

Confidence Sets for Optimal Factor Levels of a Response Surface

Fang Wan^{1,*}, Wei Liu², Frank Bretz³ and Yang Han⁴

¹Lancaster University, UK

²University of Southampton, UK

³Novartis Pharma AG, Switzerland

⁴University of Exeter, UK

**email*: f.wan@lancaster.ac.uk

SUMMARY: Construction of confidence sets for the optimal factor levels is an important topic in response surfaces methodology. In Wan *et al.* (2015), an exact $(1 - \alpha)$ confidence set has been provided for a maximum or minimum point (i.e. an optimal factor level) of a univariate polynomial function in a given interval. In this paper, the method has been extended to construct an exact $(1 - \alpha)$ confidence set for the optimal factor levels of response surfaces. The construction method is readily applied to many parametric and semi-parametric regression models involving a quadratic function. A conservative confidence set has been provided as an intermediate step in the construction of the exact confidence set. Two examples are given to illustrate the application of the confidence sets. The comparison between confidence sets indicates that our exact confidence set is better than the only other confidence set available in the statistical literature that guarantees the $(1 - \alpha)$ confidence level.

KEY WORDS: Parametric regression; Response surface; Semi-parametric regression; Statistical inference; Statistical simulation.

1 Introduction

When studying response surfaces, the researcher is often interested in finding the optimal factor levels in constrained experimental regions. For example, in medical and pharmaceutical studies, the interest often lie in finding the dosage that optimizes the treatment; in chemical reaction experiments, it is important to decide the factor levels, such as reaction temperature and proportion of chemicals combined, that optimize the reaction process. A response surface is typically expressed as a quadratic polynomial model

$$Y = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x} + e \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_q]'$ is the design variables, Y is the response, $e \sim N(0, \sigma^2)$ is the noise, and β_0 ,

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \beta_{11} & \frac{1}{2}\beta_{12} & \cdots & \frac{1}{2}\beta_{1q} \\ \frac{1}{2}\beta_{12} & \beta_{22} & \cdots & \frac{1}{2}\beta_{2q} \\ \vdots & & & \\ \frac{1}{2}\beta_{1q} & \frac{1}{2}\beta_{2q} & \cdots & \beta_{qq} \end{bmatrix}$$

are the unknown parameters. The optimal factor level of the response surface is either the maximum point or the minimum point of the response function, depending on the problem under consideration. For example, if the response is the efficacy of certain treatments, then the optimal factor level is the maximum point of the response function; on the other hand, if the response is the toxicity of a medicine, then the optimal factor level should be the minimum point. The point estimation of the optimal factor levels of the response surface is a simple calculus problem for the fitted response surface model; see, for example, Bliss (1970, pp44-50), Studier *et al.* (1975) and Zar (1999, pp458-459). Due to the sampling variation, a confidence region for the optimal factor level is much more useful (cf. Myers and Montgomery, 1995 and Cahya *et al.*, 2004). However, a confidence region is much more difficult to construct than a point estimate.

In this paper, without loss of generality, we focus on the construction of an exact $(1 - \alpha)$

confidence set for a maximum point of the response function in (1) within a pre-specified experimental region. The construction of a confidence set for a minimum point of the response function can be simply translated into the construction of a confidence set for a maximum point of $-(\beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x})$. It is noteworthy that the maximum point of the response function in (1) has nothing to do with β_0 , therefore the focus is on the construction of a confidence set for a maximum point of

$$f(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x} = \mathbf{z}(\mathbf{x})'\boldsymbol{\theta} \quad (2)$$

where $\mathbf{z}(\mathbf{x}) = [x_1, x_2, \dots, x_q, x_1^2, x_2^2, \dots, x_q^2, x_1x_2, x_1x_3, \dots, x_{q-1}x_q]'$ and

$\boldsymbol{\theta} = [\beta_1, \beta_2, \dots, \beta_q, \beta_{11}, \beta_{22}, \dots, \beta_{qq}, \beta_{12}, \beta_{13}, \dots, \beta_{q-1,q}]'$. The dimension of both $\mathbf{z}(\mathbf{x})$ and $\boldsymbol{\theta}$ is $p \times 1$ where $p = q(q + 3)/2$.

Although the method provided in this paper is based on the linear regression model in (1), it can be extended to many parametric and semi-parametric models that involve a linear function of $\boldsymbol{\theta}$ as illustrated in Section 3. If the estimates of the coefficients in the linear function of these models are normally distributed, then our method can directly be applied to produce an exact $(1 - \alpha)$ level confidence set for a maximum point. However, in many statistical models such as generalized linear models, random effects linear models and random effects generalized linear models (cf. Dobson, 2001, Pinheiro *et al.*, 2000 and McCulloch *et al.*, 2001), the maximum likelihood estimators of fixed effects regression coefficients only have large sample approximate normal distributions. In this case, our method can be applied to produce a $(1 - \alpha)$ asymptotic confidence set.

A method of constructing a confidence set for the solution of a set of simultaneous equations is given in Box and Hunter (1954). Because a stationary point can be found by solve a set of simultaneous equations, Box and Hunter's method (referred to as BH henceforth) is used to construct a confidence set for a stationary point over the whole covariate space. This confidence set is sometimes miss-used as a confidence set for a maximum point; it is clear

that a stationary point is not necessarily a maximum point, for it can be an inflection point or a minimum point (Del Castillo and Cahya, 2001). Carter *et al.* (1982) applies this method to the construction of a confidence set for the optimal treatment of cancer and discussed how the confidence set can be used to assess the therapeutic synergism of drug components. By using the Lagrange multiplier, Stablein *et al.* (1983) modifies the BH approach and provides a confidence set for a maximum point within a given covariate region. However, it has ignored the sampling variability of the multipliers and hence the nominal $(1 - \alpha)$ level is not guaranteed, as pointed out by Cahya, Del Castillo and Peterson (2004) (referred to as CDP hereafter). Peterson, Cahya and Del Castillo (2002) (referred to as PCD henceforth) provides a general approach on the construction of confidence sets for a maximum point of $f(\mathbf{x}, \boldsymbol{\theta})$, while the paper CDP proposes a method of reducing the computation time of PCD confidence set. However, the critical constant $\sqrt{qf_{q,\nu}^\alpha}$ used in PCD and CDP, is too small and does not guarantee the $(1 - \alpha)$ confidence level. Rao's method (Rao, 1973, pp.473) of constructing confidence sets can be used to produce the confidence set for a maximum point and it is the only method available in the statistical literature that guarantees the nominal $(1 - \alpha)$ confidence level. However, Rao's confidence set is often too conservative; see further discussion in Section 2.5.

