_{i,j} - m_N(\mathbf{j}^{\sim}) - m_{CN}^a(\mathbf{i}^I, \mathbf{j}^I) - m_C^a(\mathbf{i}^{\sim}) - m_N^a(\mathbf{j}^{\sim}) \right\}^2 \\ &= \sum_{\mathcal{D}} \left\{ Y_{i,j} - m_{CN}^a(\mathbf{i}^I, \mathbf{j}^I) - m_C^a(\mathbf{i}^{\sim}) - m_N^{aa}(\mathbf{j}^{\sim}) \right\}^2, \end{aligned}$$

say, where $m_N^{aa}(\mathbf{j}^{\sim}) = m_N(\mathbf{j}^{\sim}) + m_N^a(\mathbf{j}^{\sim})$. Thus we see that the OLS estimators of the terms that comprise $\boldsymbol{\mu}^*$ satisfy $\widehat{\mathbf{m}}_{CN}^* = \widehat{\mathbf{m}}_{CN}$, $\widehat{\mathbf{m}}_C^* = \widehat{\mathbf{m}}_C$, and $\widehat{\mathbf{m}}_N^* = \widehat{\mathbf{m}}_N - \mathbf{m}_N$; these equations imply that

$$\hat{\boldsymbol{\mu}}_{i,j}^* = \hat{\boldsymbol{\mu}}_{i,j} - m_N(\mathbf{j}^{\sim}). \quad (\text{A.4})$$

Equation A.4 implies two things. First, if S^2 is the residual sum of squares used to estimate of σ_{ϵ}^2 based on \mathbf{Y} and $(S^*)^2$ is the corresponding estimate based on \mathbf{Y}^* , then

$$\begin{aligned} \nu \times (S^*)^2 &= \sum_{i,j} (Y_{i,j}^* - \hat{\boldsymbol{\mu}}_{i,j}^*)^2 \\ &= \sum_{i,j} (Y_{i,j} - m_N(\mathbf{j}^{\sim}) - \hat{\boldsymbol{\mu}}_{i,j} + m_N(\mathbf{j}^{\sim}))^2 \\ &= \sum_{i,j} (Y_{i,j} - \hat{\boldsymbol{\mu}}_{i,j})^2 = \nu \times S^2, \end{aligned}$$

where ν is the degrees of freedom for S^2 and $(S^*)^2$. Second, for any $\mathbf{i} \in \mathcal{I}$ we have

$$\begin{aligned}
\widehat{\xi}_{\mathbf{i}}^* &= \min_{(\mathbf{j}^1, \mathbf{j}^2)} \{ \widehat{\mathbf{m}}_{CN}^*(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^1) + \widehat{\mathbf{m}}_C^*(\mathbf{i}^{\sim}) + \widehat{\mathbf{m}}_N^*(\mathbf{j}^2) \} \\
&= \min_{(\mathbf{j}^1, \mathbf{j}^2)} \{ \widehat{\mathbf{m}}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^1) + \widehat{\mathbf{m}}_C(\mathbf{i}^{\sim}) + \widehat{\mathbf{m}}_N(\mathbf{j}^2) \} \\
&\quad + \min_{\mathbf{j}^2} \{ \widehat{\mathbf{m}}_N(\mathbf{j}^2) - \mathbf{m}_N(\mathbf{j}^2) \} - \min_{\mathbf{j}^2} \{ \widehat{\mathbf{m}}_N(\mathbf{j}^2) \} \\
&= \widehat{\xi}_{\mathbf{i}} + \min_{\mathbf{j}^2} \{ \widehat{\mathbf{m}}_N(\mathbf{j}^2) - \mathbf{m}_N(\mathbf{j}^2) \} - \min_{\mathbf{j}^2} \{ \widehat{\mathbf{m}}_N(\mathbf{j}^2) \} \\
&= \widehat{\xi}_{\mathbf{i}} + T,
\end{aligned}$$

say, where T is independent of \mathbf{i} . Let $\widehat{\xi}_{(|\mathcal{I}|)}$ be associated with $\xi_{(|\mathcal{I}|)}$; then $\widehat{\xi}_{(|\mathcal{I}|)}^* = \widehat{\xi}_{(|\mathcal{I}|)} + T$ is associated with $\xi_{(|\mathcal{I}|)}^* = \xi_{(|\mathcal{I}|)} + Q$, where Q is given by (A.3), because the same product design \mathbf{i} simultaneously maximizes both $\xi_{\mathbf{i}}$ and $\xi_{\mathbf{i}}^* = \xi_{\mathbf{i}} + Q$. Thus we obtain

$$\begin{aligned}
P_{\boldsymbol{\mu}}\{\text{CS}\} &= P \left\{ \widehat{\xi}_{(|\mathcal{I}|)} \geq \widehat{\xi}_{\mathbf{i}} - hS \text{ for all } \mathbf{i} \right\} \\
&= P \left\{ \widehat{\xi}_{(|\mathcal{I}|)} + T \geq \widehat{\xi}_{\mathbf{i}} + T - hS \text{ for all } \mathbf{i} \right\} \\
&= P \left\{ \widehat{\xi}_{(|\mathcal{I}|)}^* \geq \widehat{\xi}_{\mathbf{i}}^* - hS \text{ for all } \mathbf{i} \right\} \\
&= P_{\boldsymbol{\mu}^*}\{\text{CS}\}
\end{aligned}$$

To prove the right-hand equality in (A.1), pick $\boldsymbol{\mu} \in \Omega_0$ with uniquely defined expansion terms $(\mathbf{m}_{CN}, \mathbf{m}_C)$; set

$$\boldsymbol{\mu}^* = \boldsymbol{\mu} - J_{|\mathcal{I}| \times 1} \otimes \mathbf{m}_C \otimes J_{1 \times |\mathcal{J}|}$$

As above, it straightforward to show that the corresponding terms for $\boldsymbol{\mu}^*$ satisfy $(\mathbf{m}_{CN}^*, \mathbf{m}_C^*, \mathbf{m}_N^*) = (\mathbf{m}_{CN}, 0, 0)$ so that $\boldsymbol{\mu}^* \in \Omega_{00}$. We show $P_{\boldsymbol{\mu}}\{\text{CS}\} = P_{\boldsymbol{\mu}^*}\{\text{CS}\}$.

