UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES
Mathematical Sciences

Some Problems Relating To Simultaneous Confidence Bands

by

Daniel Mark Tompsett

Thesis submitted for the degree of Doctor of Philosophy
April 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Mathematical Sciences

Doctor of Philosophy

Some Problems Relating To Simultaneous Confidence Bands

by Daniel Mark Tompsett

The thesis comprises two distinct areas of research involving the use of simultaneous confidence bands, specifically relating to the Scheffé type simultaneous confidence band of Scheffé (1953), and the constant width simultaneous confidence band of Gafarian (1964).

The first relates to establishing the optimal design of experiments for simultaneous confidence bands, considered over a finite range of a covariate. Methodology is developed for this area that focuses on establishing continuous optimal designs under two optimality criteria specific to simultaneous confidence bands, the Average Width and Minimum Area Confidence Set criteria. We develop a method of numeric investigation which allows the search area to be constrained over intervals. From this, we conclude that the optimal continuous designs for 95 percent simultaneous confidence bands, considered over the range $[-1, 1]$ are D-optimal for specific values of N. Also investigated is the application of traditional analytic methods used to obtain optimal continuous designs.

Secondly, simultaneous confidence bands can be used to construct simultaneous confidence sets for the Effective Doses (ED) of a the logistic regression model (cf. Walter (1983)), by inverting the bounds of a Scheffé type simultaneous confidence band. We introduce an improvement to this method, guaranteed to exhibit closer to nominal simultaneous coverage by constructing specialised simultaneous confidence sets for a specific number (k) of ED's. Two sided sets are fully defined for $k = 2$ for a multiple covariate model, and $k = 3$ or more for a one covariate model. This new methodology is then applied to construct simultaneous confidence sets for two additional situations: (i) when the ED can be assumed to lie over a finite interval, (ii) when one sided simultaneous confidence sets are sought. These methods may be applied to any generalised linear model, and the improvements over the original methods are illustrated with examples.

All numeric methods provided in the thesis, including illustrations, were carried out by custom programs made in R, which are available by request.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Daniel Mark Tompsett, declare that the thesis entitled "Some Problems Relating To Simultaneous Confidence Bands" and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

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

Signed:

Date:

Acknowledgements

I wish to firstly dedicate this work to my two supervisors, Professor Wei Liu, and Doctor Stefanie Biedermann for their constant support, without which this thesis would not have been possible. I also wish to thank the University of Southampton, as well as all the friends of whom I have met over the seven years I have been here for the best years of my life. Finally, I want to thank my family, whose continued support would have made my entire academic career unachievable.

Chapter 1

Introduction

Linear regression is ubiquitous in statistics and dates back to the early 1800's. The method seeks to evaluate the relationship between a response y, some q covariates, $x_1, \ldots, x_q$, and some random error term ϵ, through the basic relation

$$y = \beta_0 + x_1\beta_1 + \ldots + x_q\beta_q + \epsilon. \tag{1.1}$$

A simultaneous confidence band provides information on the plausible range of the unknown linear regression line of this model, $\boldsymbol{x}^\intercal\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)$ is the parameter vector, and $\boldsymbol{x} = (1, x_1, \ldots, x_q)^\intercal$ contains the q covariates. Consider a vector of covariates $x = (x_1, \ldots, x_q)^\intercal$. Then, in general (cf. Liu et al. (2008)), a two sided simultaneous confidence band consists of an lower and upper function of the covariates $l(x) < u(x)$, such that the probability of the true regression line, $\boldsymbol{x}^\intercal\boldsymbol{\beta}$, lying fully between both functions for all x over some q-dimensional rectangular region is exactly $1 - \alpha$. This rectangular region is represented by two vectors of constants, $\boldsymbol{a}$ and $\boldsymbol{A}$, with entries $a_i \leq x_i \leq A_i$ $i = 1, \ldots, q$, that is $x \in (\boldsymbol{a}, \boldsymbol{A})$. The simultaneous confidence band then sets

$$\mathrm{P}\left\{l(x) < \boldsymbol{x}^\intercal\boldsymbol{\beta} < u(x)\ \forall x \in (\boldsymbol{a}, \boldsymbol{A})\right\} = 1 - \alpha.$$

The point of providing such a broad definition is that, despite being naturally applied to, and indeed first developed, for the normal-error linear regression model around eighty five years ago, a simultaneous confidence band may be constructed for the linear regression line of **any** generalised linear model (GLM), by using the asymptotic normal distribution.

The earliest work on simultaneous confidence bands can be traced back to the work of Working and Hotelling (1929), in which a two sided simultaneous confidence band, of hyperbolic shape, was considered for a single covariate over $(a, A) = (-\infty, \infty)$, in the special case of known variance. This band was most notably expanded upon in Scheffé (1953), in which it was considered for multiple covariates over the whole real space with an unknown variance parameter. This band is known as the Scheffé type simultaneous confidence band and is the best known, and most recognisable in use. It is this band type in particular that is constructed for GLM's. The first example of this was probably in Brand et al. (1973), which constructed a Scheffé type band on the one

covariate logistic regression model under asymptotic normality conditions. This work was then extended to the multiple covariate case in Walter (1983).

There exists three main other types of confidence bands. The work of Gafarian (1964) first introduced the concept of a constant width simultaneous confidence band, taking the same width over the whole interval. This was then shown in Bowden and Graybill (1966) to be a special case of a family of two sided three segment confidence bands. These bands, and a similar family known as two segment bands (cf. Graybill and Bowden (1967)) comprise the last two main types of confidence bands, but are only covered briefly in this thesis. An excellent amalgamation of this work is given by Liu (2011) and is a recommended starting point in the area.

The thesis is motivated by two long standing areas of research. The first seeks to establish the optimal design of experiments for simultaneous confidence bands, considered over a finite range of the one covariate normal-error linear regression line. Two optimality criteria exist for simultaneous confidence bands, the Average Width (AW) criterion considered by Gafarian (1964) and Naiman (984a), and the more robust Minimum Area Confidence Set (MACS) criterion of Liu and Hayter (2007). We consider both as criteria for an optimal design for both the Scheffé type, and constant width type bands. The Scheffé type band is considered for its utility, and the constant width band for its computational simplicity.

The second is to formulate an improvement to a method first considered by Brand et al. (1973), which constructed simultaneous confidence sets on the Effective Dose (ED) using the logistic regression model, by inverting the bounds of a Scheffé type simultaneous confidence band. This method has been used by multiple authors such as Buckley and Piegorsch (2008) and Carter et al. (1986), but is often noted (cf. Li et al. (2008)) to be unduly conservative in terms of simultaneous coverage, particularly for a small number of ED's. The thesis develops new methodology to address this problem by considering confidence bands that may be inverted to obtain simultaneous confidence sets for a specific number of effective doses.

This chapter outlines the basics of normal-error linear, and generalised linear regression as well as simultaneous confidence bands, and reviews the relevant work that has gone before, and what is new that the thesis provides.

1.1 An Introduction To Normal-Error Linear Regression

Linear regression is a statistical technique which aims to evaluate the relationship between a dependent response variable y, and a set of q covariates $x_1, \ldots, x_q$. When $q = 1$, we refer to the single covariate as x. For $q > 1$, we refer to x as the vector of the covariates $x = (x_1, \ldots, x_q)$, and $\boldsymbol{x}$ is the vector x, with first entry as 1. Hence we model the relation as

$$y = \beta_0 + x_1\beta_1 + \ldots + x_q\beta_q + \epsilon = \boldsymbol{x}^\intercal\boldsymbol{\beta} + \epsilon. \tag{1.2}$$

where $\boldsymbol{x} = (1, x_1, \ldots, x_q)$, and $\boldsymbol{x}^\intercal\boldsymbol{\beta}$ is the linear regression line. We assume y is a continuous random variable following some normal distribution. The error term ϵ represents the random variation in the response not explained by the linear regression line. For normal-error linear regression we assume that the random error follows a normal distribution with mean 0 and variance $\sigma^2 > 0$. It is then obvious that the distribution of the response may be expressed as $y \sim \mathcal{N}(\boldsymbol{x}^\intercal\boldsymbol{\beta}, \sigma^2)$. In practical applications, N observations of the response and corresponding covariates are taken, giving N data points of the form $(y_i, x_{i1}, \ldots, x_{iq})$ $i = 1, \ldots, N$. We model each data point through the relation in (1.2), with a common value for the parameters $\beta_1, \ldots, \beta_q$ as follows

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq} + \epsilon_i \;\; i = 1, \ldots, N. \tag{1.3}$$

This regression model is better represented in its matrix form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1.4}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1q} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & \ldots & x_{Nq} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{pmatrix}.$$

The set of all N choices of $(x_{i1}, \ldots, x_{iq})$ is known as the **design of the experiment**, and is characterised by $\boldsymbol{X}$, known as the design matrix, which represents the input into the model controllable by the user. We assume that each y_i is a realisation of the random variable Y_i, distributed as $\mathcal{N}(\beta_0 + x_{i1}\beta_1 + \ldots + x_{iq}\beta_q, \sigma^2)$, and the ϵ_i are independent. Hence, $\boldsymbol{Y} \sim \mathcal{N}_N(\boldsymbol{X}\boldsymbol{\beta}, I\sigma^2)$ and $\boldsymbol{\epsilon} \sim \mathcal{N}_N(\boldsymbol{0}, I\sigma^2)$.

1.1.1 Parameter Estimation

We need to obtain estimates of the parameters of interest $\boldsymbol{\beta}$ and σ^2 in order to understand and utilise the model. The least squares estimator $\hat{\boldsymbol{\beta}}$ minimises the sum of squares of the errors of the model as follows

$$\min_{\beta_0, \beta_1, \ldots, \beta_q} \sum_{i=1}^{N} \epsilon_i^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{1i} - \ldots - \beta_q x_{qi})^2.$$

Note that

$$\begin{aligned} &\sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{1i} - \ldots - \beta_q x_{qi})^2 \\ =&(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\intercal(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\ =&\boldsymbol{Y}^\intercal\boldsymbol{Y} - 2\boldsymbol{\beta}^\intercal\boldsymbol{X}^\intercal\boldsymbol{Y} + \boldsymbol{\beta}^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}\boldsymbol{\beta}. \end{aligned}$$

This is since $\boldsymbol{Y}^\intercal \boldsymbol{X}\boldsymbol{\beta}$ and its transpose $\boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{Y}$ are both scalars and hence equal. Therefore $\hat{\boldsymbol{\beta}}$ satisfies

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{Y}^\intercal \boldsymbol{Y} - 2\boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{Y} + \boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X}\boldsymbol{\beta}) \\
=& -2\boldsymbol{X}^\intercal \boldsymbol{Y} + 2\boldsymbol{X}^\intercal \boldsymbol{X}\boldsymbol{\beta} \\
=& 0 \text{ at } \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}.
\end{aligned}$$

This yields the well known least squares estimator

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y}. \tag{1.5}$$

This estimator assumes the matrix $(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}$, known as the variance covariance matrix of the model, exists. Its corresponding inverse $(\boldsymbol{X}^\intercal \boldsymbol{X})$, is known as the the information matrix. The estimator of the random error $\hat{\boldsymbol{\epsilon}}$ known as the residuals is given as

$$\hat{\boldsymbol{\epsilon}} = \begin{pmatrix} \hat{\epsilon_1} \\ \vdots \\ \hat{\epsilon_N} \end{pmatrix} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

From this, an estimator for the variance σ^2 takes the form

$$\begin{aligned}
\hat{\sigma}^2 =& \frac{\sum_{i=1}^{N} \hat{\epsilon_i}^2}{N-q-1} \\
=& \frac{\hat{\boldsymbol{\epsilon}}^\intercal \hat{\boldsymbol{\epsilon}}}{N-q-1} \\
=& \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\intercal (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{N-q-1} \\
=& \frac{\|\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2}{N-q-1}.
\end{aligned}$$

These estimators may now be used to define an estimated least squares linear regression model of the form

$$\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\epsilon}}. \tag{1.6}$$

Furthermore, they may be used in equation (1.2) to define a least squares estimated regression line of the form

$$\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p = \boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \tag{1.7}$$

from which inference may be made.

1.1.2 Distributions of Parameter Estimators

To obtain any inference on the linear regression model, it is necessary to find the probability distributions of the parameter estimators. The least squares estimator

$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{Y}$ is a linear transformation of $\boldsymbol{Y} \sim \mathcal{N}_N(\boldsymbol{X}\boldsymbol{\beta}, I_N\sigma^2)$ and therefore must also be normally distributed with mean

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) =& E((\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{Y}) \\ =& (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal E(\boldsymbol{Y}) \\ =& (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{X}\boldsymbol{\beta} \\ =& \boldsymbol{\beta} \end{aligned}$$

and variance

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}) =& Cov(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}) \\ =& Cov((\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{Y}, (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{Y}) \\ =& (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal Cov(\boldsymbol{Y}, \boldsymbol{Y})(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal \\ =& \sigma^2(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}. \end{aligned}$$

Hence $\hat{\boldsymbol{\beta}} \sim \mathcal{N}_{q+1}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1})$ and thus $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$, a property that makes the least squares estimate particularly useful. It is immediate from this property that

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2} \sim \mathcal{X}^2_{q+1}, \tag{1.8}$$

where $\mathcal{X}^2_{q+1}$ is a chi square random variable with $q+1$ degrees of freedom.

We note that $\hat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = (\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal)\boldsymbol{Y}$ is also a linear combination of $\boldsymbol{Y}$, and therefore must also follow a normal distribution with mean

$$\begin{aligned} E(\hat{\boldsymbol{e}}) =& E((\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal)\boldsymbol{Y}) \\ =& (\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal)\boldsymbol{E}(\boldsymbol{Y}) \\ =& (\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal)\boldsymbol{X}\boldsymbol{\beta} \\ =& \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{X}\boldsymbol{\beta} \\ =& \boldsymbol{0} \end{aligned}$$

and variance

$$\begin{aligned} Var(\hat{\boldsymbol{e}}) =& Cov(\hat{\boldsymbol{e}}, \hat{\boldsymbol{e}}) \\ =& Cov((\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal)\boldsymbol{Y}, (\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal)\boldsymbol{Y}) \\ =& (\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal)Cov(\boldsymbol{Y}, \boldsymbol{Y})(\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal) \\ =& \sigma^2(\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal) \end{aligned}$$

since $(\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal)$ is equal to its transpose. Thus

$$\hat{\boldsymbol{e}} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2(\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal)).$$

Furthermore it can be shown (cf.Liu (2011)) that the estimator of the variance has probability distribution

$$\hat{\sigma}^2 \sim \frac{\sigma^2 \mathcal{X}^2_{N-q-1}}{N-q-1}.$$

1.1.3 Independence of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$

We note that

$$\begin{aligned} Cov(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{e}}) =& Cov((\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y}, (\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal)\boldsymbol{Y}) \\ =& \sigma^2 (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal (\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal) \\ =& 0. \end{aligned}$$

Since $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{e}}$ are normally distributed, and thus must be jointly normally distributed, it follows that they must be independent random variables. Furthermore as $\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\epsilon}}^\intercal \hat{\boldsymbol{\epsilon}}}{N-q-1}$ is a linear combination of $\hat{\boldsymbol{e}}$, it follows immediately that $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are also independent.

1.2 Introduction To Simultaneous Confidence Bands For The Normal-Error Linear Regression Model

All bands considered in this thesis are symmetric around the least squares estimated linear regression line $\boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}}$. A $1-\alpha$ Scheffé type simultaneous confidence band over a q dimensional rectangular region $x \in (\boldsymbol{a}, \boldsymbol{A})$ takes the form

$$\boldsymbol{x}^\intercal \boldsymbol{\beta} \in \boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \pm c\hat{\sigma} \sqrt{\boldsymbol{x}^\intercal (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{x}} \ \ \forall x \in (\boldsymbol{a}, \boldsymbol{A}). \tag{1.9}$$

When there is just one covariate, denoted x, the band is considered over the straight line interval $a \leq x \leq A$. We denote c as the critical constant, which is set to ensure that the confidence band has the nominal simultaneous coverage over the region of interest, that is c satisfies

$$\mathrm{P}\left\{ \boldsymbol{x}^\intercal \boldsymbol{\beta} \in \boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \pm c\hat{\sigma} \sqrt{\boldsymbol{x}^\intercal (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{x}} \ \ \forall x \in (\boldsymbol{a}, \boldsymbol{A}) \right\} = 1 - \alpha.$$

The band takes a hyperbolic shape with the width given by $2c\hat{\sigma}\sqrt{\boldsymbol{x}^\intercal (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{x}}$ which depends on the design through the matrix $\boldsymbol{X}^\intercal \boldsymbol{X}$. Hence the band is wider at more extreme values of x. The band was first considered by Working and Hotelling (1929) for one covariate over the whole real line for the case of known σ, and extended by Scheffé (1953) to the multiple covariate case over the whole real space for unknown σ.

Two sided Scheffé type bands for one covariate over a finite interval (a, A) were considered by Uusipaikka (1983) and Wynn and Bloomfield (1971), and over a rectangular region for the multiple covariate case via simulation methods in Liu et al. (2005b). For one sided Scheffé type bands, Hochberg and Quade (1975) considered such bands over the whole real space for a multiple covariate model, whilst Pan et al. (1975) constructed such bands for a single, and constrained covariate.

First considered by Gafarian (1964) for the case of $q = 1$, that is one covariate, a constant width simultaneous confidence band takes the form

$$\beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm \hat{\sigma} c \; \forall x \in (a, A) \tag{1.10}$$

where c is again set to satisfy the simultaneous coverage of the band. Constant width bands are the simplest type, with the width equal to $2\hat{\sigma}c$ over the whole range, depending implicitly on the design through c. The constant width band is a special case of a family of two sided three segment bands originally proposed for one covariate by Bowden and Graybill (1966), which take the form

$$\beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm \hat{\sigma} \boldsymbol{H_3}(x) \; \forall x \in (a, A) \tag{1.11}$$

with

$$\boldsymbol{H_3}(x) = \tfrac{1}{A-a} \left((x-a) c_1 \sqrt{(1, A)(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \begin{pmatrix} 1 \\ A \end{pmatrix}} + (A-x) c_2 \sqrt{(1, a)(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \begin{pmatrix} 1 \\ a \end{pmatrix}} \right).$$

In this instance the simultaneous coverage of the band is set by two critical constants, c_1 and c_2. A unique pair is obtained by setting one as a function of the other, which determines the type of three segment band. If we set

$$c_2 \sqrt{(1, a)(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \begin{pmatrix} 1 \\ a \end{pmatrix}} = c_1 \sqrt{(1, A)(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \begin{pmatrix} 1 \\ A \end{pmatrix}}, \tag{1.12}$$

then (1.11) becomes a constant width band with critical constant

$$c = c_2 \sqrt{(1, a)(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \begin{pmatrix} 1 \\ a \end{pmatrix}}.$$

Hence c may be obtained using a method to calculate c_1 and c_2 for three segment bands given by Bowden and Graybill (1966) under this relation. These methods are given in Liu (2011), and Liu et al. (2005a).

1.3 A Review Of Optimal Designs For Simultaneous Confidence Bands

Despite the width of each confidence band depending on the design of the experiment, there is little work relating optimal design with simultaneous confidence bands.

The first notion of an "optimal" simultaneous confidence band is found in Gafarian (1964), who argued that a band that was smaller in width on average, gave a more tightly bounded, and so better region for $\boldsymbol{x}^\intercal \boldsymbol{\beta}$. This is known as the Average Width (AW) criterion. However Gafarian gave only a passing note on the optimal design that minimised the AW of the constant width band, offering a solution but without

proof. Work on the AW criterion was followed on notably by Naiman (1983), and then Naiman (984a) and Piegorsch (985b), which developed the weighted average width of a band with respect to some probability measure. Yet this work exclusively focused on comparing different band types over specific probability measures, with no direct work on optimal design.

The Minimum Area Confidence Set (MACS) criterion was introduced by Liu and Hayter (2007). Whilst again much of the work focused on intra-band comparison (see Liu and Hayter (2007) and Liu and Ah-Kine (2010) for example), there exist two analytical results for optimal designs under MACS, an equivalence between D optimality and, (i), MACS optimal designs for Scheffé type bands over the whole real space (Atkinson et al. (2007)), and (ii), MACS optimal designs for one covariate and two sided two segment bands (Ah-Kine (2010)). The covariate in practice however is often constrained over a finite range, and for constrained three segment and Scheffé type bands, there is currently no work on optimal designs at all, save a useful numeric investigation in Ah-Kine (2010) that could not constrain the search area to intervals, and was reliant on a pre existing dataset. The motivation for this research follows directly from this work.

All current work in the area relies on an exact design framework. When an analytic proof of an optimal design is required, it is typical to instead use a continuous design framework (cf. Kiefer (1959) and Fedorov (1972)) and apply methods described, amongst others, in Silvey (1980). We explore the application of these methods to the MACS and AW criteria functions, for constrained, one covariate Scheffé type and constant width bands. We develop numeric methods for continuous designs not dependent on a dataset where the response surface, based on the elements of the continuous information matrix, is constrained over a rectangular region. This is used to show that over the range $[-1, 1]$ the best design under AW and MACS is D optimal. We further manipulate the criteria functions into a manner that permits the use of analytic methods in Silvey (1980), and discusses the difficulties in applying these methods in practice.

1.4 An Introduction To Logistic Regression

It is commonly of interest to assess data with a non normally distributed response. This can be done using generalised linear regression. We assume the response data are realisations of a random variable y following some distribution from the exponential family, whose probability density function may be written as

$$f(y) = h(y)\exp\left\{\boldsymbol{\zeta}^{\intercal}T(y) - A(\boldsymbol{\zeta})\right\},$$

with $\boldsymbol{\zeta}$ a vector of the distribution parameters, and h, T and A appropriate functions. Secondly, it is assumed that the expected value of y depends on some q independent covariates $x = (x_1, \ldots, x_q)$, that is $E(y) = \mu(x)$. Under these circumstances a function of the expectation $b(\mu(x))$ is then chosen such that $b(\mu(x))$ is both continuous and has support $-\infty < b(\mu(x)) < \infty$, known as a **link function**.

The link function can then be modelled as a linear combination of the q covariates as follows

$$b(\mu(x)) = \boldsymbol{x}^\intercal \boldsymbol{\beta}$$

$$\boldsymbol{x}^\intercal = (1, x_1, \ldots, x_q) \;\; \boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)^\intercal .$$

This is known as a generalised linear model (GLM).

