UNIVERSITY OF SOUTHAMPTON

Construction of Confidence sets with Application to Classification and Some Other Problems

by

Natchalee Srimaneekarn

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the
Mathematical Sciences
Faculty of Social, Human and Mathematical Sciences

May 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES
MATHEMATICAL SCIENCES

Doctor of Philosophy

by Natchalee Srimaneekarn

The construction of a confidence set can be applied in many problems. In this study, we are focusing on comparison and classification problems. For comparison problem, we can construct a confidence set for equivalence test and upper confidence bounds on several samples by three methods: using the theorem from Liu et al. (2009), F statistic and Studentized range statistic. For classification problem, we would like to classify a new case into its true class, based on some measurements. Five classification methods have been studied. They are logistic regression, classification tree, Bayesian method, support vector machine and the new confidence set method. The new method constructs a confidence set for the true class for a new case by inverting the acceptance sets. The advantage of this method is that the probability of correct classification is not less than $1-\alpha$. The methods are illustrated specifically with the well-known Iris data, seeds data and applied to a data set for classifying patients as normal, having fibrosis or having cirrhosis based on some measurements on blood samples. The total misclassification error and sensitivity (true positive rate) are used for comparing the methods.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my sincere gratitude to my best supervisor, Professor Wei Liu, for his great advice, guidance and a lot of encouragement and motivation. Since the beginning, he gave me this PhD study opportunity even though my background was far from Mathematics. During the study, he was patient, kind and gave me a huge knowledge when I faced with problems including some silly questions, and also in the future. You are a big part of my success.

Special thanks to Associate Professor Seksan Kiatsupaibul and Professor Anthony Hayter for everything that made me much more confident to study here. I also wish to thank the doctoral examiners, Professor Zudi Lu and and Dr.Degui Li, for all comments and good suggestions during my memorable viva exam.

I would also like to thank Faculty of Dentistry, Mahidol University that they have provided me a scholarship for my doctoral study, and also faculty staffs in Department of Anatomy, who allowed me to study Statistics.

Finally, I wish to express my heartiest thanks to my beloved family for their supporting and blessing, and to my friends for their help and support while I am far away from home.

DECLARATION OF AUTHORSHIP

I, **Miss Natchalee Srimaneekarn**, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Construction of Confidence sets with Application to Classification and Some Other Problems

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. 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;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. 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;
5. I have acknowledged all main sources of help;
6. 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;
7. None of this work has been published before submission.

Signed: ....

Date: ....

Chapter 1

Introduction

The construction of confidence set can be applied in many problems. The aim of this study is to develop efficient statistical methods using confidence set. In this study, we are focusing on comparison and classification problems. In the medical research, when developing a new drug or new treatment, the active-controlled clinical trial is used to enable the comparison of several treatments. Development of the new comparison statistical methods is an active research area (cf. Wellek, 2010). We can construct a confidence set for equivalence test by inverting technique. We are focusing in detailed in Chapter 2: Some Important Methods of Statistical Inference. Moreover, the intersection-union test can be applied with the construction of a confidence set in this problem. We also illustrate the construction of the upper confidence bounds on several samples by three methods: using the theorem from Liu et al. (2009), F statistic and Studentized range statistic in Chapter 3: Construction of the Upper Confidence Bounds on the Range of Means, and Chapter 4: on the Maximum Difference Between Two Regression Lines. For classification problem in Chapter 5, we would like to classify a new case into its true class, based on some measurements. Five classification methods have been studied. They are logistic regression, classification tree, Bayesian method, support vector machine and the new confidence set method. The new confidence set method constructs a confidence set for the true class for a new case by inverting the acceptance sets. The advantage of this method is that the probability of correct classification is not less than $1 - \alpha$.

The first chapter, introduction, is composed of the statement of the problem and introduction to the following chapters in this thesis.

The second chapter, some important methods of statistical inference, focuses on three topics:

- the intersection-union test

- construction of a confidence set by inverting a family of acceptance sets
- construction of an acceptance set by inverting a confidence set.

Each topic is composed of the mathematical proof and examples. The last part of the second topic, construction of a confidence set by inverting a family of acceptance sets, presents the comparison among the confidence sets from several methods which are constructed in this topic.

In the third chapter, the upper confidence bounds on the range of k means are constructed by three methods:

- the theorem from Liu et al. (2009)
- the F-statistic
- the studentized range statistic.

Firstly, the theorem from Liu et al. (2009) is used to construct the upper confidence bounds. Then the upper confidence bounds are constructed with two examples: $k = 2$ and $k \geq 3$. Secondly, the F-statistic is used to construct the upper confidence bounds. The mathematical proof is provided with three examples: $k = 2$, $k = 3$ and $k > 3$. The last method, the studentized range statistic is used to construct the upper confidence bounds. The mathematical proof is provided with three examples: $k = 2$, $k = 3$ and $k > 3$. The last section presents the comparison of the upper confidence bounds on several methods when σ is known. For any given σ, the F-statistic and studentized range statistic are changed in this situation. Then the Chi-squared statistic and the range statistic are the suitable adapted methods for this case. Moreover, the effects of sample sizes, unbalanced sample sizes and confidence levels on the upper confidence bounds are assessed.

In the fourth chapter, the upper confidence bounds on the maximum difference between two regression lines, $y = \beta_{01} + \beta_{11}x$ and $y = \beta_{02} + \beta_{12}x$ when σ is known, are constructed by two methods: by the theorem from Liu et al. (2009) and by Chi-squared statistic. The construction using a theorem, which is mentioned in the previous chapter, depends on several upper confidence bounds that will be computed by two methods: i) by general confidence bounds and ii) by the normal distribution method. The last section of this chapter provides some examples for comparing the upper confidence bounds from those methods.

In the fifth chapter, a confidence set is applied for classification problem. We propose the construction of a confidence set as a new alternative method for this problem. This method also use the theorem 2.2.1 in Chapter 2 for constructing a confidence set. The advantage of this method is that the probability of correct classification is not less than $1 - \alpha$. Five classification methods for classification into two classes and three classes have been studied. They are

- new confidence set method
- classification tree
- logistic regression
- Bayesian method
- support vector machine.

Moreover, three real data examples: the data of Cirrhosis patients from a hospital, the well known iris data and seeds data are used for illustrating and comparing these five classification methods. The total misclassification error and sensitivity (true positive rate) are used for comparing the methods.

In the last chapter, conclusions and future work is conclude the whole study and also suggest some points that can be further study.

Chapter 2

Some Important Methods of Statistical Inference

The construction of a confidence set in our study needs some statistical methods to support the methodology. Then, three methods of statistical inference have been considered. There are:

- The intersection-union test,
- Construction of a confidence set by inverting acceptance sets,
- Construction of an acceptance set from a confidence set.

2.1 Intersection Union Test

Since the test of equivalence is a special case of the intersection-union test, Berger and Hsu (1996) applied the intersection-union test to equivalence problems using the following theorem (Berger, 1982).

Theorem 2.1.1. Let Y have distribution function $f(y;\theta)$ with an unknown parameter θ in the the parameter space Θ with Θ_i subsets, $i \in \wedge$, where $\wedge$ is an index set. The hypotheses for the intersection-union test are

$$H_0 : \theta \in \bigcup_{i\in\wedge} \Theta_i \qquad \text{versus} \qquad H_a : \theta \in \bigcap_{i\in\wedge} \Theta_i^c,$$

where the individual hypotheses are

$$H_{0i} : \theta \in \Theta_i \quad \text{versus} \quad H_{ai} : \theta \in \Theta_i^c, \quad i \in \wedge.$$

Let R_i be a size α rejection region for testing individual hypotheses, H_{0i} versus H_{ai}. Then $R = \bigcap_{i \in \wedge} R_i$ is a size α rejection region for testing overall hypotheses, H_0 versus H_a.

Proof. Given $\theta \in \bigcup_{i \in \wedge} \Theta_i$, so $\theta \in \Theta_{i_0}$ for at least one $i_0 \in \wedge$. The size of the test is

$$P\left\{Y \in \bigcap_{i \in \wedge} R_i \middle| \theta \in \bigcup_{i \in \wedge} \Theta_i\right\} = P\left\{Y \in \bigcap_{i \in \wedge} R_i \middle| \theta \in \Theta_{i_0}\right\} \leq P\left\{Y \in R_{i_0} | \theta \in \Theta_{i_0}\right\} = \alpha.$$

□

Example 2.1.1. Two Samples

Let two samples $(X_{11}, X_{12}, \ldots, X_{1n})$ and $(X_{21}, X_{22}, \ldots, X_{2m})$ be the observations from two treatments which are independent and identically distributed as normal distributions with an equal variance σ^2 and means being θ_1 and θ_2 respectively. The hypotheses for equivalence testing are

$$H_0 : |\theta_1 - \theta_2| \geq \varepsilon \qquad \text{versus} \qquad H_a : |\theta_1 - \theta_2| < \varepsilon.$$

These can be written in the intersection-union test form as

$$H_0 : \theta \in \bigcup_{i=1}^{2} \Theta_i \qquad \text{versus} \qquad H_a : \theta \in \bigcap_{i=1}^{2} \Theta_i^c,$$

where

$$\begin{aligned}
\theta &= (\theta_1, \theta_2, \sigma^2), \\
\Theta &= \{(\theta_1, \theta_2, \sigma^2) : -\infty < \theta_1 < \infty, \ -\infty < \theta_2 < \infty, \ \sigma^2 > 0\}, \\
\Theta_1 &= \{(\theta_1, \theta_2, \sigma^2) : \theta_1 - \theta_2 \geq \varepsilon, \ \sigma^2 > 0\}, \\
\Theta_1^c &= \{(\theta_1, \theta_2, \sigma^2) : \theta_1 - \theta_2 < \varepsilon, \ \sigma^2 > 0\}, \\
\Theta_2 &= \{(\theta_1, \theta_2, \sigma^2) : \theta_1 - \theta_2 \leq -\varepsilon, \ \sigma^2 > 0\}, \\
\Theta_2^c &= \{(\theta_1, \theta_2, \sigma^2) : \theta_1 - \theta_2 > -\varepsilon, \ \sigma^2 > 0\}.
\end{aligned}$$

The individual hypotheses are

$$H_{01} : \theta \in \Theta_1 \qquad \text{versus} \qquad H_{a1} : \theta \in \Theta_1^c$$

and

$$H_{02} : \theta \in \Theta_2 \qquad \text{versus} \qquad H_{a2} : \theta \in \Theta_2^c.$$

Let R_i be a size α rejection region for testing individual hypotheses, H_{0i} against H_{ai}. The rejection region R_1 is given by

$$\frac{\frac{(\overline{x}_1-\overline{x}_2-\varepsilon)}{\sqrt{1/n+1/m}}}{\sqrt{\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2+\sum_{i=1}^{m}(x_{2i}-\overline{x}_2)^2}{n+m-2}}} < -t_{n+m-2,\alpha},$$

and the rejection region R_2 is given by

$$\frac{\frac{(\overline{x}_1-\overline{x}_2+\varepsilon)}{\sqrt{1/n+1/m}}}{\sqrt{\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2+\sum_{i=1}^{m}(x_{2i}-\overline{x}_2)^2}{n+m-2}}} > t_{n+m-2,\alpha}.$$

Individual test 1:
The individual hypotheses are

$$H_{01} : \theta \in \Theta_1 \qquad \text{versus} \qquad H_{a1} : \theta \in \Theta_1^c.$$

H_{01} will be rejected if and only if

$$\frac{\frac{(\overline{x}_1-\overline{x}_2-\varepsilon)}{\sqrt{1/n+1/m}}}{\sqrt{\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2+\sum_{i=1}^{m}(x_{2i}-\overline{x}_2)^2}{n+m-2}}} < -t_{n+m-2,\alpha}.$$

Proof. The size of the test is

$$\begin{aligned}
& P\left\{\frac{\frac{(\overline{x}_1-\overline{x}_2-\varepsilon)}{\sqrt{\sigma^2/n+\sigma^2/m}}}{\sqrt{\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2+\sum_{i=1}^{m}(x_{2i}-\overline{x}_2)^2}{\sigma^2(n+m-2)}}} < -t_{n+m-2,\alpha}\middle|\theta_1-\theta_2\geq\varepsilon\right\} \\
= \;& P\left\{\frac{\frac{(\overline{x}_1-\overline{x}_2-\varepsilon)-(\theta_1-\theta_2-\varepsilon)+(\theta_1-\theta_2-\varepsilon)}{\sqrt{\sigma^2/n+\sigma^2/m}}}{\sqrt{\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2+\sum_{i=1}^{m}(x_{2i}-\overline{x}_2)^2}{\sigma^2(n+m-2)}}} < -t_{n+m-2,\alpha}\middle|\theta_1-\theta_2-\varepsilon\geq 0\right\} \\
= \;& P\left\{\frac{Z+\frac{(\theta_1-\theta_2-\varepsilon)}{\sqrt{\sigma^2/n+\sigma^2/m}}}{\sqrt{\frac{\chi^2_{n+m-2}}{(n+m-2)}}} < -t_{n+m-2,\alpha}\middle|\theta_1-\theta_2-\varepsilon\geq 0\right\} \qquad (2.1) \\
\leq \;& P\left\{\frac{Z}{\sqrt{\frac{\chi^2_{n+m-2}}{(n+m-2)}}} < -t_{n+m-2,\alpha}\middle|\theta_1-\theta_2-\varepsilon\geq 0\right\} \qquad (2.2) \\
= \;& P\left\{\frac{Z}{\sqrt{\frac{\chi^2_{n+m-2}}{(n+m-2)}}} < -t_{n+m-2,\alpha}\right\} \\
= \;& \alpha,
\end{aligned}$$

where $Z \sim N(0,1)$. The inequality of (2.2) holds because the probability in (2.1) is maximised when $(\theta_1 - \theta_2 - \varepsilon) = 0$. Hence, the size of the individual test 1 is α. $\square$

Individual test 2:
The individual hypotheses are

$$H_{02} : \theta \in \Theta_2 \qquad \text{versus} \qquad H_{a2} : \theta \in \Theta_2^c.$$

H_{02} will be rejected if and only if

$$\frac{\frac{(\overline{x}_1 - \overline{x}_2 + \varepsilon)}{\sqrt{1/n+1/m}}}{\sqrt{\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2+\sum_{i=1}^{m}(x_{2i}-\overline{x}_2)^2}{n+m-2}}} > t_{n+m-2,\alpha}.$$

Proof is similar to individual test 1.
Overall test:
The overall hypotheses are

$$H_0 : \theta \in \bigcup_{i=1}^{2} \Theta_i \qquad \text{versus} \qquad H_a : \theta \in \bigcap_{i=1}^{2} \Theta_i^c.$$

Then H_0 will be rejected by the intersection-union test if and only if

$$\frac{\frac{(\overline{x}_1 - \overline{x}_2 - \varepsilon)}{\sqrt{1/n+1/m}}}{\sqrt{\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2+\sum_{i=1}^{m}(x_{2i}-\overline{x}_2)^2}{n+m-2}}} < -t_{n+m-2,\alpha}$$

and

$$\frac{\frac{(\overline{x}_1 - \overline{x}_2 + \varepsilon)}{\sqrt{1/n+1/m}}}{\sqrt{\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2+\sum_{i=1}^{m}(x_{2i}-\overline{x}_2)^2}{n+m-2}}} > t_{n+m-2,\alpha}$$

with the size of the test being α.

We give a proof below to show that the size of the test is exactly α.

Proof. The size of the test is

$$P\left\{\frac{\frac{(\overline{x}_1-\overline{x}_2-\varepsilon)-(\theta_1-\theta_2-\varepsilon)+(\theta_1-\theta_2-\varepsilon)}{\sqrt{\sigma^2/n+\sigma^2/m}}}{\sqrt{\frac{\chi^2_{n+m-2}}{(n+m-2)}}} < -t_{n+m-2,\alpha} \quad \text{and} \right.$$
$$\left. \frac{\frac{(\overline{x}_1-\overline{x}_2+\varepsilon)-(\theta_1-\theta_2+\varepsilon)+(\theta_1-\theta_2+\varepsilon)}{\sqrt{\sigma^2/n+\sigma^2/m}}}{\sqrt{\frac{\chi^2_{n+m-2}}{(n+m-2)}}} > t_{n+m-2,\alpha} \,\middle|\, |\theta_1-\theta_2| \geq \varepsilon \right\} \tag{2.3}$$

If σ is close to zero and $\theta_1 - \theta_2 - \varepsilon$ is zero, (2.3) will be

$$\begin{aligned} &= P\left\{ \frac{Z}{\sqrt{\frac{\chi^2_{n+m-2}}{(n+m-2)}}} < -t_{n+m-2,\alpha} \right\} \\ &= \alpha, \end{aligned}$$

where $Z \sim N(0,1)$. Hence, the size of the test is exactly α. □

Example 2.1.2. Three Samples

Let three samples, $(X_{11}, X_{12}, \ldots, X_{1n_1})$, $(X_{21}, X_{22}, \ldots, X_{2n_2})$, and $(X_{31}, X_{32}, \ldots, X_{3n_3})$, be the observations from three treatments, which are independent and identically distributed as normal distributions with an equal variance, σ^2, and means being θ_1, θ_2 and θ_3 respectively. The hypotheses we are interested in are

$$H_0 : \max_{1 \le i < j \le 3} |\theta_i - \theta_j| \ge \varepsilon \quad \text{versus} \quad H_a : \max_{1 \le i < j \le 3} |\theta_i - \theta_j| < \varepsilon.$$

These can be written in the intersection-union test form as

$$H_0 : \theta \in \bigcup_{1 \le i < j \le 3} \Theta_{ij} \qquad \text{versus} \qquad H_a : \theta \in \bigcap_{1 \le i < j \le 3} \Theta^c_{ij},$$

where

$$\begin{aligned} \theta &= (\theta_1, \theta_2, \theta_3, \sigma^2), \\ \Theta &= \{(\theta_1, \theta_2, \theta_3, \sigma^2) : -\infty < \theta_1 < \infty, \ -\infty < \theta_2 < \infty, \ -\infty < \theta_3 < \infty, \ \sigma^2 > 0\}, \\ \Theta_{ij} &= \{(\theta_1, \theta_2, \theta_3, \sigma^2) : |\theta_i - \theta_j| \ge \varepsilon, \ \sigma^2 > 0\}, \ 1 \le i < j \le 3, \\ \Theta^c_{ij} &= \{(\theta_1, \theta_2, \theta_3, \sigma^2) : |\theta_i - \theta_j| < \varepsilon, \ \sigma^2 > 0\}, \ 1 \le i < j \le 3. \end{aligned}$$

Individual test:

The individual hypotheses are

$$H_0 : |\theta_i - \theta_j| \ge \varepsilon \qquad \text{versus} \qquad H_a : |\theta_i - \theta_j| < \varepsilon.$$

The tests are given above as two groups equivalence test. Hence, $H_{0_{ij}}$ will be rejected if and only if

$$\frac{\frac{(\overline{x}_i - \overline{x}_j - \varepsilon)}{\sqrt{1/n_i + 1/n_j}}}{\sqrt{\frac{\sum_{k=1}^{n_i} (x_{ik} - \overline{x}_i)^2 + \sum_{k=1}^{n_j} (x_{jk} - \overline{x}_j)^2}{n_i + n_j - 2}}} < -t_{n_i + n_j - 2, \alpha}$$

and

$$\frac{\frac{(\overline{x}_i-\overline{x}_j+\varepsilon)}{\sqrt{1/n_i+1/n_j}}}{\sqrt{\frac{\sum_{k=1}^{n_i}(x_{ik}-\overline{x}_i)^2+\sum_{k=1}^{n_j}(x_{jk}-\overline{x}_j)^2}{n_i+n_j-2}}} > t_{n_i+n_j-2,\alpha},$$

which is of size α.

Overall test:

The overall hypotheses are

$$H_0 : \theta \in \bigcup_{1\le i<j\le 3} \Theta_{ij} \qquad \text{versus} \qquad H_a : \theta \in \bigcap_{1\le i<j\le 3} \Theta_{ij}^c.$$

Then H_0 will be rejected by the intersection-union test if and only if

$$\max_{1\le i<j\le 3} \frac{\frac{(\overline{x}_i-\overline{x}_j-\varepsilon)}{\sqrt{1/n_i+1/n_j}}}{\sqrt{\frac{\sum_{k=1}^{n_i}(x_{ik}-\overline{x}_i)^2+\sum_{k=1}^{n_j}(x_{jk}-\overline{x}_j)^2}{n_i+n_j-2}}} < -t_{n_i+n_j-2,\alpha}$$

and

$$\min_{1\le i<j\le 3} \frac{\frac{(\overline{x}_i-\overline{x}_j+\varepsilon)}{\sqrt{1/n_i+1/n_j}}}{\sqrt{\frac{\sum_{k=1}^{n_i}(x_{ik}-\overline{x}_i)^2+\sum_{k=1}^{n_j}(x_{jk}-\overline{x}_j)^2}{n_i+n_j-2}}} > t_{n_i+n_j-2,\alpha},$$

which is of size α.

We give a proof below to show the exact size of the test is α.

Proof. The size of the test is

$$P\left\{ \max_{1\le i<j\le 3} \frac{\frac{(\overline{x}_i-\overline{x}_j-\varepsilon)-(\theta_i-\theta_j-\varepsilon)+(\theta_i-\theta_j-\varepsilon)}{\sqrt{\sigma^2/n_i+\sigma^2/n_j}}}{\sqrt{\frac{\chi^2_{n_i+n_j-2}}{(n_i+n_j-2)}}} < -t_{n_i+n_j-2,\alpha} \text{ and } \min_{1\le i<j\le 3} \frac{\frac{(\overline{x}_i-\overline{x}_j+\varepsilon)-(\theta_i-\theta_j+\varepsilon)+(\theta_i-\theta_j+\varepsilon)}{\sqrt{\sigma^2/n_i+\sigma^2/n_j}}}{\sqrt{\frac{\chi^2_{n_i+n_j-2}}{(n_i+n_j-2)}}} > t_{n_i+n_j-2,\alpha} \,\middle|\, \max_{1\le i<j\le 3} |\theta_i-\theta_j| \ge \varepsilon \right\}. \tag{2.4}$$

If σ is close to zero, $\theta_1-\theta_2=\varepsilon$, and $\theta_1-\theta_3=\varepsilon/2$, (2.4) will be

$$
\begin{aligned}
& P\left\{\begin{array}{l}
\max\left(\frac{Z_{12}+0}{\sqrt{\frac{\chi^2_{n_1+n_2-2}}{(n_1+n_2-2)}}}, \frac{Z_{13}+\frac{\varepsilon/2-\varepsilon}{\sqrt{\sigma^2/n_1+\sigma^2/n_3}}}{\sqrt{\frac{\chi^2_{n_1+n_3-2}}{(n_1+n_3-2)}}}, \frac{Z_{23}+\frac{(-\varepsilon/2)-\varepsilon}{\sqrt{\sigma^2/n_2+\sigma^2/n_3}}}{\sqrt{\frac{\chi^2_{n_2+n_3-2}}{(n_2+n_3-2)}}}\right)<-t_{n_i+n_j-2,\alpha} \\
\text{and} \\
\min\left(\frac{Z_{12}+\infty}{\sqrt{\frac{\chi^2_{n_1+n_2-2}}{(n_1+n_2-2)}}}, \frac{Z_{13}+\infty}{\sqrt{\frac{\chi^2_{n_1+n_3-2}}{(n_1+n_3-2)}}}, \frac{Z_{23}+\infty}{\sqrt{\frac{\chi^2_{n_2+n_3-2}}{(n_2+n_3-2)}}}\right)>t_{n_i+n_j-2,\alpha}
\end{array}\right\} \\
= & P\left\{\max\left(\frac{Z_{12}}{\sqrt{\frac{\chi^2_{n_1+n_2-2}}{(n_1+n_2-2)}}}, \frac{Z_{13}-\infty}{\sqrt{\frac{\chi^2_{n_1+n_3-2}}{(n_1+n_3-2)}}}, \frac{Z_{23}-\infty}{\sqrt{\frac{\chi^2_{n_2+n_3-2}}{(n_2+n_3-2)}}}\right)<-t_{n_i+n_j-2,\alpha}\right\} \\
= & P\left\{\frac{Z_{12}}{\sqrt{\frac{\chi^2_{n_1+n_2-2}}{(n_1+n_2-2)}}} < -t_{n_1+n_2-2,\alpha}\right\} \\
= & \alpha,
\end{aligned}
$$

where $Z_{ij} \sim N(0,1)$. Hence, the size of the test is exactly α. □

Example 2.1.3. Two regression lines

According to Liu (2010), consider the equivalence between two regression lines,

$$m_1(x, \boldsymbol{\beta_1}) = \beta_{01} + \beta_{11}x \quad \text{and} \quad m_2(x, \boldsymbol{\beta_2}) = \beta_{02} + \beta_{12}x,$$

for $x \in [a, b]$, with $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ being the parameter vectors respectively. The intersection-union test can be used for this equivalence problem by setting the hypotheses as

$$H_0: \max_{x\in[a,b]} |m_1(x,\boldsymbol{\beta_1}) - m_2(x,\boldsymbol{\beta_2})| \geq \delta$$

versus

$$H_a: \max_{x\in[a,b]} |m_1(x,\boldsymbol{\beta_1}) - m_2(x,\boldsymbol{\beta_2})| < \delta,$$

where δ is a given positive constant.

Let

$$
\begin{aligned}
\theta &= (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2), \\
\Theta &= \{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2): \ \boldsymbol{\beta_1} \in R^2,\ \boldsymbol{\beta_2} \in R^2,\ \sigma^2 > 0\}, \\
\Theta_1 &= \left\{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2): \max_{x\in[a,b]} [m_1(x,\boldsymbol{\beta_1}) - m_2(x,\boldsymbol{\beta_2})] \geq \delta,\ \sigma^2 > 0\right\}, \\
\Theta_1^c &= \left\{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2): \max_{x\in[a,b]} [m_1(x,\boldsymbol{\beta_1}) - m_2(x,\boldsymbol{\beta_2})] < \delta,\ \sigma^2 > 0\right\}, \\
\Theta_2 &= \left\{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2): \min_{x\in[a,b]} [m_1(x,\boldsymbol{\beta_1}) - m_2(x,\boldsymbol{\beta_2})] \leq -\delta,\ \sigma^2 > 0\right\}, \\
\Theta_2^c &= \left\{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2): \min_{x\in[a,b]} [m_1(x,\boldsymbol{\beta_1}) - m_2(x,\boldsymbol{\beta_2})] > -\delta,\ \sigma^2 > 0\right\}.
\end{aligned}
$$

Individual test 1:
The individual hypotheses 1 are

$$H_{01} : \theta \in \Theta_1 \qquad \text{versus} \qquad H_{a1} : \theta \in \Theta_1^c.$$

These can be further written in the intersection-union test form as

$$H_{01} : \theta \in \Theta_1 = \bigcup_{x\in[a,b]} \Theta_{1x} \qquad \text{versus} \qquad H_{a1} : \theta \in \Theta_1^c = \bigcap_{x\in[a,b]} \Theta_{1x}^c,$$

where

$$\begin{aligned} \theta &= (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2), \\ \Theta_1 &= \left\{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2) : \max_{x\in[a,b]} [m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})] \geq \delta,\ \sigma^2 > 0\right\}, \\ \Theta_{1x} &= \{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2) : [m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})] \geq \delta,\ \sigma^2 > 0\},\ x \in [a, b], \\ \Theta_{1x}^c &= \{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2) : [m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})] < \delta,\ \sigma^2 > 0\},\ x \in [a, b]. \end{aligned}$$

The individual hypotheses $1x$ are

$$H_{01x} : m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2}) \geq \delta \quad \text{versus} \quad H_{a1x} : m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2}) < \delta,$$

which can be written as

$$H_{01x} : \theta \in \Theta_{1x} \qquad \text{versus} \qquad H_{a1x} : \theta \in \Theta_{1x}^c.$$

H_{01x} will be rejected if and only if

$$\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\boldsymbol{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\boldsymbol{2}} + t_{\nu,\alpha}\hat{\sigma}\sqrt{\boldsymbol{x}^T[({X_1}^TX_1)^{-1} + ({X_2}^TX_2)^{-1}]\boldsymbol{x}} \quad < \quad \delta.$$

Proof. The size of the test is

$$\begin{aligned} & P\left\{\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\boldsymbol{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\boldsymbol{2}} + t_{\nu,\alpha}\hat{\sigma}\sqrt{\boldsymbol{x}^T[({X_1}^TX_1)^{-1} + ({X_2}^TX_2)^{-1}]\boldsymbol{x}} < \delta \middle| \boldsymbol{x}^T\boldsymbol{\beta_1} - \boldsymbol{x}^T\boldsymbol{\beta_2} \geq \delta\right\} \\ = \;& P\left\{\frac{\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\boldsymbol{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\boldsymbol{2}} - \delta}{\hat{\sigma}\sqrt{\boldsymbol{x}^T[({X_1}^TX_1)^{-1} + ({X_2}^TX_2)^{-1}]\boldsymbol{x}}} < -t_{\nu,\alpha} \middle| \boldsymbol{x}^T\boldsymbol{\beta_1} - \boldsymbol{x}^T\boldsymbol{\beta_2} - \delta \geq 0\right\} \\ = \;& P\left\{\frac{(\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\boldsymbol{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\boldsymbol{2}}) - (\boldsymbol{x}^T\boldsymbol{\beta_1} - \boldsymbol{x}^T\boldsymbol{\beta_2}) + (\boldsymbol{x}^T\boldsymbol{\beta_1} - \boldsymbol{x}^T\boldsymbol{\beta_2} - \delta)}{\hat{\sigma}\sqrt{\boldsymbol{x}^T[({X_1}^TX_1)^{-1} + ({X_2}^TX_2)^{-1}]\boldsymbol{x}}} < -t_{\nu,\alpha} \middle| \boldsymbol{x}^T\boldsymbol{\beta_1} - \boldsymbol{x}^T\boldsymbol{\beta_2} - \delta \geq 0\right\} \\ = \;& P\left\{T_\nu + \frac{(\boldsymbol{x}^T\boldsymbol{\beta_1} - \boldsymbol{x}^T\boldsymbol{\beta_2} - \delta)}{\hat{\sigma}\sqrt{\boldsymbol{x}^T[({X_1}^TX_1)^{-1} + ({X_2}^TX_2)^{-1}]\boldsymbol{x}}} < -t_{\nu,\alpha} \middle| \boldsymbol{x}^T\boldsymbol{\beta_1} - \boldsymbol{x}^T\boldsymbol{\beta_2} - \delta \geq 0\right\} \qquad (2.5) \\ \leq \;& P\left\{T_\nu < -t_{\nu,\alpha} \mid \boldsymbol{x}^T\boldsymbol{\beta_1} - \boldsymbol{x}^T\boldsymbol{\beta_2} - \delta \geq 0\right\} \qquad (2.6) \\ = \;& P\left\{T_\nu < -t_{\nu,\alpha}\right\} \\ = \;& \alpha, \end{aligned}$$

where $T_\nu \sim t_\nu$. The inequality of (2.6) holds because the probability in (2.5) is maximised when $(\boldsymbol{x}^T\boldsymbol{\beta_1} - \boldsymbol{x}^T\boldsymbol{\beta_2} - \delta) = 0$. Hence, the size of the individual test $1x$ is α. □

Then, the individual hypothesis 1, H_{01}, will be rejected by the intersection-union test if and only if

$$\max_{x\in[a,b]} \left\{ \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_\mathbf{1} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_\mathbf{2} + t_{\nu,\alpha}\hat{\sigma}\sqrt{\boldsymbol{x}^T[({X_1}^TX_1)^{-1} + ({X_2}^TX_2)^{-1}]\boldsymbol{x}} \right\} \quad < \quad \delta,$$

which is of size α.