2 Method

2.1 Construction Method

The confidence set construction method we use is first given by Neyman (1937) and utilizes a family of acceptance sets, and has been introduced in many statistical textbooks (cf. Lehmann, 1986, pp.214, Rao, 1973, pp.471, and Casella and Berger, 2002, pp.420-422). It has been widely used and extended in the construction of intriguing confidence sets (cf. Stefansson *et al.*, 1988, Hayter and Hsu, 1994, and Finner and Strassberge, 2002). Both the

BH and PCD confidence sets use this method. Wan *et al.* (2015) also uses this method to construct a confidence set for a maximum point of a univariate polynomial function in a given covariate interval. The key idea of Neyman's (1937) method is that a $(1 - \alpha)$ level confidence set for an unknown parameter $\boldsymbol{\theta}$ or, more generally, a function of the unknown parameter $k = k(\boldsymbol{\theta})$ is given by the set of k_0 values such that the null hypothesis $H_0 : k = k_0$ is not rejected by a size α test. Specifically, let $A(k_0)$ be the acceptance set of a size α test of $H_0 : k = k_0$, i.e. $P_{\boldsymbol{\theta}_0}\{\mathbf{Y} \in A(k_0)\} \geq 1 - \alpha$ in which \mathbf{Y} is the random observation, and the probability is calculated at $\boldsymbol{\theta}_0$ such that $k_0 = k(\boldsymbol{\theta}_0)$. Then the confidence set is given by $C(\mathbf{Y}) = \{k_0 : \mathbf{Y} \in A(k_0)\}$. Furthermore, if the acceptance set $A(k_0)$ is of exact $(1 - \alpha)$ level for each k_0 then the confidence level of $C(\mathbf{Y})$ is exactly $(1 - \alpha)$.

Neyman's method can directly be applied to construct an exact $(1 - \alpha)$ confidence set for a maximum point of $f(\mathbf{x}, \boldsymbol{\theta})$ in a given covariate region \mathcal{X}_q . Let $\mathbf{k} = [k_1, k_2, \dots, k_q]'$ be a maximum point of $f(\mathbf{x}, \boldsymbol{\theta})$ in $\mathbf{x} \in \mathcal{X}_q$. Clearly, the value of \mathbf{k} depends on $\boldsymbol{\theta}$. For any given $\mathbf{k}^\circ \in \mathcal{X}_q$, if $\mathbf{k} = \mathbf{k}^\circ$, then we have

$$f(\mathbf{k}^\circ, \boldsymbol{\theta}) - f(\mathbf{x}, \boldsymbol{\theta}) \geq 0, \quad \forall \mathbf{x} \in \mathcal{X}_q.$$

Therefore, for each $\mathbf{k}^\circ \in \mathcal{X}_q$, a $(1 - \alpha)$ level acceptance set for testing the null hypothesis $H_0 : \mathbf{k} = \mathbf{k}^\circ$ can intuitively be set as

$$A(\mathbf{k}^\circ) = \{\mathbf{Y} : f(\mathbf{k}^\circ, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}, \hat{\boldsymbol{\theta}}) \geq -c(\mathbf{k}^\circ)\sqrt{\hat{v}(\mathbf{k}^\circ, \mathbf{x})}, \forall \mathbf{x} \in \mathcal{X}_q\} \quad (3)$$

where $\hat{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$ and $\hat{v}(\mathbf{k}^\circ, \mathbf{x})$ is the related estimate of the variance of $f(\mathbf{k}^\circ, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}, \hat{\boldsymbol{\theta}})$. The critical value $c(\mathbf{k}^\circ)$ is chosen such that the coverage probability of $A(\mathbf{k}^\circ)$ is equal to $(1 - \alpha)$ under H_0 . Then, according to Neyman's method, an exact $(1 - \alpha)$ level confidence set for \mathbf{k} is given by

$$\begin{aligned} \mathbf{C}_E(\mathbf{Y}) &= \{\mathbf{k}^\circ \in \mathcal{X}_q : \mathbf{Y} \in A(\mathbf{k}^\circ)\} \\ &= \left\{ \mathbf{k}^\circ \in \mathcal{X}_q : f(\mathbf{k}^\circ, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}, \hat{\boldsymbol{\theta}}) \geq -c(\mathbf{k}^\circ)\sqrt{\hat{v}(\mathbf{k}^\circ, \mathbf{x})}, \forall \mathbf{x} \in \mathcal{X}_q \right\}. \end{aligned} \quad (4)$$

The key in the construction of $\mathbf{C}_E(\mathbf{Y})$ is therefore the computation of the critical value $c(\mathbf{k}^o)$ for each $\mathbf{k}^o \in \mathcal{X}_q$. As pointed out in Wan *et al.* (2015), the PCD confidence set has a similar form, but they use the critical value $c_\alpha = \sqrt{qf_{q,\nu}^\alpha}$ in the place of $c(\mathbf{k}^o)$ which is supported by limited simulation results only. In Example 1 of Section 4, we demonstrate that c_α can be substantially smaller than $c(\mathbf{k}^o)$ for some $\mathbf{k}^o \in \mathcal{X}_q$ and therefore the PCD confidence set does not guarantee the nominal $(1 - \alpha)$ confidence level. Next, we consider the computation of the critical value $c(\mathbf{k}^o)$.

2.2 Exact Critical Value $c(\mathbf{k}^o)$

Let \mathbf{X} be the usual $n \times (p + 1)$ design matrix of the linear model $Y = \beta_0 + \mathbf{z}(\mathbf{x})'\boldsymbol{\theta} + e$, \mathbf{Y} be the vector of response and \mathbf{V} be a matrix resultant from deleting the first column and first row of the matrix $(\mathbf{X}'\mathbf{X})^{-1}$. Therefore, \mathbf{V} is a $p \times p$ symmetric and positive definite matrix which has a unique square root matrix denoted by \mathbf{P} . The least squares method gives an estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, which has a normal distribution $\hat{\boldsymbol{\theta}} \sim \mathbf{N}(\boldsymbol{\theta}, \Sigma)$ with $\Sigma = \sigma^2\mathbf{V}$. An estimator of σ^2 is given by $\hat{\sigma}^2$, the mean residual sum of squares of the linear model, which has the distribution $\hat{\sigma}^2 \sim \sigma^2\chi_\nu^2/\nu$ with $\nu = n - p - 1$. Hence we have $\mathbf{T} = \mathbf{P}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})/\hat{\sigma} \sim T_{p,\nu}$, a standard p dimensional t -distribution with ν df (cf. Genz and Bretz, 2009). For two points \mathbf{v}_1 and $\mathbf{v}_2 \in \mathcal{X}_q$, define $\mathbf{g}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{z}(\mathbf{v}_1) - \mathbf{z}(\mathbf{v}_2)$. Then the acceptance set $A(\mathbf{k}^o)$ in (3) becomes

$$A(\mathbf{k}^o) = \{\mathbf{Y} : f(\mathbf{k}^o, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}, \hat{\boldsymbol{\theta}}) \geq -c(\mathbf{k}^o)\hat{\sigma}\sqrt{\mathbf{g}(\mathbf{k}^o, \mathbf{x})'\mathbf{P}^2\mathbf{g}(\mathbf{k}^o, \mathbf{x})}, \forall \mathbf{x} \in \mathcal{X}_q\}.$$

