Journal of Statistical Planning and Inference 81 (1999) 383–392

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Simultaneous confidence intervals in the analysis of orthogonal saturated designs

Daniel T. Voss *, Weizhen Wang

Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435-0001, USA

Received 23 June 1998; received in revised form 19 February 1999; accepted 5 March 1999

Abstract

We present the first known method of constructing exact simultaneous confidence intervals for the analysis of orthogonal, saturated factorial designs. Given m independent, normally distributed, unbiased estimators of treatment contrasts, if there is an independent chi-squared estimator of error variance, then simultaneous confidence intervals based on the Studentized maximum modulus distribution are exact under all parameter configurations. In this paper, an analogous method is developed for the case of an orthogonal saturated design, for which the treatment contrasts are independently estimable but there is *no* independent estimator of error variance. Lacking an independent estimator of the error variance, the smallest sums of squares of effect estimators are pooled. The simultaneous confidence intervals are based on a probability inequality, for which the simultaneous confidence coefficient is achieved in the null case.

MSC: primary 62F25; 62J15; 62K15; secondary 62F35

Keywords: Effect sparsity; Robust; Stochastically increasing; Stochastically non-increasing

1. Introduction and summary

A design is *saturated* if the model being fit consumes all degrees of freedom, leaving none to estimate error variability. Saturated designs include highly fractionated factorial designs, such as main effect plans packed with factors, and other orthogonal arrays for sufficiently low-order models, as well as single replicate factorial designs if a full model is fit to the data.

Saturated fractional factorial designs are commonly used and highly effective as screening experiments in industry. Coupled with the limitations of known methods of analysis, this makes advancement of methods of analysis of saturated designs an

* Corresponding author. Tel.: +1-937-7752958; fax: +1-937-7752081.

E-mail address: dvoss@euler.wright.edu (D.T. Voss)

PII: S0378-3758(99)00069-5

important problem in mathematical statistics. The problem has a somewhat long history, though the most significant advances have been relatively recent.

All methods of analysis of saturated designs rely on *effect sparsity* – namely, the condition that only a small proportion of the effects under study are indeed active, or non-zero. The problem is to identify which effects are.

In the analysis of an experiment, it is prudent to use *both* individual and simultaneous inference procedures. The analyses provide different useful information, and there is no additional experimental cost to do both analyses. Individual tests are more powerful than simultaneous tests, but they are more susceptible to false positives. Similarly for confidence intervals. Thus, effects which are significantly nonzero based on individual inference procedures merit further investigation. On the other hand, effects which are significantly non-zero based on simultaneous inference procedures are more certainly real effects. Only simultaneous confidence intervals are discussed here. Corresponding individual confidence interval procedures are provided by Voss (1999).

A popular and longstanding method of analysis of saturated designs is the subjective interpretation of normal or half-normal probability plots – a method introduced by Daniel (1959). However, during the past 15 years, there has been great interest in quantitative methods of assessing significance in fractional factorial designs and other saturated designs.

The literature on hypothesis testing includes both frequentist and Bayesian methods. Frequentist methods of testing for active effects have been proposed or considered by Birnbaum (1959), Daniel (1959), Zahn (1975a,b), Voss (1988), Benski (1989), Lenth (1989), Berk and Picard (1991), Loh (1992), Juan and Peña (1992), Schneider et al. (1993), Dong (1993), Torres (1993), Haaland and O'Connell (1995), Venter and Steel (1996), and Voss (1999). For the most part, the methods proposed are intuitively appealing and have been evaluated empirically, but analytical justification is generally lacking. Concerning non-frequentist methods, Box and Meyer (1986) introduced a Bayesian method for computing posterior probabilities of effects being active. This methodology was extended by Box and Meyer (1993) and Chipman et al. (1997) to better account for possible interactions and for effects more specific than collective factor effects. Relations between main effects and interactions were accounted for by computing the posterior probabilities of competing models, from which are computed the posterior probabilities of effects being active.

The literature on confidence interval estimation of effects for saturated designs is even more sparse. Lenth (1989) proposed an heuristically appealing, adaptive method of constructing confidence intervals, which has been improved upon by Haaland and O'Connell (1995). However, the resulting intervals are empirical – control of confidence levels using adaptive methods remains an open problem, as discussed by Voss (1999). Schneider et al. (1993) proposed asymptotic confidence interval methods. However, the only confidence intervals known to maintain the specified confidence coefficient under all parameter configurations and standard model assumptions are the individual confidence intervals provided by Voss (1999), reviewed in Section 2. From his individual confidence intervals, simultaneous confidence intervals can be obtained by the Bonferroni method, but this approach yields conservative intervals.

