University of Southampton

Faculty of Social and Human Sciences

School of Mathematics

# Confidence Sets for a Maximum Point of a Regression Function

**Fang Wan**

A thesis submitted for the degree of

Doctor of Philosophy

May 2014

UNIVERSITY OF SOUTHAMPTON

## ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES
SCHOOL OF MATHEMATICS

Doctor of Philosophy

Confidence Sets for a Maximum Point of a Regression Function

by Fang Wan

A maximum point of a regression function is defined as one at which the function attains its maximum value. The determination of a maximum point of the regression function over a given covariate region is often of great importance due to its wide applications. Since the regression function needs to be estimated and its maximum point(s) can only be estimated based on random observations, the focus of this research is therefore to construct confidence sets for a maximum point of the regression function.

A confidence set for a maximum point of a regression function provides useful information on where a true maximum point lies, and so has applications in many real problems. In this thesis, an exact $(1-\alpha)$ level confidence set is provided for a maximum point of a linear regression function. It is also shown how the construction method can readily be applied to many other regression models involving a linear function. Examples are given to illustrate this confidence set and to demonstrate that it can be substantially smaller than the only other conservative confidence set that is available in the statistical literature so far.

# Contents


**Abstract** **iii**

**Author's declaration** **xv**

**1 Introduction** **1**

1.1 Confidence Sets for a Maximum Point . . . . . . . . . . . . . . . . . . 2

1.2 Some Confidence Sets in Literature . . . . . . . . . . . . . . . . . . . . 4

1.2.1 Box and Hunter's Confidence Sets . . . . . . . . . . . . . . . . 4

1.2.2 Bootstrap Confidence Sets . . . . . . . . . . . . . . . . . . . . . 5

1.2.3 Rao's Confidence Sets . . . . . . . . . . . . . . . . . . . . . . . 5

1.2.4 Confidence Sets Constructed by Using Neyman's Theorem . . 5

1.3 Outline of the Thesis . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

**2 Methods** **9**

2.1 Construction of a Confidence Set by Using Neyman's Theorem . . . 9

2.1.1 Neyman's Theorem . . . . . . . . . . . . . . . . . . . . . . . . . 10

2.1.2 Some Simple Examples . . . . . . . . . . . . . . . . . . . . . . 10

2.1.3 Confidence Set for the Solution of a Set of Simultaneous Equations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

2.2 Other Methods of Constructing a Confidence Set . . . . . . . . . . . 20

2.2.1 The Conservative Confidence Set Given by Rao . . . . . . . . 20

2.2.2 The Bootstrap Percentile Confidence Set . . . . . . . . . . . . 21

**3 Confidence Set for a Maximum Point of a Univariate Polynomial**

**Regression Function in a Given Interval** **23**
3.1 Our Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
3.1.1 A Simple Linear Regression Model . . . . . . . . . . . . . . . . 24
3.1.2 A Quadratic Polynomial Regression Model . . . . . . . . . . . 31
3.1.2.1 Theory . . . . . . . . . . . . . . . . . . . . . . . . . 32
3.1.2.2 Computation . . . . . . . . . . . . . . . . . . . . . . 43
3.1.2.3 Examples . . . . . . . . . . . . . . . . . . . . . . . 54
3.1.3 A General Polynomial Regression Model . . . . . . . . . . . 71
3.1.3.1 Theory . . . . . . . . . . . . . . . . . . . . . . . . . 72
3.1.3.2 Conservative Critical Constants . . . . . . . . . . . 78
3.1.3.3 Computation . . . . . . . . . . . . . . . . . . . . . . 82
3.1.3.4 Examples . . . . . . . . . . . . . . . . . . . . . . . 88
3.2 Rao's Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . 99
3.3 Bootstrap Method . . . . . . . . . . . . . . . . . . . . . . . . . 101
3.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 104

**4 Confidence Set for a Maximum Point of a Bivariate Quadratic Regression Function in a Given Rectangular Region** **105**
4.1 Our Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 106
4.1.1 Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . 106
4.1.2 Computation . . . . . . . . . . . . . . . . . . . . . . . . . 110
4.1.3 Examples . . . . . . . . . . . . . . . . . . . . . . . . . . . 113
4.2 Rao's Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . 122
4.3 Bootstrap Method . . . . . . . . . . . . . . . . . . . . . . . . . 124
4.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 126

**5 Extensions to other models** **127**
5.1 Generalized Linear Models . . . . . . . . . . . . . . . . . . . . . 130
5.1.1 Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . 130
5.1.2 Example . . . . . . . . . . . . . . . . . . . . . . . . . . . . 133
5.2 Cox's Proportional Hazard Models . . . . . . . . . . . . . . . . . 136

5.2.1 Method . . . . . . . . . . 136

5.2.2 Example . . . . . . . . . . 138

5.3 A Becker's H1 Model . . . . . . . . . . 142

5.3.1 Method . . . . . . . . . . 142

5.3.2 Example . . . . . . . . . . 145

5.4 Summary . . . . . . . . . . 148

**6 Conclusion and Future Works** **149**

6.1 Summary . . . . . . . . . . 149

6.2 Future Works . . . . . . . . . . 150

**Appendix** **152**

# List of Tables


3.1 Simulation data used in Example 1. . . . . . . . . . . . . . . . . . . . 55
3.2 Simulation data used in Example 2. . . . . . . . . . . . . . . . . . . . 57
3.3 Simulation data used in Example 3. . . . . . . . . . . . . . . . . . . . 58
3.4 Simulation data used in Example 4. . . . . . . . . . . . . . . . . . . . 59
3.5 Simulation data used in Example 5. . . . . . . . . . . . . . . . . . . . 61
3.6 Simulation data used in Example 6. . . . . . . . . . . . . . . . . . . . 61
3.7 Data of crop yields and fertilizer used (Sullivan and Sullivan, 2002). 63
3.8 Summary of clinical dose response data (Rom et al., 1994) . . . . . . 67
3.9 Clinical dose response data (Rom et al., 1994) . . . . . . . . . . . . . 69
3.10 Summary of clinical dose response data (Rom et al., 1994) . . . . . . 70
3.11 Perinatal mortality data for black infants (Selvin, 1998) . . . . . . . 88
3.12 Perinatal mortality data for white infants (Selvin, 1998) . . . . . . . 90
3.13 Perinatal mortality data for non-black and non-white infants (Selvin, 1998) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 95

4.1 Data of a central composite design for the chemical process (Myers, 2009) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 119

5.1 Survival data for the CTX experiment (Carter *et al.*, 1983) . . . . . 133
5.2 5FU+VM26 combination experiment (Stablein et al., 1983) . . . . . 141
5.3 The expressions of $\mathbf{z}(\boldsymbol{x})$ and $\mathbf{z}_i(\boldsymbol{x})$ in each of the 6 parts . . . . . . . 144
5.4 The formulation components (%) and the response . . . . . . . . . . 145

# List of Figures


2.1 The relationship between the acceptance set $A(\beta_0)$ and the confidence set $C(Y)$. . . . . 12

2.2 The relationship between the 95% level confidence set $C(Y)$ and the corresponding acceptance set $A(\theta_0) = \{Y : |Y| > c_{\theta_0}\}$. . . . . 13

2.3 The relationship between the 95% level confidence set $C(Y)$ and the corresponding acceptance set $A(\theta_0) = \{Y : |Y| < d_{\theta_0}\}$. . . . . 15

2.4 The relationship between the value of $m_0$ and the 95% level confidence set $C(\boldsymbol{Y})$ for $x_0$. . . . . 16

3.1 Confidence set for a maximum point, $k$, of $f(x) = \theta_0 + \theta_1 x$. . . . . 30

3.2 The region $R_{h,a}$ . . . . . 34

3.3 The region $R_{h,a} = D_1 \cup D_2 \cup D_3 \cup D_4$ . . . . . 35

3.4 The region $R_h(s)$ . . . . . 39

3.5 The region $C_2$. . . . . 41

3.6 The critical values $c(s)$, $c_1$ and $c_2$. . . . . 46

3.7 The 95% level critical values for different $x \in [-5, 5]$ in Example 1. . . 55

3.8 The 95% level confidence sets in Example 1. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. . . . . 56

3.9 The 95% level confidence sets in Example 2. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

3.10 The 95% level confidence sets in Example 3. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. . . . . . . . . . . . . . . . . . . . . . . . . . . . 59

3.11 The 95% level confidence sets in Example 4. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. . . . . . . . . . . . . . . . . . . . . . . . . . . . 60

3.12 The 95% level confidence sets in Example 5. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. All the confidence sets for a maximum point contain just one element 10, yet the BH confidence set is for the stationary point and is an interval. . . . . . . . . . . . . . . . . . . . . . . . . 60

3.13 The 95% level confidence sets in Example 6. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. . . . . . . . . . . . . . . . . . . . . . . . . . . . 62

3.14 Residual plots . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

3.15 Normal probability plot of the residuals . . . . . . . . . . . . . . . . 65

3.16 The 95% level confidence sets for the amount of fertilizer used that maximize the crop yields in Example 7. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

3.17 The 95% level confidence sets in Example 8. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 68
3.18 The 95% level confidence sets in Example 9. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70
3.19 Normal probability plot of black infant data . . . . . . . . . . . . . . 89
3.20 Residual plot of black infant data . . . . . . . . . . . . . . . . . . . . 91
3.21 The 95% level confidence sets for a maximum point based on black infant data. The horizontal line segments, from top to bottom, represent $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, $\mathbf{C}_c(\boldsymbol{Y})$ and bootstrap confidence set. . . . . . 92
3.22 Normal probability plot of white infant data . . . . . . . . . . . . . . 93
3.23 Residual plot of white infant data . . . . . . . . . . . . . . . . . . . . 93
3.24 The 95% level confidence sets for a maximum point based on white infant data. The horizontal line segments, from top to bottom, represent $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, $\mathbf{C}_c(\boldsymbol{Y})$ and bootstrap confidence set. . . . . . 94
3.25 Normal probability plot of non-black and non-white infant data . . . 97
3.26 Residual plot of non-black and non-white infant data . . . . . . . . . 97
3.27 The 95% level confidence sets for a maximum point based on non-black and non-white infant data. The horizontal line segments, from top to bottom, represent $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, $\mathbf{C}_c(\boldsymbol{Y})$ and bootstrap confidence set. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 98
3.28 Bootstrap method. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 103

4.1 The response surface and 95% confidence sets in Example 1 . . . . . 114
4.2 The response surface and 95% confidence sets in Example 2 . . . . . 115
4.3 The response surface and 95% confidence sets in Example 3 . . . . . 117
4.4 The response surface and 95% confidence sets in Example 4 . . . . . 118
4.5 The residual plot and normal probability plot in Example 5 . . . . . 119

4.6 The response surface and 95% confidence sets in Example 5 . . . . . 121

5.1 The three 95% confidence sets in Example 5.1.2; $\mathbf{C}_E(\mathbf{Y})$ is given by the top broken line, $\mathbf{C}_0(\mathbf{Y})$ is given by the middle dot line, and $\mathbf{C}_c(\mathbf{Y})$ is given by the bottom solid line. . . . . . . . . . . . . . . . . . . . . . 134

5.2 The 95% confidence sets in Example 5.2.2. . . . . . . . . . . . . . . . . 139

5.3 The partition of the covariate region $R_{cons}$. . . . . . . . . . . . . . . . 143

5.4 The 95% confidence sets in Example 5.3.2. . . . . . . . . . . . . . . . . 146

## Declaration of Authorship

I, Fang Wan, declare that the thesis entitled *Confidence Sets for a Maximum Point of a Regression Function* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of the work has been published before submission.

Signed:.................................................

Date:....................................................

# Acknowledgements

First, I would like to thank my supervisor Prof. Wei Liu for his continuous support throughout my postgraduate studies. I am extremely grateful and indebted to him for his expert, sincere and valuable guidance and encouragement. I would not have been able to complete this thesis without his supervision.

I also thank my advisor Dr. Alan Kimber for his support during my studies.

I would like to take this opportunity to thank all the staff and PhD students in S3RI for their help and encouragement.

I also place on record, my sense of gratitude to one and all who, directly or indirectly, have lent their help during this period.

Finally, I owe my warmest thanks to my parents for their love, patience and unceasing support.

# Chapter 1

# Introduction

Regression analysis is a statistical technique that models and investigates the relationship between a dependent/response variable $Y$ and one or several independent/explanatory variables $\boldsymbol{x}$. A regression model is often expressed as

$$Y = f(\boldsymbol{x}, \boldsymbol{\theta}) + e,$$

where $e$ is the random error, $\boldsymbol{\theta}$ is the unknown parameter(s) and $f(\boldsymbol{x}, \boldsymbol{\theta})$ is the regression function of a pre-specified form. A maximum point of the regression function $f(\boldsymbol{x}, \boldsymbol{\theta})$ is defined as a point in the covariate region at which $f(\boldsymbol{x}, \boldsymbol{\theta})$ attains its maximum value. A maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ over a given covariate region may represent the dose level in the experimental region that maximizes the response, or the amount of fertilizer that maximizes the crop yield, etc. Due to its wide applications, the determination of a maximum point of a regression function in a constrained covariate region is often of great importance.

If the value of $\boldsymbol{\theta}$ is known, then finding a maximum point is a simple calculus problem which can be solved according to Hancock (1960), Bliss (1970, pp44-50), Studier et al. (1975), or Zar (1999, pp458-459). The difficulty lies in that the value of $\boldsymbol{\theta}$ is unknown and only an estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ with a certain distribution is available. Hence the regression function $f(\boldsymbol{x}, \boldsymbol{\theta})$ needs to be estimated and its maximum points can only be estimated based on $\hat{\boldsymbol{\theta}}$. The focus of this research is therefore to construct a confidence set for a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$. However, the computation is much more challenging than point estimation, which naturally uses the maximum points of the estimated function $f(\boldsymbol{x}, \hat{\boldsymbol{\theta}})$. In fact, the confidence sets constructed in this thesis are valid for a given maximum point $f(\boldsymbol{x}, \boldsymbol{\theta})$, regardless of the number of maximum points $f(\boldsymbol{x}, \boldsymbol{\theta})$ has.

## 1.1 Confidence Sets for a Maximum Point

Suppose we have observations $\boldsymbol{Y}$ from the regression model defined earlier in this chapter. Let $\boldsymbol{k} = \boldsymbol{k}(\boldsymbol{\theta})$ be a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$. Then a $(1-\alpha)$ level confidence set for $\boldsymbol{k}$ is given by $C(\boldsymbol{Y})$ where the set $C(\boldsymbol{Y})$ satisfies

$$\inf_{\boldsymbol{\theta}} P_{\boldsymbol{\theta}}(\boldsymbol{k} \in C(\boldsymbol{Y})) \geq 1 - \alpha.$$

If the equality holds, then $C(\boldsymbol{Y})$ is an exact $(1-\alpha)$ level confidence set, otherwise we call it a conservative $(1-\alpha)$ level confidence set.

A confidence set for a maximum point provides useful information about the regression function and quantifies where a true maximum point lies. Carter et al. (1983) consider the relevant applications in cancer chemotherapy, while Farebrother (1998) deals with response surface models. In fact, one of the objectives of response surface methodology is to find the path of the ridge, that is, a locus of points, each of which is a maximum point of the regression function in a given sphere of certain radius and construct a confidence set for the path; see, for example, Ding et al. (2005) and Gilmour and Draper (2003). This methodology is known as ridge analysis (see Hoerl, 1985).

To be specific, in the thesis we consider the simpler situation where the regression function is a linear function of the form

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 + \mathbf{z}(\boldsymbol{x})^T \boldsymbol{\theta}^0 \tag{1.1}$$

where $\boldsymbol{x} = (x_1, \cdots, x_q)^T$, $\mathbf{z}(\boldsymbol{x})$ is a given $p \times 1$ vector-valued function of $\boldsymbol{x}$, $\boldsymbol{\theta}^0 = (\theta_1, \cdots, \theta_p)^T$ and $\boldsymbol{\theta} = (\theta_0, \cdots, \theta_p)^T$. The interest is in the maximum points of

$f(\boldsymbol{x}, \boldsymbol{\theta})$ in a given covariate region of $\boldsymbol{x}$. Note that the maximum points of $f(\boldsymbol{x}, \boldsymbol{\theta})$ do not depend on the intercept $\theta_0$ and $\mathbf{z}(\boldsymbol{x})^T\boldsymbol{\theta}^0$ contains all the information about the maximum points of $f(\boldsymbol{x}, \boldsymbol{\theta})$.

Assume that an estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is available with normal distribution

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2\Sigma) \tag{1.2}$$

where $\Sigma$ is a known positive definite matrix, and $\hat{\sigma}^2$ is an estimator of the error variance $\sigma^2$ with distribution $\hat{\sigma}^2 \sim \sigma^2\chi^2_\nu/\nu$ independent of $\hat{\boldsymbol{\theta}}$. The constant $\nu$ is the degrees of freedom (df) of the chi-squared distribution. In the special case that $\sigma^2$ is a known constant, then $\hat{\sigma}^2 = \sigma^2$ and $\nu = \infty$; hence, without loss of generality, we assume

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \Sigma)\,. \tag{1.3}$$

Our focus is on the construction of a $(1-\alpha)$ level confidence set for a maximum point of the function $f(\boldsymbol{x}, \boldsymbol{\theta})$ in a given covariate region of $\boldsymbol{x}$ based on the distributional assumption (1.2), which includes the special case (1.3).

The distributional assumption (1.3) holds asymptotically for many parametric and semi-parametric models. In the generalized linear, random effects linear and random effects generalized linear models (cf. Pinheiro and Bates, 2000; McCulloch and Searle, 2001; Faraway, 2006), the mean response $E(Y)$ may be related to $f(\boldsymbol{x}, \boldsymbol{\theta})$ by a given monotone link function. Based on the observed data, one has approximately $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \hat{\Sigma})$ where $\hat{\Sigma}$ is provided by many statistical software packages that deal with these models. Hence the method developed in this thesis (for the special case (1.3)) can readily be applied to construct an asymptotic $(1-\alpha)$ level confidence set for a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ over a given covariate region for these parametric and semi-parametric models.

It is noteworthy that the construction of a confidence set for a minimum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ in a given region of $\boldsymbol{x}$ can be transformed into the construction of a confidence set for a maximum point of $-f(\boldsymbol{x}, \boldsymbol{\theta}) = f(\boldsymbol{x}, -\boldsymbol{\theta})$ with the distributional assumption $-\hat{\boldsymbol{\theta}} \sim N(-\boldsymbol{\theta}, \sigma^2\Sigma)$. Hence the methods developed in this thesis can readily be applied to the construction of a confidence set for a minimum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ in a given region.

## 1.2 Some Confidence Sets in Literature

### 1.2.1 Box and Hunter's Confidence Sets

The method given in Box and Hunter (1954) can be used to construct a confidence set for a stationary point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ over the whole covariate region of $\boldsymbol{x} \in R^p$. A program for computing this confidence set is given by Del Castillo and Cahya (2001); see also Del Castillo (2007, Chapter 7) for details. Box and Hunter's confidence set (referred to BH confidence set henceforth) is, however, often mis-used as a confidence set for a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ over a given region; see, e.g., Bliss (1970, pp.44), Carter et al. (1983, pp.19) and Weisberg (2005, pp. 120). Based on Box and Hunter's method, Carter et al. (1982) concluded that if the confidence region does not contact either axis, then a therapeutic synergism will be claimed (the combination of some components is preferable to that of any component used alone). Stablein et al. (1983) modified Box and Hunter's method by using Lagrange multiplier and proposed a confidence region for the maximum point within the experimental region according to the signs of the eigenvalues.

However, according to the definition, a stationary point is a point of the domain of a differentiable function, where the derivative is zero. It is clear that a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ in a given covariate region is not a stationary point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ if $f(\boldsymbol{x}, \boldsymbol{\theta})$ does not have a stationary point in that region. Even if $f(\boldsymbol{x}, \boldsymbol{\theta})$ has a stationary point in the given region, this stationary point may be a turning point or a minimum point rather than a maximum point. Hence the BH confidence set is not the confidence set we want. If one uses BH confidence set for a stationary point as the confidence set for a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ only when the eigenvalues or the fitted function $f(\boldsymbol{x}, \hat{\boldsymbol{\theta}})$ indicates that the stationary point is a maximum point, see, e.g., Stablein et al. (1983), then the data snooping for the curvature at a stationary point will invalidate the $(1-\alpha)$ confidence level.

### 1.2.2 Bootstrap Confidence Sets

Bootstrap is a popular method for assigning measures of accuracy to sample estimates due to its simplicity. In regression problem, a bootstrapping procedure can be performed by resampling either the residuals or the observations $(y_1, \boldsymbol{x}_1), \cdots, (y_n, \boldsymbol{x}_n)$; see, e.g., Efron and Tibshirani (1993), Davison and Hinkley (1997) and Farebrother (1998). One can construct bootstrap confidence sets by using different bootstrap approach, such as the Basic Bootstrap, Percentile Bootstrap or Studentized Bootstrap. However, the true confidence level of a bootstrap confidence set is difficult to assess for both large and finite sample sizes.

### 1.2.3 Rao's Confidence Sets

Rao (1973, pp.473) provides a general method of constructing a conservative confidence set for $r(\boldsymbol{\theta})$ where $r(\boldsymbol{\theta})$ is any given function of $\boldsymbol{\theta}$. Applying this method, Carter et al. (1984) construct conservative confidence sets for both the location of, and the response at, the stationary point of the function $f(\boldsymbol{x}, \boldsymbol{\theta})$. This method can also be used directly for the construction of a confidence set for a maximum point; see Section 3.2 for details.

### 1.2.4 Confidence Sets Constructed by Using Neyman's Theorem

Neyman's Theorem (Neyman, 1937) can be applied to construct an exact $(1-\alpha)$ level confidence set for a maximum point of a function by inverting a family of exact $(1-\alpha)$ acceptance sets. This method of constructing a $(1-\alpha)$ confidence set has been used and generalized to construct numerous intriguing confidence sets; see, e.g., Lehmann (1986), Stefansson et al. (1988), Hayter and Hsu (1994), Finner and Strassburge (2002), Huang and Hsu (2007) and Uusipaikka (2008). Note that the BH confidence set for a stationary point is constructed by using this method too.

Indeed this method is also used in Peterson et al. (2002) and Cahya et al. (2004) (referred to as PCD and CDP, respectively, henceforth) to construct a confidence set for a minimum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ in Equation (1.1), which is also a confidence set for a maximum point of $f(\boldsymbol{x}, -\boldsymbol{\theta})$. However the critical constant recommended by them for $f(\boldsymbol{x}, -\boldsymbol{\theta})$ is shown in Chapter 5 to be too small and so the corresponding confidence level is smaller than the claimed $(1-\alpha)$.

## 1.3 Outline of the Thesis

In this thesis, we focus on the construction of confidence sets for a maximum point of a function in a given covariate region, and propose an approach based on Neyman's Theorem. The layout of the thesis is as follows:

Chapter 2 introduces Neyman's Theorem which is used for constructing a confidence set together with some examples illustrating the constructing procedure. The BH confidence set for the stationary point(s) is an important application of this theorem and is also introduced in Chapter 2.

Chapter 3 is concerned with the construction of confidence sets for a maximum point of a univariate polynomial regression function

$$f(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_p x^p$$

over the given interval $[a, b]$, and compare our confidence sets with Rao's (Rao, 1973), bootstrap and BH confidence sets by using both real and simulation data.

Chapter 4 considers the construction of a confidence set for a maximum point of a bivariate quadratic regression function

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2 + \theta_{12} x_1 x_2$$

over the rectangular region $\{\boldsymbol{x} = (x_1, x_2) : x_1 \in [a_1, b_1], x_2 \in [a_2, b_2]\}$. Rao's confidence set, bootstrap confidence set and BH confidence set are also discussed in the examples given in this chapter.

Chapter 5 extends the method of constructing a confidence set for a maximum point of a linear regression function developed in Chapters 3 and 4 to some more general models involving a linear function. Generalized linear models, the Cox-proportional hazard models and a Becker's H1 model are included as examples to illustrate the procedure.

Chapter 6 summaries the work presented in this thesis and discuss the possible future works.

# Chapter 2

# Methods

## 2.1 Construction of a Confidence Set by Using Neyman's Theorem

The purpose of this chapter is to describe a theorem which may be used to construct a confidence set for a maximum point of a regression function. This theorem is first given by Neyman (1937) and has been introduced in many statistical textbooks; see, e.g., Lehmann (1986, pp214), Rao (1973, pp471), or Casella and Berger (2002, pp420-422). In fact, Neyman's Theorem provides a general method of constructing confidence sets for the parameters or any function of the parameters by inverting a family of acceptance sets of hypothesis tests, which includes the case of a maximum point.

We introduce Neyman's Theorem in Section 2.1 followed by some examples. One important application of this theorem is the construction of a confidence set for the solution of simultaneous equations (Box and Hunter, 1954), which is also included in this section. In Section 2.2, Rao's confidence set and Bootstrap confidence set are illustrated and discussed for the applications on a maximum point of a given function.

### 2.1.1 Neyman's Theorem

**Theorem.** Suppose a random observation $\boldsymbol{Y}$ has the distribution $h(\boldsymbol{y};\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is the unknown parameter. Let $\mathbb{B}$ and $\Omega$ be the parameter space and sample space, respectively. For each $\boldsymbol{\gamma}_0 \in \mathbb{B}$, let $A(\boldsymbol{\gamma}_0) \subset \Omega$ be the acceptance set of a size $\alpha$ test of $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, i.e., $\mathrm{P}_{\boldsymbol{\gamma}_0}\{\boldsymbol{Y} \in A(\boldsymbol{\gamma}_0)\} \geq 1-\alpha$ in which the probability is calculated at $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$. For each $\boldsymbol{Y} \in \Omega$, define a set $\mathbf{C}(\boldsymbol{Y}) \subset \mathbb{B}$ by

$$\mathbf{C}(\boldsymbol{Y}) = \{\ \boldsymbol{\gamma}_0 : \boldsymbol{Y} \in A(\boldsymbol{\gamma}_0)\ \}. \tag{2.1}$$

Then the random set $\mathbf{C}(\boldsymbol{Y})$ is a $(1-\alpha)$ level confidence set for $\boldsymbol{\gamma}$.

**Proof.** Since $A(\boldsymbol{\gamma}_0)$ is the acceptance set of a size $\alpha$ test, we have

$$P_{\boldsymbol{\gamma}_0}(\boldsymbol{Y} \notin A(\boldsymbol{\gamma}_0)) \leq \alpha$$

and hence

$$P_{\boldsymbol{\gamma}_0}(\boldsymbol{Y} \in A(\boldsymbol{\gamma}_0)) \geq 1-\alpha.$$

Because $\boldsymbol{\gamma}_0$ is arbitrary, we can replace $\boldsymbol{\gamma}_0$ with $\boldsymbol{\gamma}$. Then the above inequality, together with the definition of $C(\boldsymbol{Y})$ in (2.1), shows that the coverage probability of the set $C(\boldsymbol{Y})$ is given by

$$P_{\boldsymbol{\gamma}}(\boldsymbol{\gamma} \in C(\boldsymbol{Y})) = P_{\boldsymbol{\gamma}}(\boldsymbol{Y} \in A(\boldsymbol{\gamma})) \geq 1-\alpha,$$

showing that $C(\boldsymbol{Y})$ is a $(1-\alpha)$ level confidence set for $\boldsymbol{\gamma}$.□

Note that if the size of the test of $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ is equal to $\alpha$ for each $\boldsymbol{\gamma}_0 \in \mathbb{B}$ then the confidence level of $\mathbf{C}(\boldsymbol{Y})$ is equal to $(1-\alpha)$.

### 2.1.2 Some Simple Examples

Neyman's Theorem can be applied directly to construct confidence sets. We give the following examples to illustrate the constructing method.

**Example 1.** Suppose we have a single random observation $Y$ from $N(\beta, 1)$, where $\beta$ is the unknown mean of the normal distribution. We want to construct a $(1-\alpha)$ level confidence set for $\beta$.

According to Neyman's Theorem, we can construct a confidence set for $\beta$ by inverting a family of acceptance sets for testing $H_0 : \beta = \beta_0$ for each $\beta_0 \in (-\infty, \infty)$. There are several ways to build an acceptance set for testing $H_0 : \beta = \beta_0$.

Because $Y \sim N(\beta, 1)$, that is, $Y - \beta \sim N(0, 1)$, therefore for each $\beta_0 \in (-\infty, \infty)$ we may construct an acceptance set as

$$A(\beta_0) = \{Y : |Y - \beta_0| < z_{\alpha/2}\}$$

where $z_a$ is the upper $a$ point of the standard normal distribution. Note that this test is of size $\alpha$ exactly. Then according to Neyman's Theorem, a $(1 - \alpha)$ level confidence set for $\beta$ is given by

$$\begin{aligned} C(Y) &= \{\beta_0 : Y \in A(\beta_0)\} \\ &= \{\beta_0 : |Y - \beta_0| < z_{\alpha/2}\} \\ &= \{\beta_0 : Y - z_{\alpha/2} < \beta_0 < Y + z_{\alpha/2}\} \\ &= (Y - z_{\alpha/2},\ Y + z_{\alpha/2}). \end{aligned}$$

Figure 2.1 illustrates the relationship between the acceptance set $A(\beta_0)$ and the corresponding confidence set $C(Y)$.

Alternatively, if we use an acceptance set given by

$$A(\beta_0) = \{Y : Y - \beta_0 < z_\alpha\}$$

for each $\beta_0 \in (-\infty, \infty)$, then a $(1 - \alpha)$ level confidence set for $\beta$ is given by

$$\begin{aligned} C(Y) &= \{\beta_0 : Y - \beta_0 < z_\alpha\} \\ &= (Y - z_\alpha, \infty). \end{aligned}$$

Or, if we use

$$A(\beta_0) = \{Y : Y - \beta_0 > -z_\alpha\},$$

then the $(1 - \alpha)$ level confidence set for $\beta$ is given by

$$\begin{aligned} C(Y) &= \{\beta_0 : Y - \beta_0 > -z_\alpha\} \\ &= (-\infty, Y + z_\alpha). \end{aligned}$$

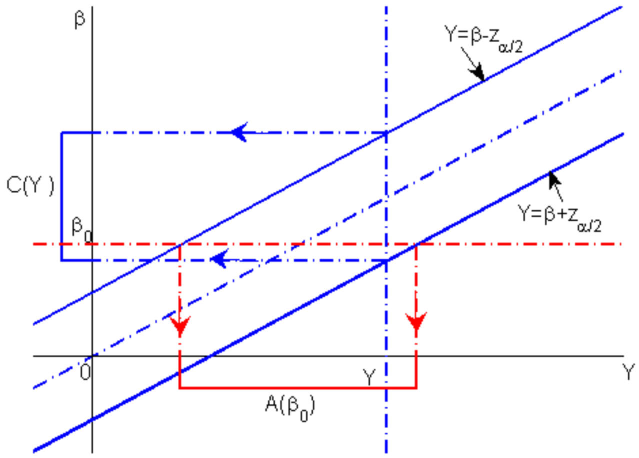



Figure 2.1: The relationship between the acceptance set $A(\beta_0)$ and the confidence set $C(Y)$.

**Example 2.** (Continued from Example 1) Now we want to construct a $(1-\alpha)$ level confidence set for $\theta = |\beta|$.

We test $H_0 : \theta = \theta_0$ by using the $(1-\alpha)$ level acceptance set

$$A(\theta_0) = \{Y : |Y| > c_{\theta_0}\}$$

where $c_{\theta_0} = c(\theta_0) > 0$ such that $P_{\theta_0}\{Y \in A(\theta_0)\} = 1-\alpha$. Let $N = Y - \beta \sim N(0,1)$. Then under the hypothesis $H_0$, we have

$$\begin{aligned} P_{\theta_0}\{Y \in A(\theta_0)\} &= P_{\theta_0}\{|Y| \geq c(\theta_0)\} \\ &= P_{\theta_0}\{Y \geq c(\theta_0)\} + P_{\theta_0}\{Y \leq -c(\theta_0)\} \\ &= P_{\theta_0}\{N \geq c(\theta_0) - \beta\} + P_{\theta_0}\{N \leq -c(\theta_0) - \beta\} \\ &= P_{\theta_0}\{N \geq c(\theta_0) - \beta\} + P_{\theta_0}\{N \geq c(\theta_0) + \beta\} \\ &= P_{\theta_0}\{N \geq c(\theta_0) - |\beta|\} + P_{\theta_0}\{N \geq c(\theta_0) + |\beta|\} \\ &= P_{\theta_0}\{N \geq c(\theta_0) - \theta_0\} + P_{\theta_0}\{N \geq c(\theta_0) + \theta_0\} \qquad (2.2) \\ &= \Phi(-c(\theta_0) + \theta_0) + \Phi(-c(\theta_0) - \theta_0) \end{aligned}$$

where Equation (2.2) follows directly from the hypothesis $\theta_0 = \theta = |\beta|$. By setting the last expression $\Phi(-c(\theta_0)+\theta_0) + \Phi(-c(\theta_0)-\theta_0) = 1-\alpha$, we have $c(\theta_0)$ for each

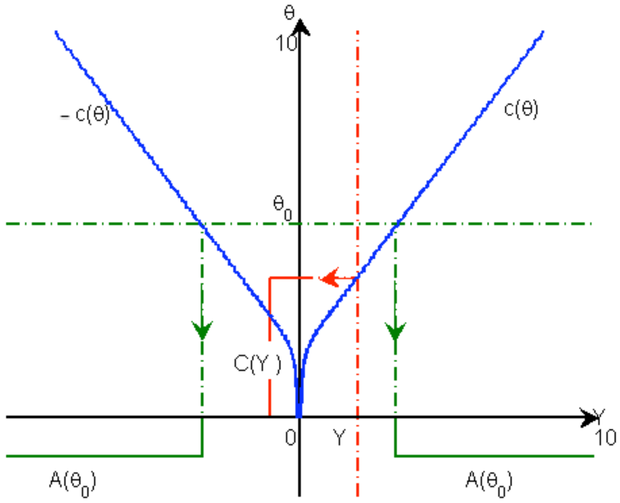



Figure 2.2: The relationship between the 95% level confidence set $C(Y)$ and the corresponding acceptance set $A(\theta_0) = \{Y : |Y| > c_{\theta_0}\}$.

$\theta_0$. It is shown that $c(\theta)$ is monotone increasing in $[0, \infty)$. Then the corresponding $(1-\alpha)$ level confidence set is given by

$$\begin{aligned} C(Y) &= \{\theta_0 : \ Y \in A(\theta_0)\} \\ &= \{\theta_0 : \ |Y| > c_{\theta_0}\} \\ &= \{\theta_0 : \ c(\theta_0) < |Y|\} \\ &= \{\theta_0 : \ \theta_0 < c^{-1}(|Y|)\} \\ &= [0, c^{-1}(|Y|)) \end{aligned}$$

where $c^{-1}(.)$ is the inverse function of $c(.)$, which exists uniquely since $c(.)$ is monotone increasing. Figure 2.2 shows the relationship between the 95% level acceptance set $A(\theta_0) = \{Y : |Y| > c_{\theta_0}\}$ and the corresponding confidence set $C(Y)$ when $\theta \in [0, 10]$.

**Example 3.** (Continued from example 2) Now we use a different acceptance set for the size $\alpha$ test $H_0 : \theta = \theta_0$. Let

$$A(\theta_0) = \{Y : |Y| < d(\theta_0)\}$$

for each $\theta_0 \in [0, \infty)$, where $d(\theta_0) > 0$ is the critical value. Under the hypothesis $H_0$, let $\beta_0 = E(Y)$ so that $\theta_0 =| \beta_0 |$. Similar to Example 2, we have

$$\begin{aligned} P\{Y \in A(\theta_0)\} &= P\{|Y| < d(\theta_0)\} \\ &= P\{-d(\theta_0) < Y < d(\theta_0)\} \\ &= P\{-d(\theta_0) - \beta_0 < Y - \beta_0 < d(\theta_0) - \beta_0\} \\ &= \Phi(d(\theta_0) - \beta_0) - \Phi(-d(\theta_0) - \beta_0) \\ &= \Phi(d(\theta_0) - \beta_0) - (1 - \Phi(d(\theta_0) + \beta_0)) \\ &= \Phi(d(\theta_0) - \theta_0) + \Phi(d(\theta_0) + \theta_0) - 1, \end{aligned}$$

where $d(\theta_0)$ can be decided by setting the coverage probability to $(1-\alpha)$. Then the corresponding $(1-\alpha)$ level confidence set for $\theta$ is given by

$$\begin{aligned} C(Y) &= \{\theta_0 : Y \in A(\theta_0)\} \\ &= \{\theta_0 : |Y| < d(\theta_0)\} \\ &= \{\theta_0 : d^{-1}(|Y|) < \theta_0\} \\ &= (d^{-1}(|Y|), \infty) \end{aligned}$$

where $d^{-1}(.)$ is the inverse function of $d(.)$, which exists uniquely since $d(.)$ is monotone increasing. Figure 2.3 shows the relationship between the 95% level acceptance set $A(\theta_0) = \{Y : |Y| < d(\theta_0)\}$ and the corresponding confidence set $C(Y)$. The critical value $d(\theta)$ is a function of $\theta$ in [0,10], and is represented by the right-hand side of the two symetric curves.

**Example 4.** For the simple linear regression model $Y = \theta_0 + \theta_1 x + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$, we have $n$ observations $(Y_i, x_i)$, $i = 1, 2, \cdots, n$. Let $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)^T$, $\mathbf{X}$ be the design matrix $\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$ and $\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$. Then

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y} \sim N\left(\boldsymbol{\theta}, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2\right)$$

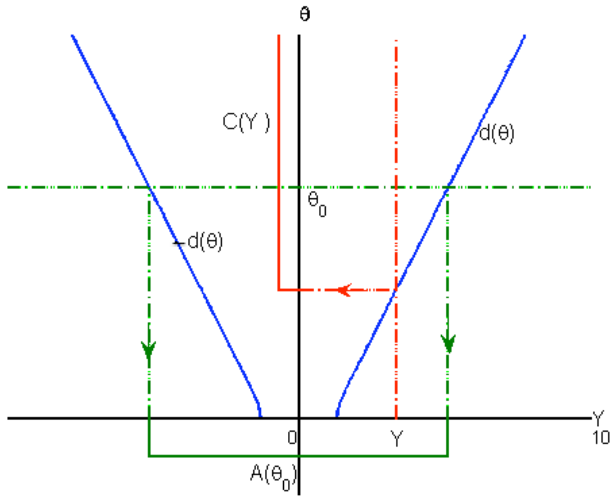



Figure 2.3: The relationship between the 95% level confidence set $C(Y)$ and the corresponding acceptance set $A(\theta_0) = \{Y : |Y| < d_{\theta_0}\}$.

is the least squares estimate of $\boldsymbol{\theta}$. Let $\hat{\sigma}^2$ be an usual estimate of $\sigma^2$ with degrees of freedom $(n-2)$, hence $\hat{\sigma}^2/\sigma^2$ has the distribution $\chi^2_{n-2}/(n-2)$ and is independent of the coefficients $\hat{\boldsymbol{\theta}}$.

We want to construct a $(1-\alpha)$ level confidence set for $x_0$, where $x_0$ is given by $m_0 = \theta_0 + \theta_1 x_0$ and $m_0$ is a given number. We need an acceptance set for testing

$$H_0 : x_0 = t_0$$

with each $t_0 \in (-\infty, \infty)$. An exact $(1-\alpha)$ level acceptance set is given by

$$A(t_0) = \left\{ \boldsymbol{Y} : \frac{\hat{\theta}_0 + \hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1,t_0)(\mathbf{X}^T\mathbf{X})^{-1}(1,t_0)^T}} < c \right\}$$

where the critical value $c$ can be determined such that $P\{\boldsymbol{Y} \in A(t_0)\} = 1-\alpha$. Since $\hat{\theta}_0 + \hat{\theta}_1 x_0 - m_0 \sim N(0, (1,t_0)(\mathbf{X}^T\mathbf{X})^{-1}(1,t_0)^T\sigma^2)$, so

$$\frac{\hat{\theta}_0 + \hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1,t_0)(\mathbf{X}^T\mathbf{X})^{-1}(1,t_0)^T}} \sim \boldsymbol{T}_{n-2},$$

where $\boldsymbol{T}_{n-2}$ is the standard $t$ distribution with degrees of freedom $(n-2)$. Hence the critical value c should be equal to $t^{\alpha}_{n-2}$. Then the $(1-\alpha)$ level confidence set

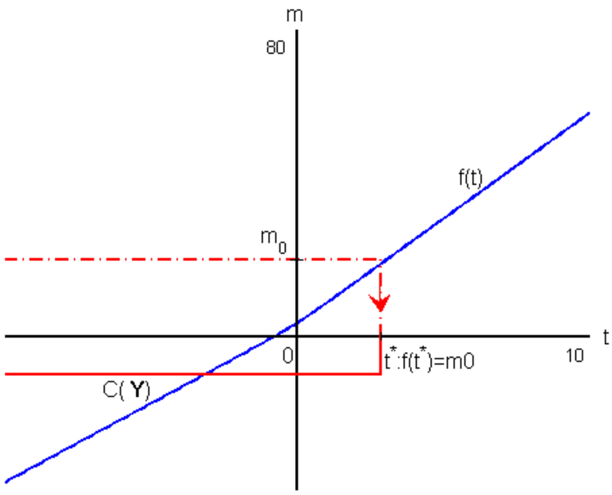



Figure 2.4: The relationship between the value of $m_0$ and the 95% level confidence set $C(\boldsymbol{Y})$ for $x_0$.

for $x_0$ is given by

$$\begin{aligned} C(\boldsymbol{Y}) &= \{t_0 : \boldsymbol{Y} \in A(t_0)\} \\ &= \left\{ t_0 : \frac{\hat{\theta}_0 + \hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1,t_0)(\mathbf{X}^T\mathbf{X})^{-1}(1,t_0)^T}} < t^{\alpha}_{n-2} \right\} \\ &= \left\{ t_0 : \hat{\theta}_0 + \hat{\theta}_1 t_0 - t^{\alpha}_{n-2}\hat{\sigma}\sqrt{(1,t_0)(\mathbf{X}^T\mathbf{X})^{-1}(1,t_0)^T} < m_0 \right\}. \end{aligned}$$

Figure 2.4 illustrates the relationship between the value of $m_0$ and the 95% confidence set $C(\boldsymbol{Y})$ for $x_0$. The curve in the figure is

$$f(t) = \hat{\theta}_0 + \hat{\theta}_1 t_0 - t^{\alpha}_{n-2}\hat{\sigma}\sqrt{(1,t_0)(\mathbf{X}^T\mathbf{X})^{-1}(1,t_0)^T},$$

and the confidence set $C(\boldsymbol{Y})$ is $(-\infty, t^*)$ where $t^*$ solves $f(t^*) = m_0$.

Since there are different acceptance sets for a size $\alpha$ test, we can construct different confidence sets for $x_0$. For example, if we use an alternative acceptance set given by

$$A(t_0) = \left\{ \boldsymbol{Y} : \frac{\hat{\theta}_0 + \hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1,t_0)(\mathbf{X}^T\mathbf{X})^{-1}(1,t_0)^T}} > -t^{\alpha}_{n-2} \right\},$$

then the corresponding $(1-\alpha)$ level confidence set for $t_0$ is given by

$$\begin{aligned} C(\boldsymbol{Y}) &= \{t_0 : \boldsymbol{Y} \in A(t_0)\} \\ &= \left\{ t_0 : \frac{\hat{\theta}_0 + \hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1,t_0)(\mathbf{X}^T\mathbf{X})^{-1}(1,t_0)^T}} > -t_{n-2}^{\alpha} \right\} \\ &= \left\{ t_0 : m_0 < \hat{\theta}_0 + \hat{\theta}_1 t_0 + t_{n-2}^{\alpha}\hat{\sigma}\sqrt{(1,t_0)(\mathbf{X}^T\mathbf{X})^{-1}(1,t_0)^T} \right\}. \end{aligned}$$

### 2.1.3 Confidence Set for the Solution of a Set of Simultaneous Equations

Suppose we are interested in the solution $\boldsymbol{x} = (x_1, \cdots, x_p)^T$ of the following simultaneous equations

$$\begin{cases} a_{10} + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p = 0 \\ a_{20} + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p = 0 \\ \cdots \\ a_{p0} + a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p = 0. \end{cases} \tag{2.3}$$

Because the true value of $(a_{ij}, i = 1, \cdots, p; j = 0, \cdots, p)$ are unknown, and we only have their estimates $(\hat{a}_{ij}, i = 1, \cdots, p; j = 0, \cdots, p)$ which are distributed multinormally with mean $(a_{ij}, i = 1, \cdots, p; j = 0, \cdots, p)$ and a $p(p+1) \times p(p+1)$ variance-covariance matrix $\Omega\sigma^2$, known apart from $\sigma^2$. Let $\hat{\sigma}^2$ be an estimate of $\sigma^2$ with degrees of freedom $\phi$, hence $\hat{\sigma}^2/\sigma^2$ is distributed as $\chi^2(\phi)/\phi$ and is independent of the coefficients $\hat{a}_{ij}$. Box and Hunter (1954) constructed a confidence set for the solution $\boldsymbol{x} = (x_1, \cdots, x_p)^T$ using Neyman's Theorem in the following way. Denote

$$\hat{\delta}_i(\boldsymbol{x}) = \hat{a}_{i0} + \hat{a}_{i1}x_1 + \hat{a}_{i2}x_2 + \cdots + \hat{a}_{ip}x_p, \qquad i = 1, 2, \cdots, p,$$

and $\hat{\boldsymbol{\delta}}(\boldsymbol{x}) = (\hat{\delta}_1(\boldsymbol{x}), \hat{\delta}_2(\boldsymbol{x}), \ldots, \hat{\delta}_p(\boldsymbol{x}))^T$. It is clear that $E(\hat{\boldsymbol{\delta}}(\boldsymbol{x})) = \mathbf{0}$. Let

$$\begin{aligned} V(\boldsymbol{x}) &= cov(\hat{\boldsymbol{\delta}}(\boldsymbol{x}), \hat{\boldsymbol{\delta}}(\boldsymbol{x}))/\sigma^2 \\ &= E[\hat{\boldsymbol{\delta}}(\boldsymbol{x})\hat{\boldsymbol{\delta}}(\boldsymbol{x})^T]/\sigma^2, \end{aligned}$$

then

$$\frac{\hat{\boldsymbol{\delta}}(\boldsymbol{x})^T V(\boldsymbol{x})^{-1}\hat{\boldsymbol{\delta}}(\boldsymbol{x})}{\sigma^2} \sim \chi^2(p) \text{ and } \frac{\hat{\boldsymbol{\delta}}(\boldsymbol{x})^T V(\boldsymbol{x})^{-1}\hat{\boldsymbol{\delta}}(\boldsymbol{x})}{p\hat{\sigma}^2} \sim F_{p,\phi},$$

where $F_{p,\phi}$ is the $F$-distribution with degrees of freedom $p$ and $\phi$. Let $\boldsymbol{x}^0 = (x_1^0, x_2^0, \cdots, x_p^0)$ be given and for testing

$$H_0 : \boldsymbol{x}^0 \text{is the solution of the simultaneous equation (2.3)},$$

we construct the $(1-\alpha)$ level acceptance set

$$A(\boldsymbol{x}^0) = \left\{ \boldsymbol{Y} : \frac{\hat{\boldsymbol{\delta}}(\boldsymbol{x}^0)^T V(\boldsymbol{x}^0)^{-1} \hat{\boldsymbol{\delta}}(\boldsymbol{x}^0)}{p\hat{\sigma}^2} \le f_{p,\phi}^{\alpha} \right\},$$

where $f_{p,\phi}^{\alpha}$ is the upper $\alpha$ point of an F distribution with degrees of freedom $p$ and $\phi$. Directly from Neyman's Theorem, a $(1-\alpha)$ level confidence set for the solution $\boldsymbol{x}$ is given by

$$C(\boldsymbol{Y}) = \left\{ \boldsymbol{x} : \frac{\hat{\boldsymbol{\delta}}(\boldsymbol{x})^T V(\boldsymbol{x})^{-1} \hat{\boldsymbol{\delta}}(\boldsymbol{x})}{p\hat{\sigma}^2} \le f_{p,\phi}^{\alpha} \right\}.$$

A program for computing this confidence set is provided by Del Castillo and Cahya (2001).