1.4.1 Logistic Regression Model

The logistic regression model is a popular generalised linear model for dichotomous response data. We assume a Bernoulli distributed response variable y, taking two values, 0 and 1, where the expectation, that is the probability of observing a success (y=1) depends on the vector of q independent explanatory variables $x = (x_1, \ldots, x_q)$, that is $y \sim Bernoulli(p(x))$. In logistic regression the link function is set to be the log odds of the probability of success, this gives the logistic model as

$$\pi(p(x)) = \log\left(\frac{p(x)}{1-p(x)}\right) = \boldsymbol{x}^\intercal \boldsymbol{\beta} \tag{1.13}$$

or equivalently,

$$p(x) = \frac{\exp\left(\boldsymbol{x}^\intercal \boldsymbol{\beta}\right)}{1 + \exp\left(\boldsymbol{x}^\intercal \boldsymbol{\beta}\right)} \tag{1.14}$$

$$\boldsymbol{x}^\intercal = (1, x_1, \ldots, x_q)\,, \;\; \boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)^\intercal .$$

The logistic regression model is notably used in drug dose response curves (Carter et al. (1986)) and hazard assessment (Piegorsch and West (2005)). A key concept in any such study is the Effective Dose (ED), which is defined as the specific combination(s) of x required to elicit a certain probability of response (or log odds of a response) in the model, that is

$$x_p = \left\{ x : \boldsymbol{x}^\intercal \boldsymbol{\beta} = \pi(p) = \log\left(\frac{p}{1-p}\right) \right\} = \left\{ x : \frac{\exp\left(\boldsymbol{x}^\intercal \boldsymbol{\beta}\right)}{1 + \exp\left(\boldsymbol{x}^\intercal \boldsymbol{\beta}\right)} = p \right\}.$$

Further detail is provided in Chapter four of the thesis.

1.4.2 Parameter Estimation

In practice, a set of observations $(y_i, x_i = (x_{i1}, \ldots, x_{iq}))$ are taken and each independently modelled as in Equation (1.13) under the assumption that the y_i are realisations of $Y_i \sim Bernoulli(p(x_i))$:

$$\pi(p(x_i)) = \log\left(\frac{p(x_i)}{1-p(x_i)}\right) = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} \quad i = 1, \ldots, N, \tag{1.15}$$

$$\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{iq})^\intercal, \;\; \boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)^\intercal.$$

This allows the estimation the parameter vector $\boldsymbol{\beta}$ using the maximum likelihood technique. In this case the estimator $\hat{\boldsymbol{\beta}}$ is set to maximise the log likelihood function of the

observed responses as follows

$$\max_{\boldsymbol{\beta}} l(\boldsymbol{\beta}|Y_1 = y_1, \ldots, Y_N = y_N) = \sum_{i=1}^{N} \log(p(x_i)^{y_i}(1 - p(x_i))^{1-y_i}), \tag{1.16}$$

where $p(x_i)$ is viewed as a function of $\boldsymbol{\beta}$, through the relation in equation (1.15). This has the effect of maximising the probability that the y_i are realisations of $Y_i \sim Bernoulli(p(x_i))$. We define the score vector as

$$\boldsymbol{u}(\boldsymbol{\beta}) = \begin{pmatrix} \frac{\partial}{\partial \beta_0} l(\boldsymbol{\beta}|Y_1, \ldots, Y_N) \\ \frac{\partial}{\partial \beta_1} l(\boldsymbol{\beta}|Y_1, \ldots, Y_N) \\ \vdots \\ \frac{\partial}{\partial \beta_q} l(\boldsymbol{\beta}|Y_1, \ldots, Y_N) \end{pmatrix}.$$

It is then immediate that the vector $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_{\mathbf{0}}, \hat{\boldsymbol{\beta}}_{\mathbf{1}}, \ldots, \hat{\boldsymbol{\beta}}_{\boldsymbol{q}})^\intercal$ which maximises the likelihood sets

$$\boldsymbol{u}(\hat{\boldsymbol{\beta}}) = \mathbf{0}.$$

The vector $\hat{\boldsymbol{\beta}}$ is known as the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$. Calculation of $\hat{\boldsymbol{\beta}}$ involves solving a series of $q+1$ equations for $q+1$ variables. Therefore in practice the MLE is found by using numerical methods such as Newton-Raphson, and is available in any statistical package performing logistic regression. Furthermore we define a matrix $\boldsymbol{\Sigma}$ as the inverse of the $(q+1) \times (q+1)$ matrix $\boldsymbol{I}(\boldsymbol{\beta})$, with (i, j) element equal to

$$\boldsymbol{I}(\boldsymbol{\beta})_{ij} = \mathbb{E}\left\{ -\frac{\partial^2}{\partial \beta_i \partial \beta_j} l(\boldsymbol{\beta}|Y_1, \ldots, Y_N) \right\}, \quad i, j = 0, \ldots, q \tag{1.17}$$

that is

$$\boldsymbol{I}(\boldsymbol{\beta})^{-1} = \boldsymbol{\Sigma}.$$

The matrix $\boldsymbol{I}(\boldsymbol{\beta})$ is known as the information matrix, and represents the variance covariance matrix of $\boldsymbol{u}(\boldsymbol{\beta})$. In this instance the expectation is in reference to the y_i. Another important construct is the observed information matrix $\boldsymbol{H}$, with entries

$$\boldsymbol{H}(\boldsymbol{\beta})_{ij} = -\frac{\partial^2}{\partial \beta_i \partial \beta_j} l(\boldsymbol{\beta}|Y_1, \ldots, Y_N) \quad i, j = 0, \ldots, q. \tag{1.18}$$

In practice both matrices are unknown, however, observable estimations can be obtained in the obvious way by substituting in the MLE $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$, giving the observable matrices $\boldsymbol{I}(\hat{\boldsymbol{\beta}})$ and $\boldsymbol{H}(\hat{\boldsymbol{\beta}})$ respectively. An estimator for $\boldsymbol{\Sigma}$ is then immediately given as

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{I}(\hat{\boldsymbol{\beta}})^{-1}.$$

We also define the estimator of the inverse of the observed information matrix as

$$\boldsymbol{H}(\hat{\boldsymbol{\beta}})^{-1} = \boldsymbol{J}^{-1}.$$

1.4.3 Large Sample Asymptotic Normality Of The MLE

When the MLE is obtained using a sufficiently large number of data points N, it can be shown (see Walter (1983)) that $\hat{\boldsymbol{\beta}}$ has, approximately, a normal distribution with variance covariance proportional to the matrix $\boldsymbol{\Sigma}$, that is

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \mathcal{N}_{q+1}\left(\mathbf{0}, \boldsymbol{\Sigma}\right). \tag{1.19}$$

This is known as the large sample asymptotic normality of the MLE. It can be shown (cf. Carter et al. (1986)) that $\boldsymbol{\Sigma}$ may be viewed as the asymptotic variance covariance matrix of $\hat{\boldsymbol{\beta}}$ and, though unknown, for large N, $\boldsymbol{\Sigma}$ is consistenly estimated by $\hat{\boldsymbol{\Sigma}} = N\boldsymbol{J}^{-1}$. Therefore for a sufficiently large N, the unknown $\boldsymbol{\Sigma}$ can be replaced by the observable $\boldsymbol{J}^{-1}$, and we have the approximate distributional results

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{d}{\approx} \mathcal{N}_{q+1}\left(\mathbf{0}, \boldsymbol{J}^{-1}\right), \tag{1.20}$$

and equivalently

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\intercal}\boldsymbol{J}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{d}{\approx} \mathcal{X}^2_{q+1}. \tag{1.21}$$

We refer to $\boldsymbol{J}^{-1}$ as the observed variance covariance matrix of the MLE, which is quoted by all statistical packages. In this thesis, it will be assumed that N is large enough that the asymptotic results above are a reasonable approximation of the distribution of the MLE.

1.5 Introduction To Simultaneous Confidence Bands For GLM's

For any GLM defined by some link function $b(\mu(x))$, the asymptotic properties of equations (1.20) and (1.21) apply in the same way as in logistic regression. Therefore, recalling that a GLM has the general form, $b(\mu(x)) = \boldsymbol{x}^{\intercal}\boldsymbol{\beta}$, one may construct an asymptotic Scheffé type simultaneous confidence band over the whole real space for any GLM as

$$b(\mu(x)) = \boldsymbol{x}^{\intercal}\boldsymbol{\beta} \in \boldsymbol{x}^{\intercal}\hat{\boldsymbol{\beta}} \pm c\sqrt{\boldsymbol{x}^{\intercal}\boldsymbol{J}^{-1}\boldsymbol{x}}\ \forall x \in (\boldsymbol{a}, \boldsymbol{A}), \tag{1.22}$$

with c setting the simultaneous coverage probability to $1 - \alpha$. By applying the inverse of the link function, $b^{-1}()$, to the simultaneous confidence band, we can define a simultaneous confidence band directly on the mean $\mu(x)$, which is given as

$$\mu(x) \in b^{-1}\left(\boldsymbol{x}^{\intercal}\hat{\boldsymbol{\beta}} \pm c\sqrt{\boldsymbol{x}^{\intercal}\boldsymbol{J}^{-1}\boldsymbol{x}}\right)\ \forall x \in (\boldsymbol{a}, \boldsymbol{A}). \tag{1.23}$$

This was first considered for the logistic regression model by Brand et al. (1973) for one covariate, and for the multiple covariate case in Walter (1983). However, asymptotic simultaneous confidence bands of this form have been applied to a wide variety of different GLMs. The preface of Liu (2011) offers some varied examples, but focus is often on bands constructed for dichotomous response models that are important in

biological and medical statistics. We cite the logistic model, the hazard function (Lin (1994)), the survival function (Hollander et al. (1997)) and less common models such as the log-logistic and Weibull models in Buckley and Piegorsch (2008) as just some of many examples.

1.6 A Review Of Confidence Sets For The Effective Dose

The importance of the estimation of the effective dose to biological assay and medical statistics has been well noted in Finney (1978), and as such, construction of confidence sets for a single ED is well studied. The two most notable methods are due to Fieller (1954) and Cox (1990), and stem from theory establishing confidence sets for the ratio of two random variables. However, there is significant motivation to identify simultaneous confidence sets for multiple ED's. One particular example is when the minimal effective dose and maximum tolerated dose are both of equal interest, as these two ED's constitute the therapeutic range of a drug.

Simultaneous inference on the effective dose often relies on a method first given by Brand et al. (1973) for the case that the regression model only has one covariate, and extended in Walter (1983) to the case of several covariates. This method involves inverting the bounds of a $(1-\alpha)$ Scheffé type simultaneous confidence band. It was first formally applied to an effective dose framework in Carter et al. (1986), which gave confidence sets on the effective dose involving multiple covariates, known as the multivariate effective dose (MED). This was simplified by Li et al. (2008) by constructing simultaneous confidence sets on the MED, conditioned on $q-1$ of the covariates. It has also seen development for a variety of circumstances and models (see for example Piegorsch and Casella (1988), Piegorsch and West (2005), Buckley and Piegorsch (2008) and Deutsch and Piegorsch (2012)). All of these methods rely on the same fundamental principle however, noted by many authors (for example Deutsch and Piegorsch (2012) and Li et al. (2008)) to be unduly conservative in terms of the simultaneous coverage for any finite number of ED's.

In this thesis we define a new adaptation to this method by constructing simultaneous confidence sets for a **specific** number of effective doses at once, by constructing specialised confidence bands that provide confidence sets offering closer to nominal coverage than the current methods. We construct both two sided and one sided improved simultaneous confidence sets for up to, but not limited to four effective doses at once. The new method is shown to offer significant improvement over previous methods, which is illustrated with some example datasets.

1.7 Outline Of The Thesis

In Chapter two, a review of the simultaneous confidence bands of interest is conducted. Chapter three contains the work on optimal continuous experimental designs for simultaneous confidence bands with a constrained covariate. This includes numeric evidence of D optimal designs being the best under the AW and MACS criteria, and theory on

how an optimal design may be found by analytical proof. Chapter four outlines the construction of one and two sided simultaneous confidence sets for a specific number of effective doses, using a method which is shown to be less conservative than current methodology. Chapter five adapts this work under the stipulation that the effective doses lie over some sensible finite range. The work of Chapters four and five is applied to both one and two sided simultaneous confidence sets, and the extent of improvement in the size of each simultaneous confidence set is assessed and illustrated with examples. In Chapter six, a review of the work is given along with conclusions and possibilities for further work.

Chapter 2

Simultaneous Confidence Bands For Normal-Error Linear And Logistic Regression Models

It is necessary to understand intimately the theory behind the construction of the simultaneous confidence bands of interest, in particular the methods of calculation of the critical constant. In this chapter we review the construction of the Scheffé type band over the whole real space for the normal-error and logistic regression models, and, for a one covariate normal-error linear model, the Scheffé type and constant width type simultaneous confidence bands over a finite interval.

2.1 Preliminaries

The material of this section closely follows section 2.1 of Liu (2011), but is adapted to apply to both normal-error linear and logistic regression models under an asymptotically normal MLE. In the case where results apply specifically to one model or the other, this will be explicitly stated. We assume a one covariate model, that is $q = 1$ for both the normal-error and logistic case.

1. Suppose $\hat{\boldsymbol{\beta}}$ follows a normal distribution with mean $\boldsymbol{\beta}$, and variance involving some covariance matrix $\boldsymbol{V}$, that is $\boldsymbol{V} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}$ for normal-error linear regression, and $\boldsymbol{V} = \boldsymbol{J}^{-1}$ for logistic regression.

   Define $\boldsymbol{P}$ as the unique square root matrix of $\boldsymbol{V}$ that is

   $$\boldsymbol{P}^2 = \boldsymbol{V} \;\Rightarrow\; \boldsymbol{P} = \boldsymbol{V}^{\frac{1}{2}}.$$

   It is clear by the properties of the variance covariance matrix that $\boldsymbol{P} = \boldsymbol{P}^\intercal$.

2. Define $\boldsymbol{N}$ as the random vector

   $$\boldsymbol{N} = \frac{\boldsymbol{P}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma}$$

in the normal-error linear regression case and

$$\boldsymbol{N} = \boldsymbol{P}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

in the asymptotic logistic regression case. It is immediate from the distributions of $\hat{\boldsymbol{\beta}}$ in both cases that $\boldsymbol{N}$ is a standard bivariate normal random vector, that is $\boldsymbol{N} \sim \mathcal{N}_2(\boldsymbol{0}, \boldsymbol{I}_2)$.

3. For the normal-error linear regression model define

$$\boldsymbol{T} = \frac{\boldsymbol{N}}{\frac{\hat{\sigma}}{\sigma}} = \frac{\boldsymbol{P}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\hat{\sigma}}.$$

We note that $\frac{\hat{\sigma}}{\sigma} \sim \frac{1}{\sqrt{N-q-1}}\sqrt{\mathcal{X}^2_{N-q-1}}$. Furthermore since $\boldsymbol{N}$ is a linear combination of $\hat{\boldsymbol{\beta}}$, and $\boldsymbol{N}$ and $\frac{\hat{\sigma}}{\sigma}$ are independent random variables, it is clear that $\boldsymbol{T}$ follows the standard bivariate $\mathcal{T}_{2,N-2}$ distribution, with degrees of freedom $N-2$.

4. Suppose a function of the form

$$v(\boldsymbol{x}) = \boldsymbol{x}^\intercal \boldsymbol{V} \boldsymbol{x}.$$

This may be written as

$$\boldsymbol{x}^\intercal \boldsymbol{V} \boldsymbol{x} = \boldsymbol{x}^\intercal \boldsymbol{P}^2 \boldsymbol{x} = \boldsymbol{x}^\intercal \boldsymbol{P}^\intercal \boldsymbol{P} \boldsymbol{x} = \{\boldsymbol{P}\boldsymbol{x}\}^\intercal \boldsymbol{P}\boldsymbol{x} = \|\boldsymbol{P}\boldsymbol{x}\|^2.$$

Additionally, it is simple to show that

$$|\boldsymbol{x}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})| = |(\boldsymbol{P}\boldsymbol{x})^\intercal \boldsymbol{N}\sigma|$$

for the normal case and

$$|(\boldsymbol{P}\boldsymbol{x})^\intercal \boldsymbol{N}|$$

for the asymptotic logistic case.

5. Represent $\boldsymbol{N}$ in terms of its polar coordinates

$$\boldsymbol{N} = \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{R_N}\cos\boldsymbol{\theta_N} \\ \boldsymbol{R_N}\sin\boldsymbol{\theta_N} \end{pmatrix} = \boldsymbol{R_N} \begin{pmatrix} \cos\boldsymbol{\theta_N} \\ \sin\boldsymbol{\theta_N} \end{pmatrix}, \quad \boldsymbol{R_N} \geq 0, \quad \boldsymbol{\theta_N} \in [0, 2\pi).$$

It can be shown (cf. Ross (1988)) that a standard bivariate normal vector in polar coordinates has the following three properties.

(a) $\boldsymbol{R_N} \sim \sqrt{\mathcal{X}_2^2}$, that is to say $\boldsymbol{R_N}$ follows a standard Rayleigh distribution with distribution function

$$\boldsymbol{F}_{\boldsymbol{R_N}}(x) = \mathrm{P}(X \leq x) = 1 - \exp\left\{-\frac{x^2}{2}\right\}. \tag{2.1}$$

(b) $\boldsymbol{\theta_N}$ follows a uniform distribution over the interval $[0, 2\pi)$.

(c) $\boldsymbol{R_N}$ and $\boldsymbol{\theta_N}$ are independent random variables.

We also take note that

$$\|\boldsymbol{N}\| = \sqrt{(\boldsymbol{R_N}\cos\boldsymbol{\theta_N})^2 + (\boldsymbol{R_N}\sin\boldsymbol{\theta_N})^2} = \boldsymbol{R_N}\sqrt{(\cos\boldsymbol{\theta_N})^2 + (\sin\boldsymbol{\theta_N})^2} = \boldsymbol{R_N}.$$

6. Specifically for the normal-error linear regression model, the polar coordinates of $\boldsymbol{T}$ are

$$\boldsymbol{T} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{R_T}\cos\boldsymbol{\theta_T} \\ \boldsymbol{R_T}\sin\boldsymbol{\theta_T} \end{pmatrix} = \boldsymbol{R_T}\begin{pmatrix} \cos\boldsymbol{\theta_T} \\ \sin\boldsymbol{\theta_T} \end{pmatrix}, \quad \boldsymbol{R_T} \geq 0, \quad \boldsymbol{\theta_T} \in [0, 2\pi).$$

These may be written in terms of $\boldsymbol{N}$ via the simple transformation

$$\boldsymbol{R_T} = \frac{\boldsymbol{R_N}}{\frac{\hat{\sigma}}{\sigma}}, \quad \boldsymbol{\theta_T} = \boldsymbol{\theta_N}.$$

The unchanged angle means that properties 5(b) and 5(c) also hold for $\boldsymbol{T}$. Also, property 5(a) implies that $\boldsymbol{R_T^2}$ is the ratio of two scaled chi square variables. Therefore $\boldsymbol{R_T^2}$ follows a scaled $\mathcal{F}$ distribution, specifically, $\boldsymbol{R_T} \sim \sqrt{2\mathcal{F}_{2,N-2}}$ with cumulative distribution function

$$\boldsymbol{F}_{\boldsymbol{R_T}}(x) = 1 - \left(1 + \frac{x^2}{N-2}\right)^{-\frac{N-2}{2}}.$$

2.2 Two sided Scheffé Type Bands For The Normal-Error Linear Regression Model

In this section we review the construction of the Scheffé type simultaneous confidence band for a multiple covariate model over the whole real space, and for one covariate over a finite interval.

2.2.1 Scheffé Type Band Over The Whole Real Space

We look to establish a band of the form in Equation (1.9) where the value of c sets

$$\mathrm{P}\left\{\boldsymbol{x}^\intercal\boldsymbol{\beta} \in \boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm c\hat{\sigma}\sqrt{\boldsymbol{x}^\intercal(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{x}} \quad \forall x \in \mathbb{R}^q\right\} = 1 - \alpha.$$

We see (cf. Liu (2011)) that

$$
\begin{aligned}
&\mathrm{P}\left\{\boldsymbol{x}^\intercal\boldsymbol{\beta} \in \boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm c\hat{\sigma}\sqrt{\boldsymbol{x}^\intercal(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{x}} \quad \forall x \in \mathbb{R}^q\right\} \\
=&\mathrm{P}\left\{\frac{\left|\boldsymbol{x}^\intercal(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right|}{\hat{\sigma}\sqrt{\boldsymbol{x}^\intercal(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{x}}} \leq c \quad \forall x \in \mathbb{R}^q\right\} \\
=&\mathrm{P}\left\{\sup_{x\in\mathbb{R}^q}\frac{|(\boldsymbol{P}\boldsymbol{x})^\intercal\sigma\boldsymbol{N}|}{\hat{\sigma}\|\boldsymbol{P}\boldsymbol{x}\|} \leq c\right\} \\
=&\mathrm{P}\left\{\frac{\|\boldsymbol{N}\|}{\hat{\sigma}/\sigma}\left(\sup_{x\in\mathbb{R}^q}\frac{|(\boldsymbol{P}\boldsymbol{x})^\intercal\boldsymbol{N}|}{\|\boldsymbol{P}\boldsymbol{x}\|\|\boldsymbol{N}\|}\right) \leq c\right\} \\
=&\mathrm{P}\left\{\frac{\|\boldsymbol{N}\|}{\hat{\sigma}/\sigma} \leq c\right\}, \qquad (2.2)
\end{aligned}
$$

which is due to Point 4 of the preliminaries. Furthermore (2.2) occurs as a result of the Cauchy-Schwartz inequality which states

$$\sup_{x\in\mathcal{R}^q}\frac{|(\boldsymbol{P}\boldsymbol{x})^\intercal\boldsymbol{N}|}{\|\boldsymbol{P}\boldsymbol{x}\|\|\boldsymbol{N}\|} = 1.$$

If we further note that

$$\|\boldsymbol{N}\| = \sqrt{\left(\frac{\boldsymbol{P}^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma}\right)^\intercal\frac{\boldsymbol{P}^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma}} = \sqrt{\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\intercal}{\sigma}\boldsymbol{P}^{-2}\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma}} = \frac{\sqrt{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}}{\sigma},$$

then Equation (2.2) implies that c is set to satisfy

$$
\begin{aligned}
&\mathrm{P}\left\{\frac{\sqrt{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}}{\sigma}\frac{1}{\hat{\sigma}/\sigma} \leq c\right\} = 1-\alpha \\
\Rightarrow&\mathrm{P}\left\{\frac{\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma^2}}{(q+1)}\frac{1}{\hat{\sigma}^2/\sigma^2} \leq \frac{c^2}{(q+1)}\right\} = 1-\alpha.
\end{aligned}
$$

Since, from Chapter one, $\frac{\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma^2}}{(q+1)}\frac{1}{\hat{\sigma}^2/\sigma^2}$ is the quotient of two chi squared random variables, divided by their degrees of freedom, $(q+1)$ and $(N-q-1)$ respectively, it must follow an $\mathcal{F}_{q+1,N-q-1}$ distribution. We therefore have

$$\mathrm{P}\left\{\frac{\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma^2}}{(q+1)}\frac{1}{\hat{\sigma}^2/\sigma^2} \leq f^\alpha_{q+1,N-q-1}\right\} = 1-\alpha, \qquad (2.3)$$

where $f^\alpha_{q+1,N-q-1}$ is the $(1-\alpha)$'th quantile of the $\mathcal{F}_{q+1,N-q-1}$ distribution. It is then immediate that

$$\frac{c^2}{(q+1)} = f^\alpha_{q+1,N-q-1} \implies c = \sqrt{(q+1)f^\alpha_{q+1,N-q-1}}.$$

Thus a Scheffé type confidence band for a normal-error linear model over the real line takes the form,

$$\boldsymbol{x}^\intercal\boldsymbol{\beta} \in \boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm \sqrt{(q+1)f^{\alpha}_{q+1,N-q-1}}\hat{\sigma}\sqrt{\boldsymbol{x}^\intercal(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{x}}, \tag{2.4}$$

where

$$\mathrm{P}\left\{\boldsymbol{x}^\intercal\boldsymbol{\beta} \in \boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm \sqrt{(q+1)f^{\alpha}_{q+1,N-q-1}}\hat{\sigma}\sqrt{\boldsymbol{x}^\intercal(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{x}} \quad \forall x \in \mathbb{R}^q\right\} = 1-\alpha.$$

We illustrate this band with an example dataset modelling blood fat content with age. This is given as file "x09.txt" from Florida State University (2011), with $N = 25$. The value of c is given as 2.616 and the band is shown in Figure 2.1 below.

