Statistica Sinica **22** (2012), 885-907
doi:http://dx.doi.org/10.5705/ss.2010.080

OPTIMAL DESIGNS FOR TWO-LEVEL FACTORIAL EXPERIMENTS WITH BINARY RESPONSE

Jie Yang[1], Abhyuday Mandal[2] and Dibyen Majumdar[1]

[1] *University of Illinois at Chicago and* [2] *University of Georgia*

Abstract: We consider the problem of obtaining locally D-optimal designs for factorial experiments with qualitative factors at two levels each and with binary response. For the 2^2 factorial experiment with main-effects model, we obtain optimal designs analytically in special cases and demonstrate how to obtain a solution in the general case using *cylindrical algebraic decomposition*. The optimal designs are shown to be robust to the choice of the assumed values of the prior, and when there is no basis to make an informed choice of the assumed values we recommend the use of the *uniform* design that assigns equal number of observations to each of the four points. For the general 2^k case we show that the uniform design has a *maximin* property.

Key words and phrases: Cylindrical algebraic decomposition, D-optimality, information matrix, full factorial design, generalized linear model, uniform design.

1. Introduction

The goal of many scientific and industrial experiments is to study a process that depends on several qualitative factors. We focus on the design of those experiments where the response is binary. When the response is quantitative and a linear model is appropriate, the design of the experiment is informed by the extensive literature on factorial experiments. On the other hand if the factors are quantitative and the response is binary, the literature on optimal design of generalized linear models in the approximate theory setup could be used. The goal of our work is to identify optimal and robust designs for factorial experiments with binary response.

Specific examples of experiments of interest are available. Smith (1932) described a bioassay for an anti-pneumococcus serum in which the explanatory variable was doses of the serum. Mice infected with pneumococcus were injected with different doses of the serum and the response was survival (or not) beyond seven days. Hamada and Nelder (1997) discussed the advantages of using a generalized linear model for discrete responses instead of linearizing the response to obtain an approximate linear model. They examined an industrial experiment on windshield molding performed at an IIT Thompson plant that was originally reported by Martin, Parker, and Zenick (1987). There were four factors each at two levels and the response was whether the part was *good* or not. A 2^{4-1} fractional factorial design was used with 1,000 runs at each experimental condition. Other examples include a seed gemination experiment described in Crowder (1978), and a sperm survival experiment in Myers, Montgomery, and Vining (2002). In a later section we examine a designed 2^2 experiment on the reproduction of plum trees, as reported by Hoblyn and Palmer (1934), with binary response.

We assume that the process under study is adequately described by a generalized linear model. While the theory we develop works for any link function, in examples and simulations we focus on the logit, probit, log-log, and complementary log-log links. The optimal designs is obtained using the D-criterion that maximizes the determinant of the inverse of the asymptotic covariance matrix of the estimators (the information matrix). In order to overcome the difficulty posed by the dependence of the design optimality criterion on the unknown parameters, we use the local optimality approach of Chernoff (1953) in which the parameters are replaced by assumed values. We refer the reader to the paper by Khuri et al. (2006) for details of the theory of designs for generalized linear models.

We assume that every factor is at two levels, a setup of particular interest in screening experiments, and that (for an experiment with k factors) we are interested in a *complete* 2^k experiment, one in which the design may be supported on all 2^k points. The model we choose may include a subset of all main effects and interactions. If we assume that the total number of observations is held fixed, then the design problem is to determine the proportion of observations allocated to each of the 2^k design points. It may be noted that if the response follows a standard linear model, then it follows from the results of Kiefer (1975) that the design which is *uniform* on the 2^k design points is universally optimal. For the problem restated in terms of weighing designs, Rao (1971) gave the optimality of the uniform design in terms of minimizing variances of each of the parameter estimators. It may be noted that the uniform design is an *orthogonal array* (Rao (1947)).

In this initial study, we focus primarily on the complete 2^2 experiment where the response is binary. While we do not find analytic solutions for D-optimal design for the general 2^2 experiment, we obtain characterizations for several special cases. For the general 2^2 experiment we demonstrate how a solution may be obtained by *Cylindrical Algebraic Decomposition* (CAD). Since local D-optimality depends on the assumed parameter values, we perform extensive simulations and demonstrate that the design is usually robust to the choice of the these values.

The optimality criterion can be written in terms of the variances, or information, at each of the 2^2 points. Note that these variances depend on the parameters through the link function. It turns out that the D-optimal design is quite different from the uniform design, especially when at least two of these variances are far from each other. Our general conclusions may be described as follows. If the experimenter has an approximate idea of the variances then the design obtained by using these values as the assumed values in the local D-criterion results is a highly efficient design. If the experimenter knows only that the variance at one point is substantially larger (this will be made precise) than the others, then the optimal design assigns one-third of observations to each of the three points with smaller variance and none to the one with the largest. In the absence of any prior idea of the variances our recommendation is to use the uniform design, which is quite robust in general. This strategy is examined for the logit, probit, log-log, and complimentary log-log link functions. It may be noted that for applications where a D-optimal design cannot be used, it can still serve as a benchmark to evaluate other designs.

For the general 2^k experiment we show that the uniform design is a maximin D-optimal design, i.e., a design that maximizes a lower bound of the D-criterion.

It may be noted that Graßhoff and Schwabe (2008) considered D-optimal designs for paired comparisons for logit links and in Section 5.2 of their paper they obtained a result for the optimality of the saturated design and uniform design.

In Section 2 we give the preliminary setup. In Section 3 we give results for the 2^2 experiment, in Section 4 we study robustness against misspecification of the assumed values and robustness of the uniform design, and in Section 5 we study an example. In Section 6 we consider the general 2^k experiment and we conclude with some remarks in Section 7. Proofs are relegated to the Appendix.

2. Preliminary Setup

Consider a 2^k experiment, an experiment with k explanatory variables at 2 levels each. Suppose n_i units are allocated to the ith experimental condition such that $n_i \geqslant 0$, $i = 1, \ldots, 2^k$, and $n_1 + \cdots + n_{2^k} = n$. We suppose that n is fixed and consider the problem of determining the "optimal" n_i's. In fact, we write our optimality criterion in terms of the proportions $p_i = n_i/n,\ i = 1, \ldots, 2^k$and determine the "optimal" p_i's. (Since n_i's are integers, an optimal design obtained in this fashion may not be "feasible" - an issue we do not deal with, except to say that a feasible solution "near" an optimal solution is expected to be "nearly optimal").

Suppose η is the linear predictor that involves main effects and interactions which are assumed to be in the model. For instance, for a 2^3 experiment with a model that includes the main effects and the two-factor interaction of factors 1 and 2, $\eta = \beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_{12} x_1 x_2$, where each $x_i \in \{-1, 1\}$. The aim of the experiment is to obtain inferences about the parameter vector of factor effects β; in the preceding example, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_{12})'$. Our main focus is the 2^2 experiment with main-effects model, in which case $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ and $\beta = (\beta_0, \beta_1, \beta_2)'$. In the framework of generalized linear models, the expectation of the response Y, $E(Y) = \pi$, is connected to the linear predictor η by the link function g: $\eta = g(\pi)$ (McCullagh and Nelder (1989)). For a binary response, the commonly used link functions are the logit, probit, log-log, and complimentary log-log links.

The maximum likelihood estimator of β has an asymptotic covariance matrix (Khuri et al. (2006)) that is the inverse of $nX'WX$, where $W = diag(w_1p_1, \ldots, w_{2^k}p_{2^k})$, $w_i = (d\pi_i/d\eta_i)^2 / (\pi_i(1-\pi_i)) \geq 0$, where η_i and π_i correspond to the ith observation for η and π defined before, and X is the "design matrix". For commonly used link functions,

$$w(\pi) = \begin{cases} \pi(1-\pi) & \text{for logit link,} \\ \frac{\left[\phi\left(\Phi^{-1}(\pi)\right)\right]^2}{\pi(1-\pi)} & \text{for probit link,} \\ \left(\frac{\pi}{1-\pi}\right)\left(\log(\pi)\right)^2 & \text{for log-log and complementary log-log link.} \end{cases}$$

For a main-effect 2^2 experiment, for instance, $X = ((1,1,1,1)', (1,1,-1,-1)', (1,-1,1,-1)')$. In this case, for the logit link, $w_1 = e^{\beta_0+\beta_1+\beta_2} / \left(1 + e^{\beta_0+\beta_1+\beta_2}\right)^2$, $w_2 = e^{\beta_0+\beta_1+\beta_2} / \left(e^{\beta_2} + e^{\beta_0+\beta_1}\right)^2$, $w_3 = e^{\beta_0+\beta_1+\beta_2} / \left(e^{\beta_1} + e^{\beta_0+\beta_2}\right)^2$, and $w_4 = e^{\beta_0+\beta_1+\beta_2} / \left(e^{\beta_0} + e^{\beta_1+\beta_2}\right)^2$. Note that the matrix $X'WX$ may be viewed as the *per-observation* information matrix. The *D-optimality* criterion maximizes the determinant $|X'WX|$.