The coverage probability of $A(\mathbf{k}^o)$ is therefore given by

$$P\{\mathbf{Y} \in A(\mathbf{k}^o)\} = P\left\{\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} \frac{f(\mathbf{k}^o, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}, \hat{\boldsymbol{\theta}})}{\hat{\sigma} \sqrt{\mathbf{g}(\mathbf{k}^o, \mathbf{x})' \mathbf{P}^2 \mathbf{g}(\mathbf{k}^o, \mathbf{x})}} \geq -c(\mathbf{k}^o)\right\} \quad (5)$$

$$\geq P\left\{\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} \frac{[f(\mathbf{k}^o, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}, \hat{\boldsymbol{\theta}})] - [f(\mathbf{k}^o, \boldsymbol{\theta}) - f(\mathbf{x}, \boldsymbol{\theta})]}{\hat{\sigma} \sqrt{\mathbf{g}(\mathbf{k}^o, \mathbf{x})' \mathbf{P}^2 \mathbf{g}(\mathbf{k}^o, \mathbf{x})}} \geq -c(\mathbf{k}^o)\right\} \quad (6)$$

$$= P\left\{\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} \frac{\mathbf{g}(\mathbf{k}^o, \mathbf{x})' (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{\hat{\sigma} \|\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})\|} \geq -c(\mathbf{k}^o)\right\}$$

$$= P\left\{\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} \frac{[\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})]' [\mathbf{P}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]}{\hat{\sigma} \|\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})\|} \geq -c(\mathbf{k}^o)\right\}$$

$$= P\left\{\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} \frac{[\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})]' \mathbf{T}}{\|\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})\|} \geq -c(\mathbf{k}^o)\right\} \quad (7)$$

where the first inequality in (6) follows from the fact that $f(\mathbf{k}^o, \boldsymbol{\theta}) - f(\mathbf{x}, \boldsymbol{\theta}) \geq 0$ for all $\mathbf{x} \in \mathcal{X}_q$ since \mathbf{k}^o is a maximum point. Also note that

$$\inf_{\boldsymbol{\theta}: \mathbf{k}=\mathbf{k}^o} P_{\mathbf{k}=\mathbf{k}^o}\{\mathbf{Y} \in A(\mathbf{k}^o)\} = P\left\{\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} \frac{[\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})]' \mathbf{T}}{\|\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})\|} \geq -c(\mathbf{k}^o)\right\},$$

with the infimum being attained at $\boldsymbol{\theta} = \mathbf{0}$. Hence the critical value $c(\mathbf{k}^o)$ is the unique solution of

$$P\left\{\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} \frac{[\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})]' \mathbf{T}}{\|\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})\|} \geq -c(\mathbf{k}^o)\right\} = 1 - \alpha. \quad (8)$$

Next we use simulation to compute the critical value $c(\mathbf{k}^o)$ from Equation (8).

Let

$$G(\mathbf{k}^o, \mathbf{x}, \mathbf{T}) = \frac{[\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})]' \mathbf{T}}{\|\mathbf{P} \mathbf{g}(\mathbf{k}^o, \mathbf{x})\|},$$

then the simulation follows the three steps below.

Step 1. Sample independent $\mathbf{T}_i \sim \mathbf{T}_{p, \nu}$, $i = 1, 2, \dots, n_T$, n_T is a specified large number.

Step 2. For each \mathbf{T}_i , $i = 1, 2, \dots, n_T$, compute

$$\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} G(\mathbf{k}^o, \mathbf{x}, \mathbf{T}_i).$$

Since the infimum of $G(\mathbf{k}^o, \mathbf{x}, \mathbf{T}_i)$ is difficult to find analytically, we execute this step by using numerical methods. One approach is first to compute the value $G(\mathbf{k}^o, \mathbf{x}, \mathbf{T}_i)$ at a finite

grid of points \mathbf{x} on \mathcal{X}_q and then use the minimum of these values as an approximation to the infimum. If $f(\mathbf{x}, \boldsymbol{\theta})$ is a bivariate quadratic function, we use a set of grid points evenly spaced in each direction with distance d_1 and d_2 chosen with respect to the region \mathcal{X}_q . If $f(\mathbf{x}, \boldsymbol{\theta})$ is a multivariate quadratic function, one can sample a set of random grid points uniformly from the region \mathcal{X}_q . When the distance is small or the sample size is large, this gives a good approximation to the region \mathcal{X}_q . An alternative way is to compute the infimum using some standard numerical minimization algorithms, such as the steepest descending or simplex methods.

Step 3. Sort the n_T values of $\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} G(\mathbf{k}^o, \mathbf{x}, \mathbf{T}_i)$, $i = 1, 2, \dots, n_T$, in an ascending order, and use the $[\alpha \times n_T]$ th value, $-\hat{c}(\mathbf{k}^o)$, as an approximation to $-c(\mathbf{k}^o)$.

It is clear that when $n_T \rightarrow \infty$, we have $\hat{c}(\mathbf{k}^o) \rightarrow c(\mathbf{k}^o)$. Based on our experience, $n_T = 20,000$ will produce $\hat{c}(\mathbf{k}^o)$ accurate enough for most applications. See example 1 of Section 4 for more information.

2.3 Conservative Critical Value c_0

Since the computation of critical constant $c(\mathbf{k}^o)$ involves simulation, it might be time consuming to construct the confidence set $\mathbf{C}_E(\mathbf{Y})$ using $c(\mathbf{k}^o)$ for each $\mathbf{k}^o \in \mathcal{X}_q$. Hence we provide a conservative critical value $c_0 = \sqrt{pf_{p,\nu}^\alpha}$ which can be used to construct a $(1 - \alpha)$ conservative confidence set $\mathbf{C}_0(\mathbf{Y})$ for the maximum point \mathbf{k} . It is noteworthy that if we construct this conservative confidence set before the exact confidence set, then the computation burden of the exact confidence set is reduced as we only need to find the critical value $c(\mathbf{k}^o)$ for each $\mathbf{k}^o \in \mathbf{C}_0(\mathbf{Y})$ rather than for each $\mathbf{k}^o \in \mathcal{X}_q$. Next, we prove that $c_0 \geq c(\mathbf{k}^o)$ for any $\mathbf{k}^o \in \mathcal{X}_q$.

Since $\mathbf{T} \sim \mathbf{T}_{p,\nu}$, it can be shown that $\frac{\|\mathbf{T}\|^2}{p} \sim F_{p,\nu}$ where $F_{p,\nu}$ denotes the F-distribution with degrees of freedom p and ν . Therefore $P\{\|\mathbf{T}\| \leq c_0\} = 1 - \alpha$ which is equivalent to

$$P\left\{\inf_{\boldsymbol{\rho} \in R^p} \frac{\boldsymbol{\rho}'\mathbf{T}}{\|\boldsymbol{\rho}\|} \geq -c_0\right\} = 1 - \alpha.$$

Since for any $\mathbf{k}^o \in \mathcal{X}_q$, $\mathbf{P}\mathbf{g}(\mathbf{k}^o, \mathbf{x})$ is a vector in R^p for all $\mathbf{x} \in \mathcal{X}_q$, we have

$$P \left\{ \inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} \frac{[\mathbf{P}\mathbf{g}(\mathbf{k}^o, \mathbf{x})]' \mathbf{T}}{\|\mathbf{P}\mathbf{g}(\mathbf{k}^o, \mathbf{x})\|} \geq -c_0 \right\} \geq P \left\{ \inf_{\boldsymbol{\rho} \in R^p} \frac{\boldsymbol{\rho}' \mathbf{T}}{\|\boldsymbol{\rho}\|} \geq -c_0 \right\} = 1 - \alpha. \quad (9)$$

By comparing the probability statements (9) and (8), we conclude $c_0 \geq c(\mathbf{k}^o)$.