The row minimums for $\boldsymbol{\mu}$, $\xi_{\mathbf{i}}$, and for $\boldsymbol{\mu}^*$, $\xi_{\mathbf{i}}^*$, satisfy the equation

$$\begin{aligned}
\xi_{\mathbf{i}} &= \min_{\mathbf{j}^1, \mathbf{j}^2} \{ m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^1) + m_C(\mathbf{i}^{\sim}) \} \\
&= \min_{\mathbf{j}^1} \{ m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^1) \} + m_C(\mathbf{i}^{\sim}) \\
&= \xi_{\mathbf{i}}^* + m_C(\mathbf{i}^{\sim});
\end{aligned}$$

notice that $\xi_{\mathbf{i}}^*$ is independent of \mathbf{i}^{\sim} . From this calculation, the optimal row levels are determined as follows. Suppose that $(\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim})$ is defined by

$$m_C(\mathbf{i}_N^{\sim}) = \max_{\mathbf{i}^{\sim}} m_C(\mathbf{i}^{\sim}) \text{ and } \min_{\mathbf{j}^1} m_{CN}(\mathbf{i}_N^{\mathcal{I}}, \mathbf{j}^1) = \max_{\mathbf{i}^{\mathcal{I}}} \left\{ m_0 + \min_{\mathbf{j}^1} m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^1) \right\}.$$

Then $(\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim})$ is the optimum product design under $\boldsymbol{\mu}$ and $(\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim})$ is the optimum product design under $\boldsymbol{\mu}^*$ for any \mathbf{i}_N^{\sim} and, in particular, for $\mathbf{i}_N^{\sim} = \mathbf{i}_N^{\sim}$. To see this we compute

$$\begin{aligned} \max_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}_N^{\sim}} \xi_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}_N^{\sim}} &= \max_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}_N^{\sim}} \left\{ \min_{\mathbf{j}^1} \{m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^1)\} + m_C(\mathbf{i}_N^{\sim}) \right\} \\ &= \max_{\mathbf{i}^{\mathcal{I}}} \left\{ \min_{\mathbf{j}^1} \{m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^1)\} \right\} + \max_{\mathbf{i}_N^{\sim}} \{m_C(\mathbf{i}_N^{\sim})\} = \xi_{\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim}} \end{aligned}$$

and

$$\begin{aligned} \max_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}_N^{\sim}} \xi_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}_N^{\sim}}^* &= \max_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}_N^{\sim}} \left\{ \min_{\mathbf{j}^1} \{m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^1)\} \right\} \\ &= \max_{\mathbf{i}^{\mathcal{I}}} \left\{ \min_{\mathbf{j}^1} \{m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^1)\} \right\} = \xi_{\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim}}^* \end{aligned}$$

for any \mathbf{i}_N^{\sim} and in particular for \mathbf{i}_N^{\sim} . Also note that $\xi_{\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim}} = \xi_{\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim}}^* + T$, say, where T is constant.

As above, let $\mathbf{Y}_{\mathcal{D}} = \boldsymbol{\mu}_{\mathcal{D}} + \boldsymbol{\epsilon}_{\mathcal{D}}$ and $\mathbf{Y}_{\mathcal{D}}^* = \boldsymbol{\mu}_{\mathcal{D}}^* + \boldsymbol{\epsilon}_{\mathcal{D}}$ where the components of $\boldsymbol{\epsilon}$ are independent $N(0, \sigma_{\epsilon}^2)$ random variables. Arguing similarly as the first part of the proof and letting quantities with(out) the superscript correspond to $\mathbf{Y}_{\mathcal{D}}^*$ ($\mathbf{Y}_{\mathcal{D}}$), the estimated row minimums for $\mathbf{Y}_{\mathcal{D}}$ and $\mathbf{Y}_{\mathcal{D}}^*$ can be shown to satisfy $\widehat{\xi}_{\mathbf{i}}^* = \widehat{\xi}_{\mathbf{i}} - m_C(\mathbf{i}_N^{\sim})$ for all \mathbf{i} and $(S^*)^2 = S^2$. Let $\widehat{\xi}_{(\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim})}$ be the estimator associated with $\xi_{(\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim})}$ and similarly for $\widehat{\xi}_{(\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim})}^*$. Thus we obtain

$$\begin{aligned} P_{\boldsymbol{\mu}}\{\text{CS}\} &= P \left\{ \widehat{\xi}_{(\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim})} \geq \widehat{\xi}_{\mathbf{i}} - hS \text{ for all } \mathbf{i} \right\} \\ &= P \left\{ \widehat{\xi}_{(\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim})}^* + m_C(\mathbf{i}_N^{\sim}) \geq \widehat{\xi}_{\mathbf{i}}^* + m_C(\mathbf{i}_N^{\sim}) - hS \text{ for all } \mathbf{i} \right\} \\ &\geq P \left\{ \widehat{\xi}_{(\mathbf{i}_N^{\mathcal{I}}, \mathbf{i}_N^{\sim})}^* \geq \widehat{\xi}_{\mathbf{i}}^* - hS \text{ for all } \mathbf{i} \right\} \\ &= P_{\boldsymbol{\mu}^*}\{\text{CS}\} \end{aligned} \tag{A.5}$$

where (A.5) holds because $m_C(\mathbf{i}_N^{\sim}) \geq m_C(\mathbf{i}_N^{\sim})$ for all \mathbf{i}_N^{\sim} . \square