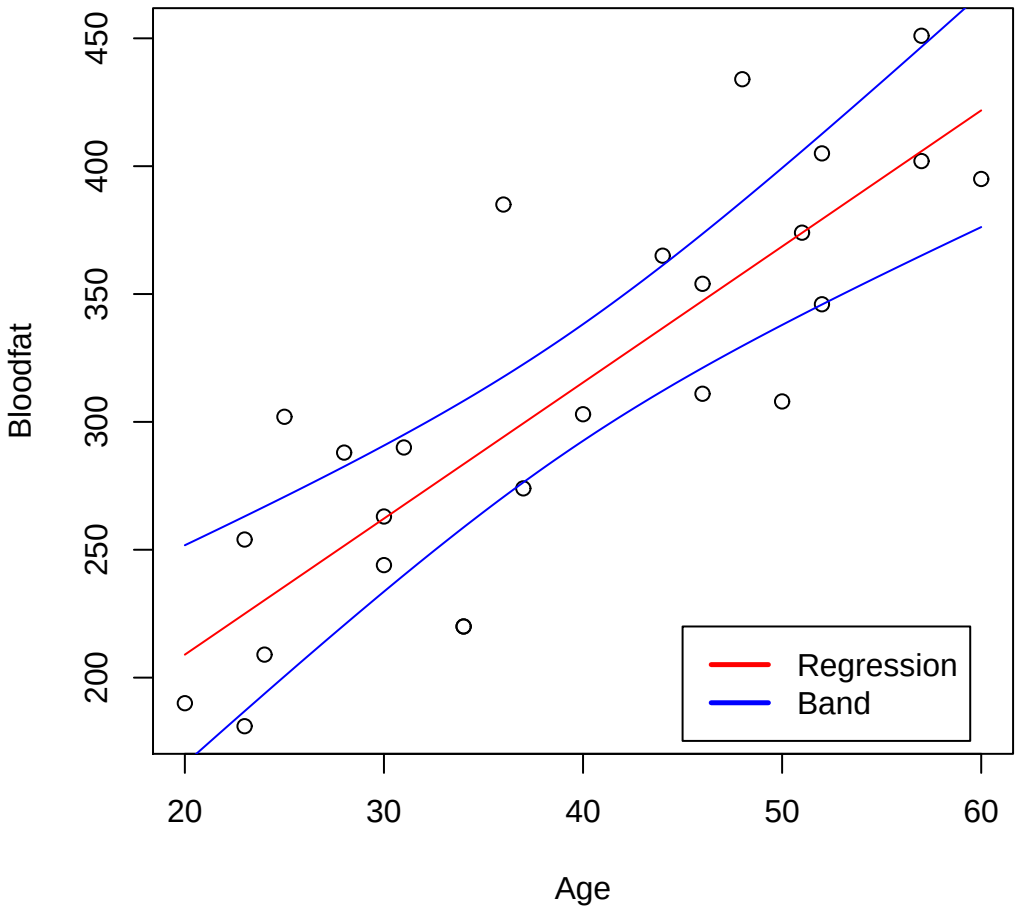


Figure 2.1: Scheffé type simultaneous confidence band for an example dataset modelling blood fat content as a function of age

2.2.2 Scheffé Type Band Over A Finite Covariate Region

A Scheffé type band considered over the whole real space is overly conservative in most practical cases, where the predictors are usually considered over some finite range. It is agreed by many researchers, such as Liu (2011), that it is more informative to consider the band over some q dimensional rectangular region of the covariates. Recall that this is represented by two vectors of constants $\boldsymbol{a}$ and $\boldsymbol{A}$ with entries $a_i \leq A_i$ $i = 1, \ldots, q$, that is $x \in (\boldsymbol{a}, \boldsymbol{A})$. Establishing a Scheffé type band of this form for $q > 1$ is a significant challenge, and is achieved by numeric simulation. Here we construct a Scheffé type band for a single covariate x, constrained over some finite interval (a, A) using the method first found in Wynn and Bloomfield (1971). The band takes the form

$$\beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm c\hat{\sigma}\sqrt{(1,x)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\x\end{pmatrix}}\ \forall x \in (a, A)$$

with c set to satisfy

$$P\{\beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm c\hat{\sigma}\sqrt{(1,x)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\x\end{pmatrix}}\ \forall x \in (a, A)\} = 1 - \alpha.$$

As before note that

$$\begin{aligned}
&P\{\beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm c\hat{\sigma}\sqrt{(1,x)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\x\end{pmatrix}}\ \forall x \in (a, A)\}\\
=&\mathrm{P}\left\{\frac{\left|(1,x)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right|}{\hat{\sigma}\sqrt{(1,x)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\x\end{pmatrix}}} \leq c \quad \forall x \in (a, A)\right\}\\
=&\mathrm{P}\left\{\frac{|(\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix})^\intercal\boldsymbol{N}|}{\frac{\hat{\sigma}}{\sigma}\|\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}\|} \leq c \quad \forall x \in (a, A)\right\}\\
=&\mathrm{P}\left\{\frac{|(\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix})^\intercal\boldsymbol{T}|}{\|\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}\|} \leq c \quad \forall x \in (a, A)\right\}
\end{aligned}$$

with $\boldsymbol{T}$ as defined in the preliminaries. By viewing $\boldsymbol{T}$ and $\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}$ as position vectors from the origin, the term

$$\frac{|\{\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}\}^{\intercal}\boldsymbol{T}|}{\|\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}\|}$$

is the absolute value of the scalar projection of $\boldsymbol{T}$ onto $\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}$. Therefore, the following set

$$R_{h,2}\{x\} = \left\{\boldsymbol{T} : \frac{|\{\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}\}^{\intercal}\boldsymbol{T}|}{\|\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}\|} < c\right\},$$

represents the set of all $\boldsymbol{T}$ of which lie in a region, formed by the two parallel lines that are perpendicular to the directional vector $\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}$ and distance c from the origin. We therefore have

$$= P\left\{\frac{|\{\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}\}^{\intercal}\boldsymbol{T}|}{\|\boldsymbol{P}\begin{pmatrix}1\\x\end{pmatrix}\|} < c; \forall x \in (a, A)\right\} = P\left\{\boldsymbol{T} \in \cap_{x\in(a,A)} R_{h,2}\{x\}\right\}$$

$$= P\left\{\boldsymbol{T} \in R_{h,2}\right\}$$

where $R_{h,2}$ is the intersection of all such sets $R_{h,2}\{x\}$ for $x \in (a, A)$. This takes the form of a spindle region on the two dimensional $\boldsymbol{T}$ plane as illustrated by Liu (2011) in Figure 2.2.

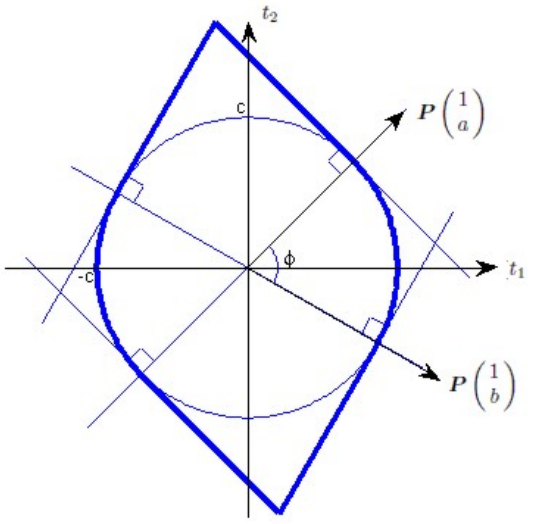


Figure 2.2: The region $R_{h,2}$ with $b = A$ from Liu (2011)

The angle between $\boldsymbol{P}\begin{pmatrix}1\\a\end{pmatrix}$ and $\boldsymbol{P}\begin{pmatrix}1\\A\end{pmatrix}$, ϕ is given as

$$cos(\phi) = \frac{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}{\sqrt{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\a\end{pmatrix}(1,A)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}}.$$

It is clear (cf. Liu (2011)) that the region is rotational invariant around the origin, hence we consider the region $R^*_{h,2}$, shown in Figure 2.3 which rotates rotates $R_{h,2}$ so that ϕ bisects the t_2 axis. We then have

$$P\left\{\boldsymbol{T} \in R_{h,2}\right\} = P\left\{\boldsymbol{T} \in R^*_{h,2}\right\}. \tag{2.5}$$

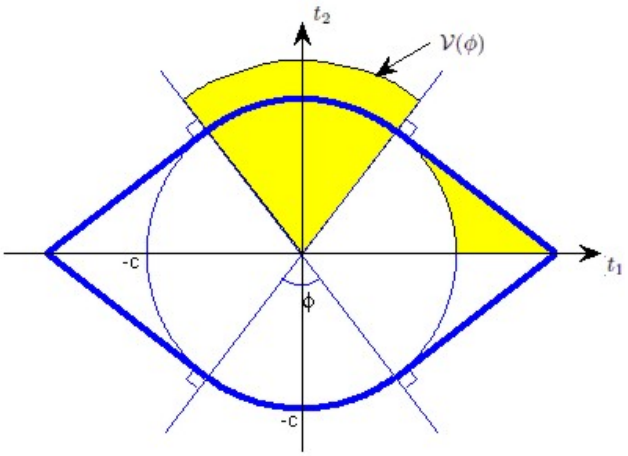


Figure 2.3: The region $R^*_{h,2}$ as taken from Liu (2011)

From Figure 2.3, Equation (2.5) may be represented as the probability of $\boldsymbol{T}$ lying within the circle of radius c, plus four times the probability of $\boldsymbol{T}$ lying within the far right yellow shaded region. Represent $\boldsymbol{T}$ in terms in of its polar coordinates, then, by using the standard formula, the probability of $\boldsymbol{T}$ lying within the circle is

$$P\left\{\sqrt{\boldsymbol{R_T}^2(sin^2(\boldsymbol{\theta_T}) + cos^2(\boldsymbol{\theta_T}))} < c\right\} = P\{\boldsymbol{R_T} < c\}.$$

Recall from the preliminaries that $\boldsymbol{R_T} \sim \sqrt{2\mathcal{F}_{2,v}}$, thus

$$P\{\boldsymbol{R_T} < c\} = \boldsymbol{F}_{\boldsymbol{R_T}}(c)$$

$$= 1 - \left(1 + \frac{c^2}{v}\right)^{-\frac{v}{2}}. \tag{2.6}$$

The probability of $\boldsymbol{T}$ lying within the yellow shaded region comprises three parts

- The angle $\boldsymbol{\theta_T} \in (0, \frac{\pi-\phi}{2})$.
- $\boldsymbol{T}$ lies outside the circle that is
$$\boldsymbol{R_T} > c.$$
- $\boldsymbol{T}$ lies below the straight line in the upper right region. This is given as
$$\begin{aligned}
&\left(cos(\frac{\pi-\phi}{2}), sin(\frac{\pi-\phi}{2})\right) \boldsymbol{R_T} \begin{pmatrix} cos(\boldsymbol{\theta_T}) \\ sin(\boldsymbol{\theta_T}) \end{pmatrix} < c \\
=&\boldsymbol{R_T} cos\left(\frac{\pi-\phi}{2} - \boldsymbol{\theta_T}\right) < c \\
=&\boldsymbol{R_T} < \frac{c}{cos\left(\frac{\pi-\phi}{2} - \boldsymbol{\theta_T}\right)}.
\end{aligned}$$

We therefore have

$$\begin{aligned}
&P\left\{\boldsymbol{T} \in \text{Yellow Region}\right\} \\
=&P\left(\boldsymbol{\theta_T} \in [0, \frac{\pi-\phi}{2}], c < \boldsymbol{R_T} < \frac{c}{cos\left(\frac{\pi-\phi}{2} - \boldsymbol{\theta_T}\right)}\right) \\
=&\frac{1}{2\pi}\int_0^{\frac{\pi-\phi}{2}} P\left(c < \boldsymbol{R_T} < \frac{c}{cos\left(\frac{\pi-\phi}{2} - \boldsymbol{\theta_T}\right)}\right) \mathrm{d}\boldsymbol{\theta_T} \\
=&\frac{1}{2\pi}\int_0^{\frac{\pi-\phi}{2}} \left\{\boldsymbol{F}_{\boldsymbol{R_T}}\left(\frac{c}{cos\left(\frac{\pi-\phi}{2} - \boldsymbol{\theta_T}\right)}\right) - \boldsymbol{F}_{\boldsymbol{R_T}}(c)\right\} \mathrm{d}\boldsymbol{\theta_T} \\
=&\frac{1}{2\pi}\int_0^{\frac{\pi-\phi}{2}} \left\{\left(1 + \frac{c^2}{N-2}\right)^{-\frac{N-2}{2}} - \left(1 + \frac{c^2}{(N-2)sin^2\left(\boldsymbol{\theta_T} + \frac{\phi}{2}\right)}\right)^{-\frac{N-2}{2}}\right\} \mathrm{d}\boldsymbol{\theta_T}.
\end{aligned}$$

This occurs due to the preliminaries that $\boldsymbol{R_T}$ and $\boldsymbol{\theta_T}$ are independent, with $\boldsymbol{\theta_T}$ having a uniform distribution over $(0, 2\pi)$ and $\boldsymbol{R_T} \sim \sqrt{2\mathcal{F}_{2,N-2}}$.

Taking both of these results gives

$$
\begin{aligned}
&P\left\{\boldsymbol{T} \in R_{h,2}\right\} \\
=&1-\left(1+\frac{c^2}{N-2}\right)^{-\frac{N-2}{2}} \\
+&\frac{4}{2\pi}\int_0^{\frac{\pi-\phi}{2}}\left\{\left(1+\frac{c^2}{N-2}\right)^{-\frac{N-2}{2}}-\left(1+\frac{c^2}{(N-2)sin^2\left(\boldsymbol{\theta_T}+\frac{\phi}{2}\right)}\right)^{-\frac{N-2}{2}}\right\}\mathrm{d}\boldsymbol{\theta_T} \\
=&1-\left(1+\frac{c^2}{N-2}\right)^{-\frac{N-2}{2}}+\frac{2}{\pi}\left[\left(1+\frac{c^2}{N-2}\right)^{-\frac{N-2}{2}}\boldsymbol{\theta_T}\right]_0^{\frac{\pi-\phi}{2}} \\
&-\frac{2}{\pi}\int_0^{\frac{\pi-\phi}{2}}\left(1+\frac{c^2}{(N-2)sin^2\left(\boldsymbol{\theta_T}+\frac{\phi}{2}\right)}\right)^{-\frac{N-2}{2}}\mathrm{d}\boldsymbol{\theta_T} \\
=&1-\frac{\phi}{\pi}\left(1+\frac{c^2}{N-2}\right)^{-\frac{N-2}{2}}-\frac{2}{\pi}\int_0^{\frac{\pi-\phi}{2}}\left(1+\frac{c^2}{(N-2)sin^2\left(\theta+\frac{\phi}{2}\right)}\right)^{-\frac{N-2}{2}}\mathrm{d}\theta \qquad (2.7)
\end{aligned}
$$

with dummy variable $\boldsymbol{\theta_T}=\theta$. The critical constant c now sets Equation (2.7) to $1-\alpha$. This constitutes a numeric search involving one dimensional integration and can is performed usng a custom search algorithm using the software package "R".

Figure 2.4 below illustrates the difference between the band considered over the whole real line, and over a finite interval for the "bloodfat" dataset. Over the whole real line we had $c=2.616$. This is reduced to 2.308 when considered for age over 20-35 years. The latter band is thus narrower on average over the considered range, and as a result more informative, at a cost of computational complexity.

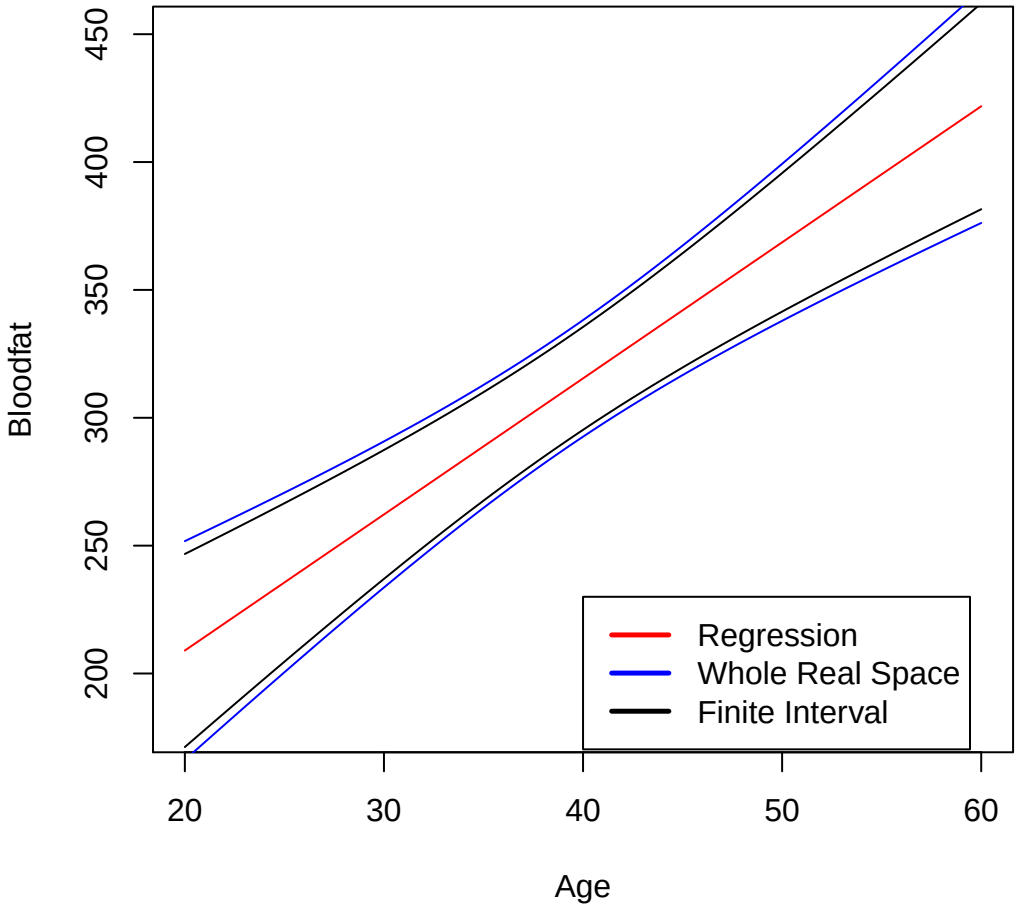


Figure 2.4: A 95 percent Scheffé simultaneous confidence band over the region $[20, 35]$, and over the whole real line for the "bloodfat" dataset.

2.3 Two Sided Constant Width Bands For Normal-Error Linear Regression

In this section we consider the constant width band, which we recall takes the form

$$\beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm \hat{\sigma} c \; \forall x \in (a, A) \tag{2.8}$$

where c is set to satisfy

$$P\left\{\beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm \hat{\sigma} c \; \forall x \in (a, A)\right\} = 1 - \alpha. \tag{2.9}$$

This is a special case of the two sided three segment band of Bowden and Graybill (1966)

$$\beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm \hat{\sigma} \boldsymbol{H_3}(x) \; \forall x \in (a, A) \tag{2.10}$$

with

$$\boldsymbol{H_3}(x) = \tfrac{1}{A-a}\left((x-a)c_1\sqrt{(1,A)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}} + (A-x)c_2\sqrt{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\a\end{pmatrix}}\right).$$

If we set

$$c_2\sqrt{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\a\end{pmatrix}} = c_1\sqrt{(1,A)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}, \tag{2.11}$$

then (2.10) becomes a constant width band with critical constant c, such that

$$c = c_2\sqrt{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\a\end{pmatrix}}.$$

Hence c may be obtained using a method to calculate c_1 and c_2 for three segment bands given first by Bowden and Graybill (1966), and then applying this relation. We require the c_1 and c_2 that satisfy the simultaneous confidence level

$$P\left\{\beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm \hat{\sigma}\boldsymbol{H_3}(x)\ \forall x \in (a,A)\right\} = 1-\alpha,$$

this probability may immediately be written as

$$\begin{aligned}
&P\left\{\frac{|(1,x)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})|}{\hat{\sigma}\boldsymbol{H_3}(x)} < 1\ \forall x \in (a,A)\right\}\\
=&P\left\{\sup_{x\in(a,A)}\frac{|(1,x)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})|}{\hat{\sigma}\boldsymbol{H_3}(x)} < 1\right\}\\
=&P\left\{\frac{|(1,x)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})|}{\hat{\sigma}\boldsymbol{H_3}(x)} < 1 \text{ for } x \in \{a,A\}\right\}.
\end{aligned}$$

This is since from Liu (2011) the function

$$\frac{\partial}{\partial x}\frac{|(1,x)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})|}{\hat{\sigma}\boldsymbol{H_3}(x)}$$

has a fixed sign, either positive or negative over $x \in (-\infty,\infty)$. This implies that the supremum must occur at one of the endpoints of the design range a or A. If we also note that

$$\boldsymbol{H_3}(a) = c_2\sqrt{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\a\end{pmatrix}} \text{ and } \boldsymbol{H_3}(A) = c_1\sqrt{(1,A)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}},$$

then we have

$$
\begin{aligned}
& P\left\{\frac{|(1,x)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})|}{\hat{\sigma}\boldsymbol{H_3}(x)}<1 \text{ for } x\in\{a,A\}\right\} \\
=&\mathrm{P}\left\{\frac{\left|(1,a)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right|}{\hat{\sigma}\sqrt{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\a\end{pmatrix}}}\leq c_2 \ \cap \ \frac{\left|(1,A)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right|}{\hat{\sigma}\sqrt{(1,A)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}}\leq c_1\right\} \\
=&\mathrm{P}\left\{\frac{|(\boldsymbol{P}\begin{pmatrix}1\\a\end{pmatrix})^\intercal\boldsymbol{T}|}{\|\boldsymbol{P}\begin{pmatrix}1\\a\end{pmatrix}\|}\leq c_2 \ \cap \ \frac{|(\boldsymbol{P}\begin{pmatrix}1\\A\end{pmatrix})^\intercal\boldsymbol{T}|}{\|\boldsymbol{P}\begin{pmatrix}1\\A\end{pmatrix}\|}\leq c_1\right\} \\
=&P\left\{\boldsymbol{T}\in R_{3,2}\right\}, \qquad (2.12)
\end{aligned}
$$

where $R_{3,2}=R_{3,2}(a)\cap R_{3,2}(A)$ with

$$
R_{3,2}(a)=\left\{\boldsymbol{T}:\frac{|\{\boldsymbol{P}\begin{pmatrix}1\\a\end{pmatrix}\}^\intercal\boldsymbol{T}|}{\|\boldsymbol{P}\begin{pmatrix}1\\a\end{pmatrix}\|}<c_2\right\}
$$

and

$$
R_{3,2}(A)=\left\{\boldsymbol{T}:\frac{|\{\boldsymbol{P}\begin{pmatrix}1\\A\end{pmatrix}\}^\intercal\boldsymbol{T}|}{\|\boldsymbol{P}\begin{pmatrix}1\\A\end{pmatrix}\|}<c_1\right\}.
$$

Similarly to the hyperbolic case, the probability of interest is reduced to the probability that $\boldsymbol{T}$ lies within the intersection of the two striped regions $R_{3,2}(a)$ and $R_{3,2}(A)$, as show in Figure 2.5, wth the angle ϕ the same as in the Scheffé band case.