Individual test 2:
The individual hypotheses 2 are

$$H_{02} : \theta \in \Theta_2 \qquad \text{versus} \qquad H_{a2} : \theta \in \Theta_2^c.$$

These can be further written in the intersection-union test form as

$$H_{02} : \theta \in \Theta_2 = \bigcup_{x\in[a,b]} \Theta_{2x} \qquad \text{versus} \qquad H_{a2} : \theta \in \Theta_2^c = \bigcap_{x\in[a,b]} \Theta_{2x}^c,$$

where

$$\begin{aligned}
\theta &= (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2), \\
\Theta_2 &= \left\{ (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2) : \min_{x\in[a,b]} [m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})] \leq -\delta,\ \sigma^2 > 0 \right\}, \\
\Theta_{2x} &= \{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2) : [m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})] \leq -\delta,\ \sigma^2 > 0\},\ x \in [a, b], \\
\Theta_{2x}^c &= \{(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \sigma^2) : [m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})] > -\delta,\ \sigma^2 > 0\},\ x \in [a, b].
\end{aligned}$$

The individual tests $2x$ are

$$H_{02x} : m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2}) \leq -\delta \quad \text{versus} \quad H_{a2x} : m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2}) > -\delta,$$

which can be written as

$$H_{02x} : \theta \in \Theta_{2x} \qquad \text{versus} \qquad H_{a2x} : \theta \in \Theta_{2x}^c.$$

H_{02x} will be rejected if and only if

$$\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_\mathbf{1} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_\mathbf{2} - t_{\nu,\alpha}\hat{\sigma}\sqrt{\boldsymbol{x}^T[({X_1}^TX_1)^{-1} + ({X_2}^TX_2)^{-1}]\boldsymbol{x}} \quad > \quad -\delta.$$

Proof. The size of the test is

$$
\begin{aligned}
& P\left\{\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{2}} - t_{\nu,\alpha}\hat{\sigma}\sqrt{\boldsymbol{x}^T[(X_1{}^TX_1)^{-1} + (X_2{}^TX_2)^{-1}]\boldsymbol{x}} > -\delta \middle| \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{2}} \le -\delta\right\} \\
= \;& P\left\{\frac{\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{2}} + \delta}{\hat{\sigma}\sqrt{\boldsymbol{x}^T[(X_1{}^TX_1)^{-1} + (X_2{}^TX_2)^{-1}]\boldsymbol{x}}} > t_{\nu,\alpha} \middle| \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{2}} + \delta \le 0\right\} \\
= \;& P\left\{\frac{(\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{2}}) - (\boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{2}}) + (\boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{2}} + \delta)}{\hat{\sigma}\sqrt{\boldsymbol{x}^T[(X_1{}^TX_1)^{-1} + (X_2{}^TX_2)^{-1}]\boldsymbol{x}}} > t_{\nu,\alpha} \middle| \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{2}} + \delta \le 0\right\} \\
= \;& P\left\{T_\nu + \frac{(\boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{2}} + \delta)}{\hat{\sigma}\sqrt{\boldsymbol{x}^T[(X_1{}^TX_1)^{-1} + (X_2{}^TX_2)^{-1}]\boldsymbol{x}}} > t_{\nu,\alpha} \middle| \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{2}} + \delta \le 0\right\} \qquad (2.7) \\
\le \;& P\left\{T_\nu > t_{\nu,\alpha} \middle| \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{2}} + \delta \le 0\right\} \qquad (2.8) \\
= \;& P\left\{T_\nu > t_{\nu,\alpha}\right\} \\
= \;& \alpha.
\end{aligned}
$$

where $T_\nu \sim t_\nu$. The inequality of (2.8) holds because the probability in (2.7) is maximised when $(\boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{1}} - \boldsymbol{x}^T\boldsymbol{\beta}_{\mathbf{2}} + \delta) = 0$. Hence, the size of the individual test $2x$ is α. □

Then, the individual hypothesis 2, H_{02}, will be rejected by the intersection-union test if and only if

$$
\min_{x\in[a,b]}\left\{\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{2}} - t_{\nu,\alpha}\hat{\sigma}\sqrt{\boldsymbol{x}^T[(X_1{}^TX_1)^{-1} + (X_2{}^TX_2)^{-1}]\boldsymbol{x}}\right\} \quad > \quad -\delta,
$$

which is of size α.

Overall test:
The overall hypothesis, H_0, will be rejected by the intersection-union test if and only if

$$
\max_{x\in[a,b]}\left\{\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{2}} + t_{\nu,\alpha}\hat{\sigma}\sqrt{\boldsymbol{x}^T[(X_1{}^TX_1)^{-1} + (X_2{}^TX_2)^{-1}]\boldsymbol{x}}\right\} \quad < \quad \delta
$$

and

$$
\min_{x\in[a,b]}\left\{\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{x}^T\hat{\boldsymbol{\beta}}_{\mathbf{2}} - t_{\nu,\alpha}\hat{\sigma}\sqrt{\boldsymbol{x}^T[(X_1{}^TX_1)^{-1} + (X_2{}^TX_2)^{-1}]\boldsymbol{x}}\right\} \quad > \quad -\delta,
$$

which is of size α.

2.2 Construction of a Confidence Set by Inverting Acceptance Sets

The confidence set can be applied to an equivalence problem if the confidence set of the difference is close to zero or less than some given positive value. A confidence set can be constructed by inverting a family of acceptance sets as given by the following theorem (cf. Lehmann, 1986).

Theorem 2.2.1. Let Y have distribution function $f(y;\theta)$ with an unknown parameter θ in parameter space Θ. For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be an acceptance set of size α for testing $H_0 : \theta = \theta_0$, that is,

$$P_{\theta_0}\{Y \in A(\theta_0)\} \geq 1 - \alpha.$$

For each Y, we can construct a $C(Y)$ as

$$C(Y) = \{\theta \in \Theta : Y \in A(\theta)\}.$$

Then this is a confidence set for θ of confidence level $1 - \alpha$.

Proof. We have

$$P_{\theta_0}\{Y \in A(\theta_0)\} \geq 1 - \alpha \quad \text{for any} \quad \theta_0 \in \Theta.$$

Then

$$P_{\theta_0}\{\theta_0 \in C(Y)\} = P_{\theta_0}\{Y \in A(\theta_0)\} \geq 1 - \alpha \quad \text{for any} \quad \theta_0 \in \Theta,$$

that is, $C(Y)$ is an $1 - \alpha$ level confidence set for θ. □

Example 2.2.1. Suppose an observation Y is taken from normal distribution with mean θ and variance 1. For null hypothesis $H_0 : \theta = \theta_0$, construct the following family of acceptance sets of size α in three cases: $\theta_0 > 0$, $\theta_0 = 0$ and $\theta_0 < 0$.

Case 1: $A(\theta_0)$ for $H_0 : \theta = \theta_0$ where $\theta_0 > 0$ is

$$A(\theta_0) = \{Y : Y - \theta_0 > -Z_\alpha\}.$$

Case 2: $A(\theta_0)$ for $H_0 : \theta = \theta_0$ where $\theta_0 = 0$ is

$$A(\theta_0) = \{Y : |Y| < Z_{\alpha/2}\}.$$

Case 3: $A(\theta_0)$ for $H_0 : \theta = \theta_0$ where $\theta_0 < 0$ is

$$A(\theta_0) = \{Y : Y - \theta_0 < Z_\alpha\}.$$

Proof. The sizes of the test for each case of θ are
Case 1: $\theta_0 > 0$

$$\begin{aligned} P_{\theta_0}\{Y \in A(\theta_0)\} &= P_{\theta_0}\{Y - \theta_0 > -Z_\alpha\} \\ &= P\{Z > -Z_\alpha\} \\ &= 1 - \alpha, \end{aligned}$$

Case 2: $\theta_0 = 0$

$$\begin{aligned} P_{\theta_0}\{Y \in A(\theta_0)\} &= P_{\theta_0}\{|Y| < Z_{\alpha/2}\} \\ &= P_{\theta_0}\{-Z_{\alpha/2} < Y < Z_{\alpha/2}\} \\ &= P\{-Z_{\alpha/2} < Z < Z_{\alpha/2}\} \\ &= 1 - \alpha, \quad \text{and} \end{aligned}$$

Case 3: $\theta_0 < 0$

$$\begin{aligned} P_{\theta_0}\{Y \in A(\theta_0)\} &= P_{\theta_0}\{Y - \theta_0 < Z_\alpha\} \\ &= P\{Z < Z_\alpha\} \\ &= 1 - \alpha, \end{aligned}$$

where $Z \sim N(0,1)$. □

Using the family of acceptance sets above, we can construct a confidence set by inverting the acceptance sets as

$$\begin{aligned} C(Y) &= \{\theta_0 : Y \in A(\theta_0)\} \\ &= \{\theta_0 : Y - \theta_0 > -Z_\alpha \quad \text{and} \quad \theta_0 > 0\} \\ &\cup \{\theta_0 : |Y| < Z_{\alpha/2} \quad \text{and} \quad \theta_0 = 0\} \\ &\cup \{\theta_0 : Y - \theta_0 < Z_\alpha \quad \text{and} \quad \theta_0 < 0\}. \end{aligned}$$

From figure (2.1), the confidence set can be divided into 5 cases of Y: $Y \geq Z_{\alpha/2}$, $Z_\alpha \leq Y < Z_{\alpha/2}$, $-Z_\alpha < Y < Z_\alpha$, $-Z_{\alpha/2} < Y \leq -Z_\alpha$ and $Y \leq -Z_{\alpha/2}$.

$$\begin{aligned} &\text{If} \quad Y \geq Z_{\alpha/2}, & C(Y) &= \{\theta_0 : 0 < \theta_0 < Y + Z_\alpha\} = (0, Y + Z_\alpha). \\ &\text{If} \quad Z_\alpha \leq Y < Z_{\alpha/2}, & C(Y) &= \{\theta_0 : 0 \leq \theta_0 < Y + Z_\alpha\} = [0, Y + Z_\alpha). \\ &\text{If} \quad -Z_\alpha < Y < Z_\alpha, & C(Y) &= \{\theta_0 : Y - Z_\alpha < \theta_0 < Y + Z_\alpha\} = (Y - Z_\alpha, Y + Z_\alpha). \\ &\text{If} \quad -Z_{\alpha/2} < Y \leq -Z_\alpha, & C(Y) &= \{\theta_0 : Y - Z_\alpha < \theta_0 \leq 0\} = (Y - Z_\alpha, 0]. \\ &\text{If} \quad Y \leq -Z_{\alpha/2}, & C(Y) &= \{\theta_0 : Y - Z_\alpha < \theta_0 < 0\} = (Y - Z_\alpha, 0). \end{aligned}$$

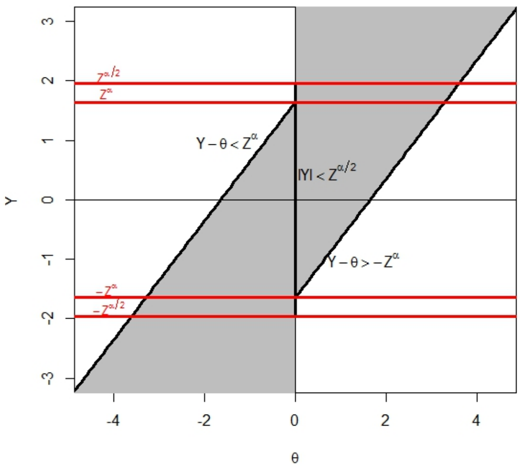

FIGURE 2.1: The relationship between confidence set and acceptance sets

We give a direct proof below to show that $C(Y)$ has a confidence level $1-\alpha$.

Proof. The confidence level of the confidence set is

$$\begin{aligned} P\{\theta \in C(Y)\} &= P\{0 < \theta < Y + Z_\alpha \ \text{ and } \ Y \geq Z_{\alpha/2}\} \\ &\quad + P\{0 \leq \theta < Y + Z_\alpha \ \text{ and } \ Z_\alpha \leq Y < Z_{\alpha/2}\} \\ &\quad + P\{Y - Z_\alpha < \theta < Y + Z_\alpha \ \text{ and } \ -Z_\alpha < Y < Z_\alpha\} \\ &\quad + P\{Y - Z_\alpha < \theta \leq 0 \ \text{ and } \ -Z_{\alpha/2} < Y \leq -Z_\alpha\} \\ &\quad + P\{Y - Z_\alpha < \theta < 0 \ \text{ and } \ Y \leq -Z_{\alpha/2}\}. \end{aligned}$$

If $\theta > 0$,

$$\begin{aligned} P\{\theta \in C(Y)\} &= P\{0 < \theta < Y + Z_\alpha \ \text{ and } \ Y \geq Z_{\alpha/2}\} \\ &\quad + P\{0 \leq \theta < Y + Z_\alpha \ \text{ and } \ Z_\alpha \leq Y < Z_{\alpha/2}\} \\ &\quad + P\{Y - Z_\alpha < \theta < Y + Z_\alpha \ \text{ and } \ -Z_\alpha < Y < Z_\alpha\} \\ &= P\{Y \geq Z_{\alpha/2}\} + P\{Z_\alpha \leq Y < Z_{\alpha/2}\} + P\{-Z_\alpha < Y < Z_\alpha\} \\ &= P\{Y > -Z_\alpha\} \\ &= P\{Z + \theta > -Z_\alpha\} \qquad (2.9) \\ &\geq P\{Z > -Z_\alpha\} \qquad (2.10) \\ &= 1 - \alpha, \end{aligned}$$

where $Z \sim N(0,1)$. The inequality of (2.10) holds because the probability in (2.9) is minimised when $\theta \to 0^+$.

If $\theta = 0$,

$$\begin{aligned}
P\{\theta \in C(Y)\} &= P\{0 \le \theta < Y + Z_\alpha \text{ and } Z_\alpha \le Y < Z_{\alpha/2}\} \\
&\quad + P\{Y - Z_\alpha < \theta < Y + Z_\alpha \text{ and } -Z_\alpha < Y < Z_\alpha\} \\
&\quad + P\{Y - Z_\alpha < \theta \le 0 \text{ and } -Z_{\alpha/2} < Y \le -Z_\alpha\} \\
&= P\{Z_\alpha \le Y < Z_{\alpha/2}\} + P\{-Z_\alpha < Y < Z_\alpha\} + P\{-Z_{\alpha/2} < Y \le -Z_\alpha\} \\
&= P\{-Z_{\alpha/2} < Y < Z_{\alpha/2}\} \\
&= P\{-Z_{\alpha/2} < Z + \theta < Z_{\alpha/2}\} \\
&= P\{-Z_{\alpha/2} < Z < Z_{\alpha/2}\} \\
&= 1 - \alpha.
\end{aligned}$$

If $\theta < 0$,

$$\begin{aligned}
P\{\theta \in C(Y)\} &= P\{Y - Z_\alpha < \theta < Y + Z_\alpha \text{ and } -Z_\alpha < Y < Z_\alpha\} \\
&\quad + P\{Y - Z_\alpha < \theta \le 0 \text{ and } -Z_{\alpha/2} < Y \le -Z_\alpha\} \\
&\quad + P\{Y - Z_\alpha < \theta < 0 \text{ and } Y \le -Z_{\alpha/2}\} \\
&= P\{-Z_\alpha < Y < Z_\alpha\} + P\{-Z_{\alpha/2} < Y \le -Z_\alpha\} + P\{Y \le -Z_{\alpha/2}\} \\
&= P\{Y < Z_\alpha\} \\
&= P\{Z + \theta < Z_\alpha\} \qquad (2.11) \\
&\ge P\{Z < Z_\alpha\} \qquad (2.12) \\
&= 1 - \alpha.
\end{aligned}$$

The inequality of (2.12) holds because the probability in (2.11) is minimised when $\theta \to 0^-$.
We have proved that $P\{\theta \in C(Y)\} \ge 1 - \alpha$. Therefore, $C(Y)$ is an $1 - \alpha$ level confidence set for θ. □

Example 2.2.2. Suppose observation Y distributes as some distribution with unknown parameter θ in parameter space Θ, $Y \sim f(y, \theta)$, $\theta \in \Theta$. For each $\theta_0 \in R$, we can construct an acceptance set for

$$\begin{aligned}
&H_0 : \theta = \theta_0 \\
\text{versus} \quad &H_a : \theta \ne \theta_0.
\end{aligned}$$

According to Uusipaikka (2008), a likelihood ratio test can be used to construct an acceptance set of size α as

$$\begin{aligned} A(\theta_0) &= \left\{ Y : \frac{f(y;\theta_0)}{\max_{\theta\in\Theta} f(y;\theta)} \geq c \right\} \\ &= \left\{ Y : \frac{L(\theta_0;y)}{\max_{\theta\in\Theta} f(y;\theta)} \geq c \right\} \\ &= \left\{ Y : \frac{L(\theta_0;y)}{L(\hat{\theta};y)} \geq c \right\}. \end{aligned} \tag{2.13}$$

The equality in equation (2.13) holds because the maximum likelihood estimator, $\hat{\theta}$, give the maximum of the likelihood function, $L(\theta, y)$, among other parameter values in parameter space.

We can construct an $1-\alpha$ level confidence set of θ using theorem 2.2.1 as

$$C(Y) = \left\{ \theta_0 : \frac{L(\theta_0;y)}{L(\hat{\theta};y)} \geq c \right\}.$$

We can calculate the value of c using the fact that $C(Y)$ is an $1-\alpha$ level confidence set as

$$\begin{aligned} P\{\theta \in C(Y)\} &= P\left\{ \frac{L(\theta_0;y)}{L(\hat{\theta};y)} \geq c \right\} \\ &= P\left\{ ln\frac{L(\theta_0;y)}{L(\hat{\theta};y)} \geq ln(c) \right\} \\ &= P\left\{ r(\theta_0;y) \geq ln(c) \right\} \\ &= 1-\alpha. \end{aligned}$$

We know that the approximate distribution of $-2r(\theta;y)$ is χ^2 distribution with a number of parameter(s), q, as a degrees of freedom. Therefore,

$$\begin{aligned} ln(c) &= -\frac{\chi^2_{q,\alpha}}{2} \\ c &= exp\left\{ -\frac{\chi^2_{q,\alpha}}{2} \right\}. \end{aligned}$$

Theorem 2.2.2. This theorem, partitioning principle (cf. Bretz et al., 2011), is the generalization of the theorem 2.2.1 by Lehmann (1986). Let θ denote an unknown parameter in parameter space Θ of data Y distribution and $\Theta = \bigcup_{i\in\wedge} \Theta_i$, where $\wedge$ is an index set. Let $A(\Theta_i)$ be an acceptance set of size α for testing $H_0 : \theta \in \Theta_i$, that is,

$$P_\theta\{Y \in A(\Theta_i)\} \geq 1-\alpha.$$

For each Y, we can construct a confidence set $C(Y)$ as

$$C(Y) = \bigcup_{i:Y\in A(\Theta_i)} \Theta_i.$$

Then this is a confidence set for θ of confidence level $1-\alpha$.

Proof. Given $\theta \in \Theta_{i_0}$ for some $\Theta_{i_0} \subset \Theta$,

$$\begin{aligned} P_\theta\{\theta \in C(Y)\} &= P_\theta\left\{\theta \in \bigcup_{i:Y\in A(\Theta_i)} \Theta_i\right\} \\ &\geq P_\theta\{Y \in A(\Theta_{i_0})\} \\ &\geq 1-\alpha, \end{aligned}$$

that is, $C(Y)$ is an $1-\alpha$ level confidence set for θ. □

Example 2.2.3. Suppose an observation Y is taken from normal distribution with mean θ and variance 1. We are interested in the confidence set for $|\theta|$. For each $|\theta| = \theta_0$, let $A(\Theta_{\theta_0})$ be an acceptance set of size α for testing $H_0 : \theta \in \Theta_{\theta_0}$, where $\Theta_{\theta_0} = \{\theta : |\theta| = \theta_0\}$ and $\theta_0 \geq 0$, of the form

$$A(\Theta_{\theta_0}) = \{Y : |Y| > c(\theta_0)\}.$$

From Theorem 2.2.2, a corresponding confidence set from the acceptance sets above is given by

$$\begin{aligned} C(Y) &= \bigcup_{i:Y\in A(\Theta_i)} \Theta_i \\ &= \{\theta_0 : Y \in A(\Theta_{\theta_0})\} \\ &= \{\theta_0 : |Y| > c(\theta_0)\} \\ &= \{\theta_0 : \theta_0 < c^{-1}(|Y|)\}. \end{aligned}$$

We can calculate the function $c(\cdot)$ and $c^{-1}(\cdot)$ using the fact that $P_{\theta_0}\{Y \in A(\Theta_{\theta_0})\} = 1-\alpha$,

$$\begin{aligned} P_{\theta_0}\{Y \in A(\Theta_{\theta_0})\} &= P_{\theta_0}\{|Y| > c(\theta_0)\} \\ &= P_{\theta_0}\{|Z+\theta| > c(\theta_0)\} \\ &= 1 - P_{\theta_0}\{|Z+\theta| \leq c(\theta_0)\} \\ &= 1 - P_{\theta_0}\{|Z+|\theta|| \leq c(\theta_0)\} \\ &= 1 - P_{\theta_0}\{|Z+\theta_0| \leq c(\theta_0)\} \\ &= 1 - P_{\theta_0}\{-c(\theta_0)-\theta_0 < Z < c(\theta_0)-\theta_0\} \\ &= 1 - [\Phi(c(\theta_0)-\theta_0) - \Phi(-c(\theta_0)-\theta_0)], \end{aligned}$$

where $Z \sim N(0,1)$. Hence, we have

$$\Phi(c(\theta_0) - \theta_0) - \Phi(-c(\theta_0) - \theta_0) = \alpha. \tag{2.14}$$

Using equation (2.14) and $\theta_0 = c^{-1}c(\theta_0)$, we can compute function $c^{-1}(\cdot)$ by

$$\Phi(w - c^{-1}(w)) - \Phi(-w - c^{-1}(w)) - \alpha = 0. \tag{2.15}$$

For any given value w, we can calculate $c^{-1}(w)$ using equation (2.15) as figure 2.2.

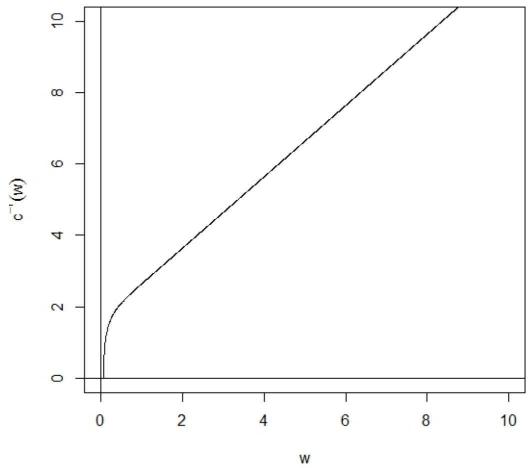

FIGURE 2.2: The relationship between given value w and $c^{-1}(w)$

Example 2.2.4. Let a simple linear regression model be $Y = \theta_0 + \theta_1 x + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$. $\boldsymbol{Y}$ is a vector of n observations from random variable Y, $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)^T$. X is a design matrix size $n \times 2$ given by $X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$. $\boldsymbol{\theta}$ is a vector of parameters θ_0 and θ_1, $\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$, and m_0 is a given number,

$$m_0 = \theta_0 + \theta_1 x_0.$$

For hypotheses

$$H_0 : x_0 = t_0 \quad \text{versus} \quad H_a : x_0 \neq t_0,$$

construct the following $A_1(t_0)$ as an acceptance set for each $t_0 \in R$,

$$A_1(t_0) = \left\{ \boldsymbol{Y} : \frac{\hat{\theta}_0 + \hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1, t_0)(X^T X)^{-1}(1, t_0)^T}} < c \right\},$$

where $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$ are parameter estimation of θ_0, θ_1 and σ. Since $A_1(t_0)$ is an $1-\alpha$ acceptance set and $\frac{\hat{\theta}_0+\hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1,t_0)(X^TX)^{-1}(1,t_0)^T}}$ has a t distribution, the value of c is $t_{n-2,\alpha}$. Then we can construct an $1-\alpha$ upper confidence set for x_0 by inverting the family of acceptance sets as

$$\begin{aligned} C_1(Y) &= \{t_0 : \boldsymbol{Y} \in A_1(t_0)\} \\ &= \left\{t_0 : \frac{\hat{\theta}_0+\hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1,t_0)(X^TX)^{-1}(1,t_0)^T}} < t_{n-2,\alpha}\right\} \\ &= \left\{t_0 : \hat{\theta}_0+\hat{\theta}_1 t_0 - t_{n-2,\alpha}\hat{\sigma}\sqrt{(1,t_0)(X^TX)^{-1}(1,t_0)^T} < m_0\right\}. \end{aligned}$$

Construct another acceptance set of size α, $A_2(t_0)$, for each $t_0 \in R$,

$$A_2(t_0) = \left\{\boldsymbol{Y} : \frac{\hat{\theta}_0+\hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1,t_0)(X^TX)^{-1}(1,t_0)^T}} > -c\right\}.$$

We can construct an $1-\alpha$ lower confidence set for x_0 by inverting the family of acceptance sets as

$$\begin{aligned} C_2(Y) &= \{t_0 : \boldsymbol{Y} \in A_2(t_0)\} \\ &= \left\{t_0 : \frac{\hat{\theta}_0+\hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1,t_0)(X^TX)^{-1}(1,t_0)^T}} > -t_{n-2,\alpha}\right\} \\ &= \left\{t_0 : \hat{\theta}_0+\hat{\theta}_1 t_0 + t_{n-2,\alpha}\hat{\sigma}\sqrt{(1,t_0)(X^TX)^{-1}(1,t_0)^T} > m_0\right\}. \end{aligned}$$

According to the two $1-\alpha$ level confidence sets above, we can construct a new $1-2\alpha$ level confidence set $C(Y) = \{x_0 \in C_1(Y) \cap C_2(Y)\}$. We can prove that the confidence level is $1-2\alpha$ by two different methods: the set property and the two-sided test.

Proof. Using the set property, we know that

$$P_{x_0}\{x_0 \in C_1(Y)\} \geq 1-\alpha$$

and

$$P_{x_0}\{x_0 \in C_2(Y)\} \geq 1-\alpha.$$

Then,

$$\begin{aligned} &P_{x_0}\{x_0 \in C_1(Y) \cap C_2(Y)\} \\ &= P_{x_0}\{x_0 \in C_1(Y)\} + P_{x_0}\{x_0 \in C_2(Y)\} - P_{x_0}\{x_0 \in C_1(Y) \cup C_2(Y)\} \\ &= P_{x_0}\{x_0 \in C_1(Y)\} + P_{x_0}\{x_0 \in C_2(Y)\} - 1 \qquad (2.16) \\ &\geq (1-\alpha) + (1-\alpha) - 1 \\ &= 1-2\alpha \end{aligned}$$

The equality of (2.16) holds because the upper bound of $C_1(Y)$ is always more than the lower bound of $C_2(Y)$ and so $C_1(Y) \cup C_2(Y) = R$ which implies $P_{x_0}\{x_0 \in C_1(Y) \cup C_2(Y)\} = P_{x_0}\{x_0 \in R\} = 1$. Then this confidence set has a confidence level $1 - 2\alpha$. □

Proof. Using a two-sided test, the $1 - 2\alpha$ acceptance set $A(\theta_0)$ is

$$A(\theta_0) = \left\{ \mathbf{Y} : \frac{\hat{\theta}_0 + \hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1, t_0)(X^T X)^{-1}(1, t_0)^T}} < t_{n-2,\alpha} \text{ and } \frac{\hat{\theta}_0 + \hat{\theta}_1 t_0 - m_0}{\hat{\sigma}\sqrt{(1, t_0)(X^T X)^{-1}(1, t_0)^T}} > -t_{n-2,\alpha} \right\}.$$

From Theorem 2.2.2, the confidence set $C(Y)$ can be produced as

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

Then this confidence set has a confidence level $1 - 2\alpha$. □

Example 2.2.5. Suppose an observation Y is taken from a normal distribution with mean θ and variance 1. Θ is a parameter space, which is divided into three subsets of parameter, $\Theta_1 = (-\infty, 0)$, $\Theta_2 = \{0\}$ and $\Theta_3 = (0, \infty)$, for three hypotheses.

Hypotheses 1:

$$H_{01} : \theta \in \Theta_1 \quad \text{versus} \quad H_{a1} : \theta \notin \Theta_1$$

Using the intersection-union principle, the hypotheses can be written as

$$H_{01} : \bigcup_{\lambda \in \Theta_1} H_{01\lambda} \qquad \text{versus} \qquad H_{a1} : \bigcap_{\lambda \in \Theta_1} H_{a1\lambda},$$

where

$$\begin{aligned} H_{01\lambda} &: \theta = \lambda, \ \lambda \in \Theta_1 \\ H_{a1\lambda} &: \theta \neq \lambda, \ \lambda \in \Theta_1. \end{aligned}$$

Construct the following $A(\lambda)$ as an acceptance set of size α for individual testing $H_{01\lambda}$ versus $H_{a1\lambda}$,

$$A(\lambda) = \{Y : Y - \lambda < Z_\alpha\}.$$

Then an acceptance set of size α, $A(\Theta_1)$, for H_{01} versus H_{a1} can be constructed as

$$\begin{aligned} A(\Theta_1) &= \bigcup_{\lambda \in \Theta_1} A(\lambda) \\ &= \{Y : Y - \max_{\Theta_1}\{\lambda\} < Z_\alpha\} \\ &= \{Y : Y - 0 < Z_\alpha\} \\ &= \{Y : Y < Z_\alpha\}. \end{aligned}$$

Proof. The size of the individual test is

$$\begin{aligned} P\{Y \in A(\lambda)\} &= P_\lambda\{Y - \lambda < Z_\alpha\} \\ &= P\{Z < Z_\alpha\} \\ &= 1 - \alpha. \end{aligned}$$

□

Below is a direct proof that the size of the overall test is α.

Proof. The size of the overall test is

$$\begin{aligned} P\{Y \in A(\Theta_1)\} &= P_{\theta_0}\{Y < Z_\alpha\}, \quad \text{for } \theta_0 \in \Theta_1 \\ &= P_{\theta_0}\{(Y - \theta_0) + \theta_0 < Z_\alpha\}, \quad \text{for } \theta_0 \in \Theta_1 \qquad (2.17) \\ &\geq P\{Z + 0 < Z_\alpha\} \qquad (2.18) \\ &= 1 - \alpha. \end{aligned}$$

The inequality of (2.18) holds because the probability in (2.17) is minimised when $\theta_0 \to 0^-$. Then $P\{Y \in A(\Theta_1)\} \geq 1 - \alpha$, and $A(\Theta_1)$ is an acceptance set of size α. □

Hypotheses 2:

$$H_{02} : \theta \in \Theta_2 \quad \text{versus} \quad H_{a2} : \theta \notin \Theta_2$$

Construct the following $A(\Theta_2)$ as an acceptance set of size α for testing H_{02} versus H_{a2},

$$A(\Theta_2) = \{Y : |Y| < Z_{\alpha/2}\}.$$

Proof. The size of the test is

$$\begin{aligned} P\{Y \in A(\Theta_2)\} &= P_{\theta_0}\{Y : |Y| < Z_{\alpha/2}\}, \quad \text{for } \theta_0 \in \Theta_2 \\ &= P_{\theta_0}\{Y : |Y + \theta_0| < Z_{\alpha/2}\}, \quad \text{for } \theta_0 \in \Theta_2 \\ &= P\{|Z| < Z_{\alpha/2}\} \\ &= 1 - \alpha. \end{aligned}$$

Hence, $A(\Theta_2)$ is an acceptance set of size α. □

Hypotheses 3:

$$H_{03} : \theta \in \Theta_3 \quad \text{versus} \quad H_{a3} : \theta \notin \Theta_3$$

Using the intersection-union principle, the hypotheses can be written as

$$H_{03} : \bigcup_{\lambda \in \Theta_3} H_{03\lambda} \qquad \text{versus} \qquad H_{a3} : \bigcap_{\lambda \in \Theta_3} H_{a3\lambda},$$

where

$$\begin{aligned} H_{03\lambda} &: \theta = \lambda,\ \lambda \in \Theta_3 \\ H_{a3\lambda} &: \theta \neq \lambda,\ \lambda \in \Theta_3. \end{aligned}$$

Construct the following $A(\lambda)$ as an acceptance set for testing $H_{03\lambda}$ versus $H_{a3\lambda}$,

$$A(\lambda) = \{Y : Y - \lambda > -Z_\alpha\}.$$

Then an acceptance set of size α, $A(\Theta_3)$, for H_{03} versus H_{a3} can be constructed as

$$\begin{aligned} A(\Theta_3) &= \bigcup_{\lambda\in\Theta_3} A(\lambda) \\ &= \{Y : Y - \min_{\Theta_3}\{\lambda\} > -Z_\alpha\} \\ &= \{Y : Y > -Z_\alpha\}. \end{aligned}$$

Proof. The size of the individual test is

$$\begin{aligned} P\{Y \in A(\lambda)\} &= P_\lambda\{Y - \lambda > -Z_\alpha\} \\ &= P\{Z > -Z_\alpha\} \\ &= 1 - \alpha. \end{aligned}$$

□

Below is a direct proof that the size of the overall test is α.