A.2 Proof of Theorem 4.1

By Lemma A.1, it suffices to consider $\boldsymbol{\mu}$ of the form $\mathbf{v} \otimes J_{|\mathcal{I}^{\sim}| \times |\mathcal{J}^{\sim}|}$ with $\mathbf{v} = (\nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}}) \in \mathbf{V}_{CN}$. In this case

$$\mu_{\mathbf{i}, \mathbf{j}} = \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} \tag{A.6}$$

for any $\mathbf{i} = (\mathbf{i}^{\mathcal{I}}, \mathbf{i}_N^{\sim})$ and so

$$\xi_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}_N^{\sim}} = \min_{\mathbf{j}^{\mathcal{I}}} \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} \tag{A.7}$$

does not depend on i^\sim . Define the function $\{j^{\mathcal{I}}(i^{\mathcal{I}}) \in \mathcal{J}^{\mathcal{I}} : i^{\mathcal{I}} \in \mathcal{I}^{\mathcal{I}}, i^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}\}$ to be any set of column indices where the minimum on the right-hand side of (A.7) is attained, i.e., $\xi_{i^{\mathcal{I}}, i^\sim} = \nu_{i^{\mathcal{I}}, j^{\mathcal{I}}(i^{\mathcal{I}})}$. Finally, let $\mathbf{I} = (\mathbf{I}^{\mathcal{I}}, \mathbf{I}^\sim) \in \mathcal{I}$ denote an optimum control setting for μ , i.e., $\xi_{\mathbf{I}} = \max_{\mathbf{i}} \xi_{\mathbf{i}}$. From the proof of Lemma A.1,

$$\begin{aligned} \widehat{\mu}_{\mathbf{i}, \mathbf{j}} &= \widehat{m}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) + \widehat{m}_C(\mathbf{i}^\sim) + \widehat{m}_N(\mathbf{j}^\sim) \\ &= \mu_{\mathbf{i}, \mathbf{j}} + \bar{\epsilon}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) + \bar{\epsilon}_C(\mathbf{i}^\sim) + \bar{\epsilon}_N(\mathbf{j}^\sim), \end{aligned} \quad (\text{A.8})$$

where $\mu_{\mathbf{i}, \mathbf{j}} = \nu_{i^{\mathcal{I}}, j^{\mathcal{I}}}$ by (A.6), and $\bar{\epsilon}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}})$ is the linear combination of the $\epsilon_{\mathbf{i}, \mathbf{j}}$ determined by the terms in (2.5). Similarly each of $\bar{\epsilon}_C(\mathbf{i}^\sim)$ and $\bar{\epsilon}_N(\mathbf{j}^\sim)$ are linear combinations of the $\epsilon_{\mathbf{i}, \mathbf{j}}$ whose values depend on i^\sim and j^\sim , respectively. Thus the estimated $\xi_{\mathbf{i}}$ are

$$\widehat{\xi}_{\mathbf{i}} = \min_{\mathbf{j}} \widehat{\mu}_{\mathbf{i}, \mathbf{j}} = \bar{\epsilon}_C(\mathbf{i}^\sim) + \min_{\mathbf{j}^\sim} \bar{\epsilon}_N(\mathbf{j}^\sim) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \nu_{i^{\mathcal{I}}, j^{\mathcal{I}}} + \bar{\epsilon}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\}. \quad (\text{A.9})$$

The probability of correct selection is

$$\begin{aligned} P_{\mu} \{\text{CS}\} &= P_{\mu} \left\{ \widehat{\xi}_{\mathbf{I}} \geq \widehat{\xi}_{\mathbf{i}} - hS \quad \forall \mathbf{i} \in \mathcal{I} \right\} \\ &= P_{\mu} \left\{ \begin{aligned} \widehat{\xi}_{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^\sim} &\geq \widehat{\xi}_{\mathbf{I}^{\mathcal{I}}, \mathbf{i}^\sim} - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ \widehat{\xi}_{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^\sim} &\geq \widehat{\xi}_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}^\sim} - hS \quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^\sim \in \mathcal{I}^\sim \end{aligned} \right\} \\ &= P \left\{ \begin{aligned} \bar{\epsilon}_C(\mathbf{I}^\sim) &\geq \bar{\epsilon}_C(\mathbf{i}^\sim) - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ \bar{\epsilon}_C(\mathbf{I}^\sim) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \nu_{\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} + \bar{\epsilon}_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} &\geq \\ \bar{\epsilon}_C(\mathbf{i}^\sim) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} + \bar{\epsilon}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} - hS &\quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^\sim \in \mathcal{I}^\sim \end{aligned} \right\} \quad (\text{A.10}) \end{aligned}$$

$$\geq P \left\{ \begin{aligned} \bar{\epsilon}_C(\mathbf{I}^\sim) &\geq \bar{\epsilon}_C(\mathbf{i}^\sim) - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ \bar{\epsilon}_C(\mathbf{I}^\sim) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \nu_{\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{I}^{\mathcal{I}})} + \bar{\epsilon}_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} &\geq \\ \bar{\epsilon}_C(\mathbf{i}^\sim) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} + \bar{\epsilon}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} - hS &\quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^\sim \in \mathcal{I}^\sim \end{aligned} \right\} \quad (\text{A.11})$$

$$\geq P \left\{ \begin{aligned} \bar{\epsilon}_C(\mathbf{I}^\sim) &\geq \bar{\epsilon}_C(\mathbf{i}^\sim) - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ \bar{\epsilon}_C(\mathbf{I}^\sim) + \nu_{\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{I}^{\mathcal{I}})} + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \bar{\epsilon}_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} &\geq \\ \bar{\epsilon}_C(\mathbf{i}^\sim) + \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})} + \bar{\epsilon}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})) - hS &\quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^\sim \in \mathcal{I}^\sim \end{aligned} \right\} \quad (\text{A.12})$$

$$\geq P \left\{ \begin{aligned} \bar{\epsilon}_C(\mathbf{I}^\sim) &\geq \bar{\epsilon}_C(\mathbf{i}^\sim) - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ \bar{\epsilon}_C(\mathbf{I}^\sim) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \bar{\epsilon}_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} &\geq \\ \bar{\epsilon}_C(\mathbf{i}^\sim) + \bar{\epsilon}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})) - hS &\quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^\sim \in \mathcal{I}^\sim \end{aligned} \right\} \quad (\text{A.13})$$

$$= P_{\mu_0} \left\{ \begin{aligned} \widehat{\xi}_{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^\sim} &\geq \min_{\mathbf{j}^\sim} \left\{ \widehat{\mu}_{\mathbf{i}^{\mathcal{I}}, \mathbf{i}^\sim, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}}), \mathbf{j}^\sim} \right\} - hS \quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ \widehat{\xi}_{\mathbf{I}^{\mathcal{I}}, \mathbf{I}^\sim} &\geq \widehat{\xi}_{\mathbf{I}^{\mathcal{I}}, \mathbf{i}^\sim} - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim \end{aligned} \right\} \quad (\text{A.14})$$