**Example: Confidence set for the stationary point of a quadratic function**

Suppose we have $n$ observations $(Y_i, x_i),\ i = 1, \cdots, n$, from a usual quadratic linear regression model

$$Y = \theta_0 + \theta_1 x + \theta_2 x^2 + \varepsilon,\ \varepsilon \sim N(0, \sigma^2)$$

with unknown parameter $\sigma^2$. Let $\hat{\boldsymbol{\theta}} = (\hat{\theta_0}, \hat{\theta_1}, \hat{\theta_2})^T$ be the least squares estimates of $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)^T$ for which we know $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2)$, where $\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}$.

Let $\hat{\sigma}$ be the usual estimate of $\sigma$, and $\hat{\sigma}^2 \sim \frac{\sigma^2}{n-3}\chi_{n-3}^2$. Moreover, $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ are independent. We are interested in constructing a confidence set for the stationary point of the quadratic function $f(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2$.

Since the stationary point of $f(x, \boldsymbol{\theta})$ is the solution $x$ of the equation $\theta_1 + 2\theta_2 x = 0$, following the approach given above, we denote $\hat{\delta}(x) = \hat{\theta_1} + 2\hat{\theta_2} x = (1, 2x) \begin{pmatrix} \hat{\theta_1} \\ \hat{\theta_2} \end{pmatrix}$.

Hence

$$\begin{aligned}
V(x) &= E[\hat{\delta}(x)\hat{\delta}(x)^T]/\sigma^2 \\
&= cov[\hat{\delta}(x)]/\sigma^2 \\
&= cov\left[(1,2x)\begin{pmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{pmatrix}\right]/\sigma^2 \\
&= (1,2x)cov[\begin{pmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{pmatrix}]\begin{pmatrix}1 \\ 2x\end{pmatrix}/\sigma^2 \\
&= (1,2x)\begin{pmatrix}v_{11} & v_{12} \\ v_{21} & v_{22}\end{pmatrix}\begin{pmatrix}1 \\ 2x\end{pmatrix}
\end{aligned}$$

where $\begin{pmatrix}v_{11} & v_{12} \\ v_{21} & v_{22}\end{pmatrix} = cov\begin{pmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{pmatrix}/\sigma^2$.

Then the $(1-\alpha)$ level acceptance set is given by

$$\begin{aligned}
A(x_0) &= \left\{\boldsymbol{Y} : \frac{\hat{\delta}(x_0)^2}{V(x_0)p\hat{\sigma}^2} \le f^{\alpha}_{1,n-3}\right\} \\
&= \left\{\boldsymbol{Y} : \frac{(\hat{\theta}_1 + 2\hat{\theta}_2 x_0)^2}{(1,2x_0)\begin{pmatrix}v_{11} & v_{12} \\ v_{21} & v_{22}\end{pmatrix}\begin{pmatrix}1 \\ 2x_0\end{pmatrix}\hat{\sigma}^2} \le f^{\alpha}_{1,n-3}\right\} \\
&= \left\{\boldsymbol{Y} : \frac{(\hat{\theta}_1 + 2\hat{\theta}_2 x_0)^2}{(v_{11} + 4x_0 v_{12} + 4x_0^2 v_{22})\hat{\sigma}^2} \le f^{\alpha}_{1,n-3}\right\}.
\end{aligned}$$

Hence the $(1-\alpha)$ level confidence set for the stationary point $x$ can be constructed as

$$\begin{aligned}
C(\boldsymbol{Y}) &= \{x : \boldsymbol{Y} \in A(x)\} \\
&= \left\{x : \frac{(\hat{\theta}_1 + 2\hat{\theta}_2 x)^2}{(v_{11} + 4xv_{12} + 4x^2 v_{22})\hat{\sigma}^2} \le f^{\alpha}_{1,n-3}\right\}.
\end{aligned}$$

This method given by Box and Hunter has been widely used in constructing the confidence sets in many practical problems, for example, in constructing confidence set for the optimal treatment of cancer (Carter et al., 1982) , for shoot regeneration protocol (Chakraborty et al., 2010) and for pigs' preference (Jensen and Pedersen, 2007).

## 2.2 Other Methods of Constructing a Confidence Set

In this section, confidence sets for a maximum point using Rao's method and bootstrap method are constructed. Suppose we have $n$ observations $(Y_i, \boldsymbol{x}_i)$, $i = 1, 2, \cdots, n$, from the linear regression model (1.1)

$$Y = f(\boldsymbol{x}, \boldsymbol{\beta}) + e.$$

Then the observations can be written as

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$$

where $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)^T$, $\mathbf{X}$ is the design matrix given by

$$\mathbf{X} = \begin{pmatrix} 1, \mathbf{z}(\boldsymbol{x}_1)^T \\ 1, \mathbf{z}(\boldsymbol{x}_2)^T \\ \vdots \\ 1, \mathbf{z}(\boldsymbol{x}_n)^T \end{pmatrix},$$

$\boldsymbol{e} = (e_1, e_2, \cdots, e_n)^T$ has the distribution $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Furthermore, the usual estimates of $\boldsymbol{\beta}$ and $\sigma^2$ are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y} \sim N(\boldsymbol{\beta}, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2) \text{ and } \hat{\sigma}^2 = \frac{\parallel \hat{\boldsymbol{e}} \parallel^2}{n-p-1} \sim \frac{\sigma^2}{n-p-1}\chi^2_{n-p-1},$$

where $\hat{\boldsymbol{e}} = (\hat{e}_1, \cdots, \hat{e}_n)^T = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$ and $\hat{\boldsymbol{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

### 2.2.1 The Conservative Confidence Set Given by Rao

Rao (1973, pp.473) suggested a conservative confidence set for any function of the parameters as follows (see also Farebrother, 1998). Let $r(.)$ be any function of $\boldsymbol{\beta}$, then a $(1-\alpha)$ level conservative confidence set for $r(\boldsymbol{\beta})$ is given by

$$\mathbf{C}_c = \{r(\boldsymbol{\beta}^*) : \boldsymbol{\beta}^* \in C_{\boldsymbol{\beta}}\},$$

where $C_{\boldsymbol{\beta}}$ is a $(1-\alpha)$ level confidence set for $\boldsymbol{\beta}$.

To use this method in constructing a $(1-\alpha)$ level confidence set for $r(\boldsymbol{\beta})$, we need a confidence set for $\boldsymbol{\beta}$ first. From the distributions of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, it is clear that

$$C_{\boldsymbol{\beta}} = \{\boldsymbol{\beta}^* : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T(\mathbf{X}^T\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) < (p+1)\hat{\sigma}^2 f^{\alpha}_{p+1,n-p-1}\}$$

is a $(1-\alpha)$ level confidence set for $\boldsymbol{\beta}$. Then, a $(1-\alpha)$ conservative confidence set for $r(\boldsymbol{\beta})$ is given by

$$\mathbf{C}_c = \{r(\boldsymbol{\beta}^*) : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T(\mathbf{X}^T\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) < (p+1)\hat{\sigma}^2 f^{\alpha}_{p+1,n-p-1}\}.$$

### 2.2.2 The Bootstrap Percentile Confidence Set

The bootstrap method is first introduced by Efron (1979), and has been used in a variety of estimation problems including the construction of a confidence set for any function $r(\boldsymbol{\beta})$. There are two ways to obtain the bootstrap datasets:

(a) Resample the residuals. Randomly choose a set of $n$ bootstrap residuals from the original residuals $\{\hat{e}_1, \hat{e}_2, \cdots, \hat{e}_n\}$ with replacement, and denote it as $\hat{\boldsymbol{e}}^B = \{\hat{e}^B_1, \hat{e}^B_2, \cdots, \hat{e}^B_n\}$. Then a bootstrap sample set is given by $\boldsymbol{Y}^B = (Y^B_1, \cdots, Y^B_n)$ where

$$\boldsymbol{Y}^B = \hat{\boldsymbol{Y}} + \hat{\boldsymbol{e}}^B.$$

The estimated parameters based on

$$\boldsymbol{Y}^B = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$$

are denoted as $\hat{\boldsymbol{\beta}}^B$. Since the residuals are supposed to distribute normally with mean 0 and variance $\sigma^2$, one can resample the residuals from the distribution $N(0, \hat{\sigma}^2)$ instead of from the original residuals. The design matrix is not changed when the sampling procedure is on the residuals.

(b) Resample the sample pairs $(y_i, \boldsymbol{x}_i)$. Randomly select $n$ data pairs from the original sample set $(y_i, \boldsymbol{x}_i), i = 1, \cdots, n$ with replacement, and denote them as $(y^B_i, \boldsymbol{x}^B_i), i = 1, \cdots, n$. Then, we re-estimate the parameters from

$$\boldsymbol{Y}^B = \mathbf{X}^B\boldsymbol{\beta} + \boldsymbol{e},$$

where $\boldsymbol{Y}^B = (y_1^B, \cdots, y_n^B)^T$, and $\mathbf{X}^B = \begin{pmatrix} \mathbf{z}(\boldsymbol{x}_1)^B \\ \vdots \\ \mathbf{z}(\boldsymbol{x}_n)^B \end{pmatrix}$. The property that the design matrix changed in each resampled data set is usually undesirable, since the design is fixed and all the observations are assumed to be from this design. Furthermore, since the $n$ pairs are selected randomly, it may happen that all the resampled pairs are the same and hence the parameters can not be estimated from this bootstrap sample set.

We form $N$ bootstrap sample sets by resampling the residuals. For the $i^{th}$ sample set, the parameters $\hat{\boldsymbol{\beta}}_i^B$ are estimated, and the value of $r(\hat{\boldsymbol{\beta}}_i^B)$, denoted as $m_i$, can be computed. To form the bootstrap percentile confidence set, we sort the vector $\boldsymbol{m} = (m_1, m_2, \cdots, m_N)$ in ascending order and then delete $[\alpha \times N]$ values (see section 3.3 for details). The remaining form the $(1-\alpha)$ percentile confidence set for $r(\boldsymbol{\beta})$.

In Chapter 3, we construct confidence sets for a maximum point of a univariate polynomial regression function in a given interval of $x$. We first apply our method of constructing a confidence set using Neyman's Theorem on a simple linear regression model, then extend this method to the quadratic and general univariate polynomial regression models. Rao's confidence set and bootstrap confidence set are also discussed in Chapter 3.

# Chapter 3

# Confidence Set for a Maximum Point of a Univariate Polynomial Regression Function in a Given Interval

In this chapter, we consider the construction of a confidence set for a maximum point of a univariate polynomial function in a given interval. Univariate functions are widely employed in real applications when the interest is on one factor alone. In this case, all other effects are assumed to be fixed and the response is related only to the factor of interest. A univariate polynomial function, where the parameters are estimated immediately using least squares estimation, is among the most popular univariate functions. Actually, many univariate functions can be well approximated by a polynomial function. Therefore, the univariate polynomial function is a good starting point in studying the confidence set for a maximum point.

There are four sections in this chapter. Section 3.1 elaborates our method of constructing a confidence set for a maximum point of a simple linear function, a quadratic function and a general polynomial function. Rao's method and bootstrap method for the general polynomial case are discussed in Section 3.2 and 3.3. Examples are given to compare the confidence sets constructed using these methods and a conclusion is drawn in Section 3.4.

## 3.1 Our Method

Suppose we have $n$ independent observations $(Y_i, x_i)$, $i = 1, 2, \cdots, n$, from a univariate polynomial regression model

$$Y = f(x, \boldsymbol{\theta}) + e = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_p x^p + e,$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_p)^T$ is the vector of the unknown regression coefficients. The random error $e$ is unobservable, and is assumed to have a normal distribution $N(0, \sigma^2)$ with unknown $\sigma^2 > 0$. Hence we have

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{e}$$

where $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)^T$, $\mathbf{X}$ is the design matrix given by

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & \cdots & x_1^p \\ 1 & x_2 & \cdots & x_2^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \cdots & x_n^p \end{pmatrix},$$

$\boldsymbol{e} = (e_1, e_2, \cdots, e_n)^T$ has the distribution $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix.

Suppose $k = k(\boldsymbol{\theta})$ is a maximum point of the regression function $f(x, \boldsymbol{\theta})$ in a given finite interval $[a, b]$. We want to construct a $(1 - \alpha)$ level confidence set for $k$ by using Neyman's Theorem. Note that $k$ must exist since the interval $[a, b]$ is finite and $f(x, \boldsymbol{\theta})$ is a continuous function.

### 3.1.1 A Simple Linear Regression Model

When $p = 1$, we have a simple linear regression model

$$Y = \theta_0 + \theta_1 x + e.$$

Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

and

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y} \sim N\left(\boldsymbol{\theta}, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2\right)$$

be the least squares estimate of $\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$. Let $\hat{\sigma}$ be the usual estimate of $\sigma$, then

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{n-2}\chi^2_{n-2}.$$

Moreover, $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ are independent.

To apply Neyman's Theorem, we require an acceptance set for testing

$$H_0 : k = k_0,$$

with each $k_0 \in [a, b]$. In the following paragraphs, two methods are given for the construction of acceptance sets.

**Method 1.**

Since the regression function is a straight line, it is clear that the slope of this line, $\theta_1$, determines the maximum point(s) in the interval $[a, b]$. If $k_0 = a$, then it is clear that $\theta_1 \leq 0$. Thus, our $(1-\alpha)$ level acceptance set is

$$A(a) = \{\boldsymbol{Y} : \hat{\theta}_1 \leq c(a)\hat{\sigma}v\}$$

where $v = \sqrt{var(\hat{\theta}_1)/\sigma^2}$, and $c(a)$ is the critical value such that A(a) is a $(1-\alpha)$ level acceptance set. Note that

$$\begin{aligned} P_{k_0=a}\{\boldsymbol{Y} \in A(a)\} &= P_{k_0=a}\{\hat{\theta}_1 \leq c(a)\hat{\sigma}v\} \\ &= P_{\theta_1 \leq 0}\{\hat{\theta}_1 \leq c(a)\hat{\sigma}v\} \\ &\geq P_{\theta_1=0}\{\hat{\theta}_1 \leq c(a)\hat{\sigma}v\}. \end{aligned}$$

When $\theta_1 = 0$, we know $\hat{\theta_1}/(\hat{\sigma}v) \sim T_{n-2}$. Hence $c(a)$ should be equal to $t^{\alpha}_{n-2}$, which is the upper $\alpha$ point of the standard t-distribution with degrees of freedom $n-2$.

Similarly, if $k_0 = b$, then $\theta_1 \geq 0$. Thus, a $(1-\alpha)$ level acceptance set is given by

$$A(b) = \{\boldsymbol{Y} : \hat{\theta_1} \geq -c(b)\hat{\sigma}v\},$$

where $c(b)$ is such that the acceptance probability of $A(b)$ is $(1-\alpha)$. Since

$$\begin{aligned} P_{k_0=b}\{\boldsymbol{Y} \in A(b)\} &= P_{k_0=b}\{\hat{\theta_1} \geq -c(b)\hat{\sigma}v\} \\ &= P_{\theta_1 \geq 0}\{\hat{\theta_1} \geq -c(b)\hat{\sigma}v\} \\ &\geq P_{\theta_1=0}\{\hat{\theta_1} \geq -c(b)\hat{\sigma}v\}, \end{aligned}$$

we have $c(b) = t^{\alpha}_{n-2}$.

Finally, if $k_0 = s$, $s \in (a,b)$, then $\theta_1 = 0$. It can be shown that for each $s \in (a,b)$, a $(1-\alpha)$ level acceptance set is

$$A(s) = \{\boldsymbol{Y} : |\hat{\theta_1}| \leq c(s)\hat{\sigma}v\},$$

where $c(s)$ is such that the acceptance probability of $A(s)$ is $(1-\alpha)$. Note that

$$\begin{aligned} P_{k_0=s}\{\boldsymbol{Y} \in A(s)\} &= P_{k_0=s}\{\mid \hat{\theta_1} \mid \leq c(s)\hat{\sigma}v\} \\ &= P_{\theta_1=0}\{\mid \hat{\theta_1} \mid \leq c(s)\hat{\sigma}v\}. \end{aligned}$$

Therefore, we have $c(s) = t^{\alpha/2}_{n-2}$ for any $s \in (a,b)$.

**Method 2.**

Intuitively, if $k_0$ is a maximum point, we should have $(\theta_0 + \theta_1 k_0) - (\theta_0 + \theta_1 x) \geq 0$ for $\forall x \in [a,b]$. This implies $(\hat{\theta_0} + \hat{\theta_1} k_0) - (\hat{\theta_0} + \hat{\theta_1} x)$ should not be too small. We therefore construct an acceptance set for $k_0$ as

$$A(k_0) = \{\boldsymbol{Y} : (\hat{\theta_0} + \hat{\theta_1} k_0) - (\hat{\theta_0} + \hat{\theta_1} x) \geq -c(k_0)\hat{\sigma}v(k_0, x), \quad \forall x \in [a,b]\},$$

where

$$\begin{aligned} v(k_0, x) &= \sqrt{var((\hat{\theta_0} + \hat{\theta_1} k_0) - (\hat{\theta_0} + \hat{\theta_1} x))/\sigma^2} \\ &= \sqrt{var(\hat{\theta_1})(k_0 - x)^2/\sigma^2} \\ &= |k_0 - x|v \end{aligned}$$

with $v = \sqrt{var(\hat{\theta_1})/\sigma^2}$. The critical value $c(k_0)$ is chosen such that the acceptance level of $A(k_0)$ is $(1-\alpha)$.

If $k_0 = a$, the acceptance set can be expressed as

$$\begin{aligned} A(a) &= \{\boldsymbol{Y} : (\hat{\theta_0} + \hat{\theta_1} a) - (\hat{\theta_0} + \hat{\theta_1} x) \geq -c(a)\hat{\sigma} v(a,x), \quad \forall x \in [a,b]\} \\ &= \{\boldsymbol{Y} : \hat{\theta_1}(a-x) \geq -c(a)\hat{\sigma}|a-x|v, \quad \forall x \in [a,b]\} \\ &= \{\boldsymbol{Y} : \hat{\theta_1} \leq c(a)\hat{\sigma} v\}. \end{aligned}$$

The last expression coincides with the acceptance set $A(a)$ constructed in Method 1. Thus we have the same critical value $c(a) = t^{\alpha}_{n-2}$ as in Method 1.

Similarly, if $k_0 = b$, a $(1-\alpha)$ level acceptance set will be

$$\begin{aligned} A(b) &= \{\boldsymbol{Y} : (\hat{\theta_0} + \hat{\theta_1} b) - (\hat{\theta_0} + \hat{\theta_1} x) \geq -c(b)\hat{\sigma} v(b,x), \quad \forall x \in [a,b]\} \\ &= \{\boldsymbol{Y} : \hat{\theta_1}(b-x) \geq -c(b)\hat{\sigma}|b-x|v, \quad \forall x \in [a,b]\} \\ &= \{\boldsymbol{Y} : \hat{\theta_1} \geq -c(b)\hat{\sigma} v\}. \end{aligned}$$

The last expression coincides with the acceptance set $A(b)$ constructed in Method 1, thus we have $c(b) = t^{\alpha}_{n-2}$.

Finally, if $k_0 = s$ with $s \in (a,b)$, then

$$\begin{aligned} A(s) &= \{\boldsymbol{Y} : (\hat{\theta_0} + \hat{\theta_1} s) - (\hat{\theta_0} + \hat{\theta_1} x) \geq -c(s)\hat{\sigma} v(s,x), \quad \forall x \in [a,s) \cup (s,b]\} \\ &= \{\boldsymbol{Y} : \hat{\theta_1}(s-x) \geq -c(s)\hat{\sigma}|s-x|v, \quad \forall x \in [a,s) \cup (s,b]\} \\ &= \{\boldsymbol{Y} : |\hat{\theta_1}| \leq c(s)\hat{\sigma} v\}. \end{aligned}$$

The last expression coincides with the acceptance set $A(s)$ constructed in Method 1, and so we have $c(s) = -t^{\alpha/2}_{n-2}$.

Therefore, we conclude that the acceptance sets given by Method 1 and Method 2 are identical. Neyman's Theorem allows us to construct a $(1-\alpha)$ level confidence set for $k$, which is given by

$$\begin{aligned}
\mathbf{C}_E(\boldsymbol{Y}) &= \{k_0 : \boldsymbol{Y} \in A(k_0)\} \\
&= \begin{cases} b, & \text{if } \hat{h} > t_{n-2}^{\alpha/2} \\ (a, b], & \text{if } t_{n-2}^{\alpha} < \hat{h} \le t_{n-2}^{\alpha/2} \\ [a, b], & \text{if } -t_{n-2}^{\alpha} \le \hat{h} \le t_{n-2}^{\alpha} \\ [a, b), & \text{if } -t_{n-2}^{\alpha/2} \le \hat{h} < -t_{n-2}^{\alpha} \\ a, & \text{if } \hat{h} < -t_{n-2}^{\alpha/2} \end{cases}
\end{aligned}$$

where $\hat{h} = \hat{\theta}/(\hat{\sigma}v)$. In fact, the coverage probability of the confidence set $\mathbf{C}_E(\boldsymbol{Y})$ can be evaluated directly as

$$\begin{aligned}
P(k \in \mathbf{C}_E(\boldsymbol{Y})) &= P\{k \in \mathbf{C}_E(\boldsymbol{Y}),\ \hat{h} > t_{n-2}^{\alpha/2}\} \\
&+ P\{k \in \mathbf{C}_E(\boldsymbol{Y}),\ t_{n-2}^{\alpha} < \hat{h} \le t_{n-2}^{\alpha/2}\} \\
&+ P\{k \in \mathbf{C}_E(\boldsymbol{Y}),\ -t_{n-2}^{\alpha} \le \hat{h} \le t_{n-2}^{\alpha}\} \\
&+ P\{k \in \mathbf{C}_E(\boldsymbol{Y}),\ -t_{n-2}^{\alpha/2} \le \hat{h} < -t_{n-2}^{\alpha}\} \\
&+ P\{k \in \mathbf{C}_E(\boldsymbol{Y}),\ \hat{h} < -t_{n-2}^{\alpha/2}\} \\
&= P\{k = b,\ \hat{h} > t_{n-2}^{\alpha/2}\} \\
&+ P\{k \in (a, b],\ t_{n-2}^{\alpha} < \hat{h} \le t_{n-2}^{\alpha/2}\} \\
&+ P\{k \in [a, b],\ -t_{n-2}^{\alpha} \le \hat{h} \le t_{n-2}^{\alpha}\} \\
&+ P\{k \in [a, b),\ -t_{n-2}^{\alpha/2} \le \hat{h} < -t_{n-2}^{\alpha}\} \\
&+ P\{k = a,\ \hat{h} < -t_{n-2}^{\alpha/2}\}.
\end{aligned}$$

Now, if $k = a$, then

$$\begin{aligned}
P_{k=a}(k \in \mathbf{C}_E(\boldsymbol{Y})) &= P_{k=a}\{a \in \mathbf{C}_E(\boldsymbol{Y})\} \\
&= P_{k=a}\{a = b,\ \hat{h} > t_{n-2}^{\alpha/2}\} \\
&+ P_{k=a}\{a \in (a, b],\quad t_{n-2}^{\alpha} < \hat{h} \le t_{n-2}^{\alpha/2}\} \\
&+ P_{k=a}\{a \in [a, b],\quad -t_{n-2}^{\alpha} \le \hat{h} \le t_{n-2}^{\alpha}\} \\
&+ P_{k=a}\{a \in [a, b),\quad -t_{n-2}^{\alpha/2} \le \hat{h} < -t_{n-2}^{\alpha}\} \\
&+ P_{k=a}\{a = a,\quad \hat{h} < -t_{n-2}^{\alpha/2}\} \\
&= P_{k=a}\{-t_{n-2}^{\alpha} \le \hat{h} \le t_{n-2}^{\alpha}\} \\
&+ P_{k=a}\{-t_{n-2}^{\alpha/2} \le \hat{h} < -t_{n-2}^{\alpha}\} \\
&+ P_{k=a}\{\hat{h} < -t_{n-2}^{\alpha/2}\} \\
&= P_{k=a}\{\hat{h} \le t_{n-2}^{\alpha}\} \\
&= 1 - \alpha.
\end{aligned}$$

Similarly, if $k = b$, then

$$\begin{aligned}
P_{k=b}(k \in \mathbf{C}_E(\boldsymbol{Y})) &= P_{k=b}\{b \in \mathbf{C}_E(\boldsymbol{Y})\} \\
&= P_{k=b}\{\hat{h} > t_{n-2}^{\alpha/2}\} \\
&+ P_{k=b}\{t_{n-2}^{\alpha} < \hat{h} \le t_{n-2}^{\alpha/2}\} \\
&+ P_{k=b}\{-t_{n-2}^{\alpha} \le \hat{h} \le t_{n-2}^{\alpha}\} \\
&= P_{k=b}\{\hat{h} \ge -t_{n-2}^{\alpha}\} \\
&= 1 - \alpha.
\end{aligned}$$

Finally, if $k = s \in (a, b)$, then

$$\begin{aligned}
P_{k=s}(k \in \mathbf{C}_E(\boldsymbol{Y})) &= P_{k=s}\{s \in \mathbf{C}_E(\boldsymbol{Y})\} \\
&= P_{k=s}\{t_{n-2}^{\alpha} < \hat{h} \le t_{n-2}^{\alpha/2}\} \\
&+ P_{k=s}\{t_{n-2}^{\alpha} \le \hat{h} \le t_{n-2}^{\alpha}\} \\
&+ P_{k=s}\{-t_{n-2}^{\alpha/2} \le \hat{h} < -t_{n-2}^{\alpha}\} \\
&= P_{k=s}\{-t_{n-2}^{\alpha/2} \le \hat{h} \le t_{n-2}^{\alpha/2}\} \\
&= 1 - \alpha.
\end{aligned}$$

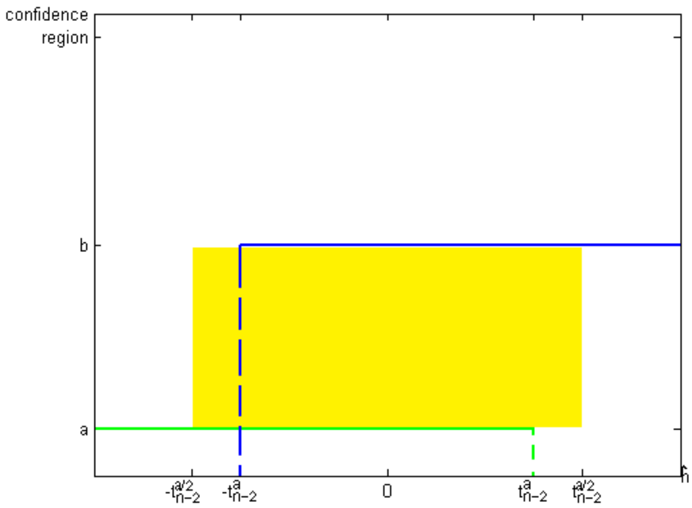



Figure 3.1: Confidence set for a maximum point, $k$, of $f(x) = \theta_0 + \theta_1 x$.

The above results show directly that the confidence level of $\mathbf{C}_E(\boldsymbol{Y})$ is indeed $(1-\alpha)$. This of course agrees with Neyman's Theorem. Figure 3.1 illustrates the confidence interval for $k$ with the value $\hat{h}$ computed based on the observations. The confidence interval is represented by the intersection of $\hat{h}$ and the shaded region (including two lines and a rectangular region). For example, if $\hat{h} = (t_{n-2}^{\alpha} + t_{n-2}^{\alpha/2})/2$, then the corresponding confidence set will be $(a, b]$. If $\hat{h} = t_{n-2}^{\alpha/2} + 1$, then the corresponding confidence set will be $\{b\}$.

### 3.1.2 A Quadratic Polynomial Regression Model

When $p = 2$, we have the quadratic regression model

$$Y = \theta_0 + \theta_1 x + \theta_2 x^2 + e,$$

with

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \text{ and } \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix}.$$

The least squares estimate of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y},$$

which has the distribution $N(\boldsymbol{\theta}, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2)$. Let $\hat{\sigma}^2$ be the usual estimate of $\sigma^2$ which has the distribution

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{n-3}\chi^2_{n-3},$$

where $\chi^2_{n-3}$ is the chi-squared distribution with degrees of freedom $(n-3)$. Moreover, $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ are independent.

We define

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} v_{1,1} & v_{1,2} & v_{1,3} \\ v_{2,1} & v_{2,2} & v_{2,3} \\ v_{3,1} & v_{3,2} & v_{3,3} \end{pmatrix} \text{ and } \mathbf{P}^2 = \begin{pmatrix} v_{2,2} & v_{2,3} \\ v_{3,2} & v_{3,3} \end{pmatrix}.$$

Since

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \hat{\boldsymbol{\theta}},$$

the variance-covariance matrix of $\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix}$ is given by

$$\sigma^2 \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \sigma^2\mathbf{P}^2.$$

Therefore

$$\mathbf{P}^{-1} \times \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} / \sigma \sim N(0, I_2),$$

and hence

$$\boldsymbol{T} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} := \mathbf{P}^{-1} \times \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} / \hat{\sigma} \sim \boldsymbol{T}_{2,n-3} \tag{3.1}$$

where $\boldsymbol{T}_{2,n-3}$ is the standard bivariate t distribution with degrees of freedom $n-3$.

Define the polar coordinates of $\boldsymbol{T}$, $(R_{\boldsymbol{T}}, \psi_{\boldsymbol{T}})$, by

$$T_1 = R_{\boldsymbol{T}} \cos \psi_{\boldsymbol{T}}, \quad T_2 = R_{\boldsymbol{T}} \sin \psi_{\boldsymbol{T}},$$

where $\psi_{\boldsymbol{T}}$ is uniformly distributed in $[0, 2\pi)$, and $R_{\boldsymbol{T}}$ and $\psi_{\boldsymbol{T}}$ are independent. Note that $R_{\boldsymbol{T}}^2/2 \sim F_{2,n-3}$, where $F_{2,n-3}$ is the $F$ distribution with degrees of freedom 2 and $n-3$. So the cumulative density function of $R_{\boldsymbol{T}}$ is given by

$$F_{R_{\boldsymbol{T}}}(x) = 1 - (1 + x^2/n - 3)^{-(n-3)/2}.$$

Moreover, we have

$$\begin{aligned} v(k_0, x) &:= \sqrt{var[(\hat{\theta}_0 + \hat{\theta}_1 k_0 + \hat{\theta}_2 k_0^2) - (\hat{\theta}_0 + \hat{\theta}_1 x + \hat{\theta}_2 x^2)]/\sigma^2} \\ &= \mid k_0 - x \mid \sqrt{var[\hat{\theta}_1 + \hat{\theta}_2(k_0 + x)]/\sigma^2} \\ &= \mid k_0 - x \mid \sqrt{(1, k_0 + x)\mathbf{P}^2(1, k_0 + x)^T} \\ &= \mid k_0 - x \mid \parallel \mathbf{P}(1, k_0 + x)^T \parallel . \end{aligned}$$

#### 3.1.2.1 Theory

To apply Neyman's Theorem, we require an acceptance set for testing

$$H_0 : k = k_0,$$

for each $k_0 \in [a, b]$. Next, we consider three cases: $k_0 = a$, $k_0 = b$ and $k_0 \in (a, b)$.

**Case I** If $k_0 = a$ is a maximum point, then for any $x \in (a, b]$, it is clear that

$$(\theta_0 + \theta_1 a + \theta_2 a^2) - (\theta_0 + \theta_1 x + \theta_2 x^2) \geq 0,$$

which implies $\theta_1 + \theta_2(a+x) \leq 0$ for all $x \in (a, b]$. Accordingly, we define an acceptance set as

$$\begin{aligned} A(a) &= \{\boldsymbol{Y} : (\hat{\theta_0} + \hat{\theta_1} a + \hat{\theta_2} a^2) - (\hat{\theta_0} + \hat{\theta_1} x + \hat{\theta_2} x^2) \geq -c(a)\hat{\sigma} v(a,x),\ \forall x \in (a,b]\} \\ &= \{\boldsymbol{Y} : (\hat{\theta_1} + \hat{\theta_2}(a+x))(a-x) \geq -c(a)\hat{\sigma} v(a,x),\ \forall x \in (a,b]\} \\ &= \{\boldsymbol{Y} : \hat{\theta_1} + \hat{\theta_2}(a+x) \leq c(a)\hat{\sigma} \parallel \mathbf{P}(1, a+x)^T \parallel,\ \forall x \in (a,b]\} \end{aligned} \quad (3.2)$$

where $c(a) > 0$ is a critical constant such that $A(a)$ is a $(1-\alpha)$ level acceptance set. To determine $c(a)$, note that

$$\begin{aligned} & P_{k_0=a} \quad \{\boldsymbol{Y} \in A(a)\} \\ = & P_{k_0=a} \quad \{\hat{\theta_1} + \hat{\theta_2}\ (a+x) \leq c(a)\hat{\sigma} \parallel \mathbf{P}(1, a+x)^T \parallel,\ \forall x \in (a,b]\} \quad (3.3) \\ \geq & P_{k_0=a} \quad \{\hat{\theta_1} + \hat{\theta_2}\ (a+x) \leq c(a)\hat{\sigma} \parallel \mathbf{P}(1, a+x)^T \parallel + (\theta_1 + \theta_2(a+x)),\ \forall x \in (a,b]\} \quad (3.4) \\ = & P_{k_0=a} \quad \{\frac{(1, a+x)(\hat{\theta_1} - \theta_1, \hat{\theta_2} - \theta_2)^T}{\hat{\sigma} \parallel \mathbf{P}(1, a+x)^T \parallel} \leq c(a),\ \forall x \in (a,b]\} \\ = & P_{k_0=a} \quad \{\sup_{x \in (a,b]} \frac{[\mathbf{P}(1, a+x)^T]^T [\mathbf{P}^{-1}(\hat{\theta_1} - \theta_1, \hat{\theta_2} - \theta_2)^T]}{\hat{\sigma} \parallel \mathbf{P}(1, a+x)^T \parallel} \leq c(a)\} \\ = & \quad P \quad \{\sup_{x \in (a,b]} \frac{[\mathbf{P}(1, a+x)^T]^T \boldsymbol{T}}{\parallel \mathbf{P}(1, a+x)^T \parallel} \leq c(a)\} \quad (3.5) \\ = & \quad P \quad \{\boldsymbol{T} \in R_{h,a}\}, \end{aligned}$$

where Equation (3.3) follows directly from Equation (3.2), Equation (3.4) follows from the fact that $\theta_1 + \theta_2(a+x) \leq 0$ for all $x \in (a, b]$, and

$$R_{h,a} = \bigcap_{x \in (a,b]} R_{h,a}(x)$$

with

$$R_{h,a}(x) = \{\boldsymbol{T} : \frac{\{\mathbf{P}(1, a+x)^T\}^T \boldsymbol{T}}{\parallel \mathbf{P}(1, a+x)^T \parallel} \leq c(a)\}.$$

Note that the inequality in (3.4) becomes an equality when $\theta_1 = \theta_2 = 0$ and so

$$\inf_{\boldsymbol{\theta}: k_0=a} P_{k_0=a}\{\boldsymbol{Y} \in A(a)\} = P\{\boldsymbol{T} \in R_{h,a}\}, \quad (3.6)$$

with the infimum being attained at $\theta_1 = \theta_2 = 0$.

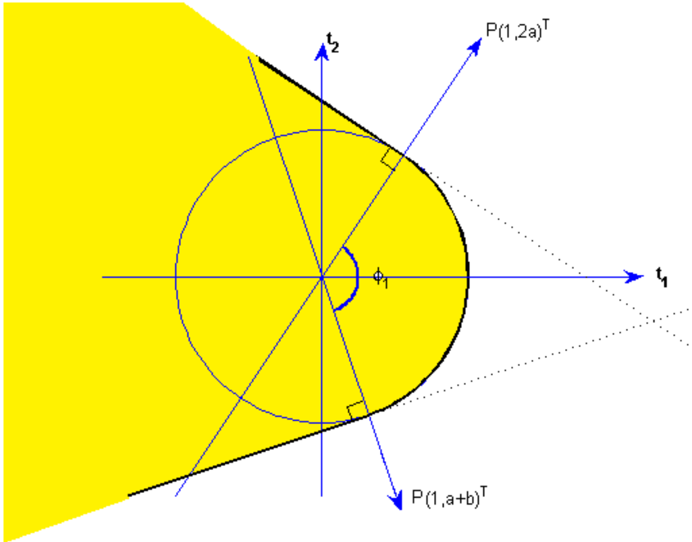



Figure 3.2: The region $R_{h,a}$

Now we consider the computation of $P\{\boldsymbol{T} \in R_{h,a}\}$. Note that $R_{h,a}$ is made up by all the points on the same side as the origin of the straight line that is perpendicular to $\mathbf{P}(1, a+x)^T$ and $c(a)$ distance from the origin in the direction of $\mathbf{P}(1, a+x)^T$. The region $R_{h,a}$ is represented by the shaded region in Figure 3.2. The angle $\phi_1$ between $\mathbf{P}(1, 2a)^T$ and $\mathbf{P}(1, a+b)^T$ satisfies

$$\cos \phi_1 = \frac{[\mathbf{P}(1, 2a)^T]^T [\mathbf{P}(1, a+b)^T]}{\| \mathbf{P}(1, 2a)^T \| \quad \| \mathbf{P}(1, a+b)^T \|}.$$

Following an argument similar to Liu (2010, pp.28-36), the region $R_{h,a}$ can be partitioned into 4 parts. Figure 3.3 depicts the partition of region $R_{h,a}$, and the 4 parts are denoted by $D_1$, $D_2$, $D_3$ and $D_4$. Thus,

$$P\{\boldsymbol{T} \in R_{h,a}\} = P\{\boldsymbol{T} \in D_1 \cup D_2\} + P\{\boldsymbol{T} \in D_3\} + P\{\boldsymbol{T} \in D_4\}.$$

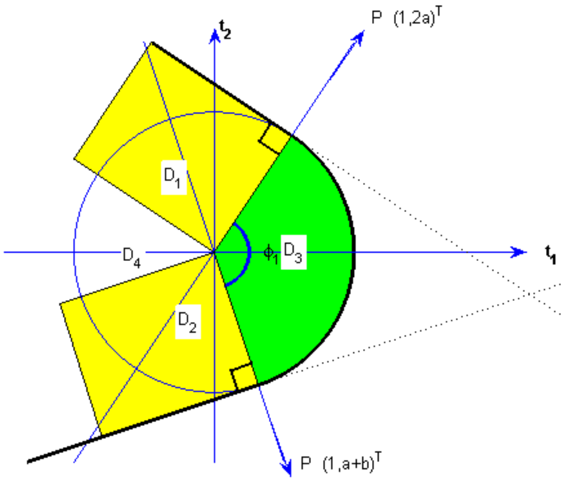



Figure 3.3: The region $R_{h,a} = D_1 \cup D_2 \cup D_3 \cup D_4$

The probability that $\boldsymbol{T}$ lies in the region $D_1 \cup D_2$ is given by

$$
\begin{aligned}
P\{\boldsymbol{T} \in D_1 \cup D_2\} &= P\{|T_2| < c(a),\ \psi_{\boldsymbol{T}} \in (\frac{\pi}{2}, \frac{3\pi}{2})\} \\
&= P\{-c(a) < T_2 < c(a),\ \psi_{\boldsymbol{T}} \in (\frac{\pi}{2}, \frac{3\pi}{2})\} \\
&= \frac{1}{2} P\{-c(a) < T_2 < c(a)\} \\
&= \frac{1}{2} P\{T_2^2 < c^2(a)\} \\
&= \frac{1}{2} F_{1,n-3}(c^2(a)),
\end{aligned}
$$

the probability that $\boldsymbol{T}$ lies in the region $D_3$ is given by

$$
\begin{aligned}
P\{\boldsymbol{T} \in D_3\} &= \frac{\phi_1}{2\pi} P\{R_{\boldsymbol{T}} < c(a)\} \\
&= \frac{\phi_1}{2\pi} P\{R_{\boldsymbol{T}}^2/2 \leq \frac{c^2(a)}{2}\} \\
&= \frac{\phi_1}{2\pi} F_{2,n-3}(\frac{c(a)^2}{2}),
\end{aligned}
$$

and finally the probability that $\boldsymbol{T}$ lies in the region $D_4$ is given by

$$
P\{\boldsymbol{T} \in D_4\} = \frac{\pi - \phi_1}{2\pi}.
$$

Therefore,

$$\begin{aligned} P\{\boldsymbol{T} \in R_{h,a}\} &= P\{\boldsymbol{T} \in D_1 \cup D_2\} + P\{\boldsymbol{T} \in D_3\} + P\{\boldsymbol{T} \in D_4\} \\ &= \frac{\phi_1}{2\pi} F_{2,n-3}(\frac{c(a)^2}{2}) + \frac{\pi - \phi_1}{2\pi} + \frac{1}{2} F_{1,n-3}(c(a)^2), \end{aligned}$$

which implies that

$$\inf_{\boldsymbol{\theta}: k_0 = a} P_{k_0=a}\{\boldsymbol{Y} \in A(a)\} = \frac{\phi_1}{2\pi} F_{2,n-3}(\frac{c(a)^2}{2}) + \frac{\pi - \phi_1}{2\pi} + \frac{1}{2} F_{1,n-3}(c(a)^2).$$

Hence $c(a) > 0$ is solved from

$$\frac{\phi_1}{2\pi} F_{2,n-3}(\frac{c(a)^2}{2}) + \frac{\pi - \phi_1}{2\pi} + \frac{1}{2} F_{1,n-3}(c(a)^2) = 1 - \alpha.$$

**Case II** If $k_0 = b$ is a maximum point, then for any $x \in [a, b)$, it is clear that

$$(\theta_0 + \theta_1 b + \theta_2 b^2) - (\theta_0 + \theta_1 x + \theta_2 x^2) \geq 0,$$

which implies $\theta_1 + \theta_2(b+x) \geq 0$ for all $x \in [a, b)$. Accordingly, we define an acceptance set as

$$\begin{aligned} A(b) &= \{\boldsymbol{Y} : (\hat{\theta}_0 + \hat{\theta}_1 b + \hat{\theta}_2 b^2) - (\hat{\theta}_0 + \hat{\theta}_1 x + \hat{\theta}_2 x^2) \geq -c(b)\hat{\sigma} v(b, x),\ \forall x \in [a, b)\} \\ &= \{\boldsymbol{Y} : \hat{\theta}_1 + \hat{\theta}_2\ (b + x) \geq -c(b)\hat{\sigma} \parallel \mathbf{P}(1, b+x)^T \parallel,\ \forall x \in [a, b)\} \end{aligned} \tag{3.7}$$

where $c(b) > 0$ is a critical constant such that $A(b)$ is a $(1 - \alpha)$ level acceptance set. To determine $c(b)$, note that

$$\begin{aligned} & P_{k_0=b} \quad \{\boldsymbol{Y} \in A(b)\} \\ = \ & P_{k_0=b} \quad \{\hat{\theta}_1 + \hat{\theta}_2\ (b + x) \geq -c(b)\hat{\sigma} \parallel \mathbf{P}(1, b+x)^T \parallel \quad \forall x \in [a, b)\} & (3.8) \\ \geq \ & P_{k_0=b} \quad \{\hat{\theta}_1 + \hat{\theta}_2\ (b + x) \geq -c(b)\hat{\sigma} \parallel \mathbf{P}(1, b+x)^T \parallel + (\theta_1 + \theta_2(b+x))\ \ \forall x \in [a, b)\} \\ & & (3.9) \\ = \ & P_{k_0=b} \quad \{\frac{(1, b+x)(\hat{\theta}_1 - \theta_1, \hat{\theta}_2 - \theta_2)^T}{\hat{\sigma} \parallel \mathbf{P}(1, b+x)^T \parallel} \geq -c(b),\ \forall x \in (a, b]\} \\ = \ & P_{k_0=b} \quad \{\inf_{x \in [a,b)} \frac{[\mathbf{P}(1, b+x)^T]^T [\mathbf{P}^{-1}(\hat{\theta}_1 - \theta_1, \hat{\theta}_2 - \theta_2)^T]}{\hat{\sigma} \parallel \mathbf{P}(1, b+x)^T \parallel} \geq -c(b)\} \\ = \ & \quad P \quad \{\sup_{x \in [a,b)} \frac{[\mathbf{P}(1, b+x)^T]^T\ (-\boldsymbol{T})}{\parallel \mathbf{P}(1, b+x)^T \parallel} \leq c(b)\} & (3.10) \\ = \ & \quad P \quad \{(-\boldsymbol{T}) \in R_{h,b}\}, \end{aligned}$$

where Equation (3.8) follows directly from Equation (3.7), Equation (3.9) follows from the fact that $\theta_1 + \theta_2(b + x) \geq 0$ for $\forall x \in [a, b)$, and

$$R_{h,b} = \bigcap_{x\in[a,b)} R_{h,b}(x),$$

where

$$R_{h,b}(x) = \{\boldsymbol{T} : \frac{\{\mathbf{P}(1, b + x)^T\}^T \boldsymbol{T}}{\| \mathbf{P}(1, b + x)^T \|} \leq c(b)\}.$$

Note that

$$\inf_{\boldsymbol{\theta}:k_0=b} P_{k_0=b}\{\boldsymbol{Y} \in A(b)\} = P\{(-\boldsymbol{T}) \in R_{h,b}\}, \tag{3.11}$$

with the infimum being attained at $\theta_1 = \theta_2 = 0$.

Now we consider the computation of $P\{(-\boldsymbol{T}) \in R_{h,b}\}$. Since $\boldsymbol{T} \sim \boldsymbol{T}_{2,n-3}$ and $(-\boldsymbol{T}) \sim \boldsymbol{T}_{2,n-3}$, the probability can be written as

$$P\{(-\boldsymbol{T}) \in R_{h,b}\} = P\{\boldsymbol{T} \in R_{h,b}\}.$$

By comparing this definition of $R_{h,b}$ with $R_{h,a}$, we note that the shape of $R_{h,b}$ is similar to that of $R_{h,a}$ but with the angle $\phi_1$ between the two stripes $D_1$ and $D_2$ in Figure 3.3 replaced by $\phi_2$ where $\phi_2$ is the angle between $\mathbf{P}(1, a+b)^T$ and $\mathbf{P}(1, 2b)^T$ and so

$$\cos\phi_2 = \frac{[\mathbf{P}(1, a + b)^T]^T [\mathbf{P}(1, 2b)^T]}{\| \mathbf{P}(1, a + b)^T \| \ \| \mathbf{P}(1, 2b)^T \|}.$$

Using a similar argument to the one used to compute $P\{\boldsymbol{T} \in R_{h,a}\}$, we have

$$P\{\boldsymbol{T} \in R_{h,b}\} = \frac{\phi_2}{2\pi} F_{2,n-3}(\frac{c(b)^2}{2}) + \frac{\pi - \phi_2}{2\pi} + \frac{1}{2} F_{1,n-3}(c(b)^2).$$

Hence, the critical value $c(b)$ is the positive solution of

$$\frac{\phi_2}{2\pi} F_{2,n-3}(\frac{c(b)^2}{2}) + \frac{\pi - \phi_2}{2\pi} + \frac{1}{2} F_{1,n-3}(c(b)^2) = 1 - \alpha.$$

**Case III** Finally, if $k_0 = s$, $s \in (a, b)$, is a maximum point, then for any $x \in [a, s) \cup (s, b]$, it is clear that

$$(\theta_0 + \theta_1 s + \theta_2 s^2) - (\theta_0 + \theta_1 x + \theta_2 x^2) \geq 0,$$

which implies $\theta_1 + \theta_2(s+x) \geq 0$ for $x \in [a,s)$, and $\theta_1 + \theta_2(s+x) \leq 0$ for $x \in (s,b]$. Accordingly, for each $s \in (a,b)$, we define an acceptance set

$$\begin{aligned} A(s) &= \{\boldsymbol{Y} : (\hat{\theta_0} + \hat{\theta_1} s + \hat{\theta_2} s^2) - (\hat{\theta_0} + \hat{\theta_1} x + \hat{\theta_2} x^2) \geq -c(s)\hat{\sigma} v(s,x),\ \forall x \in [a,s) \cup (s,b]\} \\ &= \{\boldsymbol{Y} : (\hat{\theta_1} + \hat{\theta_2}(s+x))(s-x) \geq -c(s)\hat{\sigma} v(s,x),\ \forall x \in [a,s) \cup (s,b]\} \\ &= \{\boldsymbol{Y} : \hat{\theta_1} + \hat{\theta_2}\ (s+x) \geq -c(s)\hat{\sigma} \parallel \mathbf{P}(1, s+x)^T \parallel,\ \forall x \in [a,s), \\ &\qquad \hat{\theta_1} + \hat{\theta_2}\ (s+x) \leq c(s)\hat{\sigma} \parallel \mathbf{P}(1, s+x)^T \parallel,\ \forall x \in (s,b]\} \end{aligned} \tag{3.12}$$

where $c(s) > 0$ is a critical constant such that $A(s)$ is a $(1-\alpha)$ level acceptance set.