Proof. The size of the overall test is

$$\begin{aligned} P\{Y \in A(\Theta_3)\} &= P_{\theta_0}\{Y > -Z_\alpha\}, \quad \text{for } \theta_0 \in \Theta_3 \\ &= P_{\theta_0}\{(Y - \theta_0) + \theta_0 > -Z_\alpha\}, \quad \text{for } \theta_0 \in \Theta_3 \qquad (2.19) \\ &\geq P\{Z + 0 > -Z_\alpha\} \qquad (2.20) \\ &= 1 - \alpha. \end{aligned}$$

The inequality of (2.20) holds because the probability in (2.19) is minimised when $\theta_0 \to 0^+$. Then $P\{Y \in A(\Theta_3)\} \geq 1 - \alpha$, and $A(\Theta_3)$ is an acceptance set of size α. □

Using the family of acceptance sets above, we can construct a confidence set by inverting the acceptance sets as

$$C(Y) = \bigcup_{i:Y\in A(\Theta_i)} \Theta_i.$$

From figure (2.3), the confidence set can be divided into 5 cases for different sets of Y:

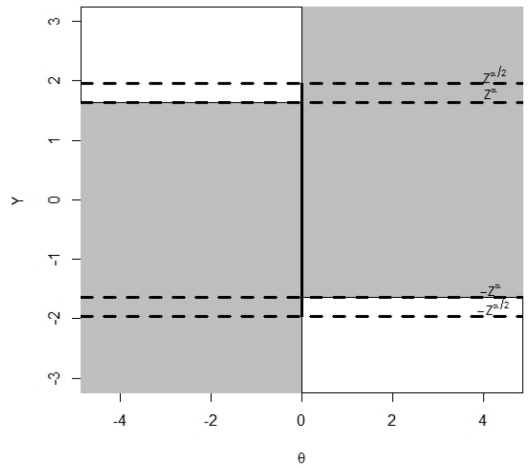

FIGURE 2.3: The relationship between confidence set and acceptance set

$Y \geq Z_{\alpha/2}$, $Z_\alpha \leq Y < Z_{\alpha/2}$, $-Z_\alpha < Y < Z_\alpha$, $-Z_{\alpha/2} < Y \leq -Z_\alpha$ and $Y \leq -Z_{\alpha/2}$.

$$
\begin{array}{llrcl}
\text{If} & Y \geq Z_{\alpha/2}, & C(Y) & = & \Theta_3 = (0, \infty). \\
\text{If} & Z_\alpha \leq Y < Z_{\alpha/2}, & C(Y) & = & \Theta_2 \cup \Theta_3 = [0, \infty). \\
\text{If} & -Z_\alpha < Y < Z_\alpha, & C(Y) & = & \Theta_1 \cup \Theta_2 \cup \Theta_3 = (-\infty, \infty). \\
\text{If} & -Z_{\alpha/2} < Y \leq -Z_\alpha, & C(Y) & = & \Theta_1 \cup \Theta_2 = (-\infty, 0]. \\
\text{If} & Y \leq -Z_{\alpha/2}, & C(Y) & = & \Theta_1 = (-\infty, 0).
\end{array}
$$

We give a direct proof below to show that $C(Y)$ has a confidence level $1 - \alpha$.

Proof. The confidence level of $C(Y)$ is

$$
\begin{aligned}
P\{\theta \in C(Y)\} = \; & P\{\theta \in (0,\ \infty) \text{ and } Y \geq Z_{\alpha/2}\} \\
& + P\{\theta \in [0,\ \infty) \text{ and } Z_\alpha \leq Y < Z_{\alpha/2}\} \\
& + P\{\theta \in (-\infty,\ \infty) \text{ and } -Z_\alpha < Y < Z_\alpha\} \\
& + P\{\theta \in (-\infty,\ 0] \text{ and } -Z_{\alpha/2} < Y \leq -Z_\alpha\} \\
& + P\{\theta \in (-\infty,\ 0) \text{ and } Y \leq -Z_{\alpha/2}\}.
\end{aligned}
$$

If $\theta > 0$,

$$\begin{aligned}
P\{\theta \in C(Y)\} &= P\{\theta \in (0,\ \infty) \ \text{ and } \ Y \geq Z_{\alpha/2}\} \\
&\quad + P\{\theta \in [0,\ \infty) \ \text{ and } \ Z_\alpha \leq Y < Z_{\alpha/2}\} \\
&\quad + P\{\theta \in (-\infty,\ \infty) \ \text{ and } \ -Z_\alpha < Y < Z_\alpha\} \\
&= P\{Y \geq Z_{\alpha/2}\} + P\{Z_\alpha \leq Y < Z_{\alpha/2}\} + P\{-Z_\alpha < Y < Z_\alpha\} \\
&= P\{Y > -Z_\alpha\} \\
&= P\{Z + \theta > -Z_\alpha\} \qquad (2.21) \\
&\geq P\{Z > -Z_\alpha\} \qquad (2.22) \\
&= 1 - \alpha.
\end{aligned}$$

The inequality of (2.22) holds because the probability in (2.21) is minimised when $\theta \to 0^+$.

If $\theta = 0$,

$$\begin{aligned}
P\{\theta \in C(Y)\} &= P\{\theta \in [0,\ \infty) \ \text{ and } \ Z_\alpha \leq Y < Z_{\alpha/2}\} \\
&\quad + P\{\theta \in (-\infty,\ \infty) \ \text{ and } \ -Z_\alpha < Y < Z_\alpha\} \\
&\quad + P\{(-\infty,\ 0] \ \text{ and } \ -Z_{\alpha/2} < Y \leq -Z_\alpha\} \\
&= P\{Z_\alpha \leq Y < Z_{\alpha/2}\} + P\{-Z_\alpha < Y < Z_\alpha\} + P\{-Z_{\alpha/2} < Y \leq -Z_\alpha\} \\
&= P\{-Z_{\alpha/2} < Y < Z_{\alpha/2}\} \\
&= P\{-Z_{\alpha/2} < Z + \theta < Z_{\alpha/2}\} \\
&= P\{-Z_{\alpha/2} < Z < Z_{\alpha/2}\} \\
&= 1 - \alpha.
\end{aligned}$$

If $\theta < 0$,

$$\begin{aligned}
P\{\theta \in C(Y)\} &= P\{\theta \in (-\infty,\ \infty) \ \text{ and } \ -Z_\alpha < Y < Z_\alpha\} \\
&\quad + P\{(-\infty,\ 0] \ \text{ and } \ -Z_{\alpha/2} < Y \leq -Z_\alpha\} \\
&\quad + P\{(-\infty,\ 0) \ \text{ and } \ Y \leq -Z_{\alpha/2}\} \\
&= P\{-Z_\alpha < Y < Z_\alpha\} + P\{-Z_{\alpha/2} < Y \leq -Z_\alpha\} + P\{Y \leq -Z_{\alpha/2}\} \\
&= P\{Y < Z_\alpha\} \\
&= P\{Z + \theta < Z_\alpha\} \qquad (2.23) \\
&\geq P\{Z < Z_\alpha\} \qquad (2.24) \\
&= 1 - \alpha.
\end{aligned}$$

The inequality of (2.24) holds because the probability in (2.23) is minimised when $\theta \to 0^-$.

We have proved that $P\{\theta \in C(Y)\} \geq 1-\alpha$. Therefore, $C(Y)$ is a confidence set for θ of confidence level $1-\alpha$. □

Example 2.2.6. This example is an generalization of the example 2.2.5. Suppose an observation Y is taken from normal distribution with mean θ and variance 1. Θ is a parameter space, which is divided into three subsets of parameter, $\Theta_1 = (-\infty, -\delta)$, $\Theta_2 = [-\delta, \delta]$ and $\Theta_3 = (\delta, \infty)$, for three hypotheses.
Hypotheses 1:

$$H_{01} : \theta \in \Theta_1 \quad \text{versus} \quad H_{a1} : \theta \notin \Theta_1$$

Using the intersection-union principle, the hypotheses can be written as

$$H_{01} : \bigcup_{\lambda \in \Theta_1} H_{01\lambda} \qquad \text{versus} \qquad H_{a1} : \bigcap_{\lambda \in \Theta_1} H_{a1\lambda},$$

where

$$\begin{aligned} H_{01\lambda} &: \theta = \lambda, \ \lambda \in \Theta_1 \\ H_{a1\lambda} &: \theta \neq \lambda, \ \lambda \in \Theta_1. \end{aligned}$$

Construct the following $A(\lambda)$ as an acceptance set of size α for individual testing $H_{01\lambda}$ versus $H_{a1\lambda}$,

$$A(\lambda) = \{Y : Y - \lambda < Z_\alpha\}.$$

Then an acceptance set of size α, $A(\Theta_1)$, for H_{01} versus H_{a1} can be constructed as

$$\begin{aligned} A(\Theta_1) &= \bigcup_{\lambda \in \Theta_1} A(\lambda) \\ &= \{Y : Y - \max_{\Theta_1}\{\lambda\} < Z_\alpha\} \\ &= \{Y : Y - (-\delta) < Z_\alpha\}. \end{aligned}$$

Proof. The size of the individual test is

$$\begin{aligned} P\{Y \in A(\lambda)\} &= P_\lambda\{Y - \lambda < Z_\alpha\} \\ &= P\{Z < Z_\alpha\} \\ &= 1-\alpha. \end{aligned}$$

□

Below is a direct proof that the size of the overall test is α.

Proof. The size of the overall test is

$$\begin{aligned} P\{Y \in A(\Theta_1)\} &= P_{\theta_0}\{Y + \delta < Z_\alpha\}, \quad \text{for } \theta_0 \in \Theta_1 \\ &= P_{\theta_0}\{(Y - \theta_0) + \theta_0 + \delta < Z_\alpha\}, \quad \text{for } \theta_0 \in \Theta_1 \\ &= P_{\theta_0}\{Z + (\theta_0 + \delta) < Z_\alpha\}, \quad \text{for } \theta_0 \in \Theta_1 \qquad (2.25) \\ &\geq P\{Z + 0 < Z_\alpha\} \qquad (2.26) \\ &= 1 - \alpha. \end{aligned}$$

The inequality of (2.26) holds because the probability in (2.25) is minimised when $\theta_0 \to -\delta^-$. Then $P\{Y \in A(\Theta_1)\} \geq 1 - \alpha$, and $A(\Theta_1)$ is an acceptance set of size α. □

Hypotheses 2:

$$H_{02} : \theta \in \Theta_2 \quad \text{versus} \quad H_{a2} : \theta \notin \Theta_2$$

Construct the following $A(\Theta_2)$ as an acceptance set of size α for testing H_{02} versus H_{a2},

$$A(\Theta_2) = \{Y : -\delta - a(\delta) < Y < \delta + a(\delta)\},$$

where $a(\delta)$ is a function of δ. We can calculate $a(\delta)$ from the fact that $P\{Y \in A(\Theta_2)\} = 1 - \alpha$,

$$\begin{aligned} P\{Y \in A(\Theta_2)\} &= P_{\theta_0}\{-\delta - a(\delta) < Y < \delta + a(\delta)\}, \quad \text{for } \theta_0 \in \Theta_2 \\ &= P_{\theta_0}\{-\delta - a(\delta) < Z + \theta_0 < \delta + a(\delta)\} \qquad (2.27) \\ &\geq P\{-\delta - a(\delta) < Z - \delta < \delta + a(\delta)\} \qquad (2.28) \\ &= P\{-a(\delta) < Z < 2\delta + a(\delta)\} \\ &= \Phi(2\delta + a(\delta)) - \Phi(-a(\delta)). \qquad (2.29) \end{aligned}$$

The inequality in (2.28) holds because the probability in (2.27) is minimised when θ_0 is $-\delta$ or δ. In this case, we give $\theta_0 = -\delta$.
For any given δ, we can calculate $a(\delta)$ using equation (2.29) equals to $1 - \alpha$ as in the figure(2.4). Then we have $A(\Theta_2)$ which is an acceptance set of size α.
Hypotheses 3:

$$H_{03} : \theta \in \Theta_3 \quad \text{versus} \quad H_{a3} : \theta \notin \Theta_3$$

Using the intersection-union principle, the hypotheses can be written as

$$H_{03} : \bigcup_{\lambda \in \Theta_3} H_{03\lambda} \qquad \text{versus} \qquad H_{a3} : \bigcap_{\lambda \in \Theta_3} H_{a3\lambda},$$

where

$$\begin{aligned} H_{03\lambda} &: \theta = \lambda, \; \lambda \in \Theta_3 \\ H_{a3\lambda} &: \theta \neq \lambda, \; \lambda \in \Theta_3. \end{aligned}$$

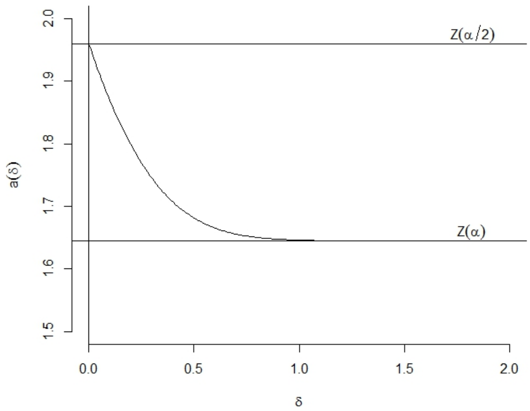

FIGURE 2.4: The relationship between given δ and $a(\delta)$

Construct the following $A(\lambda)$ as an acceptance set of size α for testing $H_{03\lambda}$ versus $H_{a3\lambda}$,

$$A(\lambda) = \{Y : Y - \lambda > -Z_\alpha\}.$$

Then an acceptance set of size α, $A(\Theta_3)$, for H_{03} versus H_{a3} is

$$\begin{aligned} A(\Theta_3) &= \bigcup_{\lambda \in \Theta_3} A(\lambda) \\ &= \{Y : Y - \min_{\Theta_3}\{\lambda\} > -Z_\alpha\} \\ &= \{Y : Y - \delta > -Z_\alpha\}. \end{aligned}$$

Proof. The size of the individual test is

$$\begin{aligned} P\{Y \in A(\lambda)\} &= P_\lambda\{Y - \lambda > -Z_\alpha\} \\ &= P\{Z > -Z_\alpha\} \\ &= 1 - \alpha. \end{aligned}$$

□

Below is a direct proof that the size of the overall test is α.

Proof. The size of the overall test is

$$\begin{aligned} P\{Y \in A(\Theta_3)\} &= P_{\theta_0}\{Y - \delta > -Z_\alpha\}, \quad \text{for } \theta_0 \in \Theta_3 \\ &= P_{\theta_0}\{(Y - \theta_0) + \theta_0 - \delta > -Z_\alpha\}, \quad \text{for } \theta_0 \in \Theta_3 \qquad (2.30) \\ &\geq P\{Z + 0 > -Z_\alpha\} \qquad (2.31) \\ &= 1 - \alpha. \end{aligned}$$

The inequality of (2.31) holds because the probability in (2.30) is minimised when $\theta_0 \to \delta^+$. Then $A(\Theta_3)$ is an acceptance set of size α. □

Using the family of acceptance sets above, we can construct a confidence set by inverting the acceptance sets above as

$$C(Y) = \bigcup_{i:Y \in A(\Theta_i)} \Theta_i.$$

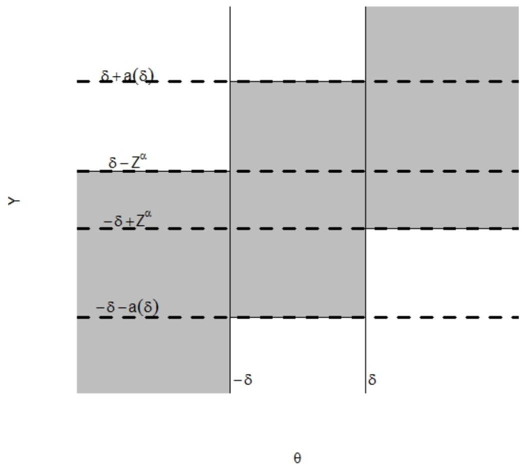

FIGURE 2.5: The relationship between confidence set and acceptance set

From figure (2.5),the confidence set can be divided into 5 cases for different sets of Y: $Y \geq \delta + a(\delta)$, $\delta - Z_\alpha < Y < \delta + a(\delta)$, $-\delta + Z_\alpha \leq Y \leq \delta - Z_\alpha$, $-\delta - a(\delta) < Y < -\delta + Z_\alpha$ and $Y \leq -\delta - a(\delta)$, where $\Phi(2\delta + a(\delta)) - \Phi(-a(\delta)) = 1 - \alpha$.

$$\begin{array}{llrcl}
\text{If} & Y \geq \delta + a(\delta), & C(Y) & = & \Theta_3 = (\delta, \infty). \\
\text{If} & \delta - Z_\alpha < Y < \delta + a(\delta), & C(Y) & = & \Theta_2 \cup \Theta_3 = [-\delta, \infty). \\
\text{If} & -\delta + Z_\alpha \leq Y \leq \delta - Z_\alpha, & C(Y) & = & \Theta_2 = (-\infty, \infty). \\
\text{If} & -\delta - a(\delta) < Y < -\delta + Z_\alpha, & C(Y) & = & \Theta_1 \cup \Theta_2 = (-\infty, \delta]. \\
\text{If} & Y \leq -\delta - a(\delta), & C(Y) & = & \Theta_1 = (-\infty, -\delta).
\end{array}$$

We give a direct proof below to show that $C(Y)$ has a confidence level $1 - \alpha$.

Proof. The confidence level of $C(Y)$ is

$$\begin{aligned}
P\{\theta \in C(Y)\} \quad = \quad & P\{\theta \in (\delta,\ \infty) \ \text{ and } \ Y \geq \delta + a(\delta)\} \\
& + P\{\theta \in [-\delta,\ \infty) \ \text{ and } \ \delta - Z_\alpha < Y < \delta + a(\delta)\} \\
& + P\{\theta \in [-\delta,\ \delta] \ \text{ and } \ -\delta + Z_\alpha \leq Y \leq \delta - Z_\alpha\} \\
& + P\{\theta \in (-\infty,\ \delta] \ \text{ and } \ -\delta - a(\delta) < Y < -\delta + Z_\alpha\} \\
& + P\{\theta \in (-\infty,\ -\delta) \ \text{ and } \ Y \leq -\delta - a(\delta)\}.
\end{aligned}$$

If $\theta > \delta$,

$$\begin{aligned} P\{\theta \in C(Y)\} &= P\{\theta \in (\delta,\ \infty) \text{ and } Y \geq \delta + a(\delta)\} \\ &\quad + P\{\theta \in [-\delta,\ \infty) \text{ and } \delta - Z_\alpha < Y < \delta + a(\delta)\} \\ &= P\{Y \geq \delta + a(\delta)\} + P\{\delta - Z_\alpha < Y < \delta + a(\delta)\} \\ &= P\{Y > \delta - Z_\alpha\} \\ &= P\{Y \in A(\Theta_3)\} \\ &\geq 1 - \alpha. \end{aligned}$$

If $\theta \in [-\delta,\ \delta]$,

$$\begin{aligned} P\{\theta \in C(Y)\} &= P\{\theta \in [-\delta,\ \infty) \text{ and } \delta - Z_\alpha < Y < \delta + a(\delta)\} \\ &\quad + P\{\theta \in [-\delta,\ \delta] \text{ and } -\delta + Z_\alpha \leq Y \leq \delta - Z_\alpha\} \\ &\quad + P\{\theta \in (-\infty,\ \delta] \text{ and } -\delta - a(\delta) < Y < -\delta + Z_\alpha\} \\ &= P\{\delta - Z_\alpha < Y < \delta + a(\delta)\} + P\{-\delta + Z_\alpha \leq Y \leq \delta - Z_\alpha\} \\ &\quad + P\{-\delta - a(\delta) < Y < -\delta + Z_\alpha\} \\ &= P\{-\delta - a(\delta) < Y < \delta + a(\delta)\} \\ &= P\{Y \in A(\Theta_2)\} \\ &\geq 1 - \alpha. \end{aligned}$$

If $\theta < -\delta$,

$$\begin{aligned} P\{\theta \in C(Y)\} &= P\{\theta \in (-\infty,\ \delta] \text{ and } -\delta - a(\delta) < Y < -\delta + Z_\alpha\} \\ &\quad + P\{\theta \in (-\infty,\ -\delta) \text{ and } Y \leq -\delta - a(\delta)\} \\ &= P\{-\delta - a(\delta) < Y < -\delta + Z_\alpha\} + P\{Y \leq -\delta - a(\delta)\} \\ &= P\{Y < -\delta + Z_\alpha\} \\ &= P\{Y \in A(\Theta_1)\} \\ &\geq 1 - \alpha. \end{aligned}$$

We have proved that $P\{\theta \in C(Y)\} \geq 1 - \alpha$. Therefore, $C(Y)$ is a confidence set for θ of confidence level $1 - \alpha$. □

Example 2.2.7. Suppose an observation Y is taken from normal distribution with mean θ and variance 1. According to Berger and Hsu (1996), they suggested an $1 - \alpha$ level confidence interval $[(Y-c)^-, (Y+c)^+]$, where $(Y-c)^-$ is the minimum of 0 and $(Y-c)$, and $(Y+c)^+$ is the maximum of 0 and $(Y+c)$. We can calculate the value of c using the fact that $P\{\theta \in C(Y)\} = 1 - \alpha$.

$$
\begin{aligned}
&P\left\{\theta \in [(Y-c)^-, (Y+c)^+]\right\} \\
&= P\left\{\{\theta \in [(Y-c)^-, \infty)\} \cap \{\theta \in (-\infty, (Y+c)^+]\}\right\} \\
&= P\{\theta \in [(Y-c)^-, \infty)\} + P\{\theta \in (-\infty, (Y+c)^+]\} - P\left\{\{\theta \in [(Y-c)^-, \infty)\} \cup \{\theta \in (-\infty, (Y+c)^+ \right. \\
&= P\{\theta \in [Y-c, \infty) \text{ and } Y-c<0\} + P\{\theta \in [0,\infty) \text{ and } Y-c>0\} \\
&\quad +P\{\theta \in (-\infty, Y+c] \text{ and } Y+c>0\} + P\{\theta \in (-\infty, 0] \text{ and } Y+c<0\} - 1 \qquad \text{(2} \\
&= P\{Y-c \le \theta \text{ and } Y-c<0\} + P\{\theta \ge 0 \text{ and } Y-c>0\} \\
&\quad +P\{Y+c \ge \theta \text{ and } Y+c>0\} + P\{\theta \le 0 \text{ and } Y+c<0\} - 1 \\
&= P\{Y \le \theta + c \text{ and } Y<c\} + P\{\theta \ge 0 \text{ and } Y>c\} \\
&\quad +P\{Y \ge \theta - c \text{ and } Y>-c\} + P\{\theta \le 0 \text{ and } Y<-c\} - 1 \qquad \text{(2}
\end{aligned}
$$

The equality in equation (2.32), $P\left\{\{\theta \in [(Y-c)^-, \infty)\} \cup \{\theta \in (-\infty, (Y+c)^+]\}\right\} = 1$, holds because $(Y-c)^-$ is always less than or equal to $(Y-c)^+$.

Let (2.33) equal to $K(\theta)$. $K(\theta)$ can be divided into three cases: $\theta < 0$, $\theta = 0$ and $\theta > 0$.

If $\theta < 0$,

$$
\begin{aligned}
K(\theta) &= P\{Y \le \theta + c\} + P\{Y > -c\} + P\{Y < -c\} - 1 \\
&= P\{Y - \theta \le c\} \\
&= P\{Z \le c\} \\
&= \Phi(c). \\
\text{Then } K(\theta) &\ge 1 - \alpha \text{ gives } c \ge Z_\alpha.
\end{aligned}
$$

If $\theta = 0$,

$$
\begin{aligned}
K(\theta) &= P\{Y < c\} + P\{Y > c\} + P\{Y > -c\} + P\{Y < -c\} - 1 \\
&= 1 + 1 - 1 \\
&= 1.
\end{aligned}
$$

If $\theta > 0$,

$$\begin{aligned} K(\theta) &= P\{Y < c\} + P\{Y > c\} + P\{Y \geq \theta - c\} - 1 \\ &= P\{Y - \theta \geq -c\} \\ &= P\{Z \geq -c\} \\ &= 1 - \Phi(-c). \\ &= \Phi(c). \\ \text{Then} \quad K(\theta) &\geq 1 - \alpha \quad \text{gives} \quad c \geq Z_\alpha. \end{aligned}$$

Therefore, the c value for the $1-\alpha$ level confidence interval $[(Y-c)^-, (Y+c)^+]$ is more than or equal to Z_α.

2.2.1 Comparing the Confidence Sets

Suppose we are interested in an unknown parameter θ of observation Y, which is taken from a normal distribution with mean θ and variance 1. From the above examples, we can construct a confidence set, $C(Y)$, for θ in four different ways.

- The usual 2-sided confidence interval for θ

$$C(Y) = [Y - Z_{\alpha/2},\ Y + Z_{\alpha/2}]$$

- The confidence set from Example 2.2.3

$$C(Y) = [0,\ c^{-1}(|Y|)]$$

- The confidence set from Example 2.2.1

$$\begin{array}{llcccc} \text{If} & Y \geq Z_{\alpha/2}, & C(Y) &=& \{\theta_0 : 0 < \theta_0 < Y + Z_\alpha\} &=& (0, Y + Z_\alpha). \\ \text{If} & Z_\alpha \leq Y < Z_{\alpha/2}, & C(Y) &=& \{\theta_0 : 0 \leq \theta_0 < Y + Z_\alpha\} &=& [0, Y + Z_\alpha). \\ \text{If} & -Z_\alpha < Y < Z_\alpha, & C(Y) &=& \{\theta_0 : Y - Z_\alpha < \theta_0 < Y + Z_\alpha\} &=& (Y - Z_\alpha, Y + Z_\alpha). \\ \text{If} & -Z_{\alpha/2} < Y \leq -Z_\alpha, & C(Y) &=& \{\theta_0 : Y - Z_\alpha < \theta_0 \leq 0\} &=& (Y - Z_\alpha, 0]. \\ \text{If} & Y \leq -Z_{\alpha/2}, & C(Y) &=& \{\theta_0 : Y - Z_\alpha < \theta_0 < 0\} &=& (Y - Z_\alpha, 0). \end{array}$$

- The confidence interval from Berger and Hsu (1996) from Example 2.2.7

$$C(Y) = [(y-c)^-, (y+c)^+]$$

From the figure 2.6, the $C(Y)$ from Example 2.2.1 and Example 2.2.7 are almost the same except the $C(Y)$ from Example 2.2.7 always includes zero in the confidence set.

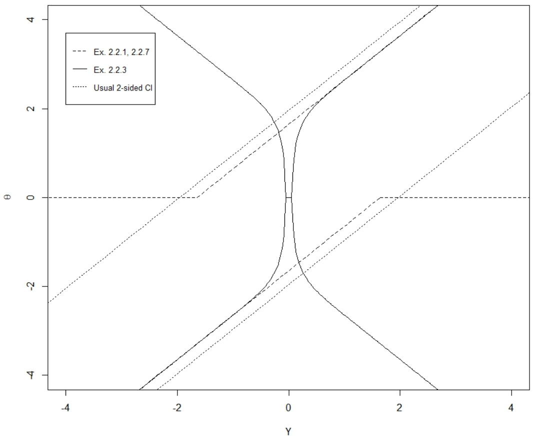

FIGURE 2.6: Comparing the confidence intervals from four methods

It is clear that, when Y is small, the $C(Y) = [0,\ c^{-1}(|Y|)]$ from Example 2.2.3 is the smallest confidence set, then the $C(Y)$ from Example 2.2.1, Example 2.2.7, and the usual 2-sided confidence interval is the largest confidence set respectively. When Y is a little bigger, the confidence intervals from Example 2.2.1, Example 2.2.7 and the usual 2-sided confidence interval are shift to positive θ. When Y is much more bigger, the usual 2-sided confidence interval is the smallest confidence set. However, if we are interested in the size of theta, $|\theta|$, the confidence set $C(Y) = [0,\ c^{-1}(|Y|)]$ from Example 2.2.3 is the smallest confidence set for $|\theta|$.

2.3 Construction of an Acceptance Set from a Confidence Set

From the previous section we can construct a confidence set by inverting the family of acceptance sets, in this section we are going to construct an acceptance set from a confidence set. Berger and Hsu (1996) constructed a test for hypotheses $H_0 : \theta \in \Theta_0$ versus $H_a : \theta \notin \Theta_0$ using an $100(1-\alpha)\%$ confidence set for θ, $C(Y)$, as the following theorem.

Theorem 2.3.1. Let Y have distribution function $f(y;\theta)$ with an unknown parameter θ in parameter space Θ. Given $C(Y)$ is an $1-\alpha$ level confidence set for θ and $\Theta_0 \subset \Theta$, then the acceptance set defined as

$$A(\Theta_0) = \{Y : \Theta_0 \cap C(Y) \neq \phi\}$$

is of size α for testing $H_0 : \theta \in \Theta_0$, where ϕ is an empty set.

Proof. Given $\theta_0 \in \Theta_0$, then

$$\begin{aligned} P_{\theta_0}\{Y \in A(\Theta_0)\} &= P_{\theta_0}\{\Theta_0 \cap C(Y) \neq \phi\} \\ &= 1 - P_{\theta_0}\{\Theta_0 \cap C(Y) = \phi\} \\ &\geq 1 - P_{\theta_0}\{\theta_0 \notin C(Y)\} \qquad (2.34) \\ &= P_{\theta_0}\{\theta_0 \in C(Y)\} \\ &\geq 1 - \alpha. \end{aligned}$$

The inequality in equation (2.34) holds because $\Theta_0 \cap C(Y) = \phi$ can imply that $\theta \notin C(Y)$. In contrast, $\theta \notin C(Y)$ cannot imply that $\Theta_0 \cap C(Y) = \phi$. Consequently, $P_\theta\{\Theta_0 \cap C(Y) = \phi\}$ is less than or equal to $P_\theta\{\theta \notin C(Y)\}$. Hence, $A(\Theta_0)$ is an acceptance set of size α for testing $H_0 : \theta \in \Theta_0$.

□

Example 2.3.1. Suppose an observation Y is taken from normal distribution with mean θ and variance 1. We are interested to test

$$H_0 : |\theta| \geq \delta \qquad \text{versus} \qquad H_0 : |\theta| < \delta.$$

Using the intersection-union principle, the hypotheses can be written as

$$H_{0\delta} : \theta \in \bigcup_{i=1}^{2} \Theta_{i\delta} \qquad \text{versus} \qquad H_{a\delta} : \theta \in \bigcup_{i=1}^{2} \Theta_{i\delta}^c,$$

where

$$\begin{aligned} \Theta &= R \\ \Theta_\delta &= \{\theta : |\theta| \geq \delta\}, \quad \delta \geq 0 \\ \Theta_{1\delta} &= \{\theta : \theta \geq \delta\}, \quad \delta \geq 0 \\ \Theta_{1\delta}^c &= \{\theta : \theta < \delta\}, \quad \delta \geq 0 \\ \Theta_{2\delta} &= \{\theta : \theta \leq -\delta\}, \quad \delta \geq 0 \\ \Theta_{2\delta}^c &= \{\theta : \theta > -\delta\}, \quad \delta \geq 0. \end{aligned}$$

The individual hypotheses 1 are

$$H_{01\delta} : \theta \in \Theta_{1\delta} \qquad \text{versus} \qquad H_{a1\delta} : \theta \in \Theta_{1\delta}^c.$$

We can construct a size α rejection region for $H_{01\delta}$ as

$$R_{1\delta} = \{Y : Y - \delta < -Z_\alpha\}.$$

The individual hypotheses 2 are

$$H_{02\delta} : \theta \in \Theta_{2\delta} \qquad \text{versus} \qquad H_{a2\delta} : \theta \in \Theta_{2\delta}^c.$$

We can construct a size α rejection region for $H_{02\delta}$ as

$$R_{2\delta} = \{Y : Y + \delta > Z_\alpha\}.$$

Then we can construct a size α test for overall hypotheses as

$$R_\delta = \{Y : Y - \delta < -Z_\alpha \text{ and } Y + \delta > Z_\alpha\}.$$

Then the acceptance set of size α for testing $H_{0\delta}$ is

$$A(\Theta_\delta) = \{Y : Y \geq \delta - Z_\alpha \text{ or } Y \leq -\delta + Z_\alpha\}.$$

For each δ, we have $A(\Theta_\delta)$ as an acceptance set of size α for testing $H_{0\delta}$. Using Theorem 2.2.2, we can construct a confidence set for θ as

$$\begin{aligned} C^*(Y) &= \bigcup_{\delta : Y \in A(\Theta_\delta)} \Theta_\delta \\ &= \Theta_{\delta_m}; \quad \delta_m = \min\{\delta : Y \in A(\Theta_\delta)\} \\ &= (-\infty, -\delta_m] \cup [\delta_m, \infty); \quad \delta_m = \min\{\delta : Y \in A(\Theta_\delta)\}. \end{aligned}$$

Because in this example one Θ_δ is a subset of another $\Theta_{\delta'}$, the confidence set we have constructed would be the biggest set of Θ_δ. Then the δ_m is probably close to zero, and so the confidence set is $(-\infty, \infty)$.