In Section 3 we improve upon the results of Voss (1999) to obtain families of simultaneous confidence intervals which are exact in the sense that the corresponding probability bound is achieved. The approach we use is essentially an extension of that based on the Studentized maximum modulus distribution used by Hochberg and Tamhane (1987, p. 136) for a non-saturated design, but in the current setting a stochastic ordering argument is needed.

The fundamental result of this paper (Theorem 1) is to establish a useful stochastic ordering of distributions. The result may seem intuitively obvious, even without certain distributional assumptions included in the theorem. To show otherwise, we present a counterexample to the theorem in Section 4 when some of those assumptions fail. Further remarks are provided in Section 5. In the next section, we provide some additional background.

2. Background and terminology

Let $f_i(x_i)$ be the p.d.f. of a continuous, unimodal distribution which is symmetric about zero with known finite non-zero variance, a_i^2 say, for $i=1,2,\ldots,m$. Consider m (stochastically) independent estimators $\hat{\theta}_i$ $(i=1,2,\ldots,m)$, where

$$\hat{\theta}_i \sim (1/\sigma) f_i((\hat{\theta}_i - \theta_i)/\sigma)$$

for unknown $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ and unknown $\sigma > 0$. Thus, (θ_i, σ) are location-scale parameters, and the estimators could for example be normally distributed. It follows that $E[\hat{\theta}_i] = \theta_i$ and $\text{Var}(\hat{\theta}_i) = a_i^2\sigma^2$. The sum of squares for $\hat{\theta}_i$ is $\text{SS}_i = (\hat{\theta}_i/a_i)^2$.

For example, consider the analysis of data collected using an orthogonal, saturated 2^{n-p} fractional factorial design, assuming the observations to be independently normally distributed with homogeneous variance σ^2. Then the usual factorial treatment contrast estimators are independently normally distributed, each with variance $4\sigma^2/2^{n-p}$. Even if the observations are not normally distributed, each treatment contrast is estimated by the difference between two terms, each being the average of 2^{n-p-1} independent observations. Hence, by the central limit theorem, each estimator is approximately normally distributed.

The problem is to construct simultaneous confidence intervals for the m parameters θ_i.

Following Casella and Berger (1990, p. 404), given data $\boldsymbol{X}=(X_1,\ldots,X_n)$, a vector of m parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, and m corresponding interval estimators $[L_i(\boldsymbol{X}), U_i(\boldsymbol{X})]$ of the parameters θ_i, the *simultaneous confidence coefficient* of the interval estimators is the infimum of the coverage probabilities,

$$\inf_{\boldsymbol{\theta}} P_{\boldsymbol{\theta}}(\theta_i \in [L_i(\boldsymbol{X}), U_i(\boldsymbol{X})],\ 1 \leqslant i \leqslant m).$$

If the infimum is $1-\alpha$, then we call the resulting intervals $[l_i(\boldsymbol{x}), u_i(\boldsymbol{x})]$ *exact simultaneous* $100(1-\alpha)\%$ *confidence intervals* for the θ_i. If the infimum exceeds $1-\alpha$, then we call the resulting intervals *conservative simultaneous* $100(1-\alpha)\%$ *confidence intervals* for the θ_i. Equivalently, we say that the confidence intervals *strongly control* the simultaneous confidence level to be at least $100(1-\alpha)\%$ if the infimum is at least $1-\alpha$.

The problem considered here is complicated by the lack of an independent estimate of error variance. Such is generally the case, however, for saturated fractional factorial designs. To circumvent this difficulty, we use the smallest sums of squares of the estimators to construct a term analogous to a mean-squared error. For example, for fixed i $(i=1,2,\ldots,m)$ and fixed v, pooling together into a mean square the v smallest sums of squares excluding SS_i, we obtain the *quasi-mean-squared error*:

$$\mathrm{QMSE}_i = (1/v)\sum_{j=1}^{v} \mathrm{SS}_{(j),i},$$

where $\mathrm{SS}_{(1),i} < \mathrm{SS}_{(2),i} < \cdots < \mathrm{SS}_{(m-1),i}$ are the order statistics of the $m-1$ sums of squares SS_j for $j \neq i$.

