Journal of Statistical Planning and Inference 147 (2014) 188–203

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Influence analysis in response surface methodology

Yufen Huang*, Chao-Yen Hsieh

Department of Mathematics, National Chung Cheng University, 160, San-Hsing, Min-Hsiung, Chia-Yi 621, Taiwan

ARTICLE INFO

Article history:
Received 30 August 2012
Received in revised form
17 July 2013
Accepted 15 November 2013
Available online 7 December 2013

Keywords:
Influence functions
Single-perturbation
Pair-perturbation

ABSTRACT

The study of response surface methodology (RSM) involves both experimental planning and data modeling and analysis. Once a design is selected, and data obtained from it, models for representing the data need to be considered and fitted. During the fitting process, observations that are suspicious (e.g. outliers and/or influential points) may cause problems. Such observations need to be detected so that appropriate adjustments can be made to analysis. Thus far, the work on influence analysis of RSM is unexplored in statistical research. This will be the focus of this paper. We not only generalize the single perturbation scheme in Hampel's (1974) method, but also implement the pair-perturbation scheme in Huang et al. (2007a–c) to develop influence functions for sensitivity analysis in RSM. A simulation study and two real data examples for illustrating the effectiveness of the proposed method are provided.

1. Introduction

Response surface methodology as pioneered by Box and Wilson (1951) is widely used in science and engineering to investigate the relationship between one or more responses and the factors of a process or system. It provides a sequential framework that involves the design of an experiment, data collection, empirical model building and response surface optimization. The main step in RSM is to find a suitable approximation to the true relationship. The most common forms are lower-order polynomials (most often first- or second-order). We focus on investigating the behavior of a response variable y over a specified region of interest by fitting a second-order response surface. During the fitting process, observations that are suspicious (e.g. outliers and/or influential points) may cause problems. For example, if there are influential points or outliers, they may influence the accuracy of the fitting response surface. Such observations need to be detected so that appropriate adjustments can be made to the analysis. Hence, our aim in this paper is to study influence analysis on the response surface.

Perturbation theory provides a useful tool to investigate the influence of abnormal observations. One major tool based on differentiation in influence analysis in statistical modeling is Hampel's (1974) influence function which used perturbation techniques. These have been applied in various contexts. For instance, influence functions in principal component analysis have been considered by Critchley (1985) and Tanaka (1988), and generalizations to pair-perturbations by Huang et al. (2007a–c). In linear discriminant analysis, influence analysis has been discussed by Fung (1992, 1993, 1995, 1996), He and Fung (2000), Poon (2004), Huang et al. (2007a–c) among others; in projection pursuit, influence analysis has been discussed by Huang et al. (2007a–c, 2008, 2011).

* Corresponding author. Tel.: +886 5 2720411x66125; fax: +886 5 2720497.
E-mail address: yfhuang@math.ccu.edu.tw (Y. Huang).

0378-3758/$ - see front matter
http://dx.doi.org/10.1016/j.jspi.2013.11.008

To our knowledge, a study in the influence analysis on the RSM has not been explored in the literature. Hence, in this paper we focus on developing influence functions based on Hampel's method for sensitivity analysis in RSM. Three types of influence functions are considered: empirical influence, deleted empirical influence and sample influence. However, single-perturbation diagnostics can suffer from the masking effect (see Riani and Atkinson, 2001). Therefore, we also focus on tackling the masking effect in RSM by implementing pair-perturbations influence functions.

The remainder of this paper is organized as follows. In Section 2, response surface methodology is reviewed. In Section 3, Hampel's influence function is revisited and three types of influence functions on RSM are derived. In Section 4, a simulation study with two scenarios and two real data examples are provided for illustration. Conclusions and discussion are given in Section 5.

2. Response surface methodology

2.1. Canonical analysis

Suppose that an adequate second-degree model has been fitted. Canonical analysis is a method of rewriting a fitted second degree equation in a form in which it can be more readily understood. This is achieved by a rotation of axes which removes all cross-product terms and gives the so-called *A canonical form*.

2.2. The A canonical form

Consider the second-degree polynomial fitted model

$$\hat{y} = b_0 + \sum_{j=1}^{k} b_j x_j + \sum\sum_{i \geq j} b_{ij} x_i x_j = b_0 + \mathbf{x}^T \hat{\mathbf{b}} + \mathbf{x}^T \hat{\mathbf{B}} \mathbf{x}, \tag{1}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \quad \hat{\mathbf{b}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}, \quad \hat{\mathbf{B}} = \begin{bmatrix} b_{11} & \frac{1}{2}b_{12} & \cdots & \frac{1}{2}b_{1k} \\ \frac{1}{2}b_{12} & b_{22} & \cdots & \frac{1}{2}b_{2k} \\ \vdots & & & \\ \frac{1}{2}b_{1k} & \frac{1}{2}b_{2k} & \cdots & b_{kk} \end{bmatrix}.$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the eigenvalues of the $\hat{\mathbf{B}}$, and let unit vectors $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_k$ be the corresponding eigenvectors. By definition

$$\hat{\mathbf{B}}\mathbf{m}_i = \mathbf{m}_i \lambda_i, \quad i = 1, 2, \ldots, k.$$

Let $\mathbf{M}$ be the $k \times k$ orthonormal matrix with $\mathbf{m}_i$ for its ith column. Then

$$\hat{\mathbf{B}}\mathbf{M} = \mathbf{M}\boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix whose ith diagonal element is λ_i. Premultiplying $\mathbf{M}\boldsymbol{\Lambda}$ by $\mathbf{M}^T (= \mathbf{M}^{-1})$, we obtain

$$\mathbf{M}^{\mathbf{T}} \hat{\mathbf{B}} \mathbf{M} = \boldsymbol{\Lambda}.$$

Since $\mathbf{M}\mathbf{M}^{\mathbf{T}} = \mathbf{I}$, we have

$$\hat{y} = b_0 + (\mathbf{x}^T \mathbf{M})(\mathbf{M}^T \hat{\mathbf{b}}) + (\mathbf{x}^T \mathbf{M}) \mathbf{M}^T \hat{\mathbf{B}} \mathbf{M} (\mathbf{M}^T \mathbf{x}). \tag{2}$$

If we let $\mathbf{x}^* = \mathbf{M}^T \mathbf{x}$ and $\boldsymbol{\theta} = \mathbf{M}^T \hat{\mathbf{b}}$, then (2) can be rewritten as

$$\hat{y} = b_0 + \mathbf{x}^{*T} \boldsymbol{\theta} + \mathbf{x}^{*T} \boldsymbol{\Lambda} \mathbf{x}^*.$$

This constitutes the *A* canonical form which is a useful tool in RSM. There are three important quantities in the canonical form. First, the sizes and the signs of the λ_i of the $\hat{\mathbf{B}}$ determine the type of second-order fitted surface. Second, the θ_i measure the slopes of the surface at the origin $\mathbf{x} = \mathbf{0}$, in the directions of the coordinate axes $x_1^*, \ldots, x_k^*$. Third, the location of the stationary point $\mathbf{x}_S$ which is a maximum (or minimum) point of the fitted surface is determined by the solution of the equations

$$-2\hat{\mathbf{B}}\mathbf{x}_S = \hat{\mathbf{b}}.$$

However, the matrix $\hat{\mathbf{B}}$ and the stationary point can be affected by outliers or influential observations. Consequently, these will influence the accuracy of the fitted response surface. Hence, we focus on developing influence functions for the matrix $\hat{\mathbf{B}}$ and influence functions for the stationary point in Section 3.