To determine $c(s)$, note that

$$\begin{aligned} & P_{k_0=s} \quad \{\boldsymbol{Y} \in A(s)\} \\ = \ & P_{k_0=s} \quad \{\hat{\theta_1} + \hat{\theta_2}\ (s+x) \geq -c(s)\hat{\sigma} \parallel \mathbf{P}(1, s+x)^T \parallel,\ \forall x \in [a,s), \\ & \qquad \hat{\theta_1} + \hat{\theta_2}\ (s+x) \leq c(s)\hat{\sigma} \parallel \mathbf{P}(1, s+x)^T \parallel,\ \forall x \in (s,b]\} \qquad (3.13) \\ \geq \ & P_{k_0=s} \quad \{\hat{\theta_1} + \hat{\theta_2}(s+x) \geq -c(s)\hat{\sigma} \parallel \mathbf{P}(1, s+x)^T \parallel + (\theta_1 + \theta_2(s+x)),\ \forall x \in [a,s), \\ & \qquad \hat{\theta_1} + \hat{\theta_2}(s+x) \leq c(s)\hat{\sigma} \parallel \mathbf{P}(1, s+x)^T \parallel + (\theta_1 + \theta_2(s+x)),\ \forall x \in (s,b]\} \\ & \qquad (3.14) \\ = \ & P_{k_0=s} \quad \{\inf_{x \in [a,s)} \frac{[\mathbf{P}(1, s+x)^T][\mathbf{P}^{-1}(\hat{\theta_1} - \theta_1, \hat{\theta_2} - \theta_2)^T]}{\hat{\sigma} \parallel \mathbf{P}(1, s+x)^T \parallel} \geq -c(s), \\ & \qquad \sup_{x \in (s,b]} \frac{[\mathbf{P}(1, s+x)^T]^T[\mathbf{P}^{-1}(\hat{\theta_1} - \theta_1, \hat{\theta_2} - \theta_2)^T]}{\hat{\sigma} \parallel \mathbf{P}(1, s+x)^T \parallel} \leq c(s)\} \\ = \ & P \quad \{\inf_{x \in [a,s)} \frac{[\mathbf{P}(1, s+x)^T]^T\ \boldsymbol{T}}{\parallel \mathbf{P}(1, s+x)^T \parallel} \geq -c(s),\ \ \sup_{x \in (s,b]} \frac{[\mathbf{P}(1, s+x)^T]^T \boldsymbol{T}}{\parallel \mathbf{P}(1, s+x)^T \parallel} \leq c(s)\} \qquad (3.15) \\ = \ & P \quad \{\boldsymbol{T} \in R_h(s)\} \qquad (3.16) \end{aligned}$$

where Equation (3.13) follows directly from Equation (3.12), Equation (3.14) follows from the fact that $\theta_1 + \theta_2(s+x) \geq 0$ for $x \in [a,s)$, and $\theta_1 + \theta_2(s+x) \leq 0$ for $x \in (s,b]$. The region $R_h(s) \subset R^2$ in (3.16) is given by

$$R_h(s) = \bigcap_{x \in [a,s) \cup (s,b]} R_h(s,x) \tag{3.17}$$

where

$$R_h(s,x) = \left\{ \boldsymbol{T} : \frac{[\mathbf{P}(1, x+s)^T]^T \boldsymbol{T}}{\parallel \mathbf{P}(1, x+s)^T \parallel} > -c(s) \right\}$$

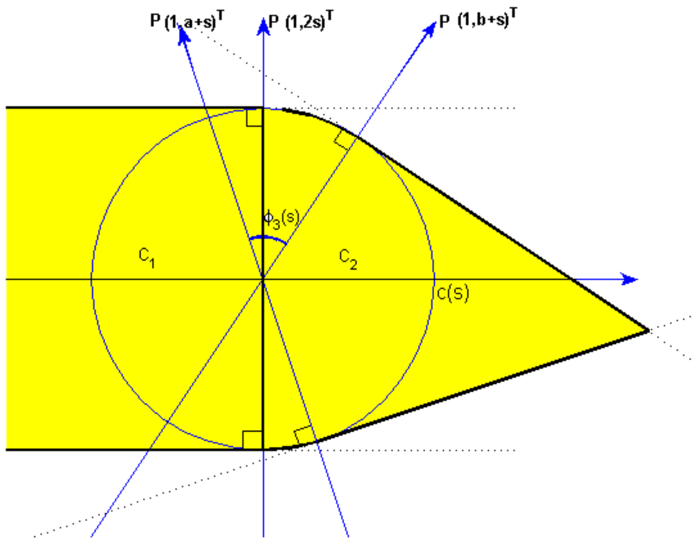



Figure 3.4: The region $R_h(s)$

if $x \in [a, s)$, and

$$R_h(s,x) = \left\{ \boldsymbol{T} : \frac{[\mathbf{P}(1, x+s)^T]^T \boldsymbol{T}}{\| \mathbf{P}(1, x+s)^T \|} < c(s) \right\}$$

if $x \in (s, b]$. Note that for each $s \in (a, b)$,

$$\inf_{\boldsymbol{\theta}:k_0=s} P_{k_0=s}\{\boldsymbol{Y} \in A(s)\} = P\{\boldsymbol{T} \in R_h(s)\}, \tag{3.18}$$

with the infimum being attained at $\theta_1 = \theta_2 = 0$.

Now we consider the computation of $P\{\boldsymbol{T} \in R_h(s)\}$. Note that the set $R_h(s,x)$ is given by the region bounded by a straight line that is perpendicular to $\mathbf{P}(1, s+x)^T$ and $c(s)$ distance away from the origin. Hence $R_h(s)$ is the intersection of all such regions with $x \in [a, s) \cup (s, b]$. Figure 3.4 depicts the region $R_h(s)$, which is bounded by two parallel straight lines, two parts of the circle and another two straight lines. Angle $\phi_3(s)$ is the angle between $\mathbf{P}(1, a+s)^T$ and $\mathbf{P}(1, b+s)^T$, and so

$$\cos \phi_3(s) = \frac{[\mathbf{P}(1, a+s)^T]^T [\mathbf{P}(1, s+b)^T]}{\| \mathbf{P}(1, a+s)^T \| \ \| \mathbf{P}(1, s+b)^T \|}.$$

The region $R_h(s)$ can be partitioned in to $C_1$ and $C_2$ , therefore

$$P\{\boldsymbol{T} \in R_h(s)\} = P\{\boldsymbol{T} \in C_1\} + P\{\boldsymbol{T} \in C_2\}.$$

Using the method of Wynn and Bloomfield (1971) following Liu (2010, pp.18-24):

$$\begin{aligned}P\{\boldsymbol{T} \in C_1\} &= \frac{1}{2}P\{|T_2| < c(s)\} \\ &= \frac{1}{2}P\{T_2^2 < c^2(s)\} \\ &= \frac{1}{2}F_{1,n-3}(c^2(s)).\end{aligned}$$

Now we consider the region $C_2$. The region $C_2$ can be further partitioned into a half disc and the remaining region. We rotate the remaining region around the disc to a position so that $C_2$ is symmetric about the $t_1$-axis. Figure 3.5 shows the region $C_2$ after rotation. The half disc is denoted as $D_1$, and the remaining region (the shaded region) is denoted as $D_2$. The probability that $\boldsymbol{T}$ lies in the region $D_1$ is given by

$$\begin{aligned}&P\{\boldsymbol{T} \in D_1\} \\ =\ & P\left\{R_{\boldsymbol{T}} < c(s),\ \psi_{\boldsymbol{T}} \in [0, \frac{\pi}{2}) \cup (\frac{3\pi}{2}, 2\pi)\right\} \\ =\ & \frac{1}{2}P\{R_{\boldsymbol{T}} < c(s)\} \\ =\ & \frac{1}{2}P\left\{\frac{\parallel \boldsymbol{T} \parallel^2}{2} \leq \frac{c^2(s)}{2}\right\} \\ =\ & \frac{1}{2}F_{2,n-3}\left(\frac{c^2(s)}{2}\right).\end{aligned}$$

Let

$$\phi(s) = \frac{\pi - \phi_3(s)}{2},$$

then the probability that $\boldsymbol{T}$ lies in the region $D_2$ is given by

$$\begin{aligned}&P\{\boldsymbol{T} \in D_2\} \\ =\ & 2P\{R_{\boldsymbol{T}} > c(s),\ \psi_{\boldsymbol{T}} \in [0, \phi(s)),\ (\cos(\phi(s)), \sin(\phi(s)))\boldsymbol{T} < c(s)\} \\ =\ & 2P\left\{\psi_{\boldsymbol{T}} \in [0, \phi(s)),\ c(s) < R_{\boldsymbol{T}} < \frac{c(s)}{\cos(\phi(s) - \psi_{\boldsymbol{T}})}\right\} \\ =\ & \frac{\phi(s)}{\pi}\left(1 + \frac{c^2(s)}{n-3}\right)^{\frac{-(n-3)}{2}} - \frac{1}{\pi}\int_0^{\phi(s)}\left(1 + \frac{c^2(s)}{(n-3)\cos^2(\phi(s) - \psi)}\right)^{\frac{-(n-3)}{2}} d\psi\end{aligned}$$

where the last equation above follows directly from the cumulative density function

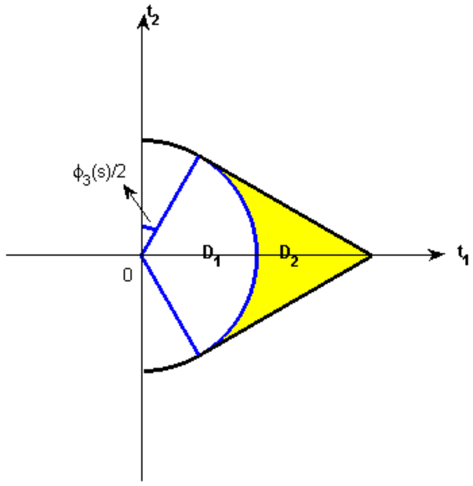



Figure 3.5: The region $C_2$.

of $R_{\boldsymbol{T}}$. Then

$$\begin{aligned} P\{\boldsymbol{T} \in C_2\} &= \frac{1}{2}F_{2,n-3}\left(\frac{c^2(s)}{2}\right) + \frac{\phi(s)}{\pi}\left(1+\frac{c^2(s)}{n-3}\right)^{\frac{-(n-3)}{2}} \\ &\quad -\frac{1}{\pi}\int_0^{\phi(s)}\left(1+\frac{c^2(s)}{(n-3)\cos^2(\phi(s)-\psi)}\right)^{\frac{-(n-3)}{2}} d\psi. \end{aligned}$$

Hence, the probability $P\{\boldsymbol{T} \in R_h(s)\}$ is equal to

$$\begin{aligned} &\frac{1}{2}F_{1,n-3}(c^2(s)) + \frac{1}{2}F_{2,n-3}\left(\frac{c^2(s)}{2}\right) + \frac{\phi(s)}{\pi}\left(1+\frac{c^2(s)}{n-3}\right)^{\frac{-(n-3)}{2}} \\ &-\frac{1}{\pi}\int_0^{\phi(s)}\left(1+\frac{c^2(s)}{(n-3)\cos^2(\phi(s)-\psi)}\right)^{\frac{-(n-3)}{2}} d\psi. \end{aligned}$$

Therefore, the critical value $c(s)$ is such that the last expression is equal to $(1-\alpha)$.

Having found $c(k_0)$ for each $k_0 \in [a, b]$, we summarize our $(1-\alpha)$ level acceptance set as

$$A(k_0) = \left\{\boldsymbol{Y} : (\hat{\theta}_1 + \hat{\theta}_2(k_0 + x))(k_0 - x) \geq -c(k_0)\hat{\sigma}v(k_0, x), \forall x \in [a, b] \setminus k_0\right\}. \tag{3.19}$$

Then according to Neyman's Theorem, a $(1-\alpha)$ level confidence set for a maximum point, k, based on $\boldsymbol{Y}$ is given by

$$\begin{aligned}\mathbf{C}_E(\boldsymbol{Y}) &= \{k_0 \in [a,b] : \boldsymbol{Y} \in A(k_0)\} \\ &= \left\{k_0 \in [a,b] : (\hat{\theta}_1 + \hat{\theta}_2(k_0+x))(k_0 - x) \geq -c(k_0)\hat{\sigma}v(k_0,x), \forall x \in [a,b] \setminus k_0\right\}.\end{aligned} \tag{3.20}$$

In other words, a point $k_0 \in [a,b]$ is in the set $\mathbf{C}_E(\boldsymbol{Y})$ if and only if for any $x \in [a,b] \setminus k_0$, we have

$$(\hat{\theta}_1 + \hat{\theta}_2(k_0+x))(k_0 - x) \geq -c(k_0)\hat{\sigma}v(k_0,x).$$

#### 3.1.2.2 Computation

In this section, we first present the direct approach of computing the confidence set in practice. Then to reduce the complexity of computation, we present a p-value method which avoids the computation of critical values.

**Direct approach.** After the critical values are computed, we construct the confidence set directly from Equation (3.20) by testing whether a point $k_0$ in $[a, b]$ is in the confidence set. Let

$$\hat{h}(z) = \frac{\hat{\theta}_1 + \hat{\theta}_2 z}{\hat{\sigma} \parallel \mathbf{P}(1, z)^T \parallel}, \tag{3.21}$$

then the confidence set in Equation (3.20) can be written as

$$\mathbf{C}_E(\boldsymbol{Y}) = \{k_0 \in [a, b] : \hat{h}(k_0 + x)\frac{k_0 - x}{|k_0 - x|} \geq -c(k_0),\ \forall x \in [a, b] \setminus k_0\}.$$

Since $\hat{h}(z)$ is a continuous function, the inequality in $\mathbf{C}_E(\boldsymbol{Y})$ implies

$$\begin{aligned}
a \in \mathbf{C}_E(\boldsymbol{Y}) \quad &\Leftrightarrow \quad \sup_{x \in [a,b]} \hat{h}(a + x) \leq c(a), \\
b \in \mathbf{C}_E(\boldsymbol{Y}) \quad &\Leftrightarrow \quad \inf_{x \in [a,b]} \hat{h}(b + x) \geq -c(b), \\
s \in \mathbf{C}_E(\boldsymbol{Y}) \quad &\Leftrightarrow \quad \begin{cases} \inf_{x \in [a,s]} \hat{h}(s + x) \geq -c(s) \\ \sup_{x \in [s,b]} \hat{h}(s + x) \leq c(s) \end{cases}.
\end{aligned}$$

Note that from the definition of $\hat{h}(z)$ in Equation (3.21) the function $\hat{h}(z)$ has a single stationary point $z_0$ over $(-\infty, \infty)$ which is given by

$$z_0 = \frac{v_{23}\hat{\theta}_1 - v_{22}\hat{\theta}_2}{v_{23}\hat{\theta}_2 - v_{33}\hat{\theta}_1}.$$

Therefore, we have the following results.
If $(z_0 - a) \notin [a, b]$, then

$$a \in \mathbf{C}_E(\boldsymbol{Y}) \Leftrightarrow \max(\hat{h}(a + a), \hat{h}(b + a)) \leq c(a); \tag{3.22}$$

alternatively, if $(z_0 - a) \in [a, b]$, then

$$a \in \mathbf{C}_E(\boldsymbol{Y}) \Leftrightarrow \max(\hat{h}(a + a), \hat{h}(b + a), \hat{h}(z_0)) \leq c(a). \tag{3.23}$$

Similarly, if $(z_0 - b) \notin [a, b]$, then

$$b \in \mathbf{C}_E(\boldsymbol{Y}) \Leftrightarrow \min(\hat{h}(a+b), \hat{h}(b+b)) \geq -c(b); \tag{3.24}$$

alternatively, if $(z_0 - b) \in [a, b]$, then

$$b \in \mathbf{C}_E(\boldsymbol{Y}) \Leftrightarrow \min(\hat{h}(a+b), \hat{h}(b+b), \hat{h}(z_0)) \geq -c(b). \tag{3.25}$$

Finally, for $s \in (a, b)$, if $(z_0 - s) \notin [a, s] \cup [s, b]$, then

$$s \in \mathbf{C}_E(\boldsymbol{Y}) \Leftrightarrow \begin{cases} \min(\hat{h}(a+s), \hat{h}(2s)) \geq -c(s) \\ \max(\hat{h}(b+s), \hat{h}(2s)) \leq c(s) \end{cases}; \tag{3.26}$$

alternatively, if $(z_0 - s) \in [a, s]$, then

$$s \in \mathbf{C}_E(\boldsymbol{Y}) \Leftrightarrow \begin{cases} \min(\hat{h}(a+s), \hat{h}(2s), \hat{h}(z_0)) \geq -c(s) \\ \max(\hat{h}(b+s), \hat{h}(2s)) \leq c(s) \end{cases}; \tag{3.27}$$

otherwise, if $(z_0 - s) \in [s, b]$, then

$$s \in \mathbf{C}_E(\boldsymbol{Y}) \Leftrightarrow \begin{cases} \min(\hat{h}(a+s), \hat{h}(2s)) \geq -c(s) \\ \max(\hat{h}(b+s), \hat{h}(2s), \hat{h}(z_0)) \leq c(s) \end{cases}. \tag{3.28}$$

Theoretically, we need to check each point $k_0 \in [a, b]$ whether it is in the confidence set by verifying the above inequalities. However, the interval $[a, b]$ contains infinite points and thus we can not check each point $k_0 \in [a, b]$ to determine whether it is in the confidence set. Hence, we choose a finite grid of points in the interval $[a, b]$ with resolution $d$, that is, $\{a = s_1, s_2, \cdots, s_J = b\}$ with $s_i - s_{i-1} = d$. If $d$ is small, then the grid of points can give a fine approximation to the set $[a, b]$. We only check each point in the grid to determine whether it is in the confidence set.

In practice, it is time consuming to construct the confidence set in this way, since the computation of the critical values takes time, especially for $c(s)$ with $s \in (a, b)$ which involves solving an equation of integration. To reduce the computation burden, we use conservative critical constants to quickly narrow down the possible points that could be in the set $\mathbf{C}_E(\boldsymbol{Y})$ and then compute the critical value $c(k_0)$ for only the remaining points.

**Conservative confidence set method.** Recall that in Section 3.1.2.1, we have Equation (3.5) equal to $(1-\alpha)$, that is,

$$P\left\{\sup_{x\in(a,b]}\frac{[\mathbf{P}(1,a+x)^T]^T\boldsymbol{T}}{\|\mathbf{P}(1,a+x)^T\|}\le c(a)\right\}=1-\alpha,$$

where $\boldsymbol{T}$ is defined to have a standard bivariate t-distribution with degrees of freedom $(n-3)$. Note that

$$\frac{\|\boldsymbol{T}\|^2}{2}\sim F_{2,n-3},$$

where $F_{2,n-3}$ is the standard $F$-distribution with degrees of freedom 2 and $(n-3)$. Let $c_1=\sqrt{2f^{\alpha}_{2,n-3}}$ where $f^{\alpha}_{2,n-3}$ is the upper $\alpha$ point of the distribution $F_{2,n-3}$, then we have

$$P\{\|\boldsymbol{T}\|\le c_1\}=1-\alpha. \tag{3.29}$$

Equation (3.29) can be written as

$$P\left\{\sup_{\boldsymbol{\rho}\in R^2}\frac{|\boldsymbol{\rho}\boldsymbol{T}|}{\|\boldsymbol{\rho}\|}\le c_1\right\}=1-\alpha,$$

hence

$$P\left\{\sup_{\boldsymbol{\rho}\in R^2}\frac{\boldsymbol{\rho}\boldsymbol{T}}{\|\boldsymbol{\rho}\|}\le c_1\right\}\ge 1-\alpha.$$

Since $\{\mathbf{P}(1,a+x)^T : x\in[a,b]\}\subset R^2$, comparing this equation with Equation (3.5), we conclude that $c(a)<c_1$.

In a similarly way, we have $c(b)<c_1$.

For $s\in(a,b)$, since the probability in Equation (3.15) is equal to $(1-\alpha)$, that is,

$$P\left\{\sup_{x\in(s,b]}\frac{[\mathbf{P}(1,s+x)^T]^T\,\boldsymbol{T}}{\|\mathbf{P}(1,s+x)^T\|}\le c(s),\quad \inf_{x\in[a,s)}\frac{[\mathbf{P}(1,s+x)^T]^T\,\boldsymbol{T}}{\|\mathbf{P}(1,s+x)^T\|}\ge -c(s)\right\}=1-\alpha.$$

Note that Equation (3.29) can be expressed in the form

$$P\left\{\sup_{\boldsymbol{\rho}\in R^2}\frac{\boldsymbol{\rho}\boldsymbol{T}}{\|\boldsymbol{\rho}\|}\le c_1,\ \inf_{\boldsymbol{\rho}\in R^2}\frac{\boldsymbol{\rho}\boldsymbol{T}}{\|\boldsymbol{\rho}\|}\ge -c_1\right\}=1-\alpha.$$

Comparing the above two equations, we therefore conclude

$$c(s)\le c_1$$

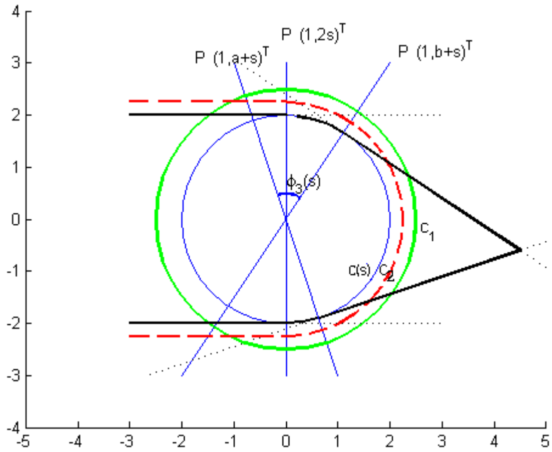



Figure 3.6: The critical values $c(s)$, $c_1$ and $c_2$.

for all $s \in (a, b)$.

Therefore, we have proved $c(k_0) < c_1$ for any $k_0 \in [a, b]$, and hence a conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ for a maximum point is given by substituting $c_1 = \sqrt{2f^{\alpha}_{2,n-3}}$ for $c(k_0)$ in Equation (3.20):

$$\begin{aligned} \mathbf{C}_0(\boldsymbol{Y}) &= \left\{k_0 \in [a,b] : (\hat{\theta}_1 + \hat{\theta}_2(k_0 + x))(k_0 - x) \geq -c_1 \hat{\sigma} v(k_0, x), \forall x \in [a,b] \setminus k_0 \right\} \\ &= \left\{k_0 \in [a,b] : \hat{h}(k_0 + x)\frac{k_0 - x}{|k_0 - x|} \geq -\sqrt{2f^{\alpha}_{2,n-3}},\ \forall x \in [a,b] \setminus k_0 \right\}. \end{aligned}$$

Yet following from Hochberg and Quade (1975) or Liu (2010, pp68), another conservative critical value $c_2$ is solved from

$$\frac{1}{2}F_{2,n-3}\left(\frac{c_2^2}{2}\right) + \frac{1}{2}F_{1,n-2}\left(c_2^2\right) = 1 - \alpha,$$

and is less conservative compared with $c_1$.

Figure 3.6 presents the relationships between the critical value $c(s)$ and the conservative critical values $c_1$ and $c_2$. The bold curve bounds the region $R_h(s)$, which is the same as shown in Figure 3.4, and the outer circle bounds the region $\{\boldsymbol{T} : \| \boldsymbol{T} \| \leq \sqrt{2f^{\alpha}_{2,n-3}}\}$. The region bounded by the dashed curve is associated with $c_2$ and is the same as region $R_{h,a}$ if the angle $\phi_1$ in $R_{h,a}$ is equal to $\pi$. This critical value is less conservative than $c_1$, thus the associated conservative confidence set contains less grid points than the one associated with $c_1$. Therefore to construct a confidence set, it is more efficient to check the grid points in the conservative confidence set with critical value $c_2$ than the one with $c_1$. It is not important which conservative critical value we choose to use in the quadratic case, because the exact critical values are not difficult to compute. But in the general polynomial case when the exact critical values need to be approximated using a large number of simulation, it saves a lot of time if we construct the confidence set by checking the grid points in the conservative confidence set with $c_2$ instead of $c_1$.

We compute this conservative confidence set directly as we did in the Direct Method, by substituting the conservative critical value (take $c_1$ for example) for $c(k_0)$, $k_0 \in [a, b]$, in Equation (3.22)-(3.28). Therefore, if $(z_0 - a) \notin [a, b]$, then

$$a \in \mathbf{C}_0(\boldsymbol{Y}) \Leftrightarrow \max(\hat{h}(2a), \hat{h}(b + a)) \leq c_1; \tag{3.30}$$

otherwise, if $(z_0 - a) \in [a, b]$, then

$$a \in \mathbf{C}_0(\boldsymbol{Y}) \Leftrightarrow \max(\hat{h}(2a), \hat{h}(b + a), \hat{h}(z_0)) \leq c_1. \tag{3.31}$$

Similarly, if $(z_0 - b) \notin [a, b]$, then

$$b \in \mathbf{C}_0(\boldsymbol{Y}) \Leftrightarrow \min(\hat{h}(a + b), \hat{h}(2b)) \geq -c_1; \tag{3.32}$$

otherwise, if $(z_0 - b) \in [a, b]$, then

$$b \in \mathbf{C}_0(\boldsymbol{Y}) \Leftrightarrow \min(\hat{h}(a + b), \hat{h}(2b), \hat{h}(z_0)) \geq -c_1. \tag{3.33}$$

Finally, for $s \in (a, b)$, if $(z_0 - s) \notin [a, s] \cup [s, b]$, then

$$s \in \mathbf{C}_0(\boldsymbol{Y}) \Leftrightarrow \left\{ \begin{array}{c} \min(\hat{h}(a + s), \hat{h}(2s)) \geq -c_1 \\ \max(\hat{h}(b + s), \hat{h}(2s)) \leq c_1 \end{array} \right. ; \tag{3.34}$$

alternatively, if $(z_0 - s) \in [a, s]$, then

$$s \in \mathbf{C}_0(\boldsymbol{Y}) \Leftrightarrow \left\{ \begin{array}{c} \min(\hat{h}(a + s), \hat{h}(2s), \hat{h}(z_0)) \geq -c_1 \\ \max(\hat{h}(b + s), \hat{h}(2s)) \leq c_1 \end{array} \right. ; \tag{3.35}$$

otherwise, if $(z_0 - s) \in [s, b]$, then

$$s \in \mathbf{C}_0(\boldsymbol{Y}) \Leftrightarrow \begin{cases} \min(\hat{h}(a+s), \hat{h}(2s)) \geq -c_1 \\ \max(\hat{h}(b+s), \hat{h}(2s), \hat{h}(z_0)) \leq c_1 \end{cases}. \tag{3.36}$$

In practice, to construct the confidence set, we still check each grid point instead of each point in the interval $[a, b]$. However, as been pointed out earlier, this conservative confidence set is not the ultimate result we want, but a way to narrow down the possible range of points that could be in the set $\mathbf{C}_E(\boldsymbol{Y})$. Then, we construct the exact $(1 - \alpha)$ level confidence set $\mathbf{C}_E(\boldsymbol{Y})$ by checking each grid point in the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$. Next, we will illustrate how to use the p-value method to construct the acceptance sets, which does not involve the computation of critical values.

**The p-value method.** Since the construction of acceptance sets using critical values directly is inefficient, we use the p-value method instead. A p-value is the probability of having a value of the test statistic that is even more extreme than the one based on the observed data, assuming the null hypothesis is true. When testing the null hypothesis, we reject $H_0$ if and only if the p-value is less than the significant level $\alpha$.

Let $\hat{\theta}_1^*$, $\hat{\theta}_2^*$, and $\hat{\sigma}^*$ be the usual estimates of $\theta_1$, $\theta_2$, and $\sigma$ based on the observation $\boldsymbol{Y}^*$. Note that $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\sigma}$ are random variables based on random observation $\boldsymbol{Y}$, while $\hat{\theta}_1^*$, $\hat{\theta}_2^*$, and $\hat{\sigma}^*$ are estimated based on the observation $\boldsymbol{Y}^*$. We distinguish these concepts only for introducing p-values in this section and Section 3.1.3.3.
Define

$$\hat{h}^*(z) = \frac{\hat{\theta}_1^* + \hat{\theta}_2^* z}{\hat{\sigma}^* \parallel \mathbf{P}(1, z)^T \parallel}.$$

Next, we discuss the computation of the p-value separately for each of the three cases: $k_0 = a$, $k_0 = b$ and $k_0 \in (a, b)$.

**Case I** Recall that in Section 3.1.2.1, our $(1 - \alpha)$ level acceptance set $A(a)$ is given by

$$\begin{aligned} A(a) &= \{\boldsymbol{Y} : (\hat{\theta_1} + \hat{\theta_2}(a+x))(a-x) \geq -c(a)\hat{\sigma}v(a,x),\ \forall x \in (a,b]\} \\ &= \{\boldsymbol{Y} : \sup_{x\in(a,b]} \hat{h}(a+x) \leq c(a)\}. \end{aligned}$$

We define

$$W_a(\boldsymbol{Y}) = \sup_{x\in(a,b]} \hat{h}(a+x)$$

and

$$w_a = W_a(\boldsymbol{Y}^*) = \sup_{x\in(a,b]} \hat{h}^*(a+x).$$

Note that $W_a(\boldsymbol{Y})$ is a random variable, but $w_a$ is the observed value of $W_a(\boldsymbol{Y})$ based on $\boldsymbol{Y}^*$. Define

$$\begin{aligned} p_a &= 1 - \inf_{\boldsymbol{\theta}:k_0=a} P\{W_a(\boldsymbol{Y}) \leq w_a\} \\ &= \sup_{\boldsymbol{\theta}:k_0=a} (1 - P\{W_a(\boldsymbol{Y}) \leq w_a\}) \\ &= \sup_{\boldsymbol{\theta}:k_0=a} P\{W_a(\boldsymbol{Y}) > w_a\}. \end{aligned}$$

As

$$\begin{aligned} a \in \mathbf{C}_E(\boldsymbol{Y}^*) &\Leftrightarrow \boldsymbol{Y}^* \in A(a) \\ &\Leftrightarrow w_a \leq c(a) \\ &\Leftrightarrow \inf_{\boldsymbol{\theta}:k_0=a} P\{W_a(\boldsymbol{Y}) \leq w_a\} \leq 1-\alpha, \end{aligned}$$

we have

$$\begin{aligned} a \notin \mathbf{C}_E(\boldsymbol{Y}^*) &\Leftrightarrow \inf_{\boldsymbol{\theta}:k_0=a} P\{W_a(\boldsymbol{Y}) \leq w_a\} > 1-\alpha \\ &\Leftrightarrow \sup_{\boldsymbol{\theta}:k_0=a} P\{W_a(\boldsymbol{Y}) > w_a\} < \alpha \\ &\Leftrightarrow p_a < \alpha. \end{aligned}$$

It is clear that $p_a$ is a p-value for testing $H_0 : k_0 = a$.

If $w_a \geq 0$, then following the mathematical derivation in Section 3.1.2.1, we have

$$\begin{aligned} 1 - p_a &= \inf_{\boldsymbol{\theta}:k_0=a} P_{k_0=a}\{W_a(\boldsymbol{Y}) \leq w_a\} \\ &= P_{k_0=a}\left\{ \sup_{x\in(a,b]} \frac{[\mathbf{P}(1,a+x)^T]^T\boldsymbol{T}}{\parallel \mathbf{P}(1,a+x)^T \parallel} \leq w_a \right\} \\ &= \frac{\phi_1}{2\pi}F_{2,n-3}(\frac{w_a^2}{2}) + \frac{1}{2}F_{1,n-3}(w_a^2) + \frac{\pi-\phi_1}{2\pi} \end{aligned}$$

where $\boldsymbol{T} \sim \boldsymbol{T}_{2,\nu}$, that is,

$$p_a = 1 - \frac{\phi_1}{2\pi} F_{2,n-3}(\frac{w_a^2}{2}) - \frac{1}{2} F_{1,n-3}(w_a^2) - \frac{\pi - \phi_1}{2\pi}. \tag{3.37}$$

We reject $H_0 : k = a$ if and only if $p_a < \alpha$.

Otherwise, if $w_a < 0$, then $w_a = - \mid w_a \mid$, we have

$$\begin{aligned} 1 - p_a &= \inf_{\boldsymbol{\theta}: k_0 = a} P_{k_0=a}\{W_a(\boldsymbol{Y}) \leq - \mid w_a \mid\} \\ &= P\left\{ \sup_{x \in (a,b]} \frac{[\mathbf{P}(1, a+x)^T]^T \boldsymbol{T}}{\parallel \mathbf{P}(1, a+x)^T \parallel} \leq - \mid w_a \mid \right\}. \end{aligned} \tag{3.38}$$

Since

$$\begin{aligned} p_a &= 1 - P\left\{ \sup_{x \in (a,b]} \frac{[\mathbf{P}(1, a+x)^T]^T \boldsymbol{T}}{\parallel \mathbf{P}(1, a+x)^T \parallel} \leq - \mid w_a \mid \right\} \\ &\geq 1 - P\left\{ \frac{[\mathbf{P}(1, a+b)^T]^T \boldsymbol{T}}{\parallel \mathbf{P}(1, a+b)^T \parallel} \leq - \mid w_a \mid \right\} \\ &\geq 1 - P\left\{ \frac{[\mathbf{P}(1, a+b)^T]^T \boldsymbol{T}}{\parallel \mathbf{P}(1, a+b)^T \parallel} \leq 0 \right\} \\ &= 1 - 0.5 \\ &= 0.5 \end{aligned}$$

and $\alpha$ is much smaller than 0.5, we can not reject $H_0 : k_0 = a$ when $w_a < 0$.

**Case II** Similarly, by using previous notations, we have

$$\begin{aligned} A(b) &= \{\boldsymbol{Y} : (\hat{\theta}_0 + \hat{\theta}_1 b + \hat{\theta}_2 b^2) - (\hat{\theta}_0 + \hat{\theta}_1 x + \hat{\theta}_2 x^2) \geq -c(b)\hat{\sigma} v(b,x), \ \forall x \in [a, b)\} \\ &= \{\boldsymbol{Y} : \inf_{x \in [a,b)} \hat{h}(b+x) \geq -c(b)\}. \end{aligned}$$

Define

$$W_b(\boldsymbol{Y}) = \inf_{x \in [a,b)} \hat{h}(b+x), \quad w_b = W_b(\boldsymbol{Y}^*) = \inf_{x \in [a,b)} \hat{h}^*(b+x),$$

and a p-value

$$p_b \;=\; 1 - \inf_{\boldsymbol{\theta}: k_0 = b} P_{k_0=b}\{W_b(\boldsymbol{Y}) \geq w_b\}$$

for testing $H_0 : k_0 = b$.

If $w_b \leq 0$, by using a similar argument in computing $p_a$, we have

$$1 - p_b = \frac{\phi_2}{2\pi} F_{2,n-3}\left(\frac{w_b^2}{2}\right) + \frac{1}{2} F_{1,n-3}(w_b^2) + \frac{\pi - \phi_2}{2\pi},$$

thus

$$p_b = 1 - \frac{\phi_2}{2\pi} F_{2,n-3}\left(\frac{w_b^2}{2}\right) + \frac{1}{2} F_{1,n-3}(w_b^2) + \frac{\pi - \phi_2}{2\pi}.$$

Alternatively, if $w_b > 0$, then

$$\begin{aligned}
p_b &= 1 - \inf_{\boldsymbol{\theta}:k_0=b} P_{k_0=b}\{W_b(\boldsymbol{Y}) \geq w_b\} \\
&= 1 - P\left\{\inf_{x\in[a,b)} \frac{[\mathbf{P}(1,b+x)^T]^T \boldsymbol{T}}{\| \mathbf{P}(1,b+x)^T \|} \geq w_b\right\} \\
&\geq 1 - P\left\{\frac{[\mathbf{P}(1,a+b)^T]^T \boldsymbol{T}}{\| \mathbf{P}(1,a+b)^T \|} \geq w_b\right\} \\
&\geq 1 - P\left\{\frac{[\mathbf{P}(1,a+b)^T]^T \boldsymbol{T}}{\| \mathbf{P}(1,a+b)^T \|} \geq 0\right\} \\
&= 1 - 0.5 \\
&= 0.5.
\end{aligned}$$

We reject $H_0 : k_0 = b$ if and only if $p_b < \alpha$. Since $\alpha$ is always much smaller than 0.5, we can not reject $H_0 : k_0 = b$.

**Case III** Finally, for $s \in (a, b)$, our $(1-\alpha)$ level acceptance set is

$$\begin{aligned}
A(s) &= \{\boldsymbol{Y} : (\hat{\theta}_0 + \hat{\theta}_1 s + \hat{\theta}_2 s^2) - (\hat{\theta}_0 + \hat{\theta}_1 x + \hat{\theta}_2 x^2) \geq -c(s)\hat{\sigma} v(s,x),\ \forall x \in [a,s) \cup (s,b]\} \\
&= \{\boldsymbol{Y} : \inf_{x\in[a,s)\cup(s,b]} \hat{h}(s+x)\frac{s-x}{| s-x |} \geq -c(s)\}.
\end{aligned}$$

Let

$$W_s(\boldsymbol{Y}, s) = \inf_{x\in[a,s)\cup(s,b]} \hat{h}(s+x)\frac{s-x}{| s-x |},$$

$$w_s(s) = W_s(\boldsymbol{Y}^*, s) = \inf_{x\in[a,s)\cup(s,b]} \hat{h}^*(s+x)\frac{s-x}{| s-x |},$$

and define

$$p_s(s) = 1 - \inf_{\boldsymbol{\theta}:k_0=s} P_{k_0=s}\{W_s(\boldsymbol{Y}, s) \geq w_s(s)\}.$$

As

$$\begin{aligned}
s \in \mathbf{C}_E(\boldsymbol{Y}^*) &\Leftrightarrow \boldsymbol{Y}^* \notin A(s) \\
&\Leftrightarrow w_s(s) \geq -c(s) \\
&\Leftrightarrow \inf_{\boldsymbol{\theta}:k_0=s} P_{k_0=s}\{W_s(\boldsymbol{Y}, s) \geq w_s(s)\} \leq 1 - \alpha,
\end{aligned}$$

we have

$$\begin{aligned} s \notin \mathbf{C}_E(\boldsymbol{Y}^*) &\Leftrightarrow \inf_{\boldsymbol{\theta}:k_0=s} P_{k_0=s}\{W_s(\boldsymbol{Y}, s) \geq w_s(s)\} > 1-\alpha \\ &\Leftrightarrow p_s(s) < \alpha. \end{aligned}$$

Thus, $p_s(s)$ is a p-value for testing $H_0 : k_0 = s$.

Following the mathematical derivation in Section 3.1.2.1, if $w_s(s) \leq 0$, then we have

$$\begin{aligned} 1-p_s(s) &= \inf_{\boldsymbol{\theta}:k_0=s} P_{k_0=s}(W_s(\boldsymbol{Y}, s) \geq w_s(s)) \\ &= \inf_{\boldsymbol{\theta}:k_0=s} P_{k_0=s}\left\{\inf_{x\in[a,s)\cup(s,b]} \hat{h}(s+x)\frac{s-x}{\mid s-x\mid} \geq w_s(s)\right\} \\ &= P\left\{\inf_{x\in[a,s)} \frac{[\mathbf{P}(1,s+x)^T]^T\,\boldsymbol{T}}{\parallel \mathbf{P}(1,s+x)^T \parallel} \geq w_s(s),\ \sup_{x\in(s,b]} \frac{[\mathbf{P}(1,s+x)^T]^T\boldsymbol{T}}{\parallel \mathbf{P}(1,s+x)^T \parallel} \leq -w_s(s)\right\} \\ &= \frac{1}{2}F_{1,n-3}(w_s(s)^2) + \frac{1}{2}F_{2,n-3}\left(\frac{w_s(s)^2}{2}\right) + \frac{\phi(s)}{\pi}\left(1+\frac{w_s(s)^2}{n-3}\right)^{\frac{n-3}{2}} \\ &\quad -\frac{1}{\pi}\int_0^{\phi(s)}\left(1+\frac{w_s(s)^2}{(n-3)\cos^2(\phi(s)-\psi)}\right)^{-\frac{n-3}{2}} d\psi \end{aligned}$$

where $\phi(s)$ is defined in Section 3.1.2.1. Thus,

$$\begin{aligned} p_s(s) &= 1-\frac{1}{2}F_{1,n-3}(w_s(s)^2) - \frac{1}{2}F_{2,n-3}\left(\frac{w_s(s)^2}{2}\right) - \frac{\phi(s)}{\pi}\left(1+\frac{w_s(s)^2}{n-3}\right)^{\frac{n-3}{2}} \\ &\quad +\frac{1}{\pi}\int_0^{\phi(s)}\left(1+\frac{w_s(s)^2}{(n-3)\cos^2(\phi(s)-\psi)}\right)^{-\frac{n-3}{2}} d\psi. \end{aligned}$$

Alternatively, if $w_s(s) > 0$, then

$$\begin{aligned} p_s(s) &= 1-\inf_{\boldsymbol{\theta}:k_0=s} P_{k_0=s}\{W_s(\boldsymbol{Y}, s) \geq w_s(s)\} \\ &= 1-P\left\{\inf_{x\in[a,s)} \frac{[\mathbf{P}(1,s+x)^T]^T\,\boldsymbol{T}}{\parallel \mathbf{P}(1,s+x)^T \parallel} \geq w_s(s),\ \sup_{x\in(s,b]} \frac{[\mathbf{P}(1,s+x)^T]^T\boldsymbol{T}}{\parallel \mathbf{P}(1,s+x)^T \parallel} \leq -w_s(s)\right\} \\ &\geq 1-P\left\{\frac{[\mathbf{P}(1,s+a)^T]^T\,\boldsymbol{T}}{\parallel \mathbf{P}(1,s+a)^T \parallel} \geq w_s(s),\ \frac{[\mathbf{P}(1,s+b)^T]^T\boldsymbol{T}}{\parallel \mathbf{P}(1,s+b)^T \parallel} \leq -w_s(s)\right\} \qquad (3.39) \\ &\geq 1-P\left\{\frac{[\mathbf{P}(1,s+a)^T]^T\,\boldsymbol{T}}{\parallel \mathbf{P}(1,s+a)^T \parallel} \geq 0,\ \frac{[\mathbf{P}(1,s+b)^T]^T\boldsymbol{T}}{\parallel \mathbf{P}(1,s+b)^T \parallel} \leq 0\right\} \qquad (3.40) \\ &\geq 1-P\left\{\frac{[\mathbf{P}(1,s+a)^T]^T\,\boldsymbol{T}}{\parallel \mathbf{P}(1,s+a)^T \parallel} \geq 0\right\} \\ &= 1-0.5 \\ &= 0.5. \end{aligned}$$

We reject $H_0 : k_0 = s$ if and only if $p_s(s) < \alpha$. Since $\alpha$ is always much smaller than 0.5, we can not reject the null hypothesis when $w_s(s) > 0$.

#### 3.1.2.3 Examples

**Example 1.** We consider the simulation data in Table 3.1, which is generated from the quadratic polynomial regression model $Y = -5x^2 + 3x + 2 + e$ where $e \sim N(0, 10^2)$, with $x$ evenly chosen from a given interval $[a, b] = [-5, 5]$. The fitted regression function is $\hat{Y} = -4.8830x^2 + 2.3958x + 2.3750$ and the standard error is $\hat{\sigma} = 10.4254$ with degrees of freedom equal to 18.

A confidence set can be constructed to quantify the plausible value of a maximum point of the true quadratic function in the given interval $[-5, 5]$. To construct a 95% level confidence set for a maximum point, we check the grid points from $-5$ to 5 with resolution $d = 0.001$. Figure 3.7 shows the conservative critical values $c_1$, $c_2$ and the exact critical value $c(k)$ over $k \in [-5, 5]$. The conservative confidence set using the critical value $c_1 = \sqrt{2f^{0.05}_{2,18}} = 2.6663$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = [0.041, 0.462]$. Then, by using the p-value on each grid point in $\mathbf{C}_0(\boldsymbol{Y})$, we finally get the exact 95% level confidence set $\mathbf{C}_E(\boldsymbol{Y}) = [0.059, 0.441]$. Computation of $\mathbf{C}_E(\boldsymbol{Y})$ takes 27 seconds. The 95% confidence sets using bootstrap method and Rao's method are $[0.111, 0.392]$ and $[0.040, 0.456]$, respectively. The BH confidence set for the stationary point is given by $[0.090, 0.410]$. The data, fitted regression function, and all the five confidence sets are plotted in Figure 3.8. The true maximum point, 0.3, is also plotted for comparison.

The conservative confidence set constructed using critical value $c_2 = 2.4487$ is given by $[0.057, 0.443]$, which is about 10% shorter than the set $\mathbf{C}_0$. Since in the quadratic case, the computation of $\mathbf{C}_E(\boldsymbol{Y})$ is very fast and hence it does not matter much whether we use critical value $c_1$ or $c_2$.

**Example 2.** We consider the simulation data in Table 3.2, which is generated from the quadratic polynomial regression model $Y = -5x^2 + 3x + 2 + e$ where $e \sim N(0, 10^2)$. The fitted regression function is $\hat{Y} = -5.3786x^2 + 3.9892x + 5.6275$ and the standard error is $\hat{\sigma} = 7.9622$ with degrees of freedom equal to 8.

A confidence set can be constructed to quantify the plausible value of a maximum

Table 3.1: Simulation data used in Example 1.

| x | y | x | y | x | y |
|---|---|---|---|---|---|
| -5 | -130.74 | -1.5 | -14.706 | 2 | -3.42 |
| -4.5 | -118.63 | -1 | -14.323 | 2.5 | -9.21 |
| -4 | -68.168 | -0.5 | 2.1941 | 3 | -49.937 |
| -3.5 | -71.114 | 0 | -11.362 | 3.5 | -63.16 |
| -3 | -50.861 | 0.5 | 9.3932 | 4 | -60.289 |
| -2.5 | -26.082 | 1 | 16.236 | 4.5 | -89.749 |
| -2 | -23.407 | 1.5 | -11.668 | 5 | -101.1 |

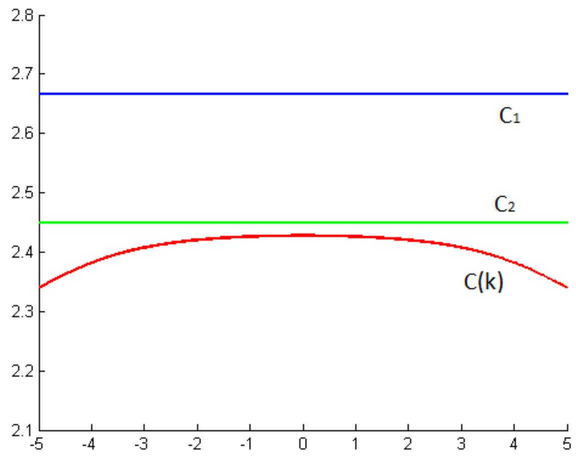



Figure 3.7: The 95% level critical values for different $x \in [-5, 5]$ in Example 1.