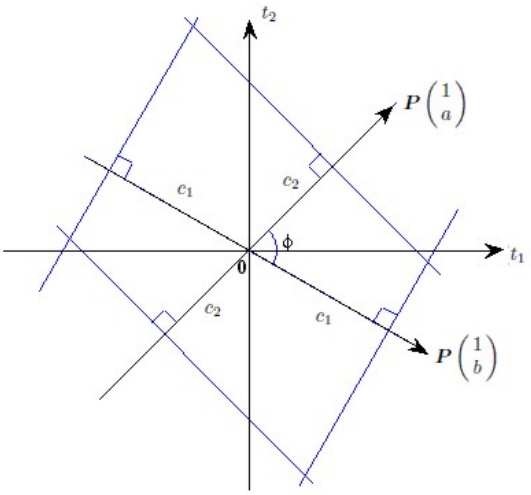


Figure 2.5: The region $R_{3,2}$, image taken from Liu (2011)

This region is also rotationally invariant around the origin, hence we may rotate this region so that the angle ϕ, sits on the t_1 axis to form the region $R^*_{3,2}$. We then have

$$P\left\{\boldsymbol{T} \in R_{3,2}\right\} = P\left\{\boldsymbol{T} \in R^*_{3,2}\right\}.$$

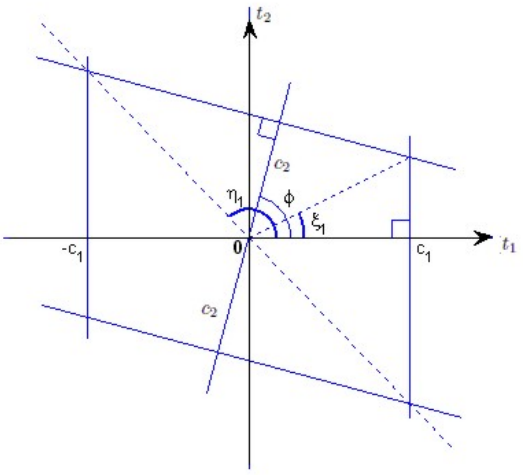


Figure 2.6: The region $R^*_{3,2}$, image taken from Liu (2011).

We define the two angles ξ_1 and η_1, shown in Liu (2011) as follows,

$$\xi_1 = arcsin\left(\frac{c_2 - c_1 cos(\phi)}{\sqrt{c_2^2 + c_1^2 - 2c_1c_2cos\phi}}\right)$$

$$\eta_1 = arccos\left(\frac{-c_1 sin(\phi)}{\sqrt{c_2^2 + c_1^2 + 2c_1c_2cos\phi}}\right).$$

The dotted lines on Figure 2.6 isolate an upper and right region of $R^*_{3,2}$. By symmetry, the relevant probability is then twice the sum of the probability of $\boldsymbol{T}$ lying in the upper and right regions respectively. It is clear from Figure 2.6 that,

$$\begin{aligned} &P\{\boldsymbol{T} \in \text{Right Region}\} \\ =&P\{\boldsymbol{\theta_T} \in (\eta - \pi, \xi_1), \boldsymbol{R_T} cos(\boldsymbol{\theta_T}) \leq c_1\}. \end{aligned}$$

Furthermore

$$\begin{aligned} &P\{\boldsymbol{T} \in \text{Upper Region}\} \\ =&P\{\boldsymbol{\theta_T} \in (\xi_1, \eta_1), \boldsymbol{R_T} cos(\boldsymbol{\theta_T} - \phi) \leq c_2\}. \end{aligned}$$

This can be seen by rotating $R^*_{3,2}$ clockwise by π, so that the upper region is perpendicular with the t_1 axis. We therefore have

$$\begin{aligned} &P\{\boldsymbol{T} \in R^*_{3,2}\} \\ =&2P\{\boldsymbol{\theta_T} \in (\eta - \pi, \xi_1), \boldsymbol{R_T} cos(\boldsymbol{\theta_T}) \leq c_1\} + 2P\{\boldsymbol{\theta_T} \in (\xi_1, \eta_1), \boldsymbol{R_T} cos(\boldsymbol{\theta_T} - \phi) \leq c_2\} \\ =&2\left[\int_{\eta_1 - \pi}^{\xi_1} \frac{1}{2\pi} P\{\boldsymbol{R_T} cos(\boldsymbol{\theta_T}) \leq c_1\} \mathrm{d}\theta_{\boldsymbol{T}} + \int_{\xi_1 - \phi}^{\eta_1 - \phi} \frac{1}{2\pi} P\{\boldsymbol{R_T} cos(u) \leq c_2\} \mathrm{d}u\right] \end{aligned}$$

with $u = \boldsymbol{\theta_T} - \phi$. As $\boldsymbol{\theta_T}$ and u are variables of integration, they are replaced by dummy variable θ. Using the preliminaries we then have

$$\begin{aligned} &P\{\boldsymbol{T} \in R^*_{3,2}\} \\ =&\frac{1}{\pi}\left\{\int_{\eta_1 - \pi}^{\xi_1} \left(1 - \left(1 + \frac{c_1^2}{(N-2)cos^2\theta}\right)^{-\frac{N-2}{2}}\right) \mathrm{d}\theta\right\} \\ +&\frac{1}{\pi}\left\{\int_{\xi_1 - \phi}^{\eta_1 - \phi} \left(1 - \left(1 + \frac{c_2^2}{(N-2)cos^2\theta}\right)^{-\frac{N-2}{2}}\right) \mathrm{d}\theta\right\}. \end{aligned} \tag{2.13}$$

We use a numeric search algorithm to calculate the c_1 and c_2 which sets Equation (2.13) to $1 - \alpha$ under the relation in (2.11). The value of c which satisfies Equation (2.9) then takes the form $c = c_2\sqrt{(1, a)(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}\begin{pmatrix} 1 \\ a \end{pmatrix}}$. We give an example of this band for the "bloodfat" dataset of Figure 2.1, for $x \in (20, 40)$. The critical constant is calculated using Equation (2.13) in "R" code, and is available on request.

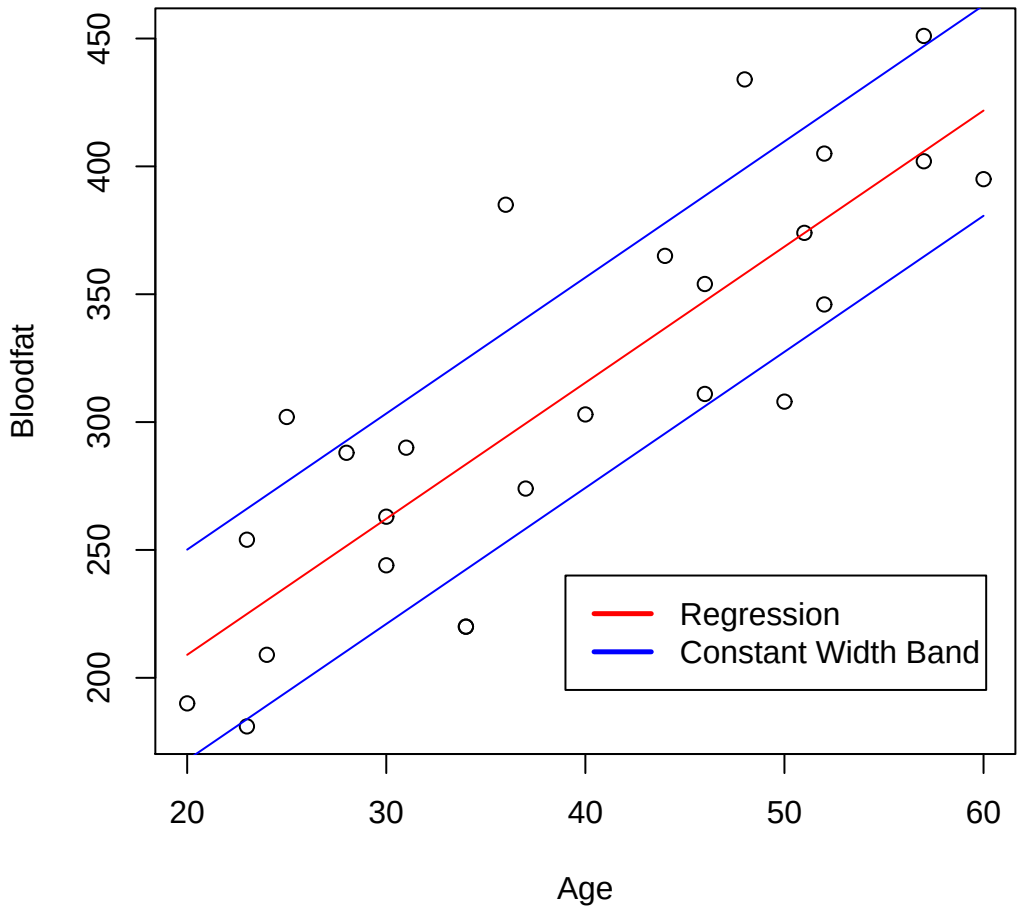


Figure 2.7: An example of a 95 percent constant width band over $(20, 60)$.

2.4 Optimality Criteria For Simultaneous Confidence Bands

An optimality criterion is a function for which its value assesses how good a statistical construct is in eliciting a property of interest. A construct is considered optimal if it minimises, or maximises the criterion function, compared to all other constructs of interest. Optimality criteria are widely applied to design of experiment problems, seeking to establish the best way to construct the matrix $\boldsymbol{X}$ of Equation (1.4), in order to obtain the best inference in regression analysis. This is extensively studied in works such as Fedorov (1972). For confidence bands, there exist two optimality criteria which are covered in this section.

2.4.1 Average Width Criterion

First considered by Gafarian (1964) and formalised by Naiman (984a), the AW criterion is based on the idea that a band that is narrower on average more tightly bounds the area in which $\boldsymbol{x}^\intercal\boldsymbol{\beta}$ could lie, and thus offers a better idea of its real value. A confidence band is therefore considered optimal if it has the smallest average width.

The average width of a Scheffé type band with corresponding critical constant c

over $[a, A]$ is given as

$$2c\hat{\sigma}\frac{1}{A-a}\int_a^A \sqrt{\boldsymbol{x}^\intercal(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{x}}\mathrm{d}x,$$

whilst constant width bands with a corresponding c have a fixed width regardless of the considered range at

$$2c\hat{\sigma}.$$

2.4.2 Minimum Area Confidence Set

The AW criterion was explored by Liu and Hayter (2007), and was critical of the fact that the range of interest can be too crucial a factor in determining the optimal simultaneous confidence band. This was demonstrated with an example involving two sided, three segment bands. Liu and Hayter (2007) instead proposed judging a simultaneous confidence band specifically by its underlying $1-\alpha$ confidence set for $\boldsymbol{\beta}$. This is defined as those $\boldsymbol{\beta}$ for which $\boldsymbol{x}^\intercal\boldsymbol{\beta}$ lies within the $1-\alpha$ simultaneous confidence band, that is

$$C = \left\{\boldsymbol{\beta} : \boldsymbol{x}^\intercal\boldsymbol{\beta} \in \boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm cg(x)\ \forall x \in [a, A]\right\} \tag{2.14}$$

with $\boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm cg(x)$ the form of a specific simultaneous confidence band. To this end, the paper defined the Minimum Area Confidence Set (MACS) for a univariate linear model. A confidence band is considered optimal under MACS if it minimises the area of this confidence set. A smaller confidence set by area reduces the number of candidates for $\boldsymbol{\beta}$, and is as such more informative.

It is clear from Chapter two that for Scheffé type and constant width type bands, the confidence set may be expressed as

$$\boldsymbol{C}_{h,2}(\hat{\boldsymbol{\beta}}, \hat{\sigma}) = \left\{\boldsymbol{\beta} : \frac{\boldsymbol{P}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\hat{\sigma}} \in R_{h,2}\right\}$$

and

$$\boldsymbol{C}_{3,2}(\hat{\boldsymbol{\beta}}, \hat{\sigma}) = \left\{\boldsymbol{\beta} : \frac{\boldsymbol{P}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\hat{\sigma}} \in R_{3,2}\right\}$$

respectively. Hence the area of each confidence set is related to areas of $R_{h,2}$ and $R_{3,2}$ via the linear transformation $\frac{\boldsymbol{P}^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\hat{\sigma}}$, and easily allows direct comparison of different band types (cf. Liu and Hayter (2007)).

The area of $\boldsymbol{C}_{h,2}$ can be given (cf. Ah-Kine (2010)) as

$$\begin{aligned}
Area(\boldsymbol{C}_{h,2}) &= \int\int_{\boldsymbol{\beta}\in\boldsymbol{C_{h,2}}=\hat{\boldsymbol{\beta}}+\hat{\sigma}\boldsymbol{P}R_{h,2}} \mathbf{1}\mathrm{d}\boldsymbol{\beta} \\
&= \int\int_{\boldsymbol{w}\in\hat{\sigma}R_{h,2}} |\boldsymbol{P}|\,\mathrm{d}\boldsymbol{w} \quad (\boldsymbol{w}=\boldsymbol{P}^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})) \\
&= |\boldsymbol{P}|\,\hat{\sigma}^2 \int\int_{\boldsymbol{p}\in R_{h,2}} \mathbf{1}\mathrm{d}\boldsymbol{p} \quad (\boldsymbol{p}=\frac{\boldsymbol{w}}{\hat{\sigma}}) \\
&= |\boldsymbol{P}|\,\hat{\sigma}^2 Area(R_{h,2}) \\
&= |\boldsymbol{P}|\hat{\sigma}^2 c^2 \left\{\phi + 2cot\left(\frac{\phi}{2}\right)\right\}
\end{aligned}$$

as the area of $R_{h,2}$ can be shown to be $c^2\left\{\phi + 2cot\left(\frac{\phi}{2}\right)\right\}$, where c and ϕ are as defined in Chapter two for Scheffé type bands. The area of $R_{3,2}$ can be shown to be, using the area of a parallelogram, $4\frac{c_1c_2}{sin(\phi)}$, and applying the same method gives

$$\begin{aligned}
\text{Area}(\boldsymbol{C}_{3,2}) =& \hat{\sigma}^2|\boldsymbol{P}|\left(4\frac{c_1c_2}{sin(\phi)}\right) \\
=& \hat{\sigma}^2|\boldsymbol{P}|\left(\frac{4c^2}{sin(\phi)\sqrt{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\a\end{pmatrix}(1,A)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}}\right),
\end{aligned}$$

substituting the values of $c_1 = \frac{c}{\sqrt{(1,A)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}}$ and $c_2 = \frac{c}{\sqrt{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\a\end{pmatrix}}}$ for the constant width band.

2.4.3 Relation between D-Optimal and MACS Criteria

Perhaps the most well known optimality criterion in general, D optimality is a criterion specifically designed to compare design of experiments. A design is considered optimal under D optimality if it maximises the determinant of the information matrix $|\boldsymbol{X}^\intercal\boldsymbol{X}|$. D optimal designs minimise the area of the implicit confidence ellipsoid for $\boldsymbol{\beta}$, formed by the least squares estimate $\hat{\boldsymbol{\beta}}$. From Equation (2.3), we have

$$\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\intercal(\boldsymbol{X}^\intercal\boldsymbol{X})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{(q+1)\hat{\sigma}^2} \sim \mathcal{F}_{q+1,N-q-1}.$$

This immediately implies a $1-\alpha$ confidence set on $\boldsymbol{\beta}$, the area of which D optimality seeks to minimise, as

$$S_{\boldsymbol{\beta}} = \{\boldsymbol{\beta} : (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\intercal(\boldsymbol{X}^\intercal\boldsymbol{X})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \leq (q+1)\hat{\sigma}^2 f^{\alpha}_{q+1,N-q-1}\}.$$

The area of this set is always proportional to $\frac{1}{\sqrt{|\boldsymbol{X}^\intercal \boldsymbol{X}|}}$ for a model with any number of covariates q. An example is given in Ah-Kine (2010) for $q = 1$. Hence the area is minimised with respect to the design when $\sqrt{|\boldsymbol{X}^\intercal \boldsymbol{X}|}$ or equivalently $|\boldsymbol{X}^\intercal \boldsymbol{X}|$ is maximised.

This criterion function leads to one of the few results for optimal designs for simultaneous confidence bands. It is clear that MACS and D optimality are closely related criteria, both looking to minimise the area of a confidence set for $\boldsymbol{\beta}$. For a Scheffé type band over the whole real space $c = \sqrt{(q+1)f^{\alpha}_{q+1,N-q-1}}$, hence the MACS criterion looks to minimise the same $\mathcal{F}$ distributed confidence set as D optimality. If we view $\boldsymbol{C}_{h,2}(\hat{\boldsymbol{\beta}}, \hat{\sigma})$ as the set of the form of Equation (2.14) for a Scheffé type band over the whole real space, then we have

$$
\begin{aligned}
&\boldsymbol{C}_{h,2}(\hat{\boldsymbol{\beta}}, \hat{\sigma}) \\
&= \left\{ \boldsymbol{\beta} : \beta_0 + \beta_1 x \in \hat{\beta}_0 + \hat{\beta}_1 x \pm \sqrt{(q+1)f^{\alpha}_{q+1,N-q-1}}\hat{\sigma}\sqrt{\boldsymbol{x}^\intercal (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}\boldsymbol{x}}\ \forall x \in (\boldsymbol{a}, \boldsymbol{A}) \right\} \\
&= \left\{ \boldsymbol{\beta} : (\boldsymbol{x}^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))^\intercal (\boldsymbol{x}^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \leq (q+1)f^{\alpha}_{q+1,N-q-1}\hat{\sigma}^2 \boldsymbol{x}^\intercal (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}\boldsymbol{x}\ \forall x \in (\boldsymbol{a}, \boldsymbol{A}) \right\} \\
&= \left\{ \boldsymbol{\beta} : \boldsymbol{x}^\intercal \boldsymbol{x} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\intercal \boldsymbol{X}^\intercal \boldsymbol{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq (q+1)f^{\alpha}_{q+1,N-q-1}\hat{\sigma}^2 \boldsymbol{x}^\intercal \boldsymbol{x}\ \forall x \in (\boldsymbol{a}, \boldsymbol{A}) \right\} \\
&= \left\{ \boldsymbol{\beta} : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\intercal \boldsymbol{X}^\intercal \boldsymbol{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq (q+1)f^{\alpha}_{q+1,N-q-1}\hat{\sigma}^2 \right\} = S_{\boldsymbol{\beta}}.
\end{aligned}
$$

Therefore in this instance D and MACS optimality are equivalent criteria. It is much more difficult to establish a similar equivalence for bands over a finite interval as c is not constant in terms of the design. This is the problem that is considered in the following Chapter.

2.5 Two Sided Scheffé Type Band For Generalised Linear Models

2.5.1 Two Sided Scheffé Type Band For The Logistic Regression Model

We construct a simultaneous confidence band of the form given by Section 1.5 for the logistic regression model, originally considered by Brand et al. (1973) for $q = 1$, and generalised to a multiple logistic model by Walter (1983). Recall for the logistic model that $b(\mu(x)) = \pi(p(x)) = log\left(\frac{p(x)}{1-p(x)}\right)$, hence the band takes the form

$$\pi(p(x)) = \boldsymbol{x}^\intercal \boldsymbol{\beta} \in \boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \pm c\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1}\boldsymbol{x}}\ \forall\ x \in \mathbb{R}^q$$

where c sets

$$\mathrm{P}\left\{ \boldsymbol{x}^\intercal \boldsymbol{\beta} \in \boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \pm c\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1}\boldsymbol{x}}\ \forall\ x \in \mathbb{R}^q \right\} = 1 - \alpha.$$

An equivalent $1 - \alpha$ confidence band for $p(x)$ is then given by

$$p(x) \in \frac{\exp\left(\boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \pm c\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1}\boldsymbol{x}}\right)}{1 + \exp\left(\boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \pm c\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1}\boldsymbol{x}}\right)}\ \forall\ x \in \mathbb{R}^q.$$

Under the asymptotic normality assumption of section 1.4.3 we have as similar to the equivalent normal-error linear simultaneous confidence band

$$\begin{aligned}
&\mathrm{P}\left\{\boldsymbol{x}^\intercal\boldsymbol{\beta} \in \boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm c\sqrt{\boldsymbol{x}^\intercal\boldsymbol{J}^{-1}\boldsymbol{x}}\ \forall\ x \in \mathbb{R}^q\right\} \\
&= \mathrm{P}\left\{\sup_{x\in\mathbb{R}^q} \frac{|(\boldsymbol{P}\boldsymbol{x})^\intercal\boldsymbol{N}|}{\|\boldsymbol{P}\boldsymbol{x}\|} \leq c\right\} \\
&= \mathrm{P}\left\{\|\boldsymbol{N}\|\left(\sup_{x\in\mathbb{R}^q} \frac{|(\boldsymbol{P}\boldsymbol{x})^\intercal\boldsymbol{N}|}{\|\boldsymbol{P}\boldsymbol{x}\|\|\boldsymbol{N}\|}\right) \leq c\right\} \\
&= \mathrm{P}\left\{\|\boldsymbol{N}\| \leq c\right\}
\end{aligned}$$

by the Cauchy-Schwartz inequality. Since $\boldsymbol{N}$ has approximately the multivariate normal distribution $N(\boldsymbol{0}, \boldsymbol{I}_{q+1})$, then $\|\boldsymbol{N}\|^2$ has approximately a chi square distribution with $q+1$ degrees of freedom, and therefore $c = \sqrt{\mathcal{X}^\alpha_{q+1}}$, where $\mathcal{X}^\alpha_{q+1}$ is the upper α point of the chi square distribution. Hence an approximate $1-\alpha$ confidence band takes the form

$$\boldsymbol{x}^\intercal\boldsymbol{\beta} \in \boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm \sqrt{\mathcal{X}^\alpha_{q+1}}\sqrt{\boldsymbol{x}^\intercal\boldsymbol{J}^{-1}\boldsymbol{x}}\ \forall\ x \in \mathbb{R}^q, \tag{2.15}$$

where

$$\mathrm{P}\left\{\boldsymbol{x}^\intercal\boldsymbol{\beta} \in \boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm \sqrt{\mathcal{X}^\alpha_{q+1}}\sqrt{\boldsymbol{x}^\intercal\boldsymbol{J}^{-1}\boldsymbol{x}}\ \forall\ x \in \mathbb{R}^q\right\} = 1-\alpha. \tag{2.16}$$

This is of course in practice an approximate $1-\alpha$ simultaneous confidence band, under an asymptotically normal MLE. An equivalent band is then given as

$$p \in \frac{\exp\left(\boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm \sqrt{\mathcal{X}^\alpha_{q+1}}\sqrt{\boldsymbol{x}^\intercal\boldsymbol{J}^{-1}\boldsymbol{x}}\right)}{1+\exp\left(\boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} \pm \sqrt{\mathcal{X}^\alpha_{q+1}}\sqrt{\boldsymbol{x}^\intercal\boldsymbol{J}^{-1}\boldsymbol{x}}\right)}\ \forall\ x \in \mathbb{R}^q. \tag{2.17}$$

This band is demonstrated for an example dataset with a Bernoulli response, found in the software package "R", called "mtcars".