With the critical value $c_0 = \sqrt{pf_{p,\nu}^\alpha}$, the conservative confidence set $\mathbf{C}_0(\mathbf{Y})$ is given by

$$\mathbf{C}_0(\mathbf{Y}) = \left\{ \mathbf{k}^o \in \mathcal{X}_q : f(\mathbf{k}^o, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}, \hat{\boldsymbol{\theta}}) \geq -c_0 \hat{\sigma} \sqrt{\mathbf{g}(\mathbf{k}^o, \mathbf{x})' \mathbf{P}^2 \mathbf{g}(\mathbf{k}^o, \mathbf{x})}, \forall \mathbf{x} \in \mathcal{X}_q \right\}. \quad (10)$$

2.4 Computation of the Confidence Sets

In order to construct the confidence sets given in (4) and (10), we need to check each $\mathbf{k}^o \in \mathcal{X}_q$ to see whether it is in the confidence set. Since the region \mathcal{X}_q is usually continuous and hence contains infinite number of points, we choose a finite set of grid points S on the region \mathcal{X}_q as an approximation of \mathcal{X}_q . If the distance d between the neighbouring grid points is small, then S gives a fine approximation to the region \mathcal{X}_q and the resultant confidence sets give an accurate approximation to $\mathbf{C}_0(\mathbf{Y})$ and $\mathbf{C}_E(\mathbf{Y})$. Therefore, we only check each point in S , but not each point in \mathcal{X}_q , in computing the conservative and exact confidence sets. The choice of d depends on the data, that is, a smaller d is needed if the covariate region is small and the plot of response surface suggests there are sharp peaks. One can try different d to see whether there is a change in the resultant confidence set. In general, we suggest to standardise each covariate to either interval $[0, 1]$ or $[-1, 1]$ and then use $d = 0.01$ or $d = 0.02$ accordingly.

As we pointed out in Section 2.3, the conservative $(1 - \alpha)$ confidence set $\mathbf{C}_0(\mathbf{Y})$ in (10) should always be constructed before $\mathbf{C}_E(\mathbf{Y})$ in order to save computation time. To construct $\mathbf{C}_0(\mathbf{Y})$, we need to compute c_0 and check each point $\mathbf{k}^o \in S \subset \mathcal{X}_q$ to see whether the inequality in (10) holds. This can be performed by comparing $\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} G(\mathbf{k}^o, \mathbf{x}, \mathbf{P}^{-1}\hat{\boldsymbol{\theta}}/\hat{\sigma})$ with $(-c_0)$, where the function G is defined in Section 2.2 and $\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} G(\mathbf{k}^o, \mathbf{x}, \mathbf{P}^{-1}\hat{\boldsymbol{\theta}}/\hat{\sigma})$ can be computed by using the numerical method as in finding $\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} G(\mathbf{k}^o, \mathbf{x}, \mathbf{T}_i)$ in Step 2 of Section 2.2. To construct the exact confidence set $\mathbf{C}_E(\mathbf{Y})$, we further check each grid

point \mathbf{k}^o in $\mathbf{C}_0(\mathbf{Y})$ to see whether

$$\inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} G(\mathbf{k}^o, \mathbf{x}, \mathbf{P}^{-1}\hat{\boldsymbol{\theta}}/\hat{\sigma}) \geq -c(\mathbf{k}^o).$$

The computation of $\mathbf{C}_E(\mathbf{Y})$ takes more time than that of $\mathbf{C}_0(\mathbf{Y})$ as it involves the computation of the ‘personalized’ critical constant for each grid point. For a large number of grid points, the computation can be intensive, as a trade-off for accuracy.

2.5 The Confidence Set of Rao (1973)

Rao (1973, pp473) provides the following $(1 - \alpha)$ confidence set for any given function $h(\cdot)$ of $\boldsymbol{\theta}$:

$$\left\{ h(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \mathbf{C}(\hat{\boldsymbol{\theta}}) \right\} \quad (11)$$

where $\mathbf{C}(\hat{\boldsymbol{\theta}})$ is a $(1 - \alpha)$ confidence set for $\boldsymbol{\theta}$. A well known $(1 - \alpha)$ confidence ellipsoid for $\boldsymbol{\theta}$ is given by $\mathbf{C}(\hat{\boldsymbol{\theta}}) = \left\{ \boldsymbol{\theta} : (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{P}^{-2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq p \hat{\sigma}^2 f_{p,\nu}^\alpha \right\}$ since $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2 \mathbf{P}^2)$. Hence a $(1 - \alpha)$ confidence set for a maximum point is given by

$$\mathbf{C}_c(\mathbf{Y}) = \left\{ \arg \max_{\mathbf{x} \in \mathcal{X}_q} f(\mathbf{x}, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbf{C}(\hat{\boldsymbol{\theta}}) \right\}. \quad (12)$$

Note that the computation of Rao’s confidence set $\mathbf{C}_c(\mathbf{Y})$ also needs a finite grid of points to approximate the continuous confidence ellipsoid $\mathbf{C}(\hat{\boldsymbol{\theta}})$. A popular approach to approximate the confidence ellipsoid is by using a set of grid points in polar co-ordinates as in Carter *et al.* (1984) and Farebrother (1998, pp.85-86) .

3 Extension to other models

The method proposed in Section 2 can be applied to other parametric or semi-parametric models that involve a quadratic function. For example, in quantile regression (cf. Koenker, 2005), the q^{th} -quantile may be modelled by

$$Q(q|\mathbf{x}) = \theta_0(q) + f(\mathbf{x}, \boldsymbol{\theta}(q)).$$

Note that the confidence set for a maximum point of $Q(q|\mathbf{x})$ is the confidence set for a maximum point of $f(\mathbf{x}, \boldsymbol{\theta})$. Hence if $f(\mathbf{x}, \boldsymbol{\theta})$ is a quadratic function, then a confidence set can be constructed using our method.

In generalized linear models, linear mixed models and generalized linear mixed models (cf. McCulloch and Searle, 2001 and Faraway, 2006), for example, the mean response $E(\mathbf{Y})$ is often related to a linear function by a given monotonic link function $h_m(\cdot)$, that is

$$h_m[E(Y)] = \theta_0 + f(\mathbf{x}, \boldsymbol{\theta}).$$

Since $h_m(\cdot)$ is monotone, a maximum point of $E(Y)$ is either a maximum point of $f(\mathbf{x}, \boldsymbol{\theta})$ or $-f(\mathbf{x}, \boldsymbol{\theta})$, depending on whether the function $h_m(\cdot)$ is increasing or decreasing. Hence our construction method of a confidence set can be used for a maximum point of the mean response $E(Y)$.

In Cox's proportional hazard model (cf. Cox, 1972 and Cox, 1975), the hazard function $h_{cp}(t, \mathbf{x})$ is related to a linear function $f(\mathbf{x}, \boldsymbol{\theta})$ by

$$h_{cp}(t, \mathbf{x}) = \lambda(t) \exp(f(\mathbf{x}, \boldsymbol{\theta})).$$

Because $h_{cp}(t, \cdot)$ is an increasing function of $f(\mathbf{x}, \boldsymbol{\theta})$ for a fixed t , the confidence set for a maximum point of $h_{cp}(t, \cdot)$ is that of $f(\mathbf{x}, \boldsymbol{\theta})$.

It is noteworthy that when the estimates of parameters are normally distributed $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \Sigma)$, then our method gives an exact $(1 - \alpha)$ confidence set. However, when the distribution of $\hat{\boldsymbol{\theta}}$ is asymptotically $N(\boldsymbol{\theta}, \hat{\Sigma})$, our procedure can directly be applied to produce an asymptotic $(1 - \alpha)$ confidence set. Specifically if we let $\mathbf{P}^2 = \hat{\Sigma}$, then $\mathbf{P}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(\mathbf{0}, \mathbf{I}_p)$ where \mathbf{I}_p is the $p \times p$ identity matrix. Hence in this case, the t-distribution $T_{p,\nu}$ in Section 2 is replaced by the p -dimensional standard normal distribution $N(\mathbf{0}, \mathbf{I}_p)$. In particular, the critical constant $c(\mathbf{k}^o)$ is solved from

$$P \left\{ \inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} \frac{[\mathbf{P}\mathbf{g}(\mathbf{k}^o, \mathbf{x})]' \mathbf{N}}{\|\mathbf{P}\mathbf{g}(\mathbf{k}^o, \mathbf{x})\|} \geq -c(\mathbf{k}^o) \right\} = 1 - \alpha, \quad (13)$$

and the conservative critical constant used in $\mathbf{C}_0(\mathbf{Y})$ should be $\sqrt{\chi_{1-\alpha, p}^2}$.