$$= P \left\{ \mathbf{I}^{\mathcal{I}}, \mathbf{I}^\sim, \mathbf{j}^{\mathcal{I}}(\cdot) \right\} \quad (\text{A.15})$$

where (A.10) holds by substituting (A.9) and cancelling common terms, (A.11) holds because $\nu_{\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}}(\mathbf{I}^{\mathcal{I}}) \leq \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}}$ for all $\mathbf{j}^{\mathcal{I}}$, (A.12) holds because

$$\nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}}(\mathbf{i}^{\mathcal{I}}) + \bar{e}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})) \geq \min_{\mathbf{j}^{\mathcal{I}}} \{\nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} + \bar{e}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}})\},$$

and (A.13) holds because $\nu_{\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}}(\mathbf{I}^{\mathcal{I}}) \geq \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}}(\mathbf{i}^{\mathcal{I}})$ holds for all $\mathbf{i}^{\mathcal{I}}$. Equation (A.14) follows from Equation (A.9). Theorem 4.1 follows directly by examining the inequality between the first and last terms in the above sequence of equations and inequalities.

A.3 Proof of Theorem 4.4

According to Equation (A.10),

$$\begin{aligned} & P_{\mu_k(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot))} \{\text{CS}\} \\ &= P \left\{ \begin{array}{l} \bar{e}_C(\mathbf{I}^{\sim}) \geq \bar{e}_C(\mathbf{i}^{\sim}) - hS \quad \forall \mathbf{i}^{\sim} \in \mathcal{I}^{\sim}; \\ \bar{e}_C(\mathbf{I}^{\sim}) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \nu_{\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} + \bar{e}_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} \geq \\ \bar{e}_C(\mathbf{i}^{\sim}) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} + \bar{e}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} - hS \\ \quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^{\sim} \in \mathcal{I}^{\sim} \end{array} \right\} \end{aligned} \quad (\text{A.16})$$

where $m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}})$ are the entries of the matrix $\mathbf{M}_k(\mathbf{I}^{\mathcal{I}}, \{\mathbf{j}^{\mathcal{I}}(\cdot)\})$. Notice that all $m_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) = 0$ and $m_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})) = 0$ for $\mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}$. Therefore, the right-hand probability in (A.16) is equal to

$$P \left\{ \begin{array}{l} \bar{e}_C(\mathbf{I}^{\sim}) \geq \bar{e}_C(\mathbf{i}^{\sim}) - hS \quad \forall \mathbf{i}^{\sim} \in \mathcal{I}^{\sim}; \\ \bar{e}_C(\mathbf{I}^{\sim}) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \bar{e}_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} \geq \\ \bar{e}_C(\mathbf{i}^{\sim}) + \min \left\{ \bar{e}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})), \min_{\mathbf{j}^{\mathcal{I}} \neq \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})} \left\{ \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} + \bar{e}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} \right\} - hS \\ \quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^{\sim} \in \mathcal{I}^{\sim} \end{array} \right\}.$$

The term $\min_{\mathbf{j}^{\mathcal{I}} \neq \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})} \left\{ \nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} + \bar{e}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} \rightarrow \infty$ a.s. as $k \rightarrow \infty$ because $\nu_{\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}} \geq k$ for $\mathbf{j}^{\mathcal{I}} \neq \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})$. Therefore, as $k \rightarrow \infty$, the probability above converges to

$$P \left\{ \begin{array}{l} \bar{e}_C(\mathbf{I}^{\sim}) \geq \bar{e}_C(\mathbf{i}^{\sim}) - hS \quad \forall \mathbf{i}^{\sim} \in \mathcal{I}^{\sim}; \\ \bar{e}_C(\mathbf{I}^{\sim}) + \min_{\mathbf{j}^{\mathcal{I}}} \left\{ \bar{e}_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \right\} \geq \\ \bar{e}_C(\mathbf{i}^{\sim}) + \bar{e}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})) - hS \quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^{\sim} \in \mathcal{I}^{\sim} \end{array} \right\} = P \{ \mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot) \}.$$

Theorem 4.4 is thus proved.

A.4 Proof of Corollary 4.2

It suffices to prove that $\min_{\mathbf{j}^{\mathcal{I}}(\cdot)} P \{ \mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot) \}$ is unchanged when the level of a *single* component of $(\mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim})$ is changed. By hypothesis, a component can change only from 0 to 1 or vice versa. Successive application of this fact leads to the general conclusion of Corollary 4.2. In the following, we discuss $\mathbf{I}^{\mathcal{I}}$ and \mathbf{I}^{\sim} separately; we first show

(1) *The value of $\min_{\mathbf{j}^{\mathcal{I}}(\cdot)} P \{ \mathbf{I}^{\mathcal{I}}, \mathbf{I}^{\sim}, \mathbf{j}^{\mathcal{I}}(\cdot) \}$ is unchanged when the level of a single component of \mathbf{I}^{\sim} changes.*

Without loss of generality, we only need to show, for any fixed $\mathbf{I}^{\mathcal{I}}$ and $(I_1^{\sim}, \dots, I_{q-1}^{\sim})$, that

$$\min_{\mathbf{j}^{\mathcal{I}}(\cdot)} P \{ \mathbf{I}^{\mathcal{I}}, (I_1^{\sim}, \dots, I_{q-1}^{\sim}, 0), \mathbf{j}^{\mathcal{I}}(\cdot) \} = \min_{\mathbf{j}^{\mathcal{I}}(\cdot)} P \{ \mathbf{I}^{\mathcal{I}}, (I_1^{\sim}, \dots, I_{q-1}^{\sim}, 1), \mathbf{j}^{\mathcal{I}}(\cdot) \}.$$