The derivation of confidence intervals given saturated designs depends on notions of stochastic ordering. Lehmann (1959, p. 73) called the distribution of a random variable $X_\theta \sim F_\theta(x)$ *stochastically decreasing* in the parameter θ if $F_\theta(x)$ is non-decreasing in θ and $\theta \neq \theta'$ implies the distributions of X_θ and $X_{\theta'}$ are not identical. A weaker notion of stochastic ordering suffices for our purposes.

The distribution of a random variable $X_\theta \sim F_\theta(x)$ is said to be *stochastically non-increasing* in the parameter θ if $F_\theta(x) = P_\theta(X \leqslant x)$ is non-decreasing in θ for all x.

Voss (1999) obtained individual $100(1-\alpha)\%$ confidence intervals as follows. The quantity

$$V_i^2 = (\hat{\theta}_i - \theta_i)^2 / a_i^2(\mathrm{QMSE}_i) \tag{1}$$

is a pivotal quantity with respect to θ_i, and the distribution of V_i^2 is stochastically non-increasing in each $|\theta_j|/\sigma$ for $j \neq i$. Define $v^2(m, v; \alpha)$ to be the upper-α quantile of the *null distribution* of V_i^2 – namely, when $\theta_1 = \theta_2 = \cdots = \theta_m = 0$, so

$$P[V_i^2 \leqslant v^2(m, v; \alpha) | \theta_j = 0\ \forall j \neq i] = 1-\alpha.$$

It follows that

$$P_{\boldsymbol{\theta},\sigma}[V_i^2 \leqslant v^2(m, v; \alpha)] \geqslant 1-\alpha$$

for all $(\boldsymbol{\theta}, \sigma)$. Hence, an exact individual $100(1-\alpha)\%$ confidence interval for θ_i is

$$\hat{\theta}_i \pm v(m, v; \alpha) a_i \sqrt{\mathrm{qmse}_i},$$

where qmse_i is the observed value of QMSE_i.

In the next section, we extend this result to show that the distribution of

$$W^2 = \max_i \{V_i^2\} \tag{2}$$

is stochastically non-increasing in each $|\theta_i|/\sigma$, from which we obtain exact simultaneous confidence intervals for the θ_i $(i=1,2,\ldots,m)$.

3. Simultaneous confidence intervals

Let V_i^2 and W^2 be as defined in (1) and (2). In this section, we prove the following theorem.

Theorem 1. *Let $\hat{\theta}_i \sim (1/\sigma) f_i((x_i-\theta_i)/\sigma)$ $(i=1,2,\ldots,m)$ be independently distributed random variables with finite variance $\mathrm{Var}(\hat{\theta}_i)=a_i^2\sigma^2$ for known constants a_i, where each $f_i(x_i)$ is a continuous, unimodal p.d.f. which is symmetric with respect to zero. Then*:

(i) *the distribution of $W^2=\max_i\{V_i^2\}$ is stochastically non-increasing with respect to each $|\theta_i|/\sigma$;*

(ii) *$P_{\boldsymbol{\theta},\sigma}[W^2 \leqslant w^2(m,v;\alpha)] \geqslant 1-\alpha$ for all $\boldsymbol{\theta}$ and σ, where $w^2(m,v;\alpha)$ is the upper-α percentile of the null distribution of W^2;*

(iii) *the intervals*

$$\hat{\theta}_i \pm w(m,v;\alpha)a_i\sqrt{\mathrm{qmse}_i} \tag{3}$$

are exact simultaneous $100(1-\alpha)\%$ confidence intervals for the θ_i $(i=1,2,\ldots,m)$.

Theorem 1 can be easily generalized in the following ways.

Corollary 1. *The intervals*

$$\hat{\theta}_i \pm w_1(m,\boldsymbol{b};\alpha)a_i\sqrt{\mathrm{gqmse}_i}$$

are exact simultaneous $100(1-\alpha)\%$ confidence intervals for the θ_i $(i=1,2,\ldots,m)$, where $w_1^2(m,\boldsymbol{b};\alpha)$ is the upper-α percentile of the null distribution of $W_1^2=\max_i\{(\hat{\theta}_i-\theta_i)^2/a_i^2(\mathrm{GQMSE}_i)\}$, gqmse_i is the observed value of ith generalized quasi-mean-squared error

$$\mathrm{GQMSE}_i=\sum_{j=1}^{m-1} b_j\,\mathrm{SS}_{(j),i}$$

and $\boldsymbol{b}=(b_1,b_2,\ldots,b_{m-1})$ for fixed non-negative scalars b_j not all zero.