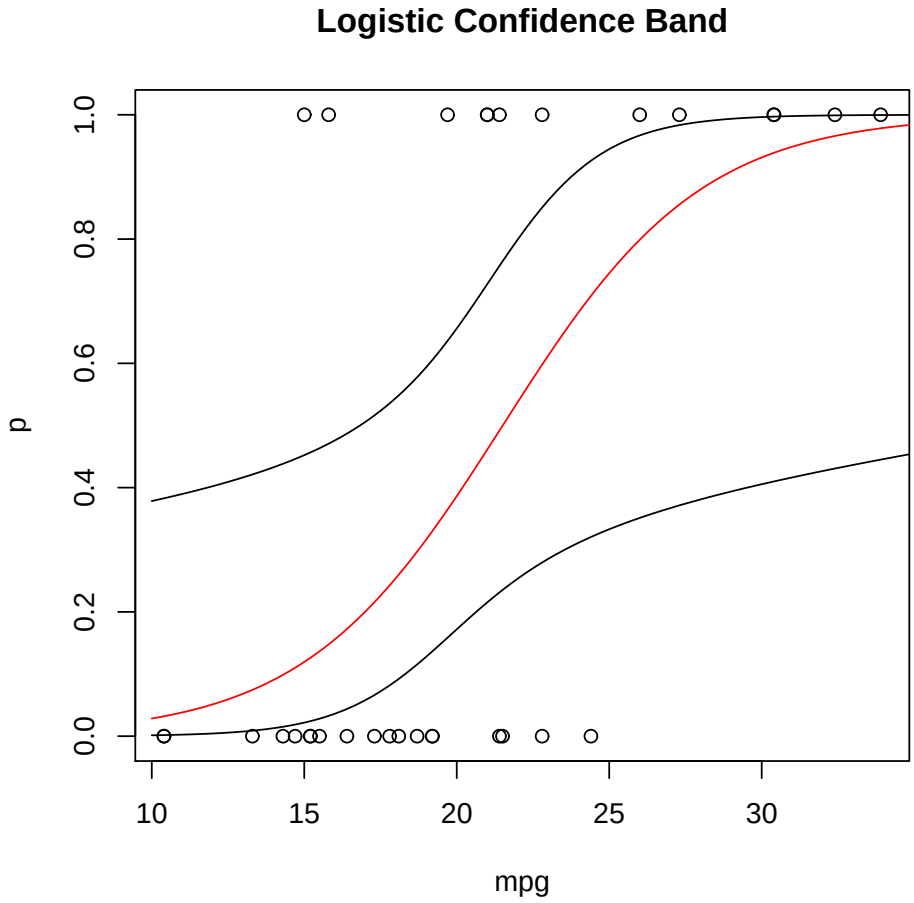


Figure 2.8: An approximate 95 percent logistic simultaneous confidence band for the "mtcars" dataset.

2.5.2 A Note On Two Sided Scheffé Type Bands for Other GLM's

To generate a simultaneous confidence band for any generalised linear model, with a linear regression line of some form $b(\mu(x)) = \boldsymbol{x}^\intercal \boldsymbol{\beta}$, we require only that the MLE of $\boldsymbol{\beta}$ is asymptotically normal, with distribution $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \mathcal{N}_{q+1}(\mathbf{0}, \boldsymbol{V})$, and $\boldsymbol{V}$ some covariance matrix. Since the large sample asymptotic normality of the MLE $\hat{\boldsymbol{\beta}}$ in Section 1.2.4 always holds true regardless of the underlying model, simultaneous confidence bands can be constructed for any linear GLM, in the exact same way as the logistic regression model.

Chapter 3

Optimal Continuous Designs For Simultaneous Confidence Bands Over A Finite Interval

Optimal design is the study of how best to assign the fixed values of the covariates in an experiment. The early groundwork is usually attributed to two papers, the very early work of Smith (1918), and that of Elfving (1952). However, perhaps the most well known papers in the area are the works of Kiefer, in particular Kiefer (1958), Kiefer (1959) and Kiefer and Wolfowitz (1960) which established the equivalence of D and G optimality criteria. In particular, these are noted for forming the style, and structure of most modern optimal design problems, in particular introducing the alphabetical labeling of optimality criteria than is now standard.

In this Chapter, we detail work on establishing the optimal designs for the univariate Scheffé type and constant width type bands, when the band, and the design lie in some finite region $x \in [a, A]$. We consider both the average width (AW) and the minimum area confidence set (MACS) criteria, detailed in Chapter two, as optimality criterion functions. A numeric approach is considered, that identifies the optimal designs as D-optimal for 95% bands over $[-1, 1]$, up to some value of N. Also considered are the best designs with design points lying on the ends of the considered range. Based on this, an analytical approach is attempted using methods given in Silvey (1980), that are based on the equivalence theorem of Kiefer and Wolfowitz (1960), but is unsuccessful. In each case the problem is approached from a continuous design standpoint.

3.1 Introduction

3.1.1 Overview

Recall that the N experimental observations of x, defined in (1.3), represent the controllable input by the user into the analysis. They are termed the **design of the experiment**, and are characterised by the design matrix $\boldsymbol{X}$. In optimal design, the objective is to judge designs against their ability to elicit a specific statistical property of interest, assessed through an **optimality criterion function**. The best design minimises, or maximises, this function. An important property of all criterion functions, including the AW and MACS criteria, are that they are functions of the design through the information matrix $\boldsymbol{X}^\intercal\boldsymbol{X}$. Some well known criteria include, but are not limited to D, G, I and A optimality criteria. We refer to Fedorov (1972) as an excellent reference point for work in the area.

Optimal design problems for simultaneous confidence bands have an additional complication. A design is considered optimal under the AW or MACS criteria, if the corresponding $1-\alpha$ simultaneous confidence band is optimal, compared to the $1-\alpha$ simultaneous confidence bands formed using all other considered designs. Hence in practice this constitutes an optimisation problem with respect to a constraint, that the critical constant c is always set to maintain the simultaneous coverage of the band at $1-\alpha$. That is for a simultaneous confidence band with bounds $l(x,c) < u(x,c)$ and criterion function $f(\boldsymbol{X}^\intercal\boldsymbol{X}, c)$, the problem is broadly given as

$$\text{Optimise } f(\boldsymbol{X}^\intercal\boldsymbol{X}, c)$$

$$\text{Such that P}\left\{l(x,c) < \boldsymbol{x}^\intercal\boldsymbol{\beta} < u(x,c)\ \forall x \in (\boldsymbol{a}, \boldsymbol{A})\right\} = 1-\alpha.$$

For Scheffé type bands over the whole real space, and two sided two segment bands, the critical constant c is independent of the design. In this instance obtaining optimal designs is a relatively simple problem with the optimal designs already available, and coinciding with the D optimal designs for the corresponding model (cf. Ah-Kine (2010)). However, as shown in Chapter 2, the simultaneous coverage probability expressions of Scheffé type and three segment bands over a finite interval are complicated functions of the design involving integration. In this instance c will change in a non obvious way with the design and makes understanding the behaviour of any criterion function significantly more difficult. There is currently no analytical work for optimal designs under these circumstances. This Chapter looks to address this problem by applying it to a continuous design framework.

3.1.2 Design Of Experiments: Exact and Continuous Designs

For one covariate, the simplest way to define a design is to map directly the value of x for each run $1, 2, \ldots, N$ of an experiment. This is an exact design, where we take a number of distinct points $n \leq N$ on the covariate, $x_1 \ldots x_n$, each given a number of

replications $r_1 \dots r_n \ \in \mathbb{Z}_+$ in the experiment such that $\sum_{i=1}^{n} r_i = N$. This defines an exact design δ as

$$\delta = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ r_1 & r_2 & \dots & r_n \end{array} \right\}. \tag{3.1}$$

The set of distinct points $x_1 \dots x_n$ are known as the **design points** of the experiment. An exact design relates to the design matrix from Chapter 1, $\boldsymbol{X}$ in the obvious way

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} r_1 \text{times} \\ \vdots \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} r_n \text{times} \end{matrix}.$$

. Let $\boldsymbol{x_i} = \begin{pmatrix} 1 \\ x_i \end{pmatrix}$, we may then relate an exact design δ to the information matrix as

$$\boldsymbol{M}\{\delta\} = \boldsymbol{X}^\intercal \boldsymbol{X} = \begin{pmatrix} \sum_{i=1}^{n} r_i & \sum_{i=1}^{n} r_i x_i \\ \sum_{i=1}^{n} r_i x_i & \sum_{i=1}^{n} r_i x_i^2 \end{pmatrix} = \sum_{i=1}^{n} r_i \boldsymbol{x_i} \boldsymbol{x_i}^\intercal.$$

Exact designs are limited in that they are reliant on the value on N. An optimal design under an exact framework typically changes based on the value, and particularly the parity of N. Outside of specific instances of distributional results, solutions are typically found numerically. In this thesis we consider the more general measure of a continuous design instead. In this case assigns a weight $0 < p_i \leq 1$ to each design point x_i, indicating the proportion of runs to be allocated to the design point in an exact framework as follows

$$\eta = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{array} \right\} \tag{3.2}$$

where $\sum_{i=1}^{n} p_i = 1$. A continuous design has no dependence on N, and the p_i, unlike the r_i are not constrained to be integers. This format does not have a direct relation to $\boldsymbol{X}$ but we can define a weighted continuous information matrix as

$$\boldsymbol{M}\{\eta\} = \sum_{i=1}^{n} p_i \boldsymbol{x_i} \boldsymbol{x_i}^\intercal = \begin{pmatrix} \sum_{i=1}^{n} p_i & \sum_{i=1}^{n} p_i x_i \\ \sum_{i=1}^{n} p_i x_i & \sum_{i=1}^{n} p_i x_i^2 \end{pmatrix}.$$

Note we may covert any exact design δ (cf. Silvey (1980)) to an equivalent continuous design by setting $p_i = \frac{r_i}{N}$, which gives

$$\eta_N = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ & & & \\ \frac{r_1}{N} & \frac{r_2}{N} & \dots & \frac{r_n}{N} \end{array} \right\} \tag{3.3}$$

where clearly $\sum_{i=1}^{n} \frac{r_i}{N} = 1$. Hence we may broadly state the exact and continuous information matrices are proportional to one another through the relation

$$\boldsymbol{M}\{\delta\} = \boldsymbol{X}^\intercal \boldsymbol{X} = N\boldsymbol{M}\{\eta_N\}.$$

As optimality criteria are functions of the information matrix, this relation means that optimal continuous design problems can be solved by simply inserting $\boldsymbol{M}\{\eta\}$ into the criterion function in replacement of $\boldsymbol{X}^\intercal \boldsymbol{X}$. Solutions for continuous designs are a more general concept and are often preferred to an exact method. Furthermore, powerful analytical methods are available using functional derivatives (cf. Silvey (1980)), that are shown later in the Chapter.

3.1.3 Support Of The Design

All experimental designs in practice are considered over some specific range, $[a, A]$, of the covariate known as the **support** of the design, or otherwise the design range. For design problems for simultaneous confidence bands, it is sensible to assume that the covariate range of the band, and the support are the same. We note that a design over some support $[a, A]$ can be converted to an equivalent design over support $[-1, 1]$, via a simple linear transformation without loss of generality. Hence, the work of this chapter will focus on designs with support $[-1, 1]$, but will also consider designs over the support $[2, 5]$, as optimal designs are not guaranteed (see Kotz and Johnson (1992)) to be invariant of this linear transformation.

3.2 AW and MACS Criteria For Continuous Designs

To search for optimal continuous designs for simultaneous confidence bands, we view the criteria as functions of $\boldsymbol{M}\{\eta\}$, instead of $\boldsymbol{X}^\intercal \boldsymbol{X}$. Noting that $\hat{\sigma}$ has no dependence on the design, it may be treated as a constant. Hence the relevant continuous AW criterion functions for Scheffé type and constant width type bands are

$$c \int_a^A \sqrt{\boldsymbol{x}^\intercal (\boldsymbol{M}\{\eta\})^{-1} \boldsymbol{x}} \mathrm{d}x \tag{3.4}$$

and

$$c \tag{3.5}$$

respectively. For the MACS criteria, we obtain

$$|\boldsymbol{M}\{\eta\}^{-\frac{1}{2}}|c^2\left\{\phi + 2cot\left(\frac{\phi}{2}\right)\right\} \tag{3.6}$$

and

$$|\boldsymbol{M}\{\eta\}^{-\frac{1}{2}}|\left(\frac{c^2}{sin(\phi)\sqrt{(1,a)(\boldsymbol{M}\{\eta\})^{-1}\begin{pmatrix}1\\a\end{pmatrix}(1,A)(\boldsymbol{M}\{\eta\})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}}\right) \tag{3.7}$$

respectively. For all criteria, c sets the simultaneous coverage probability to $1-\alpha$, using Equations (2.7) and (2.13) respectively. Recall that the AW criterion function represents an overall measure of the width of the simultaneous confidence band. A smaller average width offers a more tightly bounded region, for which to consider candidates for the true linear regression line. The MACS criterion function measures the size of the implicit confidence set on $\boldsymbol{\beta}$, formed by a simultaneous confidence band, with a smaller set reducing the number of candidates for $\boldsymbol{\beta}$. Hence, in each case, the optimal continuous design minimises the criterion function. Note that we clearly have through the relation $\boldsymbol{X}^\intercal\boldsymbol{X} = N\boldsymbol{M}\{\eta\}$

$$\begin{aligned} cos(\phi) =& \frac{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}{\sqrt{(1,a)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\a\end{pmatrix}(1,A)(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}} \\ =& \frac{(1,a)(\boldsymbol{M}\{\eta\})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}{\sqrt{(1,a)(\boldsymbol{M}\{\eta\})^{-1}\begin{pmatrix}1\\a\end{pmatrix}(1,A)(\boldsymbol{M}\{\eta\})^{-1}\begin{pmatrix}1\\A\end{pmatrix}}}. \end{aligned}$$

The dependence of ϕ on N is removed as expected, however Equations (2.7) and (2.13) still rely on N through a power of $N-2$ found in the terms inside the integrals. This cannot be removed, and we thus have an implicit dependence on N, despite a continuous framework, which must be taken as a constant. This must be taken into account in the proceeding analysis.

3.3 A Numeric Approach To Optimal Designs For Simultaneous Confidence Bands

3.3.1 Rectangular Response Surface For Numeric Analysis

To find the optimal continuous design, we need only to optimise the criterion function with respect to $\boldsymbol{M}\{\eta\}$, considered for designs lying over some support $[a, A]$. The optimal design then corresponds to the information matrix that gives the optimal value. Note that

$$\boldsymbol{M}\{\eta\} = \sum_{i=1}^{n} p_i \boldsymbol{x_i} \boldsymbol{x_i}^{\intercal} = \begin{pmatrix} \sum_{i=1}^{n} p_i & \sum_{i=1}^{n} p_i x_i \\ \sum_{i=1}^{n} p_i x_i & \sum_{i=1}^{n} p_i x_i^2 \end{pmatrix} = \begin{pmatrix} 1 & m_1 \\ m_1 & m_2 \end{pmatrix}.$$

Hence we have only two unique entries, $m_1 = \sum_{i=1}^{n} p_i x_i$, and $m_2 = \sum_{i=1}^{n} p_i x_i^2$. We may therefore view any $\boldsymbol{M}\{\eta\}$, and by extension any design η as a point on the two dimensional surface $m_1 \times m_2$. Therefore the AW or MACS criterion functions of Equations (3.4-3.7) may be viewed as functions of two variables (m_1, m_2). A similar technique was implemented in Ah-Kine (2010), Section 2.5 for an exact design framework, but the two variables could not be constrained to intervals. This limited the conclusions of the resultant analysis. We can, however, show m_1 and m_2 are constrained to intervals for a specific support in a continuous setting. This is given by the following theorem, and proven in the appendix.

Theorem 3.1. *For any continuous design η with support $[a, A]$, the two variables $m_1 = \sum_{i=1}^{n} p_i x_i$ and $m_2 = \sum_{i=1}^{n} p_i x_i^2$ must lie within the regions $[a, A]$ and $[0, A^2]$ respectively when $A + a \geq 0$, and $[a, A]$ and $[0, a^2]$ when $A + a \leq 0$. When a=-A, either case is correct.*

Hence, to numerically establish the optimal design over $[a, A]$, for a specific N, it suffices to minimise Equations (3.4-3.7) with respect to (m_1, m_2), over the surface $[m_1] \times [m_2] = [a, A] \times [0, A^2]$ or $[a, A] \times [0, a^2]$. We may then infer the corresponding optimal design. This permits the use of optimisation algorithms to numerically search for the best design. This method however has limitations.

3.3.2 Limitations of Numeric Methods

The result of Theorem 3.1 is not entirely appropriate. Although every possible design over $[a, A]$ is represented as a point on this rectangular surface, bounding m_1 and m_2 in this way assumes that they are independent variables, which is clearly not the case.

In practice, the "true" surface, consisting **only** of (m_1, m_2) that can be attained by real designs is some region, contained within the rectangular surface, with variable, as opposed to fixed bounds on the value of m_2. These bounds depend on the value of m_1. For example, over support $[2, 5]$, the upper bound of m_2 over the rectangular surface is

25, and can be achieved by only one design, which has one design point $x_1 = 5$, with $p_1 = 1$. This takes value of $m_1 = 5$, and no other design with $m_1 \neq 5$ can achieve a value of $m_2 = 25$. In fact, at $m_1 = 5$ the point $(m_1, m_2) = (5, 25)$ is the only point corresponding to a real design.

This means that there exist (m_1, m_2) over the surface given by Theorem 3.1, that do not correspond to any real design, which we term "false points", and imposes limitations on any numeric analysis in the following ways.

- The false points may affect the shape of the response surface, and could lead to difficulty in coming to general conclusions. Furthermore this is likely to affect any numeric search algorithms, leading to points and search pathways where the criterion function cannot be evaluated.
- There is also a significant risk than any optimal point found by numeric methods may be a false point, with no method to protect against this occurrence.

Hence a numeric search for the optimal will not always be appropriate, and it cannot be predicted in what situations these limitations may affect analyses. However, **provided that the optimal** (m_1, m_2) **is shown to correspond to a real design, this must be the optimal design over the support for a specific** N. Hence this remains a more powerful method than those used in previous work.

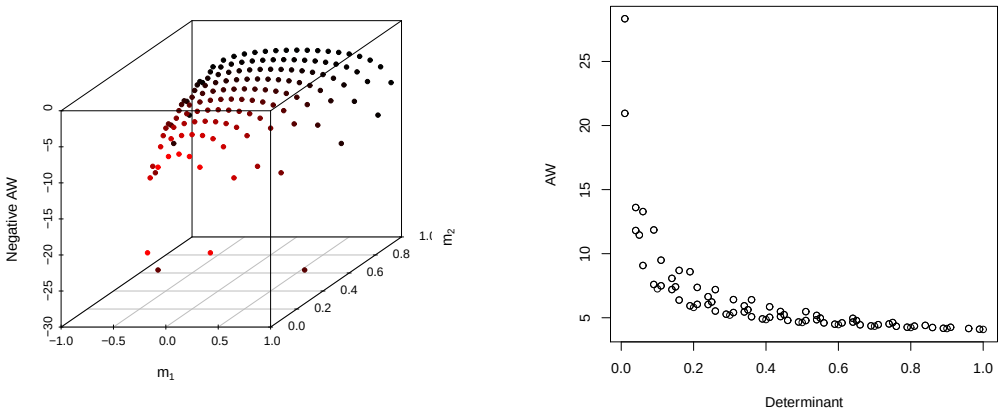
3.3.3 Numeric Investigation

We investigate the criterion functions of Equations (3.4-3.7), where c sets the simultaneous coverage to $1 - \alpha = 0.95$. We consider two regions for which to search, $[-1, 1]$ and $[2, 5]$. This means searching over the response surfaces $[m_1] \times [m_2] = [-1, 1] \times [0, 1]$, and $[2, 5] \times [0, 25]$ respectively.

We perform two numeric investigations. The first will investigate the response surface of the optimality criteria, by calculating Equations (3.4-3.7) over a grid of points on $[m_1] \times [m_2]$, with distance 0.1 between points on both axes. We provide two figures, a plot of the response surface against the negative value of the criterion function, for graphical purposes, and a plot of the determinant of $\boldsymbol{M}\{\eta\}$, corresponding to each pair of m_1 and m_2, against the value of the criterion function. This is performed for $N = 5$ and 6.