This follows from the stronger result that

$$P \{ \mathbf{I}^{\mathcal{I}}, (I_1^{\sim}, \dots, I_{q-1}^{\sim}, 0), \mathbf{j}^{\mathcal{I}}(\cdot) \} = P \{ \mathbf{I}^{\mathcal{I}}, (I_1^{\sim}, \dots, I_{q-1}^{\sim}, 1), \mathbf{j}^{\mathcal{I}}(\cdot) \} \quad (\text{A.17})$$

for any function $\mathbf{j}^{\mathcal{I}}(\cdot) : \{ \mathbf{i}^{\mathcal{I}} \in \mathcal{I}^{\mathcal{I}} : \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}} \} \rightarrow \mathcal{J}^{\mathcal{I}}$. By equalities (A.14) and (A.15), equality (A.17) is equivalent to

$$\begin{aligned} & P \left\{ \begin{array}{l} \bar{\epsilon}_C(I_1^{\sim}, \dots, I_{q-1}^{\sim}, 0) \geq \bar{\epsilon}_C(\mathbf{i}^{\sim}) - hS \quad \forall \mathbf{i}^{\sim} \in \mathcal{I}^{\sim}; \\ \bar{\epsilon}_C(I_1^{\sim}, \dots, I_{q-1}^{\sim}, 0) + \min_{\mathbf{j}^{\mathcal{I}}} \{ \bar{\epsilon}_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \} \geq \\ \bar{\epsilon}_C(\mathbf{i}^{\sim}) + \bar{\epsilon}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})) - hS \quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^{\sim} \in \mathcal{I}^{\sim} \end{array} \right\} \\ &= P \left\{ \begin{array}{l} \bar{\epsilon}_C(I_1^{\sim}, \dots, I_{q-1}^{\sim}, 1) \geq \bar{\epsilon}_C(\mathbf{i}^{\sim}) - hS \quad \forall \mathbf{i}^{\sim} \in \mathcal{I}^{\sim}; \\ \bar{\epsilon}_C(I_1^{\sim}, \dots, I_{q-1}^{\sim}, 1) + \min_{\mathbf{j}^{\mathcal{I}}} \{ \bar{\epsilon}_{CN}(\mathbf{I}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}) \} \geq \\ \bar{\epsilon}_C(\mathbf{i}^{\sim}) + \bar{\epsilon}_{CN}(\mathbf{i}^{\mathcal{I}}, \mathbf{j}^{\mathcal{I}}(\mathbf{i}^{\mathcal{I}})) - hS \quad \forall \mathbf{i}^{\mathcal{I}} \neq \mathbf{I}^{\mathcal{I}}, \mathbf{i}^{\sim} \in \mathcal{I}^{\sim} \end{array} \right\}. \end{aligned} \quad (\text{A.18})$$

To prove Equation (A.18), we argue that the probabilities on the two sides of the equation are the *same* functions of two vectors of random variables having the *same* joint distribution. To see this, consider the following two vectors of random variables from the following sets of random variables:

$$\begin{aligned} & \{ \bar{\epsilon}_C(I_1^{\sim}, \dots, I_{q-1}^{\sim}, 0), \bar{\epsilon}_C(I_1^{\sim}, \dots, I_{q-1}^{\sim}, 1) \}, \\ & \{ (\bar{\epsilon}_C(i_1^{\sim}, \dots, i_{q-1}^{\sim}, 0), \bar{\epsilon}_C(i_1^{\sim}, \dots, i_{q-1}^{\sim}, 1)), (i_1^{\sim}, \dots, i_{q-1}^{\sim}) \neq (I_1^{\sim}, \dots, I_{q-1}^{\sim}) \} \end{aligned}$$

and

$$\begin{aligned} & \{ \bar{\epsilon}_C(I_1^{\sim}, \dots, I_{q-1}^{\sim}, 1), \bar{\epsilon}_C(I_1^{\sim}, \dots, I_{q-1}^{\sim}, 0) \}, \\ & \{ (\bar{\epsilon}_C(i_1^{\sim}, \dots, i_{q-1}^{\sim}, 1), \bar{\epsilon}_C(i_1^{\sim}, \dots, i_{q-1}^{\sim}, 0)), (i_1^{\sim}, \dots, i_{q-1}^{\sim}) \neq (I_1^{\sim}, \dots, I_{q-1}^{\sim}) \} \end{aligned}$$

where the same ordering of the $(i_1^\sim, \dots, i_{q-1}^\sim)$, say lexicographic order, is used to form the top and bottom vectors. We claim that these two vectors of random variables have the same multivariate normal distribution. It is straightforward to see that both sets of random variables have a zero mean vector. The variances of corresponding random variables in the two sets are same since each is proportional to the number of relevant main effects and interaction effects in the model m_C , i.e., $|Q_C^C|$. Lastly, both sets of random variables have the same correlation structure; the correlation between any two different random variables $\bar{\epsilon}_C(i^\sim)$ and $\bar{\epsilon}_C(i^{*\sim})$ is proportional to the number of main and interaction effects that are common to both variables minus the number of main and interaction effects that are different in the two variables. Because a simultaneous change of the levels of the same component of i^\sim and $i^{*\sim}$ affects the corresponding main or interaction effects in $\bar{\epsilon}_C(i^\sim)$ and $\bar{\epsilon}_C(i^{*\sim})$, such a change leads to the same correlation between the resulting two $\bar{\epsilon}_C(\cdot)$ random variables. For example, the correlation between $\bar{\epsilon}_C(i_1^\sim, \dots, i_{q-1}^\sim, 0)$ and $\bar{\epsilon}_C(i_1^{*\sim}, \dots, i_{q-1}^{*\sim}, 1)$ is the same as that between $\bar{\epsilon}_C(i_1^\sim, \dots, i_{q-1}^\sim, 1)$ and $\bar{\epsilon}_C(i_1^{*\sim}, \dots, i_{q-1}^{*\sim}, 0)$.

In addition, it can be seen that both vectors above are independent of random variables $\{\bar{\epsilon}_{CN}(i^I, j^I) : i^I \in \{0, 1\}^p, j^I \in \{0, 1\}^r\}$ since the experiment is orthogonal. Because the probabilities on both sides of Equation (A.18) are the same functions of these two vectors of random variables augmented with random variables $\{\bar{\epsilon}_{CN}(i^I, j^I) : i^I \in \{0, 1\}^p, j^I \in \{0, 1\}^r\}$, Equation (A.18) is proved.