Corollary 2. *The intervals*

$$\hat{\theta}_i \pm w_2(m,\boldsymbol{b};\alpha)a_i(\mathrm{qse}_i)$$

are exact simultaneous $100(1-\alpha)\%$ confidence intervals for the θ_i $(i=1,2,\ldots,m)$, where $w_2(m,\boldsymbol{b};\alpha)$ is the upper-α percentile of the null distribution of $W_2=\max_i\{|\hat{\theta}_i-\theta_i|/a_i\,\mathrm{QSE}_i\}$, qse_i is the observed value of the ith quasi-standard error

$$\mathrm{QSE}_i=\left(\sum_{j=1}^{m-1} b_j|\hat{\gamma}|_{(j),i}\right)$$

computed from the order statistics $|\hat{\gamma}|_{(1),i} < |\hat{\gamma}|_{(2),i} < \cdots < |\hat{\gamma}|_{(m-1),i}$ *of the* $m-1$ *normalized absolute effect estimators* $\hat{\gamma}_j = |\hat{\theta}_j|/a_j$ *for* $j \neq i$, *and* $\boldsymbol{b} = (b_1, b_2, \ldots, b_{m-1})$ *for fixed non-negative scalars* b_j *not all zero.*

The case of Corollary 2 includes for example the method of Lenth (1989) if used non-adaptively.

The confidence intervals in the theorem and corollaries can be used for the analysis of orthogonal, saturated, single replicate and fraction factorial experiments, to augment or replace the usual subjective assessment based on normal and half-normal probability plots of the effect estimates. The method most naturally applies to experiments conducted using orthogonal, saturated 2^{n-p} designs, as well as other saturated orthogonal arrays for symmetric factorial experiments with two levels per factor. If an experiment involves factors at more than two levels, then the method can be applied to orthogonal contrasts.

An assumption in the theorem is that the distribution of each $\hat{\theta}_i$ is continuous, unimodal and symmetric. This should be at least approximately true in the analysis of 2^{n-p} experiments. In that case, each treatment contrast is estimated by the difference between two terms, each being the average of 2^{n-p-1} independent observations. Hence, by the central limit theorem, each estimator is approximately normally distributed for an experiment of at least modest size.

While the confidence intervals in Eq. (3) are generalized in Corollaries 1 and 2, we recommend use of the former. An empirical comparison of analogous testing procedures found corresponding variations on the statistics to have comparable power (Voss, 1988). Also, the confidence intervals in Eq. (3) enjoy the virtue of simplicity. The confidence interval for θ_i, for example, is obtained by pooling into qmse_i the v smallest sums of squares of the other $m-1$ effects. See Example 7.5.3 of Dean and Voss (1999, p. 216) for an illustration of the method.

The number of effects, v, to be pooled into qmse_i must be determined independently of the data in order for Theorem 1 to apply. The choice of v depends on the amount of effect sparsity anticipated, as well as on the number of estimators, m. If one knew the number of negligible effects, d say, that would be a good choice for the value of v. Any larger value of v would force qmse_i to include sums of squares of non-negligible effects, which would tend to inflate qmse_i and widen confidence intervals. Any smaller value (i.e. $v < d$) would tend to widen confidence intervals because qmse_i would be more variable, but the impact of this is small if v is not too small. For a single replicate 2^n design, if at least one factor has negligible effect on the response variable, then at least 2^{n-1} of the factorial treatment contrasts will be negligible, so setting $v = 2^{n-1}$ is a reasonable choice. For example, this corresponds to use of $v = 8$ for $m = 15$ effects, which seems to be adequate. Correspondingly, critical values $w(m, v; \alpha)$ for Eq. (3) are provided in Table A.11 of Dean and Voss (1999, p. 724) for $v = (m+1)/2$ for m odd and $v = m/2$ for m even, for $m = 2, 3, \ldots, 63$.