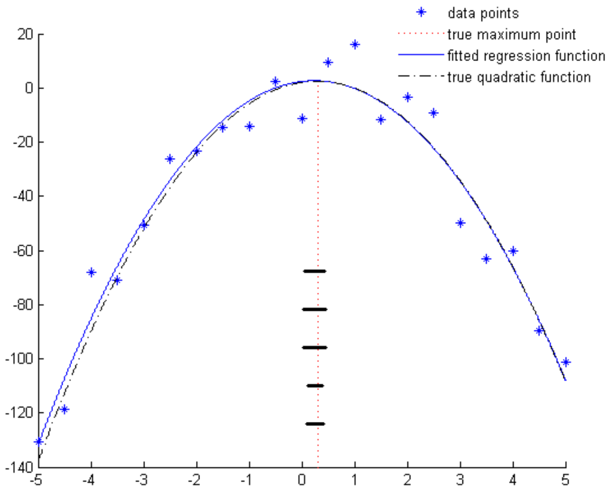



Figure 3.8: The 95% level confidence sets in Example 1. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point.

point of the true quadratic function in the given interval $[-5, 5]$. To construct a 95% level confidence set for a maximum point, we check the grid points from $-5$ to 5 with resolution $d = 0.001$. The conservative confidence set using the critical value $c_1 = \sqrt{2f_{2,8}^{0.05}} = 2.9863$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = [0.1590, 0.6000]$. Then, by using the p-value on each grid point in $\mathbf{C}_0(\boldsymbol{Y})$, we finally get the exact 95% level confidence set $\mathbf{C}_E(\boldsymbol{Y}) = [0.179, 0.577]$. The 95% confidence sets using bootstrap method and Rao's method are $[0.252, 0.487]$ and $[0.160, 0.590]$, respectively. The BH confidence set for the stationary point is given by $[0.210, 0.540]$. The data, fitted regression function, true regression function and its true maximum point, 0.3, are plotted in Figure 3.9, and so are the confidence sets.

**Example 3.** We consider the simulation data in Table 3.3, which is generated from the quadratic polynomial regression model $Y = -5x^2 + 3x + 2 + e$ where $e \sim N(0, 20^2)$. The fitted regression function is $\hat{Y} = -4.6813x^2 + 3.5333x - 7.9066$ and the standard error is $\hat{\sigma} = 21.7859$ with degrees of freedom equal to 8.

A confidence set can be constructed to quantify the plausible value of a maximum

Table 3.2: Simulation data used in Example 2.

| x | y | x | y | x | y |
|---|---|---|---|---|---|
| -5 | -142.33 | -1 | -17.465 | 3 | -30.727 |
| -4 | -106.66 | 0 | 13.909 | 4 | -64.254 |
| -3 | -50.747 | 1 | 11.892 | 5 | -109.87 |
| -2 | -21.123 | 2 | -12.376 | | |

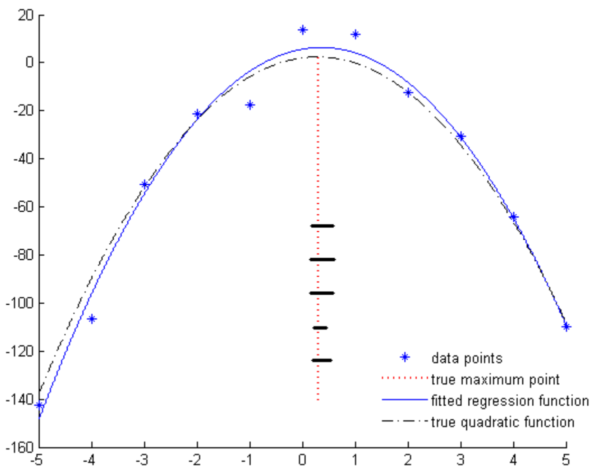



Figure 3.9: The 95% level confidence sets in Example 2. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point.

point of the true quadratic function in the given interval $[-5, 5]$. To construct a 95% level confidence set for a maximum point, we check the grid points from $-5$ to 5 with resolution $d = 0.001$. A conservative confidence set using the critical value $c_1 = \sqrt{2f_{2,8}^{0.05}} = 2.9863$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = [-0.300, 1.274]$. Then, by using the p-value on each grid point in $\mathbf{C}_0(\boldsymbol{Y})$, we finally get the exact 95% level confidence set $\mathbf{C}_E(\boldsymbol{Y}) = [-0.226, 1.149]$. The 95% confidence sets using bootstrap method and Rao's method are $[-0.001, 0.798]$ and $[-0.291, 1.129]$, respectively. The BH confidence set for the stationary point is given by $[-0.130, 1.00]$. The data, fitted regression function, true regression function and its true maximum point, 0.3, are plotted in Figure 3.10, and so are the confidence sets.

Table 3.3: Simulation data used in Example 3.

| x | y | x | y | x | y |
|---|---|---|---|---|---|
| -5 | -107.8295 | -1 | -9.7163 | 3 | -48.4298 |
| -4 | -128.9016 | 0 | 2.1787 | 4 | -70.0236 |
| -3 | -85.6109 | 1 | 16.7390 | 5 | -108.4093 |
| -2 | -35.4707 | 2 | -26.4454 | | |

**Example 4.** We consider the simulation data in Table 3.4, which is generated from the quadratic polynomial regression model $Y = -x^2 + 2x + 1 + e$ where $e \sim N(0, 10^2)$. The fitted regression function is $\hat{Y} = -0.9146x^2 + 0.9465x + 1.9889$ and the standard error is $\hat{\sigma} = 9.1076$ with degrees of freedom equal to 8.

A confidence set can be constructed to quantify the plausible value of a maximum point of the true quadratic function in the given interval $[0, 10]$. To construct a 95% level confidence set for a maximum point, we check the grid points from 0 to 10 with resolution $d = 0.001$. A conservative confidence set using the critical value $c_1 = \sqrt{2f_{2,8}^{0.05}} = 2.9863$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = [0, 2.996]$. Then, by using the p-value on each grid point in $\mathbf{C}_0(\boldsymbol{Y})$, we finally get the exact 95% level confidence set $\mathbf{C}_E(\boldsymbol{Y}) = [0, 2.851]$. The 95% confidence sets using bootstrap method and Rao's method are $[0, 2.315]$ and $[0, 2.092]$, respectively. The BH confidence set for the

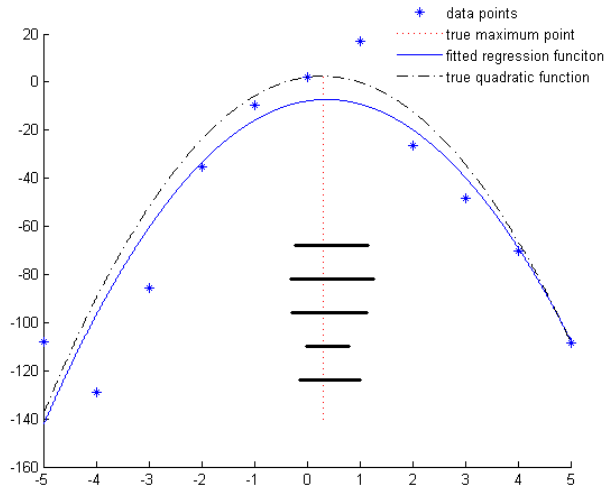



Figure 3.10: The 95% level confidence sets in Example 3. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point.

stationary point is given by $[0, 2.650]$. The data, fitted regression function, true regression function and its true maximum point, $x = 1$, are plotted in Figure 3.11, and so are the confidence sets.

Table 3.4: Simulation data used in Example 4.

| x | y | x | y | x | y |
|---|---|---|---|---|---|
| 0 | 6.9128 | 4 | -7.1951 | 8 | -36.05 |
| 1 | -4.436 | 5 | -14.482 | 9 | -80.74 |
| 2 | 4.8034 | 6 | -23 | 10 | -74.718 |
| 3 | -12.091 | 7 | -37.179 | | |

**Example 5.** We consider the simulation data in Table 3.5, which is generated from the quadratic polynomial regression model $Y = x^2 - 2x + 1 + e$ where $e \sim N(0, 10^2)$. The fitted regression function is $\hat{Y} = 0.9734x^2 + -3.4488x + 9.9499$ and the standard error is $\hat{\sigma} = 4.5741$ with degrees of freedom equal to 8.

To construct a 95% level confidence set for a maximum point in the interval $[0, 10]$, we

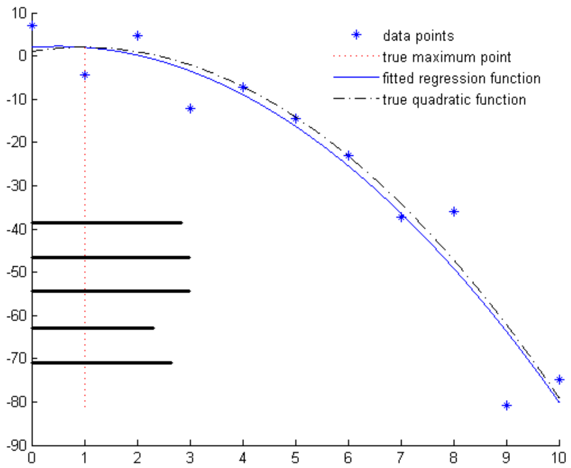



Figure 3.11: The 95% level confidence sets in Example 4. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point.

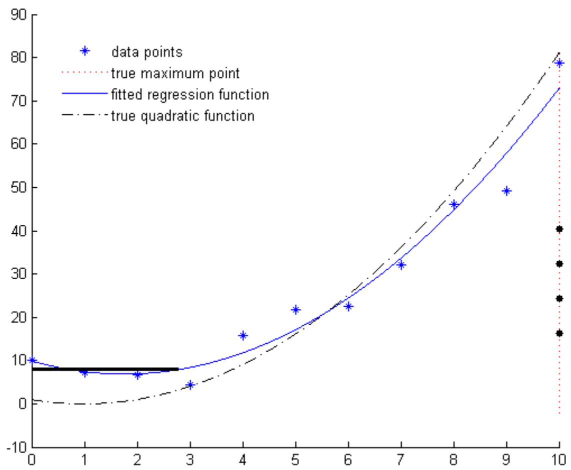



Figure 3.12: The 95% level confidence sets in Example 5. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point. All the confidence sets for a maximum point contain just one element 10, yet the BH confidence set is for the stationary point and is an interval.

check the grid points from 0 to 10 with resolution $d = 0.001$. The conservative confidence set using the critical value $c_1 = \sqrt{2f_{2,8}^{0.05}} = 2.9863$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = \{10\}$. Then, by using the p-value on $\{10\}$, we finally get the exact 95% level confidence set $\mathbf{C}_E(\boldsymbol{Y}) = \{10\}$. Computation of $\mathbf{C}_E(\boldsymbol{Y})$ takes 21 seconds. The 95% confidence sets using bootstrap method and Rao's method are both $\{10\}$, respectively. The BH confidence set for the stationary point is given by $[0, 2.75]$. The data, fitted regression function, true regression function and its true maximum point, 10, are plotted in Figure 3.12, and so are the confidence sets.

Table 3.5: Simulation data used in Example 5.

| x | y | x | y | x | y |
|---|---|---|---|---|---|
| 0 | 9.9564 | 4 | 15.771 | 8 | 46.041 |
| 1 | 7.3096 | 5 | 21.689 | 9 | 49.249 |
| 2 | 6.7786 | 6 | 22.444 | 10 | 78.66 |
| 3 | 4.4031 | 7 | 32.225 | | |

Table 3.6: Simulation data used in Example 6.

| x | y | x | y | x | y |
|---|---|---|---|---|---|
| -5 | 122.6744 | -1 | -4.4647 | 3 | 50.2729 |
| -4 | 65.3442 | 0 | 13.9092 | 4 | 83.7464 |
| -3 | 48.2533 | 1 | 18.8916 | 5 | 125.1329 |
| -2 | 24.8768 | 2 | 21.6237 | | |

**Example 6.** We consider the simulation data in Table 3.6, which is generated from the quadratic polynomial regression model $Y = 5x^2 + 2 + e$, where $e \sim N(0, 10^2)$. The fitted regression function is $\hat{Y} = 4.6214x^2 + 0.9892x + 5.6275$ and the standard error is $\hat{\sigma} = 7.9622$ with degrees of freedom equal to 8.

A confidence set can be constructed to quantify the plausible value of a maximum point of the true quadratic function in the given interval $[-5, 5]$. To construct a 95% level confidence set for a maximum point, we check the grid points from $-5$

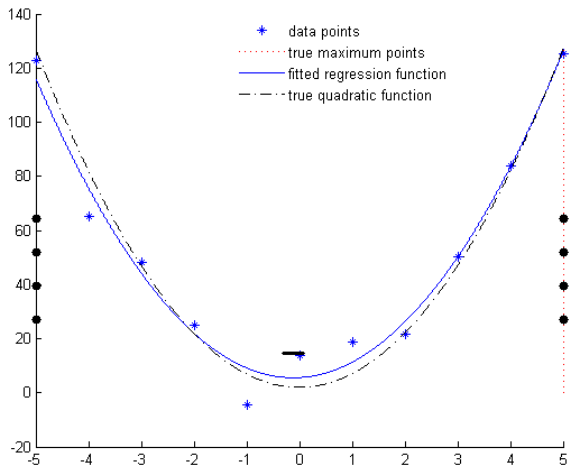



Figure 3.13: The 95% level confidence sets in Example 6. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point.

to 5 with resolution $d = 0.01$. A conservative confidence set using the critical value $c_1 = \sqrt{2f_{2,8}^{0.05}} = 2.9863$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = \{-5, 5\}$. Then, by using the p-value method on $\{-5, 5\}$, we get the 95% level confidence set $\mathbf{C}_E(\boldsymbol{Y}) = \{-5, 5\}$. The computation of $\mathbf{C}_E(\boldsymbol{Y})$ takes 24 seconds. The 95% confidence sets using bootstrap method and Rao's method are $\{-5, 5\}$ and $\{-5, 5\}$, respectively. The BH confidence set for the stationary point is given by $[-0.300, 0.080]$. The data, fitted regression function, and the confidence sets are plotted in Figure 3.13.

**Example 7.** The data of crop yields, denoted by Y, for the amounts of fertilizer used, denoted by x are shown in Table 3.7 (Sullivan and Sullivan, 2002). The fitted quadratic regression model is $\hat{Y} = 3.8939 + 1.0765x - 0.0171x^2$ and the standard error is $\hat{\sigma} = 1.4270$.

From the residual plots (Figure 3.14) and normal probability plot (Figure 3.15), the normality assumption of the residuals seems reasonable. The exploratory index $R^2 = 0.9494$, so the data fit the quadratic model well. To construct a 95% level confidence set for the amount of fertilizer that maximizes crop yields, we check the

Table 3.7: Data of crop yields and fertilizer used (Sullivan and Sullivan, 2002).

| Plot | Fertilizer,x(Pounds/100$ft^2$) | Yield(Bushels) |
|---|---|---|
| 1 | 0 | 4 |
| 2 | 0 | 6 |
| 3 | 5 | 10 |
| 4 | 5 | 7 |
| 5 | 10 | 12 |
| 6 | 10 | 10 |
| 7 | 15 | 15 |
| 8 | 15 | 17 |
| 9 | 20 | 18 |
| 10 | 20 | 21 |
| 11 | 25 | 20 |
| 12 | 25 | 21 |
| 13 | 30 | 21 |
| 14 | 30 | 22 |
| 15 | 35 | 21 |
| 16 | 35 | 20 |
| 17 | 40 | 19 |
| 18 | 40 | 19 |

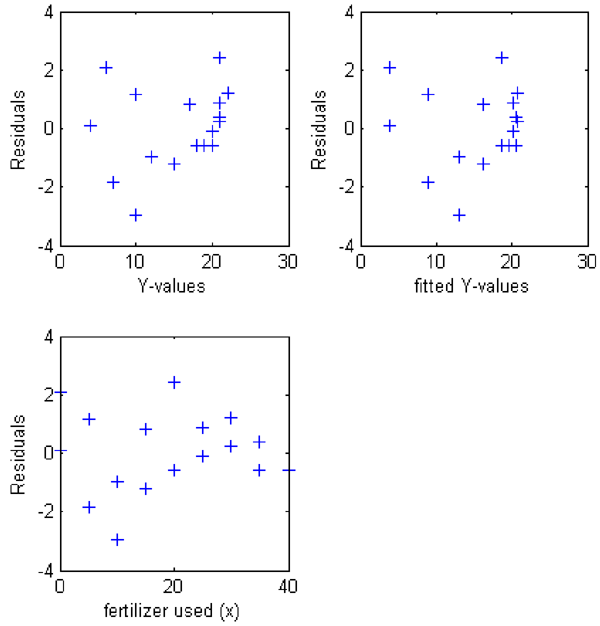



Figure 3.14: Residual plots

grid points from 0 to 40 with resolution $d = 0.1$. A conservative confidence set using the critical value $c_1 = \sqrt{2f^{0.05}_{2,15}} = 2.7138$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = [27.9, 38.4]$. Then, by using the p-value on each grid point in $\mathbf{C}_0(\boldsymbol{Y}) = [27.9, 38.4]$, we finally get the exact 95% level confidence set $\mathbf{C}_E(\boldsymbol{Y}) = [28.17, 37.32]$. The 95% confidence sets using bootstrap method and Rao's method are $[28.827, 35.308]$ and $[27.933, 37.502]$, respectively. The BH confidence set for the stationary point is given by $[28.52, 36.40]$. The data, regression function and the confidence sets are depicted in Figure 3.16. The reason that there are only 17 data points in Figure 3.16 is that the 17th and 18th data points are the same, and thus overlapped.

**Example 8.** The sample mean and standard deviation data given in Table 3.8 is the summary of the clinical dose response data given in Table 3.9. We fit the data in Table 3.9 to a quadratic regression function. The estimated parameters based on

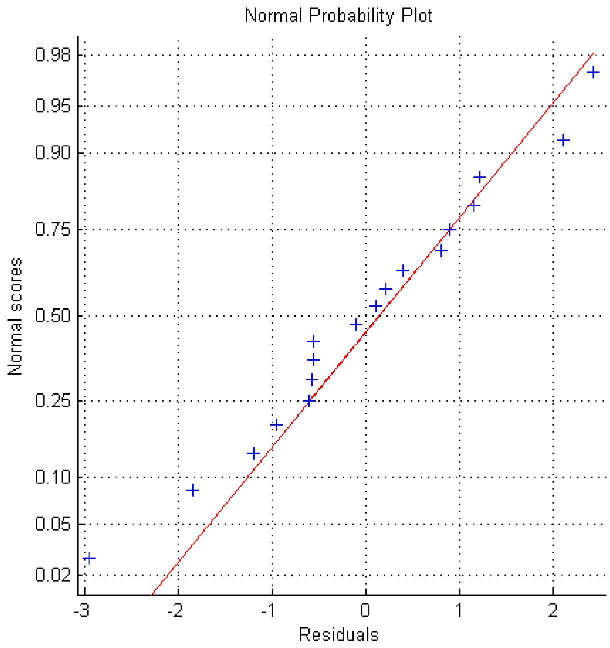



Figure 3.15: Normal probability plot of the residuals

these data are

$$\hat{\boldsymbol{\theta}}^* = \begin{pmatrix} 0.3898 \\ 1.7581 \\ -1.2204 \end{pmatrix}, \ \hat{\sigma}^* = 0.7084$$

and the exploratory index $R^2 = 0.1071$.

A confidence set can be constructed to quantify the plausible value of a maximum point of the quadratic model $Y = \theta_2 x^2 + \theta_1 x + \theta_0 + e$ in the given interval $[0, 1]$ based on the summary information in Table 3.8. The significance level is $\alpha = 0.05$. We check the grid points from 0 to 1 with resolution $d = 0.001$. A conservative confidence set using the critical value $c_1 = \sqrt{2 f_{2,97}^{0.05}} = 2.5335$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = [0.524, 1]$. Then, by using the p-value on each grid point in $\mathbf{C}_0(\boldsymbol{Y})$, we finally get the exact 95% level confidence set $\mathbf{C}_E(\boldsymbol{Y}) = [0.536, 1]$. Instead of using Table 3.9, we get the same estimates of parameters and hence the same confidence set using just Table 3.8(see Appendix).

The 95% confidence sets using bootstrap method and Rao's method are $[0.555, 1]$

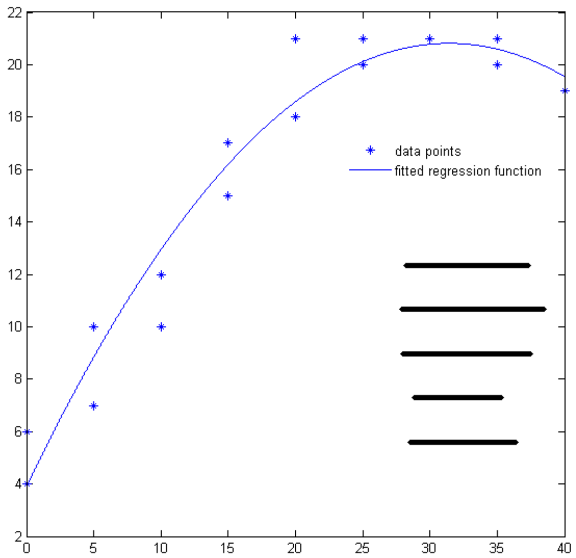



Figure 3.16: The 95% level confidence sets for the amount of fertilizer used that maximize the crop yields in Example 7. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point.

.

and $[0.51, 1]$, respectively. The BH confidence set for the stationary point is given by $[0.552, 1]$. The data points, fitted regression curve, and the confidence sets are plotted in Figure 3.17.

Table 3.8: Summary of clinical dose response data (Rom et al., 1994)

| Dose | Sample size | Sample mean | Sample Std Dev |
|---|---|---|---|
| 0 | 20 | 0.34 | 0.52 |
| 0.05 | 20 | 0.46 | 0.49 |
| 0.2 | 20 | 0.81 | 0.74 |
| 0.6 | 20 | 0.93 | 0.76 |
| 1 | 20 | 0.95 | 0.95 |

**Example 9.** Summary of a set of clinical dose response data is given in Table 3.10. We fit the data to a quadratic regression function. The estimated parameters based on these data are

$$\hat{\boldsymbol{\theta}}^* = \begin{pmatrix} -1.6669 \\ -0.0443 \\ 0.0001 \end{pmatrix}, \hat{\sigma}^* = 15.4992$$

and the exploratory index $R^2 = 0.0147$ is very small, which may be due to the large error variance of the data.

A confidence set can be constructed to quantify the plausible value of a maximum point of the quadratic model $Y = \theta_2 x^2 + \theta_1 x + \theta_0 + e$ in the given interval $[0, 400]$ based on the summary information in Table 3.10. The significance level is $\alpha = 0.05$. We check the grid points from 0 to 400 with resolution $d = 1$. Using the critical value $c_1 = \sqrt{2 f_{2,97}^{0.05}} = 2.5335$, a conservative confidence set is given by $\mathbf{C}_0(\boldsymbol{Y}) = [0, 400]$. Then, by using the p-value on each point in $\mathbf{C}_0(\boldsymbol{Y})$, we finally get the exact 95% level confidence set $\mathbf{C}_E(\boldsymbol{Y}) = [0, 400]$. The 95% confidence sets using bootstrap method and Rao's method are both $[0, 400]$, respectively. The BH confidence set for the stationary point is also given by $[0, 400]$. The data points, fitted regression curve, and the confidence sets are plotted in Figure 3.18.

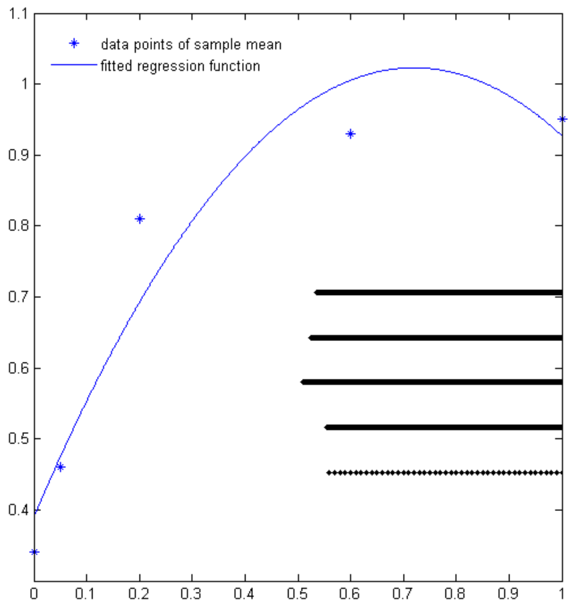



Figure 3.17: The 95% level confidence sets in Example 8. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point.

Table 3.9: Clinical dose response data (Rom et al., 1994)

| dose | response | dose | response | dose | response |
|---|---|---|---|---|---|
| 0 | 0.354621644 | 0.05 | 1.353800979 | 0.2 | 1.198081819 |
| 0 | 0.136528014 | 0.05 | 0.154980239 | 0.2 | 1.56768394 |
| 0 | -0.017221273 | 0.05 | -0.070838037 | 0.2 | -0.155607279 |
| 0 | 0.367483563 | 0.05 | 0.583740909 | 0.2 | 0.20658777 |
| 0 | 0.395127618 | 0.05 | 0.962598527 | 0.2 | 1.853465242 |
| 0 | 0.357965911 | 0.05 | 0.381388876 | 0.2 | 0.995682873 |
| 0 | 0.312936459 | 0.05 | -0.004695991 | 0.2 | 2.452234157 |
| 0 | -0.044915902 | 0.05 | 0.519304935 | 0.2 | -0.52039887 |
| 0 | 0.623091387 | 0.05 | -0.36332213 | 0.2 | 0.048856237 |
| 0 | 0.03951218 | 0.05 | 0.338903336 | 0.2 | 0.633128014 |
| 0 | -0.30840357 | 0.05 | 0.145620717 | 0.2 | 0.528413238 |
| 0 | -0.452964305 | 0.05 | 0.442803406 | 0.2 | 0.422902023 |
| 0 | -0.201709236 | 0.05 | -0.021949664 | 0.2 | 1.233325825 |
| 0 | 1.560869091 | 0.05 | 0.37259824 | 0.2 | 1.867825704 |
| 0 | 0.561286293 | 0.05 | 1.528976072 | 0.2 | 1.057835057 |
| 0 | 0.336241137 | 0.05 | 0.149471298 | 0.2 | 0.34534894 |
| 0 | -0.159375068 | 0.05 | 0.374063362 | 0.2 | 0.404851371 |
| 0 | 0.992604728 | 0.05 | 1.0116782 | 0.2 | 0.48420331 |
| 0 | 0.806987008 | 0.05 | 0.281306267 | 0.2 | 0.571199878 |
| 0 | 1.237442253 | 0.05 | 0.994656155 | 0.2 | 1.010696185 |
| 0.6 | 1.172221004 | 1 | 2.248445841 | | |
| 0.6 | 0.667558102 | 1 | 2.173688799 | | |
| 0.6 | 1.777681617 | 1 | 1.254509265 | | |
| 0.6 | 0.309735795 | 1 | 1.864626603 | | |
| 0.6 | 0.309735795 | 1 | 1.864626603 | | |
| 0.6 | 0.05646191 | 1 | -1.113082058 | | |
| 0.6 | 0.900065687 | 1 | 1.198415091 | | |
| 0.6 | 0.738302949 | 1 | 1.97485829 | | |
| 0.6 | 0.227879049 | 1 | 0.62533415 | | |
| 0.6 | 1.385106066 | 1 | 1.330258137 | | |
| 0.6 | 0.912390582 | 1 | 0.149454689 | | |
| 0.6 | 0.299300126 | 1 | 0.887139996 | | |
| 0.6 | 0.278917304 | 1 | 0.563064088 | | |
| 0.6 | 0.270043807 | 1 | 0.730863126 | | |
| 0.6 | 0.569625173 | 1 | -0.771473861 | | |
| 0.6 | 1.604328006 | 1 | 0.494778358 | | |
| 0.6 | 1.576944531 | 1 | 0.303110251 | | |
| 0.6 | 2.924508962 | 1 | 0.425231331 | | |
| 0.6 | 0.171532376 | 1 | 2.557773784 | | |
| 0.6 | 2.160523738 | 1 | 1.12889658 | | |
| 0.6 | 0.685611744 | 1 | 0.948335871 | | |

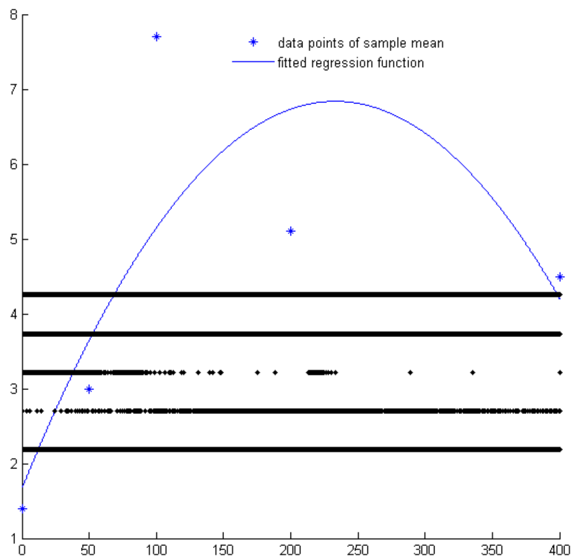



Figure 3.18: The 95% level confidence sets in Example 9. The five confidence sets plotted, from top to bottom, are $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$, bootstrap confidence set, and the BH confidence set for the stationary point.

Table 3.10: Summary of clinical dose response data (Rom et al., 1994)

| Dose | Sample size | Sample mean | Sample Std Dev |
|---|---|---|---|
| 0 | 16 | 1.4 | 16 |
| 50 | 8 | 3.0 | 15.68 |
| 100 | 8 | 7.7 | 15.68 |
| 200 | 8 | 5.1 | 15.68 |
| 400 | 7 | 4.5 | 15.75 |

### 3.1.3 A General Polynomial Regression Model

When $p \geq 2$, we have the polynomial regression model

$$Y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_p x^p + e$$

with $e \sim N(0, \sigma^2)$. Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{pmatrix} \text{ and } \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}.$$

It is clear that $\mathbf{X}$ is non-singular. The least squares estimate of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_p \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$$

and hence has the distribution $N(\boldsymbol{\theta}, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2)$. Let $\hat{\sigma}^2$ be the usual estimate of $\sigma^2$ which has the distribution

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{n-p-1}\chi^2_{n-p-1},$$

where $\chi^2_{n-p-1}$ is the chi-squared distribution with degrees of freedom $(n-p-1)$. Moreover, $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ are independent. Let

$$\boldsymbol{\theta}^0 = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}, \ \hat{\boldsymbol{\theta}}^0 = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_p \end{pmatrix}, \ \text{ and } \ \mathbf{P}^2 = \begin{bmatrix} \mathbf{0} & \mathbf{I}_p \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_p \end{bmatrix},$$

thus the variance-covariance matrix of $\hat{\boldsymbol{\theta}}^0$ is given by

$$cov\left(\begin{bmatrix} \mathbf{0} & \mathbf{I}_p \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}\right) = \sigma^2\mathbf{P}^2.$$

Therefore

$$\mathbf{P}^{-1} \times (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)/\sigma \sim N(0, \mathbf{I}_p),$$

and hence

$$\boldsymbol{T} := \mathbf{P}^{-1} \times (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)/\hat{\sigma} \sim \boldsymbol{T}_{p,n-p-1},$$

where $\boldsymbol{T}_{p,n-p-1}$ is the multivariate t-distribution with $p$ elements, and the degrees of freedom is $(n-p-1)$.

In this section, we use the subscript $p$ to indicate a function's definition related to the polynomial regression.

#### 3.1.3.1 Theory

To apply Neyman's Theorem to construct the confidence set for a maximum point of the polynomial regression function

$$f(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_p x^p,$$

we require an acceptance set for testing

$$H_0 : k = k_0$$

for each $k_0 \in [a, b]$. For $z_1 \neq z_2$, we define three functions $g_p$, $v_p$ and $h_p$ as

$$\begin{aligned} g_p(z_1, z_2, p) &:= \frac{(z_1, \cdots, z_1^p)^T - (z_2, \cdots, z_2^p)^T}{z_1 - z_2} \\ &= \frac{(z_1 - z_2, \cdots, z_1^p - z_2^p)^T}{z_1 - z_2} \\ &= (1, z_1 + z_2, \cdots, z_1^{p-1} + z_1^{p-2} z_2 + \cdots + z_1 z_2^{p-2} + z_2^{p-1})^T, \end{aligned}$$

$$\begin{aligned} v_p(z_1, z_2, \hat{\boldsymbol{\theta}}) : &= \sqrt{var[(\hat{\theta}_0 + \hat{\theta}_1 z_1 + \cdots + \hat{\theta}_p z_1^p) - (\hat{\theta}_0 + \hat{\theta}_1 z_2 + \cdots + \hat{\theta}_p z_2^p)]/\sigma^2} \\ &= \mid z_1 - z_2 \mid \sqrt{g_p(z_1, z_2, p)^T \mathbf{P}^2 g_p(z_1, z_2, p)} \end{aligned}$$

and

$$\hat{h}_p(z_1, z_2) := \frac{g_p(z_1, z_2, p)^T \hat{\boldsymbol{\theta}}^0}{\hat{\sigma}\sqrt{g_p(z_1, z_2, p)^T \mathbf{P}^2 g_p(z_1, z_2, p)}}.$$

In what follows, we consider three cases: $k_0 = a$, $k_0 = b$ and $k_0 \in (a, b)$.

**Case I** If $k_0 = a$ is a maximum point, then for any $x \in (a, b]$, we have

$$(\theta_0 + \theta_1 a + \cdots + \theta_p a^p) - (\theta_0 + \theta_1 x + \cdots + \theta_p x^p) \geq 0,$$

or equivalently

$$(a - x) g_p(a, x, p)^T \boldsymbol{\theta}^0 \geq 0.$$

This implies that $g_p(a, x, p)^T \boldsymbol{\theta}^0 \leq 0$ for all $x \in (a, b]$, since $(a - x)$ is negative for $x \in (a, b]$. Accordingly, we define an acceptance set

$$\begin{aligned} A(a) &= \{\mathbf{Y} : (\hat{\theta_0} + \hat{\theta_1} a + \cdots + \hat{\theta_p} a^p) - (\hat{\theta_0} + \hat{\theta_1} x + \cdots + \hat{\theta_p} x^p) \geq -c(a) \hat{\sigma} v_p(a, x, \hat{\boldsymbol{\theta}}), \\ &\quad \forall x \in (a, b]\} \\ &= \{\mathbf{Y} : (a - x) g_p(a, x, p)^T \hat{\boldsymbol{\theta}}^0 \geq -c(a) \hat{\sigma} v_p(a, x, \hat{\boldsymbol{\theta}}),\ \forall x \in (a, b]\} \\ &= \{\mathbf{Y} : g_p(a, x, p)^T \hat{\boldsymbol{\theta}}^0 \leq c(a) \hat{\sigma} \sqrt{g_p(a, x, p)^T \mathbf{P}^2 g_p(a, x, p)},\ \forall x \in (a, b]\} \qquad (3.41) \\ &= \{\mathbf{Y} : \sup_{x \in (a, b]} \hat{h}_p(a, x) \leq c(a)\} \end{aligned}$$

where $c(a) > 0$ is a critical constant that can be calculated such that $A(a)$ is a $(1 - \alpha)$ level acceptance set. To determine $c(a)$, note that

$$\begin{aligned} & P_{k_0=a} \quad \{\mathbf{Y} \in A(a)\} \\ = \ & P_{k_0=a} \quad \{g_p(a, x, p)^T \hat{\boldsymbol{\theta}}^0 \leq c(a) \hat{\sigma} \sqrt{g_p(a, x, p)^T \mathbf{P}^2 g_p(a, x, p)},\ \forall x \in (a, b]\} \qquad (3.42) \\ \geq \ & P_{k_0=a} \quad \{g_p(a, x, p)^T \hat{\boldsymbol{\theta}}^0 \leq c(a) \hat{\sigma} \sqrt{g_p(a, x, p)^T \mathbf{P}^2 g_p(a, x, p)} + g_p(a, x, p)^T \boldsymbol{\theta}^0,\ \forall x \in (a, b]\} \\ & \qquad (3.43) \\ = \ & P_{k_0=a} \quad \left\{ \frac{g_p(a, x, p)^T (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)}{\hat{\sigma} \sqrt{g_p(a, x, p)^T \mathbf{P}^2 g_p(a, x, p)}} \leq c(a),\ \forall x \in (a, b] \right\} \\ = \ & P_{k_0=a} \quad \left\{ \sup_{x \in (a, b]} \frac{g_p(a, x, p)^T (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)}{\hat{\sigma} \sqrt{g_p(a, x, p)^T \mathbf{P}^2 g_p(a, x, p)}} \leq c(a) \right\} \\ = \ & P_{k_0=a} \quad \left\{ \sup_{x \in (a, b]} \frac{[\mathbf{P} g_p(a, x, p)]^T [\mathbf{P}^{-1} (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)]}{\hat{\sigma} \sqrt{g_p(a, x, p)^T \mathbf{P}^2 g_p(a, x, p)}} \leq c(a) \right\} \\ = \ & \quad P \quad \left\{ \sup_{x \in (a, b]} \frac{[\mathbf{P} g_p(a, x, p)]^T \boldsymbol{T}}{\| \mathbf{P} g_p(a, x, p) \|} \leq c(a) \right\} \qquad (3.44) \end{aligned}$$

where Equation (3.42) follows directly from Equation (3.41), and Equation (3.43) follows from the fact that $g_p(a, x, p)^T \boldsymbol{\theta}^0 \leq 0$ for $\forall x \in (a, b]$. Note that

$$\inf_{\boldsymbol{\theta}: k_0 = a} P_{k_0=a}\{\mathbf{Y} \in A(a)\} = P \left\{ \sup_{x \in (a, b]} \frac{[\mathbf{P} g_p(a, x, p)]^T \boldsymbol{T}}{\| \mathbf{P} g_p(a, x, p) \|} \leq c(a) \right\},$$

with the infimum being attained at $\theta_1 = \theta_2 = \cdots = \theta_p = 0$. Then the critical value $c(a)$ is the unique solution of

$$P\left\{ \sup_{x\in(a,b]} \frac{[\mathbf{P}g_p(a,x,p)]^T \boldsymbol{T}}{\parallel \mathbf{P}g_p(a,x,p) \parallel} \le c(a) \right\} = 1-\alpha. \tag{3.45}$$

**Case II** If $k_0 = b$ is a maximum point, then for any $x \in [a,b)$,

$$(\theta_0 + \theta_1 b + \cdots + \theta_p b^p) - (\theta_0 + \theta_1 x + \cdots + \theta_p x^p) \ge 0$$

which implies $g_p(b,x,p)^T \boldsymbol{\theta}^0 \ge 0$ for all $x \in [a,b)$. Accordingly, we define an acceptance set as

$$\begin{aligned}
A(b) &= \{\mathbf{Y} : (\hat{\theta}_0 + \hat{\theta}_1 b + \cdots + \hat{\theta}_p b^p) - (\hat{\theta}_0 + \hat{\theta}_1 x + \cdots + \hat{\theta}_p x^p) \ge -c(b)\hat{\sigma} v_p(b,x,\hat{\boldsymbol{\theta}}), \\
&\quad\ \forall x \in [a,b)\} \\
&= \{\mathbf{Y} : (b-x) g_p(b,x,p)^T \hat{\boldsymbol{\theta}}^0 \ge -c(b)\hat{\sigma} v_p(b,x,\hat{\boldsymbol{\theta}}),\ \forall x \in [a,b)\} \\
&= \{\mathbf{Y} : g_p(b,x,p)^T \hat{\boldsymbol{\theta}}^0 \ge -c(b)\hat{\sigma}\sqrt{g_p(b,x,p)^T \mathbf{P}^2 g_p(b,x,p)},\ \forall x \in [a,b)\} \quad (3.46) \\
&= \{\mathbf{Y} : \inf_{x\in[a,b)} \hat{h}_p(b,x) \ge -c(b)\}
\end{aligned}$$

where $c(b) > 0$ is a critical constant such that $A(b)$ is a $(1-\alpha)$ level acceptance set. To determine $c(b)$, note that

$$\begin{aligned}
& P_{k_0=b} \quad \{\mathbf{Y} \in A(b)\} \\
= \ & P_{k_0=b} \quad \{g_p(b,x,p)^T \hat{\boldsymbol{\theta}}^0 \ge -c(b)\hat{\sigma}\sqrt{g_p(b,x,p)^T \mathbf{P}^2 g_p(b,x,p)},\ \forall x \in [a,b)\} \qquad (3.47) \\
\ge \ & P_{k_0=b} \quad \{g_p(b,x,p)^T \hat{\boldsymbol{\theta}}^0 \ge -c(b)\hat{\sigma}\sqrt{g_p(b,x,p)^T \mathbf{P}^2 g_p(b,x,p)} + g_p(b,x,p)^T \boldsymbol{\theta}^0,\ \forall x \in [a,b)\} \\
& \qquad (3.48) \\
= \ & P_{k_0=b} \quad \left\{ \frac{g_p(b,x,p)^T (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)}{\hat{\sigma}\sqrt{g_p(b,x,p)^T \mathbf{P}^2 g_p(b,x,p)}} \ge -c(b),\ \forall x \in [a,b) \right\} \\
= \ & P_{k_0=b} \quad \left\{ \inf_{x\in[a,b)} \frac{g_p(b,x,p)^T (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)}{\hat{\sigma}\sqrt{g_p(b,x,p)^T \mathbf{P}^2 g_p(b,x,p)}} \ge -c(b) \right\} \\
= \ & P_{k_0=b} \quad \left\{ \inf_{x\in[a,b)} \frac{[\mathbf{P}g_p(b,x,p)]^T [\mathbf{P}(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)]}{\hat{\sigma}\sqrt{g_p(b,x,p)^T \mathbf{P}^2 g_p(b,x,p)}} \ge -c(b) \right\} \\
= \ & \quad P \quad \left\{ \inf_{x\in[a,b)} \frac{[\mathbf{P}g_p(b,x,p)]^T \boldsymbol{T}}{\parallel \mathbf{P}g_p(b,x,p) \parallel} \ge -c(b) \right\} \qquad (3.49)
\end{aligned}$$

where Equation (3.47) follows directly from Equation (3.46), and Equation (3.48) follows from the fact that $g_p(b,x,p)^T\boldsymbol{\theta}^0 \geq 0$ for all $x \in [a,b)$. Note that

$$\inf_{\boldsymbol{\theta}:k_0=b}\{\mathbf{Y} \in A(b)\} = P\left\{\inf_{x\in[a,b)} \frac{[\mathbf{P}g_p(b,x,p)]^T\boldsymbol{T}}{\parallel \mathbf{P}g_p(b,x,p) \parallel} \geq -c(b)\right\},$$

with the infimum being attained at $\theta_1 = \theta_2 = \cdots = \theta_p = 0$. Then the critical value $c(b)$ is the unique solution of

$$P\left\{\inf_{x\in[a,b)} \frac{[\mathbf{P}g_p(b,x,p)]^T\boldsymbol{T}}{\parallel \mathbf{P}g_p(b,x,p) \parallel} \geq -c(b)\right\} = 1-\alpha. \tag{3.50}$$

**Case III** If $k_0 = s$ for $s \in (a,b)$ is a maximum point, then for any $x \in [a,s) \cup (s,b]$,

$$(\theta_0 + \theta_1 s + \cdots + \theta_p s^p) - (\theta_0 + \theta_1 x + \cdots + \theta_p x^p) \geq 0$$

which implies $g_p(s,x,p)^T\boldsymbol{\theta}^0 \geq 0$ for $x \in [a,s)$ and $g_p(s,x,p)^T\boldsymbol{\theta}^0 \leq 0$ for $x \in (s,b]$. Accordingly, for each $s \in (a,b)$, we define an acceptance set

$$\begin{aligned}
A(s) &= \{\mathbf{Y} : (\hat{\theta}_0 + \hat{\theta}_1 s + \cdots + \hat{\theta}_p s^p) - (\hat{\theta}_0 + \hat{\theta}_1 x + \cdots + \hat{\theta}_p x^p) \geq -c(s)\hat{\sigma}v_p(s,x,\hat{\boldsymbol{\theta}}), \\
&\qquad \forall x \in [a,s) \cup (s,b]\} \\
&= \{\mathbf{Y} : (s-x)g_p(s,x,p)^T\hat{\boldsymbol{\theta}}^0 \geq -c(s)\hat{\sigma}v_p(s,x,\hat{\boldsymbol{\theta}}),\ \forall x \in [a,s) \cup (s,b]\} \\
&= \{\mathbf{Y} : g_p(s,x,p)^T\hat{\boldsymbol{\theta}}^0 \geq -c(s)\hat{\sigma}\sqrt{g_p(s,x,p)^T\mathbf{P}^2 g_p(s,x,p)},\ \forall x \in [a,s) \\
&\qquad\qquad g_p(s,x,p)^T\hat{\boldsymbol{\theta}}^0 \leq c(s)\hat{\sigma}\sqrt{g_p(s,x,p)^T\mathbf{P}^2 g_p(s,x,p)},\ \forall x \in (s,b]\} \qquad (3.51) \\
&= \{\mathbf{Y} : \inf_{x\in[a,s)} \hat{h}_p(s,x) \geq -c(s), \quad \sup_{x\in(s,b]} \hat{h}_p(s,x) \leq c(s)\}
\end{aligned}$$

where $c(s) > 0$ is a critical constant such that $A(s)$ is a $(1-\alpha)$ level acceptance set.