The proof of Corollary 4.2 is completed by showing

(2) *The value of $\min_{j^I(\cdot)} P\{I^I, I^\sim, j^I(\cdot)\}$ is constant under a change of the level of any one component of I^I .*

Without loss of generality, we will show that for any $(I_1^I, \dots, I_{p-1}^I)$ and I^\sim

$$\min_{j^I(\cdot)} P\{(I_1^I, \dots, I_{p-1}^I, 0), I^\sim, j^I(\cdot)\} = \min_{j^I(\cdot)} P\{(I_1^I, \dots, I_{p-1}^I, 1), I^\sim, j^I(\cdot)\}.$$

To prove this, one need only prove that a term on the left-hand minimum corresponds to a term on the right-hand minimum, and vice versa. Specifically, we show that, for a fixed function $j^I(\cdot) : \{i^I \in \mathcal{I}^I : i^I \neq (I_1^I, \dots, I_{p-1}^I, 0)\} \rightarrow \mathcal{J}^I$, there exists a (possibly different) function $j_*^I : \{i^I \in \mathcal{I}^I : i^I \neq (I_1^I, \dots, I_{p-1}^I, 1)\} \rightarrow \mathcal{J}^I$ such that

$$P\{(I_1^I, \dots, I_{p-1}^I, 0), I^\sim, j^I(\cdot)\} = P\{(I_1^I, \dots, I_{p-1}^I, 1), I^\sim, j_*^I(\cdot)\}.$$

Equivalently, we only need to show that

$$\begin{aligned}
& P \left\{ \begin{aligned} & \bar{\epsilon}_C(\mathbf{I}^\sim) \geq \bar{\epsilon}_C(\mathbf{i}^\sim) - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ & \bar{\epsilon}_C(\mathbf{I}^\sim) + \min_{\mathbf{j}^\mathcal{I}} \{ \bar{\epsilon}_{CN}((I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 0), \mathbf{j}^\mathcal{I}) \} \geq \\ & \bar{\epsilon}_C(\mathbf{i}^\sim) + \bar{\epsilon}_{CN}(\mathbf{i}^\mathcal{I}, \mathbf{j}^\mathcal{I}(\mathbf{i}^\mathcal{I})) - hS \quad \forall \mathbf{i}^\mathcal{I} \neq (I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 0), \mathbf{i}^\sim \in \mathcal{I}^\sim \end{aligned} \right\} \\
& = P \left\{ \begin{aligned} & \bar{\epsilon}_C(\mathbf{I}^\sim) \geq \bar{\epsilon}_C(\mathbf{i}^\sim) - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ & \bar{\epsilon}_C(\mathbf{I}^\sim) + \min_{\mathbf{j}^\mathcal{I}} \{ \bar{\epsilon}_{CN}((I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 1), \mathbf{j}_*^\mathcal{I}) \} \geq \\ & \bar{\epsilon}_C(\mathbf{i}^\sim) + \bar{\epsilon}_{CN}(\mathbf{i}^\mathcal{I}, \mathbf{j}_*^\mathcal{I}(\mathbf{i}^\mathcal{I})) - hS \\ & \quad \forall \mathbf{i}^\mathcal{I} \neq (I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 1), \mathbf{i}^\sim \in \mathcal{I}^\sim \end{aligned} \right\} \tag{A.19}
\end{aligned}$$

Define $\mathbf{j}_*^\mathcal{I}(\mathbf{i}^\mathcal{I}) = \mathbf{j}^\mathcal{I}(i_1^\mathcal{I}, \dots, i_{p-1}^\mathcal{I}, 1 - i_p^\mathcal{I})$ for all $\mathbf{i}^\mathcal{I} = (i_1^\mathcal{I}, \dots, i_p^\mathcal{I}) \in \mathcal{I}^\mathcal{I}$ with $\mathbf{i}^\mathcal{I} \neq (I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 1)$. Form two vectors of random variables from the following sets of random variables

$$\begin{aligned}
& \{ \bar{\epsilon}_{CN}((I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 0), \mathbf{j}^\mathcal{I}), \mathbf{j}^\mathcal{I} \in \{0, 1\}^r \}, \\
& \{ \bar{\epsilon}_{CN}((i_1^\mathcal{I}, \dots, i_p^\mathcal{I}), \mathbf{j}^\mathcal{I}(i_1^\mathcal{I}, \dots, i_p^\mathcal{I})), (i_1^\mathcal{I}, \dots, i_p^\mathcal{I}) \neq (I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 0) \}
\end{aligned}$$

and

$$\begin{aligned}
& \{ \bar{\epsilon}_{CN}((I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 1), \mathbf{j}^\mathcal{I}), \mathbf{j}^\mathcal{I} \in \{0, 1\}^r \}, \\
& \{ \bar{\epsilon}_{CN}((i_1^\mathcal{I}, \dots, i_{p-1}^\mathcal{I}, 1 - i_p^\mathcal{I}), \mathbf{j}^\mathcal{I}(i_1^\mathcal{I}, \dots, i_p^\mathcal{I})), (i_1^\mathcal{I}, \dots, i_p^\mathcal{I}) \neq (I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 0) \}
\end{aligned}$$

by using the same ordering of $\mathbf{j}^\mathcal{I}$ and $(i_1^\mathcal{I}, \dots, i_p^\mathcal{I})$. By arguing as above, it can be seen that both vectors have the same multivariate normal distribution.