From the confidence set above, we can construct a size α test for $H_0 : \theta \in \Theta_{\delta_0}$ as

$$\begin{aligned} A^*(\Theta_0) &= \{Y : \Theta_0 \cap C^*(Y) \neq \phi\} \\ &= (-\infty, \infty). \end{aligned}$$

From above confidence set, the acceptance set we have constructed is $(-\infty, \infty)$. According to this example, if the subsets of parameter are not mutually exclusive sets, an acceptance set we constructed may sometimes become $(-\infty, \infty)$. Then the corollary below can help us to avoid that situation.

Example 2.3.2. According to Liu (2010), consider the equivalence between two regression lines,

$$m_1(x, \boldsymbol{\beta_1}) = \beta_{01} + \beta_{11}x \quad \text{and} \quad m_2(x, \boldsymbol{\beta_2}) = \beta_{02} + \beta_{12}x,$$

for $x \in [a, b]$, with $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ being the parameter vectors respectively. We can construct a confidence set for the difference between the two regression lines using the following theorem.

Theorem 2.3.2. Let $U_p(Y_1, Y_2, x)$ be a pointwise upper confidence limit of $m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})$ with $1-\alpha$ confidence level for each $x \in [a,b]$, that is,

$$P\left\{m_1\left(x, \hat{\boldsymbol{\beta_1}}\right) - m_2\left(x, \hat{\boldsymbol{\beta_2}}\right) \le U_p(Y_1, Y_2, x)\right\} \ge 1-\alpha \qquad \forall x \in [a,b].$$

Then

$$P\left\{\max_{x\in[a,b]}\{m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})\} \le \max_{x\in[a,b]} U_p(Y_1, Y_2, x)\right\} \ge 1-\alpha.$$

Proof. Let $x^* \in [a,b]$ and

$$m_1(x^*, \boldsymbol{\beta_1}) - m_2(x^*, \boldsymbol{\beta_2}) = \max_{x\in[a,b]}\{m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})\},$$

then

$$\begin{aligned}
& P\{m_1(x^*, \boldsymbol{\beta_1}) - m_2(x^*, \boldsymbol{\beta_2}) \le U_p(Y_1, Y_2, x^*)\} \\
= \; & P\left\{\max_{x\in[a,b]}\{m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})\} \le U_p(Y_1, Y_2, x^*)\right\} \\
\le \; & P\left\{\max_{x\in[a,b]}\{m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})\} \le \max_{x\in[a,b]} U_p(Y_1, Y_2, x)\right\} \\
= \; & P\left\{m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2}) \le \max_{x\in[a,b]} U_p(Y_1, Y_2, x) \quad \forall x \in [a,b]\right\} \\
\ge \; & 1-\alpha.
\end{aligned}$$

□

We have

$$P\left\{\max_{x\in[a,b]}\{m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})\} \le \max_{x\in[a,b]} U_p(Y_1, Y_2, x)\right\} \ge 1-\alpha$$

and

$$\begin{aligned}
& P\left\{\max_{x\in[a,b]}\{m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})\} \ge \min_{x\in[a,b]} L_p(Y_1, Y_2, x)\right\} \\
= \; & P\left\{\max_{x\in[a,b]}\{m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})\} \le -\min_{x\in[a,b]} L_p(Y_1, Y_2, x)\right\} \ge 1-\alpha.
\end{aligned}$$

From theorem 2.3.2, we can construct a confidence set as

$$\begin{aligned}
& P\left\{\max_{x\in[a,b]}\{(m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})), -(m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2}))\}\right. \\
& \qquad \left.\le \max\left(\max_{x\in[a,b]\}} U_p(Y_1, Y_2, x), -\min_{x\in[a,b]} L_p(Y_1, Y_2, x)\right)\right\} \\
= \; & P\left\{\max_{x\in[a,b]}|m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})| \le \max\left(\max_{x\in[a,b]} U_p(Y_1, Y_2, x), -\min_{x\in[a,b]} L_p(Y_1, Y_2, x)\right)\right\} \\
\ge \; & 1-\alpha.
\end{aligned}$$

From Theorem 2.3.1, we can construct a size α test for hypotheses

$$H_0 : \max_{x \in [a,b]} |m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})| \geq \delta$$

versus

$$H_a : \max_{x \in [a,b]} |m_1(x, \boldsymbol{\beta_1}) - m_2(x, \boldsymbol{\beta_2})| < \delta$$

as

$$R = \left\{ x : \ \max \left(\max_{x \in [a,b]} U_p(Y_1, Y_2, x), \quad - \min_{x \in [a,b]} L_p(Y_1, Y_2, x) \right) < \delta \right\}.$$

Corollary 2.3.1. From Theorem 2.2.2, we can construct an $1 - \alpha$ level confidence set for parameter θ of data Y distribution by inverting a family of acceptance sets of size α for testing $H_0 : \theta \in \Theta_0$. If we put more assumption $\Theta_i \cap \Theta_j = \phi$, where $i \neq j$, we get

$$C^*(Y) = \bigcup_{i : Y \in A(\Theta_i)} \Theta_i.$$

From Theorem 2.3.1, we can construct an acceptance set of size α for testing $H_0 : \theta \in \Theta_{i_0}$ using above confidence set,

$$\begin{aligned} A^*(\Theta_{i_0}) &= \{Y : \Theta_{i_0} \cap C^*(Y) \neq \phi\} \\ &= \{Y : Y \in A(\Theta_{i_0})\} \\ &= A(\Theta_{i_0}). \end{aligned}$$

Then the acceptance set we have constructed, $A^*(\Theta_i)$, is the same as $A(\Theta_i)$,the acceptance set for construct above confidence set, $C^*(Y)$.
It can be concluded that if the confidence set comes from Corollary 2.3.1, an acceptance set we get will be the same as the one given for constructing the confidence set.

Example 2.3.3. From example 2.2.3 we have a confidence sets for θ at $1-\alpha$ confidence level,

$$C(Y) = \left\{ \theta_0 : \theta_0 < c^{1-}(|Y|) \right\},$$

which reach an extra assumption of Corollary 2.3.1, $\Theta_i \cap \Theta_j = \phi$, where $i \neq j$. According to Theorem 2.3.1, we can construct a size α test for

$$H_0 : |\theta| \geq \delta \quad \text{versus} \quad H_a : |\theta| < \delta,$$

where δ is a given positive constant.The acceptance set for $H_0 : \theta \in \Theta_0$, where $\Theta_0 = (-\infty, -\delta] \cup [\delta, \infty)$, is given by

$$\begin{aligned} A(\Theta_0) &= \{Y : \Theta_0 \cap C(Y) \neq \phi\} \\ &= \{Y : \{(-\infty, -\delta] \cup [\delta, \infty)\} \cap (-c^{-1}(|Y|), c^{-1}(|Y|)) \neq \phi\} \\ &= \{Y : \delta \in [0, c^{-1}(|Y|))\} \\ &= \{Y : \delta < c^{-1}(|Y|))\} \\ &= \{Y : |Y| > c(\delta)\}. \end{aligned}$$

An acceptance set we have constructed here is the same as the one given for constructing the confidence set in example 2.2.3.

We give a direct proof below to show that the size of the test is α.

Proof. The size of the test is

$$\begin{aligned} P\{c^{-1}(|Y|) < \delta|\theta \geq \delta\} &= P\{c^{-1}(|Y|) - \theta + \theta - \delta < 0|\theta \geq \delta\} && (2.35) \\ &\leq P\{c^{-1}(|Y|) - \theta < 0|\theta \geq \delta\} && (2.36) \\ &= P\{c^{-1}(|Y|) < \theta|\theta \geq \delta\} \\ &= 1 - P\{c^{-1}(|Y|) > \theta|\theta \geq \delta\} \\ &\leq 1 - (1 - \alpha) && (2.37) \\ &= \alpha. \end{aligned}$$

The inequality of (2.36) holds because the probability in (2.35) is maximised when $\theta = \delta$. The inequality of (2.37) holds because $\{\theta_0 : c^{-1}(|Y|) > \theta_0\}$ is an $1 - \alpha$ confidence set. Hence, the size of the test is α. □

Example 2.3.4. Using the confidence interval from example 2.2.7, we can construct an acceptance set for hypotheses, $H_0 : \theta \in \Theta_0$ versus $H_0 : \theta \notin \Theta_0$ by Theorem 2.3.1. We have

$$C(Y) = [(Y - c)^-, (Y + c)^+].$$

$$\begin{array}{lrrcl} \text{If} & Y < -c, & C(Y) & = & [(Y - c), 0]. \\ \text{If} & -c < Y < c, & C(Y) & = & [(Y - c), (Y + c)]. \\ \text{If} & Y > c, & C(Y) & = & [0, (Y + c)]. \end{array}$$

From Theorem 2.3.1, we can construct an acceptance set of size α, $A(\Theta_0)$, for hypotheses,

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_a : \theta \notin \Theta_0,$$

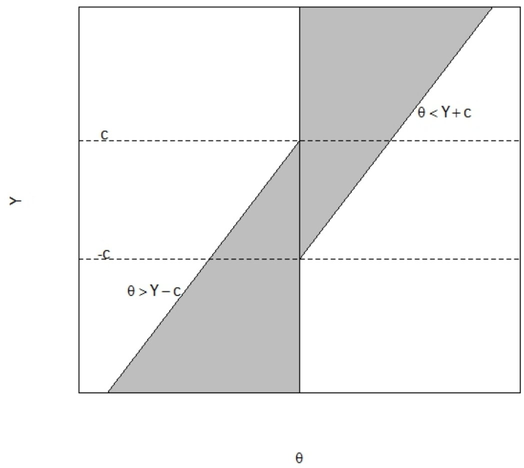

FIGURE 2.7: The relationship between confidence set and acceptance set

as

$$A(\Theta_0) = \{Y : \Theta_0 \cap C(Y) \neq \phi\}.$$

From figure 2.7, the acceptance set can be divided into three cases: $\Theta_0 = (-\infty, 0)$, $\Theta_0 = \{0\}$ and $\Theta_0 = (0, \infty)$.

$$\begin{aligned}
\text{If} \quad \Theta_0 = (-\infty, 0), \quad A(\Theta_0) &= \{Y : y \leq \theta_0 + c\}. \\
\text{If} \quad \Theta_0 = \{0\}, \quad A(\Theta_0) &= \{Y : y \in R\}. \\
\text{If} \quad \Theta_0 = (0, \infty), \quad A(\Theta_0) &= \{Y : y \geq \theta_0 - c\}.
\end{aligned}$$

We give a direct proof below to show that the size of the test is α.

Proof. The sizes of the test for each Θ_0 are

$$\begin{aligned}
\text{case 1: } \Theta_0 = (-\infty, 0), \qquad P_{\theta_0}\{Y \in A(\Theta_0)\} &= P_{\theta_0}\{Y \leq \theta_0 + c\} \\
&= P_{\theta_0}\{Y - \theta_0 \leq c\} \\
&\geq P\{Z \leq Z_\alpha\} \qquad (2.38) \\
&= 1 - \alpha. \\
\text{case 2: } \Theta_0 = \{0\}, \qquad P_{\theta_0}\{Y \in A(\Theta_0)\} &= P\{Y \in R\} = 1. \\
\text{case 3: } \Theta_0 = (0, \infty), \qquad P_{\theta_0}\{Y \in A(\Theta_0)\} &= P_{\theta_0}\{Y \geq \theta_0 - c\} \\
&= P_{\theta_0}\{Y - \theta_0 \geq -c\} \\
&\geq P\{Z \geq -Z_\alpha\} \qquad (2.39) \\
&= 1 - \alpha.
\end{aligned}$$

The inequality in equation (2.38) and equation (2.39) hold because we know that $c \geq Z_\alpha$,and so $P\{Z \leq c\} \geq P\{Z \leq Z_\alpha\}$. Therefore, $A(\Theta_0)$ is an acceptance set of size α for testing $H_0 : \theta \in \Theta_0$ versus $H_0 : \theta \notin \Theta_0$. □

2.4 Conclusion

In this chapter, we focus on three methods of statistical inference. The first method, intersection-union test allows us to construct a group of acceptance sets for further constructing a confidence set in the second method, the construction of a confidence set by inverting acceptance sets, as in example 2.2.5. Moreover, we also can construct an acceptance set from a confidence set in the last method. In the rest of this thesis, the construction of a confidence set will be used in different problems, the upper confidence bounds on the range of k means and on the maximum difference between two regression lines in Chapter 3, Chapter 4, and applied for classification problem in Chapter 5.

Chapter 3

Construction of the Upper Confidence Bounds on the Range of Means

The construction of a confidence set can be used in some equivalence problems in the previous Chapter. In this chapter, the upper confidence bounds on the range of k means are the next problem that we are considered. We propose the construction of upper confidence bounds by three methods: using the theorem from Liu et al. (2009), using the F-statistic and using the studentized range statistic. Firstly, the theorem from Liu et al. (2009) is used to construct the upper confidence bounds. Then the upper confidence bounds are constructed with two examples: $k = 2$ and $k \geq 3$. Secondly, the F-statistic is used to construct the upper confidence bounds. The mathematical proof is provided with three examples: $k = 2$, $k = 3$ and $k > 3$. The last method, the studentized range statistic is used to construct the upper confidence bounds. The mathematical proof is provided with three examples: $k = 2$, $k = 3$ and $k > 3$. The last section presents the comparison of the upper confidence bounds on several methods when σ is known. For any given σ, the F-statistic and studentized range statistic are changed in this situation. Then the Chi-squared statistic and the range statistic are the suitable adapted methods for this case. Moreover, the effects of sample sizes, unbalanced sample sizes and confidence levels on the upper confidence bounds are assessed.

3.1 Using the Theorem from Liu et al. (2009)

According to the theorem from Liu et al. (2009), given several $(1 - \alpha)$ upper confidence bounds for several parameters in parameter space. Then the upper confidence bounds on the maximum parameters can be computed by the theorem. Moreover, the data is not require to be normality or equal variance for this method. The requirement is only the upper confidence bound for each parameter θ_λ. Note that this method cannot compute the upper confidence bound of $\frac{|\mu_1-\mu_2|}{\sigma}$ if σ is unknown. However, it can be use for calculating the upper confidence bound of $|\mu_1-\mu_2|$ whether σ is known or unknown.

Theorem 3.1.1. Let $U_\lambda(X)$ is an $(1-\alpha)$ upper confidence bound for parameter θ_λ in parameter space Θ, that is

$$P\{\theta_\lambda \leq U_\lambda(X)\} \geq 1-\alpha, \quad \theta_\lambda \in \Theta.$$

Then,

$$P\left\{\max_{\lambda\in\wedge}\theta_\lambda \leq \max_{\lambda\in\wedge}U_\lambda(X)\right\} \geq 1-\alpha,$$

where $\wedge$ is an index set.

Proof. Let θ^*_λ is the maximum of $\theta_\lambda \in \Theta$. Then,

$$\begin{aligned} & P\left\{\max_{\lambda\in\wedge}\theta_\lambda \leq \max_{\lambda\in\wedge}U_\lambda(X)\right\} \\ = & P\left\{\theta^*_\lambda \leq \max_{\lambda\in\wedge}U_\lambda(X)\right\} \\ \geq & P\{\theta^*_\lambda \leq U^*_\lambda(X)\} \\ = & 1-\alpha. \end{aligned}$$

□

Example 3.1.1. Two Samples

Let two samples $(X_{11}, X_{12}, \ldots, X_{1n})$ and $(X_{21}, X_{22}, \ldots, X_{2n})$ be the observations from two treatments which are independent and identically distributed as normal distributions with a known equal variance σ^2 and means being μ_1 and μ_2 respectively. To construct the upper confidence bounds on $|\mu_1-\mu_2|/\sigma$, theorem 3.1.1 is used. We know that

$$|\mu_1-\mu_2| = \max\ \{(\mu_1-\mu_2), -(\mu_1-\mu_2)\}.$$

The upper confidence bounds and the lower confidence bounds of $\mu_1-\mu_2$ are

$$P\left\{\mu_1-\mu_2 \leq \bar{X}_1-\bar{X}_2+Z_\alpha\sqrt{\frac{2\sigma^2}{n}}\right\} \geq 1-\alpha,$$

and

$$P\left\{\mu_1-\mu_2 \geq \bar{X}_1-\bar{X}_2-Z_\alpha\sqrt{\frac{2\sigma^2}{n}}\right\} \geq 1-\alpha.$$

The lower confidence bounds can be written as

$$P\left\{-(\mu_1-\mu_2) \ \le \ -(\bar{X}_1-\bar{X}_2)+Z_\alpha\sqrt{\frac{2\sigma^2}{n}}\right\} \ge 1-\alpha.$$

Using theorem 3.1.1, we have

$$P\left\{|\mu_1-\mu_2| \ \le \ |\bar{X}_1-\bar{X}_2|+Z_\alpha\sqrt{\frac{2\sigma^2}{n}}\right\} \ge 1-\alpha.$$

Then,

$$P\left\{\frac{|\mu_1-\mu_2|}{\sigma} \ \le \ \frac{|\bar{X}_1-\bar{X}_2|}{\sigma}+Z_\alpha\sqrt{\frac{2}{n}}\right\} \ge 1-\alpha.$$

Hence, the upper confidence bounds on $|\mu_1-\mu_2|/\sigma$ is

$$\frac{|\bar{X}_1-\bar{X}_2|}{\sigma}+Z_\alpha\sqrt{\frac{2}{n}}.$$

Example 3.1.2. ***k*** **Samples**

Let k samples $(X_{11}, X_{12}, \ldots, X_{1n})$, $(X_{21}, X_{22}, \ldots, X_{2n})$, ..., $(X_{k1}, X_{k2}, \ldots, X_{kn})$ be the observations from k treatments which are independent and identically distributed as normal distributions with a known equal variance σ^2 and means being μ_1, μ_2,..., μ_k respectively. To construct the upper confidence bounds on $\max_{1\le i,j\le k}\frac{|\mu_i-\mu_j|}{\sigma}$, theorem 3.1.1 is used. From example 3.1.1 we know that

$$P\left\{\frac{|\mu_1-\mu_2|}{\sigma} \ \le \ \frac{|\bar{X}_1-\bar{X}_2|}{\sigma}+Z_\alpha\sqrt{\frac{2}{n}}\right\} \ge 1-\alpha.$$

Then

$$P\left\{\frac{|\mu_i-\mu_j|}{\sigma} \ \le \ \frac{|\bar{X}_i-\bar{X}_j|}{\sigma}+Z_\alpha\sqrt{\frac{2}{n}}\right\} \ge 1-\alpha, \quad i,j=1,2,...,k.$$

Using theorem 3.1.1, we have

$$P\left\{\max_{1\le i,j\le k}\frac{|\mu_i-\mu_j|}{\sigma} \ \le \ \max_{1\le i,j\le k}\frac{|\bar{X}_i-\bar{X}_j|}{\sigma}+Z_\alpha\sqrt{\frac{2}{n}}\right\} \ge 1-\alpha.$$

Hence, the upper confidence bounds on $\max_{1\le i,j\le k}|\mu_i-\mu_j|/\sigma$ is

$$\max_{1\le i,j\le k}\frac{|\bar{X}_i-\bar{X}_j|}{\sigma}+Z_\alpha\sqrt{\frac{2}{n}}.$$

3.2 Using the F Statistic

According to Scheffe's method for all-pairwise comparison (cf. Hsu, 1996), Let $X_{i1}, X_{i2}, ..., X_{in_i} \sim N(\mu_i, \sigma^2),\ i = 1, 2, ..., k$ and

$$F \;=\; \bar{n} \cdot \frac{(k-1)^{-1}\sum_{i=1}^{k}(n_i/\bar{n})(\bar{X}_i - \bar{X}_{..})^2/\sigma^2}{(N-k)^{-1}\sum_{i=1}^{k}\sum_{v=1}^{n_i}(X_{iv}-\bar{X}_i)^2/\sigma^2} \;\sim\; F_{k-1,N-k;\tau}, \tag{3.1}$$

with a non-central parameter, τ, where

$$\bar{n} = \frac{\sum_{i=1}^{k} n_i}{k} \quad \text{and} \quad N = \sum_{i=1}^{k} n_i.$$

Given $B(F)$ be an upper confidence bound for $\max_{1\le i,j\le k}\frac{|\mu_i-\mu_j|}{\sigma}$. Then

$$P\left\{\max_{1\le i,j\le k}\frac{|\mu_i-\mu_j|}{\sigma} \;\le\; B(F)\right\} \;=\; 1-\alpha.$$

Example 3.2.1. Two Samples
Let two samples $(X_{11}, X_{12}, \ldots, X_{1n})$ and $(X_{21}, X_{22}, \ldots, X_{2n})$ be the observations from two treatments which are independent and identically distributed as normal distributions with an equal variance σ^2 and means being μ_1 and μ_2 respectively. The F-statistic is used for constructing the upper confidence bounds on $|\mu_1 - \mu_2|/\sigma$.

Consider the numerator of F in equation (3.1) for $k = 2$, a non-central parameter, τ, is calculated,

$$\begin{aligned}
\frac{\bar{n}}{\sigma^2}\left\{\sum_{i=1}^{2}(n_i/\bar{n})(\bar{X}_i-\bar{X}_{..})^2\right\} \;&\sim\; \chi^2_{1;\frac{n}{\sigma^2}\{(\mu_1-\mu_{..})^2+(\mu_2-\mu_{..})^2\}}\\
&=\; \chi^2_{1;\frac{n}{\sigma^2}\left\{(\mu_1-\frac{\mu_1+\mu_2}{2})^2+(\mu_2-\frac{\mu_1+\mu_2}{2})^2\right\}}\\
&=\; \chi^2_{1;\frac{n}{\sigma^2}\left\{(\frac{\mu_1}{2}-\frac{\mu_2}{2})^2+(\frac{\mu_2}{2}-\frac{\mu_1}{2})^2\right\}}\\
&=\; \chi^2_{1;\frac{n(\mu_1-\mu_2)^2}{2\sigma^2}}\\
&=\; \chi^2_{1;\tau}.
\end{aligned}$$

Since $B(F)$ is an upper confidence bound for $\max_{1\le i,j\le k}\frac{|\mu_i-\mu_j|}{\sigma}$, then, for $k = 2$,

$$P\left\{\frac{|\mu_1-\mu_2|}{\sigma} \;\le\; B(F)\right\} \;=\; 1-\alpha$$

$$P\left\{B^{-1}\left(\frac{|\mu_1-\mu_2|}{\sigma}\right) \;\le\; F\right\} \;=\; 1-\alpha$$

$$1 - P\left\{B^{-1}\left(\frac{|\mu_1 - \mu_2|}{\sigma}\right) \;>\; F\right\} \;=\; 1 - \alpha.$$

Hence,

$$P\left\{F \;<\; B^{-1}\left(\frac{|\mu_1 - \mu_2|}{\sigma}\right)\right\} \;=\; \alpha$$

Let $\left(\frac{|\mu_1 - \mu_2|}{\sigma}\right)$ be δ, then the last equation becomes

$$P\left\{F \;<\; B^{-1}(\delta)\right\} \;=\; \alpha, \tag{3.2}$$

where $F \sim F_{1,N-2;\frac{n\delta^2}{2}}$. For any given δ, we know $F_{1,N-2;\frac{n\delta^2}{2}}$. Then we can compute $B^{-1}(\delta)$ which makes the probability in equation (3.2) equal to α. In figure 3.1, then function $B^{-1}(\delta)$ is plotted for $k = 2$, where $\alpha = 0.05$ and $n = 20$.

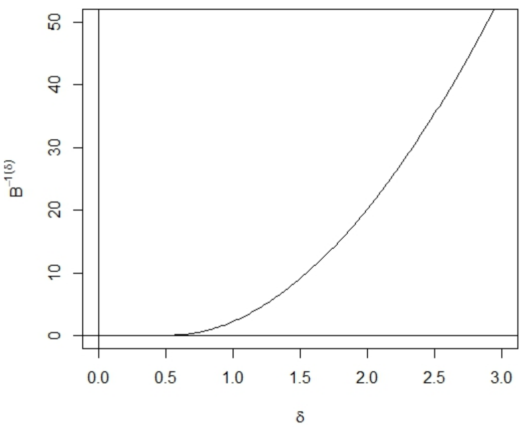

FIGURE 3.1: The relationship between given value δ and $B^{-1}(\delta)$ from the F statistic method for $k = 2$

Example 3.2.2. Three Samples

Let Three samples $(X_{11}, X_{12}, \ldots, X_{1n})$, $(X_{21}, X_{22}, \ldots, X_{2n})$ and $(X_{31}, X_{32}, \ldots, X_{3n})$ be the observations from three treatments which are independent and identically distributed as normal distributions with an equal variance σ^2 and means being μ_1, μ_2 and μ_3 respectively. The F statistic is used for constructing the upper confidence bounds on $\max_{1 \le i,j \le 3} \frac{|\mu_i - \mu_j|}{\sigma}$. That is

$$P\left\{\max_{1 \le i,j \le 3} \frac{|\mu_i - \mu_j|}{\sigma} \;\le\; B(F)\right\} \ge 1 - \alpha.$$

where $F \sim F_{2,N-3;\tau}$. Then we have

$$P\left\{F \;\le\; B^{-1}\left(\max_{1 \le i,j \le 3} \frac{|\mu_i - \mu_j|}{\sigma}\right)\right\} \le \alpha. \tag{3.3}$$

For maximised the probability in equation (3.3) to α, the τ^* is calculated by minimising

τ. Consider the numerator of F in equation (3.1) for $k = 3$, a non-central parameter, τ, is calculated,

$$\begin{aligned}\frac{\bar{n}}{\sigma^2}\left\{\sum_{i=1}^{3}(n_i/\bar{n})(\bar{X}_i - \bar{X}_{..})^2\right\} &\sim \chi^2_{2;\frac{n}{\sigma^2}\{(\mu_1-\mu_{..})^2+(\mu_2-\mu_{..})^2+(\mu_3-\mu_{..})^2\}} \\ &= \chi^2_{2;\frac{n}{\sigma^2}\left\{(\mu_1-\frac{\mu_1+\mu_2+\mu_3}{3})^2+(\mu_2-\frac{\mu_1+\mu_2+\mu_3}{3})^2+(\mu_3-\frac{\mu_1+\mu_2+\mu_3}{3})^2\right\}} \\ &= \chi^2_{2;\frac{n}{3\sigma^2}\{3(\mu_1^2+\mu_2^2+\mu_3^2)-(\mu_1+\mu_2+\mu_3)^2\}} \\ &= \chi^2_{2;\tau}.\end{aligned}$$

To calculate τ^*, τ is minimised by given $(\mu_1/\sigma, \mu_2/\sigma, \mu_3/\sigma) = (-\delta/2, 0, \delta/2)$ (cf. Bofinger et al., 1993). Then we get

$$\frac{n}{\sigma^2}\sum_{i=1}^{k}(n_i/\bar{n})(\bar{x}_i - \bar{x}_{..})^2 \sim \chi^2_{2;\frac{n\delta^2}{2}})$$

and

$$P\left\{F \;\leq\; B^{-1}(\delta)\right\} \;=\; \alpha, \qquad (3.4)$$

where $F \sim F_{2,N-3;\frac{n\delta^2}{2}}$. For any given δ, we have $F_{2,N-3;\frac{n\delta^2}{2}}$. Then we can compute $B^{-1}(\delta)$ which makes probability in equation (3.4) equal to α. In figure 3.2, then function $B^{-1}(\delta)$ is plotted for $k = 3$, where $\alpha = 0.05$ and $n = 20$.

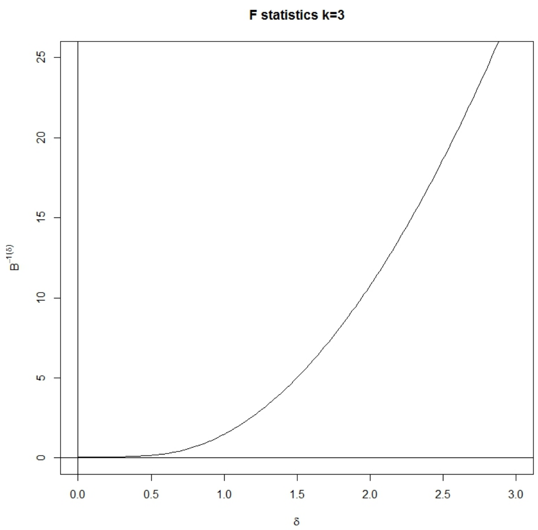

FIGURE 3.2: The relationship between given value δ and $B^{-1}(\delta)$ from the F statistic method for $k = 3$

Example 3.2.3. k Samples

Let k samples $(X_{11}, X_{12}, \ldots, X_{1n})$, $(X_{21}, X_{22}, \ldots, X_{2n})$, ..., $(X_{k1}, X_{k2}, \ldots, X_{kn})$ be the observations from k treatments which are independent and identically distributed as normal distributions with an equal variance σ^2 and means being μ_1, μ_2,..., μ_k respectively. The F statistic is used for constructing the upper confidence bounds on

$\max_{1\le i,j\le k} \frac{|\mu_i-\mu_j|}{\sigma}$. That is

$$P\left\{\max_{1\le i,j\le k} \frac{|\mu_i-\mu_j|}{\sigma} \ \le\ B(F)\right\} \ge 1-\alpha.$$

Then we have

$$P\left\{F \ \le\ B^{-1}\left(\max_{1\le i,j\le k} \frac{|\mu_i-\mu_j|}{\sigma}\right)\right\} \le \alpha. \tag{3.5}$$

For maximised the probability in equation (3.5) to α, the τ^* is calculated by minimising τ. To calculate τ^*, τ is minimised by given $(\mu_1/\sigma, \mu_2/\sigma, ..., \mu_k/\sigma) = (-\delta/2, 0, ..., 0, \delta/2)$ (cf. Bofinger et al., 1993). Consider the numerator of F in equation (3.1) for $k > 3$, a non-central parameter, τ^*, is calculated,

$$\begin{aligned}\frac{\bar{n}}{\sigma^2}\sum_{i=1}^{k}(n_i/\bar{n})(\bar{X}_i-\bar{X}_{..})^2 \ &\sim\ \chi^2_{k-1;(\frac{\sqrt{n}}{\sigma}\mu_1-\frac{\sqrt{n}}{\sigma}\mu_{..})^2+(\frac{\sqrt{n}}{\sigma}\mu_2-\frac{\sqrt{n}}{\sigma}\mu_{..})^2+...+(\frac{\sqrt{n}}{\sigma}\mu_k-\frac{\sqrt{n}}{\sigma}\mu_{..})^2}\\ &=\ \chi^2_{k-1;\frac{n\delta^2}{2}}\\ &=\ \chi^2_{k-1;\tau^*}.\end{aligned}$$

Then we get

$$\frac{\bar{n}}{\sigma^2}\sum_{i=1}^{k}(n_i/\bar{n})(\bar{x}_i-\bar{x}_{..})^2 \sim \chi^2_{k-1;\frac{n\delta^2}{2}}$$

and

$$P\left\{F \ \le\ B^{-1}(\delta)\right\} \le \alpha, \tag{3.6}$$

where $F \sim F_{k-1,N-k;\frac{n\delta^2}{2}}$. For any given δ, we have $F_{k-1,N-k;\frac{n\delta^2}{2}}$. Then we can compute $B^{-1}(\delta)$ which makes probability in equation (3.6) equal to α. In figure 3.3, then function $B^{-1}(\delta)$ is plotted for $k = 5$, where $\alpha = 0.05$ and $n = 20$.