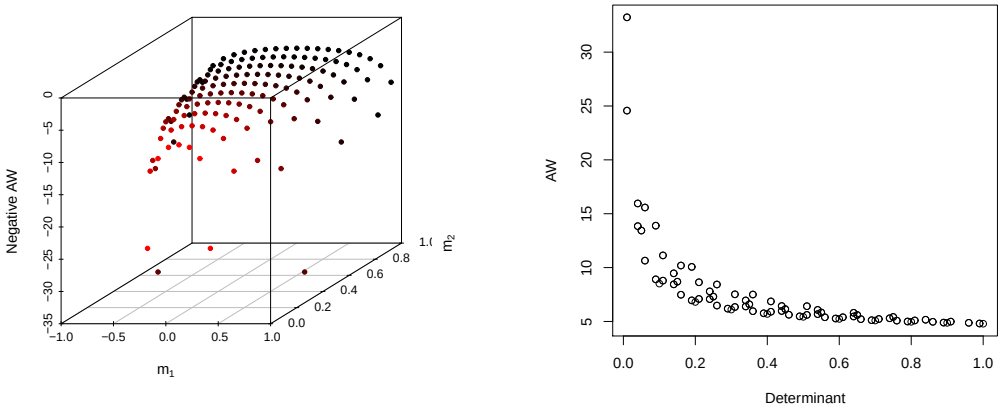
Secondly we use the optimisation algorithm "optim" from R, listed in the Appenidix and available on request, to identify the optimal m_1 and m_2 in each case for select values of N, and determine if the point corresponds to a real design. The list of R codes used for this analysis are given in the appendix.

AW Investigation For Scheffé Type Bands Over $[-1, 1]$



(a) 1a (b) 1b

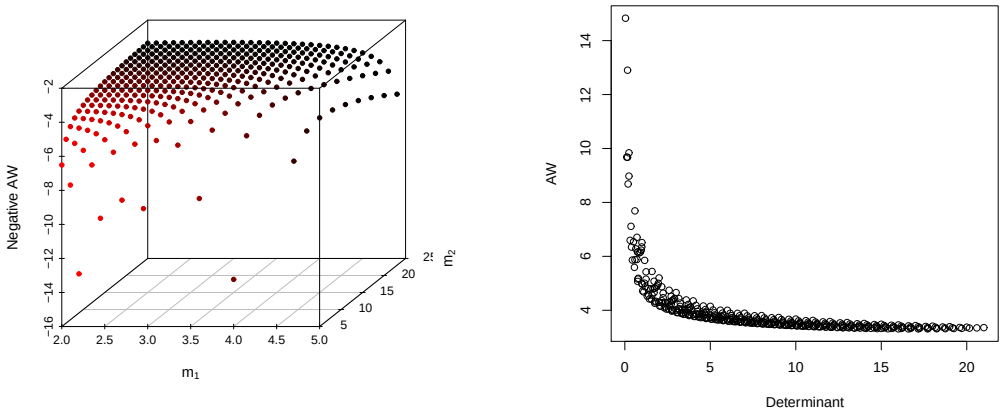
Figure 3.1: Analysis Of The AW For Scheffé Type Bands Over $[-1, 1]$ at $N = 6$



(a) 1a (b) 1b

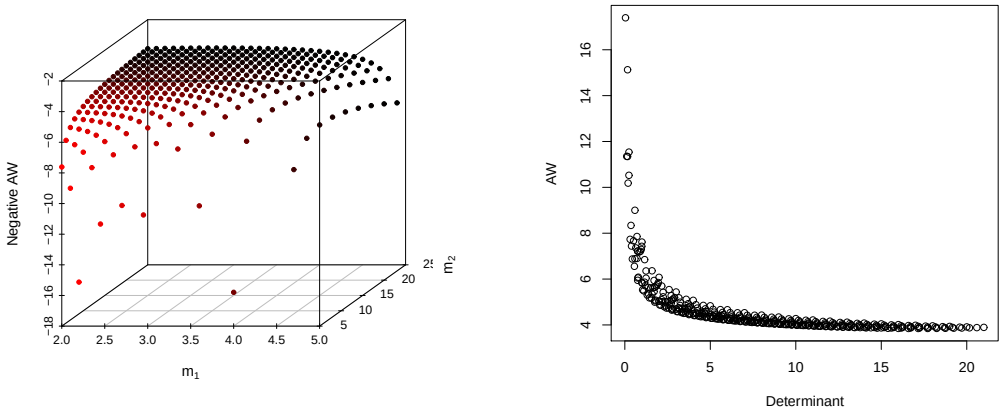
Figure 3.2: Analysis Of The AW For Scheffé Type Bands Over $[-1, 1]$ at $N = 5$

AW Investigation For Scheffé Type Bands Over $[2, 5]$



(a) 1a (b) 1b

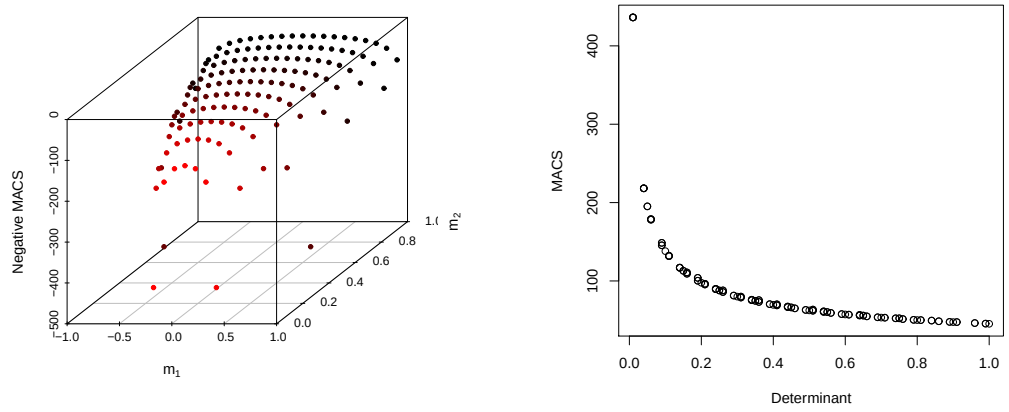
Figure 3.3: Analysis Of The AW For Scheffé Type Bands Over $[2, 5]$ at $N = 6$



(a) 1a (b) 1b

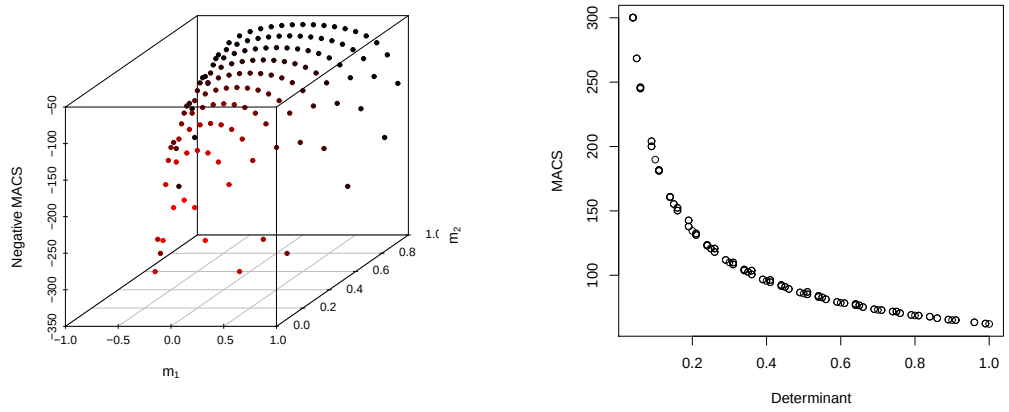
Figure 3.4: Analysis Of The AW For Scheffé Type Bands Over $[2, 5]$ at $N = 5$

MACS Investigation For Scheffé Type Bands Over $[-1, 1]$



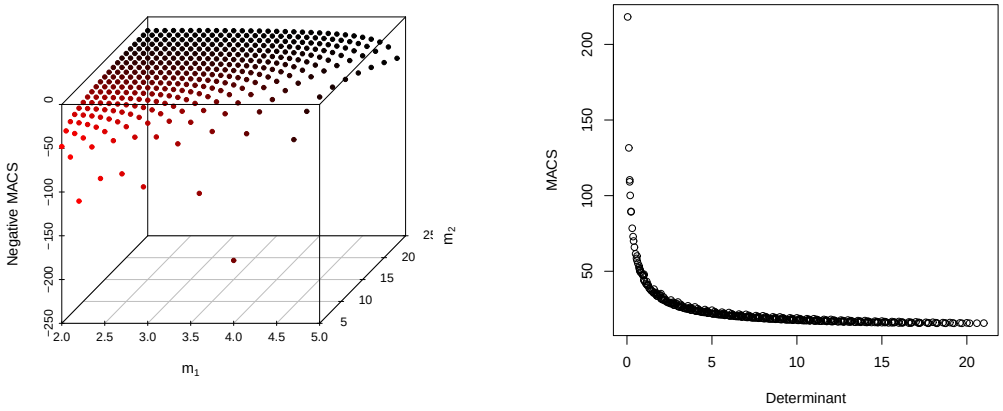
(a) 1a (b) 1b

Figure 3.5: Analysis Of The MACS Criterion For Scheffé Type Bands Over $[-1, 1]$ at $N = 6$



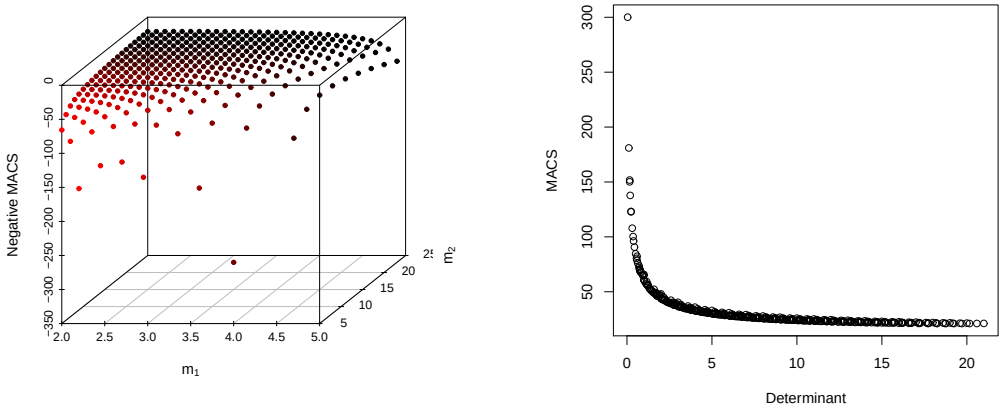
(a) 1a (b) 1b

Figure 3.6: Analysis Of The MACS Criterion For Scheffé Type Bands Over $[-1, 1]$ at $N = 5$



(a) 1a (b) 1b

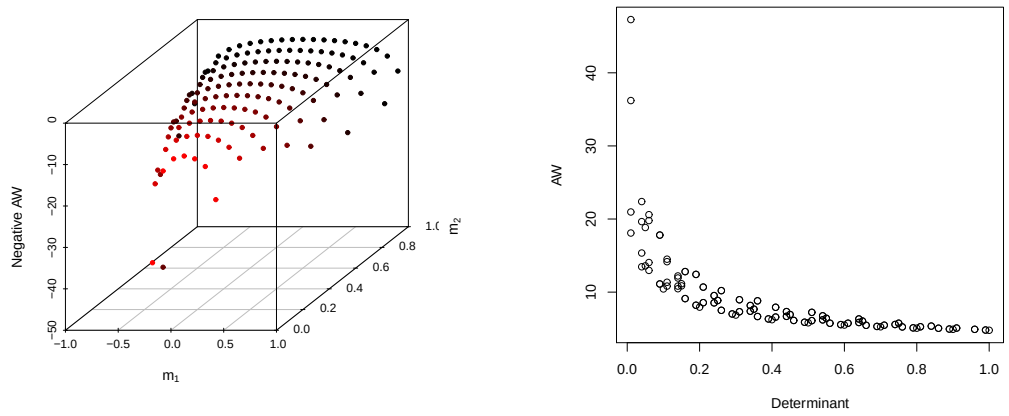
Figure 3.7: Analysis Of The MACS Criterion For Scheffé Type Bands Over $[2, 5]$ at $N = 6$



(a) 1a (b) 1b

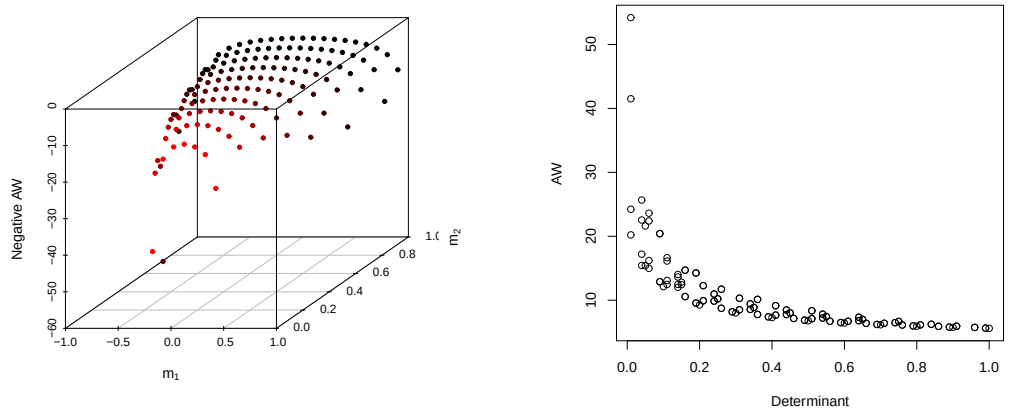
Figure 3.8: Analysis Of The MACS Criterion For Scheffé Type Bands Over $[2, 5]$ at $N = 5$

AW Investigation For Constant Width Bands Over $[-1, 1]$



(a) 1a (b) 1b

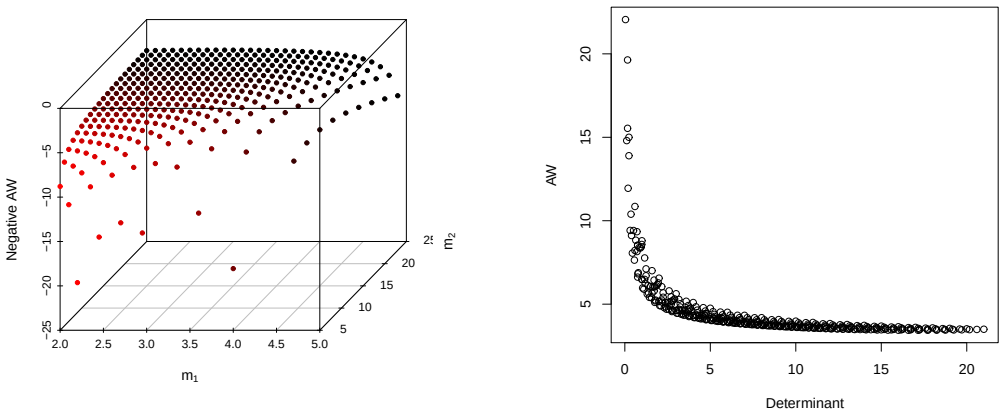
Figure 3.9: Analysis Of The AW Criterion For Constant Width Type Bands Over $[-1, 1]$ at $N = 6$



(a) 1a (b) 1b

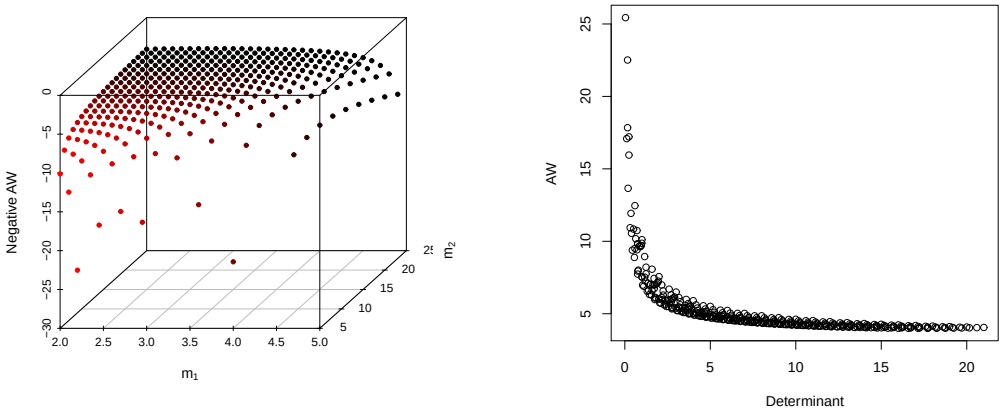
Figure 3.10: Analysis Of The AW Criterion For Constant Width Type Bands Over $[-1, 1]$ at $N = 5$

AW Investigation For Constant Width Bands Over $[2, 5]$



(a) 1a (b) 1b

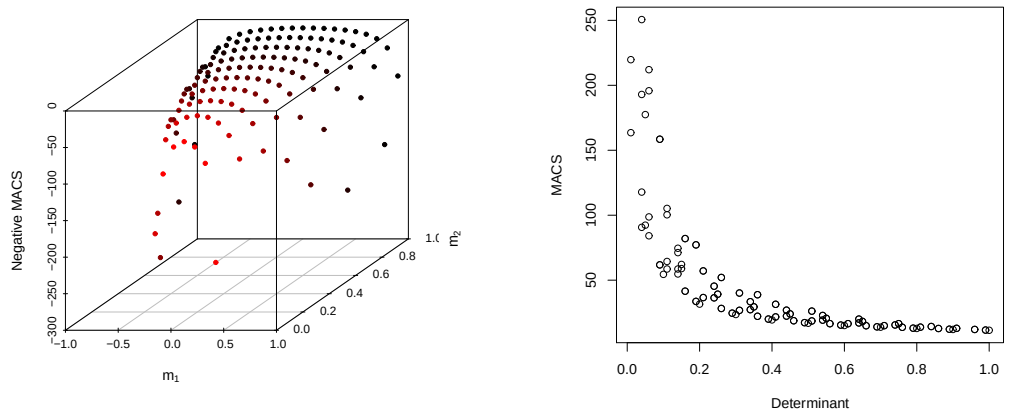
Figure 3.11: Analysis Of The AW Criterion For Constant Width Type Bands Over $[2, 5]$ at $N = 6$



(a) 1a (b) 1b

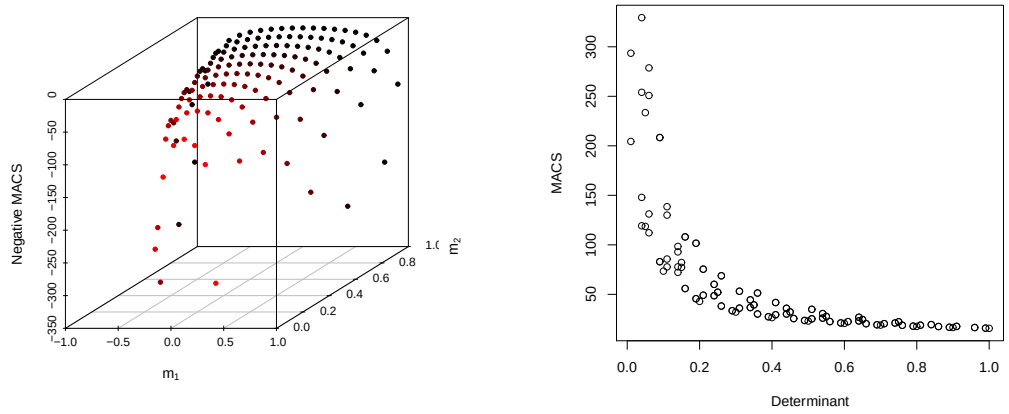
Figure 3.12: Analysis Of The AW Criterion For Constant Width Type Bands Over $[2, 5]$ at $N = 5$

MACS Investigation For Constant Width Bands Over $[-1, 1]$



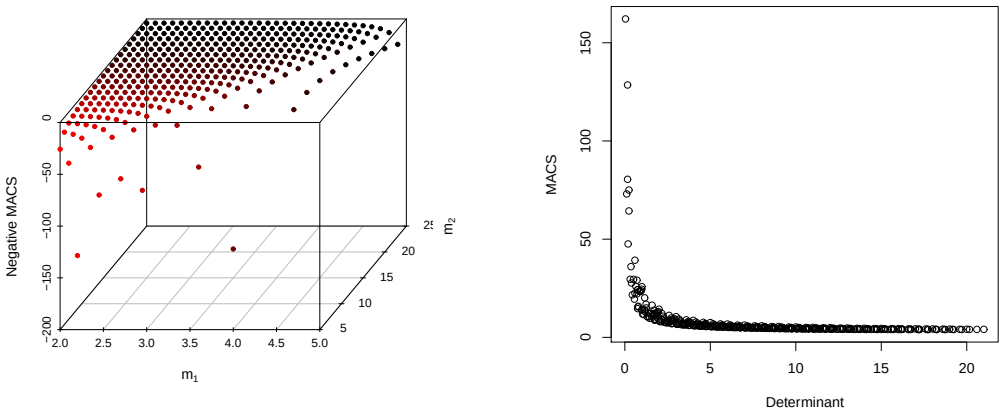
(a) 1a (b) 1b

Figure 3.13: Analysis Of The MACS Criterion For Constant Width Type Bands Over $[-1, 1]$ at $N = 6$



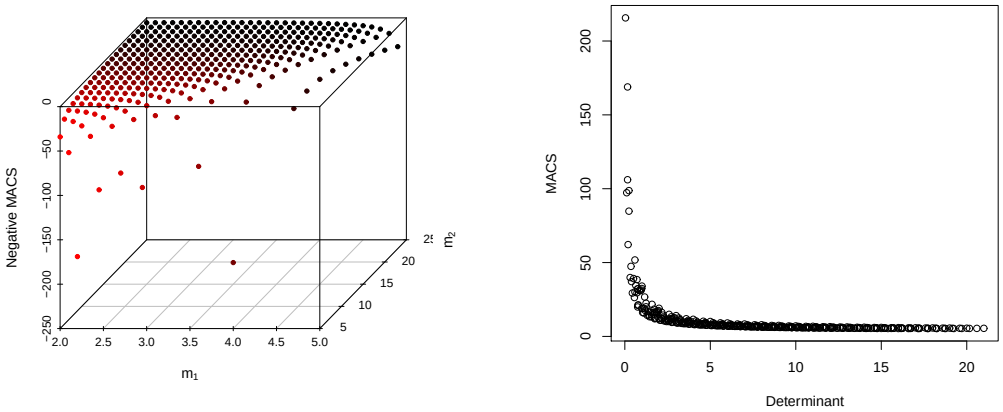
(a) 1a (b) 1b

Figure 3.14: Analysis Of The MACS Criterion For Constant Width Type Bands Over $[-1, 1]$ at $N = 5$



(a) 1a (b) 1b

Figure 3.15: Analysis Of The MACS Criterion For Constant Width Type Bands Over $[2, 5]$ at $N = 6$



(a) 1a (b) 1b

Figure 3.16: Analysis Of The MACS Criterion For Constant Width Type Bands Over $[2, 5]$ at $N = 5$

3.3.4 Optimal Designs Over $[-1, 1]$

The response surface investigation shows strong evidence that in each case, the optimal design follows D-optimality, under either AW or MACS criteria, and for both Scheffé type or constant width type bands. This supports the evidence given in a similar investigation in Ah-Kine (2010). There is a very strong trend of the AW or MACS criterion function getting smaller, when the determinant of the information matrix gets larger. This conclusion is the same for both N equal 5 and 6. Though the criteria do not monotonically decrease with increasing determinant, this is likely as a result of false points affecting the analysis. Furthermore, in each case the point on the surface

that maximises the negative of the optimality criterion (used for graphical purposes), gave $m_1 = 0$, and $m_2 = 1$, which corresponds to the information matrix which has the largest determinant over the surface. Furthermore, this corresponds to a real design, where half the points are allocated to -1 and 1 respectively as follows

$$\eta = \left\{ \begin{array}{cc} -1 & 1 \\ 0.5 & 0.5 \end{array} \right\}$$

This is well known as the D optimal design for one covariate over $[-1, 1]$. We now use optimisation algorithms using custom made code in "R", which are listed in the appendix, and available on request, to search for the optimal (m_1, m_2) for each case at a number of specific values of N. The results are shown in the following tables.