To determine $c(s)$, note that

$$
\begin{aligned}
& P_{k_0=s} \quad \{\mathbf{Y} \in A(s)\} \\
= \;& P_{k_0=s} \quad \{g_p(s,x,p)^T \hat{\boldsymbol{\theta}}^0 \geq -c(s)\hat{\sigma}\sqrt{g_p(s,x,p)^T \mathbf{P}^2 g_p(s,x,p)},\ \forall x \in [a,s), \\
& \qquad g_p(s,x,p)^T \hat{\boldsymbol{\theta}}^0 \leq c(s)\hat{\sigma}\sqrt{g_p(s,x,p)^T \mathbf{P}^2 g_p(s,x,p)},\ \forall x \in (s,b]\} \qquad (3.52) \\
\geq \;& P_{k_0=s} \quad \{g_p(s,x,p)^T \hat{\boldsymbol{\theta}}^0 \geq -c(s)\hat{\sigma}\sqrt{g_p(s,x,p)^T \mathbf{P}^2 g_p(s,x,p)} + g_p(s,x,p)^T \boldsymbol{\theta}^0,\ \forall x \in [a,s), \\
& \qquad g_p(s,x,p)^T \hat{\boldsymbol{\theta}}^0 \leq c(s)\hat{\sigma}\sqrt{g_p(s,x,p)^T \mathbf{P}^2 g_p(s,x,p)} + g_p(s,x,p)^T \boldsymbol{\theta}^0,\ \forall x \in (s,b]\} \\
& \qquad (3.53) \\
= \;& P_{k_0=s} \quad \left\{ \inf_{x\in[a,s)} \frac{g_p(s,x,p)^T(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)}{\hat{\sigma}\sqrt{g_p(s,x,p)^T \mathbf{P}^2 g_p(s,x,p)}} \geq -c(s), \right. \\
& \qquad \left. \sup_{x\in(s,b]} \frac{g_p(s,x,p)^T(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)}{\hat{\sigma}\sqrt{g_p(s,x,p)^T \mathbf{P}^2 g_p(s,x,p)}} \leq c(s) \right\} \\
= \;& P_{k_0=s} \quad \left\{ \inf_{x\in[a,s)} \frac{[\mathbf{P} g_p(s,x,p)]^T [\mathbf{P}^{-1}(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)]}{\hat{\sigma}\sqrt{g_p(s,x,p)^T \mathbf{P}^2 g_p(s,x,p)}} \geq -c(s), \right. \\
& \qquad \left. \sup_{x\in(s,b]} \frac{[\mathbf{P} g_p(s,x,p)]^T [\mathbf{P}^{-1}(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)]}{\hat{\sigma}\sqrt{g_p(s,x,p)^T \mathbf{P}^2 g_p(s,x,p)}} \leq c(s) \right\} \\
= \;& \quad P \quad \left\{ \inf_{x\in[a,s)} \frac{[\mathbf{P} g_p(s,x,p)]^T \boldsymbol{T}}{\| \mathbf{P} g_p(s,x,p) \|} \geq -c(s),\ \sup_{x\in(s,b]} \frac{[\mathbf{P} g_p(s,x,p)]^T \boldsymbol{T}}{\| \mathbf{P} g_p(s,x,p) \|} \leq c(s) \right\} \qquad (3.54)
\end{aligned}
$$

where Equation (3.52) follows directly from Equation (3.51), and Equation (3.53) follows from the fact that $g_p(s,x,p)^T \boldsymbol{\theta}^0 \geq 0$ for $x \in [a,s)$ and $g_p(s,x,p)^T \boldsymbol{\theta}^0 \leq 0$ for $x \in (s,b]$. Note that

$$
\begin{aligned}
& \inf_{\boldsymbol{\theta}:k_0=s} P_{k_0=s}\{\mathbf{Y} \in A(s)\} \\
= \;& P \left\{ \inf_{x\in[a,s)} \frac{[\mathbf{P} g_p(s,x,p)]^T \boldsymbol{T}}{\| \mathbf{P} g_p(s,x,p) \|} \geq -c(s), \quad \sup_{x\in(s,b]} \frac{[\mathbf{P} g_p(s,x,p)]^T \boldsymbol{T}}{\| \mathbf{P} g_p(s,x,p) \|} \leq c(s) \right\},
\end{aligned}
$$

with the infimum being attained at $\theta_1 = \theta_2 = \cdots = \theta_p = 0$. Then the critical value $c(s)$ is the unique solution of

$$
P \left\{ \inf_{x\in[a,s)} \frac{[\mathbf{P} g_p(s,x,p)]^T \boldsymbol{T}}{\| \mathbf{P} g_p(s,x,p) \|} \geq -c(s), \quad \sup_{x\in(s,b]} \frac{[\mathbf{P} g_p(s,x,p)]^T \boldsymbol{T}}{\| \mathbf{P} g_p(s,x,p) \|} \leq c(s) \right\} = 1 - \alpha. \quad (3.55)
$$

Having found $c(k_0)$ for each $k_0 \in [a,b]$, we summarize our $(1-\alpha)$ level acceptance set as

$$
A(k_0) = \left\{ \boldsymbol{Y} : \frac{k_0 - x}{|k_0 - x|} \hat{h}_p(k_0, x) \geq -c(k_0), \forall x \in [a,b] \setminus k_0 \right\}. \qquad (3.56)
$$

Then according to Neyman's Theorem, a $(1-\alpha)$ level confidence set for a maximum point, $k$, based on $\boldsymbol{Y}$ is given by

$$\begin{aligned}\mathbf{C}_E(\boldsymbol{Y}) &= \{k_0 \in [a,b] : \boldsymbol{Y} \in A(k_0)\} \\ &= \left\{k_0 \in [a,b] : \frac{k_0 - x}{|k_0 - x|}\hat{h}_p(k_0, x) \geq -c(k_0), \forall x \in [a,b] \setminus k_0\right\}.\end{aligned} \tag{3.57}$$

In other words, a point $k_0 \in [a,b]$ is in the set $\mathbf{C}_E(\boldsymbol{Y})$ if and only if for any $x \in [a,b] \setminus k_0$,

$$\frac{k_0 - x}{|k_0 - x|}\hat{h}_p(k_0, x) \geq -c(k_0).$$

The interval $[a,b]$ contains infinite many points and thus we can not check whether each point $k_0 \in [a,b]$ is in the confidence set. Hence, like in the quadratic case, we choose a finite grid of points on the interval $[a,b]$ with resolution $d$, that is, $\{a = s_1, s_2, \cdots, s_J = b\}$ with $s_i - s_{i-1} = d$. If $d$ is small, then the grid of points can give a fine approximation to the set $[a,b]$. We only check each point in the grid to see whether it is in the confidence set.

#### 3.1.3.2 Conservative Critical Constants

Recall that in Section 3.1.3.1, we need to find $c(k_0)$ for each grid point $k_0 \in [a, b]$ in order to decide the confidence set. However, the computation for each $c(k_0)$ is time consuming, since it requires a large number $m$ of replications to reach a certain accuracy level (the computation details are given later in Section 3.1.3.3). Similar to the quadratic case, we reduce the computational expense by using a conservative critical value, which reduces the number of points for which we must compute $c(k_0)$ from all the grid points $k_0$ in $[a, b]$ to only the grid points in the conservative confidence set $C_0(\boldsymbol{Y}^*)$.

**The $\boldsymbol{F}$ critical constants.** Note that $\boldsymbol{T} \sim \boldsymbol{T}_{p,n-p-1}$, therefore

$$\frac{\| \boldsymbol{T} \|^2}{p} \sim F_{p,n-p-1},$$

where $F_{p,n-p-1}$ is the standard F-distribution with degrees of freedom $p$ and $n-p-1$. Denote $f^{\alpha}_{p,n-p-1}$ as the upper $\alpha$ point of the distribution $F_{p,n-p-1}$, then we have

$$P\{\| \boldsymbol{T} \| \leq \sqrt{p f^{\alpha}_{p,n-p-1}}\} = 1 - \alpha.$$

Let $c_p = \sqrt{p f^{\alpha}_{p,n-p-1}}$, then we have

$$P\left\{\sup_{\boldsymbol{\rho} \in R^p} \frac{\boldsymbol{\rho}\boldsymbol{T}}{\| \boldsymbol{\rho} \|} \leq c_p\right\} = P\left\{\sup_{\boldsymbol{\rho} \in R^p} \frac{|\boldsymbol{\rho}\boldsymbol{T}|}{\| \boldsymbol{\rho} \|} \leq c_p\right\} = P\left\{\inf_{\boldsymbol{\rho} \in R^p} \frac{\boldsymbol{\rho}\boldsymbol{T}}{\| \boldsymbol{\rho} \|} \geq -c_p\right\} = 1 - \alpha.$$

Hence, we have

$$\begin{aligned} & P\left\{\sup_{x \in (-\infty,\infty)} \frac{[\mathbf{P} g_p(a,x,p)^T]^T \boldsymbol{T}}{\| \mathbf{P} g_p(a,x,p)^T \|} \leq c_p\right\} \\ \geq \;& P\left\{\sup_{\boldsymbol{\rho} \in R^p} \frac{\boldsymbol{\rho}\boldsymbol{T}}{\| \boldsymbol{\rho} \|} \leq -c_p\right\} \\ = \;& 1 - \alpha. \end{aligned} \tag{3.58}$$

As $(a, b] \subset (-\infty, \infty)$, comparing the probability statement (3.58) with the probability statement (3.45), we conclude that $c(a) < c_p$.

Similarly, we have

$$\begin{aligned}
& P\left\{\inf_{x\in(-\infty,\infty)} \frac{[\mathbf{P}g_p(b,x,p)^T]^T\boldsymbol{T}}{\parallel \mathbf{P}g_p(b,x,p)^T \parallel} \geq -c_p\right\} \\
\geq\ & P\left\{\inf_{\boldsymbol{\rho}\in R^p} \frac{\boldsymbol{\rho}\boldsymbol{T}}{\parallel \boldsymbol{\rho} \parallel} \geq -c_p\right\} \\
=\ & 1-\alpha. \qquad (3.59)
\end{aligned}$$

Comparing the probability statement (3.59) with the probability statement (3.50), we conclude that $c(b) < c_p$.

Finally, for $s \in (a,b)$, we have

$$\begin{aligned}
& P\left\{\sup_{x\in(-\infty,\infty)} \frac{[\mathbf{P}g_p(s,x,p)^T]^T\boldsymbol{T}}{\parallel \mathbf{P}g_p(s,x,p)^T \parallel} \leq c_p, \inf_{x\in(-\infty,\infty)} \frac{[\mathbf{P}g_p(s,x,p)^T]^T\boldsymbol{T}}{\parallel \mathbf{P}g_p(s,x,p)^T \parallel} \geq -c_p\right\} \\
=\ & P\left\{\sup_{x\in(-\infty,\infty)} \frac{|[\mathbf{P}g_p(s,x,p)^T]^T\boldsymbol{T}|}{\parallel \mathbf{P}g_p(s,x,p)^T \parallel} \leq c_p\right\} \\
\geq\ & P\left\{\sup_{\boldsymbol{\rho}\in R^p} \frac{|\boldsymbol{\rho}\boldsymbol{T}|}{\parallel \boldsymbol{\rho} \parallel} \leq c_p\right\} \\
=\ & 1-\alpha. \qquad (3.60)
\end{aligned}$$

Comparing the probability statement (3.60) with the probability statement (3.55), we conclude that $c(s) < c_p$ for all $s \in (a,b)$.

Thus, substituting $\sqrt{pf^{\alpha}_{p,n-p-1}}$ for $c(k_0)$ in Equation (3.57), the conservative confidence set is given by

$$\mathbf{C}_0(\boldsymbol{Y}) = \left\{k_0 \in [a,b] : \frac{k_0 - x}{|k_0 - x|}\hat{h}_p(k_0,x) \geq -c_p, \forall x \in [a,b] \setminus k_0\right\}. \qquad (3.61)$$

Instead of checking all the grid points in [a,b], we only check the grid points in the conservative confidence set to determine the exact confidence set and thus get significant savings in computing time.

**An improved conservative critical constant.** Similarly to the univariate quadratic case, an improved conservative critical constant $c_h$ is solved from

$$\frac{1}{2}F_{p,n-p-1}\left(\frac{c_h^2}{p}\right) + \frac{1}{2}F_{p-1,n-p-1}\left(\frac{c_h^2}{p-1}\right) = 1-\alpha.$$

The proof is as follows.

First, we show that there exists at least one vector $\boldsymbol{u} = (u_1, u_2, \cdots, u_p)^T$ such that

$\boldsymbol{u}^T g_p(a, x, p) > 0$ for all $x \in (-\infty, \infty)$. Note that

$$\begin{aligned} & \boldsymbol{u}^T g_p(a, x, p) \\ = \;& \boldsymbol{u}^T [1, a + x, a^2 + ax + x^2, \cdots, a^{p-1} + a^{p-2}x + \cdots + ax^{p-2} + x^{p-1}]^T \\ = \;& u_1 + u_2(a + x) + u_3(a^2 + ax + x^2) + \cdots + u_p(a^{p-1} + a^{p-2}x + \cdots + ax^{p-2} + x^{p-1}) \\ = \;& (u_1 + u_2 a + u_3 a^2 + \cdots + u_p a^{p-1}) + (u_2 + u_3 a + \cdots + u_p a^{p-2})x + \cdots + u_p x^{p-1}, \end{aligned}$$

and the last expression is a polynomial function of $x$. Therefore, a sufficient condition for a $\boldsymbol{u}$ satisfying $\boldsymbol{u}^T g_p(a, x, p) > 0$ for all $x \in (-\infty, \infty)$ is $\begin{cases} u_2 = 0, \cdots, u_p = 0 \\ u_1 > 0. \end{cases}$ We choose $u_1 = 1$ and so $\boldsymbol{u} = [1, 0, 0, \ldots, 0]^T$. Let $\boldsymbol{v} = \mathbf{P}^{-1}\boldsymbol{u}$, hence $\boldsymbol{v}^T \mathbf{P} g_p(a, x, p) > 0$ for all $x \in (-\infty, \infty)$. Therefore, all $\mathbf{P} g_p(a, x, p)$ are in the same side of a plane (hyperplane) that is perpendicular to $\boldsymbol{v}$ through $\mathbf{0}$. Note that all the possible values of the random vector $\boldsymbol{T} \sim \boldsymbol{T}_{p,n-p-1}$ such that

$$\sup_{x \in (a,b]} \frac{[\mathbf{P} g_p(a, x, p)^T]^T \boldsymbol{T}}{\| \mathbf{P} g_p(a, x, p)^T \|} \leq c(a)$$

must fall in the region bounded by all the planes (hyperplanes) that are perpendicular to $\mathbf{P} g_p(a, x, p)$, $x \in (a, b]$, and $c(a)$ distance away from the origin in the direction of $\mathbf{P} g_p(a, x, p)$. We define this region to be $R(a, p, c(a))$. Clearly $R(a, p, c(a))$ is larger than the region bounded by all the planes that are perpendicular to $\{\boldsymbol{w} : \boldsymbol{v}^T \boldsymbol{w} > 0\}$ and $c(a)$ distance away in the direction of $\boldsymbol{w}$, denoted by $R_H(\boldsymbol{v}, p, c(a))$, which is actually a cylinder of radius $c(a)$ connected to a semisphere of the same radius. Hence if $P\{\boldsymbol{T} \in R(a, p, c(a))\} = P\{\boldsymbol{T} \in R_H(\boldsymbol{v}, p, c_h)\}$ then the critical constant $c_h$ must be larger than $c(a)$.

Next, we show that there exists at least one vector $\boldsymbol{u}$, such that $\boldsymbol{u}^T g_p(s, x, p) > 0$ for all $x < s$ and $\boldsymbol{u}^T g_p(s, x, p) < 0$ for all $x > s$. Note that

$$\begin{aligned} \boldsymbol{u}^T g_p(s, x, p) \;=\;& (u_1 + u_2 s + u_3 s^2 + \cdots + u_p s^{p-1}) + (u_2 + u_3 s + \cdots + u_p s^{p-2})x \\ & + \cdots + u_p x^{p-1}, \end{aligned}$$

and the last expression is a polynomial function of $x$. Therefore, a sufficient condition for a $\boldsymbol{u}$ satisfying $\boldsymbol{u}^T g_p(s,x,p) > 0, \forall x < s$ and $\boldsymbol{u}^T g_p(s,x,p) < 0, \forall x > s$ is

$$\begin{cases} u_2 < 0, u_3 = 0, \cdots, u_p = 0 \\ u_1 + u_2 s + u_2 s = 0 \end{cases}.$$

We choose $u_2 = -1$ to give $u_1 = 2s$ and so $\boldsymbol{u} = [2s, -1, 0, 0, \ldots, 0]^T$. Recall that $\boldsymbol{v} = \mathbf{P}^{-1}\boldsymbol{u}$, we have

$$\begin{cases} \boldsymbol{v}^T \mathbf{P} g_p(s,x,p) > 0, \ \forall x \in (-\infty, s) \\ \boldsymbol{v}^T \mathbf{P} g_p(s,x,p) < 0, \ \forall x \in (s, +\infty) \end{cases}. \tag{3.62}$$

Therefore, $\mathbf{P}g_p(s,x,p), \forall x \in [a,s)$ and $-\mathbf{P}g_p(s,x,p), \forall x \in (s,b]$ are in the same side of a plane (hyperplane) that is perpendicular to $\boldsymbol{v}$. Note that, all the possible values of the random vector $\boldsymbol{T} \sim \boldsymbol{T}_{p,n-p-1}$ that satisfy

$$\inf_{x\in[a,s)} \frac{[\mathbf{P}g_p(s,x,p)^T]^T \boldsymbol{T}}{\| \mathbf{P}g_p(s,x,p)^T \|} \geq -c(s), \ \sup_{x\in(s,b]} \frac{[\mathbf{P}g_p(s,x,p)^T]^T \boldsymbol{T}}{\| \mathbf{P}g_p(s,x,p)^T \|} \leq c(s)$$

must fall in the region bounded by all the planes (hyperplanes) each with a condition either (1) perpendicular to $\mathbf{P}g_p(s,x,p)$ and $c(s)$ distance away from the origin in the direction of $\mathbf{P}g_p(s,x,p)$ if $x \in [a,s)$, or (2) perpendicular to $\mathbf{P}g_p(s,x,p)$ and $c(s)$ distance away from the origin in the opposite direction of $\mathbf{P}g_p(s,x,p)$ if $x \in (s,b]$. We define this region to be $R(s,p,c(s))$. Clearly $R(s,p,c(s))$ is larger than $R_H(\boldsymbol{v},p,c(s))$. Hence if $P\{\boldsymbol{T} \in R(s,p,c(s))\} = P\{\boldsymbol{T} \in R_H(\boldsymbol{v},p,c_h)\}$ then the critical constant $c_h$ must be larger than $c(s)$. The same conclusion is reached by Hochberg and Quade (1975), but using a different approach.

Hence, $c_h$ is a conservative critical constant and we use it to construct a conservative confidence set for a maximum point

$$C_h(\boldsymbol{Y}) = \left\{ k_0 \in [a,b] : \frac{k_0 - x}{|k_0 - x|} \hat{h}_p(k_0, x) \geq -c_h, \forall x \in [a,b] \setminus k_0 \right\}. \tag{3.63}$$

Note that $c_h$ is less conservative than $c_p$, thus the confidence set $C_h(\boldsymbol{Y})$ contains less grid points than the one associated with $c_p$, $\mathbf{C}_0(\boldsymbol{Y})$, and so is more efficient in the construction of $\mathbf{C}_E(\boldsymbol{Y})$.

#### 3.1.3.3 Computation

Similar to the computation of the confidence set for a maximum point of a quadratic regression function, we first present the direct approach by using simulated critical values. Then in order to reduce the complexity of computation, we present the conservative critical value method to reduce the possible points that could be in the confidence set. We also present the p-value method here, but in this case, the p-value method takes approximately the same computation time as the direct approach.

**Simulating critical values.** Since the critical value $c(k_0)$ for $k_0 \in [a, b]$ is difficult to solve from Equations (3.45), (3.50) and (3.55), we use simulation to find an approximation to $c(k_0)$.

Let

$$G_a(\boldsymbol{T}) = \sup_{x \in (a,b]} \frac{[\mathbf{P} g_p(a,x,p)^T]^T \boldsymbol{T}}{\| \mathbf{P} g_p(a,x,p)^T \|},$$

$$G_b(\boldsymbol{T}) = -\inf_{x \in [a,b)} \frac{[\mathbf{P} g_p(b,x,p)^T]^T \boldsymbol{T}}{\| \mathbf{P} g_p(b,x,p)^T \|},$$

and for each $s \in (a, b)$

$$G_s(s, \boldsymbol{T}) = \max \left( -\inf_{x \in [a,s)} \frac{[\mathbf{P} g_p(s,x,p)^T]^T \boldsymbol{T}}{\| \mathbf{P} g_p(s,x,p)^T \|}, \sup_{x \in (s,b]} \frac{[\mathbf{P} g_p(s,x,p)^T]^T \boldsymbol{T}}{\| \mathbf{P} g_p(s,x,p)^T \|} \right).$$

Then from Equations (3.45), (3.50) and (3.55), respectively, we have

$$P\{G_a(\boldsymbol{T}) \le c(a)\} = 1 - \alpha,$$

$$P\{G_b(\boldsymbol{T}) \le c(b)\} = 1 - \alpha$$

and

$$P\{G_s(s, \boldsymbol{T}) \le c(s)\} = 1 - \alpha.$$

Here we only illustrate the simulation method for finding the critical value $c(a)$. The critical constants $c(b)$ and $c(s)$ are found in a similar way.

**Step 1.** Simulate independent $\boldsymbol{T}_i \sim \boldsymbol{T}_{p,n-p-1}$, $i = 1, 2, \cdots, m$.

**Step 2.** For each $\boldsymbol{T}_i$, $i = 1, 2, \cdots, m$, we compute $G_a(\boldsymbol{T}_i)$. From the definition of

$G_a(\boldsymbol{T}_i)$, the supremum can only be attained at the end points $a$, $b$, and the stationary points in the interval $(a, b)$ of

$$m_i(x) = \frac{[\mathbf{P}g_p(a,x,p)^T]^T \boldsymbol{T}_i}{\| \mathbf{P}g_p(a,x,p)^T \|}.$$

Let $U_i(x) = [\mathbf{P}g_p(a,x,p)^T]^T \boldsymbol{T}_i$ and $V_i(x) = \| \mathbf{P}g_p(a,x,p)^T \|^2$, then the stationary points satisfy

$$\frac{dm_i(x)}{dx} = \frac{d}{dx}\left[\frac{U_i(x)}{\sqrt{V_i(x)}}\right] = 0,$$

or

$$\frac{dU_i(x)}{dx}V_i(x) - \frac{1}{2}U_i(x)\frac{dV_i(x)}{dx} = 0$$

since

$$\begin{aligned}\frac{d}{dx}\left[\frac{U_i(x)}{\sqrt{V_i(x)}}\right] &= \left[\frac{dU_i(x)}{dx}\sqrt{V_i(x)} - U_i(x)\frac{d\sqrt{V_i(x)}}{dx}\right]V_i^{-1}(x) \\ &= \left[\frac{dU_i(x)}{dx}V_i(x) - \frac{1}{2}U_i(x)\frac{dV_i(x)}{dx}\right]V_i^{-\frac{3}{2}}(x).\end{aligned}$$

Note that $U_i(x)$, $V_i(x)$, $dU_i(x)/dx$ and $dV_i(x)/dx$ are all polynomial functions with orders $(p-1)$, $(2p-2)$, $(p-2)$ and $(2p-3)$, respectively, so $dm_i(x)/dx$ is of order $(3p-5)$. Therefore, there are $(3p-5)$ roots, which we find numerically by using MATLAB. The stationary points of $m_i(x)$ in the interval $[a, b]$ are the real roots among the $(3p-5)$ roots in $[a, b]$, and are denoted by $r_1, \cdots, r_q$, with $0 \le q \le (3p-5)$. It follows that

$$G_a(\boldsymbol{T}_i) = \max\{m_i(a), m_i(b), m_i(r_1), \cdots, m_i(r_q)\}.$$

**Step 3.** Use the $[(1-\alpha)\times m]th$ largest value of $G_a(\boldsymbol{T}_i)$, $i = 1, 2, \cdots, m$, as the approximation of $c(a)$.

In this way, we simulate $c(k_0)$ for each $k_0 \in [a, b]$ and then determine the confidence set according to Equation (3.57).

**The P-Value Method.** Instead of computing the critical values, we can use the p-values to construct confidence sets similar to that in the quadratic case. However, simulation is still necessary for computing the p-value.

Let

$$\hat{h}^*_p(z_1, z_2) := \frac{g_p(z_1, z_2, p)^T \hat{\boldsymbol{\theta}}^{*0}}{\hat{\sigma}^* \sqrt{g_p(z_1, z_2, p)^T \mathbf{P}^2 g_p(z_1, z_2, p)}}$$

where $\hat{\boldsymbol{\theta}}^{*0}$ and $\hat{\sigma}^*$ are the estimates of $\boldsymbol{\theta}^0$ and $\sigma$ based on the observation $\boldsymbol{Y}^*$. Next, we discuss the computation of the p-value separately for each of the three cases: $k_0 = a$, $k_0 = b$ and $k_0 \in (a, b)$.

**Case I.** Recall that in section 3.1.3.1, our $(1-\alpha)$ level acceptance set $A(a)$ is given by

$$A(a) \quad = \quad \{\boldsymbol{Y} : \sup_{x \in (a,b]} \hat{h}_p(a, x) \leq c(a)\}.$$

We define

$$W^{(p)}_a(\boldsymbol{Y}) := \sup_{x \in (a,b]} \hat{h}_p(a, x)$$

and

$$w^{(p)}_a := W_a(\boldsymbol{Y}^*) = \sup_{x \in (a,b]} \hat{h}^*_p(a, x).$$

Note that $W^{(p)}_a(\boldsymbol{Y})$ is a random variable, but $w^{(p)}_a$ is the observed value of $W^{(p)}_a(\boldsymbol{Y})$ based on $\boldsymbol{Y}^*$. Define

$$p^{(p)}_a \quad = \quad 1 - \inf_{\boldsymbol{\theta}: k_0 = a} P\{W^{(p)}_a(\boldsymbol{Y}) \leq w^{(p)}_a\}.$$

As

$$\begin{aligned} a \in \mathbf{C}_E(\boldsymbol{Y}^*) \quad &\Leftrightarrow \quad \boldsymbol{Y}^* \in A(a) \\ &\Leftrightarrow \quad w^{(p)}_a \leq c(a) \\ &\Leftrightarrow \quad \inf_{\boldsymbol{\theta}: k_0 = a} P\{W^{(p)}_a(\boldsymbol{Y}) \leq w^{(p)}_a\} \leq 1 - \alpha, \end{aligned}$$

we have

$$\begin{aligned} a \notin \mathbf{C}_E(\boldsymbol{Y}^*) \quad &\Leftrightarrow \quad \inf_{\boldsymbol{\theta}: k_0 = a} P\{W^{(p)}_a(\boldsymbol{Y}) \leq w^{(p)}_a\} > 1 - \alpha \\ &\Leftrightarrow \quad 1 - \inf_{\boldsymbol{\theta}: k_0 = a} P\{W^{(p)}_a(\boldsymbol{Y}) \leq w^{(p)}_a\} < \alpha \\ &\Leftrightarrow \quad p^{(p)}_a < \alpha. \end{aligned}$$

It is clear that $p_a^{(p)}$ is a p-value for testing $H_0 : k_0 = a$. From the definition of $G_a(\boldsymbol{T})$, we have

$$\begin{aligned} p_a^{(p)} &= 1 - \inf_{\boldsymbol{\theta}:k_0=a} P_{k_0=a}\{W_a^{(p)}(\boldsymbol{Y}) \le w_a^{(p)}\} \\ &= 1 - P\left\{\sup_{x\in(a,b]} \frac{[\mathbf{P}g_p(a,x,p)^T]^T\boldsymbol{T}}{\parallel \mathbf{P}g_p(a,x,p)^T \parallel} \le w_a^{(p)}\right\} \\ &= P\left\{G_a(\boldsymbol{T}) > w_a^{(p)}\right\} \end{aligned}$$

where $\boldsymbol{T} \sim \boldsymbol{T}_{p,n-p-1}$. Next, we employ the following steps to simulate the p-value $p_a^{(p)}$.

**Step 1.** Sample independent $\boldsymbol{T}_i \sim \boldsymbol{T}_{p,n-p-1}$, $i = 1, 2, \cdots, m$.

**Step 2.** For each $\boldsymbol{T}_i$, $i = 1, 2, \cdots, m$, we compute $G_a(\boldsymbol{T}_i)$ as we did earlier in this section.

**Step 3.** Compare $G_a(\boldsymbol{T}_i)$ with $w_a^{(p)}$ for each $i$, and let $n_a$ denote the number of $G_a(\boldsymbol{T}_i)$ that is larger than $w_a^{(p)}$. Then, we use $n_a/m$ as an approximation to $p_a^{(p)}$.

**Case II.** Recall that in Section 3.1.3.1, the $(1-\alpha)$ level acceptance set $A(b)$ is given by

$$A(b) = \{\boldsymbol{Y} : \inf_{x\in[a,b)} \hat{h}_p(b,x) \ge -c(b)\}.$$

We define

$$W_b^{(p)}(\boldsymbol{Y}) := \inf_{x\in[a,b)} \hat{h}_p(b,x)$$

and

$$w_b^{(p)} := W_b(\boldsymbol{Y}^*) = \inf_{x\in[a,b)} \hat{h}_p^*(b,x).$$

Note that $W_b^{(p)}(\boldsymbol{Y})$ is a random variable, but $w_b^{(p)}$ is the observed value of $W_b^{(p)}(\boldsymbol{Y})$ when $\boldsymbol{Y} = \boldsymbol{Y}^*$. Define

$$p_b^{(p)} = 1 - \inf_{\boldsymbol{\theta}:k_0=b} P\{W_b^{(p)}(\boldsymbol{Y}) \ge w_b^{(p)}\}.$$

As

$$\begin{aligned} b \in \mathbf{C}_E(\boldsymbol{Y}^*) &\Leftrightarrow \boldsymbol{Y}^* \in A(b) \\ &\Leftrightarrow w_a^{(p)} \ge -c(b) \\ &\Leftrightarrow \inf_{\boldsymbol{\theta}:k_0=b} P\{W_b^{(p)}(\boldsymbol{Y}) \ge w_b^{(p)}\} \le 1-\alpha, \end{aligned}$$

we have

$$\begin{aligned} b \notin \mathbf{C}_E(\boldsymbol{Y}^*) &\Leftrightarrow \inf_{\boldsymbol{\theta}:k_0=b} P\{W_b^{(p)}(\boldsymbol{Y}) \geq w_b^{(p)}\} > 1-\alpha \\ &\Leftrightarrow 1 - \inf_{\boldsymbol{\theta}:k_0=b} P\{W_b^{(p)}(\boldsymbol{Y}) \geq w_b^{(p)}\} < \alpha \\ &\Leftrightarrow p_b^{(p)} < \alpha. \end{aligned}$$

It is clear that $p_b^{(p)}$ is a p-value for testing $H_0 : k_0 = b$. From the definition of $G_b(\boldsymbol{T})$ we have

$$\begin{aligned} p_b^{(p)} &= 1 - \inf_{\boldsymbol{\theta}:k_0=b} P_{k_0=b}\{W_b^{(p)}(\boldsymbol{Y}) \geq w_b^{(p)}\} \\ &= 1 - P_{k_0=b}\left\{ \inf_{x\in[a,b)} \frac{[\mathbf{P} g_p(b,x,p)^T]^T \boldsymbol{T}}{\| \mathbf{P} g_p(b,x,p)^T \|} \geq w_b^{(p)} \right\} \\ &= P\left\{ G_b(\boldsymbol{T}) > -w_b^{(p)} \right\} \end{aligned}$$

where $\boldsymbol{T} \sim \boldsymbol{T}_{p,n-p-1}$. Then, we employ the following steps to simulate the p-value $p_b^{(p)}$.

**Step 1.** Simulate independent $\boldsymbol{T}_i \sim \boldsymbol{T}_{p,n-p-1}$, $i = 1, 2, \cdots, m$.

**Step 2.** For each $\boldsymbol{T}_i$, $i = 1, 2, \cdots, m$, we compute $G_b(\boldsymbol{T}_i)$.

**Step 3.** Compare $G_b(\boldsymbol{T}_i)$ with $-w_b^{(p)}$ for each $i$, and let $n_b$ denote the number of $G_b(\boldsymbol{T}_i)$ that is larger than $-w_b^{(p)}$. Then, we use $n_b/m$ as the approximation to $p_b^{(p)}$.

**Case III.** Recall that in Section 3.1.3.1, for $s \in (a, b)$, our $(1-\alpha)$ level acceptance set $A(s)$ is given by

$$\begin{aligned} A(s) &= \{\boldsymbol{Y} : \hat{h}_p(s,x) \geq -c(s)\ \forall x \in [a,s),\ \hat{h}_p(s,x) \leq c(s)\ \forall x \in (s,b]\} \\ &= \{\boldsymbol{Y} : \frac{s-x}{|s-x|}\hat{h}_p(s,x) \geq -c(s), \forall x \in [a,s) \cup (s,b]\}. \end{aligned}$$

Let

$$W_s^{(p)}(\boldsymbol{Y}, s) = \inf_{x\in[a,s)\cup(s,b]} \frac{s-x}{| s-x |}\hat{h}_p(s,x),$$

$$w_s^{(p)}(s) = W_s(\boldsymbol{Y}^*, s) = \inf_{x\in[a,s)\cup(s,b]} \frac{s-x}{| s-x |}\hat{h}_p^*(s,x),$$

and

$$p_s^{(p)}(s) = 1 - \inf_{\boldsymbol{\theta}:k_0=s} P_{k_0=s}\{W_s^{(p)}(\boldsymbol{Y}, s) \geq w_s^{(p)}(s)\}.$$

Because

$$
\begin{aligned}
s \in \mathbf{C}_E(\boldsymbol{Y}^*) &\Leftrightarrow \boldsymbol{Y}^* \notin A(s) \\
&\Leftrightarrow w_s^{(p)}(s) \geq -c(s) \\
&\Leftrightarrow \inf_{\boldsymbol{\theta}: k_0 = s} P_{k_0=s}\{W_s^{(p)}(\boldsymbol{Y}, s) \geq w_s^{(p)}(s)\} \leq 1-\alpha,
\end{aligned}
$$

we have

$$
\begin{aligned}
s \notin \mathbf{C}_E(\boldsymbol{Y}^*) &\Leftrightarrow \inf_{\boldsymbol{\theta}: k_0 = s} P_{k_0=s}\{W_s^{(p)}(\boldsymbol{Y}, s) \geq w_s^{(p)}(s)\} > 1-\alpha \\
&\Leftrightarrow p_s^{(p)}(s) < \alpha.
\end{aligned}
$$

Thus, $p_s^{(p)}(s)$ is a p-value for testing $H_0 : k_0 = s$. We have

$$
\begin{aligned}
p_s^{(p)}(s) &= 1 - \inf_{\boldsymbol{\theta}: k_0 = s} P_{k_0=s}(W_s^{(p)}(\boldsymbol{Y}, s) \geq w_s^{(p)}(s)) \\
&= 1 - \inf_{\boldsymbol{\theta}: k_0 = s} P_{k_0=s}\left\{\inf_{x \in [a,s)\cup(s,b]} \hat{h}(s+x)\frac{s-x}{\mid s-x \mid} \geq w_s^{(p)}(s)\right\} \\
&= 1 - P\left\{\inf_{x \in [a,s)} \frac{[\mathbf{P}g_p(s,x,p)^T]^T\ \boldsymbol{T}}{\parallel \mathbf{P}g_p(s,x,p)^T \parallel} \geq w_s^{(p)}(s),\ \sup_{x \in (s,b]} \frac{[\mathbf{P}g_p(s,x,p)^T]^T \boldsymbol{T}}{\parallel \mathbf{P}g_p(s,x,p)^T \parallel} \leq -w_s^{(p)}(s)\right\} \\
&= P\left\{G_s(s, \boldsymbol{T}) \geq -w_s^{(p)}(s)\right\}
\end{aligned}
$$

We reject $H_0 : k_0 = s$ if and only if $p_s^{(p)}(s) < \alpha$. Then, we employ the following steps to simulate the p-value $p_s^{(p)}(s)$.

**Step 1.** Simulate independent $\boldsymbol{T}_i \sim \boldsymbol{T}_{p,n-p-1}$, $i = 1, 2, \cdots, m$.

**Step 2.** For each $\boldsymbol{T}_i$, $i = 1, 2, \cdots, m$, we compute $G_s(s, \boldsymbol{T}_i)$.

**Step 3.** Compare $G_s(s, \boldsymbol{T}_i)$ with $-w_s^{(p)}(s)$ for each $i$, and let $n_s(s)$ denote the number of $G_s(s, \boldsymbol{T}_i)$ that is larger than $-w_s^{(p)}(s)$. Then, we use $n_s(s)/m$ as an approximation to $p_s^{(p)}(s)$.

#### 3.1.3.4 Examples

Table 3.11: Perinatal mortality data for black infants (Selvin, 1998)

| $i$ | $x_i$ | $Y_i$ | $i$ | $x_i$ | $Y_i$ | $i$ | $x_i$ | $Y_i$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.85 | -0.3556 | 13 | 2.05 | 1.0972 | 25 | 3.25 | 1.7774 |
| 2 | 0.95 | 0.1089 | 14 | 2.15 | 1.3382 | 26 | 3.35 | 1.7538 |
| 3 | 1.05 | 0.3880 | 15 | 2.25 | 1.3254 | 27 | 3.45 | 2.0933 |
| 4 | 1.15 | 0.4399 | 16 | 2.35 | 1.4241 | 28 | 3.55 | 1.7594 |
| 5 | 1.25 | 0.6513 | 17 | 2.45 | 1.4632 | 29 | 3.65 | 1.7538 |
| 6 | 1.35 | 0.7022 | 18 | 2.55 | 1.4906 | 30 | 3.75 | 1.9478 |
| 7 | 1.45 | 0.7706 | 19 | 2.65 | 1.6324 | 31 | 3.85 | 1.8351 |
| 8 | 1.55 | 0.7523 | 20 | 2.75 | 1.5383 | 32 | 3.95 | 1.9830 |
| 9 | 1.65 | 0.7934 | 21 | 2.85 | 1.6955 | 33 | 4.05 | 1.7429 |
| 10 | 1.75 | 1.0233 | 22 | 2.95 | 1.7538 | 34 | 4.15 | 1.8827 |
| 11 | 1.85 | 0.8918 | 23 | 3.05 | 1.6998 | 35 | 4.25 | 1.8269 |
| 12 | 1.95 | 1.0959 | 24 | 3.15 | 1.7903 | | | |

**Example 1.** The data of perinatal mortality rate (PMR) and birth weight (BW) is given by Selvin (1998). Selvin considered fitting a 4th order polynomial regression model between $Y = log(-log(PMR))$ and $x = BW$. Table 3.11 shows the data $(x, Y)$ for 35 black infants. The fitted 4th order polynomial regression function is

$$\hat{Y} = -2.861 + 4.809x - 2.316x^2 + 0.568x^3 - 0.054x^4,$$

and $\hat{\sigma} = 0.108$.

From the normal probability plot (Figure 3.19) and the residual plots (Figure 3.20), the normality assumption of the error seems reasonable. The exploratory index $R^2 = 0.909$, so the 4th order polynomial model fit the data well. To construct a 95% level confidence set for a maximum point, we check the grid points from 0.85 to 4.25

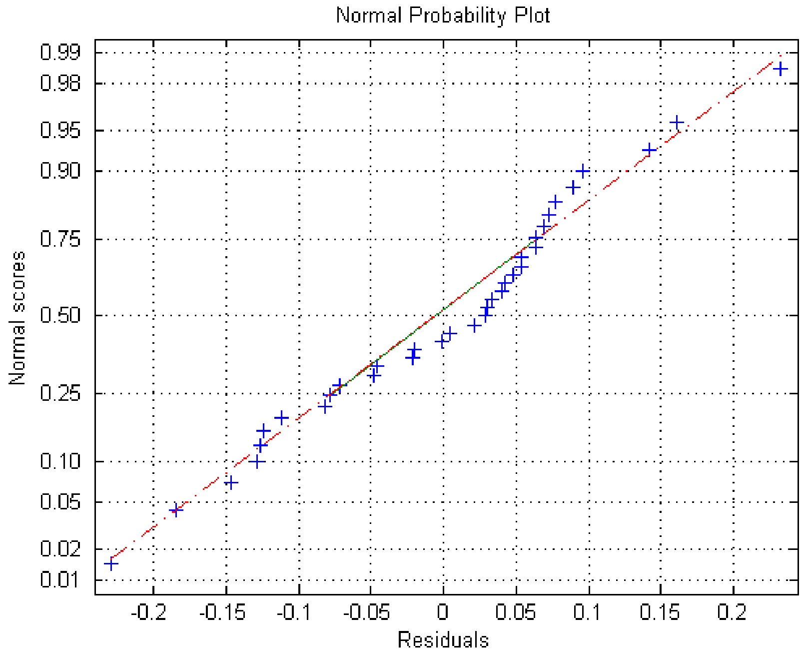



Figure 3.19: Normal probability plot of black infant data

with resolution $d = 0.01$. A conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ using its conservative critical value $c_2 = 3.1424$ is given by $[3.60, 4.25]$. Then, we compute the critical values (based on $10,000$ simulations) for the grid points in $\mathbf{C}_0(\boldsymbol{Y})$ and decide whether each grid point is in the exact confidence set $\mathbf{C}_E(\boldsymbol{Y})$. The set $\mathbf{C}_E(\boldsymbol{Y})$ is given by $[3.62, 4.25]$ and the computation takes 338 seconds. Then we use $100,000$ simulations to calculate the confidence set again, we still get $[3.62, 4.25]$. Computation of this new $\mathbf{C}_E(\boldsymbol{Y})$ takes 3369 seconds. It seems $10,000$ simulations in this case is sufficient. The bootstrap confidence set and Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$ are given by $[3.65, 4.11]$ and $[3.58, 4.25]$, respectively. The data points, fitted regression curve and the 95% confidence sets are plotted in Figure 3.21. The horizontal line segments, from top to bottom, represent $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, $\mathbf{C}_c(\boldsymbol{Y})$ and bootstrap confidence set.

**Example 2.** Table 3.12 shows the data $Y = log(-log(PMR))$ and $x = BW$ for 35 white infants. The fitted 4th order polynomial regression function is

$$\hat{Y} = -2.842 + 4.179x - 1.803x^2 + 0.421x^3 - 0.039x^4,$$

Table 3.12: Perinatal mortality data for white infants (Selvin, 1998)

| $i$ | $x_i$ | $Y_i$ | $i$ | $x_i$ | $Y_i$ | $i$ | $x_i$ | $Y_i$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.85 | -0.4761 | 13 | 2.05 | 1.1204 | 25 | 3.25 | 1.7429 |
| 2 | 0.95 | -0.1950 | 14 | 2.15 | 1.0919 | 26 | 3.35 | 1.8114 |
| 3 | 1.05 | 0.0849 | 15 | 2.25 | 1.2771 | 27 | 3.45 | 1.8269 |
| 4 | 1.15 | 0.2464 | 16 | 2.35 | 1.2771 | 28 | 3.55 | 1.8351 |
| 5 | 1.25 | 0.3791 | 17 | 2.45 | 1.3731 | 29 | 3.65 | 1.8351 |
| 6 | 1.35 | 0.4715 | 18 | 2.55 | 1.4241 | 30 | 3.75 | 1.8437 |
| 7 | 1.45 | 0.5364 | 19 | 2.65 | 1.4775 | 31 | 3.85 | 1.8527 |
| 8 | 1.55 | 0.6340 | 20 | 2.75 | 1.5165 | 32 | 3.95 | 1.8722 |
| 9 | 1.65 | 0.7391 | 21 | 2.85 | 1.6018 | 33 | 4.05 | 1.8437 |
| 10 | 1.75 | 0.7551 | 22 | 2.95 | 1.6751 | 34 | 4.15 | 1.8939 |
| 11 | 1.85 | 0.8042 | 23 | 3.05 | 1.6830 | 35 | 4.25 | 1.8527 |
| 12 | 1.95 | 0.9128 | 24 | 3.15 | 1.7429 | | | |

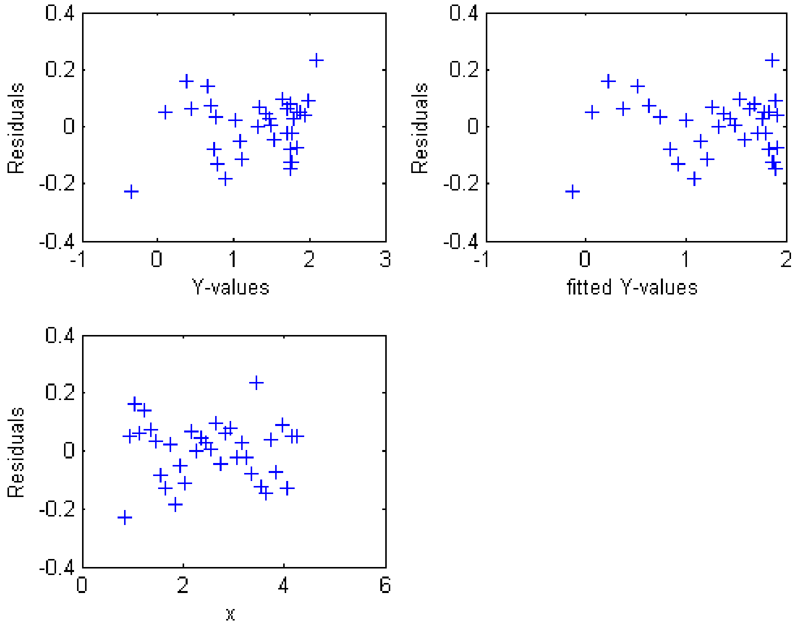



Figure 3.20: Residual plot of black infant data

and $\hat{\sigma} = 0.058$.

From the normal probability plot (Figure 3.22) and the residual plots (Figure 3.23), the normality assumption of the error seems reasonable. The exploratory index $R^2 = 0.994$, so the 4th order polynomial model fit the data well. To construct a 95% level confidence set for a maximum point, we check the grid points from 0.85 to 4.25 with resolution $d = 0.01$. A conservative confidence set using its conservative critical value $c_2 = 3.1424$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = [3.75, 4.20]$. Then, we compute the critical values (based on $100,000$ simulations) for the grid points in $\mathbf{C}_0(\boldsymbol{Y})$ and decide whether each grid point is in the exact confidence set $\mathbf{C}_E(\boldsymbol{Y})$. The set $\mathbf{C}_E(\boldsymbol{Y})$ is given by $[3.76, 4.19]$ and the computation takes 2395 seconds. The bootstrap confidence set and Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$ are given by $[3.76, 4.05]$ and $[3.72, 4.24]$, respectively. The data points, fitted regression curve and the 95% confidence sets are plotted in Figure 3.24. The horizontal line segments, from top to bottom, represent $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, $\mathbf{C}_c(\boldsymbol{Y})$ and bootstrap confidence set.

**Example 3.** Table 3.13 shows the data $Y = log(-log(PMR))$ and $x = BW$ for

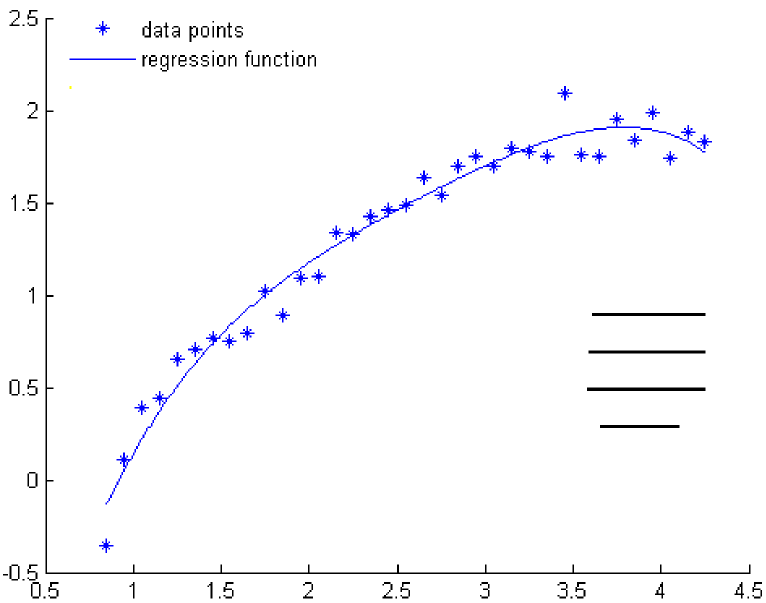



Figure 3.21: The 95% level confidence sets for a maximum point based on black infant data. The horizontal line segments, from top to bottom, represent $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, $\mathbf{C}_c(\boldsymbol{Y})$ and bootstrap confidence set.