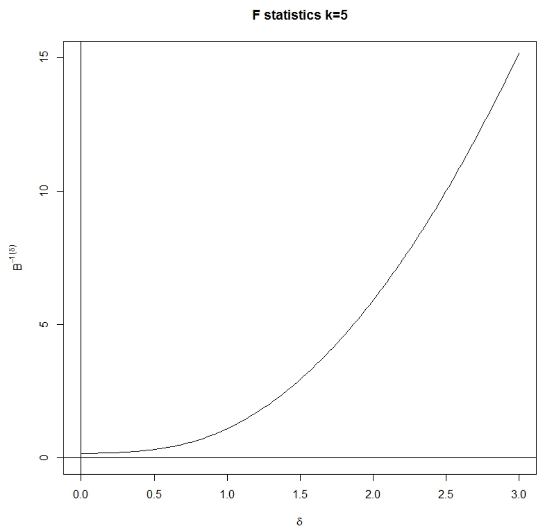

FIGURE 3.3: The relationship between given value δ and $B^{-1}(\delta)$ from the F statistic method for $k = 5$

3.3 Using the Studentized Range Statistic

This method is applied from the studentized range or Tukey's technique (Miller 1980, Hochberg and Tamhane 2011, Hsu 1996).

3.3.1 The Studentized Range or Tukey's Technique

Let $X_{i1}, X_{i2}, ..., X_{in} \sim N(\mu_i, \sigma^2), i = 1, 2, ..., k$. The studentized range technique provides a $1 - \alpha$ simultaneous confidence interval for all $\mu_i - \mu_j$, $1 \leq i, j \leq k$ which are given by

$$\mu_i - \mu_j \in \bar{X}_i - \bar{X}_j \pm q^{\alpha}_{k,k(n-1)} \frac{SD}{\sqrt{n}} \quad \text{for all } i \neq j, \tag{3.7}$$

where

$$SD = \sqrt{\frac{\sum_{i=1}^{k} \sum_{l=1}^{n} (X_{il} - \bar{X}_i)^2}{k(n-1)}}, \quad \bar{X}_i = \frac{\sum_{l=1}^{n} X_{il}}{n}$$

and $q^{\alpha}_{k,k(n-1)}$ is an upper α point of the studentized range distribution with k and $k(n-1)$ degrees of freedom. The null hypothesis we are interested in is

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k = \mu$$

or

$$H_0 : \mu_i - \mu_j = 0, \quad i \neq j.$$

H_0 will be rejected if and only if at least one interval of (3.7) not containing zero. This rejection can be written in the term of

$$\max_{1 \leq i,j \leq k} \frac{|\sqrt{n}(\bar{X}_i - \bar{X}_j)|}{SD} > q^{\alpha}_{k,k(n-1)}. \tag{3.8}$$

Proof. The size of the test is

$$\begin{aligned}
& P\left\{ \max_{1 \leq i,j \leq k} \frac{|\sqrt{n}(\bar{X}_i - \bar{X}_j)|}{SD} > q^{\alpha}_{k,k(n-1)} \middle| \mu_i = \mu_j = \mu \right\} \\
= \; & P\left\{ \max_{1 \leq i,j \leq k} \frac{\left| \frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma} - \frac{\sqrt{n}(\bar{X}_j - \mu)}{\sigma} \right|}{\sqrt{SD^2/\sigma^2}} > q^{\alpha}_{k,k(n-1)} \middle| \mu_i = \mu_j = \mu \right\} \\
= \; & P\left\{ \max_{1 \leq i,j \leq k} \frac{|Z_i - Z_j|}{\sqrt{SD^2/\sigma^2}} > q^{\alpha}_{k,k(n-1)} \right\} \\
= \; & P\left\{ \frac{\max_{1 \leq i \leq k} Z_i - \min_{1 \leq i \leq k} Z_i}{\sqrt{SD^2/\sigma^2}} > q^{\alpha}_{k,k(n-1)} \right\} \qquad (3.9) \\
= \; & \alpha,
\end{aligned}$$

where $Z_i, Z_{i'} \sim N(0,1)$. Equation (3.9) is equal to α because the studentized range distribution is the distribution of the range of k standard normal variables divided by $\sqrt{\chi^2_\nu/\nu}$ variable with degrees of freedom $\nu = k(n-1)$. □

From the proof above, we can develop the R code for computing the critical value of the Studentized range distribution using the numerical calculation below. Given

$$s = \sqrt{\frac{SD^2}{\sigma^2}} = \sqrt{\sigma^{-2}(k(n-1))^{-1}\sum_{i=1}^{2}\sum_{v=1}^{n_i}(X_{iv} - \bar{X}_i)^2}.$$

$$\begin{aligned} H(c) &= P\left\{\frac{\max_{1\le i\le k} Z_i - \min_{1\le i\le k} Z_i}{\sqrt{SD^2/\sigma^2}} < c\right\} \\ &= P\left\{\frac{\max_{1\le i\le k} Z_i - \min_{1\le i\le k} Z_i}{s} < c\right\} \\ &= P\left\{\max_{1\le i\le k} Z_i - \min_{1\le i\le k} Z_i < cs\right\} \\ &= \int_0^\infty f_s(s) P\left\{\max_{1\le i\le k} Z_i - \min_{1\le i\le k} Z_i < cs\right\} ds \end{aligned}$$

where $f_s(s)$ is the distribution function of χ^2_v/v and given by

$$f_s(s) = \frac{2(v/2)^{v/2} s^{v-1} e^{-vs^2/2}}{\Gamma(v/2)}, \quad v = k(n-1). \tag{3.10}$$

Then we focus on the probability,

$$\begin{aligned}
& P\left\{\max_{1\le i\le k} Z_i - \min_{1\le i\le k} Z_i < cs\right\} \\
= \; & P\left\{\min_{1\le i\le k} Z_i = Z_1, \quad \max_{1\le i\le k} Z_i - Z_1 < cs\right\} \\
& + \; P\left\{\min_{1\le i\le k} Z_i = Z_2, \quad \max_{1\le i\le k} Z_i - Z_2 < cs\right\} \\
& + \; ... \\
& + \; P\left\{\min_{1\le i\le k} Z_i = Z_k, \quad \max_{1\le i\le k} Z_i - Z_k < cs\right\} \\
= \; & k\cdot P\left\{\min_{1\le i\le k} Z_i = Z_1, \quad \max_{1\le i\le k} Z_i - Z_1 < cs\right\} \\
= \; & k\cdot P\left\{Z_1 \le \min(Z_2,...,Z_k), \quad \max(Z_2,...,Z_k) < Z_1 + cs\right\} \\
= \; & k\int_{-\infty}^{\infty} \phi(z_1) P\left\{z_1 \le \min(Z_2,...,Z_k), \quad \max(Z_2,...,Z_k) < z_1 + cs\right\} dz_1 \\
= \; & k\int_{-\infty}^{\infty} \phi(z_1) P\left\{z_1 \le Z_2 < z_1 + cs, \; z_1 \le Z_3 < z_1 + cs, ..., \; z_1 \le Z_k < z_1 + cs\right\} dz_1 \\
= \; & k\int_{-\infty}^{\infty} \phi(z_1) P\left\{z_1 \le Z_2 < z_1 + cs\right\}^{k-1} dz_1 \\
= \; & k\int_{-\infty}^{\infty} \phi(z_1)\left[\phi(z_1+cs) - \phi(z_1)\right]^{k-1} dz_1.
\end{aligned}$$

Then

$$H(c) = k\int_0^{\infty} f_s(s)\int_{-\infty}^{\infty} \phi(z_1)\left[\phi(z_1+cs) - \phi(z_1)\right]^{k-1} dz_1 ds = 1-\alpha$$

is used for writing the R code to compute the critical value c of the studentized range distribution. However, the R code for the quantile of studentized range distribution is existed as *qtukey* in package *Stats*.

3.3.2 Construct the Upper Confidence Bounds Using the Studentized Range Statistic

Then we can construct the upper confidence bounds on $\max_{1\le i,j\le k}\frac{|\mu_i-\mu_j|}{\sigma}$ using the studentized range method. Let $X_{i1}, X_{i2}, ..., X_{in} \sim N(\mu_i, \sigma^2), i = 1, 2, ..., k$. Given $B(R)$ be an $1-\alpha$ upper confidence bound on $\max_{1\le i,j\le k}\frac{|\mu_i-\mu_j|}{\sigma}$, which R is

$$R = \frac{\max_{1\le i,j\le k}|\sqrt{n}(\bar{X}_i - \bar{X}_j)|}{\sqrt{(N-k)^{-1}\sum_{i=1}^{k}\sum_{v=1}^{n}(X_{iv}-\bar{X}_i)^2}}.$$

Then

$$P\left\{\frac{\max_{1\le i,j\le k}|(\mu_i-\mu_j)|}{\sigma} \;\le\; B(R)\right\} \ge 1-\alpha.$$

Example 3.3.1. Two Samples

Let two samples $(X_{11}, X_{12}, \ldots, X_{1n})$ and $(X_{21}, X_{22}, \ldots, X_{2n})$ be the observations from two treatments which are independent and identically distributed as normal distributions with an equal variance σ^2 and means being μ_1 and μ_2 respectively. The studentized range statistic is used for constructing the upper confidence bounds on $|\mu_1-\mu_2|/\sigma$. Since $B(R)$ is an upper confidence bound for $\max_{1\le i,j\le k}\frac{|\mu_i-\mu_j|}{\sigma}$, then

$$P\left\{\frac{|\mu_1-\mu_2|}{\sigma} \le B(R)\right\} = 1-\alpha$$

$$P\left\{B^{-1}\left(\frac{|\mu_1-\mu_2|}{\sigma}\right) \le R\right\} = 1-\alpha$$

$$1-P\left\{B^{-1}\left(\frac{|\mu_1-\mu_2|}{\sigma}\right) \ge R\right\} = 1-\alpha.$$

Hence,

$$P\left\{R \le B^{-1}\left(\frac{|\mu_1-\mu_2|}{\sigma}\right)\right\} = \alpha.$$

Given $\delta = \frac{|\mu_1-\mu_2|}{\sigma}$, $Z_1, Z_2 \sim N(0,1)$ and

$$s = \sqrt{\sigma^{-2}(k(n-1))^{-1}\sum_{i=1}^{2}\sum_{v=1}^{n_i}(X_{iv}-\bar{X}_i)^2}.$$

Then

$$\begin{aligned}
& P\left\{\frac{|\sqrt{n}(\bar{X}_1-\bar{X}_2)|}{\sqrt{(N-k)^{-1}\sum_{i=1}^{2}\sum_{v=1}^{n_i}(X_{iv}-\bar{X}_i)^2}} \le B^{-1}\left(\frac{|\mu_1-\mu_2|}{\sigma}\right)\right\} \\
= \;& P\left\{|Z_1-Z_2+\sqrt{n}\left(\frac{\mu_1-\mu_2}{\sigma}\right)| \le B^{-1}\left(\frac{|\mu_1-\mu_2|}{\sigma}\right)s\right\} \\
= \;& P\left\{|Z_1-Z_2+\sqrt{n}\delta| \le B^{-1}(\delta)s\right\} \\
= \;& P\left\{-B^{-1}(\delta)s \le Z_1-Z_2+\sqrt{n}\delta \le B^{-1}(\delta)s\right\} \\
= \;& \int_0^\infty f_s(s)P\left\{-B^{-1}(\delta)s \le Z_1-Z_2+\sqrt{n}\delta \le B^{-1}(\delta)s\right\}ds \\
= \;& \int_0^\infty f_s(s)P\left\{\frac{1}{\sqrt{2}}\left(-B^{-1}(\delta)s-\sqrt{n}\delta\right) \le \frac{Z_1-Z_2}{\sqrt{2}} \le \frac{1}{\sqrt{2}}\left(B^{-1}(\delta)s-\sqrt{n}\delta\right)\right\}ds \\
= \;& \int_0^\infty f_s(s)\left[\Phi\left(\frac{1}{\sqrt{2}}\left(-B^{-1}(\delta)s-\sqrt{n}\delta\right)\right)-\Phi\left(\frac{1}{\sqrt{2}}\left(B^{-1}(\delta)s-\sqrt{n}\delta\right)\right)\right]ds \quad (3.11) \\
= \;& \alpha,
\end{aligned}$$

where $f_s(s)$ is the distribution function of χ_v^2/v and given by equation (3.10). For any given δ, we can compute $B^{-1}(\delta)$ which makes the probability in equation (3.11) equals to α. In figure 3.4, then function $B^{-1}(\delta)$ is plotted for $\alpha = 0.05$ and $n = 20$.

Example 3.3.2. Three Samples

Let Three samples $(X_{11}, X_{12}, \ldots, X_{1n})$, $(X_{21}, X_{22}, \ldots, X_{2n})$ and $(X_{31}, X_{32}, \ldots, X_{3n})$ be the observations from three treatments which are independent and identically distributed

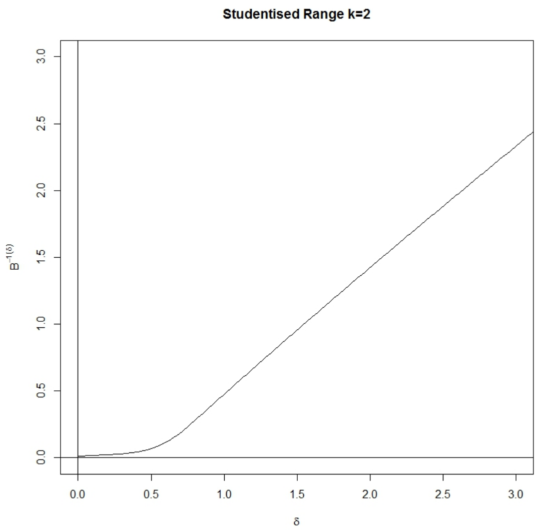

FIGURE 3.4: The relationship between given value δ and $B^{-1}(\delta)$ from the studentized range statistic method for k=2

as normal distributions with an equal variance σ^2 and means being μ_1, μ_2 and μ_3 respectively. The studentized range statistic is used for constructing the upper confidence bounds on $\max_{1\leq i,j\leq 3}|\mu_i-\mu_j|/\sigma$. Given $\delta=\max_{1\leq i,j\leq 3}\frac{|\mu_i-\mu_j|}{\sigma}$, $Z_1, Z_2, Z_3 \sim N(0,1)$, and

$$s=\sqrt{\sigma^{-2}(N-k)^{-1}\sum_{i=1}^{3}\sum_{v=1}^{n_i}(X_{iv}-\bar{X}_i)^2}.$$

Then

$$\begin{aligned} & P\left\{\frac{\max_{1\leq i,j\leq 3}\sqrt{n}(\bar{X}_i-\bar{X}_j)}{\sqrt{(N-k)^{-1}\sum_{i=1}^{3}\sum_{v=1}^{n_i}(X_{iv}-\bar{X}_i)^2}} \;\leq\; B^{-1}\left(\max_{1\leq i,j\leq 3}\frac{|\mu_i-\mu_j|}{\sigma}\right)\right\} \\ = \; & P\left\{\sqrt{n}\left(\max_{1\leq i,j\leq 3}\frac{\bar{X}_i-\bar{X}_j}{\sigma}\right) \;\leq\; B^{-1}(\delta)s\right\} \\ = \; & \int_0^\infty f_s(s)P\left\{\sqrt{n}\left(\max_{1\leq i,j\leq 3}\frac{\bar{X}_i-\bar{X}_j}{\sigma}\right) \;\leq\; B^{-1}(\delta)s\right\}ds \qquad (3.12) \\ \leq \; & \int_0^\infty f_s(s)P_{\mu^*}\left\{\sqrt{n}\left(\max_{1\leq i,j\leq 3}\frac{\bar{X}_i-\bar{X}_j}{\sigma}\right) \;\leq\; B^{-1}(\delta)s\right\}ds, \qquad (3.13) \end{aligned}$$

where $f_s(s)$ is the distribution function of χ_v^2/v given in equation (3.10). The inequality of (3.13) holds because the integrand in (3.12) is maximised when $\mu=\mu^*$ (cf. Bofinger et al., 1993), where

$$\mu^*=(\mu_1/\sigma,\mu_2/\sigma,\mu_3/\sigma,)=(-\delta/2,0,\delta/2).$$

Then we focus on the probability,

$$
\begin{aligned}
& P_{\mu^*}\left\{\sqrt{n}\left(\max_{1\le i,j\le 3}\frac{\bar{X}_i-\bar{X}_j}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\} \\
= \;& P_{\mu^*}\left\{\min_{1\le i,j\le 3}\bar{X}_i=\bar{X}_1,\ \sqrt{n}\left(\max_{1\le i,j\le 3}\frac{\bar{X}_i-\bar{X}_1}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\} \\
& +P_{\mu^*}\left\{\min_{1\le i,j\le 3}\bar{X}_i=\bar{X}_2,\ \sqrt{n}\left(\max_{1\le i,j\le 3}\frac{\bar{X}_i-\bar{X}_2}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\} \\
& +P_{\mu^*}\left\{\min_{1\le i,j\le 3}\bar{X}_i=\bar{X}_3,\ \sqrt{n}\left(\max_{1\le i,j\le 3}\frac{\bar{X}_i-\bar{X}_3}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\} \\
= \;& P_{\mu^*}\left\{\frac{\sqrt{n}\bar{X}_1}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_2}{\sigma},\frac{\sqrt{n}\bar{X}_3}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_1}{\sigma}+B^{-1}(\delta)s\right\} \\
& +P_{\mu^*}\left\{\frac{\sqrt{n}\bar{X}_2}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_1}{\sigma},\frac{\sqrt{n}\bar{X}_3}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_2}{\sigma}+B^{-1}(\delta)s\right\} \\
& +P_{\mu^*}\left\{\frac{\sqrt{n}\bar{X}_3}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_1}{\sigma},\frac{\sqrt{n}\bar{X}_2}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_3}{\sigma}+B^{-1}(\delta)s\right\} \\
= \;& P_{\mu^*}\left\{Z_1-\frac{\sqrt{n}\delta}{2} \;\le\; Z_2, Z_3+\frac{\sqrt{n}\delta}{2} \;\le\; Z_1-\frac{\sqrt{n}\delta}{2}+B^{-1}(\delta)s\right\} \\
& +P_{\mu^*}\left\{Z_2 \;\le\; Z_1-\frac{\sqrt{n}\delta}{2}, Z_3+\frac{\sqrt{n}\delta}{2} \;\le\; Z_2+B^{-1}(\delta)s\right\} \\
& +P_{\mu^*}\left\{Z_3+\frac{\sqrt{n}\delta}{2} \;\le\; Z_1-\frac{\sqrt{n}\delta}{2}, Z_2 \;\le\; Z_3+\frac{\sqrt{n}\delta}{2}+B^{-1}(\delta)s\right\} \\
= \;& \int_{-\infty}^{\infty}\phi(z_1)P_{\mu^*}\left\{z_1-\frac{\sqrt{n}\delta}{2} \;\le\; Z_2 \;\le\; z_1-\frac{\sqrt{n}\delta}{2}+B^{-1}(\delta)s\right\}\cdot \\
& \quad P_{\mu^*}\left\{z_1-\frac{\sqrt{n}\delta}{2} \;\le\; Z_3+\frac{\sqrt{n}\delta}{2} \;\le\; z_1-\frac{\sqrt{n}\delta}{2}+B^{-1}(\delta)s\right\}dz_1 \\
& +\int_{-\infty}^{\infty}\phi(z_2)P_{\mu^*}\left\{z_2 \;\le\; Z_1-\frac{\sqrt{n}\delta}{2} \;\le\; z_2+B^{-1}(\delta)s\right\}\cdot \\
& \quad P_{\mu^*}\left\{z_2 \;\le\; Z_3+\frac{\sqrt{n}\delta}{2} \;\le\; z_2+B^{-1}(\delta)s\right\}dz_2 \\
& +\int_{-\infty}^{\infty}\phi(z_3)P_{\mu^*}\left\{z_3+\frac{\sqrt{n}\delta}{2} \;\le\; Z_1-\frac{\sqrt{n}\delta}{2} \;\le\; z_3+\frac{\sqrt{n}\delta}{2}+B^{-1}(\delta)s\right\}\cdot \\
& \quad P_{\mu^*}\left\{z_3+\frac{\sqrt{n}\delta}{2} \;\le\; Z_2 \;\le\; z_3+\frac{\sqrt{n}\delta}{2}+B^{-1}(\delta)s\right\}dz_3
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \phi(z_1) P_{\mu^*}\left\{z_1 - \frac{\sqrt{n}\delta}{2} \ \le\ Z_2 \ \le\ z_1 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} \cdot \\
&\quad P_{\mu^*}\left\{z_1 - \sqrt{n}\delta \ \le\ Z_3 \ \le\ z_1 - \sqrt{n}\delta + B^{-1}(\delta)s\right\} dz_1 \\
&\quad + \int_{-\infty}^{\infty} \phi(z_2) P_{\mu^*}\left\{z_2 + \frac{\sqrt{n}\delta}{2} \ \le\ Z_1 \ \le\ z_2 + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} \cdot \\
&\quad P_{\mu^*}\left\{z_2 - \frac{\sqrt{n}\delta}{2} \ \le\ Z_3 \ \le\ z_2 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} dz_2 \\
&\quad + \int_{-\infty}^{\infty} \phi(z_3) P_{\mu^*}\left\{z_3 + \sqrt{n}\delta \ \le\ Z_1 \ \le\ z_3 + \sqrt{n}\delta + B^{-1}(\delta)s\right\} \cdot \\
&\quad P_{\mu^*}\left\{z_3 + \frac{\sqrt{n}\delta}{2} \ \le\ Z_2 \ \le\ z_3 + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} dz_3 \\
&= \int_{-\infty}^{\infty} \phi(z_1)\left[\phi(z_1 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s) - \phi(z_1 - \frac{\sqrt{n}\delta}{2})\right] \cdot \\
&\quad \left[\phi(z_1 - \sqrt{n}\delta + B^{-1}(\delta)s) - \phi(z_1 - \sqrt{n}\delta)\right] dz_1 \\
&\quad + \int_{-\infty}^{\infty} \phi(z_2 + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s) - \phi(z_2)\left[\phi(z_2 + \frac{\sqrt{n}\delta}{2})\right] \cdot \\
&\quad \left[\phi(z_2 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s) - \phi(z_2 - \frac{\sqrt{n}\delta}{2})\right] dz_2 \\
&\quad + \int_{-\infty}^{\infty} \phi(z_3)\left[\phi(z_3 + \sqrt{n}\delta + B^{-1}(\delta)s) - \phi(z_3 + \sqrt{n}\delta)\right] \cdot \\
&\quad \left[\phi(z_3 + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s) - \phi(z_3 + \frac{\sqrt{n}\delta}{2})\right] dz_3.
\end{aligned}
$$

For any given δ, we can compute $B^{-1}(\delta)$ which makes probability in equation (3.13) equal to α. In figure 3.5, then function $B^{-1}(\delta)$ is plotted for $\alpha = 0.05$ and $n = 20$.

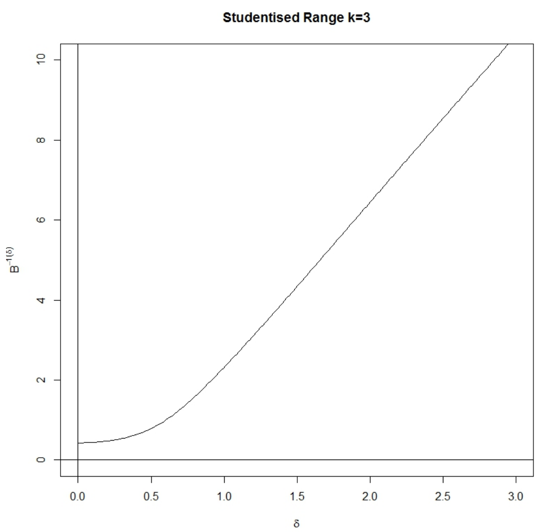

FIGURE 3.5: The relationship between given value δ and $B^{-1}(\delta)$ from the studentized range statistic method for k=3

Example 3.3.3. k **Samples**

Let k samples $(X_{11}, X_{12}, \ldots, X_{1n})$, $(X_{21}, X_{22}, \ldots, X_{2n})$,...,$(X_{k1}, X_{k2}, \ldots, X_{kn})$ be the observations from k treatments which are independent and identically distributed as normal distributions with an equal variance σ^2 and means being μ_1, μ_2,...,μ_k respectively. The studentized range statistic is used for constructing the upper confidence bounds on $\max_{1\le i,j\le k} |\mu_i - \mu_j|/\sigma$. Given $\delta = \max_{1\le i,j\le k} \frac{|\mu_i-\mu_j|}{\sigma}$, $Z_1, Z_2, ..., Z_k \sim N(0,1)$, and

$$s = \sqrt{\sigma^{-2}(N-k)^{-1}\sum_{i=1}^{k}\sum_{v=1}^{n_i}(X_{iv}-\bar{X}_i)^2}.$$

Then

$$\begin{aligned}
& P\left\{\frac{\max_{1\le i,j\le k}\sqrt{n}(\bar{X}_i-\bar{X}_j)}{\sqrt{(N-k)^{-1}\sum_{i=1}^{k}\sum_{v=1}^{n_i}(X_{iv}-\bar{X}_i)^2}} \;\le\; B^{-1}\left(\max_{1\le i,j\le k}\frac{|\mu_i-\mu_j|}{\sigma}\right)\right\} \\
= \; & P\left\{\sqrt{n}\left(\max_{1\le i,j\le k}\frac{\bar{X}_i-\bar{X}_j}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\} \\
= \; & \int_0^\infty f_s(s)P\left\{\sqrt{n}\left(\max_{1\le i,j\le k}\frac{\bar{X}_i-\bar{X}_j}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\}ds \qquad (3.14) \\
\le \; & \int_0^\infty f_s(s)P_{\mu^*}\left\{\sqrt{n}\left(\max_{1\le i,j\le k}\frac{\bar{X}_i-\bar{X}_j}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\}ds, \qquad (3.15)
\end{aligned}$$

where $f_s(s)$ is the distribution function of χ^2_v/v given in equation (3.10). The inequality of (3.15) holds because the integrand in (3.14) is maximised when $\mu = \mu^*$ (cf. Bofinger et al., 1993), where

$$\mu^* = (\mu_1/\sigma, \mu_2/\sigma, ..., \mu_3/\sigma,) = (-\delta/2, 0, ..., 0, \delta/2).$$

Then we focus on the probability,

$$\begin{aligned}
& P_{\mu^*}\left\{\sqrt{n}\left(\max_{1\le i,j\le k}\frac{\bar{X}_i-\bar{X}_j}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\} \\
= \; & P_{\mu^*}\left\{\min_{1\le i,j\le k}\bar{X}_i=\bar{X}_1,\; \sqrt{n}\left(\max_{1\le i,j\le k}\frac{\bar{X}_i-\bar{X}_1}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\} \\
& +P_{\mu^*}\left\{\min_{1\le i,j\le k}\bar{X}_i=\bar{X}_2,\; \sqrt{n}\left(\max_{1\le i,j\le k}\frac{\bar{X}_i-\bar{X}_2}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\} \\
& +\ldots \\
& +P_{\mu^*}\left\{\min_{1\le i,j\le k}\bar{X}_i=\bar{X}_k,\; \sqrt{n}\left(\max_{1\le i,j\le k}\frac{\bar{X}_i-\bar{X}_k}{\sigma}\right) \;\le\; B^{-1}(\delta)s\right\} \\
= \; & P_{\mu^*}\left\{\frac{\sqrt{n}\bar{X}_1}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_2}{\sigma}, ..., \frac{\sqrt{n}\bar{X}_k}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_1}{\sigma}+B^{-1}(\delta)s\right\} \\
& +P_{\mu^*}\left\{\frac{\sqrt{n}\bar{X}_2}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_1}{\sigma}, ..., \frac{\sqrt{n}\bar{X}_k}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_2}{\sigma}+B^{-1}(\delta)s\right\} \\
& +\ldots \\
& +P_{\mu^*}\left\{\frac{\sqrt{n}\bar{X}_k}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_1}{\sigma}, ..., \frac{\sqrt{n}\bar{X}_{k-1}}{\sigma} \;\le\; \frac{\sqrt{n}\bar{X}_k}{\sigma}+B^{-1}(\delta)s\right\}
\end{aligned}$$

$$
\begin{aligned}
&= P_{\mu^*}\left\{Z_1 - \frac{\sqrt{n}\delta}{2} \ \le\ Z_2, ..., Z_k + \frac{\sqrt{n}\delta}{2} \ \le\ Z_1 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} \\
&\quad +(k-2)P_{\mu^*}\left\{Z_2 \ \le\ Z_1 - \frac{\sqrt{n}\delta}{2}, Z_3, ..., Z_k + \frac{\sqrt{n}\delta}{2} \ \le\ Z_2 + B^{-1}(\delta)s\right\} \\
&\quad +P_{\mu^*}\left\{Z_k + \frac{\sqrt{n}\delta}{2} \ \le\ Z_1 - \frac{\sqrt{n}\delta}{2}, Z_2, ..., Z_{k-1} \ \le\ Z_k + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} \\
&= \int_{-\infty}^{\infty} \phi(z_1)P_{\mu^*}\left\{z_1 - \frac{\sqrt{n}\delta}{2} \ \le\ Z_2 \ \le\ z_1 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} \cdot \\
&\quad\ P_{\mu^*}\left\{z_1 - \frac{\sqrt{n}\delta}{2} \ \le\ Z_3 \ \le\ z_1 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} \cdot \\
&\quad\ \cdots \\
&\quad\ P_{\mu^*}\left\{z_1 - \frac{\sqrt{n}\delta}{2} \ \le\ Z_k + \frac{\sqrt{n}\delta}{2} \ \le\ z_1 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} dz_1 \\
&\quad + \int_{-\infty}^{\infty} \phi(z_2)P_{\mu^*}\left\{z_2 \ \le\ Z_1 - \frac{\sqrt{n}\delta}{2} \ \le\ z_2 + B^{-1}(\delta)s\right\} \cdot \\
&\quad\ P_{\mu^*}\left\{z_2 \ \le\ Z_3 \ \le\ z_2 + B^{-1}(\delta)s\right\} \cdot \\
&\quad\ \cdots \\
&\quad\ P_{\mu^*}\left\{z_2 \ \le\ Z_k + \frac{\sqrt{n}\delta}{2} \ \le\ z_2 + B^{-1}(\delta)s\right\} dz_2 \\
&\quad + \int_{-\infty}^{\infty} \phi(z_k)P_{\mu^*}\left\{z_k + \frac{\sqrt{n}\delta}{2} \ \le\ Z_1 - \frac{\sqrt{n}\delta}{2} \ \le\ z_k + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} \cdot \\
&\quad\ P_{\mu^*}\left\{z_k + \frac{\sqrt{n}\delta}{2} \ \le\ Z_2 \ \le\ z_k + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} dz_3 \\
&\quad\ \cdots \\
&\quad\ P_{\mu^*}\left\{z_k + \frac{\sqrt{n}\delta}{2} \ \le\ Z_{k-1} \ \le\ z_k + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} dz_k \\
&= \int_{-\infty}^{\infty} \phi(z_1)P_{\mu^*}\left\{z_1 - \frac{\sqrt{n}\delta}{2} \ \le\ Z_2 \ \le\ z_1 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\}^{k-2} \cdot \\
&\quad\ P_{\mu^*}\left\{z_1 - \sqrt{n}\delta \ \le\ Z_k \ \le\ z_1 - \sqrt{n}\delta + B^{-1}(\delta)s\right\} dz_1 \\
&\quad +(k-2)\int_{-\infty}^{\infty} \phi(z_2)P_{\mu^*}\left\{z_2 + \frac{\sqrt{n}\delta}{2} \ \le\ Z_1 \ \le\ z_2 + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} \cdot \\
&\quad\ P_{\mu^*}\left\{z_2 \ \le\ Z_3 \ \le\ z_2 + B^{-1}(\delta)s\right\}^{k-3} \\
&\quad\ P_{\mu^*}\left\{z_2 - \frac{\sqrt{n}\delta}{2} \ \le\ Z_k \ \le\ z_2 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\} dz_2 \\
&\quad + \int_{-\infty}^{\infty} \phi(z_k)P_{\mu^*}\left\{z_k + \sqrt{n}\delta \ \le\ Z_1 \ \le\ z_k + \sqrt{n}\delta + B^{-1}(\delta)s\right\} \cdot \\
&\quad\ P_{\mu^*}\left\{z_k + \frac{\sqrt{n}\delta}{2} \ \le\ Z_2 \ \le\ z_k + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s\right\}^{k-2} dz_k
\end{aligned}
$$

$$
\begin{aligned}
= & \int_{-\infty}^{\infty} \phi(z_1) \left[\phi(z_1 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s) - \phi(z_1 - \frac{\sqrt{n}\delta}{2})\right]^{k-2} \cdot \\
& \left[\phi(z_1 - \sqrt{n}\delta + B^{-1}(\delta)s) - \phi(z_1 - \sqrt{n}\delta)\right] dz_1 \\
& +(k-2) \int_{-\infty}^{\infty} \phi(z_2) \left[\phi(z_2 + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s) - \phi(z_2 + \frac{\sqrt{n}\delta}{2})\right] \cdot \\
& \left[\phi(z_2 + B^{-1}(\delta)s) - \phi(z_2)\right]^{k-3} \cdot \\
& \left[\phi(z_2 - \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s) - \phi(z_2 - \frac{\sqrt{n}\delta}{2})\right] dz_2 \\
& + \int_{-\infty}^{\infty} \phi(z_k) \left[\phi(z_k + \sqrt{n}\delta + B^{-1}(\delta)s) - \phi(z_k + \sqrt{n}\delta)\right] \cdot \\
& \left[\phi(z_k + \frac{\sqrt{n}\delta}{2} + B^{-1}(\delta)s) - \phi(z_k + \frac{\sqrt{n}\delta}{2})\right]^{k-2} dz_k.
\end{aligned}
$$

For any given δ, we can compute $B^{-1}(\delta)$ which makes probability in equation (3.15) equal to α. In figure 3.6, then function $B^{-1}(\delta)$ is plotted for $k = 5$, $\alpha = 0.05$ and $n = 20$.