The confidence intervals using the quasi-mean-squared error, qmse_i, can easily be computed using standard statistical software as follows. To do so, first specify the number of effects v to be pooled into qmse_i, as discussed previously. Secondly, fit the full model to the data, generating the m sums of squares, each with one degree of freedom, corresponding to the m orthogonal treatment contrasts under study. Finally, fit a reduced model which excludes the treatment contrasts corresponding to the v smallest sums of squares. For each treatment contrast θ_i *not* excluded from the model, the confidence interval is then

$$\hat{\theta}_i \pm w(m, v; \alpha)\, s(\hat{\theta}_i),$$

where $s^2(\hat{\theta}_i) = a_i^2(\text{qmse}_i)$ is a *quasi-standard error* of $\hat{\theta}_i$ computed using qmse_i equal to the mean-squared error, mse, of the fitted model. This is illustrated by Dean and Voss (1999, p. 224) using the SAS software. Clearly, since mse is computed from the v smallest sums of squares, it is generally a biased estimator of σ^2. So, $s(\hat{\theta}_i)$ does *not* estimate the standard deviation of $\hat{\theta}_i$. However, this is adjusted for by the quantile $w(m, v; \alpha)$.

Proof of Theorem 1. Parts (ii) and (iii) of Theorem 1 follow directly from (i).

We now prove (i). It is trivial that the distribution of W^2 depends on $\theta_1, \theta_2, \ldots, \theta_m$ and σ only through $|\theta_1|/\sigma, |\theta_2|/\sigma, \ldots, |\theta_m|/\sigma$. Without loss of generality, assume $\sigma = 1$. Let $X_i = \hat{\theta}_i - \theta_i$ $(i = 1, 2, \ldots, m)$. It follows that the X_i are independently distributed, continuous random variables, each with a unimodal distribution which is symmetric with respect to zero and not dependent on θ_i, with $\text{Var}(X_i) = a_i^2$. Observe that $\text{SS}_i = (X_i + \theta_i)^2/a_i^2$.

Fix $X_2, X_3, \ldots, X_m$. Then $\text{SS}_2, \text{SS}_3, \ldots, \text{SS}_m$ are also fixed. So is $\text{QMSE}_1 = \sum_{j=1}^{v} v^{-1}\, \text{SS}_{(j),1}$ – denote its value by $c = \text{QMSE}_1$. For each $i > 1$, consider

$$\text{QMSE}_i = \sum_{j=1}^{v} v^{-1}\, \text{SS}_{(j),i}.$$

Each $\text{SS}_{(j),i}$ is non-decreasing in SS_1 for fixed $\text{SS}_2, \text{SS}_3, \ldots, \text{SS}_m$, so QMSE_i is non-decreasing in SS_1. For fixed constant w^2, if there exists a value of $\text{SS}_1 = (X_1 + \theta_1)^2/a_1^2$ such that

$$V_i^2 = (\hat{\theta}_i - \theta_i)^2/a_i^2(\text{QMSE}_i) = X_i^2/a_i^2\text{QMSE}_i \leqslant w^2 \tag{4}$$

is possible, then (4) holds if and only if $|X_1 + \theta_1| \geqslant d_i$ for some constant d_i. Otherwise, set $d_i = \infty$.

Now, consider the stochastic ordering of the distribution of W^2 with respect to $|\theta_1|$, conditioned on the values $(X_2, X_3, \ldots, X_m)$, for fixed $(\theta_2, \theta_3, \ldots, \theta_m)$. Specifically, consider the conditional probability $G(w; \theta_1) = P[W^2 \leqslant w^2 | (X_2, X_3, \ldots, X_m)]$. Then

$$\begin{aligned} G(w; \theta_1) &= P_\theta[V_i^2 \leqslant w^2 \quad \forall i = 1, 2, \ldots, m | (X_2, X_3, \ldots, X_m)] \\ &= P_\theta[X_i^2 \leqslant a_i^2 w^2(\text{QMSE}_i) \quad \forall i = 1, 2, \ldots, m | (X_2, X_3, \ldots, X_m)] \end{aligned}$$

$$= P_{\boldsymbol{\theta}}[|X_1| \leqslant c, |X_1 + \theta_1| \geqslant d_2, |X_1 + \theta_1| \geqslant d_3, \ldots, |X_1 + \theta_1| \geqslant d_m \\ |(X_2, X_3, \ldots, X_m)] \\ = P_{\boldsymbol{\theta}}[|X_1| \leqslant c, |X_1 + \theta_1| \geqslant d | (X_2, X_3, \ldots, X_m)], \quad (5)$$

where $d = \max\{d_2, d_3, \ldots, d_m\}$.