3. D-Optimal 2^2 Designs

Suppose $k = 2$. If we consider both the main effects and the interaction, the design problem is straightforward, because $|X'WX| = 256p_1p_2p_3p_4w_1w_2w_3w_4$. In this case, the D-optimal 2^2 design is the uniform, $p_1 = p_2 = p_3 = p_4 = 1/4$, regardless of the w_i's.

From now on, we consider the main effects only. For a main-effect plan with $k = 2$, the asymptotic information matrix is proportional to $X'WX$. It can be shown that $|X'WX|$ can be written as (except for the constant 16):

$$\begin{aligned} det(\mathbf{w}, \mathbf{p}) = G(\mathbf{p}) &= p_2w_2 \cdot p_3w_3 \cdot p_4w_4 + p_1w_1 \cdot p_3w_3 \cdot p_4w_4 \\ &\quad + p_1w_1 \cdot p_2w_2 \cdot p_4w_4 + p_1w_1 \cdot p_2w_2 \cdot p_3w_3, \end{aligned} \tag{3.1}$$

where $\mathbf{w} = (w_1, w_2, w_3, w_4)'$ and $\mathbf{p} = (p_1, p_2, p_3, p_4)'$. In this section, we consider the problem of maximizing $G(\mathbf{p})$ over all vectors $\mathbf{p}$ with $p_i \geq 0$ and $\sum_i p_i = 1$.

3.1. Analytic solutions to special cases and saturated design

It follows from Kiefer (1975) that if all the w_i's are equal then the uniform design ($p_1 = p_2 = p_3 = p_4 = 1/4$) is D-optimal. If one and only one of the w_i's is zero, then the optimal design is uniform over the design points that correspond to the nonzero w_i's, and if two or more w_i's are zero, then $G(\mathbf{p}) \equiv 0$.

From now on, we assume $w_i > 0,\ i = 1, 2, 3, 4$, which is always true under logit, probit, log-log, or complementary log-log link functions. Let $L(\mathbf{p}) = G(\mathbf{p})/(w_1w_2w_3w_4)$ and $v_i = 1/w_i,\ i = 1, 2, 3, 4$. The maximization problem (3.1) can be rewritten as that of maximizing

$$L(\mathbf{p}) = v_4p_1p_2p_3 + v_3p_1p_2p_4 + v_2p_1p_3p_4 + v_1p_2p_3p_4. \tag{3.2}$$

The solution always exists and is unique. Its existence is due to the continuity of $L(\mathbf{p})$, defined on a closed convex region, and its uniqueness is due to the strict concavity of $\log L(\mathbf{p})$. Although the objective function (3.2) is elegant, an analytic solution with general $v_i > 0$ is not available. In this subsection, analytic solutions are obtained for some special cases.

Theorem 1. *$L(\mathbf{p})$ has a unique maximum at $\mathbf{p} = (0, 1/3, 1/3, 1/3)$ if and only if $v_1 \geq v_2 + v_3 + v_4$.*

Graßhoff and Schwabe (2008) obtained essentially the same result. In the Appendix, we provide a direct proof.

Lemma 1. *If $v_1 > v_2$, then any solution to maximizing (3.2) satisfies $p_1 \leq p_2$; if $v_1 = v_2$, then any solution satisfies $p_1 = p_2$.*

Theorem 2. *Suppose $v_1 \geq v_2$, $v_3 = v_4 = v$, and $v_1 < v_2 + 2v$. Then the solution to maximizing (3.2) is*

$$p_1 = \frac{1}{2} - \frac{v_1 - v_2 + 4v}{2(-2\delta + D)},\ p_2 = \frac{1}{2} + \frac{v_1 - v_2 - 4v}{2(-2\delta + D)},\ p_3 = p_4 = \frac{2v}{-2\delta + D} \tag{3.3}$$

with $L = 2v^2\left(\delta^2 + 4v_1v_2 - \delta D\right)/(-2\delta + D)^3$, where $\delta = v_1 + v_2 - 4v$ and $D = \sqrt{\delta^2 + 12v_1v_2}$.

Corollary 1. *Suppose $v_2 = v_3 = v_4 = v$ and $v_1 < 3v$. Then the solution to maximizing (3.2) is*

$$p_1 = \frac{3v - v_1}{9v - v_1},\ p_2 = p_3 = p_4 = \frac{2v}{9v - v_1},$$

with maximum $L = 4v^3/(9v - v_1)^2$.

Corollary 2. *Suppose* $v_1 = v_2 = u$, $v_3 = v_4 = v$, *and* $u > v$. *Then the solution to maximizing* (3.2) *is*

$$p_1 = p_2 = \frac{2u - v - d}{6(u - v)},\ p_3 = p_4 = \frac{u - 2v + d}{6(u - v)}$$

with maximum $L = (2u - v - d)(u - 2v + d)(u + v + d)/\left(108(u - v)^2\right)$, *where* $d = \sqrt{u^2 - uv + v^2}$.

The analytic solutions in Theorem 2 and the two subsequent corollaries are obtained by forcing some of the v_i's to be equal. The equivalent restrictions in terms of regression coefficients $\beta_0, \beta_1, \beta_2$, can be obtained accordingly. As an illustration, under the logit link, the conditions of Corollary 1 are satisfied if and only if *either* at least two of the β_i's are zeros (the case $v_1 = v_2 = v_3 = v_4$) *or* $\beta_0 = \beta_1 = \beta_2$ and $|\beta_0| < \log\left((1 + \sqrt{3})/2 + \sqrt[4]{3}/\sqrt{2}\right) \approx 0.8314$, while the conditions of Corollary 2 are satisfied if and only if $\beta_2 = 0$ and $\beta_0\beta_1 > 0$. Figure 1(a) marks the regions satisfying the conditions of Theorem 2 in terms of β_i's under the logit link. More detailed results can be found in the Supplementary Materials of this paper.

Note that Theorem 1 does not correspond to a complete 2^2 experiment. It corresponds to a situation where the number of support points of the design is the number of parameters, *i.e.*, the design is saturated. We call $2\max_i v_i \geq v_1 + v_2 + v_3 + v_4$ the *saturation condition.* In terms of w's, it is

$$\frac{2}{\min\{w_1, w_2, w_3, w_4\}} \geq \frac{1}{w_1} + \frac{1}{w_2} + \frac{1}{w_3} + \frac{1}{w_4}. \tag{3.4}$$

3.2. Saturation condition in terms of the regression parameters

Now we investigate the saturation condition in terms of the regression coefficients.

Theorem 3. *For the logit link, the saturation condition is true if and only if*

$$\beta_0 \neq 0, \quad |\beta_1| > \frac{1}{2}\log\left(\frac{e^{2|\beta_0|} + 1}{e^{2|\beta_0|} - 1}\right), \quad \textit{and}$$

$$|\beta_2| \geq \log\left(\frac{2e^{|\beta_0| + |\beta_1|} + \sqrt{\left(e^{4|\beta_0|} - 1\right)\left(e^{4|\beta_1|} - 1\right)}}{\left(e^{2|\beta_0|} - 1\right)\left(e^{2|\beta_1|} - 1\right) - 2}\right).$$

Figure 1(b) shows how the region satisfying the saturation condition changes with β_0. For fixed β_0, a pair (β_1, β_2) satisfies the saturation condition if and only if the corresponding point in Figure 1(b) is above the curve labelled by β_0.

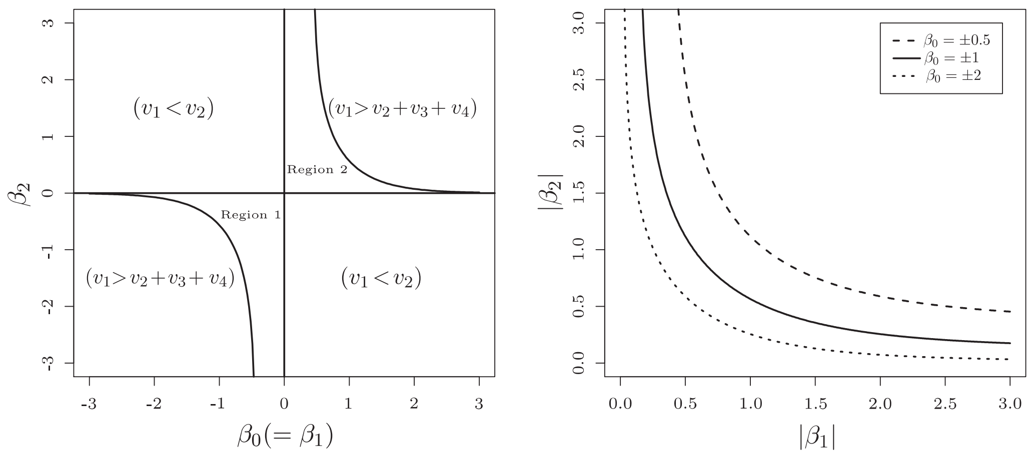

Figure 1. (a) Partitioning of the parameter space where the expressions in parenthesis indicate the corresponding properties of the v_i's besides $v_3 = v_4$. Region 1 and Region 2 satisfy the conditions $v_1 > v_2$, $v_3 = v_4$, and $v_1 < v_2 + v_3 + v_4$. (b) Lower boundary of the region satisfying the saturation condition.