4 EXAMPLES

Example 1

[Table 1 about here.]

This example is used in PCD and considers the survival data in a murine cancer chemotherapy experiment that used the drugs 5-Fluorouracil and Teniposide originally given in Stablein *et al.* (1983). A series combinations of 5-Fluorouracil (5FU) and Teniposide (VM26) were given to treat 127 mice with leukaemia. The original data of combinations and survival times were recorded in Table 1. The respective doses of 5FU and VM26 were scaled to give

$$x_1 = \frac{5FU(mg/kg) - 130}{130}$$

$$x_2 = \frac{VM26(mg/kg) - 13}{13}.$$

Following Stablein *et al.* (1983) and PCD, a cox-proportional hazard model

$$h_{cp}(t, \mathbf{x}) = \lambda(t) \exp(f(\mathbf{x}, \boldsymbol{\theta}))$$

is fitted to the data, where $f(\mathbf{x}, \boldsymbol{\theta}) = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2$.

[Figure 1 about here.]

The parameter estimates are given by

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} -1.2312 \\ -1.5084 \\ 0.5467 \\ 0.8850 \\ -0.7186 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 0.0558 & 0.0409 & 0.0147 & 0.0121 & 0.0531 \\ * & 0.0552 & 0.0063 & 0.0136 & 0.0538 \\ * & * & 0.0753 & -0.0235 & 0.0337 \\ * & * & * & 0.0703 & 0.0347 \\ * & * & * & * & 0.1037 \end{pmatrix}.$$

It is clearly of interest to construct a confidence set for the dose combination \mathbf{k} at which the hazard function $h_{cp}(t, \mathbf{x})$ is minimized, i.e., $f(\mathbf{x}, -\boldsymbol{\theta})$ is maximized. The method given in Section 3 allows us to construct a confidence set for the optimal dose combination level within the constrained region $\{\mathbf{x}'\mathbf{x} \leq 1\}$. Using distance $d_1 = d_2 = 0.02$, we have in total 10201 grid points representing the region $\{\mathbf{x}'\mathbf{x} \leq 1\}$. Let $\alpha = 0.05$ and simulation number $n_T = 20,000$,

the asymptotic confidence regions $\mathbf{C}_E(\mathbf{Y})$, $\mathbf{C}_0(\mathbf{Y})$ and the asymptotic confidence set using PCD's critical value $\sqrt{\chi_{1-\alpha,2}^2} = 2.4477$ (denoted by $\mathbf{C}_P(\mathbf{Y})$) are computed and depicted in Figure 1. It is clear that $\mathbf{C}_E(\mathbf{Y})$ is smaller than $\mathbf{C}_0(\mathbf{Y})$ and larger than $\mathbf{C}_P(\mathbf{Y})$. In fact, there are 840 grid points in $\mathbf{C}_0(\mathbf{Y})$, 451 in $\mathbf{C}_E(\mathbf{Y})$ and 381 in $\mathbf{C}_P(\mathbf{Y})$. This indicates $\mathbf{C}_E(\mathbf{Y})$ is about 46.31% smaller than $\mathbf{C}_0(\mathbf{Y})$ and $\mathbf{C}_P(\mathbf{Y})$ is about 15.52% smaller than $\mathbf{C}_E(\mathbf{Y})$. Rao's confidence set $\mathbf{C}_c(\mathbf{Y})$ is computed by using 5 evenly spaced grid points in the radial coordinate and 16 evenly spaced grid points in each angular coordinate and is depicted in Figure 1(c). The computation time of each critical constant $c(\mathbf{k}^o)$ on an iMac (Core i5 @3.4GHz) is 17 seconds.

To assess the accuracy of the critical constants $c(\mathbf{k})$ computed, we use the methods given by Edwards and Berry (1987, Lemma 2) and Liu *et al.* (2005). Let

$$W(\mathbf{T}, \mathbf{k}^o) = P\left\{ \inf_{\mathbf{x} \in \mathcal{X}_q \setminus \mathbf{k}^o} G(\mathbf{k}^o, \mathbf{x}, \mathbf{T}) \geq -\hat{c}(\mathbf{k}^o) \right\}.$$

According to Edwards and Berry, $W(\mathbf{T}, \mathbf{k}^o)$ has a beta distribution with mean $(1 - \alpha)$ and variance $\alpha(1 - \alpha)/(n_T + 2)$. Hence when $\alpha = 0.05$ and $n_T = 20,000$, the 3- σ rule indicates that $W(\mathbf{T}, \mathbf{k}^o)$ has its value in $1 - \alpha \pm 3\sqrt{\alpha(1 - \alpha)/(n_T + 2)} = (0.9454, 0.9546)$ with a probability almost equal to one. Because the random variable $\hat{c}(\mathbf{k}^o)$ has an asymptotic normal distribution, the method provided by Liu *et al.* (2005) (see also Liu, 2010, pp243-244) assesses the variation in $\hat{c}(\mathbf{k}^o)$. For $n_T = 20,000$, $\hat{c}(\mathbf{k}^o)$ varies often only at the second decimal place. For example, using the data given by Example 1 and eight different random seeds, we computed $c(\mathbf{k}^o)$ with $\mathbf{k}^o = [0.6, 0.7]'$: 2.5526, 2.5521, 2.5531, 2.5751, 2.5520, 2.5449, 2.5738, 2.5646 and $c(\mathbf{k}^o)$ with $\mathbf{k}^o = [0.4, 0.8]'$: 2.5766, 2.5722, 2.5632, 2.5886, 2.5692, 2.5662, 2.5903, 2.5894. These indicate that the critical constants are accurate to the second decimal place. Based on these observations, the simulation-based critical constants can be regarded as exact for practical purpose.

From the values of $c(\mathbf{k}^o)$ given in the last paragraph, it is clear that the critical value

$\sqrt{\chi_{1-\alpha,2}^2} = 2.4477$ used in PCD can be substantially smaller than the exact $c(\mathbf{k}^o)$. In fact, with the PCD acceptance set

$$A_p(\mathbf{k}^o) = \{\mathbf{Y} : f(\mathbf{k}^o, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}, \hat{\boldsymbol{\theta}}) \geq -2.4477\hat{\sigma}\sqrt{\mathbf{g}(\mathbf{k}^o, \mathbf{x})'\mathbf{P}^2\mathbf{g}(\mathbf{k}^o, \mathbf{x})}, \forall \mathbf{x} \in \mathcal{X}_2\},$$

the size of testing $H_0 : \mathbf{k} = \mathbf{k}^o$ is given by

$$\begin{aligned} \alpha_p(\mathbf{k}^o) &= \sup_{\boldsymbol{\theta}} P \left\{ \inf_{\mathbf{x} \in \mathcal{X}_2 \setminus \mathbf{k}^o} \frac{f(\mathbf{k}^o, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}, \hat{\boldsymbol{\theta}})}{\hat{\sigma}\sqrt{\mathbf{g}(\mathbf{k}^o, \mathbf{x})'\mathbf{P}^2\mathbf{g}(\mathbf{k}^o, \mathbf{x})}} < -2.4477 \mid \mathbf{k} = \mathbf{k}^o \right\} \\ &= P \left\{ \inf_{\mathbf{x} \in \mathcal{X}_2 \setminus \mathbf{k}^o} \frac{[\mathbf{P}\mathbf{g}(\mathbf{k}^o, \mathbf{x})]'\mathbf{N}}{\|\mathbf{P}\mathbf{g}(\mathbf{k}^o, \mathbf{x})\|} < -2.4477 \right\}. \end{aligned}$$