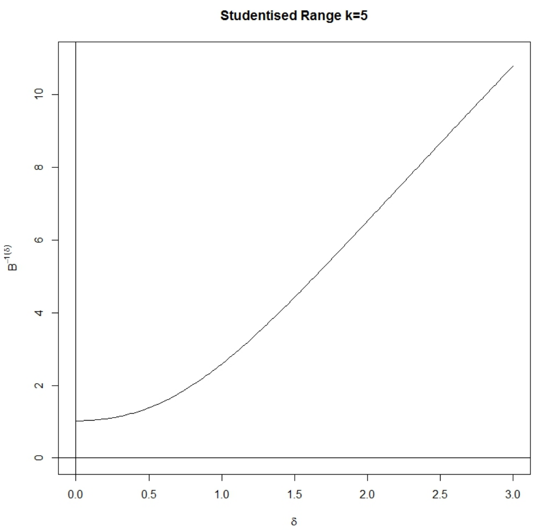

FIGURE 3.6: The relationship between given value δ and $B^{-1}(\delta)$ from the studentized range statistic method for k=5

3.4 Comparison of the Upper Confidence Bounds from Several Methods

The method in section 3.1 cannot give an upper confidence bound for $\frac{|\mu_1-\mu_2|}{\sigma}$ if σ is unknown. Hence, for comparing the upper confidence bounds among three methods, in sections 3.1 - 3.3, we consider the situation of known σ, and without loss of generality, assume $\sigma = 1$. When σ is given, the F statistic method changes the F distribution to Chi-squared distribution. Then the Chi-squared distribution is used in the second method instead of the F distribution. The studentized range statistic method is also changed to the range statistic when σ is known. Then the range statistic is used in the third method instead of the studentized range statistic.

Example 3.4.1. Two Samples
Let two samples $(X_{11}, X_{12}, \dots, X_{1n})$ and $(X_{21}, X_{22}, \dots, X_{2n})$ be the observations from two treatments which are independent and identically distributed as normal distributions with a known equal variance $\sigma^2 = 1$ and means being μ_1 and μ_2 respectively. The upper confidence bounds on $|\mu_1 - \mu_2|/\sigma$ from three methods, in section 3.1 - 3.3, are computed,

$$\begin{aligned}
\frac{|\mu_1-\mu_2|}{\sigma} &\leq \frac{|\bar{X}_1-\bar{X}_2|}{\sigma} + Z^{\alpha}\sqrt{\frac{2}{n}}, \\
\frac{|\mu_i-\mu_j|}{\sigma} &\leq B^F\left(\frac{n\sum_{i=1}^{2}(\bar{X}_i-\bar{X}_.)^2}{\sigma^2}\right), \\
\frac{|\mu_i-\mu_j|}{\sigma} &\leq B^R\left(\frac{\sqrt{n}|\bar{X}_1-\bar{X}_2|}{\sigma}\right).
\end{aligned}$$

For any given $\delta = |\bar{X}_1 - \bar{X}_2|/\sigma$, we can compute the upper confidence bounds on $|\mu_1 - \mu_2|/\sigma$. However, the upper confidence bounds from the Chi-squared statistic method is exactly the same as the one from the range statistic method. In figure 3.7, three upper confidence bounds is plotted for any given δ, where $\alpha = 0.05$ and $n = 20$.

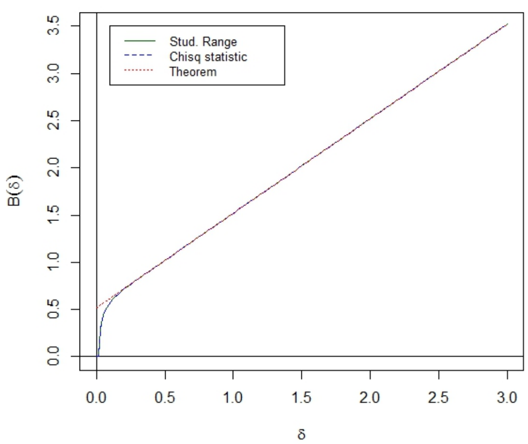

FIGURE 3.7: The upper confidence bounds from three different methods for $k = 2$

From figure 3.7, the upper confidence bounds from the first method, in section 3.1, are always above the other methods especially when δ is very small. Hence, we can conclude that for $k = 2$ case, The upper confidence bounds from the Chi-squared statistic method and the range statistic method are better than the method in section 3.1. However, the upper confidence bounds from these three methods are almost the same when δ is large.

Example 3.4.2. Three Samples

Let Three samples $(X_{11}, X_{12}, \ldots, X_{1n})$, $(X_{21}, X_{22}, \ldots, X_{2n})$ and $(X_{31}, X_{32}, \ldots, X_{3n})$ be the observations from three treatments which are independent and identically distributed as normal distributions with a known equal variance $\sigma^2 = 1$ and means being μ_1, μ_2 and μ_3 respectively. The upper confidence bounds on $\max_{1\le i,j\le 3} |\mu_i - \mu_j|/\sigma$ from three methods, in section 3.1 - 3.3, are computed,

$$\begin{aligned}
\max_{1\le i,j\le 3} \frac{|\mu_i - \mu_j|}{\sigma} &\le \max_{1\le i,j\le 3} \frac{|\bar{X}_i - \bar{X}_j|}{\sigma} + Z^{\alpha}\sqrt{\frac{2}{n}}, \\
\max_{1\le i,j\le 3} \frac{|\mu_i - \mu_j|}{\sigma} &\le B^F\left(\frac{n\sum_{i=1}^{k}(\bar{X}_i - \bar{X}_.)^2/(3-1)}{\sigma^2}\right), \\
\max_{1\le i,j\le 3} \frac{|\mu_i - \mu_j|}{\sigma} &\le B^R\left(\max_{1\le i,j\le 3} \frac{\sqrt{n}|\bar{X}_i - \bar{X}_j|}{\sigma}\right).
\end{aligned}$$

Given

$$\begin{aligned}
\frac{\bar{x}_1 - \bar{x}_2}{\sigma} &= u, \\
\frac{\bar{x}_2 - \bar{x}_3}{\sigma} &= v, \\
\text{and} \quad \frac{\bar{x}_1 - \bar{x}_3}{\sigma} &= u + v.
\end{aligned}$$

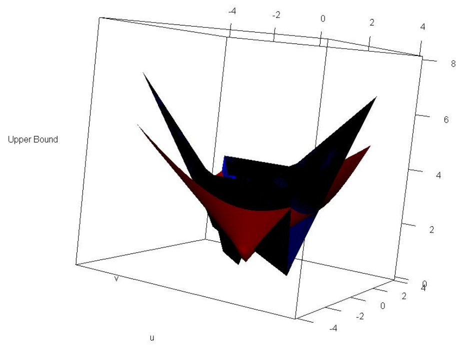

FIGURE 3.8: The upper confidence bounds from three different methods for $k = 3$

For any given u and v, we can compute the upper confidence bounds on $\max_{1\le i,j\le 3} |\mu_i - \mu_j|/\sigma$. In figure 3.8, the three upper confidence bounds are plotted for any given u and v, where $\alpha = 0.05$ and $n = 20$. The different colours represent the different upper confidence bounds. The black colour shows the upper confidence bounds from the range

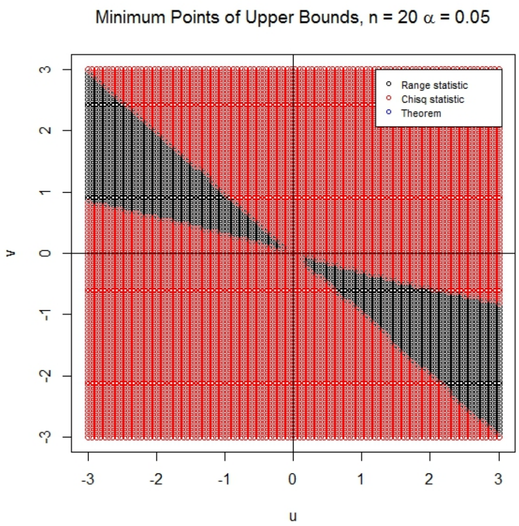

FIGURE 3.9: The minimum of the upper confidence bounds from three different methods for $k = 3$

statistic method. The red colour shows the upper confidence bounds from the Chi-squared statistic method. The blue colour shows the upper confidence bounds from the method in section 3.1. Figure 3.9 presents the minimum point of the upper confidence bounds from three methods. We can see that no point from the method in section 3.1 is the minimum of the bounds. Hence, this method always provides the largest upper confidence bounds compared with other methods. The Chi-squared statistic method in red dots is presented the most area of the graph. That is the Chi-squared statistic method provides most of the minimum upper confidence bounds in this given region $-3 \leq u, v \leq 3$.

Example 3.4.3. Different Sample Sizes

The effect of sample sizes, $n = 5, 20, 100$ and 1000, is evaluated. Let $X_{i1}, X_{i2}, ..., X_{in} \sim N(\mu_i, \sigma^2)$, $i = 1, 2, 3$, $n_1 = n_2 = n_3 = n$ and $\sigma^2 = 1$. The upper confidence bounds on $\max_{1 \leq i,j \leq 3} |\mu_i - \mu_j|/\sigma$ from three methods: the range statistic, the Chi-squared statistic and the method in section 3.1, are computed for any given u and v as in example 3.4.2. The figure 3.10, 3.11, 3.12 and 3.13 show the methods which provide the minimum point of the upper confidence bounds for different sample sizes: $n = 5, 20, 100$ and 1000, on the (u, v) plots. The different colour dots present the different methods for computing the upper confidence bounds. The first plot from all figures show the minimum point among three methods. We can see that the Chi-squared statistic method provides the most minimum points in this given region $-3 \leq u, v \leq 3$ whether n is changed or not. The upper right plot for each figure presents the Chi-squared statistic versus the range statistic methods, and the lower left plots present the Chi-squared statistic method versus the method in section 3.1, which provide the minimum point of the upper confidence bounds. From these three plots, the difference of sample sizes does not much effect to any methods. However, the lower right plots, which present the range statistic method versus the method in section 3.1, show the effect of sample sizes to the difference between the upper confidence bounds from two methods. The yellow colour dots reveal that the difference between the upper confidence bounds from the range statistic method and the method in section 3.1 is less than 0.001. It is clear that when the sample size is increased, the upper confidence bounds from these two methods are closer in this given region $-3 \leq u, v \leq 3$.

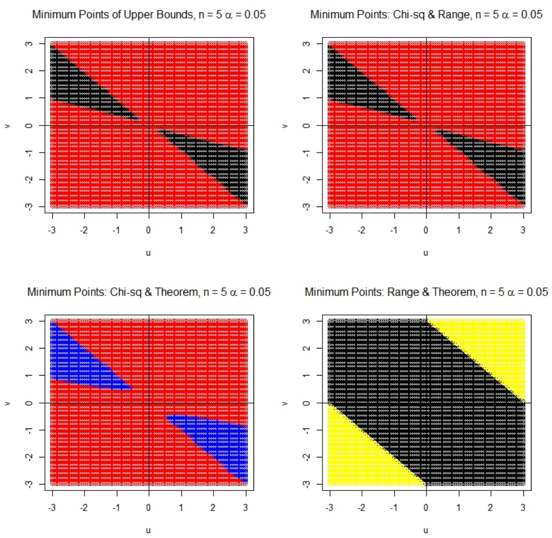

FIGURE 3.10: Comparison of the upper confidence bounds for $k = 3,\ n = 5$

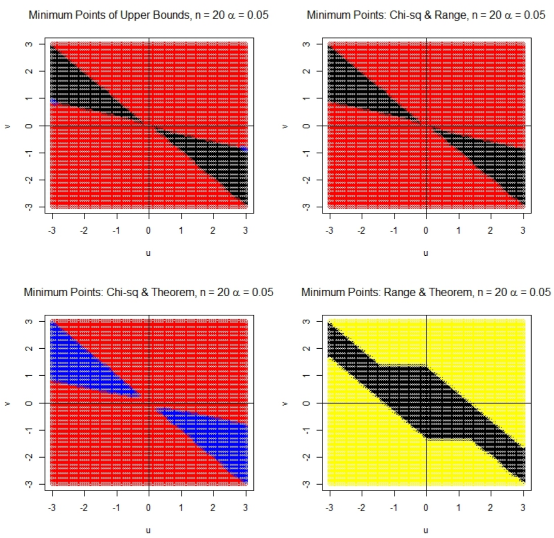

FIGURE 3.11: Comparison of the upper confidence bounds for $k = 3,\ n = 20$

Example 3.4.4. Different Confidence Levels

The effect of confidence levels, $\alpha = 0.01, 0.05, 0.10$ and 0.20, is evaluated. Let $X_{i1}, X_{i2}, ..., X_{in} \sim N(\mu_i, \sigma^2),\ i = 1, 2, 3,\ n_1 = n_2 = n_3 = n = 20$ and $\sigma^2 = 1$. The upper confidence bounds on $\max_{1 \leq i,j \leq 3} |\mu_i - \mu_j|/\sigma$ from three methods: the range statistic, the Chi-squared statistic and the method in section 3.1, are computed for any given u and v as example 3.4.2. The figure 3.14 shows the methods which provide the minimum point of the upper confidence bounds for different confidence levels, $\alpha = 0.01, 0.05, 0.10$ and 0.20, on the (u, v)

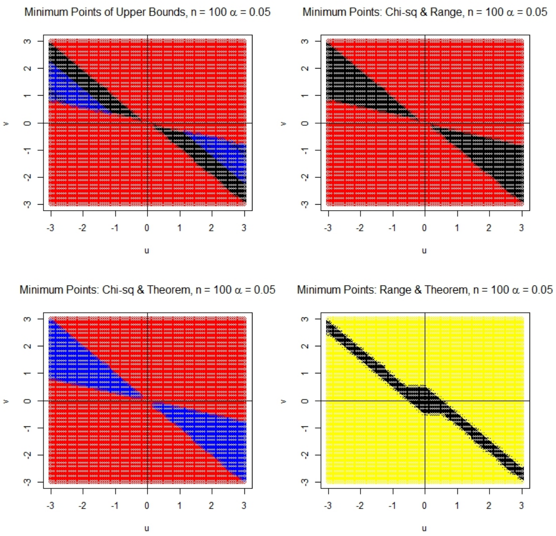

FIGURE 3.12: Comparison of the upper confidence bounds for $k = 3,\ n = 100$

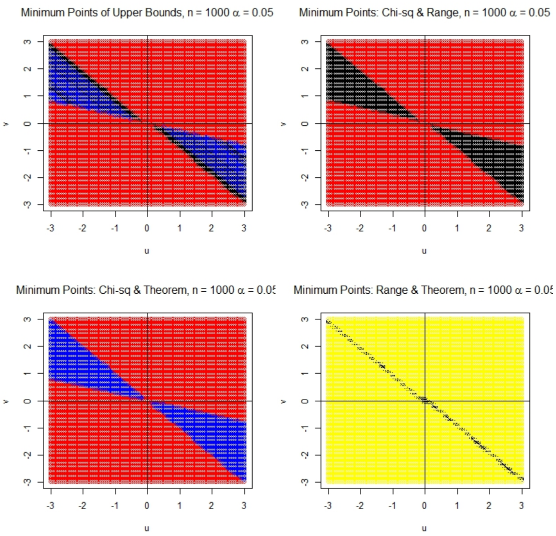

FIGURE 3.13: Comparison of the upper confidence bounds for $k = 3,\ n = 1000$

plots. The different colour dots present the different methods for computing the upper confidence bounds. From those four plots, the difference of confidence levels does not much effect to the methods which provide the minimum point of the upper confidence bounds in this given region $-3 \leq u, v \leq 3$.

Example 3.4.5. Unbalanced Sample Sizes

In this study, the effect of unbalanced sample sizes is evaluated for only two sample groups, $k = 2$. For $k > 2$, a complicated numerical computation is required. Let $X_{i1}, X_{i2}, ..., X_{in_i} \sim N(\mu_i, \sigma^2),\ i = 1, 2$ and $\sigma^2 = 1$. The range statistic method cannot construct the upper confidence bounds for unbalanced sample sizes. Then the upper confidence bounds on $|\mu_1 - \mu_2|/\sigma$ from two methods: the Chi-squared statistic and the method in section 3.1, are computed. Given $\delta^* = |\bar{x_1} - \bar{x_2}|/\sigma$. The figure 3.15 shows the upper confidence bounds for different unbalance sample sizes, $n_1 : n_2 = 10 : 10, 10 : 20, 10 : 50$ and $10 : 100$. The red colour line shows the upper confidence bounds from

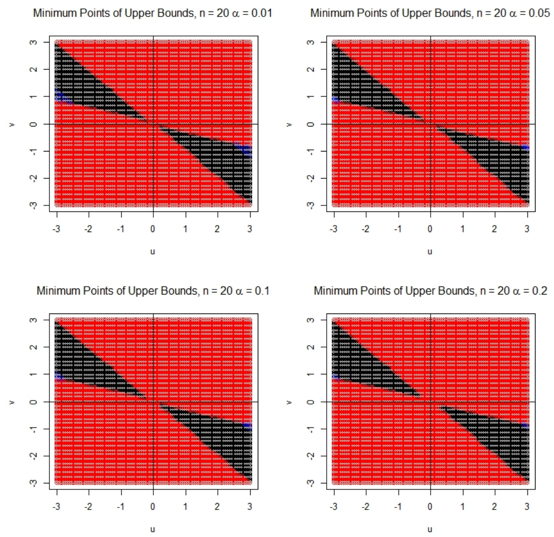

FIGURE 3.14: Comparison of the upper confidence bounds for different α

the Chi-squared statistic method, and the blue colour line shows the upper confidence bounds from the method in section 3.1. From those four plots, the different unbalanced sample sizes does not effect to the methods which provide the minimum upper confidence bounds. That is the upper confidence bounds from the Chi-squared statistic are always lower than the ones from the theorem for $k = 2$. However, the sample size effects on the upper confidence bounds, which are smaller when sample size is increased.

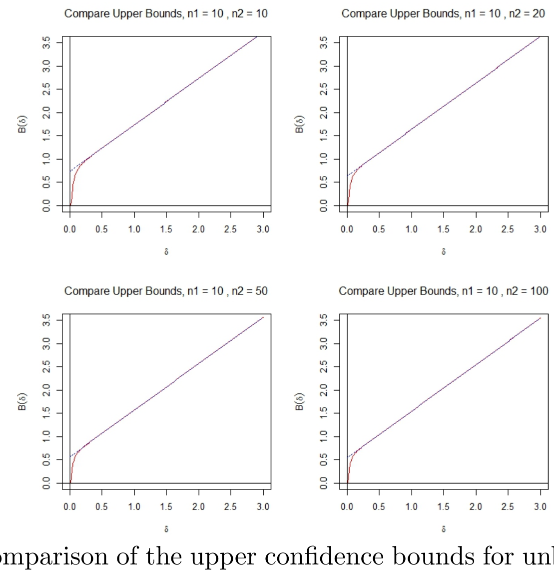

FIGURE 3.15: Comparison of the upper confidence bounds for unbalanced sample sizes

3.5 Conclusion

In this chapter, we propose the construction of upper confidence bounds on the range of k means by three methods: using the theorem from Liu et al. (2009), using the F-statistic and using the studentized range statistic. The first method, using the theorem from Liu et al. (2009), cannot compute the upper confidence bound of $\frac{|\mu_1-\mu_2|}{\sigma}$ if σ is unknown. However, it can be use for calculating the upper confidence bound of $|\mu_1 - \mu_2|$ whether σ is known or unknown, and the data is not necessary to be normal distribution for any situation. The other two methods, using the F-statistic and using the studentized range statistic, the data needs to be normal distribution, but σ is not necessary for computation. However, if σ is know, the F-statistic and studentized range statistic are changed in this situation. Then the Chi-squared statistic and the range statistic are the suitable adapted methods for this case. For $k = 2$, we can conclude that, The upper confidence bounds from the Chi-squared statistic method and the range statistic method are better than the first method. However, the upper confidence bounds from these three methods are almost the same when δ is large. It is also the same for $k = 3$ in the given region $-3 \leq u, v \leq 3$, the upper confidence bounds from the Chi-squared statistic method and the range statistic method are better than the first method. There is no effects of sample sizes, unbalanced sample sizes and confidence levels on the minimum point of upper confidence bounds among three methods. However, when the sample size is increased, the upper confidence bounds from the range statistic method and the first method are closer. Moreover, the increasing of the sample size effects on the smaller upper confidence bounds. Then, we are moving to the next problem in next Chapter, the construction of upper confidence bounds on the maximum difference between two regression lines.

Chapter 4

Construction of the Upper Confidence Bounds on the Maximum Difference Between Two Regression Lines

The construction of upper confidence bounds on the range of k means was discussed in the previous Chapter. In this chapter, the upper confidence bounds on the maximum difference between two regression lines, $y = \beta_{01} + \beta_{11}x$ and $y = \beta_{02} + \beta_{12}x$, are the next problem that we have considered. We propose the construction of upper confidence bounds by two methods: by the theorem 3.1.1 from Liu et al. (2009) and by the Chi-squared statistic. The construction using the first method, which is mentioned in the previous chapter, depends on several upper confidence bounds that will be computed by two methods in this study: general confidence bounds and the normal distribution method. The last section provides some examples for comparing the upper confidence bounds from those methods. All examples will be presented as two regression lines from two different sample groups. As in the previous Chapter, the theorem 3.1.1 from Liu et al. (2009) cannot give an upper confidence bound for $\frac{|\mu_1-\mu_2|}{\sigma}$ if σ is unknown. Hence, we consider the situation of known σ for all the examples in this Chapter to compare the upper confidence bounds on

$$\max_{x\in[a,b]} \frac{|(\beta_{01} + \beta_{11}x) - (\beta_{02} + \beta_{12}x)|}{\sigma}.$$

In each example, $\hat{\sigma}$ from the data set is assumed to be the known σ.

4.1 Using the Theorem from Liu et al. (2009)

According to the theorem 3.1.1 from Liu et al. (2009), given several $(1-\alpha)$ upper confidence bounds for several parameters in parameter space. Then the upper confidence bound on

$$\max_{x\in[a,b]} |(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|$$

can be computed, by theorem 3.1.1 in previous chapter, using the upper confidence bound on

$$|(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)| \ \text{ for each given } x\in[a,b].$$

This upper confidence bound for each given $x\in[a,b]$ will be computed by two methods: by general confidence bounds and by the normal distribution method.

4.1.1 by General Confidence Bounds

To compute the upper confidence bound on $\max_{x\in[a,b]} |(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|$ using the theorem 3.1.1, we need the upper confidence bound for each x. In this part, the upper confidence bounds are computed by general confidence bounds. When σ is known, we know that the $(1-\alpha)$ upper confidence bound of $(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)$ is

$$P\left\{(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)\le(\hat{\beta_{01}}+\hat{\beta_{11}}x)-(\hat{\beta_{02}}+\hat{\beta_{12}}x)+Z^{\alpha}\sigma\sqrt{x^T\Delta x}\right\}=1-\alpha. \tag{4.1}$$

The $(1-\alpha)$ lower confidence bound of $(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)$ is

$$P\left\{(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)\ge(\hat{\beta_{01}}+\hat{\beta_{11}}x)-(\hat{\beta_{02}}+\hat{\beta_{12}}x)-Z^{\alpha}\sigma\sqrt{x^T\Delta x}\right\}=1-\alpha,$$

or equivalently

$$P\left\{-\left[(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)\right]\le-\left[(\hat{\beta_{01}}+\hat{\beta_{11}}x)-(\hat{\beta_{02}}+\hat{\beta_{12}}x)\right]+Z^{\alpha}\sigma\sqrt{x^T\Delta x}\right\}=1-\alpha, \tag{4.2}$$

where $\Delta=(X_1^TX_1)^{-1}+(X_2^TX_2)^{-1}$. Equation (4.1), equation (4.2) and the theorem 3.1.1 are used to construct the upper confidence bound of $|(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|$, that is

$$P\left\{|(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|\le|(\hat{\beta_{01}}+\hat{\beta_{11}}x)-(\hat{\beta_{02}}+\hat{\beta_{12}}x)|+Z^{\alpha}\sigma\sqrt{x^T\Delta x}\right\}\ge1-\alpha. \tag{4.3}$$

Here we get the upper confidence bound of $|(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|$ for each x. Then we can construct the upper confidence bound on $\max_{x\in[a,b]} |(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|$

using equation (4.3) and theorem 3.1.1:

$$P\left\{\max_{x\in[a,b]}|(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|\le\max_{x\in[a,b]}\left[|(\hat{\beta_{01}}+\hat{\beta_{11}}x)-(\hat{\beta_{02}}+\hat{\beta_{12}}x)|+Z^{\alpha}\sigma\sqrt{x^T\Delta x}\right]\right\}\ge 1-\alpha.$$

That is

$$P\left\{\max_{x\in[a,b]}\frac{|(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|}{\sigma}\le\max_{x\in[a,b]}\left[\frac{|(\hat{\beta_{01}}+\hat{\beta_{11}}x)-(\hat{\beta_{02}}+\hat{\beta_{12}}x)|}{\sigma}+Z^{\alpha}\sqrt{x^T\Delta x}\right]\right\}\ge 1-\alpha$$

Hence, the upper confidence bound of $\max_{x\in[a,b]}|(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|/\sigma$ is

$$\max_{x\in[a,b]}\left[\frac{|(\hat{\beta_{01}}+\hat{\beta_{11}}x)-(\hat{\beta_{02}}+\hat{\beta_{12}}x)|}{\sigma}+Z^{\alpha}\sqrt{x^T\Delta x}\right].$$

4.1.2 by the Normal Distribution Method

To compute the upper confidence bound on $\max_{x\in[a,b]}|(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|/\sigma$ using the theorem 3.1.1, we need the upper confidence bound for each x. In this part, the upper confidence bounds are computed by the normal distribution method. Let $\mathbf{x}^T=(1,x)$, $\mathbf{x}^T\boldsymbol{\beta}_1=(\beta_{01}+\beta_{11}x)$, $\mathbf{x}^T\boldsymbol{\beta}_2=(\beta_{02}+\beta_{12}x)$, and given $B(|\mathbf{x}^T\hat{\boldsymbol{\beta}}_1-\mathbf{x}^T\hat{\boldsymbol{\beta}}_2|/\sigma)$ be an $1-\alpha$ upper confidence bound on $|\mathbf{x}^T\boldsymbol{\beta}_1-\mathbf{x}^T\boldsymbol{\beta}_2|/\sigma$ for each given x. That is

$$P\left\{\frac{|\mathbf{x}^T\boldsymbol{\beta}_1-\mathbf{x}^T\boldsymbol{\beta}_2|}{\sigma}\le B\left(\frac{|\mathbf{x}^T\hat{\boldsymbol{\beta}}_1-\mathbf{x}^T\hat{\boldsymbol{\beta}}_2|}{\sigma}\right)\right\}=1-\alpha$$

Let $d=\mathbf{x}^T\boldsymbol{\beta}_1-\mathbf{x}^T\boldsymbol{\beta}_2/\sigma$, $w^2=\mathbf{x}^T\Delta\mathbf{x}$ and

$$\frac{\mathbf{x}^T\hat{\boldsymbol{\beta}}_1-\mathbf{x}^T\hat{\boldsymbol{\beta}}_2}{\sigma}\sim N(d,w^2).$$

Then we can compute function $B^{-1}(\cdot)$ by inverting technique,

$$\begin{aligned}
& P\left\{\frac{|\mathbf{x}^T\boldsymbol{\beta}_1-\mathbf{x}^T\boldsymbol{\beta}_2|}{\sigma}\le B\left(\frac{|\mathbf{x}^T\hat{\boldsymbol{\beta}}_1-\mathbf{x}^T\hat{\boldsymbol{\beta}}_2|}{\sigma}\right)\right\} \\
= & P\left\{B^{-1}(d)\le\frac{|\mathbf{x}^T\hat{\boldsymbol{\beta}}_1-\mathbf{x}^T\hat{\boldsymbol{\beta}}_2|}{\sigma}\right\} \\
= & 1-P\left\{B^{-1}(d)\ge\frac{|\mathbf{x}^T\hat{\boldsymbol{\beta}}_1-\mathbf{x}^T\hat{\boldsymbol{\beta}}_2|}{\sigma}\right\} \\
= & 1-P\left\{\frac{-B^{-1}(d)-d}{w}\le Z\le\frac{B^{-1}(d)-d}{w}\right\} \qquad (4.4)\\
= & 1-\left[\Phi\left(\frac{B^{-1}(d)-d}{w}\right)-\Phi\left(\frac{-B^{-1}(d)-d}{w}\right)\right] \\
= & 1-\alpha,
\end{aligned}$$

where $Z \sim N(0,1)$. Equation (4.4) holds because $(\mathbf{x}^T\hat{\boldsymbol{\beta}}_1 - \mathbf{x}^T\hat{\boldsymbol{\beta}}_2)/\sigma$ distributes as normal distribution with mean $d = (\mathbf{x}^T\boldsymbol{\beta}_1 - \mathbf{x}^T\boldsymbol{\beta}_2)/\sigma$ and variance $w^2 = \mathbf{x}^T\Delta\mathbf{x}$. Then the $1-\alpha$ upper confidence bound for each x can be computed. For each x we have

$$P\left\{\frac{|\mathbf{x}^T\boldsymbol{\beta}_1 - \mathbf{x}^T\boldsymbol{\beta}_2|}{\sigma} \leq B\left(\frac{|\mathbf{x}^T\hat{\boldsymbol{\beta}}_1 - \mathbf{x}^T\hat{\boldsymbol{\beta}}_2|}{\sigma}\right)\right\} = 1-\alpha. \tag{4.5}$$

Then the $1-\alpha$ upper confidence bound on $\max_{x\in[a,b]} |\mathbf{x}^T\boldsymbol{\beta}_1 - \mathbf{x}^T\boldsymbol{\beta}_2|/\sigma$ is constructed using equation (4.5) and theorem 3.1.1:

$$P\left\{\max_{x\in[a,b]}\frac{|\mathbf{x}^T\boldsymbol{\beta}_1 - \mathbf{x}^T\boldsymbol{\beta}_2|}{\sigma} \leq \max_{x\in[a,b]} B\left(\frac{|\mathbf{x}^T\hat{\boldsymbol{\beta}}_1 - \mathbf{x}^T\hat{\boldsymbol{\beta}}_2|}{\sigma}\right)\right\} = 1-\alpha.$$

4.2 Using the Chi-squared Statistic

Liu et al. (2007) proposed the method for constructing the upper confidence bound on the maximum difference between two regression lines for $x \in [a,b]$ using the Chi-squared statistic. Given $\boldsymbol{\delta} = |\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2|/\sigma$, then $C = (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)^T\Delta^{-1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)/\sigma^2$ has a Chi-squared distribution with 2 degrees of freedom and non-centrality parameter $\tau = \boldsymbol{\delta}^T\Delta^{-1}\boldsymbol{\delta}$. Let $U(a)$ be the non-centrality parameter of $\chi^2_{2,U(a)}$ such that a is an $1-\alpha$ upper point of $\chi^2_{2,U(a)}$ random variable. Then, the $1-\alpha$ confidence interval for τ is $[0, U(C)]$ as follows

$$\begin{aligned} P\{\tau \leq U(C)\} &= P\left\{U^{-1}(\tau) \leq C\right\} \\ &= P\left\{U^{-1}(\tau) \leq \chi^2_{2,\tau}\right\} \\ &= P\left\{U^{-1}(\tau) \leq \chi^2_{2,U[U^{-1}(\tau)]}\right\} \\ &= 1-\alpha. \end{aligned}$$

According to Liu et al. (2007), the $1-\alpha$ upper confidence bound for $\max_{x\in R} |(\mathbf{x}^T\boldsymbol{\beta}_1 - \mathbf{x}^T\boldsymbol{\beta}_2)|/\sigma$ for a given R is

$$\max_{x\in R}\left(\sqrt{\mathbf{x}^T\Delta\mathbf{x}}\right) U\left(\frac{(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)^T\Delta^{-1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)}{\sigma^2}\right).$$

Then, the upper confidence bound on the maximum difference between two regression lines, $\max_{x\in[a,b]} |(\mathbf{x}^T\boldsymbol{\beta}_1 - \mathbf{x}^T\boldsymbol{\beta}_2)|/\sigma$, for a given $[a,b]$ is

$$\max_{x\in[a,b]}\left(\sqrt{\mathbf{x}^T\Delta\mathbf{x}}\right) U\left(\frac{(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)^T\Delta^{-1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)}{\sigma^2}\right).$$

4.3 Some Examples for Comparing the Upper Confidence Bounds from Several Methods

Example 4.3.1. The data set of growth rates at different vitamin B dosage levels between male and female chicks from Kleinbaum et al. (2008) is plotted in Figure 4.1. All the methods from the previous sections are used to compute the upper confidence bounds on the maximum different between male and female regression lines at different α, $\alpha = 0.01, 0.05, 0.10, 0.20$. Given $x \in [0.301, 1.505]$ be an area of interested, which is the range of vitamin B dosage levels in the data set. Table 4.1 shows the $1 - \alpha$ upper confidence bounds from several methods at different α. From the table, the upper confidence bounds from method 4.2 at all confidence levels are the minimum bounds for this data set. Moreover, the upper confidence bounds from method 4.1.1 and 4.1.2 are different in the third decimal place, that is for this data set method 4.1.1 and 4.1.2 present close upper confidence bounds.