Notice that, by definition of $\mathbf{j}_*^\mathcal{I}$, the bottom vector of random variables is the same as that formed from

$$\begin{aligned}
& \{ \bar{\epsilon}_{CN}((I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 1), \mathbf{j}^\mathcal{I}), \mathbf{j}^\mathcal{I} \in \{0, 1\}^r \}, \\
& \{ \bar{\epsilon}_{CN}((i_1^\mathcal{I}, \dots, i_{p-1}^\mathcal{I}, 1 - i_p^\mathcal{I}), \mathbf{j}_*^\mathcal{I}((i_1^\mathcal{I}, \dots, i_{p-1}^\mathcal{I}, 1 - i_p^\mathcal{I}))), (i_1^\mathcal{I}, \dots, i_p^\mathcal{I}) \neq (I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 0) \}
\end{aligned}$$

which, by a change in notation, is the same as that formed from

$$\begin{aligned}
& \{ \bar{\epsilon}_{CN}((I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 1), \mathbf{j}^\mathcal{I}), \mathbf{j}^\mathcal{I} \in \{0, 1\}^r \}, \\
& \left\{ \bar{\epsilon}_{CN}((i_1^{*\mathcal{I}}, \dots, i_p^{*\mathcal{I}}), \mathbf{j}_*^\mathcal{I}((i_1^{*\mathcal{I}}, \dots, i_p^{*\mathcal{I}}))), (i_1^{*\mathcal{I}}, \dots, i_p^{*\mathcal{I}}) \neq (I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 1) \right\},
\end{aligned}$$

where $(i_1^{*\mathcal{I}}, \dots, i_{p-1}^{*\mathcal{I}}) = (i_1^\mathcal{I}, \dots, i_{p-1}^\mathcal{I})$ and $i_p^{*\mathcal{I}} = 1 - i_p^\mathcal{I}$. In forming the vector, the ordering of the indices in $\{(i_1^{*\mathcal{I}}, \dots, i_p^{*\mathcal{I}}) \neq (I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 1)\}$ must correspond to those in $\{(i_1^\mathcal{I}, \dots, i_p^\mathcal{I}) \neq (I_1^\mathcal{I}, \dots, I_{p-1}^\mathcal{I}, 0)\}$.

It is easy to see that the two vectors of random variables above, are independent of the set of random variables $\{\bar{\epsilon}_C(\mathbf{i}^\sim), \mathbf{i}^\sim \in \mathcal{I}^\sim\}$. It can be checked that the probabilities on the two sides of Equation (A.19) are the *same* functions of these two vectors of random variables augmented by $\{\bar{\epsilon}_C(\mathbf{i}^\sim), \mathbf{i}^\sim \in \mathcal{I}^\sim\}$. Thus, Equation (A.19) and Corollary 4.2 are proved.

A.5 Proofs of Corollary 4.3 and Theorem 3.1

A.5.1 Proof of Corollary 4.3

It follows from Corollary 4.2 that we need only prove that all the probabilities involved in the minimum (4.5) are equal. Without loss of generality, take $\mathbf{I}^\mathcal{I} = 1$. Thus each function $\mathbf{j}^\mathcal{I}(\cdot)$ has degenerate domain $\{\mathbf{i}^\mathcal{I} \neq \mathbf{I}^\mathcal{I}\} = \{0\}$ and $\mathbf{j}^\mathcal{I}(0)$ identifies a single point in $\mathcal{J}^\mathcal{I}$.

To prove Corollary 4.3, it suffices to show that for any fixed \mathbf{I}^\sim , any $s = 1, \dots, r$, and any choice $(j_1^\mathcal{I}, \dots, j_{s-1}^\mathcal{I}, j_{s+1}^\mathcal{I}, \dots, j_r^\mathcal{I})$, that

$$\begin{aligned} & P \{1, \mathbf{I}^\sim, \mathbf{j}^\mathcal{I}(0) = (j_1^\mathcal{I}, \dots, j_{s-1}^\mathcal{I}, 0, j_{s+1}^\mathcal{I}, \dots, j_r^\mathcal{I})\} \\ &= P \{1, \mathbf{I}^\sim, \mathbf{j}^\mathcal{I}(0) = (j_1^\mathcal{I}, \dots, j_{s-1}^\mathcal{I}, 1, j_{s+1}^\mathcal{I}, \dots, j_r^\mathcal{I})\}. \end{aligned} \quad (\text{A.20})$$

Successive application of Equation (A.20) leads to

$$P \{1, \mathbf{I}^\sim, \mathbf{j}^\mathcal{I}(0) = (j_1^\mathcal{I}, \dots, j_r^\mathcal{I})\} = P \{1, \mathbf{I}^\sim, \mathbf{j}^\mathcal{I}(0) = (0, \dots, 0)\}$$

for any $(j_1^\mathcal{I}, \dots, j_r^\mathcal{I}) \in \{0, 1\}^r$. Corollary 4.3 then follows immediately from this fact and Theorem 4.2.

Without loss of generality, to prove Equation (A.20), we only provide details for the case when $s = r$ and $(j_1^\mathcal{I}, \dots, j_{r-1}^\mathcal{I}) = (1, \dots, 1)$. We show, for any fixed \mathbf{I}^\sim that

$$P \{1, \mathbf{I}^\sim, \mathbf{j}^\mathcal{I}(0) = (1, \dots, 1, 0)\} = P \{1, \mathbf{I}^\sim, \mathbf{j}^\mathcal{I}(0) = (1, \dots, 1, 1)\}.$$

By (A.14)-(A.15), this equation is equivalent to

$$\begin{aligned} & P \left\{ \begin{array}{l} \bar{\epsilon}_C(\mathbf{I}^\sim) \geq \bar{\epsilon}_C(\mathbf{i}^\sim) - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ \bar{\epsilon}_C(\mathbf{I}^\sim) + \min_{\mathbf{j}^\mathcal{I}} \{\bar{\epsilon}_{CN}(1, \mathbf{j}^\mathcal{I})\} \geq \\ \bar{\epsilon}_C(\mathbf{i}^\sim) + \bar{\epsilon}_{CN}(0, (1, \dots, 1, 0)) - hS, \quad \mathbf{i}^\sim \in \mathcal{I}^\sim \end{array} \right\} \\ &= P \left\{ \begin{array}{l} \bar{\epsilon}_C(\mathbf{I}^\sim) \geq \bar{\epsilon}_C(\mathbf{i}^\sim) - hS \quad \forall \mathbf{i}^\sim \in \mathcal{I}^\sim; \\ \bar{\epsilon}_C(\mathbf{I}^\sim) + \min_{\mathbf{j}^\mathcal{I}} \{\bar{\epsilon}_{CN}(1, \mathbf{j}^\mathcal{I})\} \geq \\ \bar{\epsilon}_C(\mathbf{i}^\sim) + \bar{\epsilon}_{CN}(0, (1, \dots, 1, 1)) - hS, \quad \mathbf{i}^\sim \in \mathcal{I}^\sim \end{array} \right\}. \end{aligned} \quad (\text{A.21})$$