Since the distribution of X_1 is continuous, symmetric and unimodal with respect to zero, it follows from (5) that the (conditional) probability $G(w; \theta_1)$ is nondecreasing in $|\theta_1|$ for each w. Equivalently, the conditional distribution of W^2 given $(X_2, X_3, \ldots, X_m)$ is stochastically non-increasing in $|\theta_1|$. It follows that the unconditional distribution of W^2 is stochastically non-increasing in $|\theta_1|$. Similarly, the unconditional distribution of W^2 is stochastically non-increasing in each $|\theta_i|$. □

4. Insufficient conditions

In Theorem 1, for each i, V_i^2 is a pivotal quantity with respect to θ_i, and the distribution of V_i^2 is stochastically non-increasing with respect to $|\theta_j|/\sigma$ for each $j \neq i$. One might expect these conditions to be sufficient for the distribution of $W^2 = \max_i\{V_i^2\}$ to be stochastically non-increasing with respect to each $|\theta_i|/\sigma$, which would establish the theorem. In this section, we show that these conditions are *not* sufficient. Thus, one or more of the additional distributional assumptions of the theorem are important, these being that the distribution of each estimator is continuous, symmetric, and unimodal.

The following is a counterexample. Consider the following family of bivariate discrete distributions:

$$p(x_1, x_2; \theta_1, \theta_2) = \begin{cases} 0.3 - 0.1\theta_1 - 0.05\theta_2, & (x_1, x_2) = (1 + \theta_1, 1 + \theta_2), \\ 0.2 + 0.1\theta_1 + 0.10\theta_2, & (x_1, x_2) = (1 + \theta_1, 2 + \theta_2), \\ 0.1 + 0.3\theta_1 + 0.05\theta_2, & (x_1, x_2) = (2 + \theta_1, 1 + \theta_2), \\ 0.4 - 0.3\theta_1 - 0.10\theta_2, & (x_1, x_2) = (2 + \theta_1, 2 + \theta_2), \end{cases}$$

where $0 \leqslant \theta_i \leqslant 1$ for $i = 1, 2$. Then $V_1 = X_1 - \theta_1$ is a pivotal quantity with respect to θ_1 and is stochastically non-increasing with respect to θ_2. Likewise, $V_2 = X_2 - \theta_2$ is a pivotal quantity with respect to θ_2 and is stochastically non-increasing with respect to θ_1. Hence, one might expect that $W = \max\{V_1, V_2\}$ is also stochastically non-increasing with respect to each θ_i. Instead, it can be verified that W is stochastically increasing with respect to each θ_i. For simplicity, the reader may restrict attention to $\theta_i \in \{0, 1\}$.

5. Remarks

We have solved an open problem by providing the first exact simultaneous confidence intervals (3) for the analysis of saturated factorial designs. The probability of simultaneous coverage exceeds the specified confidence coefficient when any of the effects θ_i are non-zero, but this seems to be a necessary consequence of the fact that each effect θ_i is a nuisance parameter with respect to the construction of the confidence intervals for each of the other effects θ_j, $j \neq i$. This 'conservative' nature also applies to exact individual tests and confidence intervals, and even to adaptive methods as discussed below. The critical coefficients $w(m, v; \alpha)$ required in Eq. (3) are available in Table A.11 of Dean and Voss (1999, p. 724).

Adaptive methods along the lines of Lenth (1989) are quite popular in the literature, perhaps because of their heuristic appeal. One would anticipate non-adaptive methods to be preferable when there are fewer non-zero effects but adaptive methods to be more effective otherwise. Somewhat surprisingly, a simulation study by Voss (1999) shows that the operating characteristics of adaptive and non-adaptive methods are similar even when four of 15 effects are non-zero, using $v = 8$. Since Lenth (1989) (adaptively) estimates the standard error using the median of the absolute effect estimates, if one takes v to be half m (rounded to an integer) for the method of this paper, then the two methods have the same break-down point – having more than half of the effects non-negligible causes the performance of both methods to deteriorate. It is an interesting but challenging open problem to show that adaptive methods along the lines of Lenth (1989) provide exact or conservative confidence intervals, either for individual or simultaneous confidence intervals. Voss (1999) gave a counterexample to show that the stochastic ordering argument he used to obtain exact individual non-adaptive confidence intervals fails in the case of adaptive confidence intervals.