Table 3.1: Numeric AW and MACS Optimal Designs For 95 Percent Scheffé type bands over $[-1, 1]$

(a)			(b)		
N	Optimal (m_1, m_2)	Optimal AW	N	Optimal (m_1, m_2)	Optimal MACS
5	(0,1)	4.789	5	(0,1)	62.171
6	(0,1)	4.088	6	(0,1)	44.305
10	(0,1)	3.283	10	(0,1)	29.220
20	(0,1)	2.936	20	(0,1)	23.365
50	(0,1)	2.784	50	(0,1)	21.009
100	(0,1)	2.740	100	(0,1)	20.352

Table 3.2: Numeric AW and MACS Optimal Designs For 95 Percent Constant Width type bands over $[-1, 1]$

(a)			(b)		
N	Optimal (m_1, m_2)	Optimal AW	N	Optimal (m_1, m_2)	Optimal MACS
5	(0,1)	5.601	5	(0,1)	15.683
6	(0,1)	4.783	6	(0,1)	11.438
10	(0,1)	3.844	10	(0,1)	7.388
20	(0,1)	3.439	20	(0,1)	5.913
50	(0,1)	3.262	50	(0,1)	5.320
100	(0,1)	3.211	100	(0,1)	5.154

The optimisation algorithm was quite unstable, with the false points causing numeric errors in calculations. However, with a stable choice of initial points $(m_1, m_2) = (0, 0.5)$, all cases converged exactly to the optimal point which was given as $(0, 1)$ for all chosen N, and corresponds to the (m_1, m_2) of the D optimal design. Hence we may conclude that the optimal AW and MACS continuous designs, for 95 percent Scheffé type and constant width type bands over $[-1, 1]$ are D optimal, up to a specific value of N. We performed the same analysis for differing values of α, such as 0.01 and 0.1

and gave the exact same conclusions. Resultantly, the numeric approach infers that the values of both α and N do not affect the optimal design, we may therefore, tentatively, offer this analysis as a full proof of the optimal design. In particular, any analytical optimal must, with almost certainty, be D-optimal.

3.3.5 Optimal Designs Over $[2, 5]$

For a design over support $[2, 5]$ the limitations of the numeric analysis causes significant problems. Despite a clear correlation that the criterion function values are smaller as the determinant increases, in this case, the point that minimises the criteria was **not** that with the largest determinant. Once again, the optimality criteria do not monotonically decrease with increasing determinant, and so in this case the analysis has been more strongly influenced by false design points. Hence, evidence of the best design being D optimal is less compelling than over $[-1, 1]$. Applying an optimisation algorithm in the same way as before gave the following results, with initial point $(m_1, m_2) = (3, 10)$.

Table 3.3: Numeric AW and MACS Optimal Designs For 95 Percent Scheffé type bands over $[2, 5]$

	(a)			(b)	
N	Optimal (m_1, m_2)	Optimal AW	N	Optimal (m_1, m_2)	Optimal MACS
5	(2.808,25)	3.845	5	(2.427,25)	21.009
6	(2.855,25)	3.308	6	(2.477,25)	15.578
10	(2.931,25)	2.693	10	(2.556,25)	10.349
20	(2.971,25)	2.428	20	(2.597,25)	8.426
50	(2.991,25)	2.312	50	(2.618,25)	7.648
100	(2.996,25)	2.279	100	(2.624,25)	7.430

Table 3.4: Numeric AW and MACS Optimal Designs For 95 Percent Constant Width type bands over $[2, 5]$

	(a)			(b)	
N	Optimal (m_1, m_2)	Optimal AW	N	Optimal (m_1, m_2)	Optimal MACS
5	(2.785,25)	3.992	5	(2.785,25)	5.312
6	(2.838,25)	3.436	6	(2.838,25)	3.935
10	(2.919,25)	2.798	10	(2.919,25)	2.609
20	(2.962,25)	2.523	20	(2.962,25)	2.121
50	(2.982,25)	2.402	50	(2.982,25)	1.924
100	(2.988,25)	2.368	100	(2.988.25)	1.869

Note that it can be shown that D optimality is invariant under linear transformation of the support (see Kotz and Johnson (1992)) and therefore, for any support $[a, A]$, the D optimal design for simple linear regression always assigns weight $p = 0.5$ at the end points of the design range, a and A. In this case, the optimal value of m_2 was always

at 25 which as previously shown, is only possible for the special case of a design with one point $x_1 = 5$ and $p_1 = 1$, that is when $m_1 = 5$. Hence none of the optimal points found represent any real design.

To investigate, we performed the same search to various other design ranges, and found that we could converge to a "real" optimal (m_1, m_2) only when the support was symmetric around zero, that is when of the form $[-A, A]$. This point was always D optimal. For every case, the optimal value of m_2 was at its maximum, and in fact, the criterion function seems to always decrease as m_2 gets larger. It is therefore likely that the search algorithm is favouring points with the maximum value of m_2, which would suggest the method is unsuitable for non symmetric supports. For symmetric supports, multiple designs can attain the maximum value of m_2, with one of these happening to be D optimal. This is not the case for non symmetric supports however, with only one special design, with one support point, hitting the maximal m_2. This cannot reasonably be an optimal design. Hence any search algorithm will, with near certainty, converge to false points with a a non symmetric support. A different approach is needed in this instance.

3.3.6 Optimal End Point Designs For Non Symmetric Supports

We may consider a simpler approach for designs with support not symmetric around the origin. In this section, we search for the optimal end point design, that is a design with two fixed support points at the end points of the design range, with some weights p, and $1 - p$ as follows

$$\eta = \left\{ \begin{array}{cc} a & A \\ p & 1-p \end{array} \right\}.$$

This is a design of similar type to the D optimal design, where the support points also lie at the ends of the design range with $p = 0.5$. This would not constitute a formal proof of the optimal design, however if these end point designs can be shown to be D optimal, it would strengthen the limited evidence of the previous section that optimal designs, with non symmetric support are also D optimal. Note that with designs of this type, m_1 and m_2 become $ap + A(1-p)$ and $pa^2 + (1-p)A^2$ respectively. Hence the problem reduces to a one dimensional search in terms of $0 \leq p \leq 1$, which requires only a minor change to the algorithms used in the previous analyses. A numeric search is performed for a non symmetric support $[2, 5]$, for select N, at $\alpha = 0.05$ with initial point $p = 0.1$. This gave the following results with the code titles given in the appendix.

Table 3.5: AW and MACS Optimal End Point Designs For 95 Percent Scheffé type bands over $[2, 5]$

(a)

N	Optimal p	Optimal AW
5	0.5	4.789
6	0.5	4.088
10	0.5	3.283
20	0.5	2.936
50	0.5	2.784
100	0.5	2.740

(b)

N	Optimal p	Optimal MACS
5	0.5	41.447
6	0.5	30.203
10	0.5	19.480
20	0.5	15.576
50	0.5	14.006
100	0.5	13.568

Table 3.6: AW and MACS Optimal End Point Designs For 95 Percent Constant Width type bands over $[2, 5]$

(a)

N	Optimal p	Optimal AW
5	0.5	5.60
6	0.5	4.782
10	0.5	3.844
20	0.5	3.434
50	0.5	3.261
100	0.5	3.211

(b)

N	Optimal p	Optimal MACS
5	0.5	10.455
6	0.5	7.625
10	0.5	4.925
20	0.5	3.942
50	0.5	3.546
100	0.5	3.436

Tables 3.5 and 3.6 show that the optimal value of p in all cases is at 0.5, that is the optimal end point designs up to some N are always D-optimal. The criterion function values are larger than those given by the "false" optimal designs in Tables 3.3 and 3.4, but the difference is relatively small, being between only $2 - 3$ units larger for the AW, and around $5 - 20$ units larger for the MACS criterion. The response surface investigation suggests these designs would be among those with the smallest criterion function value. This adds additional weight to the best designs d=being D optimal,however the best chance of a full proof of the optimal non symmetric design remains an analytical approach. A natural extension of this work is to consider the best design supported by any two design points, x_1 and x_2 over $[a, A]$, which constitutes the optimal **minimally supported design** (see Lin and Chang (2007)), but is not considered here.

3.4 An Analytical Approach To Optimal Designs For Simultaneous Confidence Bands

Based on the numeric evidence, it is sensible to try and analytically show that the optimal design under AW or MACS is also D optimal. This is attempted in this section applying the method of calculus of variations as found in Silvey (1980). A summary of

this work is given, and we show how to manipulate each problem into a format such that these methods may be applied. However a full analytical proof is not given and the difficulties discussed.

3.4.1 Preliminary Theory

A continuous design η may be viewed as a probability distribution on some random variable $\tilde{x}$, where $\tilde{x}$ takes value x_i, with probability p_i, through the probability mass function

$$f_{\tilde{x}}(x) = \begin{cases} p_1, & x = x_1 \\ p_2, & x = x_2 \\ \cdots, & \cdots \\ p_n, & x = x_n. \end{cases}$$

We may then define the set of all continuous designs over the support $[a, A]$ as the set of all probability distributions on $\tilde{x}$ over $[a, A]$

$$H_{[a,A]} = \{\eta : x_i \in [a, A], i \in 1, \ldots, n\}.$$

If we now consider the random matrix

$$\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\intercal} = \begin{pmatrix} 1 \\ \tilde{x} \end{pmatrix} (1, \tilde{x}) = \begin{pmatrix} 1 & \tilde{x} \\ \tilde{x} & \tilde{x}^2 \end{pmatrix},$$

it is clear that

$$\boldsymbol{M}\{\eta\} = \mathrm{E}\{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\intercal}\}$$

as

$$\mathrm{E}\{\tilde{x}\} = \sum_{i=1}^{n} p_i x_i \quad \text{and} \quad \mathrm{E}\{\tilde{x}^2\} = \sum_{i=1}^{n} p_i x_i^2.$$

We may then define

$$\mathcal{M}_{[a,A]} = \{\boldsymbol{M}\{\eta\} = \mathrm{E}\{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\intercal}\} : \eta \in H_{[a,A]}\}$$

as the set of all continuous information matrices for designs over the design range $[a, A]$. This framework carries an advantage that the set $\mathcal{M}_{[a,A]}$ is a convex set, (see Silvey (1980)). That is, for two elements $M_1, M_2 \in \mathcal{M}_{[a,A]}$ and $\lambda \in [0, 1]$, $(1 - \lambda)M_1 + \lambda M_2 \in \mathcal{M}_{[a,A]}$. Searching for an optimal design over support $[a, A]$ is therefore reduced to searching over the set $\mathcal{M}_{[a,A]}$.

Now define a functional $\Psi(\boldsymbol{M}\{\eta\})$ as the criterion function acting on the continuous information matrices $\boldsymbol{M}\{\eta\} \in \mathcal{M}_{[a,A]}$, and define $\psi = -\Psi$. To minimise Ψ, it is sufficient to **maximise** ψ, hence we say that a design η^* is optimal over the set $H_{[a,A]}$ if it's corresponding information matrix $\boldsymbol{M}\{\eta^*\}$ is such that

$$\max_{\boldsymbol{M}\{\eta\} \in \mathcal{M}_{[a,A]}} \psi(\boldsymbol{M}\{\eta\}) = \psi(\boldsymbol{M}\{\eta^*\}).$$

It is assumed that ψ satisfies two properties.

1. ψ is a strictly increasing functions of the information matrices, that is for $M_1, M_2 \in \mathcal{M}_{[a,A]}$, if $M_1 - M_2$ is a non negative definite matrix (in the sense that $M_1 - M_2 \geq \mathbf{0}$) then
$$\psi(M_1) \geq \psi(M_2).$$

2. ψ is concave on $\mathcal{M}_{[a,A]}$, that is for any $0 \leq \lambda \leq 1$ and $M_1, M_2 \in \mathcal{M}_{[a,A]}$ we have
$$\psi(\lambda M_1 + (1-\lambda)M_2) \geq \lambda\psi(M_1) + (1-\lambda)\psi(M_2).$$

We define two generalised directional derivatives. The Gâteaux derivative is directly analogous to a directional derivative for ψ. The Gâteaux directional derivative at the matrix M_1, in the direction of M_2 is

$$G_\psi\{M_1, M_2\} = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \left[\psi\{M_1 + \epsilon M_2\} - \psi\{M_1\}\right].$$

The Fréchet derivative is more specialised, and is given as

$$F_\psi\{M_1, M_2\} = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \left[\psi\{(1-\epsilon)M_1 + \epsilon M_2\} - \psi\{M_1\}\right].$$

It is easy to see that

$$\begin{aligned} F_\psi\{M_1, M_2\} =& \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \left[\psi\{(1-\epsilon)M_1 + \epsilon M_2\} - \psi\{M_1\}\right] \\ =& \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \left[\psi\{(M_1 + \epsilon(M_2 - M_1)\} - \psi\{M_1\}\right] = G_\psi\{M_1, M_2 - M_1\}, \end{aligned}$$

hence the Fréchet derivative exists as long as the Gâteaux derivative does. The maximisation of ψ then relies on the following theorem.

Theorem 3.2. *If ψ is a strictly increasing and concave function on $\mathcal{M}_{[a,A]}$, then $M\{\eta^*\}$ is optimal over $\mathcal{M}_{[a,A]}$ if, and only if $F_\psi\{M\{\eta^*\}, \boldsymbol{x}\boldsymbol{x}^\intercal\} \leq 0 \; \forall \; x \in [a, A]$, provided that $G_\psi\{M\{\eta^*\}, \boldsymbol{x}\boldsymbol{x}^\intercal\}$ always exists.*

It is then immediate that η^* must be the optimal design over $H_{[a,A]}$ for Ψ.

3.4.2 Framework for Analytical Proof

The primary difficulty in applying Theorem 3.2 to the AW or MACS criteria again lies in the complicated relationship between the critical constant c, and $\boldsymbol{M}\{\eta\}$, as a result of the simultaneous coverage probability expressions. Suppose we take constant width bands, each problem takes the the general form

$$\text{Min}_{\boldsymbol{M}\{\eta\} \in \mathcal{M}_{[a,A]}} \; \Psi(\boldsymbol{M}\{\eta\}, c)$$

$$w.r.t \int_{q(\boldsymbol{M}\{\eta\})}^{r(\boldsymbol{M}\{\eta\})} T(\theta, \boldsymbol{M}\{\eta\}, c) \, \mathrm{d}\theta + \int_{l(\boldsymbol{M}\{\eta\})}^{u(\boldsymbol{M}\{\eta\})} S(\theta, \boldsymbol{M}\{\eta\}, c) \, \mathrm{d}\theta = 1 - \alpha.$$

Note that $\Psi(\boldsymbol{M}\{\eta\}, c)$ is deliberately labeled to note that we cannot represent c directly as a function of $\boldsymbol{M}\{\eta\}$, and in this form, we cannot apply Theorem 3.2. It is easier to consider the dual problem,

$$\text{Max}_{\boldsymbol{M}\{\eta\}\in\mathcal{M}_{[a,A]}} \int_{q(\boldsymbol{M}\{\eta\})}^{r(\boldsymbol{M}\{\eta\})} T(\theta, \boldsymbol{M}\{\eta\}, c)\, \mathrm{d}\theta + \int_{l(\boldsymbol{M}\{\eta\})}^{u(\boldsymbol{M}\{\eta\})} S(\theta, \boldsymbol{M}\{\eta\}, c)\, \mathrm{d}\theta$$

$$w.r.t\ \Psi(\boldsymbol{M}\{\eta\}, c) = K$$

where K is some constant. Noting that c is a multiplicative constant under the AW or MACS criteria, we may easily set the dual constraint as a function directly of c, that is

$$c = K\Psi^{-1}(\boldsymbol{M}\{\eta\}).$$

We can then immediately substitute in for c for the dual objective function and we get

$$\begin{aligned}\text{Max}_{\boldsymbol{M}\{\eta\}\in\mathcal{M}_{[a,A]}} &\int_{q(\boldsymbol{M}\{\eta\})}^{r(\boldsymbol{M}\{\eta\})} T(\theta, \boldsymbol{M}\{\eta\}, K\Psi^{-1}(\boldsymbol{M}\{\eta\}))\, \mathrm{d}\theta \\ &+ \int_{l(\boldsymbol{M}\{\eta\})}^{u(\boldsymbol{M}\{\eta\})} S(\theta, \boldsymbol{M}\{\eta\}, K\Psi^{-1}(\boldsymbol{M}\{\eta\}))\, \mathrm{d}\theta.\end{aligned}$$

Intuitively the problem can be viewed as maximising the simultaneous coverage probability of the confidence band, given a fixed AW or area of confidence set, which are used to get an expression for c directly in terms of the information matrix. This gives a single expression one must maximise to obtain η^*, which we may set to ψ and then apply theorem 3.2. For example, the the dual problem for a constant width band under the AW is

$$\begin{aligned}\text{Max}_{\boldsymbol{M}\{\eta\}\in\mathcal{M}_{[a,A]}} \frac{1}{\pi} &\int_{\eta_1-\pi}^{\xi_1} \left(1 - \left(1 + \frac{c_1^2}{(N-2)cos^2\theta}\right)^{-\frac{N-2}{2}}\right) \mathrm{d}\theta \\ &+ \int_{\xi_1-\phi}^{\eta_1-\phi} \left(1 - \left(1 + \frac{c_2^2}{(N-2)cos^2\theta}\right)^{-\frac{N-2}{2}}\right) \mathrm{d}\theta\end{aligned}$$

$$w.r.t\ c = K.$$

It is then immediate that this is equivalent to solving

$$\begin{aligned}\text{Max}_{\boldsymbol{M}\{\eta\}\in\mathcal{M}_{[a,A]}} &\int_{\eta_1-\pi}^{\xi_1} \left(1 - \left(1 + \frac{K^2}{(1,A)(\boldsymbol{M}\{\eta\})^{-1}\begin{pmatrix}1\\A\end{pmatrix}(N-2)cos^2\theta}\right)^{-\frac{N-2}{2}}\right) \mathrm{d}\theta \\ &+ \int_{\xi_1-\phi}^{\eta_1-\phi} \left(1 - \left(1 + \frac{K^2}{(1,a)(\boldsymbol{M}\{\eta\})^{-1}\begin{pmatrix}1\\a\end{pmatrix}(N-2)cos^2\theta}\right)^{-\frac{N-2}{2}}\right) \mathrm{d}\theta\end{aligned}$$

using the relation between c_1, c_2 and c for a constant width band. Even with the simplification the simultaneous coverage probability expressions are complicated functions of the information matrix. For example, once c is expressed as a function of $\boldsymbol{M}\{\eta\}$, the Gâteaux derivative for $\int_{l(\boldsymbol{M}\{\eta\})}^{u(\boldsymbol{M}\{\eta\})} S(\theta, \boldsymbol{M}\{\eta\})\,\mathrm{d}\theta$ is

$$G_{\int_{l(\boldsymbol{M}\{\eta\})}^{u(\boldsymbol{M}\{\eta\})} S(\theta,\boldsymbol{M}\{\eta\})\,\mathrm{d}\theta}\{M_1, M_2\} = \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\int_{l(M_1+\epsilon M_2)}^{u(M_1+\epsilon M_2)} S(\theta, M_1+\epsilon M_2)\,\mathrm{d}\theta - \int_{l(M_1)}^{u(M_1)} S(\theta, M_1)\,\mathrm{d}\theta,$$

and cannot be easily evaluated. The term inside the limit can be simplified as

$$\begin{aligned}
&\int_{l(M_1+\epsilon M_2)}^{u(M_1+\epsilon M_2)} S(\theta, M_1+\epsilon M_2)\,\mathrm{d}\theta - \int_{l(M_1)}^{u(M_1)} S(\theta, M_1)\,\mathrm{d}\theta \\
&= \int_{l(M_1+\epsilon M_2)}^{u(M_1+\epsilon M_2)} \{S(\theta, M_1+\epsilon M_2) - S(\theta, M_1)\}\,\mathrm{d}\theta + \int_{l(M_1+\epsilon M_2)}^{u(M_1+\epsilon M_2)} S(\theta, M_1)\,\mathrm{d}\theta \\
&- \int_{l(M_1)}^{u(M_1)} S(\theta, M_1)\,\mathrm{d}\theta \\
&= \int_{l(M_1+\epsilon M_2)}^{u(M_1+\epsilon M_2)} \{S(\theta, M_1+\epsilon M_2) - S(\theta, M_1)\}\,\mathrm{d}\theta \\
&+ \left(\int_{l(M_1+\epsilon M_2)}^{u(M_1)} S(\theta, M_1)\,\mathrm{d}\theta + \int_{u(M_1)}^{u(M_1+\epsilon M_2)} S(\theta, M_1)\,\mathrm{d}\theta\right) \\
&- \left(\int_{l(M_1)}^{l(M_1+\epsilon M_2)} S(\theta, M_1)\,\mathrm{d}\theta + \int_{l(M_1+\epsilon M_2)}^{u(M_1)} S(\theta, M_1)\,\mathrm{d}\theta\right) \\
&= \int_{l(M_1+\epsilon M_2)}^{u(M_1+\epsilon M_2)} \{S(\theta, M_1+\epsilon M_2) - S(\theta, M_1)\}\,\mathrm{d}\theta \\
&+ \int_{u(M_1)}^{u(M_1+\epsilon M_2)} S(\theta, M_1)\,\mathrm{d}\theta - \int_{l(M_1)}^{l(M_1+\epsilon M_2)} S(\theta, M_1)\,\mathrm{d}\theta,
\end{aligned}$$

since for sufficiently small ϵ, we can write $l(M_1) < l(M_1+\epsilon M_2) < u(M_1) < u(M_1+\epsilon M_2)$. The Gâteaux derivative hence involves three expressions, it is clear that, for the first term

$$\begin{aligned}
&\lim_{\epsilon\to 0^+} \int_{l(M_1+\epsilon M_2)}^{u(M_1+\epsilon M_2)} \frac{1}{\epsilon}\{S(\theta, M_1+\epsilon M_2) - S(\theta, M_1)\}\,\mathrm{d}\theta \\
&= \int_{l(M_1)}^{u(M_1)} G_{S(\theta,\boldsymbol{M}\{\eta\})}\{M_1, M_2\}\mathrm{d}\theta.
\end{aligned}$$

Applying the first mean value mean value theorem for integrals to the final two terms gives

$$\int_{u(M_1)}^{u(M_1+\epsilon M_2)} S(\theta, M_1)\,\mathrm{d}\theta = (u(M_1+\epsilon M_2) - u(M_1))S(s, M_1),$$

for some real value $s \in (u(M_1), u(M_1 + \epsilon M_2))$. Clearly as $\epsilon \to 0^+,\ s \to u(M_1)$ this is expressed as

$$\lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\{(u(M_1 + \epsilon M_2) - u(M_1))S(c, M_1)\} = G_{u(\boldsymbol{M}\{\eta\})}\{M_1, M_2\}S(u(M_1), M_1).$$

Hence the full derivative is given as

$$G_{\int_{l(\boldsymbol{M}\{\eta\})}^{u(\boldsymbol{M}\{\eta\})} S(\theta, \boldsymbol{M}\{\eta\})\, \mathrm{d}\theta}\{M_1, M_2\} = \int_{l(M_1)}^{u(M_1)} G_{S(\theta, \boldsymbol{M}\{\eta\})}\{M_1, M_2\}\mathrm{d}\theta$$
$$+G_{u(\boldsymbol{M}\{\eta\})}\{M_1, M_2\}\ S(u(M_1), M_1) - G_{l(\boldsymbol{M}\{\eta\})}\{M_1, M_2\}\ S(l(M_1), M_1).$$

Clearly, the dual function will always consist of at least one integral of the form above, hence we always require the evaluation of at least three separate Gâteaux derivatives, of which is a significant challenge given its complicated nature. We found little or no cancellation of terms in our efforts, which does not even consider an attempt to further show the concavity of the function, or that it is strictly increasing. A proof of this manner may be possible, but very difficult.