The key step to prove (A.21) is the fact that the vectors formed from

$$\begin{aligned} & \bar{\epsilon}_{CN} (0, (1, \dots, 1, 0)), \\ & \{ [\bar{\epsilon}_{CN} (1, (j_1^{\mathcal{I}}, \dots, j_{r-1}^{\mathcal{I}}, 0)), \bar{\epsilon}_{CN} (0, (j_1^{\mathcal{I}}, \dots, j_{r-1}^{\mathcal{I}}, 1))] , (j_1^{\mathcal{I}}, \dots, j_{r-1}^{\mathcal{I}}) \in \{0, 1\}^{r-1} \}, \\ & \{ \bar{\epsilon}_C(\mathbf{i}^\sim), \mathbf{i}^\sim \in \mathcal{I}^\sim \} \end{aligned}$$

and

$$\begin{aligned} & \bar{\epsilon}_{CN} (0, (1, \dots, 1, 1)), \\ & \{ [\bar{\epsilon}_{CN} (1, (j_1^{\mathcal{I}}, \dots, j_{r-1}^{\mathcal{I}}, 1)), \bar{\epsilon}_{CN} (0, (j_1^{\mathcal{I}}, \dots, j_{r-1}^{\mathcal{I}}, 0))] , (j_1^{\mathcal{I}}, \dots, j_{r-1}^{\mathcal{I}}) \in \{0, 1\}^{r-1} \}, \\ & \{ \bar{\epsilon}_C(\mathbf{i}^\sim), \mathbf{i}^\sim \in \mathcal{I}^\sim \} \end{aligned}$$

have the same (multivariate normal) distribution. The same ordering of the $j_{r-1}^{\mathcal{I}} \in \{0, 1\}^{r-1}$ and $\mathbf{i}^\sim \in \mathcal{I}^\sim$ must be used in forming the vectors. To see this, first notice that each random variable in these vectors has mean vector zero. Second, any two variables in corresponding positions from the two vectors have the same variance; this is because the variance of any random variable in the two vectors is proportional to the number of relevant main effects and interaction effects in Model (2.4)–(2.5), i.e., $|Q_{\sim}^{CN}|$ or $|Q_{\sim}^C|$. Also notice that, in each vector, any random variables taken from the third set are mutually independent of random variables from the first two groups. The correlation between any two different $\bar{\epsilon}_{CN}$ random variables is proportional to the number of main and interaction effects identical to both variables minus the number of main and interaction effects different for both variables. As an example, the correlation between $\bar{\epsilon}_{CN} (0, (j_1^{\mathcal{I}}, \dots, j_{r-1}^{\mathcal{I}}, 0))$ and $\bar{\epsilon}_{CN} (0, (j_1^{*\mathcal{I}}, \dots, j_{r-1}^{*\mathcal{I}}, 0))$ is the same as that between $\bar{\epsilon}_{CN} (0, (j_1^{\mathcal{I}}, \dots, j_{r-1}^{\mathcal{I}}, 1))$ and $\bar{\epsilon}_{CN} (0, (j_1^{*\mathcal{I}}, \dots, j_{r-1}^{*\mathcal{I}}, 1))$, because a change of the level of $j_r^{\mathcal{I}}$ from 0 to 1 affects the same main or interaction effects involved in the two random variables in a pair.

Finally, it can be checked that the two sides of Equation (A.21) are probabilities of the same functions of these two vectors of random variables above and thus Equation (A.21) is proved.

A.5.2 Proof of Theorem 3.1

According to Corollary 4.3, the probability

$$P \{ \mathbf{I}^{\mathcal{I}}, \mathbf{I}^\sim, \mathbf{j}^{\mathcal{I}}(\cdot) \} = P \{ 1, \mathbf{1}_q, \mathbf{j}^{\mathcal{I}}(0) = 0 \}$$

is a lower bound for the infimum of the probability of correct selection. To investigate whether $P \{ 1, \mathbf{1}_q, \mathbf{j}^{\mathcal{I}}(0) = 0 \}$ corresponds to the $P_{\boldsymbol{\mu}}\{\text{CS}\}$ at some $\boldsymbol{\mu}$ satisfying the model (2.4)–(2.5)

parameter configuration, Theorem 4.4 states that one need only determine if it is possible to construct a sequence of

$$\mathbf{M}_k(1, \mathbf{j}^{\mathcal{I}}(0) = 0) = \begin{bmatrix} 0 & \star & \cdots & \star \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{2 \times 2^r},$$

matrices where each nonzero entry \star (possibly assuming different values) is at least k and satisfies model m_{CN} . If this can be done for a sequence of $k \rightarrow \infty$, then, for sufficiently large k , $\boldsymbol{\mu}_k(1, \mathbf{1}_q, \mathbf{j}^{\mathcal{I}}(0) = 0) = \mathbf{M}_k(1, \mathbf{j}^{\mathcal{I}}(0) = 0) \otimes J_{2^q \times 2^s}$ is a mean configuration for which the $P\{\text{CS}\}$ (effectively) equals the lower bound $P\{1, \mathbf{1}_q, \mathbf{j}^{\mathcal{I}}(0) = 0\}$. In other words,

$$\lim_{k \rightarrow \infty} \{\mathbf{M}_k(1, \mathbf{j}^{\mathcal{I}}(0) = 0)\} \otimes J_{2^q \times 2^s} = \begin{bmatrix} 0 + \infty & \cdots & + \infty \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{2 \times 2^r} \otimes J_{2^q \times 2^s},$$

is a least favorable configuration corresponding to $P\{1, \mathbf{1}_q, \mathbf{j}^{\mathcal{I}}(0) = 0\}$. Theorem 3.1 is thus proved.

A.6 Proof of Corollary 4.5

Corollary 4.5 follows from Corollary 4.3 and Theorem 4.4.