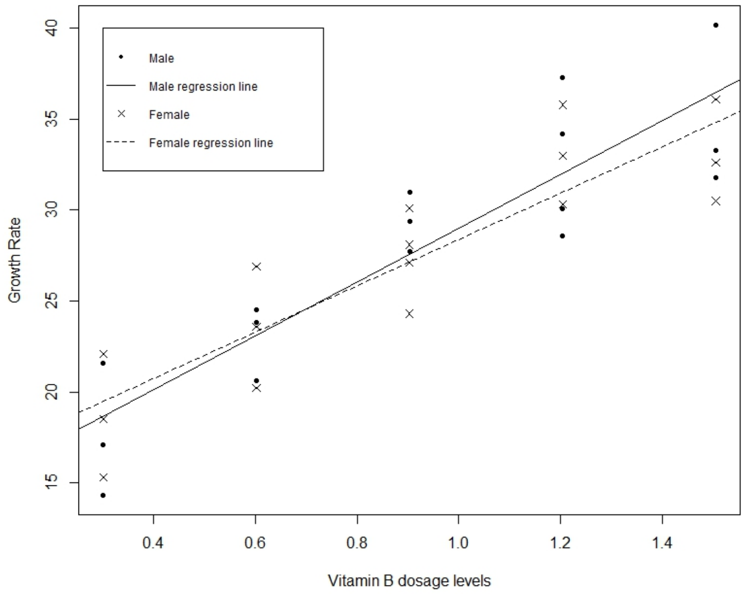

FIGURE 4.1: The data points and fitted regression lines of growth rates at different vitamin B dosage levels

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
Method 4.1.1	1.974	1.554	1.330	1.059
Method 4.1.2	1.973	1.552	1.326	1.048
Method 4.2	1.818	1.339	1.065	0.686

TABLE 4.1: The $1 - \alpha$ upper confidence bounds from several methods at different α for example 4.3.1

Example 4.3.2. The data set of systolic blood pressure at different age between male and female from Kleinbaum et al. (2008) is plotted in Figure 4.2. All the methods from the previous sections are used to compute the upper confidence bounds on the maximum difference between male and female regression lines at different α, $\alpha = 0.01, 0.05, 0.10, 0.20$. Given $x \in [17, 70]$ be an area of interested, which is the range of age in the data set. Table 4.2 shows the $1 - \alpha$ upper confidence bounds from several methods at different α. From the table, the upper confidence bounds from method 4.1.1 and 4.1.2 are not different in the third decimal place, and they are the minimum bounds for this data set.

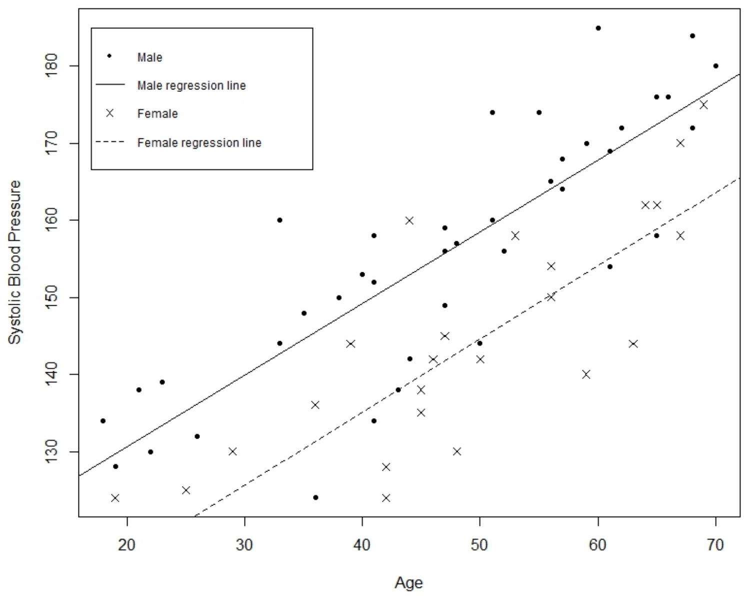

FIGURE 4.2: The data points and fitted regression lines of systolic blood pressure at different age

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
Method 4.1.1 and 4.1.2	2.819	2.459	2.267	2.034
Method 4.2	4.495	4.133	3.940	3.706

TABLE 4.2: The $1 - \alpha$ upper confidence bounds from several methods at different α for example 4.3.2

Example 4.3.3. The data set of drug concentration at different shelf-life time between batch 2 and 3 from Ruberg and Stegeman (1991) is plotted in Figure 4.3. All the methods from the previous sections are used to compute the upper confidence bounds on the maximum different between batch 2 and 3 regression lines at different α, $\alpha = 0.01, 0.05, 0.10, 0.20$. Given $x \in [0.022, 3.077]$ be an area of interested, which is the range of time in the data set. Table 4.3 shows the $1 - \alpha$ upper confidence bounds from several methods at different α. From the table, the upper confidence bounds from method 4.1.1 and 4.1.2 are not different in the third decimal place, and they are the minimum bounds for this data set.

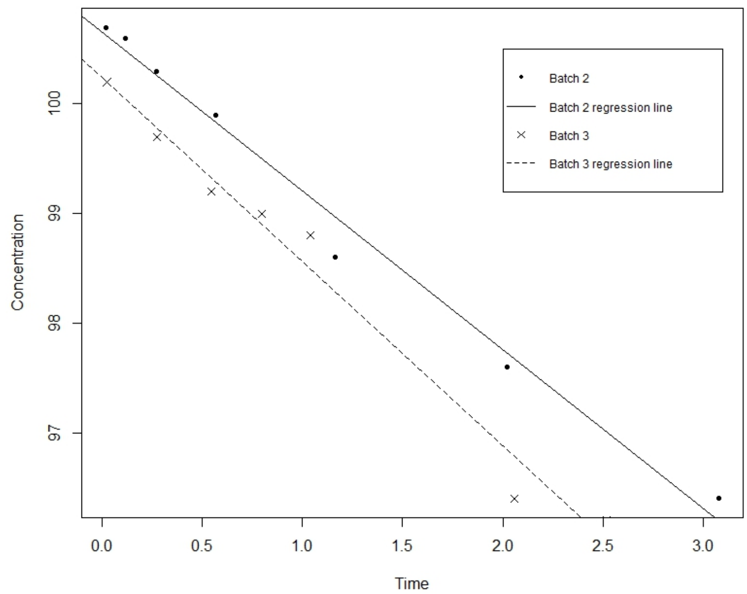

FIGURE 4.3: The data points and fitted regression lines of drug concentration at different shelf-life time

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
Method 4.1.1 and 4.1.2	7.890	7.021	6.558	5.996
Method 4.2	10.040	9.166	8.700	8.135

TABLE 4.3: The $1 - \alpha$ upper confidence bounds from several methods at different α for example 4.3.3

Furthermore, the difference of the area of interested, $x \in [a, b]$, is considered. In figure 4.4, the upper confidence bounds at different areas of interested by method 4.2 for example 4.3.3 at confidence level $\alpha = 0.05$ is plotted. The dash line presents the upper confidence bounds over $x \in [a, b]$ at different $a \in [-2, 4]$, but $b = 3.077$ is fixed. The dot line presents the upper confidence bounds over $x \in [a, b]$ at different $b \in [-2, 4]$, but $a = 0.022$ is fixed. The graph shows that the upper confidence bound is changed when an area of interested, $[a, b]$, is changed. Consider the horizontal part of dash line, the upper confidence bound is not changed when a is changed. It can be implied that at this area, the upper confidence bound is depended on b which is fixed at 3.077. The explanation of the horizontal part of dot line is the same as dash line, the upper confidence bound is not changed when b is changed. It can be implied that at this area, the upper confidence bound is depended on a which is fixed at 0.022.

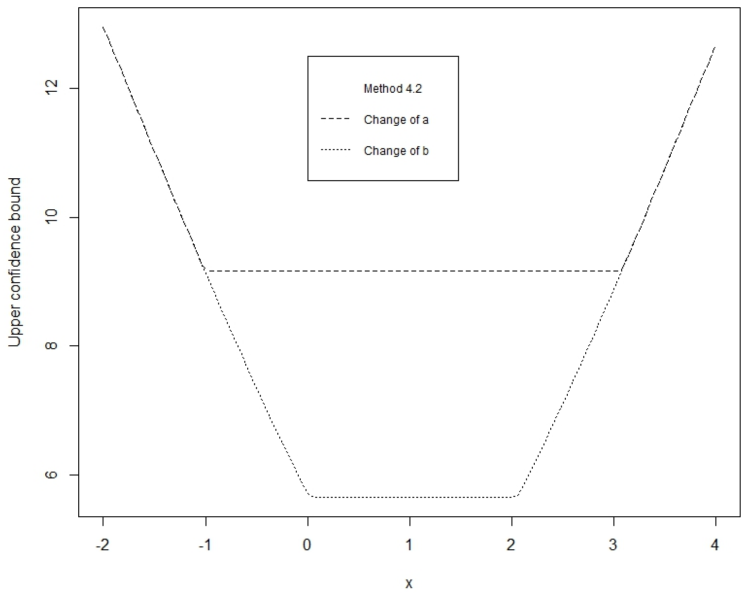

FIGURE 4.4: The upper confidence bounds at different areas of interested by method 4.2 for example 4.3.3

In figure 4.5, the upper confidence bounds at different areas of interested, $x \in [a, b]$, by method 4.1.1 and 4.1.2 for example 4.3.3 at confidence level $\alpha = 0.05$ are plotted. The dash line presents the upper confidence bounds over $x \in [a, b]$ at different $a \in [-5, 4]$, but $b = 3.077$ is fixed. The dot curve and solid curve present the upper confidence bounds over $x \in [a, b]$ at different $b \in [-5, 4]$, but $a = 0.022$ is fixed from method 4.1.1 and 4.1.2 respectively. The graph shows that the upper confidence bounds is changed when an area of interested, $[a, b]$, is changed. Consider the horizontal part of dash line, the upper confidence bound is not changed when a is changed. It can be implied that at this area, the upper confidence bound is depended on b which is fixed at 3.077. The explanation of the horizontal part of dot line and solid line are the same as dash line, the upper confidence bound is not changed when b is changed. It can be implied that at this area of the graph, the upper confidence bound is depended on a which is fixed at 0.022.

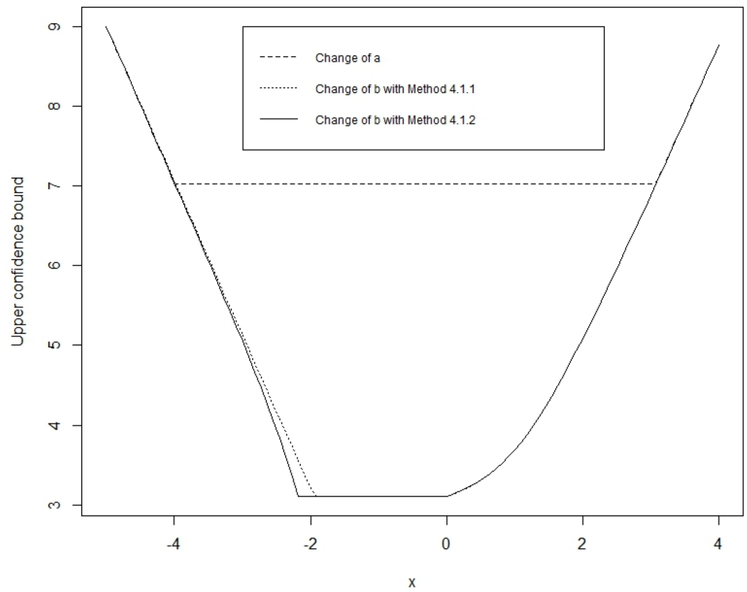

FIGURE 4.5: The upper confidence bounds at different areas of interested by method 4.1.1 and 4.1.2 for example 4.3.3

4.4 Conclusion

In this chapter, we propose the construction of upper confidence bounds on the maximum difference between two regression lines by two methods: by the theorem 3.1.1 from Liu et al. (2009) and by the Chi-squared statistic. The construction using the first method depends on several upper confidence bounds that can be computed by several methods. This method cannot compute the upper confidence bound of $\max_{x\in[a,b]} \frac{|(\beta_{01}+\beta_{11}x)-(\beta_{02}+\beta_{12}x)|}{\sigma}$ if σ is unknown. However, it can be use for calculating the upper confidence bound of

$\max_{x\in[a,b]} |(\beta_{01} + \beta_{11}x) - (\beta_{02} + \beta_{12}x)|$ whether σ is known or unknown. In conclusion, the three examples show that the minimum of the upper confidence bound for each example is computed from different methods. In the first example, method 4.2 provides the minimum upper confidence bound. For the last two examples the minimum upper confidence bounds are calculated by method 4.1.1 and 4.1.2. Hence, we cannot conclude that which method is the best method which provides the minimum upper confidence bound. Moreover, all three methods depend on the area of interested, $[a, b]$. The change of this area has an effect on the upper confidence bound as in figure 4.4 and 4.5. Then, we are moving to the classification problem in the next Chapter.

Chapter 5

Classification Methods

In this Chapter, a confidence set is applied for classification problem. To classify a patient in term of whether he or she has a disease, based on some diagnostic measurements, is an important problem of statistical classification. We propose the construction of a confidence set as an alternative method for this problem. This method also use the theorem 2.2.1 in Chapter 2 for constructing a confidence set. Five classification methods for classification into two classes and three classes have been studied. They are

- new confidence set method
- classification tree
- logistic regression
- Bayesian method
- support vector machine.

We have some measurements among two groups of patient, normal and disease groups. Then a new patient is coming with those measurements have been measured. We would like to classify this new patient whether he or she has a disease. The methods in this chapter will help us to deal with this problem. Moreover, The methods are illustrated specifically with the well known Iris data and some other examples.

5.1 Confidence set method

We have some measurements among several groups of interested. Then these measurements for a new case has been measured. We would like to classify this new case to its true group, for example, to classify new patient whether he or she has a disease. The new confidence set method constructs a confidence set for the true class for a new case by inverting the acceptance sets. The result from this method can be one-class, multi-classes or no-class. The advantage of this method is that the probability of correct classification is not less than 1-alpha.

One measurement for two classes

Let X_{ij} be a measurement for the j^{th} patient in group i for $i = 0, 1$, $j = 1, ..., n_i$ and $X_{i1}, ..., X_{in_i} \sim N(\mu_i, \sigma_i^2)$. We would like to classify a new patient with that measurement, Y, to the true group, θ. Let $A(\theta_0)$ be an acceptance set of size α for testing $H_0 : \theta = 0$, the true class of a new patient is the normal or non-disease group, and $A(\theta_1)$ be an acceptance set of size α for testing $H_0 : \theta = 1$, the true class of a new patient is the disease group. The $A(\theta_0)$ is given by

$$A(\theta_0) = \left\{ Y : \frac{|Y - \bar{X}_0|}{\hat{\sigma}_0 \sqrt{1 + 1/n_0}} < t_{n_0 - 1, \alpha/2} \right\},$$

where $\bar{X}_0 = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{0j}$ and $\hat{\sigma}_0 = \sqrt{\frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (X_{0j} - \bar{X}_0)^2}$. The $A(\theta_1)$ is given by

$$A(\theta_1) = \left\{ Y : \frac{|Y - \bar{X}_1|}{\hat{\sigma}_1 \sqrt{1 + 1/n_1}} < t_{n_1 - 1, \alpha/2} \right\},$$

where $\bar{X}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} X_{1j}$ and $\hat{\sigma}_1 = \sqrt{\frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2}$.

Proof. The sizes of the test for each case of θ are

$$\begin{aligned}
P_{\theta_0}\{Y \in A(\theta_0)\} &= P_{\theta_0} \left\{ \frac{|Y - \bar{X}_0|}{\hat{\sigma}_0 \sqrt{1 + 1/n_0}} < t_{n_0 - 1, \alpha/2} \right\} \\
&= P_{\theta_0} \left\{ \frac{|\frac{Y - \mu_0}{\sigma_0} - \frac{\bar{X}_0 - \mu_0}{\sigma_0}| / \sqrt{1 + 1/n_0}}{\sqrt{\hat{\sigma}_0^2 / \sigma_0^2}} < t_{n_0 - 1, \alpha/2} \right\} \\
&= P \left\{ \frac{|Z|}{\sqrt{w_0}} < t_{n_0 - 1, \alpha/2} \right\} \\
&= P \left\{ |T| < t_{n_0 - 1, \alpha/2} \right\} \\
&= 1 - \alpha
\end{aligned}$$

and

$$\begin{aligned} P_{\theta_1}\{Y \in A(\theta_1)\} &= P_{\theta_1}\left\{\frac{|Y-\bar{X}_1|}{\hat{\sigma_1}\sqrt{1+1/n_1}} < t_{n_1-1,\alpha/2}\right\} \\ &= P_{\theta_1}\left\{\frac{|\frac{Y-\mu_1}{\sigma_1} - \frac{\bar{X}_1-\mu_1}{\sigma_1}|/\sqrt{1+1/n_1}}{\sqrt{\hat{\sigma}_1^2/\sigma_1^2}} < t_{n_1-1,\alpha/2}\right\} \\ &= P\left\{\frac{|Z|}{\sqrt{w_1}} < t_{n_1-1,\alpha/2}\right\} \\ &= P\left\{|T| < t_{n_1-1,\alpha/2}\right\} \\ &= 1-\alpha \end{aligned}$$

where $Z \sim N(0,1)$, $w_0 \sim \chi^2_{n_0-1}/(n_0-1)$ and $w_1 \sim \chi^2_{n_1-1}/(n_1-1)$. Hence, the sizes of both tests are α. □

Using theorem 2.2.1, we can construct a $C(Y)$ as

$$C(Y) = \{\theta \in \Theta : Y \in A(\theta)\}, \text{ with } \Theta = \{0,1\}.$$

Then this is a confidence set for the true class, θ, of confidence level $1-\alpha$.

Multiple measurements for multiple classes

Let $\boldsymbol{X_{ij}}$ be the k measurements vector for the j^{th} patient in group i for $i = 1, 2, ..., p$, $j = 1, ..., n_i$ and $\boldsymbol{X_{i1}, ..., X_{in_i}} \sim N(\boldsymbol{\mu_i}, \Sigma_i)$. We would like to classify a new patient with the measurements $\boldsymbol{Y}$ to the true group, θ. Let $A(\theta_i)$ be an acceptance set of size α for testing $H_0 : \theta = i$. The $A(\theta_i)$ is given by

$$A(\theta_i) = \left\{\boldsymbol{Y} : \frac{n_i - p}{p(1+1/n_i)}(\boldsymbol{Y} - \boldsymbol{\bar{X}_i})^T A_i^{-1}(\boldsymbol{Y} - \boldsymbol{\bar{X}_i}) < f^{1-\alpha}_{p,n_i-p}\right\},$$

where $A_i^{-1} = \frac{1}{n_i-1}\sum_{a=1}^{n_i}(\boldsymbol{X_a} - \boldsymbol{\bar{X}_i})(\boldsymbol{X_a} - \boldsymbol{\bar{X}_i})^T$.

Proof. According to the following theorem from Anderson (2003),

Theorem 5.1.1. let $\boldsymbol{V}$ is distributed as $N(\boldsymbol{\mu}, \Sigma)$, nS is distributed as $\sum_{a=1}^n \boldsymbol{Z_a Z_a}^T$ with $\boldsymbol{Z_1}, ..., \boldsymbol{Z_n}$ are independent and identically distributed as $N(\boldsymbol{0}, \Sigma)$ and $T^2 = \boldsymbol{V}' S^{-1} \boldsymbol{V}$. Then,

$$\left(\frac{T^2}{n}\right)\frac{[n-p+1]}{p} \sim F_{p,n-p+1;\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}},$$

non-central F distribution with $\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$ non-central parameter. If $\boldsymbol{\mu} = \boldsymbol{0}$, the distribution will be central F.

The sizes of the test for each case of θ is

$$\begin{aligned}
A(\theta_i) &= \left\{\boldsymbol{Y} : \frac{n_i - p}{p(1+1/n_i)}(\boldsymbol{Y} - \bar{\boldsymbol{X}}_{\boldsymbol{i}})^T A_i^{-1}(\boldsymbol{Y} - \bar{\boldsymbol{X}}_{\boldsymbol{i}}) < f_{p,n_i-p}^{1-\alpha}\right\} \\
&= \left\{\boldsymbol{Y} : \frac{n_i - p}{p}\left(\frac{\boldsymbol{Y} - \bar{\boldsymbol{X}}_{\boldsymbol{i}}}{\sqrt{(1+1/n_i)}}\right)^T A_i^{-1}\left(\frac{\boldsymbol{Y} - \bar{\boldsymbol{X}}_{\boldsymbol{i}}}{\sqrt{(1+1/n_i)}}\right) < f_{p,n_i-p}^{1-\alpha}\right\} \\
&= \left\{\boldsymbol{Y} : \frac{n_i - p}{p(n_i-1)}\left(\frac{\boldsymbol{Y} - \bar{\boldsymbol{X}}_{\boldsymbol{i}}}{\sqrt{(1+1/n_i)}}\right)^T \left(\frac{A_i}{n_i-1}\right)^{-1}\left(\frac{\boldsymbol{Y} - \bar{\boldsymbol{X}}_{\boldsymbol{i}}}{\sqrt{(1+1/n_i)}}\right) < f_{p,n_i-p}^{1-\alpha}\right\} \\
&= \left\{\boldsymbol{Y} : \frac{n_i - p}{p}\frac{T^2}{(n_i-1)} < f_{p,n_i-p}^{1-\alpha}\right\} \\
&= 1-\alpha,
\end{aligned}$$

where $n = n_i - 1$, $nS = A_i$ and using theorem 5.1.1. Hence, the size of the test is α. □

Using theorem 2.2.1, we can construct a $C(Y)$ as

$$C(Y) = \{\theta \in \Theta : Y \in A(\theta)\}, \quad \Theta = \{1, 2, ..., p\}.$$

Then this is a confidence set for the true class, θ, of confidence level $1-\alpha$.

5.2 Classification Tree

Classification tree is a simple method to use as primary data inspection (Crawley, 2003). Moreover, this method presents clear diagram result which shows some variables interaction. This method can be divided into three steps (Martinez and Martinez, 2008):

- Growing the tree,
- Pruning the tree,
- Choosing the tree.

Growing the tree

Binary recursive partitioning technique is used in this step. Start with the root of the tree, which is all cases are in one node. Then, the best split is investigated by the most decreasing of the node impurity $I(t)$. The impurity of node t is calculated as

$$I(t) = \prod_{i=1}^{p} P(\theta = i|t),$$

where $P(\theta = i|t)$ is the probability of belonging to class i, $\theta = i$, in node t,

$$P(\theta = i|t) = \frac{n_{ti}}{n_t},$$

where n_{ti} is the number of cases in node t which belong to class i, and n_t is the total number of cases in node t. Then, repeat split until all terminal nodes has only one class ($n_{ti} = n_t$). At the end of this step, we will get the largest tree, $T_{k=1}$ or T_{max}.

Pruning the tree

When the tree is growing until all terminal nodes have only one class , it might be over fitted. Then, we need this step to reduce the complexity. Pruning the tree bases on the risk, $R(t)$, which is composed of misclassification rate and a cost for complexity. Start with the largest tree, $T_{k=1}$ or T_{max}, $g_k(t)$ of subtree T_k is calculated for all internal nodes t as

$$g_k(t) = \frac{R(t) - R(T_{tk})}{|\tilde{T}_t| - 1}, \tag{5.1}$$

where $R(T_{tk})$ is the summation of the risk of terminal node in branch T_t of node t in subtree T_k and $|\tilde{T}_t|$ is the number of terminal node in internal node t. The risk of node t, $R(t)$, the probability of misclassification for node t, $r(t)$ and the probability of belonging to node t, P(t) are calculated by

$$\begin{aligned} R(t) &= r(t)P(t) \\ r(t) &= 1 - \max_{\theta} P(\theta|t) \\ P(t) &= \sum_{\theta=1}^{p} P(\theta, t) \end{aligned}$$

Then, the weakest link, t_k^* in subtree T_k is investigated using equation 5.1

$$g_k(t^*) = \min{}_t\{g_k(t)\}.$$

The branch that has the weakest link is pruned off as

$$T_{k+1} = T_k - T_{t_k^*}.$$

Repeat this step until only the root node is left.

Choosing the tree

After we grow the tree to the largest tree, and prune the tree to smallest root. We needs to choose the best tree. There are several methods for choosing the tree (Martinez and Martinez, 2008). An independent test sample method has been studied. Firstly, randomly partition the data into L_{tree}, L_{case}. Grow a largest tree, T_1, using L_{tree}. Then, prune the tree to set the sequence of subtree, T_k. take L_{case} to each subtree and calculate the estimate of the misclassification rate, $\hat{M}(T_k)$, as

$$\hat{M}(T_k) = \frac{1}{n^{case}} \sum_{ij} n_{ij}^{case},$$

where, n^{case} is number of observation in L_{case}, and n_{ij}^{case} is the number of observation belong to group j, $\theta = j$, but classified to group i. Find the minimum misclassification rate as

$$\hat{M}_{min} = \min_k\{\hat{M}(T_k)\}.$$

Then, find the smallest tree such that misclassification rate is less than $\hat{M}_{min} + \hat{SE}(\hat{M}_{min})$,

$$\max(\hat{M}(T_k^*)) \leq \hat{M}_{min} + \hat{SE}(\hat{M}_{min}).$$

The T_k^* tree is the best tree. Then this tree is to be used to classify a new case.

5.3 Logistic Regression

Logistic Regression is a classification method using conditional probability for predicting a new data. The dependent variable is the probability that the data is belong to a particular group that is constrained from zero to one. The advantage of this method is the odds ratio for the model predictor can be estimated (Hilbe, 2009; Hosmer et al., 2013).

One measurement for two classes

Let X_{ij} be a measurement for the j^{th} case in group i for $i = 0, 1$, $j = 1, ..., n_i$. We would like to classify a new case with that measurement to the true group, θ. The binary logistic regression is used for classify a new case. The logit function is the link function of this generalized linear regression model as:

$$\begin{aligned} log\frac{P(\theta = 1|X)}{1 - P(\theta = 1|X)} &= \beta_0 + \beta_1 X, \\ P(\theta = 1|X) &= \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}}. \end{aligned}$$

Then the cut point probability is applied for classifying a new case.

Multiple measurements for multiple classes

Let $\boldsymbol{X}_{ij}$ be a vector of measurements for the j^{th} case in group i for $i = 1, 2, ..., p$, $j = 1, ..., n_i$. We would like to classify a new case with the measurements $\boldsymbol{Y}$ to the true group, $\theta \in \{\mathbf{1}, 2, ..., p\}$. According to Faraway (2006), Let P_{ij} is the probability of case j belongs to class i as

$$\begin{aligned} P_{ij} &= P(\theta_j = i), \\ \text{so,} \sum_{i=1}^{p} P_{ij} &= 1, \end{aligned} \tag{5.2}$$

and Y_{ij} be the number of observations falling into categories i for group or individual j (for individual j in this study, $Y_{ij} = 1$ or 0 and $n_j = \sum_{i=1}^{p} Y_{ij} = 1$). The Y_{ij}, conditional on n_j follow a multinomial distribution:

$$P(Y_{1j} = y_{1j},\ Y_{2j} = y_{2j},\ ...,\ Y_{pj} = y_{pj}) = \frac{n_j!}{y_{1j}!y_{2j}!...y_{pj}!} P_{1j}^{y_{1j}} P_{2j}^{y_{2j}} ...\ P_{pj}^{y_{pj}}$$

The logit function is the link function of $p-1$ models as:

$$\begin{aligned}
\eta_{ij} &= log\frac{P_{ij}}{P_{1j}} = \boldsymbol{\beta_i X_j}, \quad i = 2, 3, ..., p \\
P_{ij} &= \frac{exp(\eta_{ij})}{1+\sum_{i=2}^{p} exp(\eta_{ij})}, \quad i = 2, 3, ..., p
\end{aligned}$$

where $\boldsymbol{\beta_i}$ is the linear coordinator parameters for group i of the predictors. According to equation (5.2) P_{1j} can be computed as

$$\begin{aligned}
P_{1j} &= 1 - \sum_{i=2}^{p} P_{ij} \\
&= \frac{1}{1+\sum_{i=2}^{p} exp(\eta_{ij})}.
\end{aligned}$$

After parameter estimation procedure, the models can be used for predicting a new case with the probability of belonging to each class. Then these p models will be used for classifying a new case.

5.4 Bayesian method

The Bayesian method is used for classifying a new case in this section. For this method, the posterior probabilities is used to classify a new case. The advantage is the decision boundary from the Bayesian method is computed from the boundary that gives the minimum missclassification error (Martinez and Martinez, 2008).

One measurement for two classes

Let X_{ij} be a measurement for the j^{th} case in group i for $i = 0, 1,\ j = 1, ..., n_i$ and $X_{i1}, ..., X_{in_i} \sim N(\mu_i, \sigma_i^2)$. We would like to classify a new case with that measurement to the true group, θ. Given the posterior probability of $P(\theta|X)$ for $\theta = 0$ and $\theta = 1$ are

$$\begin{aligned} P(\theta = 0|X) &= \frac{P(X|\theta = 0)P(\theta = 0)}{P(X)} \\ &= \frac{\Phi(\frac{X-\hat{\mu}_0}{\hat{\sigma}_0})(n_0/N)}{P(X)} \\ \text{and} \quad P(\theta = 1|X) &= \frac{P(X|\theta = 1)P(\theta = 1)}{P(X)} \\ &= \frac{\Phi(\frac{x-\hat{\mu}_1}{\hat{\sigma}_1})(n_1/N)}{P(X)}. \end{aligned}$$

Using this method, a new case will be classified to class i if and only if

$$P(\theta_i|X) > P(\theta_j|X); \quad i, j = 0, 1, \quad i \neq j.$$

The decision boundary can be computed by finding X^*, which results in equal posterior probability of two classes as equation below,

$$P(\theta = 0|X = X^*) = P(\theta = 1|X = X^*).$$

Multiple measurements for multiple classes

Let $\boldsymbol{X}_{ij}$ be a vector of measurements for the j^{th} case in group i for $i = 1, 2, ..., p$, $j = 1, ..., n_i$ and $\boldsymbol{X_{i1}, ..., X_{in_i}} \sim N(\boldsymbol{\mu_i}, \Sigma_i)$. We would like to classify a new case with the measurements $\boldsymbol{Y}$ to the true group, θ. Given the posterior probability, $P(\theta|\boldsymbol{X})$, for $\theta = i, \quad i = 1, 2, ..., p$ are

$$P(\theta = i|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|\theta = i)P(\theta = i)}{P(\boldsymbol{X})}$$

Using this method, a new case with the measurements $\boldsymbol{Y}$ will be classified to class i if and only if

$$P(\theta_i|\boldsymbol{Y}) > P(\theta_j|\boldsymbol{Y}); \quad i, j = 1, ..., p, \quad i \neq j.$$

5.5 Support vector machine

Support vector machines technique was initially developed for binary classification using linear separator. The optimum hyperplane between two classes is used for class separation by maximized margin. However, for higher dimensional and the non-linear separable data, kernel techniques are used for adjusting the data to be appropriated for linear separator or classifier, l (Cortes and Vapnik, 1993).