The probability can be approximated accurately by using Monte Carlo method with a large number N_{sim} of random samples $\mathbf{N} \sim \mathbf{N}(\mathbf{0}, I_5)$. With $N_{sim} = 1,000,000$, we have $\alpha_p([0.6, 0.7]') = 6.56\%$ and $\alpha_p([0.4, 0.8]') = 6.52\%$, both substantially larger than $\alpha = 5\%$. Hence the confidence level of PCD confidence set $\mathbf{C}_P(\mathbf{Y})$, which is $1 - \max_{\mathbf{k}^o \in \mathcal{X}_2} \alpha_p(\mathbf{k}^o)$, is substantially smaller than the nominated $(1 - \alpha) = 95\%$.

Example 2

[Table 2 about here.]

This example considers the formulation of a controlled-release drug substance to aid in obtaining more uniform blood levels (Frisbee *et al.*, 1994). The formulation components are recorded in Table 2, where x_1 is the percentage of Pluronic F68, x_2 is the percentage of polyoxyethylene 40 monostearate, x_3 is the percentage of polyoxyethylene sorbitan fatty acid ester NF and $x_1 + x_2 + x_3 = 1$. The response Y is the observed glass transition temperature ($^{\circ}C$) for which a smaller value is better. The response surface model that gives a good fit was a Becker's H1 model (Becker, 1968; Becker 1978):

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} \min(x_1, x_2) + \beta_{13} \min(x_1, x_3) + \beta_{23} \min(x_2, x_3) + e \quad (14)$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}]'$ is the unknown parameter vector and $e \sim N(0, \sigma^2)$ is the

random error. Let $\mathbf{x} = (x_1, x_2)$ and $f(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{z}(\mathbf{x})'\boldsymbol{\theta}$, where

$$\boldsymbol{\theta} = \begin{pmatrix} \beta_1 - \beta_3 \\ \beta_2 - \beta_3 \\ \beta_{12} \\ \beta_{13} \\ \beta_{23} \end{pmatrix} \text{ and } \mathbf{z}(\mathbf{x}) = \begin{pmatrix} x_1 \\ x_2 \\ \min(x_1, x_2) \\ \min(x_1, 1 - x_1 - x_2) \\ \min(x_2, 1 - x_1 - x_2) \end{pmatrix},$$

then the Becker's H1 Model can be written as

$$Y = \beta_3 + f(\mathbf{x}, \boldsymbol{\theta}) + e \quad (15)$$

by replacing x_3 by $1 - x_1 - x_2$. The parameter estimates are given by

$$\begin{pmatrix} \hat{\beta}_3 \\ \hat{\boldsymbol{\theta}} \end{pmatrix} = \begin{pmatrix} 36.2719 \\ -17.6436 \\ -20.8057 \\ -2.8033 \\ -18.0089 \\ 9.7532 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 0.7936 & -0.8008 & -0.8008 & 0.1624 & -0.6384 & -0.6384 \\ * & 1.6013 & 0.8006 & -0.7999 & 0.0007 & 0.8014 \\ * & * & 1.6013 & -0.7999 & 0.8014 & 0.0007 \\ * & * & * & 4.0156 & -0.7844 & -0.7844 \\ * & * & * & * & 4.0170 & -0.7837 \\ * & * & * & * & * & 4.0170 \end{pmatrix}.$$

[Figure 2 about here.]

The interest lies in constructing a confidence set for the formulation factors that minimize the response in the covariate region constrained by $x_1 + x_2 + x_3 = 1$ and $x_1, x_2, x_3 \geq 0$, that is, a maximum point of function $f(\mathbf{x}, -\boldsymbol{\theta})$, in the constrained region

$$R_{cons} = \{\mathbf{x} = (x_1, x_2) : x_1 + x_2 \leq 1, x_1, x_2 \geq 0\}.$$

Using $\alpha = 0.05$, distance $d = 0.01$ and simulation number $n_T = 20,000$, the confidence regions $\mathbf{C}_0(\mathbf{Y})$, $\mathbf{C}_E(\mathbf{Y})$ and $\mathbf{C}_P(\mathbf{Y})$ using PCD's critical value $\sqrt{f_{2,5}^\alpha} = 3.4018$ have been computed and depicted on the $x_1 + x_2 + x_3 = 1$ plane in Figure 2. It is clear that $\mathbf{C}_E(\mathbf{Y})$ is smaller than $\mathbf{C}_0(\mathbf{Y})$ and $\mathbf{C}_c(\mathbf{Y})$ but larger than $\mathbf{C}_P(\mathbf{Y})$. In fact, among the total 5151 grid points in R_{cons} , there're 2601 fell in $\mathbf{C}_0(\mathbf{Y})$, 1864 in $\mathbf{C}_E(\mathbf{Y})$ and 810 in $\mathbf{C}_P(\mathbf{Y})$. This indicates that $\mathbf{C}_E(\mathbf{Y})$ is about 28.34% smaller than $\mathbf{C}_0(\mathbf{Y})$ and the set $\mathbf{C}_P(\mathbf{Y})$ is 56.55% smaller than $\mathbf{C}_E(\mathbf{Y})$. $\mathbf{C}_P(\mathbf{Y})$ is smaller than $\mathbf{C}_E(\mathbf{Y})$ because it uses an unduly small critical

constant which does not guarantee the $(1 - \alpha)$ confidence level. It is also worth noting that $f(\mathbf{x}, \boldsymbol{\theta})$ is not of the form (2) exactly but our method works just as well.

5 CONCLUSION AND DISCUSSION

In this paper, the construction of confidence sets for the optimal factor levels of response surfaces is discussed. The only confidence set available in literature that guarantees the $(1 - \alpha)$ confidence level is Rao's confidence set $\mathbf{C}_c(\mathbf{Y})$, but it is usually conservative. In this paper one exact confidence set $\mathbf{C}_E(\mathbf{Y})$ and one conservative confidence set $\mathbf{C}_0(\mathbf{Y})$ are constructed. It is shown in the examples that the confidence set $\mathbf{C}_E(\mathbf{Y})$ is always, and can be substantially, smaller and so better than $\mathbf{C}_0(\mathbf{Y})$ and Rao's confidence set $\mathbf{C}_c(\mathbf{Y})$. This is not surprising because $\mathbf{C}_E(\mathbf{Y})$ is proved to have an exact $(1 - \alpha)$ confidence level theoretically, while $\mathbf{C}_0(\mathbf{Y})$ and $\mathbf{C}_c(\mathbf{Y})$ are conservative. In both the examples, PCD's confidence set $\mathbf{C}_P(\mathbf{Y})$ is smaller than $\mathbf{C}_E(\mathbf{Y})$. This is due to that the critical constants used in $\mathbf{C}_P(\mathbf{Y})$ are too small and hence the resultant confidence set $\mathbf{C}_P(\mathbf{Y})$ do not guarantee the nominal $(1 - \alpha)$ confidence level.