Corollary 3. *Assume $\beta_0, \beta_1, \beta_2$ all have continuous distributions with support $(-a, a)$, $a > 0$, and $\beta_0, \beta_1, \beta_2$ are independent. The probability of a saturated design under the logit link is greater than 0 if and only if*

$$a > \log\Big(\frac{1+\sqrt{3}}{2} + \frac{\sqrt[4]{3}}{\sqrt{2}}\Big) \approx 0.8314.$$

For example, if $\beta_0, \beta_1, \beta_2$ are iid $\sim$ Uniform$(-a, a)$, the probabilities of obtaining saturation condition for different values of a are given in Table 1. In many applications, it is reasonable to assume that the coefficients β_1, β_2 are non-negative. In that case the probability of obtaining saturation condition is given by the next corollary.

Corollary 4. *Assume β_0 has a continuous distribution with support $(-a, a)$, $a > 0$, β_1 and β_2 have continuous distributions with support $[0, b)$, $b > 0$, and $\beta_0, \beta_1, \beta_2$ are independent. The probability of a saturated design under the logit link is greater than 0 if and only if*

$$b > \frac{1}{2}\log\Big(\frac{e^a + 1 + \sqrt{2(e^{2a}+1)}}{e^a - 1}\Big).$$

For example, the lower bound of b in Corollary 4 is $1.06,\ 0.76,\ 0.54,\ 0.48$, or 0.45 for $a = 0.5,\ 1,\ 2,\ 3$, or 5, respectively. If the support of β_0 is $(-\infty, \infty)$, then $b > \log(1+\sqrt{2})/2 \approx 0.4407$ guarantees a positive chance of a saturated design.

Table 1. Probability of obtaining saturation condition under uniform prior

a	logit	probit	log-log	c-log-log
0.5	0	0	0	0.21
1	0.02	0.24	0.30	0.63
2	0.51	0.79	0.79	0.80
3	0.76	0.92	0.92	0.84
4	0.87	0.96	0.96	0.90
5	0.92	0.98	0.98	0.94

Table 2. Probability of obtaining saturation condition under normal prior (logit link)

μ_1 \ μ_2	-2	-1	0	1	2
-2	0.81	0.64	0.54	0.64	0.81
-1	0.64	0.47	0.36	0.47	0.64
0	0.53	0.37	0.27	0.36	0.54
1	0.64	0.47	0.37	0.47	0.64
2	0.81	0.64	0.54	0.64	0.81

We have also investigated the occurrence of the saturation condition when the β's are assumed normal. Table 2 gives the probabilities for the logit link under the assumption that β_0 is standard normal, whereas β_1 and β_2 follow normal distribution with common variance 1 and means μ_1 and μ_2, respectively. The numbers are even higher under probit, log-log, or complementary log-log links. These simulations indicate that under reasonable distributional assumptions on the β's, the probability of attaining the saturation condition is quite high for the all common link functions.

3.3. Exact solution using cylindrical algebraic decomposition

Since analytic solutions for the optimization problem in (3.2) are not available, here we investigate computer-aided exact solutions. One option is to use Lagrange multipliers or the Karush-Kuhn-Tucker (KKT) conditions (Karush (1939), Kuhn and Tucker (1951)). It leads to intractable polynomial equations. Another option is to use numerical search algorithms, such as Nelder-Mead, quasi-Newton, conjugate-gradient, or simply a grid search (for a comprehensive reference, see Nocedal and Wright (1999)). Those numerical methods are computationally intensive in general when an accurate solution is needed. We suggest using the *Cylindrical Algebraic Decomposition* (CAD) algorithm to find the exact global solution.

Fotiou, Parrilo, and Morari (2005) provide the details for using CAD for general constrained optimization problems. Our optimization problem (3.2) is associated with the *boolean expression*

$$(L_3 - f \geq 0) \bigwedge (p_1 \geq 0) \bigwedge (p_2 \geq 0) \bigwedge (p_3 \geq 0) \bigwedge (p_1 + p_2 + p_3 \leq 1),$$

where $L_3 = L(p_1, p_2, p_3, 1 - p_1 - p_2 - p_3)$, and f is a new parameter indicating the value of the objective function. Given specific values of v_1, v_2, v_3, v_4, the CAD can represent the feasible domain of (f, p_1, p_2, p_3) in $\mathbb{R}^4$ as a finite union of disjoint cells. Each cell takes the form

$$\left\{ \begin{array}{c|l} & f = a_0 \text{ or } a_0 < f < b_0, \\ (f, p_1, p_2, p_3) & p_1 = g_1(f) \text{ or } g_1(f) < p_1 < h_1(f), \\ \in \mathbb{R}^4 & p_2 = g_2(f, p_1) \text{ or } g_2(f, p_1) < p_2 < h_2(f, p_1), \\ & p_3 = g_3(f, p_1, p_2) \text{ or } g_3(f, p_1, p_2) < p_3 < h_3(f, p_1, p_2) \end{array} \right\}$$

for some constants a_0, b_0 and some functions g_i, h_i, $i = 1, 2, 3$. Since f indicates the value of the objective function L (or L_3), the cell with greatest f provides the maximum of L. As an illustration, consider that $v_i = i \times v_1$ for $i = 2, 3, 4$, and the objective function is proportional to

$$4p_1p_2p_3 + 3p_1p_2p_4 + 2p_1p_3p_4 + p_2p_3p_4 \tag{3.5}$$

with $p_i \geq 0$ and $p_1 + p_2 + p_3 + p_4 = 1$. Using the software `Mathematica`, we obtain that the maximum of (3.5) based on CAD is the negative first root of $-96 + 800x + 5220x^2 - 19035x^3 + 2187x^4 = 0$ and

$$\begin{aligned} &p_1 \text{ is the 4th root of } -2 - 13x + 18x^2 + 126x^3 + 54x^4 = 0, \\ &p_2 \text{ is the 2nd root of } -2 + 2x + 28x^2 - 39x^3 + 9x^4 = 0, \\ &p_3 \text{ is the 2nd root of } -3 + 13x + 2x^2 - 26x^3 + 6x^4 = 0. \end{aligned}$$

Here the numerical maximum of (3.5) is 0.1645 with $p_1 = 0.3112, p_2 = 0.2849, p_3 = 0.2508$ and $p_4 = 0.1531$. In this example, the negative roots of the equation $-96 + 800x + 5220x^2 - 19035x^3 + 2187x^4 = 0$ provide candidates for the maximum of (3.5). The negative first root is the largest among them and hence is chosen. The choices of roots for p_i's are determined by the equalities/inequalities for the cell that provides the maximum.

Note that the CAD algorithm can be used to find the exact solution for general v_1, v_2, v_3, v_4, although an explicit formula is not available. This technique will be used in the next section in a robustness study.

3.4. Analytic approximate solution

In this section, we propose an analytic approximate solution for (3.2) in the general case. To simplify notation, write $L^*[v_1, v_2, v_3, v_4]$ for $\max_{\mathbf{p}} L$, given

v_1, v_2, v_3, v_4. For example, Theorem 2 corresponds to $L^*[v_1, v_2, v, v]$. Without any loss of generality, we assume $v_1 < v_2 < v_3 < v_4$ and $v_4 < v_1 + v_2 + v_3$. Let

$$L_{34} = L^*[v_1, v_2, \frac{v_3+v_4}{2}, \frac{v_3+v_4}{2}]$$

and take L_{12}, L_{13}, L_{14}, L_{23}, L_{24} accordingly. The strategy is to use $\max\{L_{12}, L_{23}, L_{34}\}$ to approximate $\max_{\mathbf{p}} L$ based on the following results.

Theorem 4. *Assume $v_1 < v_2 < v_3 < v_4$ and $v_4 < v_1 + v_2 + v_3$. Then*

$$\max\{L_{13}, L_{14}, L_{24}\} \le \max\{L_{12}, L_{23}, L_{34}\},$$
$$\max_{\mathbf{p}} L - \max\{L_{12}, L_{23}, L_{34}\} \le \min\left\{\frac{v_2-v_1}{216}, \frac{v_3-v_2}{96\sqrt{3}}, \frac{v_4-v_3}{54}\right\}.$$

We call the best $\mathbf{p}$ among the solutions to L_{12}, L_{23}, or L_{34} the *analytic approximate solution*, and denote it by $\mathbf{p_a}$. Thus $L(\mathbf{p_a}) = \max\{L_{12}, L_{23}, L_{34}\}$. Theorem 4 provides a theoretical upper bound for the difference $\max_{\mathbf{p}} L - L(\mathbf{p_a})$.