For saturated designs, it is a more difficult problem to obtain exact or conservative confidence intervals than to obtain size-α tests. Voss (1988) provided simultaneous tests based on test statistics of the form $\mathrm{SS}_i/\mathrm{QMSE}$, where each test statistic denominator QMSE is a quasi-mean-squared error constructed using the order statistics of the sums of squares of *all* m estimators. That suggests consideration of the quantities

$$Q_i = (\hat{\theta}_i - \theta_i)^2 / a_i^2 \,\mathrm{QMSE} \tag{6}$$

$(i = 1, 2, \ldots, m)$ for the construction of confidence intervals. However, this approach fails because the required stochastic ordering is lost. For example, suppose $m = 2$ and $\mathrm{QMSE} = (\mathrm{SS}_{(1)} + \mathrm{SS}_{(2)})/2$, and without loss of generality take $a_1 = a_2 = 1$. If $\theta_1 = \theta_2 = 0$, then $Q_i = (\hat{\theta}_i - \theta_i)^2/\mathrm{QMSE} = \mathrm{SS}_i/\mathrm{QMSE} \leqslant 2$ almost surely, for $i = 1, 2$. However, for $|\theta_i| \neq 0$, Q_i as defined in (6) can exceed 2 with positive probability. Hence, $\max_i\{Q_i\}$ fails to be stochastically non-increasing in each $|\theta_i|/\sigma$.

Acknowledgements

The authors would like to thank an anonymous referee for helpful comments.

References

Benski, H.C., 1989. Use of a normality test to identify significant effects in factorial designs. J. Qual. Technol. 21, 174–178.

Berk, K.N., Picard, R.R., 1991. Significance tests for saturated orthogonal arrays. J. Qual. Technol. 23, 79–89.

Birnbaum, A., 1959. On the analysis of factorial experiments without replication. Technometrics 1, 343–357.
Box, G.E.P., Meyer, R.D., 1986. An analysis for unreplicated fractional factorials. Technometrics 28, 11–18.
Box, G.E.P., Meyer, R.D., 1993. Finding the active factors in fractionated screening experiments. J. Qual. Technol. 25, 94–105.
Casella, G., Berger, R.L., 1990. Statistical Inference. Wadsworth & Brooks/Cole, Pacific Grove.
Chipman, H., Hamada, M., Wu, C.F.J., 1997. A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. Technometrics 39, 372–381.
Daniel, C., 1959. Use of half-normal plots in interpreting factorial two-level experiments. Technometrics 1, 311–341.
Dean, A.M., Voss, D.T., 1999. Design and Analysis of Experiments. Springer, New York.
Dong, F., 1993. On the identification of active contrasts in unreplicated fractional factorials. Statist. Sin. 3, 209–217.
Haaland, P.D., O'Connell, M.A., 1995. Inference for effect-saturated fractional factorials. Technometrics 37, 82–93.
Hochberg, Y., Tamhane, A.C., 1987. Multiple Comparison Procedures. Wiley, New York.
Juan, J., Peña, D., 1992. Commun. Statist. Theory Methods 21, 1383–1403.
Lehmann, E.L., 1959. Testing Statistical Hypotheses. Wiley, New York.
Lenth, R.V., 1989. Quick and easy analysis of unreplicated factorials. Technometrics 31, 469–473.
Loh, W.Y., 1992. Identification of active contrasts in unreplicated factorial experiments. Comput. Statist. Data Anal. 14, 135–148.
Schneider, H., Kasperski, W.J., Weissfeld, L., 1993. Finding significant effects for unreplicated fractional factorials using the *n* smallest contrasts. J. Qual. Technol. 25, 18–27.
Torres, V.A., 1993. A simple analysis of unreplicated factorials with possible abnormalities. J. Qual. Technol. 25, 183–187.
Venter, J.H., Steel, S.J., 1996. A hypothesis-testing approach toward identifying active contrasts. Technometrics 38, 161–169.
Voss, D.T., 1988. Generalized modulus-ratio tests for analysis of factorial designs with zero degrees of freedom for error. Commun. Statist. Theory Methods 17, 3345–3359.
Voss, D.T., 1999. Analysis of orthogonal saturated designs. J. Statist. Plan. Inference 78, 111–130.
Zahn, D.A., 1975a. Modifications of and revised critical values for the half-normal plots. Ph.D. Thesis, Harvard Univ.
Zahn, D.A., 1975b. An empirical study of the half-normal plot. Technometrics 17, 201–211.