Bootstrap method is versatile and it is not too difficult to devise a bootstrap method for constructing a confidence set. The problem with a bootstrap confidence set is that its true confidence level is difficult to pin down for a finite sample size; the best one can claim is that the confidence level is approximately $(1 - \alpha)$. While one can assume a specific model in terms of the β and σ and apply the bootstrap (sampling) method from this specific model to assess the coverage probability of the confidence set, the conclusion is only for this particular model. This is characteristically different from the $\mathbf{C}_E(\mathbf{Y})$ proposed in this paper: so long as the model in (1) is true and the critical constants $c(\mathbf{k}^o)$ are computed accurately then $\mathbf{C}_E(\mathbf{Y})$ has $(1 - \alpha)$ confidence level.

As shown in Section 3, the construction method is readily applied to other regression models where the parameter estimators have an approximate multivariate normal distribution.

Examples include generalized linear models, semi-parametric proportional hazard models and quantile regression models. Therefore the method proposed in this paper is widely applicable. One possible future work is to construct confidence set for the ridge of a response surface, a problem that has considered by many researchers (cf. Gilmour *et al.*, 2003 and Peterson, 1993). A ridge is the path of the minimum or maximum point in the response when varying the radius of the locus of factors. Hence each point on the ridge represents the optimal factor levels of the response on a certain ellipse of the factors, that is, a constrained experimental region. Our work in this paper can be extended to construct confidence sets for the ridge of a response surface. It would also be useful to explore other maximization/minimization methods to reduce the computation time in computing the critical constant for each grid point \mathbf{k}^o .

6 Supplementary Materials

Matlab code implementing our method is available with this paper at the Biometrics website on Wiley Online Library.

REFERENCES

- Becker, N.G. (1968). Models for the response of a mixture. *Journal of the Royal Statistical Society, Series B* **30**, 349-358.
- Becker, N.G. (1978). Models and designs for experiments with mixtures. *The Australian journal of statistics* **20**, 195-208.
- Bliss, C. I. (1970). *Statistics in Biology*. McGraw-Hill Book Company.
- Box, G. E. P. and Hunter, J. S. (1954). A confidence region for the solution of a set of simultaneous equations with an application to experimental design. *Biometrika* **41**, 190-199.
- Cahya, S., Del Castillo, E., and Peterson, J. J. (2004). Computation of Confidence Regions

- for Optimal Factor Levels in Constrained Response Surface Problems. *Journal of Computational and Graphical Statistics* **13**, 499-518.
- Carter, W.H. and Wampler, G. L. and Stablein, D.M. and Campbell, E.D. (1982). Drug activity and therapeutic synergism in cancer treatment. *Cancer research* **13**, 2963-2971.
- Carter, W. H., Chinchilli, V. M., Campbell, E. D., and Wampler, G.L. (1984). Confidence Interval about the Response at the Stationary Point of a Response Surface, with an Application to Preclinical Cancer Therapy. *Biometrics* **40**, 1125-1130.
- Casella, G. and Berger, R.L. (2002). *Statistical Inference*. Duxbury.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B* **34**, 187-220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.
- Del Castillo, E. and Cahya, S. (2001). A Tool for Computing Confidence Regions on the Stationary Point of a Response Surface. *The American Statistician* **55**, 4, 358-365.
- Dobson, A. J. (2001). *An Introduction to Generalized Linear Models, Second Edition*. CRC Press.
- Edwards, D. and Berry, J. J. (1987). The Efficiency of Simulation-based Multiple Comparisons. *Biometrics* **43**, 913-928.
- Faraway, J. J. (2006). *Extending the Linear Model with R*. CRC Press.
- Farebrother, E. (1998). *Statistical design and analysis of factorial combination Drug trials*. PhD thesis, University of Reading.
- Finner, H. and Strassburge, K. (2002). The partitioning principle: a powerful tool in multiple decision theory. *Annals of Statistics* **30**, 1194-1213.
- Frisbee, S.E. and McGinity, J.W. (1994). Influence of nonionic surfactants on the physical and chemical properties of biodegradable pseudolatex. *Journal of pharmaceuticals and biopharmaceuticals* **40**, 355-363.

- Genz, A. and Bretz, F. (2009). *Computation of multivariate and t probabilities*. Springer.
- Gilmour, S. G. and Draper, N.R. (2003) Confidence regions around the ridge of optimal response on fitted second-order response surfaces. *Technometrics* **45**, 333-339.
- Hayter, A. J. and Hsu, J. C. (1994). On the relationship between stepwise decision procedures and confidence sets. *Journal of the American Statistical Association* **89**, 425.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses, 2nd ed.* John Wiley.
- Liu, W., Jamshidian, M., Zhang, Y., and Donnelly, J. (2005). Simulation-based simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals. *Journal of Computational and Graphical Statistics* **14**, 459-484.
- Liu, W. (2010). *Simultaneous Inference in Regression*. CRC Press.
- McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. John Wiley.
- Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transaction of the Royal Society of London, Series A, Mathematical and Physical Sciences* **236**, 333-380.
- Myers, R.H. and Montgomery, D.C. (1995). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley.
- Peterson, J. J. (1993). A general approach to ridge analysis with confidence intervals. *Technometrics* **35**, 204-214.
- Peterson, J. J., Cahya, S., and Del Castillo, E. (2002). A general approach to confidence regions for optimal factor levels of response surface. *Biometrics* **58**, 422-431.
- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus*. Springer.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Application, 2nd ed.* Oxford University Press.

- Stablein, D. M., Carter, W. H. Jr., and Wampler, G. L. (1983). Confidence regions for constrained optima in Response-Surface experiments. *Biometrics* **39**, 759-763.
- Stefansson, G., Kim, W. C., and Hsu, J. C. (1988). On confidence sets in multiple comparisons, in *Statistical Decision Theory and Related TopicsIV*, eds. S.S. Gupta and J.O. Berger, New York: Academic Press, 89-104.
- Studier, E.H., Dapson, R.W., and Bigelow, R.E. (1975). Analysis of polynomial functions for determining maximum or minimum conditions in biological systems. *Comparative Biochemistry and Physiology, Part A: Physiology* **52**,19-20.
- Wan, F., Liu, W., Bretz, F., and Han, Y. (2015). An Exact Confidence Set for a Maximum Point of a Univariate Polynomial Function in a Given Interval. *Technometrics* **57**, 559-565.
- Zar, J.H. (1999). *Biostatistical Analysis (4th edition)*. Prentice Hall.