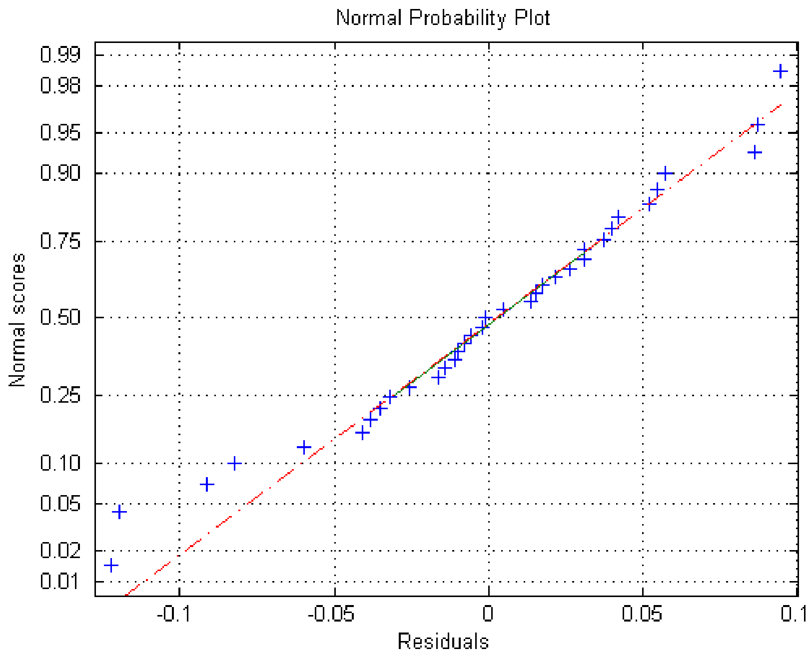



Figure 3.22: Normal probability plot of white infant data

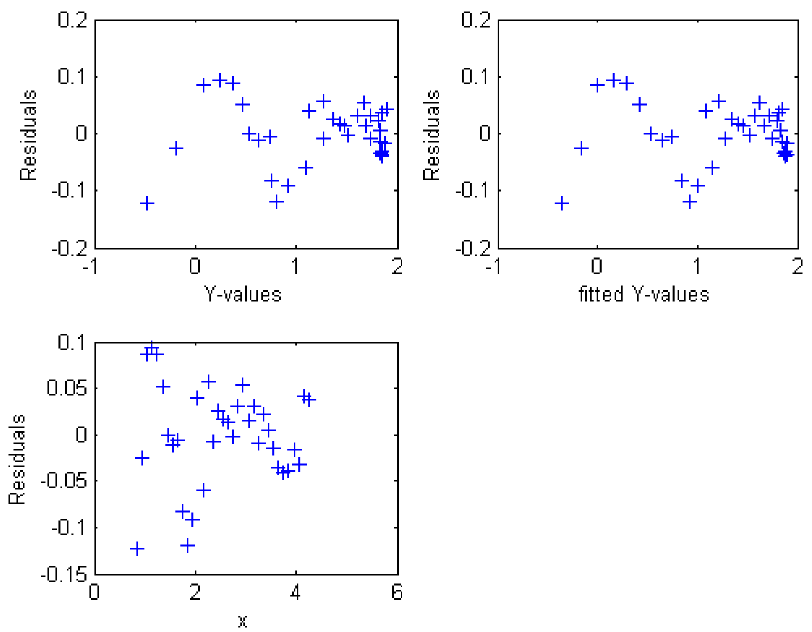



Figure 3.23: Residual plot of white infant data

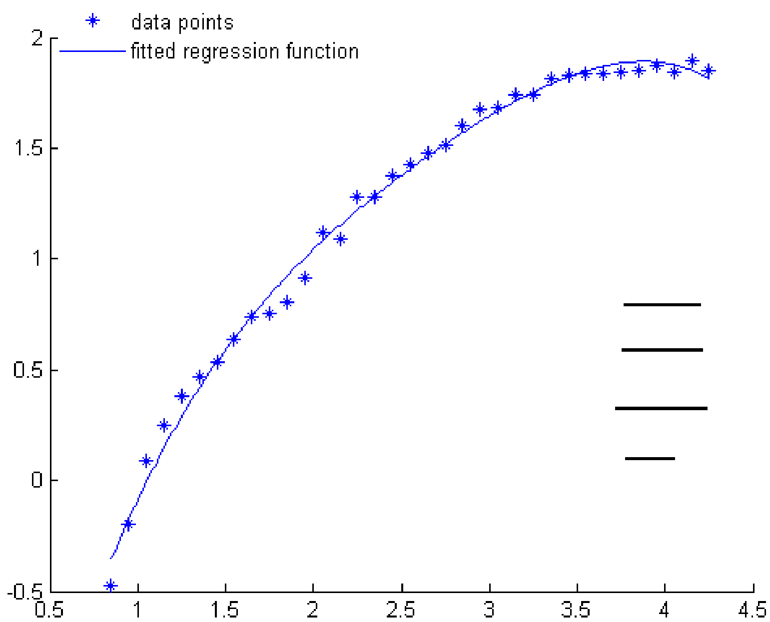



Figure 3.24: The 95% level confidence sets for a maximum point based on white infant data. The horizontal line segments, from top to bottom, represent $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, $\mathbf{C}_c(\boldsymbol{Y})$ and bootstrap confidence set.

Table 3.13: Perinatal mortality data for non-black and non-white infants (Selvin, 1998)

| $i$ | $x_i$ | $Y_i$ | $i$ | $x_i$ | $Y_i$ | $i$ | $x_i$ | $Y_i$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.85 | -0.4337 | 13 | 2.05 | 1.1329 | 25 | 3.25 | 1.5126 |
| 2 | 0.95 | -0.2376 | 14 | 2.15 | 1.1161 | 26 | 3.35 | 1.5905 |
| 3 | 1.05 | 0.0508 | 15 | 2.25 | 1.1209 | 27 | 3.45 | 1.7130 |
| 4 | 1.15 | 0.4103 | 16 | 2.35 | 1.2330 | 28 | 3.55 | 1.6367 |
| 5 | 1.25 | 0.2727 | 17 | 2.45 | 1.1727 | 29 | 3.65 | 1.6358 |
| 6 | 1.35 | 0.5511 | 18 | 2.55 | 1.1759 | 30 | 3.75 | 1.8273 |
| 7 | 1.45 | 0.6673 | 19 | 2.65 | 1.4988 | 31 | 3.85 | 1.5862 |
| 8 | 1.55 | 0.6860 | 20 | 2.75 | 1.4763 | 32 | 3.95 | 1.6953 |
| 9 | 1.65 | 0.7556 | 21 | 2.85 | 1.3615 | 33 | 4.05 | 1.7289 |
| 10 | 1.75 | 0.8515 | 22 | 2.95 | 1.3539 | 34 | 4.15 | 1.7840 |
| 11 | 1.85 | 0.7294 | 23 | 3.05 | 1.5560 | 35 | 4.25 | 1.5445 |
| 12 | 1.95 | 1.0599 | 24 | 3.15 | 1.5707 | | | |

35 non-black and non-white infants. The fitted 4th order polynomial regression function is

$$\hat{Y} = -4.108 + 6.758x - 3.485x^2 + 0.851x^3 - 0.078x^4,$$

and $\hat{\sigma} = 0.093$.

From the normal probability plot (Figure 3.25) and the residual plots (Figure 3.26), the normality assumption of the error seems reasonable. The exploratory index $R^2 = 0.978$, so the 4th order polynomial model fit the data well. To construct a 95% level confidence set for a maximum point, we check the grid points from 0.85 to 4.25 with resolution $d = 0.01$. A conservative confidence set using its conservative critical value $c_2 = 3.1424$ is given by $\mathbf{C}_0(\boldsymbol{Y}) = [3.71, 4.25]$. Then, we compute the critical values (based on 100000 simulations) for the grid points in $\mathbf{C}_0(\boldsymbol{Y})$, and determine whether a grid point is in the exact confidence set $\mathbf{C}_E(\boldsymbol{Y})$. The set $\mathbf{C}_E(\boldsymbol{Y})$ is given by $[3.71, 4.24]$ and the computation takes 2721 seconds. The bootstrap confidence set and Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$ are given by $[3.74, 4.10]$ and $[3.67, 4.25]$, respectively. The data points, fitted regression curve and the 95% confidence sets are plotted in Figure 3.27. The horizontal line segments, from top to bottom, represent $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, $\mathbf{C}_c(\boldsymbol{Y})$ and bootstrap confidence set.

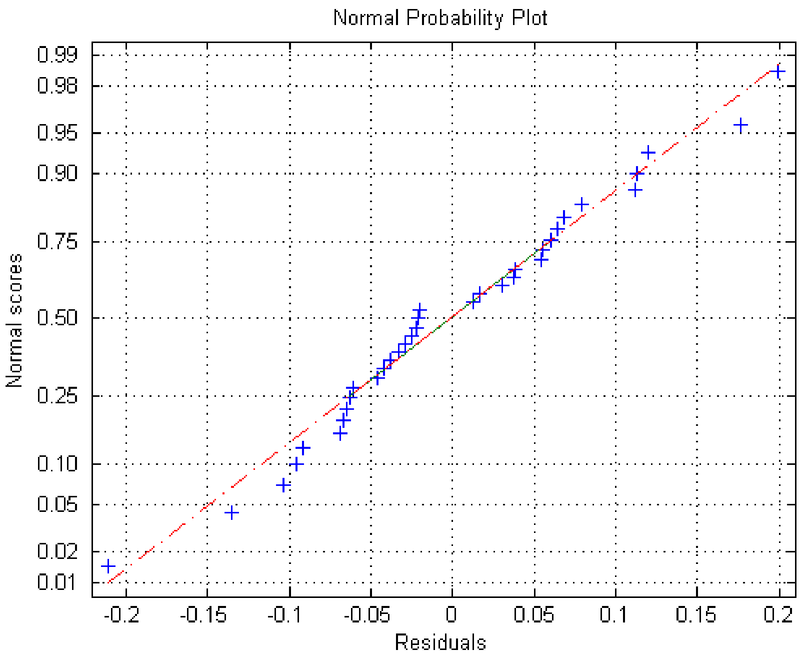



Figure 3.25: Normal probability plot of non-black and non-white infant data

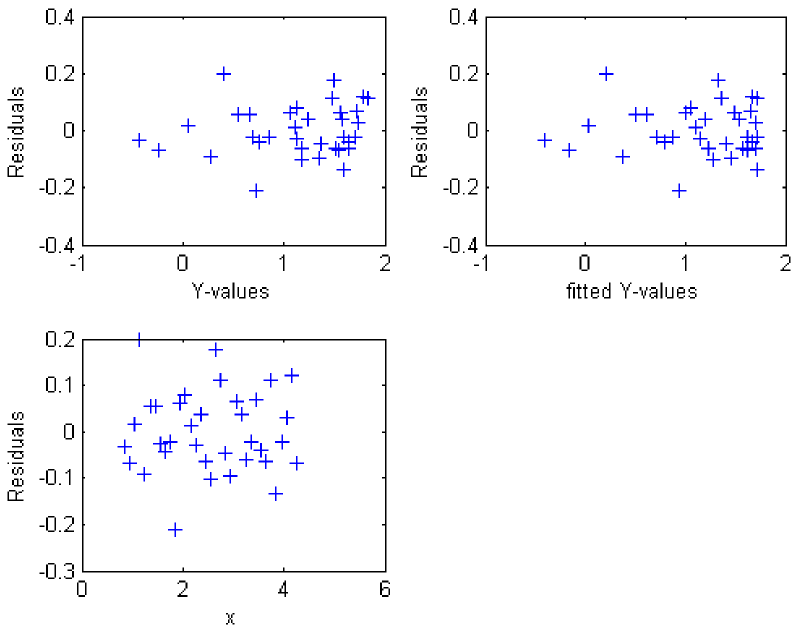



Figure 3.26: Residual plot of non-black and non-white infant data

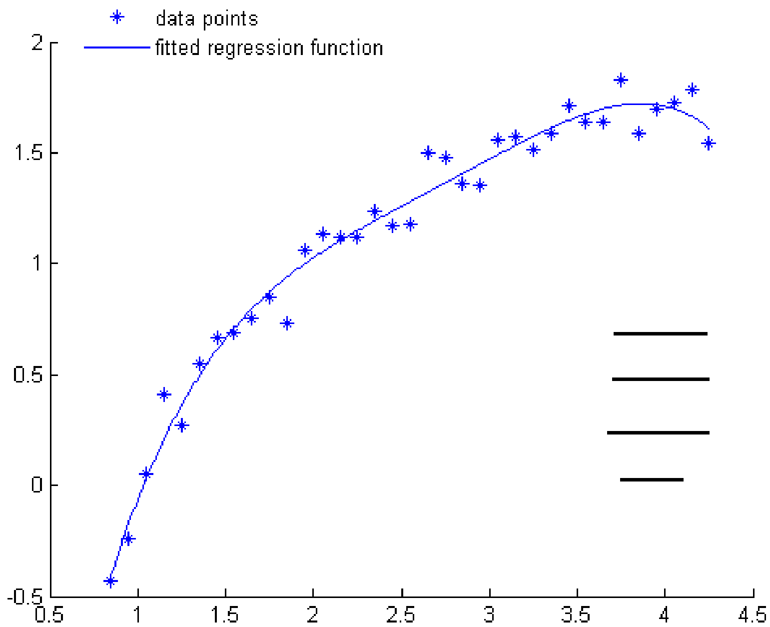



Figure 3.27: The 95% level confidence sets for a maximum point based on non-black and non-white infant data. The horizontal line segments, from top to bottom, represent $\mathbf{C}_E(\boldsymbol{Y})$, $\mathbf{C}_0(\boldsymbol{Y})$, $\mathbf{C}_c(\boldsymbol{Y})$ and bootstrap confidence set.

## 3.2 Rao's Method

Recall that in Chapter 2, a $(1-\alpha)$ level conservative confidence set for any given function $r(\boldsymbol{\theta})$ can be constructed by using Rao's method. In this section, we focus on the univariate polynomial function

$$f(x,\boldsymbol{\theta}) := \theta_0 + \theta_1 x + \cdots + \theta_p x^p$$

and elaborate the method on the construction of confidence sets for the maximum point.

Let $k(\boldsymbol{\theta})$ be a maximum point of the function $f(x,\boldsymbol{\theta})$ in a given interval $[a,b]$, then a $(1-\alpha)$ level conservative confidence set for $k(\boldsymbol{\theta})$ is given by

$$\{k(\boldsymbol{\beta}) \in [a,b] : \boldsymbol{\beta} \in C_{\boldsymbol{\theta}}\}$$

where

$$C_{\boldsymbol{\theta}} = \{\boldsymbol{\beta} : (\hat{\boldsymbol{\theta}} - \boldsymbol{\beta})^T (\mathbf{X}^T\mathbf{X})(\hat{\boldsymbol{\theta}} - \boldsymbol{\beta}) \leq (p+1)\hat{\sigma}^2 f^{\alpha}_{p+1,n-p-1}\}.$$

However, note that a maximum point of $f(x,\boldsymbol{\theta}) = \theta_0 + \theta_1 x + \cdots + \theta_p x^p$ depends only on $\boldsymbol{\theta}^0 := (\theta_1, \theta_2, \cdots, \theta_p)$, but not $\theta_0$. Therefore, the confidence set should be given by

$$\mathbf{C}_c = \{k(\boldsymbol{\beta}^0) \in [a,b] : \boldsymbol{\beta}^0 \in C^0_{\boldsymbol{\theta}}\}$$

where

$$C^0_{\boldsymbol{\theta}} = \{\boldsymbol{\beta}^0 : (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\beta}^0)^T \mathbf{P}^2_R (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\beta}^0) \leq p\hat{\sigma}^2 f^{\alpha}_{p,n-p-1}\}$$

with $\hat{\boldsymbol{\theta}}^0 := (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_p)$ and $\mathbf{P}^2_R = \begin{bmatrix} \mathbf{0} & \mathbf{I}_p \end{bmatrix} (\mathbf{X}^T\mathbf{X}) \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_p \end{bmatrix}$.

According to Carter et al. (1984), the confidence set $C^0_{\boldsymbol{\theta}}$ can be computed using polar co-ordinates. Note that the constraint

$$(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\beta}^0)^T \mathbf{P}^2_R (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\beta}^0) \leq p\hat{\sigma}^2 f^{\alpha}_{p,n-p-1}$$

is just

$$\boldsymbol{z}^T \boldsymbol{z} \leq p\hat{\sigma}^2 f^{\alpha}_{p,n-p-1},$$

which means $\boldsymbol{z}$ lies inside a sphere of radius $r_{rad} = \sqrt{p\hat{\sigma}^2 f^{\alpha}_{p,n-p-1}}$. The vector $\boldsymbol{z}$ can be written in the polar co-ordinates of the following form:

$$\begin{aligned}
z_1 &= r\ \cos(\phi_1) \\
z_2 &= r\ \sin(\phi_1)\cos(\phi_2) \\
\vdots &\quad \vdots \\
z_{p-1} &= r\ \sin(\phi_1)\sin(\phi_2)\cdots\cos(\phi_{p-1}) \\
z_p &= r\ \sin(\phi_1)\sin(\phi_2)\cdots\sin(\phi_{p-1}),
\end{aligned}$$

for $r \in [0, r_{rad}],\ \phi_1 \in [0,\pi),\ \phi_2 \in [0,\pi),\ \cdots, \phi_{p-2} \in [0,\pi)$ and $\phi_{p-1} \in [0, 2\pi)$.

In practice, to compute the confidence set $C^0_{\boldsymbol{\theta}}$, we replace each of the intervals $[0, r_{rad}]$, $[0, 2\pi)$ and $[0,\pi)$ by a set of grid points and denote these sets as $G_r$, $G_\phi$ and $G_{p-1}$, respectively. There are a total of $N = (\#G_r) \times (\#G_\phi)^{p-2} \times (\#G_{p-1})$ different points $\{r, \phi_1, \phi_2, \cdots, \phi_{p-1}\}$, where $r$ takes a value in $G_r$, $\phi_1$, $\phi_2$, $\cdots$, $\phi_{p-2}$ take values in $G_\phi$ and $\phi_{p-1}$ takes a value in $G_{p-1}$. For the $i^{th}$ point, we compute $\boldsymbol{z} = (z_1,\ z_2,\ \cdots,\ z_p)$ using the definition of the polar co-ordinate, and $\boldsymbol{\beta}^0_i = \hat{\boldsymbol{\theta}}^0 - \mathbf{P}^{-1}_R \boldsymbol{z}_i$. If the grid points of $G_r$, $G_\phi$ and $G_{p-1}$ are dense enough, then the $(1-\alpha)$ level confidence set $C^0_{\boldsymbol{\theta}}$ can be represented by $\{\boldsymbol{\beta}^0_i, i = 1, 2, \cdots, N\}$.

Now for each $\boldsymbol{\beta}^0_i$, we compute the maximum point(s) of $f(x, \boldsymbol{\beta}^0_i) = [x, \cdots, x^p]\boldsymbol{\beta}^0_i$. If $\boldsymbol{\beta}^0_i = \mathbf{0}$, then any point in the interval $[a, b]$ is a maximum point. Otherwise, the function $f(x, \boldsymbol{\beta}^0_i)$ has at most $(p-1)$ stationary points, which can be found by solving the equation

$$\frac{df(x, \boldsymbol{\beta}^0_i)}{dx} = [1, 2x, \cdots, (p-1)x^{p-2}, px^{p-1}]\boldsymbol{\beta}^0_i = 0.$$

Denote the stationary point(s) in $(a, b)$ by $s^i_1, s^i_2, \cdots, s^i_k$, where $k \leq (p-1)$. Then the maximum point(s) of $f(x, \boldsymbol{\beta}^0_i)$ must lie within the finite set

$$\{a, b, s^i_1, s^i_2, \cdots, s^i_k\}.$$

For each $\boldsymbol{\beta}^0_i \in C^0_{\boldsymbol{\theta}}$, we compute the maximum point(s) using the method above, and all these maximum points form a $(1-\alpha)$ level Rao's confidence set for $k(\boldsymbol{\theta})$.

## 3.3 Bootstrap Method

When using Bootstrap method to construct a $(1-\alpha)$ level confidence set for a maximum point of the univariate polynomial regression function

$$f(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \cdots + \theta_p x^p$$

in a given interval $[a, b]$, we employ the following steps:

**Step 1.** Randomly choose a set of $n$ bootstrap residuals from the original residuals $\hat{e}_1, \hat{e}_2, \cdots, \hat{e}_n$ with replacement, and denote these bootstrap residuals as $\hat{\boldsymbol{e}}^B = (\hat{e}_1^B, \hat{e}_2^B, \cdots, \hat{e}_n^B)^T$.

**Step 2.** Form the bootstrap sample set $y_1^B, y_2^B, \cdots, y_n^B$, where

$$\boldsymbol{Y}^B = \begin{pmatrix} y_1^B \\ y_2^B \\ \vdots \\ y_n^B \end{pmatrix} := \mathbf{X}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{e}}^B.$$

The design matrix $\mathbf{X}$ remains the same as in the original data set.

**Step 3.** Estimate the parameter $\hat{\boldsymbol{\theta}}^B$ based on the bootstrapped data $\boldsymbol{Y}^B$, that is,

$$\hat{\boldsymbol{\theta}}^B = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}^B.$$

**Step 4.** Find the maximum points of the function

$$f(x, \hat{\boldsymbol{\theta}}^B) = \hat{\theta}_0^B + \hat{\theta}_1^B x + \cdots + \hat{\theta}_p^B x^p$$

in the interval $[a, b]$. This can be done by comparing the values of $f(x, \hat{\boldsymbol{\theta}}^B)$ at $a, b$ and all the stationary points of $f(x, \hat{\boldsymbol{\theta}}^B)$ that lie in $(a, b)$.

Repeat the above steps $N$ times, where $N$ is usually 1000 or larger, and define the set $S_{max}$ to be the set that contains all the maximum points from the $N$ repetitions. Then, the number of points in $S_{max}$, $n_s$ say, should be equal to or larger than $N$. Sort the points in $S_{max}$ in ascending order, and drop the smallest $[\frac{\alpha}{2}n_s]$ and the largest $[\frac{\alpha}{2}n_s]$ values. Then use the remaining points in $S_{max}$ as a $(1-\alpha)$ confidence set for a maximum point.

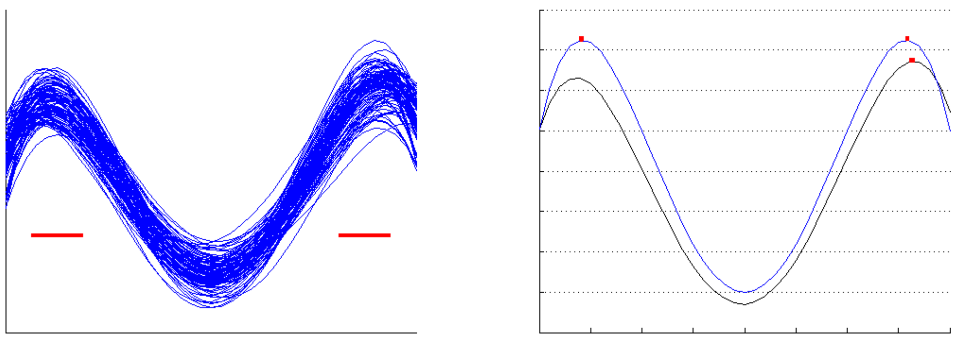
The advantage of bootstrap percentile confidence set for a maximum point is that it is simple and hence is much faster to construct than other confidence sets. However, there are a number of disadvantages in using bootstrap percentile confidence set.

Firstly, there are no fixed rules on which $[n_s\alpha]$ points should be dropped from all of the maximum points $S_{max}$ computed from bootstrapped samples. In Figure 3.28(a), the 2 line segments represent all of the maximum points $S_{max}$ of the fitted regression functions (which is represented by the curves) computed from $N$ bootstrapped samples. Clearly in this case, it is not appropriate to drop the smallest $[\frac{\alpha}{2}n_s]$ and the largest $[\frac{\alpha}{2}n_s]$ points. Instead, it is more appropriate to drop the smallest and largest $[\frac{\alpha}{4}n_s]$ points from the points represented by the line segment in the left and the smallest and largest $[\frac{\alpha}{4}n_s]$ points from the points represented by the line segment in the right.

A plausible way of selecting the $[n_s(1-\alpha)]$ points is choosing the smallest content that captures any $[n_s(1-\alpha)]$ points. Or in other words, we choose the $[n_s(1-\alpha)]$ points with the highest fitted density (see Gibb et al., 2007).

Secondly, it is not clear whether we should treat all of the maximum points in $S_{max}$ equally, if there are different numbers of maximum points of $f(x, \hat{\boldsymbol{\theta}}^B)$ for different $\hat{\boldsymbol{\theta}}^B$. Figure 3.28(b) illustrates this situation. For one bootstrapped parameter $\hat{\boldsymbol{\theta}}^B$, the fitted regression function is depicted by the upper curve, which has 2 maximum points. For another $\hat{\boldsymbol{\theta}}^B$, the fitted regression function is depicted by the lower curve, which has only 1 maximum point. It is not clear whether we should treat these three maximum points in $S_{max}$ equally.

Finally, due to the problems identified above, it might be difficult to establish the large sample asymptotic coverage. For a finite sample, the coverage probability may not be close to the nominal $(1-\alpha)$ level.



(a) (b)

Figure 3.28: Bootstrap method.

## 3.4 Summary

In this chapter, we have considered constructing a confidence set for a maximum point of the regression function in univariate polynomial models. When $p \leq 2$, the critical constants in the confidence set can be computed using analytical method. For a general $p \geq 3$, finding the critical constants involves a $(p-1)$ dimensional integration and so becomes difficult by using a numerical quadrature. Hence we employ a simulation-based method to find $c(k_0)$ for each $k_0$ for a general $p$.

From the examples given in Section 3.1, we conclude that our confidence set $\mathbf{C}_E(\boldsymbol{Y})$ is always smaller thus better than Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$ and of course better than the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$. Although Bootstrap confidence set seems even smaller, its coverage probabilities is hard to define and may not be close to the nominal $1-\alpha$. As explained earlier, the BH confidence set is for the stationary points, but not for the maximum point. Therefore, we recommend the set $\mathbf{C}_E(\boldsymbol{Y})$ when a confidence set for a maximum point of a univariate polynomial function is of interest.

As for the computation time, the set $\mathbf{C}_E(\boldsymbol{Y})$ takes the longest time in the general univariate case when $p \geq 3$ due to the simulation. However, the computation time increases dramatically for Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$ when the order of the regression function increases.

## Chapter 4

# Confidence Set for a Maximum Point of a Bivariate Quadratic Regression Function in a Given Rectangular Region

We have discussed the construction of a $(1-\alpha)$ level confidence set for a maximum point of the general univariate polynomial regression function in Chapter 3. In this chapter, we extend the method to a bivariate quadratic regression function, which is the most used regression function in response surface methodology. A confidence set for a maximum point of a bivariate function has important applications in clinical trials, when the interest is often the combination of components in drugs to reach an optimal performance of treatments.

We propose the method of constructing a confidence set for a maximum point in Section 4.1.1 and elaborate the computation in Section 4.1.2. Rao's method and bootstrap method for the bivariate quadratic case are illustrated in Sections 4.2 and 4.3. Examples are given to compare our confidence sets with other confidence sets. Section 4.4 concludes this chapter with a discussion.

## 4.1 Our Method

### 4.1.1 Theory

Suppose we have the regression model

$$Y = f(\boldsymbol{x}, \boldsymbol{\theta}) + e$$

in the rectangular region $\chi_2 = \{\boldsymbol{x} = (\boldsymbol{x}(1), \boldsymbol{x}(2)) \in R^2 : \boldsymbol{x}(1) \in [a_1, b_1], \quad \boldsymbol{x}(2) \in [a_2, b_2]\}$, where

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 + \theta_1 \boldsymbol{x}(1) + \theta_2 \boldsymbol{x}(2) + \theta_{11} \boldsymbol{x}(1)^2 + \theta_{22} \boldsymbol{x}(2)^2 + \theta_{12} \boldsymbol{x}(1)\boldsymbol{x}(2),$$

$e \sim N(0, \sigma^2)$ and $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_{11}, \theta_{22}, \theta_{12})^T$ is the vector of unknown regression coefficients. We are interested in constructing a $(1 - \alpha)$ level confidence set for a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$.

Let $\boldsymbol{Y} = (y_1, y_2, \cdots, y_n)^T$ be a vector of $n$ observations corresponding to the design matrix

$$\mathbf{X} = \begin{bmatrix} 1 & \boldsymbol{x}_1(1) & \boldsymbol{x}_1(2) & \boldsymbol{x}_1(1)^2 & \boldsymbol{x}_1(2)^2 & \boldsymbol{x}_1(1)\boldsymbol{x}_1(2) \\ 1 & \boldsymbol{x}_2(1) & \boldsymbol{x}_2(2) & \boldsymbol{x}_2(1)^2 & \boldsymbol{x}_2(2)^2 & \boldsymbol{x}_2(1)\boldsymbol{x}_2(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \boldsymbol{x}_n(1) & \boldsymbol{x}_n(2) & \boldsymbol{x}_n(1)^2 & \boldsymbol{x}_n(2)^2 & \boldsymbol{x}_n(1)\boldsymbol{x}_n(2) \end{bmatrix},$$

and $\boldsymbol{e} = (e_1, e_2, \cdots, e_n)^T \sim N(\boldsymbol{0}, I_n\sigma^2)$ be the random errors. Then the regression model can be written in the following matrix form

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{e}.$$

The least squares estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y} \sim N(\boldsymbol{\theta}, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2)$$

and $\hat{\sigma}^2 = \| \hat{\boldsymbol{e}} \|^2 / (n-6)$, where $\hat{\boldsymbol{e}} = \boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}$. Define a $5 \times 5$ symmetric matrix $\mathbf{P}$ such that

$$\mathbf{P}^2 = (\mathbf{0} \quad I_5)\, (\mathbf{X}^T\mathbf{X})^{-1} \begin{pmatrix} \mathbf{0} \\ I_5 \end{pmatrix}.$$

Let $\boldsymbol{k} = \boldsymbol{k}(\boldsymbol{\theta}) \in \chi_2$ be a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ in $\chi_2$. For $\boldsymbol{k}^o \in \chi_2$, if $\boldsymbol{k} = \boldsymbol{k}^o$, we have

$$f(\boldsymbol{k}^o, \boldsymbol{\theta}) - f(\boldsymbol{x}, \boldsymbol{\theta}) \geq 0, \quad \forall \boldsymbol{x} \in \chi_2 \setminus \boldsymbol{k}^o.$$

Therefore, for each $\boldsymbol{k}^o = (k_1^o, k_2^o) \in \chi_2$, a $(1-\alpha)$ level acceptance set for testing the null hypothesis

$$H_0: \quad \boldsymbol{k} = \boldsymbol{k}^o$$

is given by

$$\begin{aligned} A(\boldsymbol{k}^o) &= \{\boldsymbol{Y} : f(\boldsymbol{k}^o, \hat{\boldsymbol{\theta}}) - f(\boldsymbol{x}, \hat{\boldsymbol{\theta}}) \geq -c(\boldsymbol{k}^o)\hat{\sigma}\sqrt{\boldsymbol{g}_b(\boldsymbol{k}^o, \boldsymbol{x})^T \mathbf{P}^2 \boldsymbol{g}_b(\boldsymbol{k}^o, \boldsymbol{x})}, \forall \boldsymbol{x} \in \chi_2 \setminus \boldsymbol{k}^o\} \\ &= \{\boldsymbol{Y} : \inf_{\boldsymbol{x} \in \chi_2 \setminus \boldsymbol{k}^o} \frac{f(\boldsymbol{k}^o, \hat{\boldsymbol{\theta}}) - f(\boldsymbol{x}, \hat{\boldsymbol{\theta}})}{\hat{\sigma}\sqrt{\boldsymbol{g}_b(\boldsymbol{k}^o, \boldsymbol{x})^T \mathbf{P}^2 \boldsymbol{g}_b(\boldsymbol{k}^o, \boldsymbol{x})}} \geq -c(\boldsymbol{k}^o)\}, \end{aligned} \tag{4.1}$$

where

$$\boldsymbol{g}_b(\boldsymbol{k}^o, \boldsymbol{x}) = [k_1^o - x_1, k_2^o - x_2, k_1^{o2} - {x_1}^2, k_2^{o2} - {x_2}^2, k_1^o k_2^o - x_1 x_2]^T$$

with $\boldsymbol{x} = (x_1, x_2)$. It is clear from the definition above that

$$f(\boldsymbol{k}^o, \boldsymbol{\theta}) - f(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{g}_b(\boldsymbol{k}^o, \boldsymbol{x})\boldsymbol{\theta}^0$$

where $\boldsymbol{\theta}^0 = (\theta_1, \theta_2, \theta_{11}, \theta_{22}, \theta_{12})^T$. The estimate of $\boldsymbol{\theta}^0$ is denoted by $\hat{\boldsymbol{\theta}}^0$ and given by $\hat{\boldsymbol{\theta}}^0 = (\mathbf{0} \quad I_5)\hat{\boldsymbol{\theta}}$.

The critical value $c(\boldsymbol{k}^o)$ in Equation (4.1) is chosen such that the coverage probability of $A(\boldsymbol{k}^o)$ is equal to $(1-\alpha)$ under $H_0$.

Next, we determine the critical value $c(\boldsymbol{k}^o)$. Note that

$$\begin{aligned}
& P\{\boldsymbol{Y} \in A(\boldsymbol{k}^o)\} \\
= \;& P\left\{\inf_{\boldsymbol{x}\in\chi_2\setminus\boldsymbol{k}^o} \frac{f(\boldsymbol{k}^o,\hat{\boldsymbol{\theta}}) - f(\boldsymbol{x},\hat{\boldsymbol{\theta}})}{\hat{\sigma}\sqrt{\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T\mathbf{P}^2\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})}} \geq -c(\boldsymbol{k}^o)\right\} \quad (4.2) \\
\geq \;& P\left\{\inf_{\boldsymbol{x}\in\chi_2\setminus\boldsymbol{k}^o} \frac{[f(\boldsymbol{k}^o,\hat{\boldsymbol{\theta}}) - f(\boldsymbol{x},\hat{\boldsymbol{\theta}})] - [f(\boldsymbol{k}^o,\boldsymbol{\theta}) - f(\boldsymbol{x},\boldsymbol{\theta})]}{\hat{\sigma}\sqrt{\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T\mathbf{P}^2\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})}} \geq -c(\boldsymbol{k}^o)\right\} \quad (4.3) \\
= \;& P\left\{\inf_{\boldsymbol{x}\in\chi_2\setminus\boldsymbol{k}^o} \frac{\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)}{\hat{\sigma} \parallel \mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T \parallel} \geq -c(\boldsymbol{k}^o)\right\} \\
= \;& P\left\{\inf_{\boldsymbol{x}\in\chi_2\setminus\boldsymbol{k}^o} \frac{[\mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T]^T[\mathbf{P}^{-1}(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)]}{\hat{\sigma} \parallel \mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T \parallel} \geq -c(\boldsymbol{k}^o)\right\} \\
= \;& P\left\{\inf_{\boldsymbol{x}\in\chi_2\setminus\boldsymbol{k}^o} \frac{[\mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T]^T\boldsymbol{T}}{\parallel \mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T \parallel} \geq -c(\boldsymbol{k}^o)\right\} \quad (4.4)
\end{aligned}$$

where $\boldsymbol{T} \sim T_{5,n-6}$, Equation (4.2) follows directly from Equation (4.1), and Equation (4.3) follows from the fact that $f(\boldsymbol{k}^o,\boldsymbol{\theta}) - f(\boldsymbol{x},\boldsymbol{\theta}) \geq 0 \quad \forall \boldsymbol{x} \in \chi_2 \setminus \boldsymbol{k}^o$. Note that

$$\inf_{\boldsymbol{\theta}:\boldsymbol{k}=\boldsymbol{k}^o} P_{\boldsymbol{k}=\boldsymbol{k}^o}\{\mathbf{Y} \in A(\boldsymbol{k}^o)\} = P\left\{\inf_{\boldsymbol{x}\in\chi_2\setminus\boldsymbol{k}^o} \frac{[\mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T]^T\boldsymbol{T}}{\parallel \mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T \parallel} \geq -c(\boldsymbol{k}^o)\right\},$$

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

$$P\left\{\inf_{\boldsymbol{x}\in\chi_2\setminus\boldsymbol{k}^o} \frac{[\mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T]^T\boldsymbol{T}}{\parallel \mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T \parallel} \geq -c(\boldsymbol{k}^o)\right\} = 1 - \alpha. \quad (4.5)$$

According to Neyman's Theorem, a $(1-\alpha)$ level confidence set for a maximum point, $\boldsymbol{k}$, based on the observation $\boldsymbol{Y}$ is given by

$$\begin{aligned}
\mathbf{C}_E(\boldsymbol{Y}) &= \{\boldsymbol{k}^o \in \chi_2 : \boldsymbol{Y} \in A(\boldsymbol{k}^o)\} \\
&= \left\{\boldsymbol{k}^o \in \chi_2 : \inf_{\boldsymbol{x}\in\chi_2\setminus\boldsymbol{k}^o} \frac{f(\boldsymbol{k}^o,\hat{\boldsymbol{\theta}}) - f(\boldsymbol{x},\hat{\boldsymbol{\theta}})}{\hat{\sigma}\sqrt{\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T\mathbf{P}^2\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})}} \geq -c(\boldsymbol{k}^o)\right\}. \quad (4.6)
\end{aligned}$$

Peterson et al. (2002) considered this problem and proposed an approach to the confidence region for a maximum point. Their confidence set for a maximum point is given by

$$\left\{\boldsymbol{k}^o \in \chi_2 : \inf_{\boldsymbol{x}\in\chi_2\setminus\boldsymbol{k}^o} \frac{f(\boldsymbol{k}^o,\hat{\boldsymbol{\theta}}) - f(\boldsymbol{x},\hat{\boldsymbol{\theta}})}{\hat{\sigma}\sqrt{\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T\mathbf{P}^2\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})}} \geq -c_\alpha\right\},$$

where the critical value is $c_\alpha = \sqrt{q f^{\alpha}_{q,n-p-1}}$ with $q = 2$ and $p = 5$. However, this critical value does not seem correct to us. Our approach above constructs a confidence set which has the same form but with different critical values from theirs.

### 4.1.2 Computation

In order to construct the confidence set given in Equation (4.6), we need to check whether $\boldsymbol{Y}$ is in the acceptance set $A(\boldsymbol{k}^o)$ for each $\boldsymbol{k}^o \in \chi_2$. Since the region $\chi_2$ is continuous and thus contains infinite number of points, we choose a finite grid $S$ on the region $\chi_2$ as a substitute for $\chi_2$. The resolution of the grid is $(d_1, d_2)$, that is, $S = \{(x_1, x_2) : x_1 \in S_1, x_2 \in S_2\}$ where

$$S_1 = \{s_1 = a_1, s_2, \cdots, s_{k_1} = b_1\} \text{ and } S_2 = \{ss_1 = a_2, ss_2, \cdots, ss_{k_2} = b_2\}$$

with $s_i - s_{i-1} = d_1$ and $ss_j - ss_{j-1} = d_2$ for $i = 2, \cdots, k_1$; $j = 2, \cdots, k_2$. If $d_1$ and $d_2$ are small, then $S$ gives a fine approximation to the region $\chi_2$. Therefore, in this section, we only check each point in $S$ and not each point in $\chi_2$ in computing the conservative confidence set and the exact confidence set.

In what follows, we first approximate the critical values in Equation (4.6) by simulated critical values. Then, in order to reduce the complexity of computation, we present the conservative critical value method, which serves to narrow down the range of points that could be in the exact confidence set before we compute the critical values.

**Simulating critical values.** To construct the confidence set in Equation (4.6), we need the critical value $c(\boldsymbol{k}^o)$ for each $\boldsymbol{k}^o \in S$ first. Since the exact critical value $c(\boldsymbol{k}^o)$ is difficult to solve from Equation (4.5), we use simulation to find an approximation to $c(\boldsymbol{k}^o)$.

Let

$$G(\boldsymbol{k}^o, \boldsymbol{x}, \boldsymbol{T}) = \frac{[\mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o, \boldsymbol{x})^T]^T \boldsymbol{T}}{\| \mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o, \boldsymbol{x})^T \|},$$

then from Equation (4.5), we have

$$P\{\inf_{\boldsymbol{x} \in \chi_2 \setminus \boldsymbol{k}^o} G(\boldsymbol{k}^o, \boldsymbol{x}, \boldsymbol{T}) \geq -c(\boldsymbol{k}^o)\} = 1 - \alpha.$$

The simulation method for finding the critical value $c(\boldsymbol{k}^o)$ follows three steps.

**Step 1.** Sample independent $\boldsymbol{T}_i \sim \boldsymbol{T}_{5,n-6}$, $i = 1, 2, \cdots, m$.

**Step 2.** For each $\boldsymbol{T}_i$, $i = 1, 2, \cdots, m$, compute

$$\inf_{\boldsymbol{x} \in \chi_2 \setminus \boldsymbol{k}^o} G(\boldsymbol{k}^o, \boldsymbol{x}, \boldsymbol{T}_i).$$

Since the form of $G(\boldsymbol{k}^o, \boldsymbol{x}, \boldsymbol{T}_i)$ is complicated and the infimum on $\boldsymbol{x} \in \chi_2 \setminus \boldsymbol{k}^o$ is difficult to find analytically, we use numerical methods. One way to approximate this infimum is to compute $G(\boldsymbol{k}^o, \boldsymbol{x}, \boldsymbol{T}_i)$ at each $\boldsymbol{x}$ in the set $S \setminus \boldsymbol{k}^o$, or a fine grid of points on the region $\chi_2$, and then use the minimum of these values as an approximation to the infimum. Alternatively, by using the method of computing the infimum of a univariate function in Section 3.1.3.2, we can compute $\inf_{\boldsymbol{x}(2) \in [a_2, b_2]} G(\boldsymbol{k}^o, \boldsymbol{x}, \boldsymbol{T}_i)$ at each fixed $\boldsymbol{x}(1) \in S_1$, and use the minimum of these values as an approximation to $\inf_{\boldsymbol{x} \in \chi_2 \setminus \boldsymbol{k}^o} G(\boldsymbol{k}^o, \boldsymbol{x}, \boldsymbol{T}_i)$.

An alternative way to compute the infimum is through the direct use of the Matlab function *fmincon*. However, *fmincon* function only finds a local infimum near the starting point chosen by the users. Therefore, the result is not guaranteed to be the global infimum in the region $\chi_2$. To search for the infimum, we vary the starting points in *fmincon*, and use the minimum of all the *fmincon* outputs as an approximation to the global infimum. However, the computation will take longer when there are more starting points. We suggest using the maximum point(s) of the fitted regression function in the given region $\chi_2$ and $\boldsymbol{k}^o$ as the starting points. This seems to work well based on empirical experience.

**Step 3.** Sort the values of $\inf_{\boldsymbol{x} \in \chi_2 \setminus \boldsymbol{k}^o} G(\boldsymbol{k}^o, \boldsymbol{x}, \boldsymbol{T}_i)$, $i = 1, 2, \cdots, m$, in increasing order. Use the $[\alpha \times m]th$ value as an approximation to $-c(\boldsymbol{k}^o)$.

In this way, we simulate $c(\boldsymbol{k}^o)$ for each $\boldsymbol{k}^o \in S$ and then decide the confidence set according to Equation (4.6).

**The Conservative Confidence Set.** From Section 4.1.1, we need to find $c(\boldsymbol{k}^o)$ for each grid point $\boldsymbol{k}^o \in S$ in order to decide the confidence set. However, the simulation for each $c(\boldsymbol{k}^o)$ is time consuming, since it involves a large number of replications, $m$, to reach a certain accuracy. As in the univariate case, we reduce the computation by first using a conservative critical value to construct a conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$, and then only computing $c(\boldsymbol{k}^o)$ for the grid points $\boldsymbol{k}^o$ in $\mathbf{C}_0(\boldsymbol{Y})$. By constructing $\mathbf{C}_0(\boldsymbol{Y})$ first, we only need to compute the critical values for each of the grid points that lie in $\mathbf{C}_0(\boldsymbol{Y})$, not for each grid point in $S$. In this way, the computation burden is greatly reduced.

Note that $\boldsymbol{T} \sim \boldsymbol{T}_{5,n-6}$, therefore

$$\frac{\| \boldsymbol{T} \|^2}{5} \sim F_{5,n-6},$$

where $F_{5,n-6}$ is the standard F-distribution with degrees of freedom 5 and $n-6$. Denote $f^{\alpha}_{5,n-6}$ as the upper $\alpha$ point of the distribution $F_{5,n-6}$, then we have

$$P\{\| \boldsymbol{T} \| \leq \sqrt{5f^{\alpha}_{5,n-6}}\} = 1-\alpha,$$

that is

$$P\left\{\inf_{\boldsymbol{\rho}\in R^5} \frac{\boldsymbol{\rho}\boldsymbol{T}}{\| \boldsymbol{\rho} \|} \geq -\sqrt{5f^{\alpha}_{5,n-6}}\right\} = 1-\alpha.$$

Hence, we have

$$\begin{aligned} & P\left\{\inf_{\boldsymbol{x}\in S\backslash \boldsymbol{k}^o} \frac{[\mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T]^T\boldsymbol{T}}{\| \mathbf{P}\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T \|} \geq -\sqrt{5f^{\alpha}_{5,n-6}}\right\} \\ \geq \;& P\left\{\inf_{\boldsymbol{\rho}\in R^5} \frac{\boldsymbol{\rho}\boldsymbol{T}}{\| \boldsymbol{\rho} \|} \geq -\sqrt{5f^{\alpha}_{5,n-6}}\right\} \\ = \;& 1-\alpha. \end{aligned} \tag{4.7}$$

By comparing the probability statement (4.7) with the probability statement (4.5), we conclude $c(\boldsymbol{k}^o) < \sqrt{5f^{\alpha}_{5,n-6}}$.

Thus, by substituting $\sqrt{5f^{\alpha}_{5,n-6}}$ for $c(\boldsymbol{k}^o)$ in Equation (4.6), the conservative confidence set is given by

$$\mathbf{C}_0(\boldsymbol{Y}) = \left\{\boldsymbol{k}^o \in S : \inf_{\boldsymbol{x}\in\chi_2\backslash \boldsymbol{k}^o} \frac{f(\boldsymbol{k}^o,\hat{\boldsymbol{\theta}}) - f(\boldsymbol{x},\hat{\boldsymbol{\theta}})}{\hat{\sigma}\sqrt{\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})^T\mathbf{P}^2\boldsymbol{g}_b(\boldsymbol{k}^o,\boldsymbol{x})}} \geq -\sqrt{5f^{\alpha}_{5,n-6}}\right\}. \tag{4.8}$$

By using the conservative confidence set first, we do not need to compute the critical value $c(\boldsymbol{k}^o)$ for each $\boldsymbol{k}^o \in S$, but only for $\boldsymbol{k}^o$ in the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ to determine the confidence set $\mathbf{C}_E(\boldsymbol{Y})$. The computational expense is significantly reduced.

### 4.1.3 Examples

**Example 1.** Suppose we have $21 \times 21$ observations $(x_1, x_2, y)$ from the regression model

$$Y = 2x_1 + 3x_2 - x_1^2 - 5x_2^2 - 2x_1x_2 + e$$

where $e \sim N(0, 1)$, $x_1 \in \{-1, -0.9, \cdots, 1\}$ and $x_2 \in \{-1, -0.9, \cdots, 1\}$. The fitted bivariate quadratic regression model is

$$\hat{Y} = -0.0212 + 2.0545x_1 + 3.0370x_2 - 1.1933x_1^2 - 4.8154x_2^2 - 1.9195x_1x_2,$$

and $\hat{\sigma} = 0.9539$. To construct a 95% level confidence set for a maximum point in the region $[-1, 1] \times [-1, 1]$, we construct the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ first. Then, we compute the critical values (based on 1000 simulations) for the grid points in $\mathbf{C}_0(\boldsymbol{Y})$ and determine whether each grid point is in the exact confidence set $\mathbf{C}_E(\boldsymbol{Y})$. Figure 4.1(a) shows the fitted response surface. In Figure 4.1 (b)-(d), the shaded regions represent the 95% level corresponding confidence sets. In Figure 4.1(e), the whole dotted region represents the confidence set $\mathbf{C}_0(\boldsymbol{Y})$ in Equation (4.8) while the light dotted region represents the confidence set $\mathbf{C}_E(\boldsymbol{Y})$. The cross in each confidence set represents the true maximum point. Computation of $\mathbf{C}_E(\boldsymbol{Y})$ takes approximately 50 minutes.