To see how our approximation works numerically, we simulated the regression coefficients as $\beta_0, \beta_1, \beta_2$ iid $\sim N(0,1)$, and calculated the corresponding $\mathbf{w} = (w_1, w_2, w_3, w_4)$ under the logit link. Since we know the exact solution if $\mathbf{w}$ satisfies the saturation condition, we omitted those $\mathbf{w}$'s and collected 1,000 non-saturated cases. For each randomly chosen non-saturated $\mathbf{w}$, we calculated the optimal $\mathbf{p}$ using CAD and denoted it by $\mathbf{p_o}$. We also calculated the analytic solution based on Theorems 4 and 2, $\mathbf{p}_*$. Then we calculated the determinant of the information matrix, denoted D_o and D_* for $\mathbf{p_o}$ and $\mathbf{p}_*$, respectively. Numerical results show that 99% of the relative losses $\left(D_o^{1/3} - D_*^{1/3}\right) \Big/ D_o^{1/3} \times 100\%$ are less than 0.01% and the maximum relative loss is about 0.02%. If we change the distribution $N(0,1)$ to $N(0,a^2)$ or Uniform$(-a,a)$, $a = 2,3,4,5$, the relative losses are roughly the same.

Since the v_i's depend on the assumed values, they are in general not quantified accurately at the planning stage. The results in this section show that setting some v_i's to be equal is not a bad strategy. The analytic approximation is also potentially useful for future research in this area.

4. Robustness for 2^2 Designs

Since locally optimal designs depend on the assumed values of the parameters, it is important to study the robustness of the designs to these values. For experiments where there is no basis for making an informed choice of the assumed values, the natural design choice is the uniform design. In this section, we study the robustness of the optimal design for misspecification of assumed values.

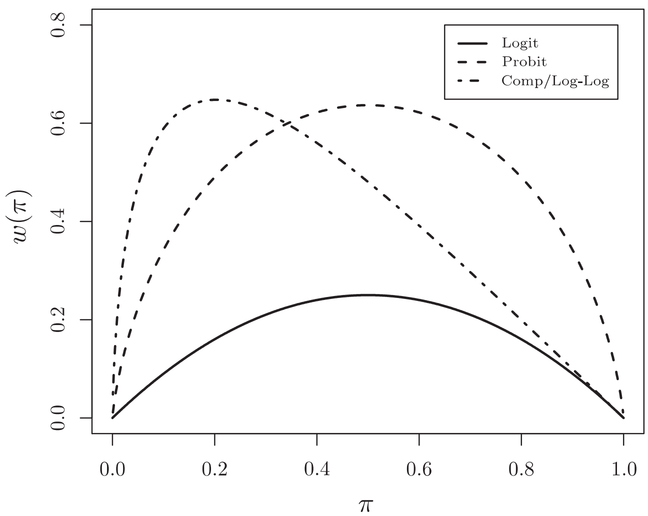

Figure 2. Plot of w versus π.

4.1. Robustness for misspecification of w

Write $\mathbf{w} = (w_1, w_2, w_3, w_4)$, and take $\mathbf{w}_t = (w_{t1}, w_{t2}, w_{t3}, w_{t4})$ as the true $\mathbf{w}$, and $\mathbf{w}_c = (w_{c1}, w_{c2}, w_{c3}, w_{c4})$ as the chosen (assumed) $\mathbf{w}$. Let $\mathbf{p}_t = (p_{t1}, p_{t2}, p_{t3}, p_{t4})$ and $\mathbf{p}_c = (p_{c1}, p_{c2}, p_{c3}, p_{c4})$ be the optimal designs corresponding to $\mathbf{w}_t$ and $\mathbf{w}_c$, respectively. The relative loss of efficiency of choosing $\mathbf{w}_c$ instead of $\mathbf{w}_t$ is taken as

$$R(\mathbf{w}_t, \mathbf{w}_c) = \frac{det(\mathbf{w}_t, \mathbf{p}_t)^{1/3} - det(\mathbf{w}_t, \mathbf{p}_c)^{1/3}}{det(\mathbf{w}_t, \mathbf{p}_t)^{1/3}}, \tag{4.1}$$

where $det(\mathbf{w}, \mathbf{p})$ was defined in (3.1). Note that $R(\mathbf{w}_t, \mathbf{w}_c)$ remains invariant under scalar multiplication of determinants. The maximum relative loss of efficiency is

$$R_{max}(\mathbf{w}_c) = \max_{\mathbf{w}_t} \Big\{ R(\mathbf{w}_t, \mathbf{w}_c) \Big\}. \tag{4.2}$$

This maximum corresponds to the worst case scenario. This tells us, for each $\mathbf{w}$, how bad the design can perform if we do not choose the $\mathbf{w}$ correctly.

For a binary response, commonly used link functions include logit, probit, log-log and complimentary log-log links. Figure 2 illustrates the range of w for specific link functions, as mentioned in Section 2. The logit link corresponds to $0 < w \leq 0.25$ whereas for the probit link $0 < w \leq 2/\pi$, and for the (complementary) log-log links $0 < w \leq 0.648$. It should also be noted that the w-curve is symmetric for logit and probit links but asymmetric for the (complementary) log-log link. To examine the robustness for mis-specification of w for different links, we take $0 < \xi \le w \le \zeta$ for some constants ξ and ζ, since $w = 0$ leads to trivial cases.

Fixing a chosen $\mathbf{w}_c = (w_{c1}, w_{c2}, w_{c3}, w_{c4})$, let $v_{ci} = 1/w_{ci}$, $i = 1, 2, 3, 4$. Without loss of generality, $v_{c1} \le v_{c2} \le v_{c3} \le v_{c4}$. It follows from Lemma 1 that $p_{c1} \ge p_{c2} \ge p_{c3} \ge p_{c4}$. For the true $\mathbf{w}_t = (w_{t1}, w_{t2}, w_{t3}, w_{t4})$, take $\mathbf{v}_t = (v_{t1}, v_{t2}, v_{t3}, v_{t4})$ with $v_{ti} = 1/w_{ti}$, $i = 1, 2, 3, 4$. In practice, the experimenter might have some rough idea about the range of the parameter values. For example, if the experimenter believes that the regression coefficients $\beta_0, \beta_1, \beta_2$ all take values in $[-1, 1]$, then w_1, w_2, w_3, w_4 fall into $[0.045, 0.25]$ under the logit link. Our next theorem specifies the worst possible performance of a chosen design for an assumed range. To simplify notation, we write $R^{(c)}_{\max} = R_{\max}(\mathbf{w}_c)$, the maximum relative loss of efficiency as at (4.2).

Theorem 5. *Suppose* $0 < a \le v_{c1} \le v_{c2} \le v_{c3} \le v_{c4} \le b$ *and* $a \le v_{ti} \le b$, $i = 1, 2, 3, 4$.

(i) *If* $v_{c4} \ge v_{c1} + v_{c2} + v_{c3}$, *then* $R^{(c)}_{\max} = 1 - ((9\theta - 1)/2)^{2/3}/(3\theta)$, *where* $\theta = b/a \ge 3$, *and the maximum can only be attained at* $\mathbf{v}_t = (b, b, b, a)$.

(ii) *If* $v_{c4} < v_{c1} + v_{c2} + v_{c3}$, *then* $R^{(c)}_{\max}$ *can only be attained at* $\mathbf{v}_t = (b, a, a, a)$, (b, b, a, a), *or* (b, b, b, a).

Note that in Theorem 5, both v_{ci}'s and v_{ti}'s are restricted to a bounded region $[a,\ b]$ excluding 0 and ∞. However, w's can be arbitrarily close to 0, and the upper bound b for v can go to ∞. For large enough b, case (i) of Theorem 5 leads to $R^{(c)}_{\max} = 1$, indicating total loss of efficiency, while $R^{(c)}_{\max}$ in case (ii) can only be attained at $\mathbf{v}_t = (b, a, a, a)$.

Corollary 5. *Suppose* $0 < a \le v_{c1} \le v_{c2} \le v_{c3} \le v_{c4} < \infty$ *and* $a \le v_{ti} < \infty$, $i = 1, 2, 3, 4$.

(i) *If* $v_{c4} \ge v_{c1} + v_{c2} + v_{c3}$, *then* $R^{(c)}_{\max} = 1$.

(ii) *If* $v_{c4} < v_{c1} + v_{c2} + v_{c3}$, *then* $R^{(c)}_{\max} = 1 - 3\,(p_{c2}p_{c3}p_{c4})^{1/3}$.