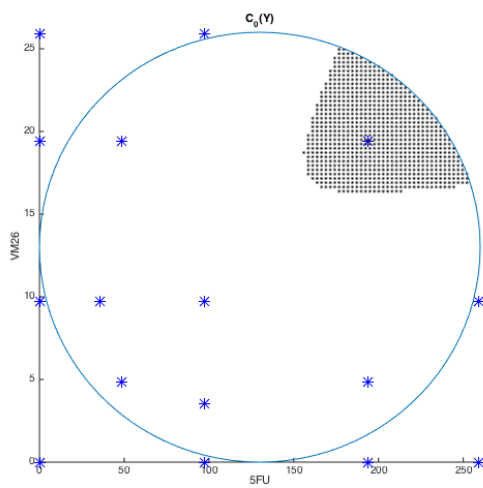
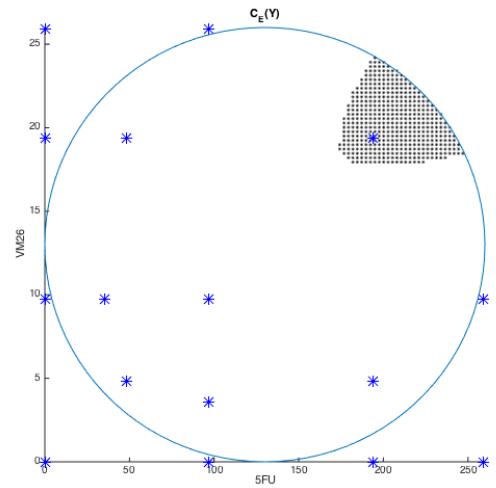
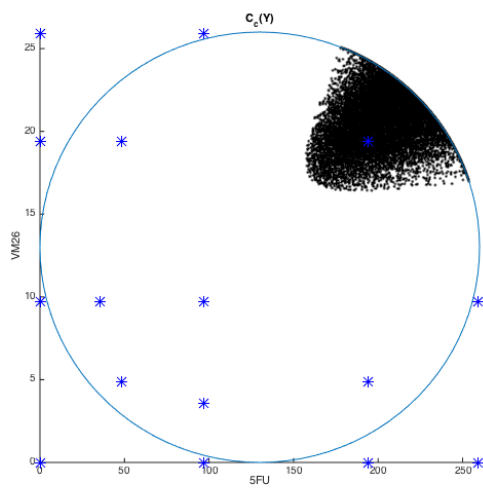
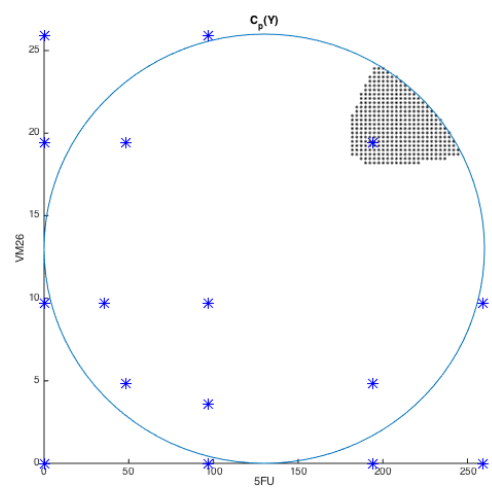
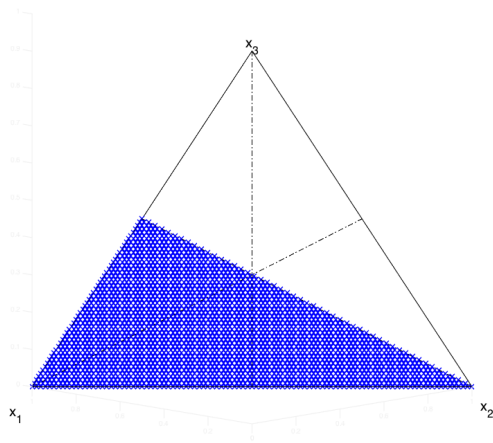
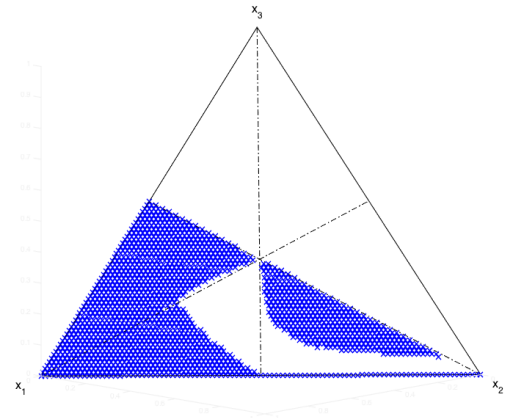
(a) Confidence set $\mathbf{C}_0(\mathbf{Y})$ (b) Confidence set $\mathbf{C}_E(\mathbf{Y})$ (c) Confidence set $\mathbf{C}_c(\mathbf{Y})$ (d) Confidence set $\mathbf{C}_P(\mathbf{Y})$

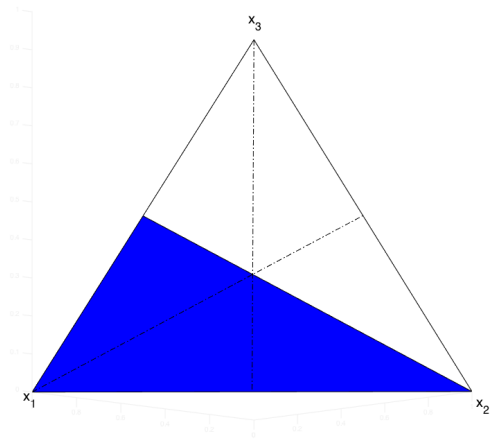
Figure 1. The 95% confidence sets in Example 1. The asterisks represent the observation and the dot points represent the confidence sets.



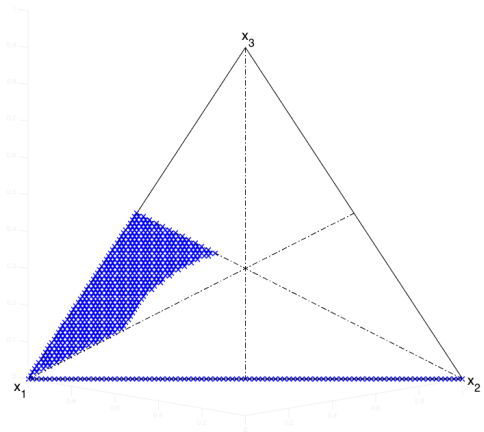
(a) Confidence set $C_0(\mathbf{Y})$



(b) Confidence set $C_E(\mathbf{Y})$



(c) Confidence set $C_c(\mathbf{Y})$



(d) Confidence set $C_P(\mathbf{Y})$

Figure 2. The 95% confidence sets in Example 2.

Table 1
5FU+VM26 combination experiment (Stablein et al., 1983)

Treatment levels		Days of survival
5FU(mg/kg)	VM26(mg/kg)	
0.0	0.00	8,9(2),10(5)
0.0	9.71	10,13(5),14(2)
0.0	19.40	8,10,13,14(4),15
0.0	25.90	9,14(4),15(3)
35.6	9.71	13,14(3),15(3),17
48.5	4.85	9,13(2),14(3),15(2)
48.5	19.40	14(2),15(2),16(4)
97.1	0.00	8(2),10,11,12(2),14,16
97.1	3.56	8,9(2),11(2),13(2),16
97.1	9.71	8,10,11,16(2),17(2),18
97.1	25.9	16(3),17,18(3),19
194.0	0.00	10, 13(6),14
194.0	4.85	11(2),14(3),16,17
194.0	19.40	8,14,16,20(4),21
259.0	0.00	9,11,12(3),13(3)
259.0	9.71	16(2),17,18(2),19(2),20

The number in the parentheses indicates the number of animals dead on that day.

Table 2
The formulation components (%) and the response (Frisbee et al., 1994)

x_1	1	0	0	0.5	0.5	0	0.333	0.666	0.167	0.167	0.333
x_2	0	1	0	0.5	0	0.5	0.333	0.167	0.666	0.167	0.333
x_3	0	0	1	0	0.5	0.5	0.333	0.167	0.167	0.666	0.333
Y	18.9	15.2	35.0	16.1	18.9	31.2	19.3	18.2	17.7	30.1	19.0