In this research, multiple binary linear classifiers, $l's$, using the one-versus-rest classifier technique is the method for classifying a new case. Given we have a p classes data set. Then, the number of classifiers is p, which are i^{th} class versus the rest;

$$\begin{aligned} l_1 &= (1^{st}\text{class})\ vs.\ (\text{ all classes except } 1^{st}\text{class}) \\ l_2 &= (2^{nd}\text{class})\ vs.\ (\text{ all classes except } 2^{nd}\text{class}) \\ &\vdots \\ l_p &= (p^{th}\text{class})\ vs.\ (\text{ all classes except } i^{th}\text{class}). \end{aligned}$$

according to Flach (2012), the margin of linear classifier is the distance between the decision boundary and the closest examples. These nearest examples are called support vectors. The support vector machine boundary is a linear combination of the support vectors. Let margin be defined as $c(X)\hat{s}(X)$, where $c(X) = +1$ if example is in the interesting class, and $c(X) = -1$ if example is not in the interesting class. This margin is positive when the prediction is correct, and negative for incorrect prediction. $\hat{s}(X)$ is indicating score to determine how likely the case is in the interesting class. Given $\hat{s}(X) = W \cdot X - t$, where t is the adjustable decision threshold. Then, the example in the interesting class will have the margin as $W \cdot X - t > 0$, and the example in the non-interesting class will have the margin as $-(W \cdot X - t) > 0$. Let m^+ be the smallest margin of any example in the interesting class, and m^- be the smallest margin of any example in the non-interesting class. Then, we would like to have the maximum of the summation of m^+ and m^-. The t is re-adjusted for equal m^+ and m^-.

The nearest examples to the decision boundary are supports vectors, which the distance between the decision boundary and the nearest examples along W is m. we define the margin as $m/||W||$. We are free to re-adjust $t, ||W||$ and m, then assume $m = 1$. The margin is maximised using the method of Lagrange multipliers to minimise $||W||$ as

$$W^*, t^* = \underset{W,t}{\text{argmin}} \frac{1}{2}||W||^2$$

with constrain optimisation to

$$C(X_i)(W \cdot X_i - t) \geq 1,\ 1 \leq i \leq n.$$

If our data is not linear separable data, the constrain $W \cdot X_i - t \geq 1$ will not be satisfied. Then the slack variable ϵ_i is add to each samples to allow them to be in the margin or in the wrong side of decision boundary. This term is called margin error. Then, the constrains are change to

$$W \cdot X_i - t \geq 1 - \epsilon_i,\ \epsilon_i > 0,\ 1 \leq i \leq n$$

and the sum of all slack variables is add to the objective function to minimise as

$$W^*, t^*, \epsilon_i^* = \underset{W,t,\epsilon_i}{\operatorname{argmin}} \frac{1}{2}||W||^2 + C\sum_{i=1}^{n} \epsilon_i,$$

where C is user defined parameter: higher C means higher penalty for margin error, and lower C will have more margin error to get large margin. This call soft margin optimisation problem.

Moreover, for higher dimensional and the non-linear separable data, kernel techniques are used for adjusting the data to be appropriated for hyperplane classifier.

5.6 Examples

Three real data examples: the data of cirrhosis patients from a hospital, the well known iris data and seeds data are used for illustrating and comparing those five classification methods. The total misclassification error and sensitivity (true positive rate) are used for comparing the methods.

5.6.1 Evaluation Criteria

Total misclassification error (TME)

Total misclassification error (TME) is the number of total incorrect classified divided by the total number of cases as

$$\text{TME} = \frac{\text{number of } \textbf{\textit{incorrect}} \text{ classified}}{\sum_{i=1}^{p} n_i},$$

where n_i is the total number of cases in class i and $\sum_{i=1}^{p} n_i$ is the total number of cases.

Sensitivity or true positive rate (SEN)

Sensitivity or true positive rate (SEN) is the number of correct classified in specific class divided by the total number of cases in that class as

$$\text{SEN}_i = \frac{\text{number of } \textbf{\textit{correct}} \text{ classified in class } i}{n_i}.$$

5.6.2 Cirrhosis data

The cirrhosis data set of three classes: normal, fibrosis or cirrhosis, with 14 measurements on blood samples (A-M,U), is from a hospital. The data of 12,842 cases is randomly divided in to two sets, S_1, S_2. The first set, S_1, with 5,000 cases is for constructing the five classification methods, and the second set, S_2 with 7,842 cases acts as new cases for evaluation of the methods.

Confidence set method

For confidence set method, using theorem 5.1.1 and theorem 2.2.1 with the S_1 data set to classify new 7,842 cases in S_2 by `R` program shows the result in the Table 5.1. The result from this method will be a confidence set for the true class, which can be $\{0\}, \{1\}, \{2\}, \{0,1\}, \{1,2\}, \{0,2\}, \{0,1,2\}$ and ϕ, empty set. Some new cases, which fall in ϕ is reclassified by Bayesian method for this study. The reclassification presents in the Table 5.2 does not decrease the probability of correct classification, which is not less than $1 - \alpha$.

Table 5.1: result from the new confidence set method on cirrhosis data

Observed Class	{0}	{1}	{2}	{0,1}	{1,2}	{0,2}	{0,1,2}	ϕ
Normal (0)	145	0	288	0	30	1179	4194	180
Fibrosis (1)	0	0	8	0	5	7	225	6
Cirrhosis (2)	3	0	528	0	74	118	681	171

Table 5.2: result from the new confidence set method with reclassification

Observed Class	{0}	{1}	{2}	{0,1}	{1,2}	{0,2}	{0,1,2}	ϕ
Normal (0)	252	0	361	0	30	1179	4194	0
Fibrosis (1)	0	0	14	0	5	7	225	0
Cirrhosis (2)	32	0	670	0	74	118	681	0

Classification tree

The `tree()` function of package `tree` in `R` program was used to construct tree model using 5,000 cases in S_1. The classification tree diagram composed of only 4 measurements: A, C, E, F as in figure 5.1. Then, this tree was applied to classify 7,842 new cases in S_2. The result presents in Table 5.3.

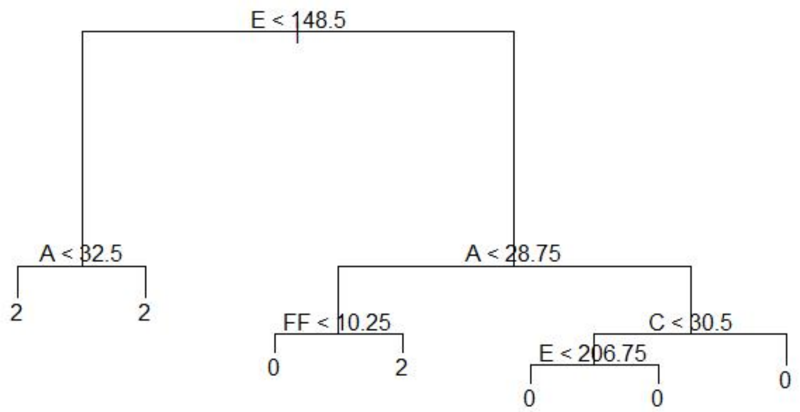

Figure 5.1: classification tree diagram for classifying a new case for cirrhosis data

Table 5.3: result from classification tree on cirrhosis data

Observed	Classified class		
Class	Normal(0)	Fibrosis(1)	Cirrhosis(2)
Normal (0)	5569	0	447
Fibrosis (1)	213	0	38
Cirrhosis (2)	568	0	1007

Logistic regression

The `multinom()` function of package `nnet` in `R` program was used to construct multiple logistic regression model using 5,000 cases in S_1. The multiple logistic regression model composed of 13 measurements: A, B, C, D, E, F, G, I, J, K, L, M, U. Only one measurement, H, was removed by backward elimination method. Then, this model was applied to classify 7,842 new cases in S_2. The result presents in Table 5.4.

Table 5.4: result from multiple logistic regression on cirrhosis data

Observed	Classified class		
Class	Normal(0)	Fibrosis(1)	Cirrhosis(2)
Normal (0)	5830	0	186
Fibrosis (1)	231	1	19
Cirrhosis (2)	706	5	864

Bayesian method

The `naiveBayes()` function of package `e1071` in `R` program was used to construct Bayes' classifier using 5,000 cases in S_1. Backward elimination was the method for variable selection with the smallest total misclassification error. The Bayes' classifier composed of 11 measurements: A, B, D, E, F, G, I, J, K, M, U. Three measurements, L, C and H, was removed. Then, this model was applied to classify 7,842 new cases in S_2. The result presents in Table 5.5.

Table 5.5: result from Bayesian method on cirrhosis data

Observed	Classified class		
Class	Normal(0)	Fibrosis(1)	Cirrhosis(2)
Normal (0)	5729	29	258
Fibrosis (1)	236	2	13
Cirrhosis (2)	850	15	710

Support vector machine

The `svm()` function of package `e1071` in `R` program was used to construct support vector machine using 5,000 cases in S_1. Transform data using the kernel function was selected with the smallest total misclassification error. The support vector machine with radial kernel function and $\gamma = 0.2$ was the best model for this method. Then, this model was applied to classify 7,842 new cases in S_2. The result presents in Table 5.6.

TABLE 5.6: result from support vector machine on cirrhosis data

Observed Class	Classified class		
	Normal(0)	Fibrosis(1)	Cirrhosis(2)
Normal (0)	5868	1	147
Fibrosis (1)	209	8	34
Cirrhosis (2)	580	3	992

Comparison of classification methods

TABLE 5.7: Total misclassification error (TME) and sensitivity for cirrhosis data

	TME	Sensitivity	
		Fibrosis	Cirrhosis
Multinomial logistic regression	0.1169	0.0040	0.5486
Classification tree	0.1343	0.0000	0.6394
Bayesian method	0.1816	0.0199	0.4406
Support vector machine	0.0986	0.0319	0.6298
Confidence set method	0.0597	0.0000	0.3352
Confidence set method (reclassification)	0.0566	0.0000	0.4254

Total misclassification error (TME) and sensitivity from five classification methods for cirrhosis data is present in Table 5.7. For this data, the new confidence set method provides the smallest total misclassification error, and also controls the probability of correct classified. Support vactor machine technique gives the highest sensitivity for both fibrosis and cirrhosis classes. The new confidence set method shows the lowest sensitivity because the result of multi-classes is not include in correct classified for some cases, which cannot be clearly classified into one single class. As a result, the probability of correct classified was controlled.

5.6.3 Iris data

The Fisher's Iris data set is a classified multivariate data of three species (three classes) of Iris: *Iris setosa*, *Iris virginica* and *Iris versicolor* 50 samples each, with four measurements: the length and the width of the sepals and petals, in centimetres (Fisher, 1936). Figure 5.2 shows four boxplot of four measurements for each class. Five classification

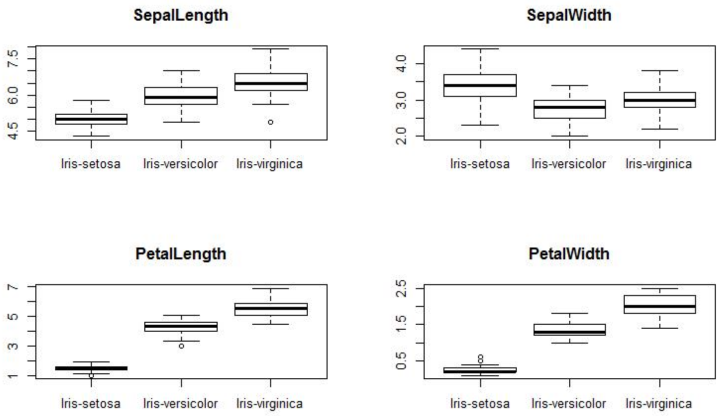

FIGURE 5.2: The boxplot of four measurements for each class of Iris data

methods were applied to classify a new case by using leave one out method for evaluation. Leave one out method is taken one case out for testing as a new case, and the rest for constructing the five classification methods.

Confidence set method

TABLE 5.8: the number of result from the new confidence set method

	{0}	{1}	{2}	{0,1}	{1,2}	{0,2}	{0,1,2}	ϕ
Iris-setosa(0)	45	0	0	0	0	0	0	5
Iris-vesicolor(1)	0	38	0	0	10	0	0	2
Iris-virginica (2)	0	0	40	0	7	0	0	3

Table 5.8 presents the number of result from the new confidence set method. The result from this method will be a confidence set for the true class, which can be $\{0\}, \{1\}, \{2\}, \{0, 1\}, \{1, 2\}, \{0, 2\}, \{0, 1, 2\}$ and ϕ, empty set. Some new cases, which fall in ϕ is reclassified by Bayesian method for this study. The reclassification presents in the Table 5.9 does not decrease the probability of correct classification, which is not less than $1 - \alpha$.

Table 5.9: the number of result from the new confidence set method with reclassified by Bayesian method

	{1}	{2}	{3}	{1,2}	{2,3}	{1,3}	{1,2,3}	ϕ
Iris-setosa(1)	50	0	0	0	0	0	0	0
Iris-vesicolor(2)	0	40	0	0	10	0	0	0
Iris-virginica (3)	0	0	45	0	7	0	0	0

Classification tree

The `tree()` function of package `tree` in `R` program was used to construct tree model using 149 cases (one case was taken out for test). The classification tree diagram composed of two measurements: the length and the width of petals as in figure 5.3. Then, this tree was applied to a cases that was taken out. We repeated for all 150 cases. The result presents in Table 5.10.

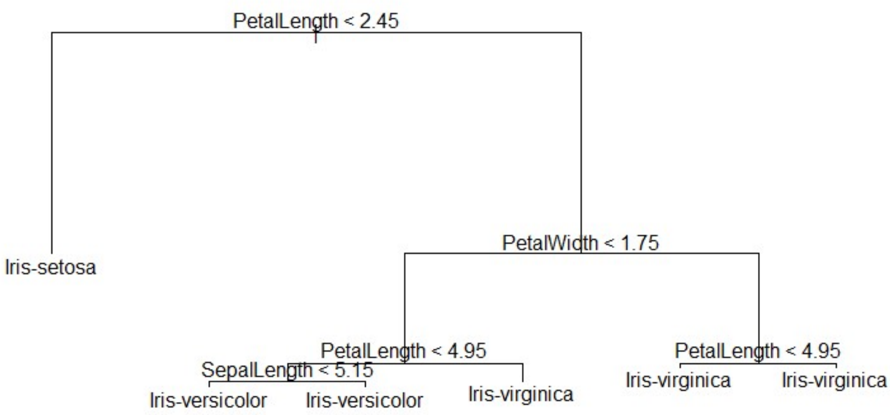

Figure 5.3: classification tree diagram for classifying a new case for iris data

Table 5.10: result from classification tree on iris data

Observed Class	Classified class		
	Iris-setosa	Iris-vesicolor	Iris-virginica
Iris-setosa	50	0	0
Iris-vesicolor	0	47	3
Iris-virginica	0	2	48

Logistic regression

The `multinom()` function of package `nnet` in `R` program was used to construct multiple logistic regression model using 149 cases (one case was taken out for test). The multiple logistic regression model composed of three measurements: the width of sepals, the length and the width of petals. Only one measurement, the length of sepals, was removed by backward elimination method. Then, this model was applied to a cases that was taken out. We repeated for all 150 cases. The result presents in Table 5.11, which is the same result as using classification tree.

TABLE 5.11: result from multiple logistic regression on iris data

Observed	Classified class		
Class	Iris-setosa	Iris-vesicolor	Iris-virginica
Iris-setosa	50	0	0
Iris-vesicolor	0	47	3
Iris-virginica	0	2	48

Bayesian method

The `naiveBayes()` function of package `e1071` in `R` program was used to construct Bayes' classifier using 149 cases (one case was taken out for test). Backward elimination was the method for variable selection with the smallest total misclassification error. The Bayes' classifier composed of two measurements: the length and the width of petals. Two measurements, the length and the width of sepals, was removed. Then, this model was applied to classify a cases that was taken out. We repeated for all 150 cases. The result presents in Table 5.12.

TABLE 5.12: result from Bayesian method on iris data

Observed	Classified class		
Class	Iris-setosa	Iris-vesicolor	Iris-virginica
Iris-setosa	50	0	0
Iris-vesicolor	0	47	3
Iris-virginica	0	1	49

Support vector machine

The `svm()` function of package `e1071` in `R` program was used to construct support vector machine using 149 cases (one case was taken out for test). Transform data using the kernel function was selected with the smallest total misclassification error. The support vector machine with radial kernel function and $\gamma = 6$ was the best model for this method. Then, this model was applied to classify a cases that was taken out. We repeated for all 150 cases. The result presents in Table 5.13. This method presents 100% correction for this iris data.

Table 5.13: result from support vector machine on iris data

Observed Class	Classified class		
	Iris-setosa	Iris-vesicolor	Iris-virginica
Iris-setosa	50	0	0
Iris-vesicolor	0	50	0
Iris-virginica	0	0	50

Table 5.14: Total misclassification error (TME) and sensitivity for Iris data

	TME	Sensitivity		
		Iris-setosa	Iris-vesicolor	Iris-virginica
Multinomial logistic regression	0.0333	1	0.94	0.96
Classification tree	0.0333	1	0.94	0.96
Bayesian method	0.0267	1	0.94	0.98
Support vector machine	0	1	1	1
Confidence set method	0.0667	0.90	0.76	0.80
Confidence set method (reclassification)	0	1	0.80	0.90

Comparison of classification methods

Total misclassification error (TME) and sensitivity from five classification methods for iris data is present in Table 5.14. For this data, the new confidence set method with reclassification provides the zero total misclassification error. Support vactor machine technique gives 100% correct classified for this data. The new confidence set method shows the lowest sensitivity because the result of multi-classes is not include in correct classified for some cases, which cannot be clearly classified into one single class. As a result, the probability of correct classified was controlled.

5.6.4 Seeds data

The seeds data set is a classified multivariate data of three varieties (three classes) of wheat: Kama, Rosa and Canadian 70 samples each, with seven measurements of internal kernel structure:

- area A,
- perimeter P,
- compactness C,
- length of kernel,
- width of kernel,
- asymmetry coefficient,
- length of kernel groove.

The five classification methods were applied to classify a new case by using leave one out method for evaluation.

Confidence set method

TABLE 5.15: the number of result from the new confidence set method

	{1}	{2}	{3}	{1,2}	{2,3}	{1,3}	{1,2,3}	ϕ
Kama (1)	56	0	1	3	0	5	0	5
Rosa (2)	2	62	0	3	0	0	0	3
Canadian (3)	2	0	49	0	0	16	0	3

Table 5.15 presents the number of result from the new confidence set method. The result from this method will be a confidence set for the true class, which can be $\{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}$ and ϕ, empty set. Some new cases, which fall inϕ is reclassified by Bayesian method for this study. The reclassification presents in the Table 5.16 does not decrease the probability of correct classification, which is not less than $1 - \alpha$.

TABLE 5.16: the number of result from the new confidence set method with reclassified by Bayesian method

	{1}	{2}	{3}	{1,2}	{2,3}	{1,3}	{1,2,3}	ϕ
Kama (1)	59	1	2	3	0	5	0	0
Rosa (2)	3	64	0	3	0	0	0	0
Canadian (3)	2	0	52	0	0	16	0	0

Classification tree

The `tree()` function of package `tree` in `R` program was used to construct tree model using 209 cases (one case was taken out for test). The classification tree diagram composed of three measurements: length of kernel groove, area A and asymmetry coefficient as in figure 5.4. Then, this tree was applied to a cases that was taken out. We repeated for all 210 cases. The result presents in Table 5.17.

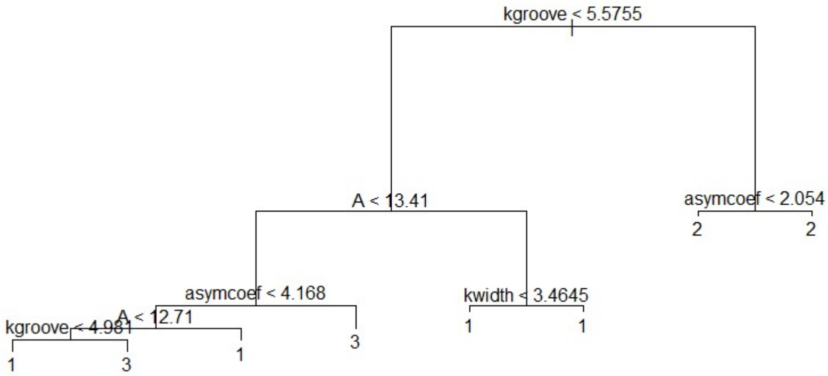

FIGURE 5.4: classification tree diagram for classifying a new case for seeds data

TABLE 5.17: result from classification tree on seeds data

Observed	Classified class		
Class	Kama	Rosa	Canadian
Kama	64	1	5
Rosa	2	68	0
Canadian	6	0	64

Logistic regression

The `multinom()` function of package `nnet` in `R` program was used to construct multiple logistic regression model using 209 cases (one case was taken out for test). The multiple logistic regression model composed of four measurements: length of kernel, length of kernel groove, area A and asymmetry coefficient. Three measurements, perimeter P, compactness C and the width of kernel, were removed by backward elimination method. Then, this model was applied to a cases that was taken out. We repeated for all 210 cases. The result presents in Table 5.18.

TABLE 5.18: result from multiple logistic regression on seeds data

Observed Class	Classified class		
	Kama	Rosa	Canadian
Kama	67	1	2
Rosa	0	70	0
Canadian	3	0	67

Bayesian method

The `naiveBayes()` function of package `e1071` in `R` program was used to construct Bayes' classifier using 209 cases (one case was taken out for test). Backward elimination was the method for variable selection with the smallest total misclassification error. The Bayes' classifier composed of five measurements: perimeter P, compactness C, the length of kernel,asymmetry coefficient and length of kernel groove. Two measurements, the width of kernel and area A, was removed. Then, this model was applied to classify a cases that was taken out. We repeated for all 210 cases. The result presents in Table 5.19.

TABLE 5.19: result from Bayesian method on seeds data

Observed Class	Classified class		
	Kama	Rosa	Canadian
Kama	60	3	7
Rosa	2	68	0
Canadian	3	0	67

Support vector machine

The `svm()` function of package `e1071` in `R` program was used to construct support vector machine using 209 cases (one case was taken out for test). Transform data using the kernel function was selected with the smallest total misclassification error. The support vector machine with radial kernel function and $\gamma = 3$ was the best model for this method. Then, this model was applied to classify a cases that was taken out. We repeated for all 210 cases. The result presents in Table 5.20.

TABLE 5.20: result from support vector machine on seeds data

Observed Class	Classified class		
	Kama	Rosa	Canadian
Kama	65	1	4
Rosa	5	65	0
Canadian	4	0	66

Comparison of classification methods

Table 5.21: Total misclassification error (TME) and sensitivity for seeds data

	TME	Sensitivity		
		Kama	Rosa	Canadian
Multinomial logistic regression	0.0286	0.9571	1	0.9571
Classification tree	0.0667	0.9143	0.9714	0.9143
Bayesian method	0.0714	0.8571	0.9714	0.9571
Support vector machine	0.0667	0.9286	0.9286	0.9429
Confidence set method	0.0762	0.80	0.8857	0.70
Confidence set method (reclassification)	0.0381	0.8429	0.9143	0.7429

Total misclassification error (TME) and sensitivity from five classification methods for seeds data is present in Table 5.21 For this data, logistic regression provides the smallest total misclassification error and also the highest sensitivity for all three classes. The new confidence set method shows the lowest sensitivity because the result of multi-classes is not include in correct classified for some cases, which cannot be clearly classified into one single class. As a result, the probability of correct classified was controlled.

5.7 Conclusion

In this Chapter, we have proposed a new confidence set method for classification problem, which can control the probability of correct classified. We have briefly reviewed four classical classification methods. Classification tree method is a simple method for initial variables selection, also observing variables interaction. Logistic regression is one of the popular method for classification because the odds ratio can be calculated from this method. Support vector machine use kernel technique to adjust the explanatory variables. Bayesian method can minimize the missclassification error, but can not control it. For important problem, such as, classifying patients, we need suitable methods which can provide the most accurate result. The new confidence set method is the only one method in this study that can controls the probability of correct classified. Moreover, when some cases cannot be classified into one specific class with good confidence, this method provides a set of class(es), which indicates that further evaluation is requires.

The three real data examples; the data of cirrhosis patients from a hospital, the well known iris data and seeds data are used for comparing between those five classification methods. The total misclassification error and sensitivity (true positive rate) are used for illustrating and comparing the methods. Support vector machine and logistic regression show the low misclassification error with high sensitivity. However, the confidence set method also shows the low total misclassification error as this method controls probability of correct classified. Some cases which got the empty set, ϕ, result can be reclassified with other methods. The final result does not decrease the probability of correct classification. For some important problem, such as, classifying patients, we need suitable methods which can provide the most accurate result. The new confidence set method controls the probability of correct classified. Moreover, when some cases cannot be classified into one specific class with good confidence, this method provides a set of class(es), which indicates that further medical evaluation is required.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

The construction of confidence set can be applied in many problems. In this study, we are focusing on comparison and classification problems. We can construct a confidence set for equivalence test by inverting technique. The intersection-union test can be applied with the construction of a confidence set in this problem. We also illustrate the construction of the upper confidence bounds on several samples by three methods: using the theorem from Liu et al. (2009), F statistic and Studentized range statistic. The first method, using the theorem from Liu et al. (2009), cannot compute the upper confidence bound of $\frac{|\mu_1-\mu_2|}{\sigma}$ if σ is unknown. However, it can be use for calculating the upper confidence bound of $|\mu_1 - \mu_2|$ whether σ is known or unknown, and the data does not need to be normal distribution for any situation. The other two methods, using the F-statistic and using the studentized range statistic, the data needs to be normal distribution, but σ is not necessary for computation.

For classification problem, we would like to classify a new case into its true class, based on some measurements. Five classification methods have been studied. They are logistic regression, classification tree, Bayesian method, support vector machine and the new confidence set method. The new method constructs a confidence set for the true class for a new case by inverting the acceptance sets. The advantage of this method is that the probability of correct classification is not less than $1-\alpha$. For important problem of classifying patients, we need suitable methods which can provide the most accurate result. The new confidence set method controls the probability of correct classified. Moreover, when some cases cannot be classified into one specific class with good confidence, this method provides a set of class(es), which indicates that further information or futher medical evaluation for patient is required.

The suggestion for using each classification method depends on the purpose of the classifying. The new confidence set method can control the probability of correct classification. This method is suitable for the classification that needs more accuracy, such as, diagnostic problem for patients in hospital. When the classified is questionable, we cannot make a classification with confidence. However, some situation we just needs to make some decision. Most of the methods in this study can be used for this purpose. Classification tree present the diagram result, which reviews the important variables and some variable interaction. Logistic regression can compute the odds ratio of belonging to each class. Bayesian method give the minimum error result based on computational method. Support vector machine has the kernel technique and adjustable parameters this make the method more flexible.

6.2 Future Work

The construction of upper confidence bonds on the range of means still have a lot of points for further study. The first method, using the theorem from Liu et al. (2009), cannot compute the upper confidence bound of $\frac{|\mu_1-\mu_2|}{\sigma}$ if σ is unknown. This is an interesting topic to be studied when σ is unknown. The other two methods, using the F-statistic and using the studentized range statistic, the data needs to be normal distribution. The further study of these two methods when the data is not normal distribution is needed. All methods also require to assume the equal variance in all groups. Then, the situation that the equal variance is not assume is also another point to consider. Moreover, the effect of unbalance sample sizes for more than two samples groups when we construct the upper confidence bonds on the range of means requires a further more study with a complicated numerical computation.

The construction of upper confidence bonds on the maximum difference between two regression lines, we have illustrated only for two groups of data (two regression lines). Then, the study of several regression lines still be required.

When the new confidence set method for classification is used in this study, it requires that the data is distributed as normal distribution because of the construction of the acceptance sets. The extended study can be focused on the data which is not distributed as normal distributions. Moreover, the limitation of this study is that most of the computation needs some complex codding. Then, It can be the further study to develop the package bases on each method to be more user friendly package.

Bibliography

T.W. Anderson. *An Introduction to Multivariate Statistical Analysis.* Wiley-Blackwel, New York, 3rd edition, 2003.

R.L. Berger. Multiparameter hypothesis testing and acceptance sampling. *Technometrics*, 24(4):295–300, 1982.

R.L. Berger and J.C. Hsu. Bioequivalence trials intersection-union tests and equivalence confidence sets. *Statistical Science*, 11(4):283–302, 1996.

E. Bofinger, A.J. Hayter, and W. Liu. The construction of upper confidence bounds on the range of several location parameters. *Journal of the American Statistical Association*, 88(423):906–911, 1993.

F. Bretz, T. Hothorn, and P. Westfall. *Multiple Comparison Using R.* Chapman and Hall/ CRC, Boca Raton, 2011.

C. Cortes and V. Vapnik. Support-vector network. *Machine Learning*, 20(3):273–297, 1993.

M.J. Crawley. *The R Book.* John Wiley & Sons Ltd, West Sussex, 2nd edition, 2003.

J.J. Faraway. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models.* Chapman and Hall/ CRC, Boca Raton, 2006.

R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

P. Flach. *Machine Learning, The Art and Science of Algorithms that Make Sense of Data.* Cambridge University Press., Cambridge, 2012.

M.J. Hilbe. *Logistic Regression Model.* Chapman and Hall/CRC, Boca Raton, 2009.

Y. Hochberg and A.C. Tamhane. *Multiple Comparison Procedures.* Wiley-Blackwel, New York, 2011.

D.W. Hosmer, S. Lemeshow, and R.X. Sturdivant. *Applied Logistic Regresion.* John Wiley & Sons Inc., New Jersey, 3rd edition, 2013.

J.C. Hsu. *Multiple Comparisons Theory and Methods.* Chapman and Hall/ CRC, Boca Raton, 1996.

D.G. Kleinbaum, L.L. Kupper, K.E. Muller, and A. Nizam. *Applied Regression Analysis and Other Multivariable Methods.* Wiley, New York, 4th edition, 2008.

E.L. Lehmann. *Testing Statistical Hypotheses.* Duxbury Press., California, 2rd edition, 1986.

W. Liu. *Simultaneous Inference in Regression.* Chapman and Hall/ CRC, Boca Raton, 2010.

W. Liu, F. Bretz, A.J. Hayter, and H.P. Wynn. Assessingnonsuperiority, noninferiority, or equivalence when comparing two regression models over a restricted covariate region. *Biometrics*, 65(4):1279–1287, 2009.

W. Liu, A.J. Hayter, and H.P. Wynn. Operability region equivalence: simultaneous confidence bands for the equivalence of two regression models over restricted regions. *Biometrical Journal*, 49(1):144–150, 2007.

W.L. Martinez and A.R. Martinez. *Computational Statistics Handbook with MATLAB.* Chapman and Hall/ CRC, Boca Raton, 2nd edition, 2008.

R.G. Miller. *Simultaneous Statistical Inference.* pringer-Verlag, New York, 2nd edition, 1980.

S.J. Ruberg and J.W. Stegeman. Pooling data for stability studies: testing the equality of batch degradation slopes. *Biometrics.*, 47(3):1059–1069, 1991.

E. Uusipaikka. *Confidence Intervals in Generalized Regression Models.* Chapman and Hall/ CRC, Boca Raton, 2008.

S. Wellek. *Testing Statistical Hypotheses of Equivalence and Noninferiority.* Chapman and Hall/ CRC, Boca Raton, 2nd edition, 2010.