The proof of Corollary 5 is relegated to the Supplementary Materials. The next theorem asserts that the uniform design is the most robust design in terms of maximum relative loss. Note that the conclusion of Theorem 6 is still true even if $[a, b]$ is replaced with $[a, \infty)$.

Theorem 6. *Suppose* $v_{ci}, v_{ti} \in [a, b]$, $i = 1, 2, 3, 4$, $0 < a \le b$. *Then* $R^{(c)}_{\max}$ *attains its minimum if and only if* $\mathbf{p}_c = (1/4, 1/4, 1/4, 1/4)$.

Remark 1. Uniform design is the most robust design in the absence of any knowledge about the parameters. However, if the experimenter has some prior knowledge about the model parameters, say some parameters are nonnegative,

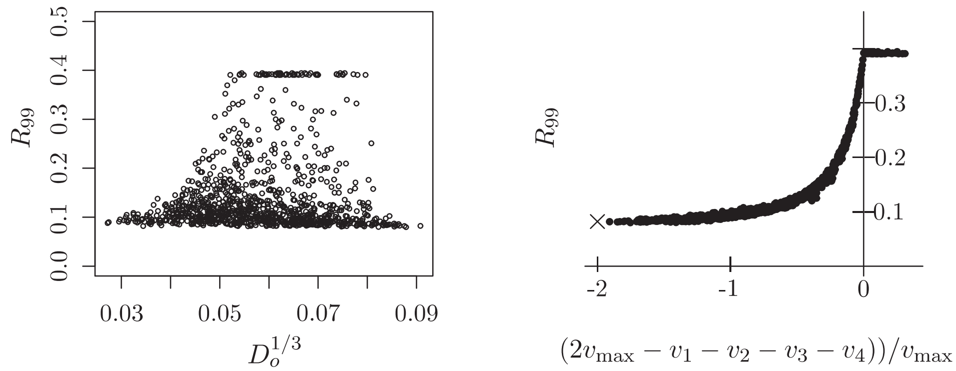

Figure 3. Robustness study : performance of the "worst 1%" design.

then the uniform design may not be the most robust one. For example, if $\beta_0 \sim$ Uniform$(-1, 1)$, $\beta_1, \beta_2 \sim$ Uniform$[0, 1)$, and $\beta_0, \beta_1, \beta_2$ are independent, the theoretical $R_{\max}$ of the uniform design is 0.134, while the theoretical $R_{\max}$ of the design given by $\mathbf{p} = (0.19, 0.31, 0.31, 0.19)$ is 0.116. This does not violate Theorem 6 since $w_1, w_4 \in (0.045, 0.25]$ and $w_2, w_3 \in (0.105, 0.25]$ under the logit link.

4.2. Simulation study

To study the robustness measured by percentiles of $\{R(\mathbf{w}_t, \mathbf{w}_c)\}_{\mathbf{w}_t}$ other than the maximum $R_{\max}(c)$, we randomly selected 1,000 vectors $\mathbf{w}_i = (w_{i1}, w_{i2}, w_{i3}, w_{i4})$, $i = 1, \ldots,$ 1,000. For the logit link, $0 < w \leq 0.25$. If we randomly select w_i's between 0 and 0.25, the chance of getting a saturated design can be as high as 48% when some w_i is close to 0 and the condition of Theorem 1 applies. We try to reduce the cases that give a saturated design since in those cases both the exact solution and robustness are clearly known. So here we consider $w \geq 0.05$ only. Then the chance of saturated design drops to 6% for uniformly distributed w_i's. So, for the logit link, we consider $0.05 \leq w \leq 0.25$.

In our robustness study, each one of the 1,000 $\mathbf{w}_i$'s is chosen in turn as $\mathbf{w}_c$, the assumed $\mathbf{w}$, and the remaining 999 cases are regarded as $\mathbf{w}_t$, the true $\mathbf{w}$, respectively. We use CAD to determine the optimal designs $\mathbf{p}_t = (p_{t1}, p_{t2}, p_{t3}, p_{t4})$ and $\mathbf{p}_c = (p_{c1}, p_{c2}, p_{c3}, p_{c4})$ corresponding to $\mathbf{w}_t$ and $\mathbf{w}_c$, respectively.

For the numerical computations in this section, we considered the upper 99$^{\text{th}}$ percentile of $\{R(\mathbf{w}_t, \mathbf{w}_c)\}_{t=1}^{1000}$, denoted by $R_{99}^{(c)}$. This corresponds to the worst 1% case scenario. The left panel of Figure 3 illustrates that this relative loss will range roughly between 0.1 and 0.4, whereas the right panel helps us identify those $\mathbf{w}$'s with non-robust optimal solution $\mathbf{p}$. The horizontal axis corresponds to the distance between $v_{\max}$ and $\sum v_i - v_{\max}$ divided by $v_{\max}$, where $v_{\max} = \max\{v_1, v_2, v_3, v_4\}$. The vertical axis is our robustness measurement R_{99}. There is a clearly positive association between the relative loss and the distance. We have examined the other quantiles such as the 25^{th} quantile, median, 75^{th} quantile, and 95^{th} quantile of $R(\mathbf{w}_t, \mathbf{w}_c)$. The patterns are similar for all of them. From this, we conclude that the locally D-optimal designs are quite robust and the farther the $\mathbf{w}$'s are from the saturation condition (3.4), the smaller is the relative loss of efficiency. It is interesting to note that the left most point (denoted by $\times$) on the right panel of Figure 3 corresponds to the uniform design. It can be verified that the standardized distance $(2v_{\max} - \sum v_i)/v_{\max}$ attains its minimum -2 if and only if $v_1 = v_2 = v_3 = v_4$ which leads to the uniform design. While the left panel of Figure 3 indicates that the performance of the "worst 1%" designs is not too bad in terms of robustness, the right panel (as well as figures of other quantiles not shown here) clearly demonstrates the significance of the saturation condition. The points with R_{99} values greater than 0.15 either satisfy or almost satisfy the saturation condition. This figure also suggests that the uniform design is highly robust.

Note that although our simulation model assumes independence of the w_i, they are associated through dependence on the β_i's. We have performed several simulations starting with random β_i's and then computing w_i's via a link function. The results are similar to the ones reported here except for a slightly smaller relative loss of efficiency.

The simulation study based on a model for the w_i's has the advantage that the results do not depend on the link function. Moreover, since w_i's are inverses of the variances of the y_i's, in some applications, it may be easier to suggest initial values of w_i's.

5. Robustness of Uniform Design

If the experimenter is unable to make an informed choice of the values for local optimality, the natural design choice is the uniform design $\mathbf{p}_u = (1/4, 1/4, 1/4, 1/4)$. The relative loss of efficiency of $\mathbf{p}_u$ with respect to the true $\mathbf{w}_t$ can be obtained from (4.1), say

$$R_u(\mathbf{w}_t) = 1 - \frac{1}{4}\left(\frac{v_{t1} + v_{t2} + v_{t3} + v_{t4}}{L(\mathbf{p}_t)}\right)^{1/3},$$

where $v_{ti} = 1/w_{ti}$, $i = 1, 2, 3, 4$, $\mathbf{p}_t$ is the D-optimal design with respect to $\mathbf{w}_t$, and $L(\mathbf{p}_t)$ is defined in (3.2).

5.1. Maximum relative loss of uniform design

We denote by $R_{\max}^{(u)} = \max_{\mathbf{w}_t} R_u(\mathbf{w}_t)$ the maximum loss of efficiency of the uniform design. The following theorem formulates $R_{\max}^{(u)}$ with different restricted regions of $\mathbf{w}_t$'s.

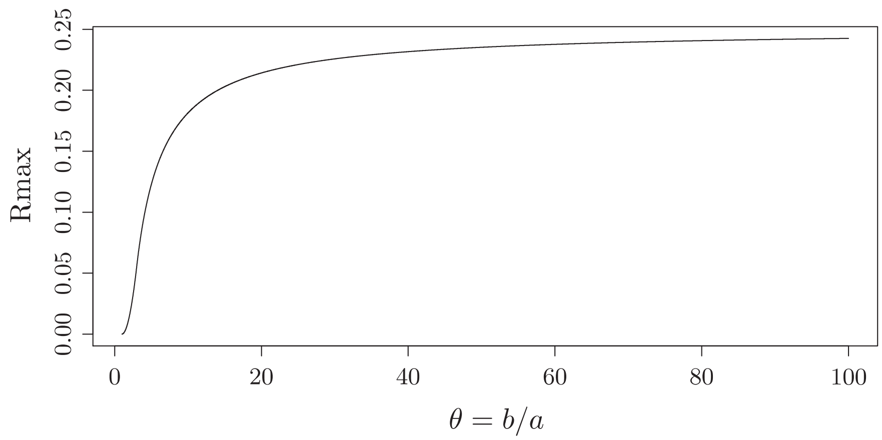

Figure 4. Plot of $R^{(u)}_{\max}$ versus θ.