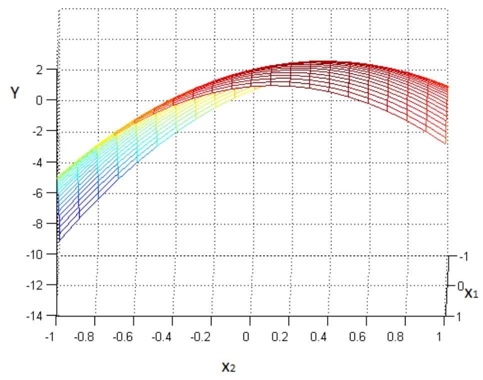
**Example 2.** Suppose we have $6 \times 6$ observations $(x_1, x_2, y)$ from the regression model

$$Y = 2 + 2x_1 + 3x_2 + x_1^2 + 5x_2^2 + 2x_1x_2 + e,$$

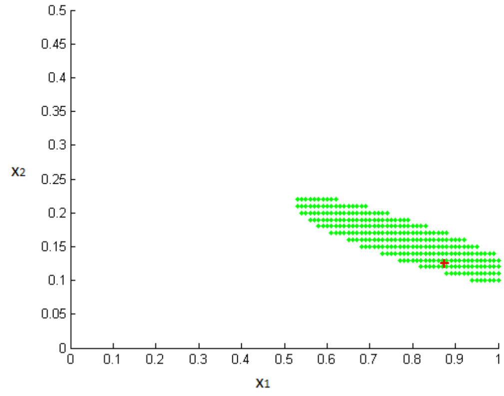
where $e \sim N(0, 10^2)$, $x_1 \in \{0, 2, \cdots, 10\}$ and $x_2 \in \{0, 1, \cdots, 5\}$. The fitted bivariate quadratic regression model is

$$\hat{Y} = 0.1773 + 1.5707x_1 + 4.2067x_2 + 1.1001x_1^2 + 4.9186x_2^2 + 1.9834x_1x_2,$$

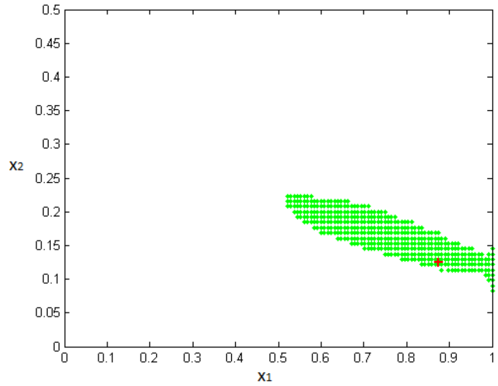
and $\hat{\sigma} = 9.7201$. To construct a 95% level confidence set for a maximum point in the region $[0, 10] \times [0, 5]$, we construct the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ first. Then, we compute the critical values (based on 10000 simulations) for the grid points in $\mathbf{C}_0(\boldsymbol{Y})$ and determine whether each grid point is in the confidence set



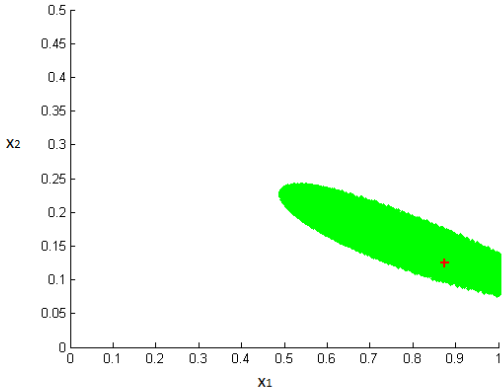

(a) Fitted response surface



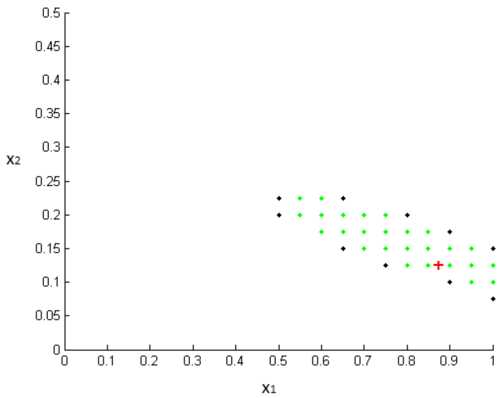

(b) BH confidence set




(c) Bootstrap confidence set

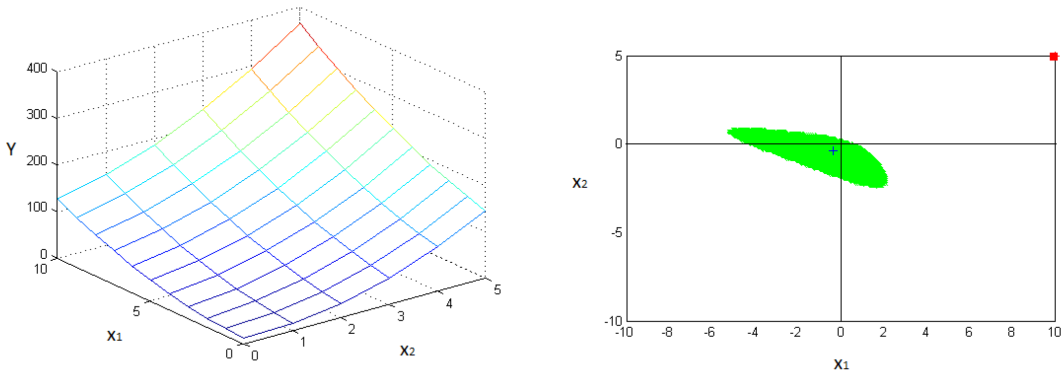



(d) Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$




(e) Confidence set $\mathbf{C}_0(\boldsymbol{Y})$ and $\mathbf{C}_E(\boldsymbol{Y})$

Figure 4.1: The response surface and 95% confidence sets in Example 1




(a)Fitted response surface (b) The confidence sets

Figure 4.2: The response surface and 95% confidence sets in Example 2

$\mathbf{C}_E(\boldsymbol{Y})$. Figure 4.2(a) shows the fitted response surface. In Figure 4.2(b), the shaded region represents the BH confidence set, the cross inside the region represents the true stationary point, and the point in the upper-right conner represents all other confidence sets computed in this example.

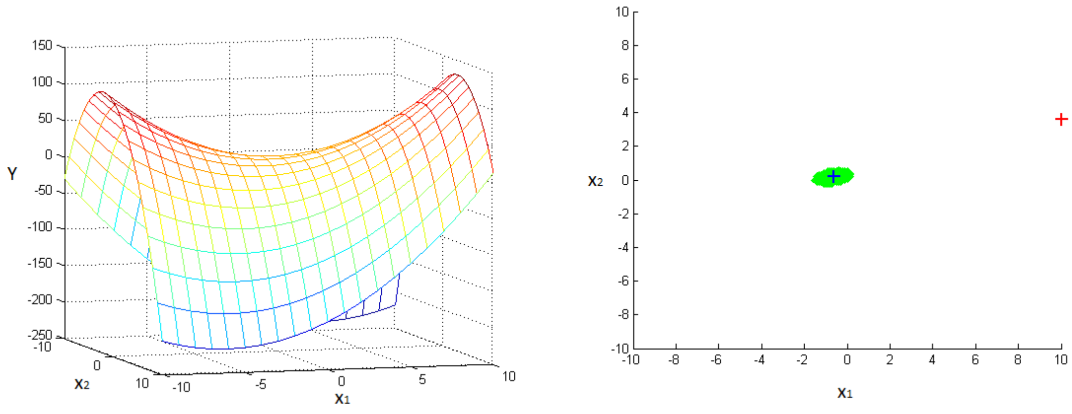
**Example 3.** Suppose we have $5 \times 5$ observations $(x_1, x_2, y)$ from the regression model

$$Y = 1 + x_1 + x_2 + x_1^2 - 2x_2^2 + x_1x_2 + e,$$

where $e \sim N(0, 25^2)$. The fitted bivariate quadratic regression function is

$$\hat{Y} = -1.8403 + 0.95602x_1 + 1.4354x_2 + 0.94671x_1^2 - 1.8555x_2^2 + 1.1818x_1x_2,$$

and $\hat{\sigma} = 25.095$, $x_1 \in \{-10, -5, \cdots, 10\}$ and $x_2 \in \{-10, -5, \cdots, 10\}$. To construct a 95% level confidence set for a maximum point in the region $[-10, 10] \times [-10, 10]$, we construct the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ first. Then, we compute the critical values (based on 1000 simulations) for the grid points in $\mathbf{C}_0(\boldsymbol{Y})$ and determine whether each grid point is in the confidence set $\mathbf{C}_E(\boldsymbol{Y})$. Figure 4.3 shows the fitted response surface, the confidence set using our method and using other methods. The shaded region in each figure represents the confidence set and the cross near the edge on the right represents the true maximum point. The other cross in Figure 4.3(b)




(a) Fitted response surface (b) BH confidence set

represents the true stationary point.

**Example 4.** Suppose we have $5 \times 5$ observations $(x_1, x_2, y)$ from the regression model

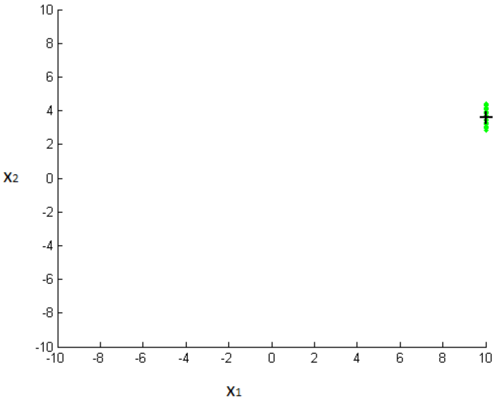
$$Y = 10x_1 + 10x_2 - 5x_1^2 - 5x_2^2 - 10x_1x_2 + e,$$

where $e \sim N(0, 1)$, $x_1 \in \{-1, -0.5, \cdots, 1\}$ and $x_2 \in \{-1, -0.5, \cdots, 1\}$. The fitted bivariate quadratic regression function is

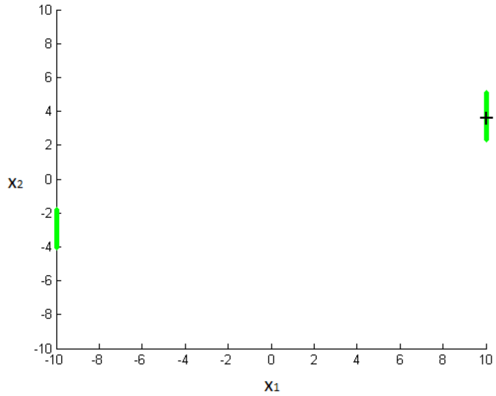
$$\hat{Y} = 0.0023851 + 10.144x_1 + 10.266x_2 - 5.2619x_1^2 - 4.5582x_2^2 - 9.6445x_1x_2,$$

and $\hat{\sigma} = 0.84128$. To construct a 95% level confidence set for a maximum point in the region $[-1, 1] \times [-1, 1]$, we construct the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ first. Then, we compute the critical values (based on 1000 simulations) for the grid points in $\mathbf{C}_0(\boldsymbol{Y})$ and determine whether each grid point is in the confidence set $\mathbf{C}_E(\boldsymbol{Y})$. Figure 4.4 shows the fitted response surface, the confidence set using our method and using other methods. The shaded region in each figure represents the corresponding confidence set and the line lies within the confidence set represents the true maximum points. In Figure 4.4(e), the light dotted region represents the set $\mathbf{C}_E(\boldsymbol{Y})$ while the whole dotted region represents $\mathbf{C}_0(\boldsymbol{Y})$.

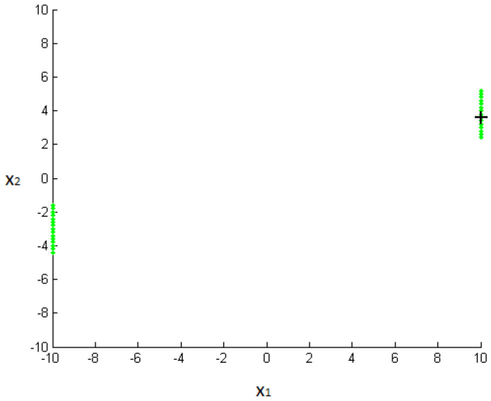
**Example 5.** Table 4.1 shows the data of a central composite design for a chemical process (Myers, 2009, pp 48, table 2.8), resulting from an investigation into the effect of two variables, reaction temperature $(x_1)$ and reactant concentration $(x_2)$,



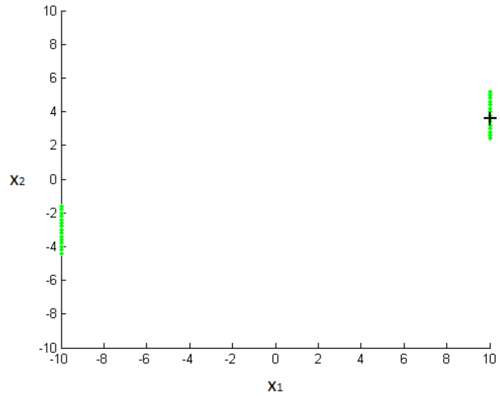

(c) Bootstrap confidence set




(d) Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$

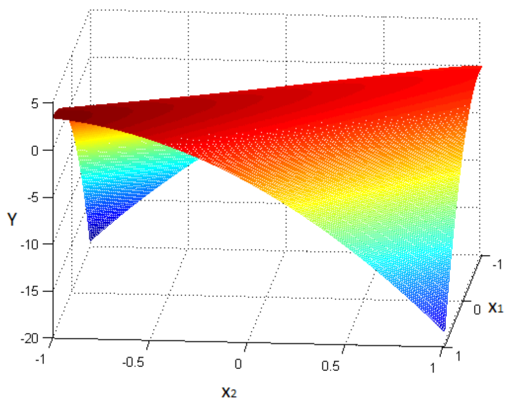



(e) Confidence set $\mathbf{C}_0(\boldsymbol{Y})$

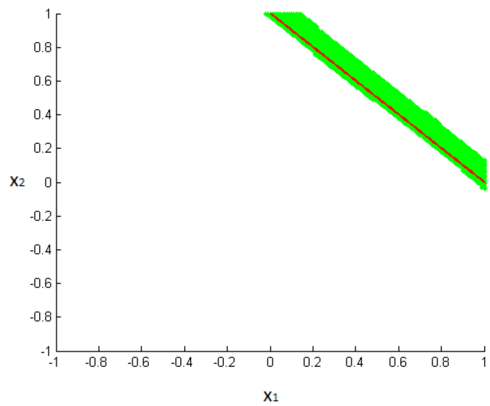



(f) Confidence set $\mathbf{C}_E(\boldsymbol{Y})$

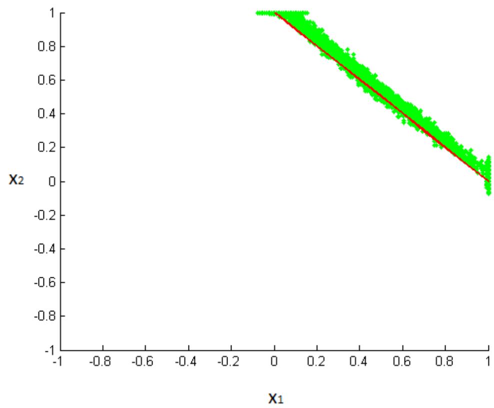
Figure 4.3: The response surface and 95% confidence sets in Example 3



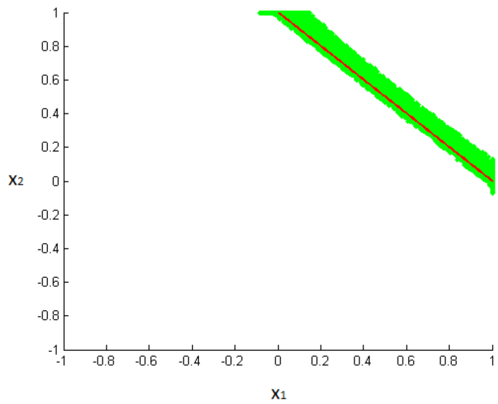

(a) Fitted response surface



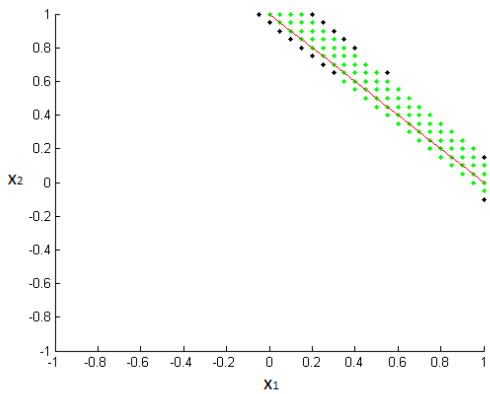

(b) BH confidence set




(c) Bootstrap confidence set




(d) Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$




(e) Confidence set $\mathbf{C}_0(\boldsymbol{Y})$ and $\mathbf{C}_E(\boldsymbol{Y})$

Figure 4.4: The response surface and 95% confidence sets in Example 4

Table 4.1: Data of a central composite design for the chemical process (Myers, 2009)

| observation | run | temperature($^oC$) | conc.(%) | $x_1$ | $x_2$ | $y$ |
|---|---|---|---|---|---|---|
| 1 | 4 | 200 | 15 | -1 | -1 | 43 |
| 2 | 12 | 250 | 15 | 1 | -1 | 78 |
| 3 | 11 | 200 | 25 | -1 | 1 | 69 |
| 4 | 5 | 250 | 25 | 1 | 1 | 73 |
| 5 | 6 | 189.65 | 20 | -1.414 | 0 | 48 |
| 6 | 7 | 260.35 | 20 | 1.414 | 0 | 78 |
| 7 | 1 | 225 | 12.93 | 0 | -1.414 | 65 |
| 8 | 3 | 225 | 27.07 | 0 | 1.414 | 74 |
| 9 | 8 | 225 | 20 | 0 | 0 | 76 |
| 10 | 10 | 225 | 20 | 0 | 0 | 79 |
| 11 | 9 | 225 | 20 | 0 | 0 | 83 |
| 12 | 2 | 225 | 20 | 0 | 0 | 81 |

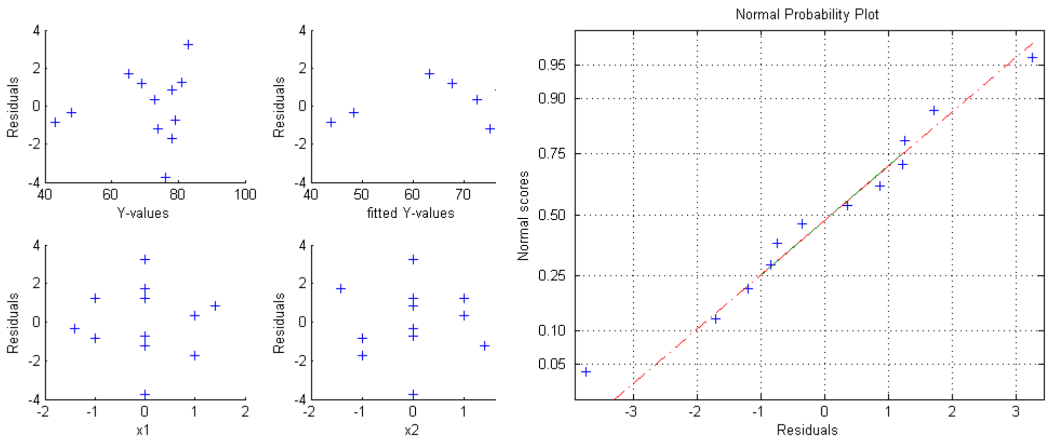
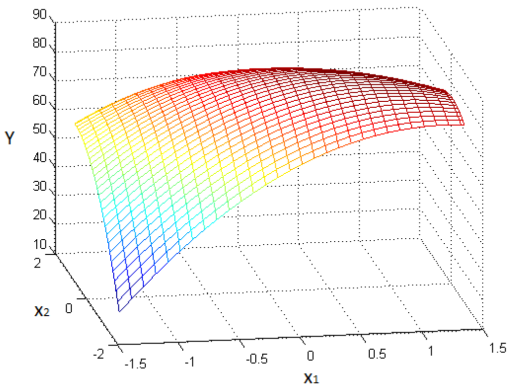
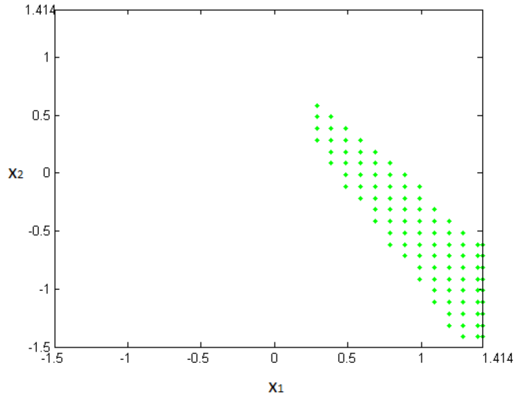



Figure 4.5: The residual plot and normal probability plot in Example 5




(a) Fitted response surface

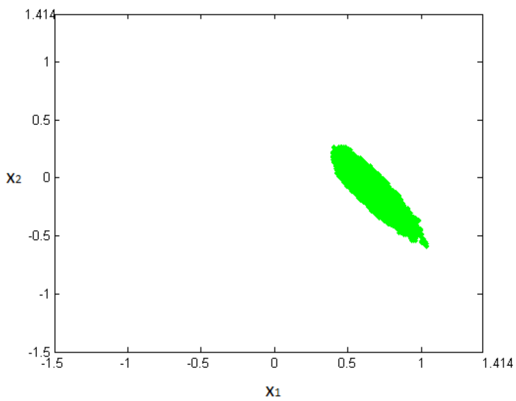



(b) Box and Hunter's confidence set

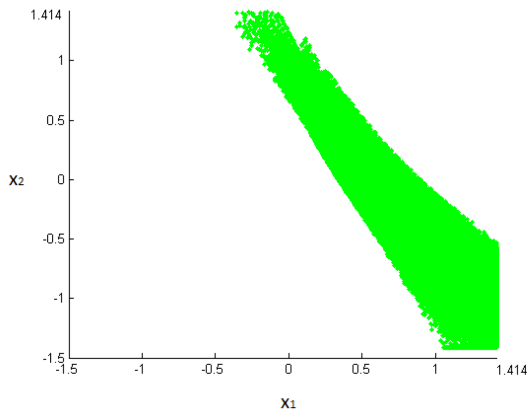
on the percentage conversion of a chemical process ($y$). The fitted bivariate quadratic regression model is

$$\hat{Y} = 79.75 + 10.18x_1 + 4.22x_2 - 8.50x_1^2 - 5.25x_2^2 - 7.75x_1x_2,$$

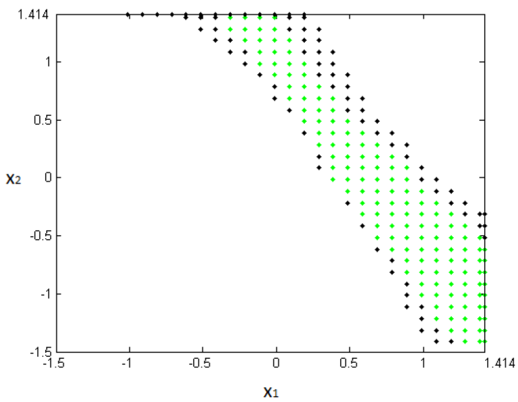
and $\hat{\sigma} = 2.49$. From the residual plots and normal probability plot (Figure 4.5), the normality assumption of the residuals seemed reasonable. To construct a 95% level confidence set for a maximum point, we check the grid points in the rectangular region edged by $\pm 1.414$ first and construct the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$. Then, we compute the critical values (based on 1000 simulations) for the grid points in $\mathbf{C}_0(\boldsymbol{Y})$ and determine whether each grid point is in the confidence set $\mathbf{C}_E(\boldsymbol{Y})$. Figure 4.6 shows the fitted response surface, the confidence set using our method and using other methods. Computation of the confidence set $\mathbf{C}_E(\boldsymbol{Y})$ takes approximately 5 hours.




(c) Bootstrap confidence set




(d) Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$




(e) Confidence set $\mathbf{C}_0(\boldsymbol{Y})$ and $\mathbf{C}_E(\boldsymbol{Y})$

Figure 4.6: The response surface and 95% confidence sets in Example 5

## 4.2 Rao's Method

The construction of confidence sets for a maximum point of a general univariate polynomial function using Rao's method has been illustrated in Section 3.2. In this section, we construct a confidence set for the bivariate quadratic function

$$f(\boldsymbol{x},\boldsymbol{\theta}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2 + \theta_{12} x_1 x_2$$

in a given covariate region $\chi_2$ following the same steps. We briefly describe the procedure next.

Let $k(\boldsymbol{\theta})$ be a maximum point of the function $f(\boldsymbol{x},\boldsymbol{\theta})$ in $\chi_2$, then a $(1-\alpha)$ level conservative confidence set for $k(\boldsymbol{\theta})$ is given by

$$\mathbf{C}_c = \{k(\boldsymbol{\beta}^0) \in [a,b] : \boldsymbol{\beta}^0 \in C^0_{\boldsymbol{\theta}}\}$$

where

$$C^0_{\boldsymbol{\theta}} = \{\boldsymbol{\beta}^0 : (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\beta}^0)^T \mathbf{P}_R^2 (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\beta}^0) \leq 5\hat{\sigma}^2 f^{\alpha}_{5,n-6}\}$$

with $\hat{\boldsymbol{\theta}}^0 := (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_{11}, \hat{\theta}_{22}, \hat{\theta}_{12})$ and $\mathbf{P}_R^2 = \begin{bmatrix} \mathbf{0} & \mathbf{I}_5 \end{bmatrix} (\mathbf{X}^T\mathbf{X}) \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_5 \end{bmatrix}$.

The constraint

$$(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\beta}^0)^T \mathbf{P}_R^2 (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\beta}^0) \leq 5\hat{\sigma}^2 f^{\alpha}_{5,n-6}$$

can be written as

$$\boldsymbol{z}^T \boldsymbol{z} \leq 5\hat{\sigma}^2 f^{\alpha}_{5,n-6},$$

where $\mathbf{z} = \mathbf{P}_R(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\beta}^0)$. Following Section 3.2, the vector $\boldsymbol{z}$ can be written in the following polar co-ordinates:

$$\begin{aligned}
z_1 &= r\ \cos(\phi_1) \\
z_2 &= r\ \sin(\phi_1)\cos(\phi_2) \\
z_3 &= r\ \sin(\phi_1)\sin(\phi_2)\cos(\phi_3) \\
z_4 &= r\ \sin(\phi_1)\sin(\phi_2)\sin(\phi_3)\cos(\phi_4) \\
z_5 &= r\ \sin(\phi_1)\sin(\phi_2)\sin(\phi_3)\sin(\phi_4),
\end{aligned}$$

for $r \in [0, r_{rad}]$, $\phi_1, \ \phi_2, \ \phi_3 \in [0, \pi)$ and $\phi_4 \in [0, 2\pi)$.

We replace each of the intervals $[0, r_{rad}]$, $[0, 2\pi)$ and $[0, \pi)$ by a set of grid points and denote these sets as $G_r$, $G_\phi$ and $G_4$, respectively. There are a total of $N = (\#G_r) \times (\#G_\phi)^3 \times (\#G_4)$ points $\{r, \phi_1, \phi_2, \phi_3, \phi_4\}$, where $r$ takes a value in $G_r$, $\phi_1$, $\phi_2$ and $\phi_3$ take values in $G_\phi$ and $\phi_4$ takes a value in $G_4$. For the $i^{th}$ point, we compute $\boldsymbol{z} = (z_1, \ z_2, \ \cdots, \ z_5)$ using the definition of the polar co-ordinate, and $\boldsymbol{\beta}_i^0 = \hat{\boldsymbol{\theta}}^0 - \mathbf{P}_R^{-1}\boldsymbol{z}_i$. If the grid points of $G_r$, $G_\phi$ and $G_4$ are densely chosen, then the $(1-\alpha)$ level confidence set $C_{\boldsymbol{\theta}}^0$ can be represented by $\{\boldsymbol{\beta}_i^0, i = 1, 2, \cdots, N\}$.

Now for each $\boldsymbol{\beta}_i^0$, we compute the maximum point(s) of $f(\boldsymbol{x}, \boldsymbol{\beta}_i^0) = [x_1, x_2, x_1^2, x_2^2, x_1x_2]\boldsymbol{\beta}_i^0$ in $\chi_2$. Note that the maximum point(s) are either at the stationary point(s) of the function $f(\boldsymbol{x}, \boldsymbol{\beta}_i^0)$, or on the boundary of the rectangular region $\chi_2$. The stationary point(s) can be found, if exist, by solving the equation

$$\frac{\partial f(\boldsymbol{x}, \boldsymbol{\beta}_i^0)}{\partial x_j} = 0, \ j = 1, 2,$$

that is,

$$\begin{cases} [1, 0, 2x_1, 0, x_2]\boldsymbol{\beta}_i^0 = 0 \\ [0, 1, 0, 2x_2, x_1]\boldsymbol{\beta}_i^0 = 0 \end{cases}.$$

The boundary of $\chi_2$ is made of 4 line segments and at each line segment $f(\boldsymbol{x}, \boldsymbol{\beta}_i^0)$ is a univariate quadratic polynomial function whose maximum point(s) can be found according to Section 3.2. Hence the maximum point(s) of $f(\boldsymbol{x}, \boldsymbol{\beta}_i^0)$ in the covariate region $\chi_2$ is determined by comparing the function values at the maximum points of the 4 line segments and at the stationary point(s).

For each $\boldsymbol{\beta}_i^0 \in C_{\boldsymbol{\theta}}^0$, we compute the maximum point(s) using the method above, and all these maximum points form a $(1-\alpha)$ level Rao's confidence set for $k(\boldsymbol{\theta})$.

## 4.3 Bootstrap Method

When using Bootstrap method to construct a $(1-\alpha)$ level confidence set for a maximum point of the bivariate quadratic function

$$f(\boldsymbol{x},\boldsymbol{\theta}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2 + \theta_{12} x_1 x_2$$

in a given covariate region $\chi_2$, we modify the constructing method in Section 3.3. The resampling and parameter estimation steps are the same as in the univariate case:

**Step 1.** Randomly choose a set of $n$ bootstrap residuals from the original residuals $\hat{e}_1, \hat{e}_2, \cdots, \hat{e}_n$ with replacement, and denote these bootstrap residuals as $\hat{\boldsymbol{e}}^B = (\hat{e}_1^B, \hat{e}_2^B, \cdots, \hat{e}_n^B)^T$.

**Step 2.** Form the bootstrap sample set $y_1^B, y_2^B, \cdots, y_n^B$, where

$$\boldsymbol{Y}^B = \begin{pmatrix} y_1^B \\ y_2^B \\ \vdots \\ y_n^B \end{pmatrix} := \mathbf{X}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{e}}^B.$$

The design matrix $\mathbf{X}$ remains the same as in the original data set.

**Step 3.** Estimate the parameter $\hat{\boldsymbol{\theta}}^B$ based on the bootstrapped data $\boldsymbol{Y}^B$, that is,

$$\hat{\boldsymbol{\theta}}^B = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}^B.$$

**Step 4.** Find the maximum points of the function $f(\boldsymbol{x},\boldsymbol{\theta}^B)$ in $\chi_2$. This can be done by comparing the values of $f(\boldsymbol{x},\hat{\boldsymbol{\theta}}^B)$ at the stationary point(s) that lie in $\chi_2$ and on the boundary as discussed in Section 4.2.

Repeat the above steps $N$ times, where $N$ is usually 1000 or larger, and define the set $S_{max}$ to be the set that contains all the maximum points from the $N$ repetitions. Then, the number of points in $S_{max}$, $n_s$ say, should be equal to or larger than $N$. Any $[(1-\alpha)n_s]$ points in $S_{max}$ form a $(1-\alpha)$ level confidence set for $k(\boldsymbol{\theta})$.

A plausible way of selecting the $[(1-\alpha)n_s]$ points is to first estimate the density at each of the $n_s$ points in $S_{max}$ and then select the $[(1-\alpha)n_s]$ points with the highest density. This density estimation can be implemented by using the matlab function *kde2d* which uses a Gaussian kernel.

## 4.4 Summary

In this chapter, we have investigated the confidence sets for the popular quadratic response surface models. The theory of our method covers the bivariate case as well as the univariate case. But computing the critical constants becomes even more difficult than that in the univariate case. We use the Matlab function *fmincon* to find the minimum values. However, *fmincon* does not guarantee that the minimization leads to a global minimum. This applies to the calculation of the confidence sets for Equation (4.6). The calculated critical values may therefore be larger than the true critical values and the infimum of the left-hand side of the inequality in (4.6) may also be larger than the true values. To improve the accuracy, one can vary the starting point when using *fmincon* and select the smallest one as an approximation to the global minimum in the constrained covariate region.

From the examples given in Section 4.1.3, we note that the smallest confidence sets are bootstrap confidence sets. But due to the reasons stated earlier, we do not recommend bootstrap confidence set and BH confidence set. The confidence set $\mathbf{C}_E(\boldsymbol{Y})$ is always appeared to be smaller than or equal to Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$ and of course smaller than the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$. Hence we conclude that it is better to use our method in constructing a confidence set in the bivariate case.

As for the computation time, our method generally takes the longest time due to the use of simulation in computing the critical constants.

# Chapter 5

# Extensions to other models

In Chapters 3 and 4, we have constructed an exact $(1-\alpha)$ level confidence set for a maximum point for normal-error univariate polynomial regression models and bivariate quadratic regression models. This method can be extended to some other models which involve a linear function. 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 the corresponding $(1-\alpha)$ 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 and Bates, 2000; McCulloch and Searle, 2001), only the maximum likelihood estimators of fixed effects regression coefficients can be obtained. In that case, our method can still be applied to produce the corresponding $(1-\alpha)$ asymptotic confidence sets via the large sample approximate normal distribution of the maximum likelihood estimators.

In this chapter, we consider to construct a confidence set for a maximum point of the mean response $E(Y)$ for models other than ordinary linear regression models. Sections 5.1-5.3 consider three specific models to illustrate the method.

Suppose $(Y_1, \boldsymbol{x}_1), (Y_2, \boldsymbol{x}_2), \cdots, (Y_n, \boldsymbol{x}_n)$ are $n$ observations from a pre-specified model which involves a linear function

$$f(\boldsymbol{x},\boldsymbol{\theta}) = \theta_0 + \mathbf{z}(\boldsymbol{x})^T\boldsymbol{\theta}^0$$

where $\mathbf{z}$ is a given $p\times 1$ vector-valued function of the observation $\boldsymbol{x}$, $\boldsymbol{\theta} = (\theta_0, \cdots, \theta_p)^T$ and $\boldsymbol{\theta}^0 = (\theta_1, \cdots, \theta_p)^T$. Furthermore, suppose the mean response $E(Y)$ is related to $f(\boldsymbol{x},\boldsymbol{\theta})$ by a given monotone function. Therefore the confidence set for $E(Y)$ can be derived from that of $f(\boldsymbol{x},\boldsymbol{\theta})$. If the assumption holds, then the problem of constructing a confidence set for a maximum point of $E(Y)$ is translated to that of a linear function $f(\boldsymbol{x},\boldsymbol{\theta})$. Next, we consider the construction of a confidence set for a maximum point $\boldsymbol{k}$ of $f(\boldsymbol{x},\boldsymbol{\theta})$ in a given region $\chi$.

Following Neyman's Theorem in Chapter 2, a $(1-\alpha)$ level confidence set for $\boldsymbol{k}$ is given by

$$C(\boldsymbol{Y}) = \{\boldsymbol{k}^o \in \chi : \boldsymbol{Y} \in A(\boldsymbol{k}^o)\}$$

where $A(\boldsymbol{k}^o)$ is a $(1-\alpha)$ level acceptance set for testing $H : \boldsymbol{k} = \boldsymbol{k}^o$ for each $\boldsymbol{k}^o \in \chi$. A $(1-\alpha)$ level acceptance set can be constructed as

$$A(\boldsymbol{k}^o) = \left\{\boldsymbol{Y} : f(\boldsymbol{k}^o,\hat{\boldsymbol{\theta}}) - f(\boldsymbol{x},\hat{\boldsymbol{\theta}}) \geq -c(\boldsymbol{k}^o)\sqrt{v\hat{a}r[f(\boldsymbol{k}^o,\hat{\boldsymbol{\theta}}) - f(\boldsymbol{x},\hat{\boldsymbol{\theta}})]}, \forall \boldsymbol{x} \in \chi\right\}$$

where each critical value $c(\boldsymbol{k}^o)$ can be determined such that $P\{\boldsymbol{Y} \in A(\boldsymbol{k}^o)\} = 1-\alpha$. In fact, if an estimator $\hat{\boldsymbol{\theta}}^0$ of $\boldsymbol{\theta}^0$ is available with normal distribution

$$\hat{\boldsymbol{\theta}}^0 \sim N(\boldsymbol{\theta}^0, \Sigma)$$

then the above acceptance set becomes

$$\begin{aligned} A(\boldsymbol{k}^o) &= \left\{\boldsymbol{Y} : f(\boldsymbol{k}^o,\hat{\boldsymbol{\theta}}) - f(\boldsymbol{x},\hat{\boldsymbol{\theta}}) \geq -c(\boldsymbol{k}^o)\hat{\sigma}\sqrt{[z(\boldsymbol{k}^o) - z(\boldsymbol{x})]^T\mathbf{V}[z(\boldsymbol{k}^o) - z(\boldsymbol{x})]}, \forall \boldsymbol{x} \in \chi\right\} \\ &= \left\{\boldsymbol{Y} : f(\boldsymbol{k}^o,\hat{\boldsymbol{\theta}}) - f(\boldsymbol{x},\hat{\boldsymbol{\theta}}) \geq -c(\boldsymbol{k}^o)\hat{\sigma} \parallel [z(\boldsymbol{k}^o) - z(\boldsymbol{x})]^T\mathbf{P} \parallel, \forall \boldsymbol{x} \in \chi\right\} \qquad (5.1) \end{aligned}$$

where $\Sigma = \sigma^2\mathbf{V}$ is a positive definite variance-covariance matrix and $\mathbf{P}^2 = \mathbf{V}$. The estimator of the error variance $\sigma^2$ has a distribution of $\hat{\sigma}^2 \sim \sigma^2\chi^2_\nu/\nu$ and is denoted by $\hat{\sigma}^2$. The determination of each critical constant $c(\boldsymbol{k}^o)$ employs the same procedure discussed in Chapters 3 and 4.

However, when the distribution of $\hat{\boldsymbol{\theta}}^0$ is otherwise not normal or not available, the procedure is adapted to produce an asymptotic $(1-\alpha)$ confidence set for $\boldsymbol{k}$. From the Central Limit Theory, if the sample size is large, approximately we have

$$\hat{\boldsymbol{\theta}}^0 \sim N(\boldsymbol{\theta}^0, \Sigma)$$

with $\Sigma = J^{-1}(\boldsymbol{\theta}^0)$ where $J(\boldsymbol{\theta}^0)$ is the Fisher Information matrix of $\boldsymbol{\theta}^0$. Furthermore, $\hat{\Sigma} = J^{-1}(\hat{\boldsymbol{\theta}}^0)$ is usually used to approximate the variance-covariance matrix $\Sigma$. Let $\mathbf{P}^2 = \hat{\Sigma}$ and $\hat{\sigma} = 1$, then the acceptance set is the same as we constructed in Equation (5.1)

$$A(\boldsymbol{k}^o) \quad = \quad \left\{\boldsymbol{Y} : f(\boldsymbol{k}^o, \hat{\boldsymbol{\theta}}) - f(\boldsymbol{x}, \hat{\boldsymbol{\theta}}) \geq -c(\boldsymbol{k}^o) \parallel [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \parallel, \forall \boldsymbol{x} \in \chi \right\}.$$

Note that

$$\begin{aligned}
& P\{\boldsymbol{Y} \in A(\boldsymbol{k}^o)\} \\
= \;& P\left\{f(\boldsymbol{k}^o, \hat{\boldsymbol{\theta}}) - f(\boldsymbol{x}, \hat{\boldsymbol{\theta}}) \geq -c(\boldsymbol{k}^o) \parallel [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \parallel, \forall \boldsymbol{x} \in \chi \right\} \\
\geq \;& P\left\{[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0) \geq -c(\boldsymbol{k}^o) \parallel [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \parallel, \forall \boldsymbol{x} \in \chi \right\} \quad (5.2) \\
= \;& P\left\{ \inf_{\forall \boldsymbol{x} \in \chi \setminus \boldsymbol{k}^o} \frac{[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)}{\parallel [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \parallel} \geq -c(\boldsymbol{k}^o) \right\} \\
= \;& P\left\{ \inf_{\forall \boldsymbol{x} \in \chi \setminus \boldsymbol{k}^o} \frac{[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} [\mathbf{P}^{-1} (\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)]}{\parallel [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \parallel} \geq -c(\boldsymbol{k}^o) \right\} \\
= \;& P\left\{ \inf_{\forall \boldsymbol{x} \in \chi \setminus \boldsymbol{k}^o} \frac{\{[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P}\} \boldsymbol{N}}{\parallel [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \parallel} \geq -c(\boldsymbol{k}^o) \right\}
\end{aligned}$$

where Equation (5.2) follows from the fact that $[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \boldsymbol{\theta}^0 \geq 0$ under the hypothesis $\boldsymbol{k} = \boldsymbol{k}^o$,

$$\boldsymbol{N} = \mathbf{P}^{-1}(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0) \sim N(0, \mathbf{I}_p)$$

and $\mathbf{I}_p$ is the $p \times p$ identity matrix. Therefore, $c(\boldsymbol{k}^o)$ is solved from

$$P\left\{ \inf_{\forall \boldsymbol{x} \in \chi \setminus \boldsymbol{k}^o} \frac{\{[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P}\} \boldsymbol{N}}{\parallel [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \parallel} \geq -c(\boldsymbol{k}^o) \right\} = 1 - \alpha \qquad (5.3)$$

using a similar inference and computation procedure to that used in Chapters 3 and 4. We take generalized linear models, Cox's proportional hazard models and a Becker's H1 model as examples to illustrate the method.

## 5.1 Generalized Linear Models

In statistics, the term generalized linear models (GLMs) refers to a class of models which are generalized from ordinary linear regression. The GLMs allow the linear model to be related to the response variable via a link function and allow the magnitude of the variance of each measurement to be a function of its predicted value.

A generalized linear model has three components:

1. Response $Y$ with a probability distribution from the exponential family

$$f_{exp}(y; \beta) = \exp[yb(\beta) + c(\beta) + d(y)].$$

2. A linear predictor $f(\boldsymbol{x}, \boldsymbol{\theta})$ .

3. A monotone link function $h$ such that $h[E(Y)] = f(\boldsymbol{x}, \boldsymbol{\theta})$.

Suppose we have $n$ observations $(Y_1, \boldsymbol{x}_1), (Y_2, \boldsymbol{x}_2), \cdots, (Y_n, \boldsymbol{x}_n)$ from a generalized linear model in the form

$$h[E(Y)] = f(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 + \mathbf{z}(\boldsymbol{x})^T \boldsymbol{\theta}, \tag{5.4}$$

where $h$ is a strictly monotone, differentiable function and $Y_i$ has a distribution

$$f_{exp}(y_i; \beta_i) = \exp[y_i b(\beta_i) + c(\beta_i) + d(y_i)], \tag{5.5}$$

with unknow parameter $\beta_i$ (the $\beta_i$s don't need to be identical). We are interested in constructing a confidence set for a maximum point of $E(Y) = h^{-1}(f(\boldsymbol{x}, \boldsymbol{\theta}))$ in a given region $\chi$ based on observations $(Y_1, \boldsymbol{x}_1), (Y_2, \boldsymbol{x}_2), \cdots, (Y_n, \boldsymbol{x}_n)$.

### 5.1.1 Method

Since $h$ is a monotone function, a maximum point of $E(Y)$ is either a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ or a maximum point of $f(\boldsymbol{x}, -\boldsymbol{\theta})$, depending on whether the function $h$ is increasing or decreasing. Without loss of generality, we assume $h$ is a strictly increasing function. Therefore, a point $\boldsymbol{k}$ is a maximum point of $E(Y)$ if and only if it is a maximum point of $f(\boldsymbol{x},\boldsymbol{\theta})$. Hence a $(1-\alpha)$ level confidence set for a maximum point of $E(Y)$ is a $(1-\alpha)$ level confidence set for a maximum point of $f(\boldsymbol{x},\boldsymbol{\theta})$. Our problem of constructing a confidence set for a maximum point of $E(Y)$ is therefore translated to that of constructing a confidence set for a maximum point of a polynomial function $f(\boldsymbol{x},\boldsymbol{\theta})$, which we have already discussed in Chapters 3 and 4. A similar argument clearly goes through for other fixed or random effects linear or generalized linear models so long as the regression function involved is a linear function.

**Parameter and its covariance matrix**

To construct a confidence set for a maximum point $\boldsymbol{k}$ of $f(\boldsymbol{x},\boldsymbol{\theta})$, we first need an estimator of $\boldsymbol{\theta}$ and its distribution. This step can be done by using the maximum likelihood method (see Dobson(2002) for details). From Equation (5.5) we have

$$\begin{aligned} \mu_i &:= E(Y_i) = -c^{'}(\beta_i)/b^{'}(\beta_i) \\ \eta &:= h(\mu_i) \\ \text{and } var(Y_i) &= [b^{''}(\beta_i)c^{'}(\beta_i) - c^{''}(\beta_i)b^{'}(\beta_i)]/[b^{'}(\beta_i)]^3. \end{aligned}$$

The log-likelihood function for $Y_i$ is given by

$$l_i = y_i b(\beta_i) + c(\beta_i) + d(y_i),$$

and the log-likelihood function for $\boldsymbol{Y}$ is

$$l = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} [y_i b(\beta_i) + c(\beta_i) + d(y_i)].$$

The maximum likelihood estimator for $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, is the solution of

$$\frac{\partial l}{\partial \theta_j} = U_j = \sum_{i=1}^{n} \left[\frac{\partial l_i}{\partial \theta_j}\right] = 0, \tag{5.6}$$

$i = 0, 1, \cdots, n;\ j = 1, \cdots, p$, with $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ and $var[\hat{\boldsymbol{\theta}}] = J^{-1}(\boldsymbol{\theta})$ where $J(\boldsymbol{\theta})$ is the Fisher Information matrix. Although the maximum likelihood estimator can be obtained by solving Equation (5.6) analytically, it is difficult to give an expression in most of the cases. Therefore numerical methods are usually used in parameter estimation. We simply use here the procedure illustrated by Dobson (2002).

Note that Equation (5.6) can be simplified by using the chain rule for differentiation, that is,

$$U_j = \sum_{i=1}^{n} \left[ \frac{\partial l_i}{\partial \beta_i} \cdot \frac{\partial \beta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \theta_j} \right] \quad (5.7)$$

$$= \sum_{i=1}^{n} \left[ b^{'}(\beta_i)(y_i - \mu_i) \cdot b^{'}(\beta_i) var(Y_i) \cdot \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \right] \quad (5.8)$$

$$= \sum_{i=1}^{n} \left[ \frac{(y_i - \mu_i)}{var(Y_i)} x_{ij} (\frac{\partial \mu_i}{\partial \eta_i}) \right]. \quad (5.9)$$

Hence the $(j,k)$th element of the variance-covariance matrix of $[U_1, \cdots, U_p]$ is given by

$$J_{j,k} := E(U_j U_k) = \sum_{i=1}^{n} \left[ \frac{x_{ij} x_{ik}}{var(Y_i)} (\frac{\partial \mu_i}{\partial \eta_i})^2 \right]. \quad (5.10)$$

To solve

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \boldsymbol{U} = \mathbf{0},$$

Newton-Raphson method is applied. The estimate is given by iteration

$$\hat{\boldsymbol{\theta}}^{(m)} = \hat{\boldsymbol{\theta}}^{(m-1)} + [J^{(m-1)}]^{-1} \boldsymbol{U}^{(m-1)} \quad (5.11)$$

where $\hat{\boldsymbol{\theta}}^{(m)}$ is the estimate of $\boldsymbol{\theta}$ at the $m$-th iteration, $J^{(m)} = J(\hat{\boldsymbol{\theta}}^{(m)}) = (J_{i,k})$ is the Fisher information and $\boldsymbol{U}^{(m)}$ is the score vector evaluated at $\hat{\boldsymbol{\theta}}^{(m)}$. Let $W$ be an $n \times n$ diagonal matrix with elements

$$w_{ii} = \frac{1}{var(Y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2,$$

then Equation (5.10) can be expressed as

$$J = \mathbf{X}^T W \mathbf{X}.$$

Let $\boldsymbol{z}$ be a vector with elements

$$z_i = \sum_{k=1}^{p} x_{ik} \theta_k^{(m-1)} + (y_i - \mu_i) \left( \frac{\partial \eta_i}{\partial \mu_i} \right)$$

with $\mu_i$ and $\partial \eta_i / \partial \mu_i$ evaluated at $\boldsymbol{\theta}^{(m-1)}$, then Equation (5.11) becomes

$$\mathbf{X}^T W \mathbf{X} \boldsymbol{\theta}^{(m)} = \mathbf{X}^T W \boldsymbol{z}.$$

Thus $\boldsymbol{\theta}^{(m)}$ is obtained by an iterative weighted least squares procedure and $\hat{\Sigma}$ is approximated by $J^{-1}(\hat{\boldsymbol{\theta}}^0)$. In Matlab, this procedure is implemented by the function *glmfit*

$$[\hat{\boldsymbol{\theta}}, dev, stats] = glmfit(\mathbf{X}, \boldsymbol{y}, distr),$$

where the output *dev* is the deviance of the fit at the solution $\hat{\boldsymbol{\theta}}$, and *stats* includes the summary statistics such as the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, the residuals and the degrees of freedom for error.