Theorem 7. *Suppose* $0 < a \le v_{ti} \le b$, $i = 1, 2, 3, 4$. *Let* $\theta = b/a \ge 1$. *Then*

$$R^{(u)}_{\max} = \begin{cases} 1 - \frac{3}{4}\left(1+\frac{3}{\theta}\right)^{1/3}, & \text{if } \theta \ge 3, \\ 1 - \frac{1}{8}\left(2(\theta+3)(9-\theta)^2\right)^{1/3}, & \text{if } \theta_* \le \theta < 3, \\ 1 - \frac{3}{2}\left(\frac{(\theta+1)(\theta-1)^2}{(2\theta-1-\rho)(\theta-2+\rho)(\theta+1+\rho)}\right)^{1/3}, & \text{if } 1 < \theta < \theta_*, \\ 0, & \text{if } \theta = 1, \end{cases}$$

where $\rho = \sqrt{\theta^2 - \theta + 1}$, *and* $\theta_* \approx 1.32$ *is the 3rd root of the equation*

$$3456 - 51184\theta + 3561\theta^2 + 596\theta^3 - 1506\theta^4 + 100\theta^5 + \theta^6 = 0.$$

Figure 4 reveals the nature of association between the maximum relative loss of uniform design and the ratio between the upper and lower limits of range of **w**'s. It can be seen that the performance of uniform design is worst when θ is 10 or more, but even in that case $R^{(u)}_{max}$ is less than 1/4. Uniform designs perform moderately well when θ lies between 3 and 10, and extremely well if $\theta < 3$.

5.2. An example

The data given in Table 3, reported by Collett (1991), were originally obtained from an experiment conducted at the East Malling Research Station (Hoblyn and Palmer (1934)). The experimenters investigated the vegetative reproduction of plum trees. Cuttings from the roots of older trees of the palms named Common Mussel were taken between October, 1931 and February, 1932. This experiment involved two factors each at two levels. The first factor was time of planting (root stocks were either planted as soon as possible after they were taken, or they were imbedded in sand under cover and were planted in the

Table 3. Survival rate of plum root-stock cutting.

Length of cutting	Time of planting	Number of surviving out of 240
Short	At once	107
	In Spring	31
Long	At once	156
	In Spring	84

next spring). The second factor was the length of root cuttings (6 cm or 12 cm). Hoblyn and Palmer used a uniform design, and a total of 240 cuttings were taken for each of the four combinations. The response was the condition of each plant (alive or dead) in October, 1932.

After fitting the logit model we get $\hat{\beta} = (-0.5088, -0.5088, 0.7138)'$, and the corresponding $\mathbf{w} = (0.244, 0.128, 0.221, 0.221)'$. If we use this $\mathbf{w}$, the optimal proportions are $\mathbf{p_o} = (0.2818, 0.1686, 0.2748, 0.2748)'$. The corresponding determinant of $X'WX$ is 8.197×10^{-3}. On the other hand, the determinant of the information matrix corresponding to the uniform design is 7.975×10^{-3}. Thus the uniform design is 99.1% efficient. If this was the first of a series of experiments, then the result supports continued use of the uniform design. Similar calculations with the probit link show that the uniform design is 99.9% efficient.

6. General Case: 2^k Experiment

The general case of 2^k factorial is technically complicated. We have some results for particular cases that will be reported in future publications. In this section we show that the uniform design ($p_1 = \cdots = p_{2^k} = 1/2^k$) has a *maximin* optimality property in that it maximizes a lower bound of the *D-criterion*.

Suppose there are q parameters (main effects and interactions) in the chosen model for the 2^k experiment, β is a $q \times 1$ vector and the design matrix X is $2^k \times q$. The matrix X can be extended to the $2^k \times 2^k$ design matrix F of the "full model" that includes all main effects and interactions. Note that F is a Hadamard matrix, i.e., $F'F = 2^k I_{2^k}$, where I_{2^k} is the identity matrix of order 2^k. Moreover, we can partition F as $F = (X|R)$ where R contains all effects that are not in the chosen model. Clearly,

$$\left|F'WF\right| = \left|X'WX\right| \left|R'WR - R'WX\left(X'WX\right)^{-1}X'WR\right| \leqslant \left|X'WX\right| \left|R'WR\right|.$$

Hence

$$\left|X'WX\right| \geqslant \frac{|F'WF|}{|R'WR|}.$$

Note that $|F'WF| = 2^{kq}\Pi w_i \Pi p_i$, and

$$\left|R'WR\right| \leqslant (w_M)^{2^k - q} \left|R'PR\right|,$$

where $w_M = \max\{w_1, \ldots, w_{2^k}\}$, and $P = diag\,(p_1, \ldots, p_{2^k})$. It follows from Proposition $1'$ of Kiefer (1975) that the uniform design $p_1 = \cdots = p_{2^k} = 1/2^k$ maximizes $|R'PR|$. Hence we obtain a *lower bound* to the *D-criterion*

$$\left|X'WX\right| \geqslant \frac{2^{k2^k}\Pi w_i \Pi p_i}{(w_M)^{2^k-q}}.$$

This lower bound is a maximum when $p_1 = \cdots = p_{2^k} = 1/2^k$, which gives a maximin property of the uniform design.

How good is the uniform design? The loss of efficiency corresponding to the uniform design is

$$R_u(w) = 1 - \left(\frac{\left|X'diag\left(w_1/2^k, \ldots, w_{2^k}/2^k\right)X\right|}{\max|X'WX|}\right)^{1/q}.$$

Then it can be shown that

$$R_u(w) \leqslant 1 - \frac{|X'W_0X|^{1/q}}{2^k w_M} \leqslant 1 - \frac{w_m}{w_M},$$

where $w_m = \min\{w_1, \ldots, w_{2^k}\}$ and $W_0 = diag\,(w_1, \ldots, w_{2^k})$.

For example, suppose it is known that $0.14 \leqslant w_i \leqslant 0.20$. Then the uniform design is not less than 70% efficient, irrespective of the value of k or the X matrix (the model). While this bound enables us to make general statements like this, the results in Theorem 7 for the 2^2 case show that this bound can be quite conservative.

7. Discussion

Locally optimal designs require assumed values of the parameters w that may not be readily available at the planning stage. The expression of w_i given in Section 2 depends on the assumed values of the parameter β and the link function g. In situations where there is overdispersion, this expression (the nominal variance) may not be adequate to describe variation in the model, and a more realistic representation for w may be

$$w_i = c_i \frac{(d\pi_i/d\eta_i)^2}{(\pi_i(1-\pi_i))},$$

where c_i is a function of the factor levels $x_1, \ldots, x_k$. This makes the specification of w_i even more difficult. For the design problem, however, it can be seen that we need only the relative magnitudes $w_i^* = w_i/w_M$ that may be easier to specify in some applications. Note that w_i^* take values in the interval $(0, 1]$.

The overall conclusion for the 2^2 factorial experiment with main-effects model is that, for the link functions we studied, the locally optimal designs are robust in the sense that the loss of efficiency due to misspecification of the assumed values is not large. If the (assumed) variance at one point is substantially larger than the others, then the D-optimal design is based on only three of the four points. In applications, however, an experimenter would rarely feel confident to not allocate observations at a point based solely on assumed values, and this is not our recommendation for practice. However, the D-optimal design still provides a useful benchmark for the efficiency of designs, and to the extent feasible it is wise to dedicate more resources to points that we believe have small variance, and less resources to points with large variance. If there is no basis to make an informed choice of the assumed values, we recommend the use of the uniform design.

Acknowledgement

The authors thank Professor John P Morgan for his valuable suggestions. They would also like to thank the Editor, an associate editor and the reviewers for comments and suggestions that substantially improved the manuscript. The research was in part supported by NSF Grant DMS-09-05731.

Appendices

A.1. Proof of Theorem 1

(1) If $v_1 \geq v_2 + v_3 + v_4$, then

$$
\begin{aligned}
L &= v_4(p_1p_2p_3 + p_2p_3p_4) + v_3(p_1p_2p_4 + p_2p_3p_4) + v_2(p_1p_3p_4 + p_2p_3p_4) \\
&\quad +(v_1 - v_2 - v_3 - v_4)p_2p_3p_4 \\
&= v_4(p_1 + p_4)p_2p_3 + v_3(p_1 + p_3)p_2p_4 + v_2(p_1 + p_2)p_3p_4 \\
&\quad +(v_1 - v_2 - v_3 - v_4)p_2p_3p_4 \\
&\leq v_4\left(\frac{(p_1 + p_4) + p_2 + p_3}{3}\right)^3 + v_3\left(\frac{(p_1 + p_3) + p_2 + p_4}{3}\right)^3 \\
&\quad +v_2\left(\frac{(p_1 + p_2) + p_3 + p_4}{3}\right)^3 + (v_1 - v_2 - v_3 - v_4)p_2p_3p_4 \qquad \text{(A.1)} \\
&= \frac{v_4}{27} + \frac{v_3}{27} + \frac{v_2}{27} + (v_1 - v_2 - v_3 - v_4)p_2p_3p_4 \\
&\leq \frac{v_2 + v_3 + v_4}{27} + (v_1 - v_2 - v_3 - v_4)\left(\frac{p_2 + p_3 + p_4}{3}\right)^3 \qquad \text{(A.2)} \\
&\leq \frac{v_2 + v_3 + v_4}{27} + (v_1 - v_2 - v_3 - v_4)\left(\frac{p_1 + p_2 + p_3 + p_4}{3}\right)^3 = \frac{v_1}{27}. \qquad \text{(A.3)}
\end{aligned}
$$

By the inequality of arithmetic and geometric means, equality in (A.1) holds if and only if

$$p_1 + p_4 = p_2 = p_3,\ \ p_1 + p_3 = p_2 = p_4,\ \ \text{and}\ \ p_1 + p_2 = p_3 = p_4,$$

which implies $p_1 = 0$, $p_2 = p_3 = p_4 = 1/3$. Note that the equality in (A.2) holds if $p_2 = p_3 = p_4$, and equality in (A.3) holds if $p_1 = 0$. After all, $L = v_1/27$ if and only if $p_1 = 0$, $p_2 = p_3 = p_4 = 1/3$.

(2) If $\mathbf{p} = (0, 1/3, 1/3, 1/3)$ maximizes (3.2), we claim that $v_1 \geq v_2 + v_3 + v_4$. Otherwise, if $v_1 < v_2+v_3+v_4$, the solution $\mathbf{p}_\epsilon = (\epsilon, (1-\epsilon)/3, (1-\epsilon)/3, (1-\epsilon)/3)$ is better than $\mathbf{p}$ for small enough $\epsilon > 0$. It can be shown that $L(\mathbf{p}_\epsilon) > L(\mathbf{p})$ if and only if $3d_1 > \epsilon(3+6d_1-2\epsilon-3d_1\epsilon)$, where $d_1 = (v_2+v_3+v_4-v_1)/v_1 > 0$.

A.2. Proof of Theorem 2

Based on Lemma 3.1.1, $v_3 = v_4$ implies $p_3 = p_4$ in any solution maximizing (3.2). Write $x = p_1$, $y = p_2$, and $z = p_3 = p_4$. Then $x = 1 - y - 2z$. Now (3.2) is equivalent to maximizing

$$L(y, z) = z\left[2vy - 2vy^2 + v_2 z + (v_1 - v_2 - 4v)\, yz - 2v_2 z^2\right]$$

with restrictions $y \geq 0$, $z \geq 0$, and $y + 2z \leq 1$. It can be verified that none of the boundary cases could be a solution maximizing (3.2).

Assume $y > 0$, $z > 0$, and $y + 2z < 1$. The solution to $\partial L(y, z)/\partial y = 0$ is $y = y_* := z(v_1 - v_2 - 4v)/(4v) + 1/2$. Plug $y = y_*$ into $L(y, z)$ and get

$$L(z) = z\left(\frac{\Delta}{8v} z^2 + \frac{v_1 + v_2 - 4v}{2} z + \frac{v}{2}\right),$$

where $\Delta = (v_1 + v_2 - 4v)^2 - 4v_1 v_2$. So we only need to maximize $L(z)$ with the restriction $0 \leq z \leq z_*$, where $z_* = 2v/(v_1 - v_2 + 4v)$. Assume $v_1 > v_2$ first.

Case (i): Suppose $\Delta = 0$. Then $dL(z)/dz = (v_1 + v_2 - 4v)z + v/2$. Since $v_1 < v_2 + 2v$, then $\sqrt{v_1} = 2\sqrt{v} - \sqrt{v_2}$, which implies $9v_2 > v_1 > v > v_2$ and $v_1 + v_2 - 4v < 0$. $L(z)$ attains its maximum $\left(\sqrt{v_1} + \sqrt{v_2}\right)^4 / [256\sqrt{v_1 v_2}]$ at $z_0 = v/[2(4v - v_1 - v_2)]$. Then the solution maximizing (3.2) is

$$p_1 = \frac{3\sqrt{v_2} - \sqrt{v_1}}{8\sqrt{v_2}},\ p_2 = \frac{3\sqrt{v_1} - \sqrt{v_2}}{8\sqrt{v_1}},\ p_3 = p_4 = \frac{\left(\sqrt{v_1} + \sqrt{v_2}\right)^2}{16\sqrt{v_1 v_2}}. \qquad \text{(A.4)}$$

It can be verified that the solution (A.4) is a special case of (3.3) when $\sqrt{v_1} = 2\sqrt{v} - \sqrt{v_2}$.

From now on, assume $\Delta \neq 0$. Let $\delta = v_1 + v_2 - 4v$. Then $\Delta = \delta^2 - 4v_1v_2$ and the solutions to $dL(z)/dz = 0$ are

$$z_1 = -\frac{2v}{2\delta + \sqrt{\delta^2 + 12v_1v_2}}, \quad z_2 = \frac{2v}{-2\delta + \sqrt{\delta^2 + 12v_1v_2}}.$$

Therefore, $dL(z)/dz = 3\Delta(z - z_1)(z - z_2)/(8v)$.

Case (ii): Suppose $\Delta > 0$. Then $v_1 < v_2 + 2v$ implies $\delta < -2\sqrt{v_1v_2}$, that is, $\sqrt{v_1} < 2\sqrt{v} - \sqrt{v_2}$. It can be verified that $0 < z_2 < z_* < z_1$. $L(z)$ attains its maximum at z_2. The solution leads to (3.3), which maximizes (3.2).

Case (iii): Suppose $\Delta < 0$. Then $v < 4v_2$ and $z_1 < 0 < z_2 < z_*$. $L(z)$ attains its maximum at z_2. The solution leads to (3.3) as well.

So far we have shown that the solution to (3.3) maximizes (3.2) when $v_1 > v_2$, $v_3 = v_4 = v$ and $v_1 < v_2 + 2v$. It can be verified that the (3.3) holds if $v_1 > v_2$ is replaced with $v_1 \geq v_2$.

A.3. Proof of Theorem 5.

To simplify notation, let $q_4 = p_{c1}p_{c2}p_{c3}$, $q_3 = p_{c1}p_{c2}p_{c4}$, $q_2 = p_{c1}p_{c3}p_{c4}$, and $q_1 = p_{c2}p_{c3}p_{c4}$. Then $q_4 \geq q_3 \geq q_2 \geq q_1$. If

$$Q_c(v_{t1}, v_{t2}, v_{t3}, v_{t4}) = \frac{v_{t4}q_4 + v_{t3}q_3 + v_{t2}q_2 + v_{t1}q_1}{v_{t4}p_{t1}p_{t2}p_{t3} + v_{t3}p_{t1}p_{t2}p_{t4} + v_{t2}p_{t1}p_{t3}p_{t4} + v_{t1}p_{t2}p_{t3}p_{t4}},$$

then $R(\mathbf{w}_t, \mathbf{w}_c) = 1 - Q_c(v_{t1}, v_{t2}, v_{t3}, v_{t4})^{1/3}$. Also, we use (v_1, v_2, v_3, v_4) instead of $(v_{t1}, v_{t2}, v_{t3}, v_{t4})$. Note that $Q_c(\lambda v_1, \lambda v_2, \lambda v_3, \lambda v_4) = Q_c(v_1, v_2, v_3, v_4)$ for any $\lambda > 0$. To minimize $Q_c(v_1, v_2, v_3, v_4)$, we only need to consider those cases with $v_4 = a$.

If $v_{c4} \geq v_{c1} + v_{c2} + v_{c3}$, then $p_{c4} = 0$ and $p_{c1} = p_{c2} = p_{c3} = 1/3$. Thus $q_3 = q_2 = q_1 = 0$ and $q_4 = 1/27$. Fixing $v_4 = a$, it can be verified that

$$Q_c(v_1, v_2, v_3, a) = \frac{a/27}{v_1p_2p_3p_4 + v_2p_1p_3p_4 + v_3p_1p_2p_4 + ap_1p_2p_3} \geq Q_c(b, b, b, a),$$

where equality holds if and only if $v_1 = v_2 = v_3 = b$. Let $\theta = b/a \geq 3$. In this case,

$$R_{\max}^{(c)} = 1 - \frac{1}{3\theta}\left(\frac{9\theta - 1}{2}\right)^{2/3}.$$

Suppose $v_{c4} < v_{c1} + v_{c2} + v_{c3}$. Then $p_{c1} \geq p_{c2} \geq p_{c3} \geq p_{c4} > 0$ and $q_4 \geq q_3 \geq q_2 \geq q_1 > 0$. Again, we fix $v_4 = a$ and assume $0 < a \leq v_i \leq b$, $i = 1, 2, 3, 4$.