As long as $\hat{\boldsymbol{\theta}}$ and $\hat{\Sigma}$ are obtained, the acceptance set (5.1) is constructed using the critical constant $c(\boldsymbol{k}^o)$ solved from Equation (5.3). However, because the standard Normal distribution rather than the T-distribution is used in Equation (5.3), the critical value in computing the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ should be $\sqrt{\chi^2_p}$ following a similar inference given in Section 3.1.3.2.

### 5.1.2 Example

Table 5.1: Survival data for the CTX experiment (Carter *et al.*, 1983)

| group number | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| CTX(mg/kg) | 0.0 | 65.7 | 92.0 | 164 | 230 | 296 | 414 |
| ♯ of mice | 16 | 8 | 8 | 8 | 8 | 8 | 8 |
| ♯ surviving 21 days | 0 | 3 | 5 | 8 | 8 | 5 | 0 |

This example considers the survival data for the cyclophosphamide (CTX) experiment given in Carter *et al.* (1983). Female mice were randomly divided into a control group (16 mice) and six treatment groups (8 mice per group) at different levels of CTX. Carter *et al.* (1983) were particularly interested in the probability of survival of at least 21 days. Table 5.1 summarizes the observations on 21-day survival from the experiment. The usual model fitting procedure suggests that the following logistic regression model of 21-day survival probability $p$ on dose level $x$ =CTX fits

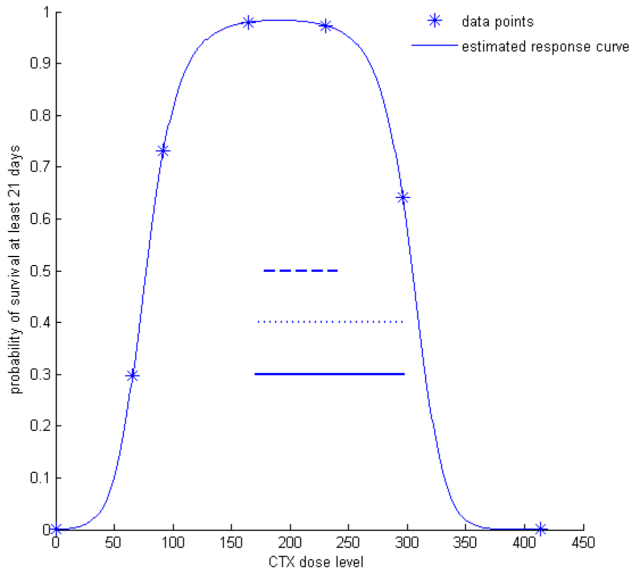



Figure 5.1: The three 95% confidence sets in Example 5.1.2; $\mathbf{C}_E(\mathbf{Y})$ is given by the top broken line, $\mathbf{C}_0(\mathbf{Y})$ is given by the middle dot line, and $\mathbf{C}_c(\mathbf{Y})$ is given by the bottom solid line.

the observed data very well:

$$p(x, \boldsymbol{\theta}) = (1 + \exp(-f(x, \boldsymbol{\theta})))^{-1} \quad \text{with } f(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2.$$

Figure 5.1 plots the observed 21-day survival relative frequencies (given by the $*$'s) and the fitted logistic regression curve. The parameter estimates are given by

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} -7.4333 \\ 0.1209 \\ -0.0003 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 8.1865 & -0.1249 & 0.0003 \\ -0.1249 & 0.0020 & -0.0000 \\ 0.0003 & -0.0000 & 0.0000 \end{pmatrix}.$$

It is clearly of interest to construct a confidence set for the CTX level in the observed covariate range $x \in [0, 414]$ that maximizes the 21-day survival probability. Using $\alpha = 0.05$, resolution $d = 0.1$ and numerical quadrature, the asymptotic confidence sets are computed to be $\mathbf{C}_E(\mathbf{Y}) = [173.5, 239.9]$, $\mathbf{C}_0(\mathbf{Y}) = [169.7, 297.1]$ and $\mathbf{C}_c(\mathbf{Y}) = [169.7, 297.0]$. The three confidence sets are indicated by the three line segments in Figure 5.1. It is clear that the width 66.4 of $\mathbf{C}_E(\mathbf{Y})$ is much smaller than both the widths 127.4 and 127.3 of $\mathbf{C}_0(\mathbf{Y})$ and $\mathbf{C}_c(\mathbf{Y})$ respectively. The critical constants in this example were computed by using numerical quadrature, and the computation time of $\mathbf{C}_E(\mathbf{Y})$ was 47 seconds on the same PC.

## 5.2 Cox's Proportional Hazard Models

In survival analysis, there are three ways to describe the population

- $f(t)$-the density function, which gives the instantaneous probability of an event at time $t$.
- $h(t)$-the hazard function, which gives the instantaneous potential for an event to occur given that an event has not occurred yet up to time $t$.
- $S(t)$-the survivor function, which gives the probability an individual survives longer than time $t$.

Let the distribution function be $F(t) = \int_0^t f(s)ds$, then the hazard function and survivor function can be expressed as

$$S(t) = 1 - F(t) \quad \text{and} \quad h(t) = \frac{f(t)}{S(t)}.$$

It is clear that the hazard function $h(t)$ is non-negative for all $t \in [0, \infty)$.

A proportional hazard model assumes that the hazard function takes the following form

$$h(t, \boldsymbol{x}) = \lambda(t) exp(f(\boldsymbol{x}, \boldsymbol{\theta})) = \lambda(t) exp(g(\boldsymbol{x})^T \boldsymbol{\beta}), \tag{5.12}$$

where $\lambda(t)$ is a smooth function of $t$ and is positive for all $t \geq 0$. However, a Cox's proportional hazard model (cf. Cox, 1972, 1975; Kleinbaum and Klein, 2005) assume $\lambda(t)$ can take any value since it is irrelevant to the parameter estimation using Cox's partial maximum likelihood (Cox, 1975).

Given observations $(Y_1, \boldsymbol{x}_1), (Y_2, \boldsymbol{x}_2), \cdots, (Y_n, \boldsymbol{x}_n)$, the interest is on the construction of a confidence set for a minimum point of the hazard function $h(t, \boldsymbol{x})$ at any time $t$ in a covariate region $\chi$.

### 5.2.1 Method

Because $\lambda(t)$ is positive for all $t \geq 0$, hence a minimum point of $h(t, \boldsymbol{x})$ is a minimum point of $f(\boldsymbol{x}, \boldsymbol{\beta})$, that is, a $1-\alpha$ level confidence set for a maximum point of $f(\boldsymbol{x}, -\boldsymbol{\theta})$.

Our problem is therefore translated to that of constructing a confidence set for a maximum point of $f(\boldsymbol{x}, -\boldsymbol{\theta})$ in a given covariate region $\chi$.

According to the inference at the beginning of this chapter, to construct a confidence set for a maximum point of $f(\boldsymbol{x}, -\boldsymbol{\theta})$, we need an estimator of $\boldsymbol{\theta}$ and its distribution first. The parameter estimation can be done by using Cox's partial maximum likelihood method.

**Cox's Partial Maximum Likelihood**

Let $t_1, t_2, \cdots, t_n$ be $n$ distinct failure times and there is no tie between these times. The Cox's partial likelihood is given by

$$L = \prod_{j=1}^{n} L_j,$$

where

$$L_j = \frac{\lambda(t_j)exp(f(\boldsymbol{x}, \boldsymbol{\theta})}{\sum_{\boldsymbol{x}\in R(t_j)} \lambda(t_j)exp(f(\boldsymbol{x}, \boldsymbol{\theta}))} = \frac{exp(f(\boldsymbol{x}, \boldsymbol{\theta}))}{\sum_{\boldsymbol{x}\in R(t_j)} exp(f(\boldsymbol{x}, \boldsymbol{\theta}))}$$

with $R(t_j)$ being the set of individuals any of whom may be found fail at time $t_j$. When there are ties in the survival time, several methods are provided to approximate the partial likelihood, see Breslow (1970) and Efron (1977) for more information.

The parameters are solved from

$$\frac{\partial \log L}{\partial \theta_i} = 0, \; i = 1, 2, \cdots, p+1.$$

The estimates can be obtained by using the Matlab command

$$[\hat{\boldsymbol{\theta}}, \log L, H, stats] = coxphfit(X, T)$$

where $X$ is the matrix of the observations on predictor variables and $T$ is the vector of time-to-event data. The output $\log L$ is the log-likelihood, $H$ is a two-column matrix which contains the $T$ values in the first column and the estimated baseline cumulative hazard in the second column. *stats* is a structure which contains the summary statistics such as the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$.

According to the Central Limit Theory, approximately we have

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \hat{\Sigma}).$$

Then the critical values can be determined by solving Equation (5.3) and a $(1-\alpha)$ level confidence set for a maximum point of $f(\boldsymbol{x}, -\boldsymbol{\theta})$ can then be constructed following the previous inference directly.

### 5.2.2 Example

This example considers the survival data in a murine cancer chemotherapy experiment that used the drugs 5-Fluorouracil and Teniposide given by Stablein et al. (1983). A series combinations of 5-Fluorouracil (5FU) and Teniposide (VM26) was given to treat 127 mice with leukemia. The original data of combinations and survival times were recorded in Table 5.2. These data were scaled to be

$$x_1 = \frac{5FU(mg/kg) - 130}{130}$$
$$x_2 = \frac{VM26(mg/kg) - 13}{13}$$

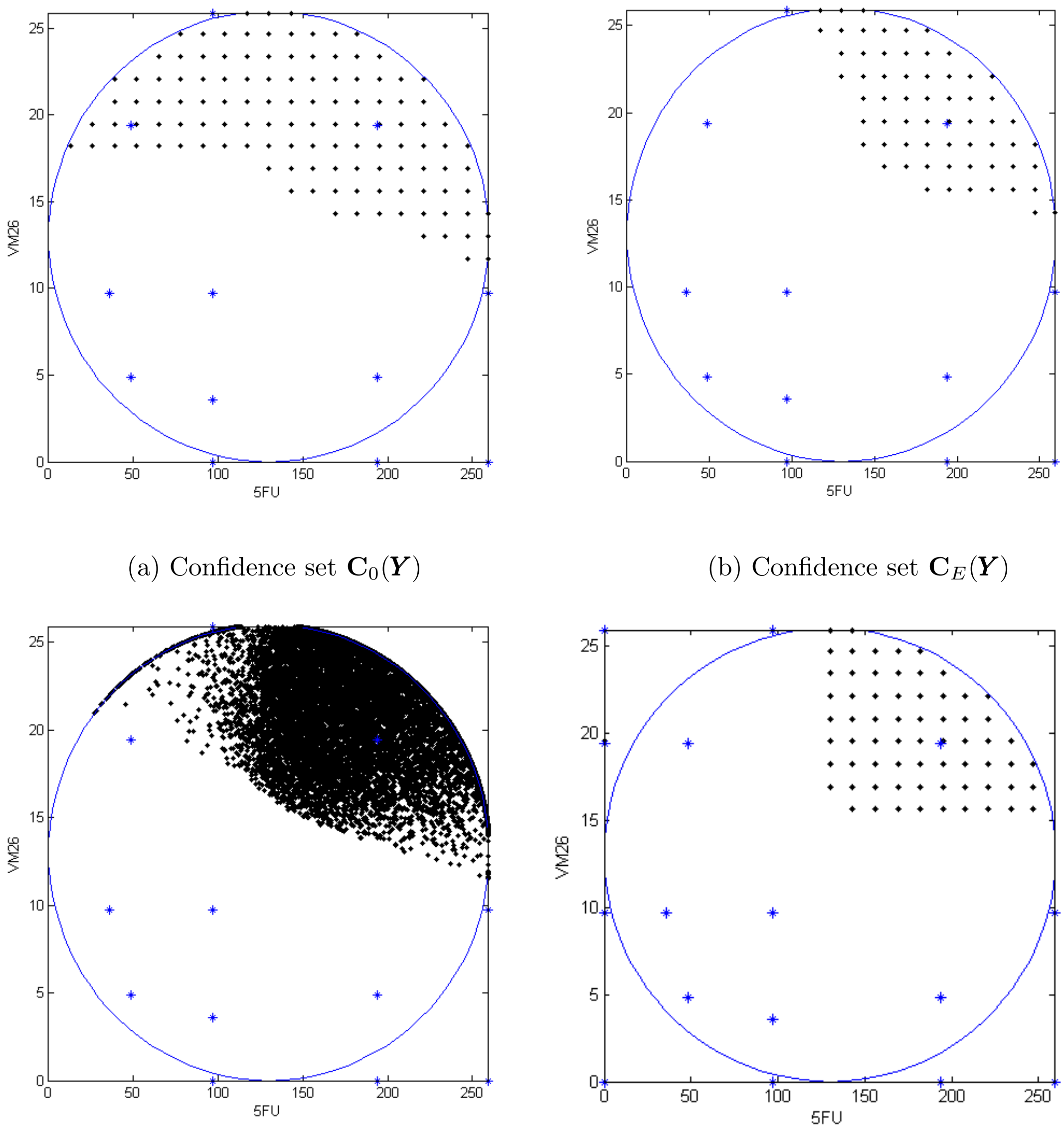
and analyzed using a proportional hazard model

$$h(t, \boldsymbol{x}) = \lambda(t) exp(f(\boldsymbol{x}, \boldsymbol{\theta}))$$

where $f(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2$. The usual model fitting procedure suggests that this proportional hazard model fits the observed data well. The parameter estimates are given by

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} 0.5702 \\ 0.7712 \\ 0.0279 \\ -0.2695 \\ 0.2196 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 0.0284 & 0.0124 & 0.0198 & 0.0044 & 0.0199 \\ 0.0124 & 0.0297 & 0.0041 & 0.0205 & 0.0200 \\ 0.0198 & 0.0041 & 0.0699 & -0.0037 & 0.0240 \\ 0.0044 & 0.0205 & -0.0037 & 0.0747 & 0.0220 \\ 0.0199 & 0.0200 & 0.0240 & 0.0220 & 0.0532 \end{pmatrix}.$$

It is clearly of interest to construct a confidence set for the dose combination $\boldsymbol{k}$ at which the hazard function $h(t, \boldsymbol{x})$ is minimized, i.e., $f(\boldsymbol{x}, -\boldsymbol{\theta})$ is maximized. We consider the constrained region $\{\boldsymbol{x}^T \boldsymbol{x} \leq 1\}$ following Stablein et al. (1983). Using $\alpha = 0.05$, resolution $d = 0.1$ and simulation $N = 10,000$, the asymptotic confidence




(a) Confidence set $\mathbf{C}_0(\boldsymbol{Y})$

(b) Confidence set $\mathbf{C}_E(\boldsymbol{Y})$

(c) Confidence set $\mathbf{C}_c(\boldsymbol{Y})$

(d) Confidence set using PCD's critical value

Figure 5.2: The 95% confidence sets in Example 5.2.2.

regions $\mathbf{C}_E(\boldsymbol{Y})$ and $\mathbf{C}_0(\boldsymbol{Y})$ are computed and depicted in Figure 5.2. Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$ is also depicted. It is clear that $\mathbf{C}_E(\boldsymbol{Y})$ is smaller than $\mathbf{C}_0(\boldsymbol{Y})$. The computation time for $\mathbf{C}_0(\boldsymbol{Y})$ is 74 seconds while for $\mathbf{C}_E(\boldsymbol{Y})$ is 11.2 hours and for $\mathbf{C}_c(\boldsymbol{Y})$ is 55 minutes. Figure 5.2(d) depicts the asymptotic confidence set using PCD's critical value. However, it is too small due to the incorrect critical value.

Table 5.2: 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.

## 5.3 A Becker's H1 Model

In this section, we consider a Becker's H1 Model (Becker, 1968, 1978)

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} min(x_1, x_2) + \beta_{13} min(x_1, x_3) + \beta_{23} min(x_2, x_3) + e \quad (5.13)$$

where $x_1, x_2, x_3 \geq 0$ satisfy $x_1 + x_2 + x_3 = 1$, $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}]^T$ is the unknown parameter and $e \sim N(0, \sigma^2)$ is the random error. Let $\boldsymbol{x} = (x_1, x_2)$ and $f(\boldsymbol{x}, \boldsymbol{\theta}) = \beta_3 + [\mathbf{z}(\boldsymbol{x})]^T \boldsymbol{\theta}^0$, where

$$\boldsymbol{\theta}^0 = [\beta_1 - \beta_3, \beta_2 - \beta_3, \beta_{12}, \beta_{13}, \beta_{23}]^T,$$
$$\mathbf{z}(\boldsymbol{x}) = [x_1, x_2, min(x_1, x_2), min(x_1, 1 - x_1 - x_2), min(x_2, 1 - x_1 - x_2)]^T,$$

then the Becker's H1 Model can be written as

$$Y = f(\boldsymbol{x}, \boldsymbol{\theta}) + e. \quad (5.14)$$

Note that unlike Generalized Linear Models and Cox Proportional Hazard model, the regression function in Becker's H1 Model is linear in the parameter $\boldsymbol{\theta}$ and so the parameters can be estimated from least squares estimation as usual. However, it is different from ordinary linear functions since $f(\boldsymbol{x}, \boldsymbol{\theta})$ is not differentiable in the covariate region $R_{cons}$ where

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

We want to find the confidence set for a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ in $R_{cons}$.

### 5.3.1 Method

Neyman's Theorem can be applied directly to construct a confidence set for $\boldsymbol{k}$, a maximum point of $f(\boldsymbol{x}, \boldsymbol{\theta})$ in $\boldsymbol{x} \in R_{cons}$. Following the acceptance set in Equation (5.1), a natural $(1-\alpha)$ level acceptance set for testing $H_0 : \boldsymbol{k} = \boldsymbol{k}^o$ for each $\boldsymbol{k}^o \in R_{cons}$ can be constructed as

$$\begin{aligned} A(\boldsymbol{k}^o) &= \left\{ \boldsymbol{Y} : f(\boldsymbol{k}^o, \hat{\boldsymbol{\theta}}) - f(\boldsymbol{x}, \hat{\boldsymbol{\theta}}) \geq -c(\boldsymbol{k}^o)\hat{\sigma} \parallel [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \parallel, \ \forall \boldsymbol{x} \in R_{cons} \right\} \\ &= \left\{ \boldsymbol{Y} : \inf_{\boldsymbol{x} \in R_{cons} \backslash \boldsymbol{k}^o} \frac{[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \boldsymbol{T}}{\parallel [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \parallel} \geq -c(\boldsymbol{k}^o) \right\}. \end{aligned}$$

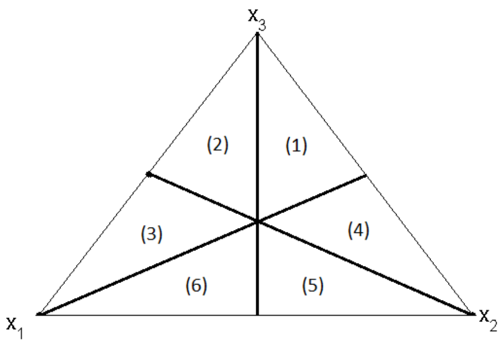



Figure 5.3: The partition of the covariate region $R_{cons}$.

To compute the critical value $c(\boldsymbol{k}^o)$, we employ the simulation-based method described in Section 4.1.2. Let

$$h(\boldsymbol{k}^o, \boldsymbol{x}) = \frac{[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P}\boldsymbol{T}}{\| [\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P} \|}, \tag{5.15}$$

where $\boldsymbol{T} = \mathbf{P}^{-1}(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}^0)/\hat{\sigma} \sim T_{p,n-p-1}$. Next, we compute the infimum of $h(\boldsymbol{k}^o, \boldsymbol{x})$ in each realization of $\boldsymbol{T}$.

Note that, the covariate region $R_{cons}$ can be partitioned into the following 6 parts: (1) $x_1 \le x_2 \le 1 - x_1 - x_2$; (2) $x_2 \le x_1 \le 1 - x_1 - x_2$; (3) $x_2 \le 1 - x_1 - x_2 \le x_1$; (4) $x_1 \le 1 - x_1 - x_2 \le x_2$; (5) $1 - x_1 - x_2 \le x_1 \le x_2$; (6) $1 - x_1 - x_2 \le x_2 \le x_1$, as shown in Figure 5.3 and in each of the 6 parts function $\mathbf{z}(\boldsymbol{x})$ is differentiable and has a specific expression, so does its derivative $\mathbf{z}_i(\boldsymbol{x})$ (see Table 5.3). Therefore, in each of the parts, the infimum of $h(\boldsymbol{k}^o, \boldsymbol{x})$ can be attained either at its stationary point(s) that lies in $R_{cons}$, or on the boundary.

We first investigate the stationary points of $h(\boldsymbol{k}^o, \boldsymbol{x})$ by letting

$$\partial h(\boldsymbol{k}^o, \boldsymbol{x})/\partial x_i = 0, i = 1, 2,$$

which gives

$$[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P}\boldsymbol{T}\mathbf{z}_i(\boldsymbol{x})\mathbf{P}^2[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})] - [\mathbf{P}\mathbf{z}_i(\boldsymbol{x})]^T \boldsymbol{T}[\mathbf{z}(\boldsymbol{k}^o) - \mathbf{z}(\boldsymbol{x})]^T \mathbf{P}^2[\mathbf{z}(\boldsymbol{k}^o)$$
$$-\mathbf{z}(\boldsymbol{x})] = 0 \tag{5.16}$$

Table 5.3: The expressions of $\mathbf{z}(\boldsymbol{x})$ and $\mathbf{z}_i(\boldsymbol{x})$ in each of the 6 parts

| Part | $\mathbf{z}(\boldsymbol{x})$ | $\mathbf{z}_1(\boldsymbol{x})$ | $\mathbf{z}_2(\boldsymbol{x})$ |
|---|---|---|---|
| (1) | $[x_1, x_2, x_1, x_1, x_2]^T$ | $[1, 0, 1, 1, 0]^T$ | $[0, 1, 0, 0, 1]^T$ |
| (2) | $[x_1, x_2, x_2, x_1, x_2]^T$ | $[1, 0, 0, 1, 0]^T$ | $[0, 1, 1, 0, 1]^T$ |
| (3) | $[x_1, x_2, x_2, 1 - x_1 - x_2, x_2]^T$ | $[1, 0, 0, -1, 0]^T$ | $[0, 1, 1, -1, 1]^T$ |
| (4) | $[x_1, x_2, x_1, x_1, 1 - x_1 - x_2]^T$ | $[1, 0, 1, 1, -1]^T$ | $[0, 1, 0, 0, -1]^T$ |
| (5) | $[x_1, x_2, x_1, 1 - x_1 - x_2, 1 - x_1 - x_2]^T$ | $[1, 0, 1, -1, -1]^T$ | $[0, 1, 0, -1, -1]^T$ |
| (6) | $[x_1, x_2, x_2, 1 - x_1 - x_2, 1 - x_1 - x_2]^T$ | $[1, 0, 0, -1, -1]^T$ | $[0, 1, 1, -1, -1]^T$ |

where $\mathbf{z}_i(\boldsymbol{x}) = \partial \mathbf{z}(\boldsymbol{x})/\partial x_i$. The left-hand side of Equation (5.16) can be further expressed as

$$\begin{aligned}
&\mathbf{z}(\boldsymbol{x})^T\{\mathbf{P}\boldsymbol{T}\mathbf{z}_i(\boldsymbol{x})\mathbf{P}^2 - \mathbf{P}^2[\mathbf{P}\mathbf{z}_i(\boldsymbol{x})]^T\boldsymbol{T}\}\mathbf{z}(\boldsymbol{x}) + [\mathbf{P}\mathbf{z}(\boldsymbol{k}^o)]^T\boldsymbol{T}\mathbf{z}_i(\boldsymbol{x})^T\mathbf{P}^2\mathbf{z}(\boldsymbol{k}^o) \\
&-[\mathbf{P}\mathbf{z}_i(\boldsymbol{x})]^T\boldsymbol{T}\mathbf{z}(\boldsymbol{k}^o)^T\mathbf{P}^2\mathbf{z}(\boldsymbol{k}^o) + \{2[\mathbf{P}\mathbf{z}_i(\boldsymbol{x})]^T\boldsymbol{T}\mathbf{z}(\boldsymbol{k}^o)^T\mathbf{P}^2 - \mathbf{z}_i(\boldsymbol{x})\mathbf{P}^2\mathbf{z}(\boldsymbol{k}^o)(\mathbf{P}\boldsymbol{T})^T \\
&-[\mathbf{P}\mathbf{z}(\boldsymbol{k}^o)]^T\boldsymbol{T}\mathbf{z}_i(\boldsymbol{x})\mathbf{P}^2\}\mathbf{z}(\boldsymbol{x}). \qquad (5.17)
\end{aligned}$$

Note that Equation (5.17) is a bivariate quadratic function of $\boldsymbol{x} = (x_1, x_2)$ in each of the 6 parts, hence the analytical solution(s) of Equation (5.16), if exist, can be found. However, one can still use numerical method to solve the equations.

Then we check the value of $h(\boldsymbol{k}^o, \boldsymbol{x})$ on the boundary of each part. Note that the boundary is made of three segments on which $h(\boldsymbol{k}^o, \boldsymbol{x})$ is a univariate function. Therefore its infimum can be found according to Section 3.1.

Finally, the infimum of $h(\boldsymbol{k}^o, \boldsymbol{x})$ in the covariate region $R_{cons}$ is attained by the minimum value of $h(\boldsymbol{k}^o, \boldsymbol{x})$ evaluated at its stationary point and that on the boundary in all of the 6 parts.

We generate $N$ independent $\boldsymbol{T} \sim T_{p,n-p-1}$ and use the $[\alpha \times N]th$ infimum as an approximation to $-c(\boldsymbol{k}^o)$. Then a $(1-\alpha)$ level confidence set can be constructed by using Neyman's Theorem, that is, by inverting $A(\boldsymbol{k}^o)$ with $\boldsymbol{k}^o \in R_{cons}$.

### 5.3.2 Example

Table 5.4: The formulation components (%) and the response

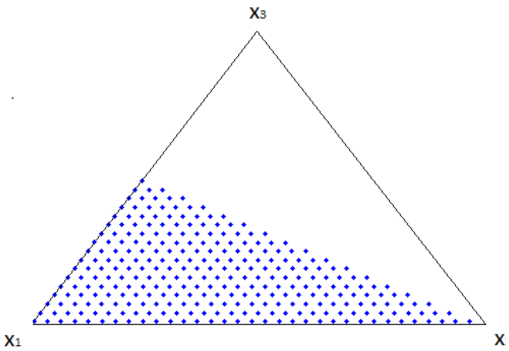
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | 0.5 | 0.5 | 0 | 0.333 | 0.666 | 0.167 | 0.167 | 0.333 |
| $x_2$ | 0 | 1 | 0 | 0.5 | 0 | 0.5 | 0.333 | 0.167 | 0.666 | 0.167 | 0.333 |
| $x_3$ | 0 | 0 | 1 | 0 | 0.5 | 0.5 | 0.333 | 0.167 | 0.666 | 0.666 | 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 |

This example considers the formulation of a controlled-release drug substance to aid in obtaining more uniform blood levels (Frisbee and McGinity, 1994). The formulation components were recorded in Table 5.4, with $x_1$ is the percentage of Pluronic F68, $x_2$ is the percentage of polyoxyethlene 40 monostearate, and $x_3$ is the percentage of polyoxyethylene sorbitan fatty acid ester NF. The response $Y$ is the observed glass transition temperature ($^{O}C$) (the smaller the better). The interest lies in constructing a confidence set for the formulation factors that minimize the response in the region constrained by $x_1 + x_2 + x_3 = 1$ and $x_1, x_2, x_3 \geq 0$. Hence, we construct a confidence set for a minimum point of function $f(\boldsymbol{x}, \boldsymbol{\theta})$ in Equation (5.14), that is, a maximum point of function $f(\boldsymbol{x}, -\boldsymbol{\theta})$, in the constrained region

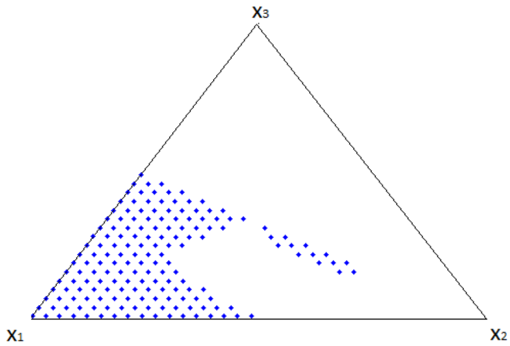
$$R_{cons} = \{\boldsymbol{x} = (x_1, x_2) : x_1 + x_2 \leq 1, \quad x_1, x_2 \geq 0\}.$$

The response surface model that gives a good fit was a Becker's H1 model in Equation (5.13) (Becker, 1968). Following the inference earlier, we use its equivalent expression (5.14) instead. The parameter estimates are given by

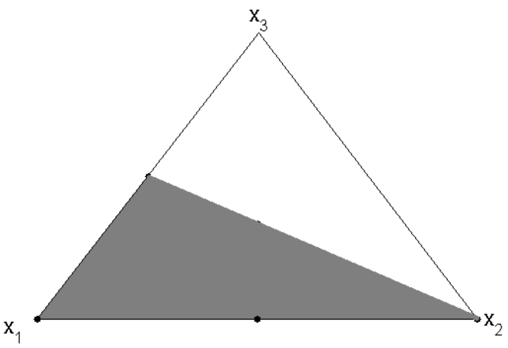
$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} -36.2719 \\ 17.6436 \\ 20.8057 \\ 2.8033 \\ 18.0089 \\ -9.7532 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 0.7936 & -0.8008 & -0.8008 & 0.1624 & -0.6384 & -0.6384 \\ -0.8008 & 1.6013 & 0.8006 & -0.7999 & 0.0007 & 0.8014 \\ -0.8008 & 0.8006 &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



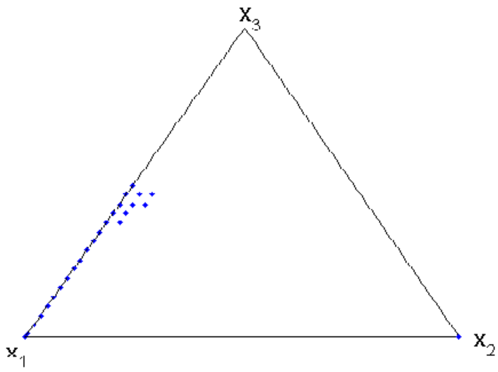

(a) Confidence set $\mathbf{C}_0(\boldsymbol{Y})$




(b) Confidence set $\mathbf{C}_E(\boldsymbol{Y})$




(c) Confidence set $\mathbf{C}_c(\boldsymbol{Y})$




(d) Confidence set using PCD's critical value

Figure 5.4: The 95% confidence sets in Example 5.3.2.

Using $\alpha = 0.05$, resolution $d = 0.03$ and simulation $N = 5,000$, the confidence regions $\mathbf{C}_E(\boldsymbol{Y})$ and $\mathbf{C}_0(\boldsymbol{Y})$ are computed and depicted in Figure 5.4. It is clear that $\mathbf{C}_E(\boldsymbol{Y})$ is smaller than $\mathbf{C}_0(\boldsymbol{Y})$ and $\mathbf{C}_c(\boldsymbol{Y})$ (see Figure 5.4(c)). The computation time for $\mathbf{C}_0(\boldsymbol{Y})$ is 44 seconds while for $\mathbf{C}_E(\boldsymbol{Y})$ is 24 hours and for $\mathbf{C}_c(\boldsymbol{Y})$ is 1.5 hour. Figure 5.4(d) depicts the asymptotic confidence set using PCD's critical value. However, it is too small due to the incorrect critical value.

## 5.4 Summary

In this chapter, we have extended the method of constructing a confidence set for a maximum point of functions in linear regression models to other models. In particular, the distributional assumption

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \Sigma)$$

holds asymptotically for many parametric and semi-parametric models. As long as the estimates of parameters are obtained, an asymptotic confidence set for a maximum point can be constructed by using our method.

Generalized linear models, Cox's proportional hazard models and a Becker's H1 model are studied and analyzed in this chapter. From the examples in each section, we conclude that our confidence set $\mathbf{C}_E(\boldsymbol{Y})$ is always smaller thus better than Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$ and the conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$. The computation of $\mathbf{C}_E(\boldsymbol{Y})$ takes generally the longest time among the three confidence sets. However, the difference is subject to the linear function involved in the model, that is, if the linear function involved is a univariate quadratic one then the time difference in computation will not be substantial.

# Chapter 6

# Conclusion and Future Works

In this chapter, we give a summary of the work presented in this thesis and provide some possible areas for future studies.

## 6.1 Summary

The construction of a confidence set for a maximum point of a function is an important statistical issue with immediate applications in many real problems. The only confidence set available in the statistical literature so far that guarantees a $(1-\alpha)$ confidence level is Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$. However, Rao's confidence set is a conservative one and sometimes it can be very conservative. In this thesis one exact confidence set $\mathbf{C}_E(\boldsymbol{Y})$ and one conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ for a maximum point of a function in a given covariate region are provided. These two new confidence sets are constructed by using Neyman's Theorem, that is, by inverting a family of acceptance sets of hypothesis tests.

Chapters 3, 4 and 5 have considered three types of models separately. Starting with a simple linear regression model, the construction method of an exact confidence set has been elaborated with details and then the theory for the general univariate polynomial regression models are derived. The bivariate quadratic regression models has been investigate in Chapter 4 and the method can be generalized to multiple regression functions. Although in that case the computation is expected to be much more intensive. In Chapter 5, we have extended the method to other models which involve a linear function. An exact $(1-\alpha)$ level confidence set can be constructed if the estimates of the parameters are normally distributed. An asymptotic $(1-\alpha)$ level confidence set is constructed if the estimates of the parameters are asymptotically normally distributed.

Examples of confidence sets constructed using both simulated and real data sets are provided and discussed throughout Chapters 3 to 5. These examples serve to illustrate the method and demonstrate the wide applicability of the confidence sets developed in this thesis. In all the examples, the new exact confidence set $\mathbf{C}_E(\boldsymbol{Y})$ is almost always (and in some cases substantially) smaller and so better than the new conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ and Rao's conservative confidence set $\mathbf{C}_c(\boldsymbol{Y})$. This is expected since the new exact confidence set $\mathbf{C}_E(\boldsymbol{Y})$ is purpose-built while the new conservative confidence set $\mathbf{C}_0(\boldsymbol{Y})$ uses a conservative critical constant and Rao's confidence set $\mathbf{C}_c(\boldsymbol{Y})$ is a by-product of the standard confidence ellipsoid for the coefficient vector $\boldsymbol{\theta}$.

The new exact confidence set $\mathbf{C}_E(\boldsymbol{Y})$ is in general most demanding computationally due to the simulation in computing the critical constants. However, it is the only confidence set for a maximum point that guarantees an exact $(1-\alpha)$ confidence level. Therefore, the construction method developed in this thesis is recommended when an exact confidence set for a maximum point is of interest.

## 6.2 Future Works

As we mentioned earlier, the method developed in this thesis can be applied to the construction of a confidence set for a maximum point of a function in some models of a more general form, such as a multiple quadratic regression function. Indeed, they can also be applied to many parametric and semi-parametric models that involve a linear regression function as well as quantile regression models (cf. Koenker, 2005) where the $q$-quantile may be modeled by $f(\boldsymbol{x}, \boldsymbol{\theta})$. Hence one possible area of the future work is to investigate and construct confidence sets for a large variety of models.

However, the computation for the confidence set of a maximum point for a multivariate function will be much more intensive than for the univariate and bivariate functions. Therefore, another possible area of the future work is to construct exact confidence sets for a maximum point by inverting a range of acceptance sets in different forms, such as

$$A(\boldsymbol{k}_0) = \left\{ \boldsymbol{Y} : f(\boldsymbol{k}_0, \hat{\boldsymbol{\theta}}) - f(\boldsymbol{x}, \hat{\boldsymbol{\theta}}) \geq -c(\boldsymbol{k}_0)\hat{\sigma}, \forall \boldsymbol{x} \in \chi \right\}. \tag{6.1}$$

The inequality in the acceptance set (6.1) is simpler than the one we used in this thesis, say, acceptance set (4.1), because the right-hand side of (6.1) is of a constant form while (4.1) has a hyperbolic form. Hence the computation should be less intensive in constructing an exact confidence set.

# Appendix
# The construction of confidence sets using only the summary statistics in polynomial regression

In this section, we demonstrate how to construct a confidence set using only the summary statistics (as given in Example 8 of Chapter 3) in univariate polynomial regression.

Suppose we have observations $(Y_{i,j}, x_i)$, $i = 1, 2, \cdots, k$; $j = 1, 2, \cdots, n_i$. The regression model is

$$Y_{i,j} = f(x_i, \boldsymbol{\theta}) = \boldsymbol{x}_i \boldsymbol{\theta} + e_{i,j}$$

for $i = 1, 2, \cdots, k$; $j = 1, 2, \cdots, n_i$, where $e_{i,j} \sim N(0, \sigma^2)$ and

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}, \ \boldsymbol{x}_i = (1, x_i, x_i^2, \cdots, x_i^p).$$

Alternatively, we write the model in the matrix form

$$\boldsymbol{Y} = X\boldsymbol{\theta} + \boldsymbol{e}$$

where

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} y_{1,1} \\ \vdots \\ y_{1,n_1} \\ \vdots \\ y_{k,1} \\ \vdots \\ y_{k,n_k} \end{pmatrix}, \ \boldsymbol{e} = \begin{pmatrix} e_{1,1} \\ \vdots \\ e_{1,n_1} \\ \vdots \\ e_{k,1} \\ \vdots \\ e_{k,n_k} \end{pmatrix},$$

and

$$X_i = \begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{x}_i \\ \vdots \\ \boldsymbol{x}_i \end{pmatrix} = \begin{pmatrix} 1 & x_i & x_i^2 & \cdots & x_i^p \\ 1 & x_i & x_i^2 & \cdots & x_i^p \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_i & x_i^2 & \cdots & x_i^p \end{pmatrix}$$

is a $n_i \times (p+1)$ matrix. Then, the least squares estimate of $\boldsymbol{\theta}$ is given by

$$
\begin{aligned}
\hat{\boldsymbol{\theta}} &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{Y} \\
&= \left( (X_1^T X_2^T \cdots X_k^T) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix} \right) (X_1^T X_2^T \cdots X_k^T) \begin{pmatrix} y_{1,1} \\ y_{1,2} \\ \vdots \\ y_{k,n_k} \end{pmatrix} \\
&= (\sum_{i=1}^{k} X_i^T X_i^T)^{-1} (\sum_{i=1}^{k} X_i{}^T \boldsymbol{y}_i) \\
&= (\sum_{i=1}^{k} X_i{}^T X_i)^{-1} \begin{pmatrix} \sum_{i=1}^{k} n_i \bar{y}_i \\ \sum_{i=1}^{k} n_i x_i \bar{y}_i \\ \vdots \\ \sum_{i=1}^{k} n_i x_i^p \bar{y}_i \end{pmatrix}
\end{aligned}
$$

where

$$
\boldsymbol{y}_i = \begin{pmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,n_i} \end{pmatrix}, \bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{i,j} \text{ and } \bar{y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{i,j} = \frac{1}{n}\sum_{i=1}^{k} n_i \bar{y}_i.
$$

Denote $n = \sum_{i=1}^{k} n_i$ and $\bar{\boldsymbol{Y}} = \bar{y}\,(1, 1, \cdots, 1)^T$ is a $n \times 1$ vector, then the estimate of $\sigma^2$ is given by

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-p-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})^T(\boldsymbol{Y} - \hat{\boldsymbol{Y}}) \\
&= \frac{1}{n-p-1} SSE \\
&= \frac{1}{n-p-1}(SST - SSR)
\end{aligned}
$$

where $\hat{\boldsymbol{Y}} = X\hat{\boldsymbol{\theta}} = X(\sum_{i=1}^{k} X_i{}^T X_i)^{-1} \begin{pmatrix} \sum_{i=1}^{k} n_i \bar{y}_i \\ \sum_{i=1}^{k} n_i x_i \bar{y}_i \\ \vdots \\ \sum_{i=1}^{k} n_i x_i^p \bar{y}_i \end{pmatrix}$, the total sum of squares is given by

$$\begin{aligned} SST &= (\boldsymbol{Y} - \bar{\boldsymbol{Y}})^T(\boldsymbol{Y} - \bar{\boldsymbol{Y}}) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{i,j} - \bar{y})^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{i,j} - \bar{y}_i + \bar{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{i,j} - \bar{y}_i)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^{k}(n_i - 1)s_i^2 + \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^{k}(n_i - 1)s_i^2 + \sum_{i=1}^{k} n_i\bar{y}_i^2 - n\bar{y}^2, \end{aligned}$$

and the residual sum of squares is given by

$$\begin{aligned} SSR &= (\bar{\boldsymbol{Y}} - \hat{\boldsymbol{Y}})^T(\bar{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}) \\ &= \bar{\boldsymbol{Y}}^T\bar{\boldsymbol{Y}} - 2\hat{\boldsymbol{Y}}^T\bar{\boldsymbol{Y}} + \hat{\boldsymbol{Y}}^T\hat{\boldsymbol{Y}} \\ &= -n\bar{y}^2 + \hat{\boldsymbol{Y}}^T\hat{\boldsymbol{Y}}. \end{aligned}$$

Thus, we have

$$\hat{\sigma}^2 \;=\; \frac{1}{n-p-1}\left(\sum_{i=1}^{k}(n_i - 1)s_i^2 + \sum_{i=1}^{k} n_i\bar{y}_i^2 - \hat{\boldsymbol{Y}}^T\hat{\boldsymbol{Y}}\right).$$

Therefore, we obtain the estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ based only on the summary statistics of a data set (observations of the independent variable, sample size, sample mean and sample standard deviation). Then, a $(1-\alpha)$ level confidence set for a maximum point of the regression function can be constructed as

$$C(\boldsymbol{Y}) = \{k_0 \in [a, b] : \boldsymbol{Y} \in A(k_0)\}$$

where

$$A(k_0) = \{\boldsymbol{Y} : f(k_0, \hat{\boldsymbol{\theta}}) - f(x, \hat{\boldsymbol{\theta}}) \geq -c(k_0)\hat{\sigma} v_p(k_0, x, \hat{\boldsymbol{\theta}}), \forall x \in [a, b] \setminus k_0\}$$

is the corresponding $(1 - \alpha)$ level acceptance set.

# References

Becker, N. (1968) Models for the response of a mixture. *J. Roy. Statist. Soc*, **B**, 349–358.

— (1978) Models and designs for experiments with mixtures. *Austral. J. Statist.*, **20**, 195–208.

Bliss, C. I. (1970) *Statistics in biology.* McGraw-Hill Book Company.

Box, G. 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.

Breslow, N. (1970) A generalized kruskal-wallace test for comparing k samples subjects tu unequal patterns of censorship. *Biometrika*, 579–594.

Cahya, S., Del Cstillo, E. and Peterson, 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., Chinchilli, V., Campbell, E. and Wampler, G. (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.

Carter, W., Wampler, G. and Stablein, D. (1983) *Regression Analysis of Survival Data in Cancer Chemotherapy.* Marcel Dekker.

Carter, W., Wampler, G. L., Stablein, D. and Campbell, E. (1982) Drug activity and therapeutic synergism in cancer treatment. *Cancer research*, **42**, 2963–2971.

Casella, G. and Berger, R. (2002) *Statistical Inference.* Duxbury.

Chakraborty, D., Bandyopadhyay, A., Bandopandhyay, S., Gupta, K. and Chatterjee, A. (2010) Use of response surface methodology for optimization of a shoot regeneration protocol in basilicum polystachyon. *In Vitro Cellular and Developmental Biology - Plant*, **46**, 451–459.

Cox, D. R. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.

— (1975) Partial likelihood. *Biometrika.*, **62**, 269–276.

Davison, A. and Hinkley, D. (1997) *Bootstrap Methods and their Application.* Cambridge University Press.

Del Castillo, E. (2007) *Process Optimization: A Statistical Approach.* Springer.

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**, 358–365.

Ding, R., Lin, D. and Peterson, J. (2005) A large-sample confidence band for a multi-response ridge path. *Quality and Reliability Engineering International*, **61**, 669–675.

Dobson, A. (2001) *An Introduction to Generalized Linear Models, Second Edition.* CRC Press.

Efron, B. (1977) The efciency of cox s likelihood function for censored data. *Journal of the American Statistical Association*, 557 565.

— (1979) Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, **7**, 1–26.

Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap.* CRC Press.

Faraway, J. J. (2006) *Extending the Linear Model with R.* CRC Press.

Farebrother, J. 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. and McGinity, J. (1994) Influence of nonionic surfactants on the physical and chemical properties of biodegradable pseudolatex. *Journal of pharmaceuticals and biopharmaceuticals*, 355–363.

Gibb, R., Lu, I.-L. and Carter, W. (2007) Bootstrap confidence regions for optimal operating conditions in response surface methodology. *COBRA Preprint Series.*

Gilmour, S. and Draper, N. (2003) Confidence regions around the ridge of optimal response on fitted second-order response surface. *technometrics*, **45**, 333–339.

Hancock, H. (1960) *Theory of maxima and minima.* Dover Publications.

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**, 128–136.

Hochberg, Y. and Quade, D. (1975) One-sided simultaneous confidence bounds on regression surfaces with intercepts. *Journal of the American statistical association*, **70**, 889–891.

Hoerl, R. W. (1985) Ridge analysis 25 years later. *The American Statistician*, **39**, 186–192.

Huang, Y. and Hsu, J. C. (2007) Hochberg's step-up method: cutting corners off holm's step-down method. *Biometrika*, **94**, 965–975.

Jensen, M. and Pedersen, L. (2007) The value assigned to six different rooting materials by growing pigs. *Applied animal behaviour science*, **108**, 31–34.

Kleinbaum, D. G. and Klein, M. (2005) *Survival Analysis : A Self-Learning Text, 2nd ed.* Springer.

Koenker, R. (2005) *Quantile Regression.* Cambridge University Press.

Lehmann, E. L. (1986) *Testing Statistical Hypotheses (second edition).* John Wiley.

Liu, W. (2010) *Simultaneous Inference in Regression.* CRC Press.

McCulloch, C. E. and Searle, S. R. (2001) *Generalized, Linear, and Mixed Models.* John Wiley.

Myers, R. (2009) *Response surface methodology, 3rd edition.* 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*, 333–380.

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. (1973) *Linear Statistical Inference and Its Application, Second Edition.* John Wiley.

Rom, D., Costello, R. and Connell, L. (1994) On close test procedures for dose-response analysis. *Statistics in medicine*, **13**, 1583–1596.

Selvin, S. (1998) *Modern Applied Biostatistical Methods Using S-Plus.* Oxford University Press.

Stablein, D. M., Carter, W. H. 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 Topics IV.* eds. S.S. Gupta and J.O. Berger, New York: Academic Press.

Studier, E., Dapson, R. and Bigelow, R. (1975) Analysis of polynomial functions for determining maximum or minimum conditions in biological systems. *Comparative Biochemistry and Physiology, Part A: Physiology*, **52**, 19–20.

Sullivan, M. and Sullivan, M. (2002) *College Algebra Enhanced with Graphing Utilities (3rd edition ).* Prentice Hall.

Uusipaikka, E. (2008) *Confidence Intervals in Generalized Regression Models.* CRC Press.

Weisberg, S. (2005) *Applied Linear Regression, 3rd ed.* John Wiley.

Wynn, H. P. and Bloomfield, P. (1971) Simultaneous confidence bands in regression analysis. *Journal of the Royal Statistical Society (B)*, **33**, 202–217.

Zar, J. (1999) *Biostatistical Analysis (4th edition).* Prentice Hall.