Case 1*:* If $v_1 \geq v_2 + v_3 + a$, then $Q_c(v_1, v_2, v_3, a) \geq Q_c(b, a, a, a)$, where equality holds if and only if $v_1 = b$ and $v_2 = v_3 = v_4 = a$. In this case, $\theta = b/a \geq 3$.

Case 2*:* If $v_1 < v_2 + v_3 + a$ and $v_1' > v_1$, then $Q_c(v_1', v_2, v_3, a) < Q_c(v_1, v_2, v_3, a)$.

Actually, in this case, $0 < p_1 \leq p_2 \leq p_3 \leq p_4$. For small enough $\epsilon > 0$,

$$Q_c(v_1, v_2, v_3, a) \geq \frac{v_1' q_1 + v_2 q_2 + v_3 q_3 + a q_4}{v_1' p_2' p_3' p_4' + v_2 p_1' p_3' p_4' + v_3 p_1' p_2' p_4' + a p_1' p_2' p_3'} \geq Q_c(v_1', v_2, v_3, a),$$

where $p_i' = p_i + \epsilon$, $i = 2, 3, 4$ and $p_1' = p_1 - 3\epsilon$.

From now on, we only need to consider $Q_c(b, v_2, v_3, a)$ with $b \geq v_2 \geq v_3 \geq a$ and $b < v_2 + v_3 + a$. In this case, $0 < p_1 \leq p_2 \leq p_3 \leq p_4$. It can be verified that

(1) if $b > v_2 > a$, then $Q_c(b, v_2, a, a) > \min\{Q_c(b, a, a, a), Q_c(b, b, a, a)\}$,

(2) if $b > v_3 > a$, then $Q_c(b, b, v_3, a) > \min\{Q_c(b, b, b, a), Q_c(b, b, a, a)\}$,

(3) if $b > v_2 > v_3 > a$, then $Q_c(b, v_2, v_3, a) > \min\{Q_c(b, v_2, v_2, a), Q_c(b, v_2, a, a)\}$,

(4) if $b > v = v > a$, then $Q_c(b, v, v, a) > \min\{Q_c(b, b, b, a), Q_c(b, a, a, a)\}$.

In summary, if $v_{c4} < v_{c1} + v_{c2} + v_{c3}$ then

$$Q_c(v_1, v_2, v_3, v_4) \geq \min\{Q_c(b, b, b, a), Q_c(b, b, a, a), Q_c(b, a, a, a)\}.$$

Based on the proof, the minimum of $Q_c(v_1, v_2, v_3, v_4)$ can only be obtained at (b, a, a, a), (b, b, a, a), or (b, b, b, a).

A.4. Proof of Theorem 6

To find out the most robust design in terms of maximum relative loss, we need explicit formulas of $R_{\max}^{(c)}$ for part (ii) of Theorem 5. Following the proof

of Theorem 5,

$$R_{\max}^{(c)} = \max\left\{1 - Q_c(b, a, a, a)^{1/3},\ 1 - Q_c(b, b, a, a)^{1/3},\ 1 - Q_c(b, b, b, a)^{1/3}\right\}.$$

Let $\theta = b/a \geq 1$ and $\rho = \sqrt{\theta^2 - \theta + 1}$. Based on Corollary 1 and Corollary 2,

$$Q_c(b, a, a, a) = Q_c(\theta, 1, 1, 1) = \begin{cases} \frac{27}{\theta}(\theta q_1 + q_2 + q_3 + q_4), & \text{if } \theta \geq 3, \\ \frac{(9-\theta)^2}{4}(\theta q_1 + q_2 + q_3 + q_4), & \text{if } 1 \leq \theta < 3, \end{cases}$$

$$Q_c(b, b, a, a) = Q_c(\theta, \theta, 1, 1) = \frac{108(\theta - 1)^2(\theta q_1 + \theta q_2 + q_3 + q_4)}{(2\theta - 1 - \rho)(\theta - 2 + \rho)(\theta + 1 + \rho)},$$

$$Q_c(b, b, b, a) = Q_c(\theta, \theta, \theta, 1) = \frac{(9\theta - 1)^2}{4\theta^3}(\theta q_1 + \theta q_2 + \theta q_3 + q_4).$$

Given $\theta \geq 1$ and $p_{c1} \geq p_{c2} \geq p_{c3} \geq p_{c4}$, it can be verified that

$$\theta q_1 + q_2 + q_3 + q_4 = \theta p_{c2} p_{c3} p_{c4} + p_{c1} p_{c3} p_{c4} + p_{c1} p_{c2} p_{c4} + p_{c1} p_{c2} p_{c3} \leq \frac{\theta + 3}{64},$$

where the second equality holds if and only if $p_{c1} = p_{c2} = p_{c3} = p_{c4} = 1/4$. Similarly, $\theta q_1 + \theta q_2 + q_3 + q_4 \leq 2(\theta + 1)/64$, where equality holds if and only if $p_{c1} = p_{c2} = p_{c3} = p_{c4} = 1/4$; $\theta q_1 + \theta q_2 + \theta q_3 + q_4 \leq (3\theta + 1)/64$ where equality holds if and only if $p_{c1} = p_{c2} = p_{c3} = p_{c4} = 1/4$. Therefore, if $v_{c4} < v_{c1} + v_{c2} + v_{c3}$, $\min\{Q_c(b,a,a,a), Q_c(b,b,a,a), Q_c(b,b,b,a)\}$ attains its maximum only at $p_{c1} = p_{c2} = p_{c3} = p_{c4} = 1/4$. In other words, the uniform design has smaller $R^{(c)}_{\max}$ than any other design $\mathbf{p}_c$ with $v_{c4} < v_{c1} + v_{c2} + v_{c3}$.

On the other hand, it can be verified that if $\theta \geq 3$,

$$R^{(u)}_{\max} = 1 - \frac{3}{4}\left(1 + \frac{3}{\theta}\right)^{1/3} > 1 - \frac{1}{3\theta}\left(\frac{9\theta - 1}{2}\right)^{2/3} = R^{(c)}_{\max}$$

for any design with $v_{c4} \geq v_{c1} + v_{c2} + v_{c3}$.

In short, $R^{(c)}_{\max}$ attains its minimum only at the uniform design.

References

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586-602.

Collett, D. (1991). *Modelling Binary Data*. Chapman and Hall/CRC, New York.

Crowder, M. J. (1978). Beta-binomial anova for proportions. *Appl. Statist.* **27**, 34-37.

Fotiou, I. A., Parrilo, P. A. and Morari, M. (2005). Nonlinear parametric optimization using cylindrical algebraic decomposition. In *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005*, 3735-3740.

Graßhoff, U. and Schwabe, R. (2008). Optimal design for the Bradley-Terry paired comparison model. *Statist. Methods Appl.* **17**, 275-289.

Hamada, M. and Nelder, J. A. (1997). Generalized linear models for quality-improvement experiments. *J. Quality Tech.* **29**, 292-304.

Hoblyn, T. N. and Palmer, R. C. (1934). A complex experiment in the propagation of plum rootstocks from root cuttings: season, 1931-1932 *J. Pomology and Horticultural Sci.* **12**, 36-56.

Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. M.Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois.

Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models*, (Edited by J. N. Srivastava), 333-353. North-Holland, Amsterdam .

Khuri, A. I., Mukherjee, B., Sinha, B. K. and Ghosh, M. (2006). Design issues for generalized linear models: A Review. *Statistical Science* **21**, 376-399.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. *Proceedings of* 2nd *Berkeley Symposium*, 481-492. Berkeley: University of California Press.

Martin, B., Parker, D. and Zenick, L. (1987). Minimize slugging by optimizing controllable factors on topaz windshield molding. In *Fifth Symposium on Taguchi Methods, American Supplier Institute, Inc.* 519-526. Dearborn, MI.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models.* 2nd edition. Chapman and Hall/CRC, Boca Raton.

Myers, R. M., Montgomery, D. C. and Vining, G. G. (2002). *Generalized Linear Models with Applications in Engineering and Statistics.* John Wiley, New York.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization.* Springer, New York.

Rao, C. R. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *J. Roy. Statist. Soc.* **9**, 128-139.

Rao, C. R. (1971). Unified theory of linear estimation (Corr: 72V34 p194; 72V34 p477). *Sankhyā, Ser. A, Indian J. Statist.* **33**, 371-394.

Smith, W. (1932). The titration of antipneumococcus serum. *J. Pathology* **35**, 509-526.

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, IL 60607-7045, USA.

E-mail: jyang06@math.uic.edu

Department of Statistics, University of Georgia, Athens, GA 30602-1952, USA.

E-mail: amandal@stat.uga.edu

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, IL 60607-7045, USA.

E-mail: dibyen@uic.edu

(Received April 2010; accepted March 2011)