3.5 Concluding Remarks

The work of the Chapter obtains for the first time, results for the optimal design of experiments for Scheffé type and constant width type bands over a finite interval, not specific to a particular dataset. However the work has more limitations than was intended, and presents new problems for further work.

For bands considered over the normalised covariate range $x \in [-1, 1]$, numeric analysis, for which the response surface may be constrained over intervals, found that, conditioned on a value of N and α, the optimal continuous design for both bands was D-optimal. This was the case for both MACS, and of particular note, AW criteria. The analysis suggests that the optimal design is the same for any chosen N or α, from which we may deduce that the analytical solution is almost certainly D-optimal. However this method is limited by its search area being flawed, and was not, in its current state, appropriate to obtain the optimal design over any non symmetric support.

Work has also been developed for the first time on obtaining an analytical optimal design, using the method of calculus of variations. The relevant criterion functions were manipulated into a correct form, but an analytic proof remains unfulfilled. Further work into the analytical optimal designs, building on these functions is necessary.

Chapter 4

An Improvement To Simultaneous Confidence Sets For Several Effective Doses

For the second part of the thesis we construct one and two sided simultaneous confidence sets on the effective doses of the logistic regression model. This is achieved by developing an adaption of an existing method, involving the construction of simultaneous confidence bands. These sets are shown to be less conservative than the current methodology. This method is immediately applicable to any GLM under an asymptotically normal MLE, with a band of the form in section 1.5, and some link function $b(\mu(x))$. However, logistic regression has seen wide applications to medical and biological statistics over the years, a point noted early on by Finney (1971). It is a key statistical tool in quantitative risk and hazard assessment, (Piegorsch and West (2005)) and in cohort and clinical trial studies assessing drug dose response curves Carter et al. (1986), for which the Effective Dose (ED) is a particularly important construct. The following work would have most impact, and is hence established, under the assumption of a logistic regression model.

4.1 Introduction

4.1.1 The Effective Dose

Recall that the effective dose is defined as the value of the covariate(s) required to elicit a specific probability of response, p, in the true regression model, or equivalently the stimulus required to observe a response in a fixed proportion of the population. For some p, the effective dose is defined as ED_{100_p}, which we term x_p, and is defined as

$$x_p = \left\{ x : \boldsymbol{x}^\intercal \boldsymbol{\beta} = \pi(p) = \log\left(\frac{p}{1-p}\right) \right\} = \left\{ x : \frac{\exp\left(\boldsymbol{x}^\intercal \boldsymbol{\beta}\right)}{1+\exp\left(\boldsymbol{x}^\intercal \boldsymbol{\beta}\right)} = p \right\}. \tag{4.1}$$

For a single covariate ($q = 1$), the ED is a single point often viewed as the ratio

$$x_p = \frac{\pi(p) - \beta_0}{\beta_1} \tag{4.2}$$

with the most well known effective dose the ED_{50}, that is $x_{0.5}$. This represents the point at which the odds $\frac{p}{1-p}$ of observing a success is equal to 1. As $\pi(0.5) = 0$, the ED_{50} can be simplified to

$$x_{0.5} = \frac{-\beta_0}{\beta_1}$$

which is perhaps the most recognisable form of an effective dose. For a multiple covariate model ($q > 1$), Equation (4.1) is the multivariate effective dose (MED) and is a surface, consisting of the set of $x = (x_1, \ldots, x_q)$ which satisfies the equation. This can be a challenging construct, and in response Li et al. (2008) considered the MED for a single x_i, conditioned on values of the remaining $q - 1$ covariates. This is the Conditioning Effective Dose (CED) and is given as

$$x_p^{CED} = \left\{ x^* = (x_1, x_2^*, \ldots, x_q^*) : \boldsymbol{x}^{*\intercal}\boldsymbol{\beta} = \pi(p) = \log\left(\frac{p}{1-p}\right) \right\}.$$

Point estimation of the ED can be achieved by plugging in the MLE. It is of greater interest, however, to establish a confidence region around the true ED's.

4.1.2 Confidence Sets For A Single Effective Dose

The most well known methods to establish a confidence set for one ED are the Fieller and Delta methods of Fieller (1954) and Cox (1990) respectively, which rely on separate approximations of the variance of the ED, viewed as the ratio in Equation (4.2). Other popular methods include the likelihood ratio method, as well as bootstrap and MCMC numeric algorithms. These are widely used in particular for one covariate models, but may be applied in a multiple covariate setting and, can be easily adapted for the CED, as shown in Li et al. (2008). However, these methods are an aside to the current topic as none can be suitably adapted to account for the Bonferroni correction of multiple comparisons. Hence we refer to the following method. For the large sample logistic regression model, recall from Equation (1.20) that for a specific value of x

$$\begin{aligned} \boldsymbol{x}^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &\overset{d}{\approx} \mathcal{N}(0, \boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}) \\ \implies \frac{\boldsymbol{x}^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{\boldsymbol{x}^{*\intercal} \boldsymbol{J}^{-1} \boldsymbol{x}^*}} &\overset{d}{\approx} \mathcal{N}(0, 1). \end{aligned}$$

It is then immediate that an approximate $1 - \alpha$ confidence interval on $\boldsymbol{\beta}$ is given as

$$\mathrm{P}\left\{ \frac{|\boldsymbol{x}^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|}{\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}}} < z^{\frac{\alpha}{2}} \right\} \approx 1 - \alpha \tag{4.3}$$

for any fixed q dimensional value of x, with $z^{\frac{\alpha}{2}}$ the $(1 - \frac{\alpha}{2})$'th quantile of the standard normal distribution. Using this result, we consider the following confidence set on an

unknown x with a specific value of p

$$\mathbf{C}_p = \left\{ \boldsymbol{x} : \pi(p) \in \boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \pm z^{\frac{\alpha}{2}} \sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}} \right\} = \left\{ \boldsymbol{x} : \frac{|\boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} - \pi(p)|}{\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}}} < z^{\frac{\alpha}{2}} \right\}. \tag{4.4}$$

We then see that this set constitutes an approximate $1-\alpha$ confidence set for the ED_p, that is x_p, as

$$\begin{aligned} \mathrm{P}\left\{x_p \in \mathbf{C}_p\right\} =& \mathrm{P}\left\{ \frac{|\boldsymbol{x_p}^\intercal \hat{\boldsymbol{\beta}} - \pi(p)|}{\sqrt{\boldsymbol{x_p}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_p}}} < z^{\frac{\alpha}{2}} \right\} \\ =& \mathrm{P}\left\{ \frac{|\boldsymbol{x_p}^\intercal (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|}{\sqrt{\boldsymbol{x_p}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_p}}} < z^{\frac{\alpha}{2}} \right\} \approx 1 - \alpha. \end{aligned}$$

This is since $\boldsymbol{x}_p^\intercal \boldsymbol{\beta} = \pi(p)$, with $\boldsymbol{x_p} = (1, x_p)^\intercal$, by the logistic regression model. Note that the term inside $\mathbf{C}_p$ is a Scheffé type confidence band over the whole range with $c = z^{\frac{\alpha}{2}}$. By definition, the coverage coverage probability of this band at any single point on x is $1-\alpha$. This is known as a **point wise** confidence band. We may therefore interpret $\mathbf{C}_p$ as the bounds on a region of the inverse dose response curve, which contains all such x, that the bounds of the $1-\alpha$ point-wise Scheffé type confidence band at x, includes $\pi(p)$. By definition of the confidence band, any point on this region is a plausible candidate for x_p with confidence level $1-\alpha$, illustrated in Figure 4.1 below.

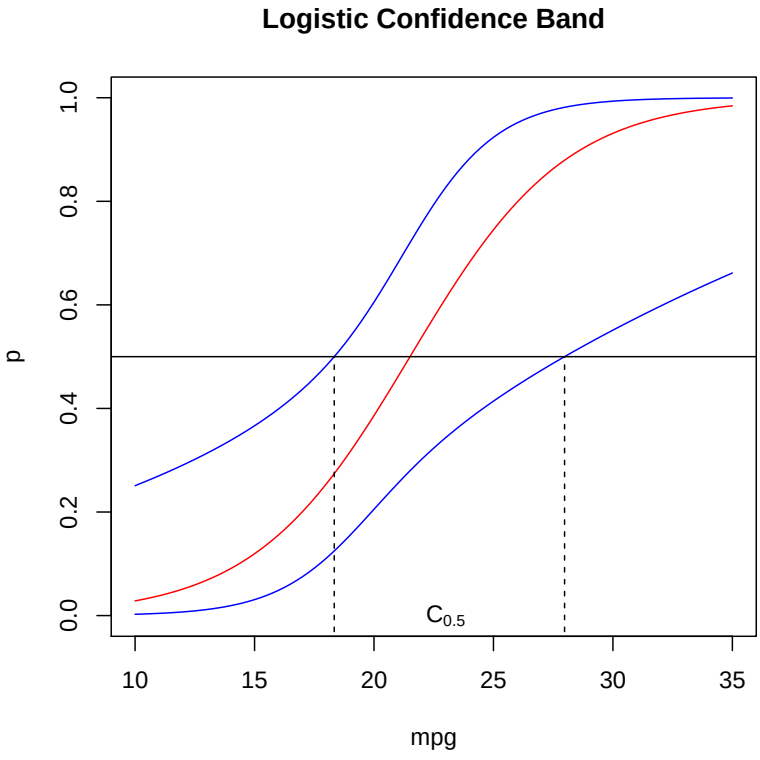


Figure 4.1: An illustration of the Scheffé band method for the ED_{50} of the "mtcars" dataset.

We term this the Scheffé band method. It can be shown (see Li et al. (2008)), that applying this method to a point wise Scheffé type confidence band is equivalent to Fiellers theorem. It is however more intuitive, and can be immediately applied to the Multivariate or Conditional effective doses. Most of all, the Scheffé Band method is easily adaptable to multiple comparisons simply by changing the critical constant of the band c, which is the focus of this Chapter.

4.1.3 Simultaneous Confidence Sets For Several Effective Doses

For simultaneous inference on the effective dose, the methods of Brand et al. (1973) and Walter (1983) invert the bounds of a Scheffé type simultaneous confidence band over the whole real line, at particular values of $\pi(p)$ (or equivalently p). Formally we define this confidence set as

$$\mathbf{C}_p = \left\{ \boldsymbol{x} : \pi(p) \in \boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \pm \sqrt{\chi^\alpha_{q+1}} \sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}} \right\} = \left\{ \boldsymbol{x} : \frac{|\boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} - \pi(p)|}{\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}}} < \sqrt{\chi^\alpha_{q+1}} \right\}. \quad (4.5)$$

Like before, $\mathbf{C}_p$ contains all such x that the Scheffé type simultaneous confidence band at x includes $\pi(p)$. In this case, however, it can be shown that the simultaneous coverage probability of **any** k effective doses x_{p_i}, lying in the corresponding $\mathbf{C}_{p_i}$, for $i = 1, \ldots, k$ is at least $1 - \alpha$ in the following way:

$$\begin{aligned}
&\mathrm{P}\left\{x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, \ldots, k\right\} \\
=&\mathrm{P}\left\{ \frac{|(\boldsymbol{x_{p_i}})^\intercal \hat{\boldsymbol{\beta}} - \pi(p_i)|}{\sqrt{(\boldsymbol{x_{p_i}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{p_i}}}} < \sqrt{\chi^\alpha_{q+1}} \text{ for } i = 1, \ldots, k \right\} \\
\geq&\mathrm{P}\left\{ \frac{|(\boldsymbol{x_p})^\intercal (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|}{\sqrt{(\boldsymbol{x_p})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_p}}} < \sqrt{\chi^\alpha_{q+1}} \; \forall p \in (0,1) \right\} \quad (4.6) \\
=&\mathrm{P}\left\{ \frac{|\boldsymbol{x}^\intercal (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|}{\sqrt{(\boldsymbol{x})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}}} < \sqrt{\chi^\alpha_{q+1}} \; \forall x \in \mathbb{R}^q \right\} \quad (4.7) \\
=&\mathrm{P}\left\{ \boldsymbol{x}^\intercal \boldsymbol{\beta} \in \boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} \pm \sqrt{\chi^\alpha_{q+1}} \sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}} \; \forall \; x \in \mathbb{R}^q \right\} = 1 - \alpha.
\end{aligned}$$

The inequality in (4.6) occurs since $\boldsymbol{x_{p_i}}^\intercal \boldsymbol{\beta} = \pi(p_i)$, and the inequality in (4.7) is formed because the set of all effective doses is the same as the set of all $x \in \mathbb{R}^p$. Consequently we must have

$$\mathrm{P}\left\{x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, \ldots, k\right\} \geq 1 - \alpha.$$

Note that just as in the case for one ED, in practice, the coverage probability of any set of this form is approximate, based on the accuracy of the large sample probability distribution of $\hat{\boldsymbol{\beta}}$ to the normal. We assume for the simplicity of mathematical labelling in this thesis that the approximation is close enough to be considered exact. Nonetheless, this is a powerful method. We are guaranteed at least the nominal simultaneous coverage $1 - \alpha$ **regardless** of the choice of k. However these confidence sets are unduly

conservative in terms of simultaneous coverage, particularly when k is small. From the expression in (4.6), we see that exact simultaneous coverage is only possible when sets are sought for **every** effective dose at once, that is for all $p \in (0, 1)$. The smaller the k, the larger this inequality tends to be, and thus the more conservative it is. Many papers utilising this method, such as Piegorsch and West (2005) note this problem. A simulation study conducted in Li et al. (2008) in particular showed that, for small a k, coverage sometimes reached nearly100%.

4.1.4 Methodology for Improved Simultaneous Confidence Sets

Applying the scheffé band method to the point-wise or simultaneous confidence band represent two extremes. In this thesis, we develop the following adaptation to address this problem. The simultaneous coverage of any set of the form $\mathbf{C}_p$ is determined by the critical constant of the band, hence we establish simultaneous confidence sets of the same form as in (4.5), but with a general critical constant c;

$$\mathbf{C}_p = \left\{ x : \frac{|\boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} - \pi(p)|}{\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}}} < \mathrm{c} \right\}, \tag{4.8}$$

where c is specifically set, so that at least the nominal simultaneous coverage $1 - \alpha$ is guaranteed for a **specific** number, k, of effective doses at once, that is

$$\mathrm{P}\left\{ x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, \ldots, k \right\} \geq 1 - \alpha. \tag{4.9}$$

In summary, we replace a single confidence set applicable to any value of k, with a series of specialised sets for each k. The methods derived in the following sections all hinge on the following methodology. To establish the required c, it is sufficient to find the k effective doses $-\infty < x_{p_1}, \ldots, x_{p_k} < \infty$ that minimise the probability in (4.9), for any value of $\boldsymbol{\beta} \in \mathbb{R}^{q+1}$ or $c > 0$, that is

$$\min_{-\infty < x_{p_1}, \ldots, x_{p_k} < \infty} \mathrm{P}\left\{ x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, \ldots, k \right\} \quad \forall \boldsymbol{\beta} \in \mathbb{R}^{q+1} \text{ and } \forall c > 0.$$

If the value of c is then set to satisfy

$$\min_{-\infty < x_{p_1}, \ldots, x_{p_k} < \infty} \mathrm{P}\left\{ x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, \ldots, k \right\} = 1 - \alpha$$

then Equation (4.9) must also be satisfied. Sets of this form are guaranteed, by design, to exhibit closer to nominal simultaneous coverage than the existing methods. For the case $k = 2$, the methods developed can be applied to the CED for a multiple model. For $k \geq 3$, our results are restricted to a univariate model with $q = 1$.

4.2 Two Sided Simultaneous Confidence Sets For Several Effective Doses

We first apply the methodology of Section 4.1.4 to establish two sided simultaneous confidence sets for specific numbers of effective doses.

4.2.1 Simultaneous Confidence Sets For Two Effective Doses

We begin with the simplest situation, to construct a confidence set of the form in (4.8) for $k = 2$. In this case, c is chosen such that for any two effective doses x_{p_1} and x_{p_2} we have

$$\mathrm{P}\left\{x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, 2\right\} \geq 1 - \alpha. \tag{4.10}$$

The value of c is set to satisfy

$$\min_{-\infty < x_{p_1}, x_{p_2} < \infty} \mathrm{P}\left\{x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, 2\right\} = 1 - \alpha.$$

Note that

$$\mathrm{P}\left\{x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, 2\right\} = \mathrm{P}\left\{\frac{|(\boldsymbol{x_p})^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|}{\sqrt{(\boldsymbol{x_p})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_p}}} < c \; i = 1, 2\right\} = \mathrm{P}\left\{|\boldsymbol{Z}_i| < \mathrm{c} \; i = 1, 2\right\}$$

where $\boldsymbol{Z}_i = \frac{(\boldsymbol{x_{p_i}})^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{(\boldsymbol{x_{p_i}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{p_i}}}}$ is a standard normal variable due to (1.20). We then appeal to Sidaks Inequality (cf. Hsu (1996)) which states that for any n values of $\boldsymbol{Z}_i$ and c_i

$$\mathrm{P}\left\{|\boldsymbol{Z}_i| < c_i \text{ for } i = 1, \ldots, n\right\} \geq \prod_{i=1}^{n} \mathrm{P}\left\{|\boldsymbol{Z}_i| < c_i\right\}.$$

It is then immediate that

$$\min_{-\infty < x_{p_1}, x_{p_2} < \infty} \mathrm{P}\left\{x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, 2\right\} = \left(\mathrm{P}\left\{|\boldsymbol{Z}_i| < \mathrm{c}\right\}\right)^2,$$

provided there exists $\boldsymbol{x}_{p_1}$ and $\boldsymbol{x}_{p_2}$ for which Sidaks lower bound is reached. By the principles of the normal distribution this must occur when $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are independent random variables, that is when

$$\begin{aligned} Cov(\boldsymbol{Z}_1, \boldsymbol{Z}_2) &= Cov\left(\frac{(\boldsymbol{x_{p_1}})^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{(\boldsymbol{x_{p_1}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{p_1}}}}, \frac{(\boldsymbol{x_{p_2}})^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{(\boldsymbol{x_{p_2}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{p_2}}}}\right) \\ &= \frac{\boldsymbol{x}_{p_1}^\intercal \, Cov(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}) \boldsymbol{x}_{p_2}}{\sqrt{(\boldsymbol{x_{p_1}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{p_1}}} \sqrt{(\boldsymbol{x_{p_2}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{p_2}}}} \\ &= \frac{\boldsymbol{x}_{p_1}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}_{p_2}}{\sqrt{(\boldsymbol{x_{p_1}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{p_1}}} \sqrt{(\boldsymbol{x_{p_2}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{p_2}}}} = 0 \end{aligned}$$

i.e when $\boldsymbol{x}_{p_1}^{\intercal}\boldsymbol{J}^{-1}\boldsymbol{x}_{p_2}=0$. Since we allow the effective doses $\boldsymbol{x}_{p_1}$ and $\boldsymbol{x}_{p_2}$ to lie over the whole real space, there must exist $\boldsymbol{x}_{p_1}$ and $\boldsymbol{x}_{p_2}$ such that $\boldsymbol{x}_{p_1}^{\intercal}\boldsymbol{J}^{-1}\boldsymbol{x}_{p_2}=0$, and as such the lower bound can be reached. Therefore, c is set to satisfy,

$$\left\{\mathrm{P}\left\{|\boldsymbol{Z}_i|<\mathrm{c}\right\}\right\}^2=1-\alpha$$

and so

$$\mathrm{c}=z^{\frac{\alpha_2}{2}}=z^{\frac{1-\sqrt{1-\alpha}}{2}},$$

where (z^{β}) represents the $1-\beta$'th quantile of the standard normal distribution. We note that this method applies for any two effective doses for a univariate model, and for any two CED's for multiple covariates, which offers significant utility. As the MED is a set, establishing a confidence bound would be a significantly different challenge. Clearly, for an equivalent model, a smaller c generates a more tightly bounded confidence set, hence this method is a clear improvement. Noting that c is independent of the design, for $q=1$ and $\alpha=0.05$, the confidence set of the original Scheffé band method of Section 4.1.3 sets c at 2.44, whilst this new confidence set has $c=2.236477$, or approximately 2.24.

4.2.2 Simultaneous Confidence Sets For Three Effective Doses

We construct confidence sets of the the same form as in (4.8) for a univariate logistic model ($q=1$), where now c is chosen so that for $k=3$.

$$\mathrm{P}\left\{x_{p_i}\in\mathbf{C}_{p_i}\text{ for }i=1,2,3\right\}\geq 1-\alpha.$$

Once again c is set to satisfy

$$\min_{-\infty<x_{p_1},x_{p_1},x_{p_3}<\infty}\mathrm{P}\left\{x_{p_i}\in\mathbf{C}_{p_i}\text{ for }i=1,2,3\right\}=1-\alpha. \tag{4.11}$$

We note that the method for $k=2$ relies on the independence of $\boldsymbol{Z}_k$. It is not possible however to achieve this for $k\geq 3$. Hence we consider a different method, specific to a one covariate model. Recall that we may write

$$\begin{aligned}
&\mathrm{P}\left\{x_{p_i}\in\mathbf{C}_{p_i}\text{ for }i=1,2,3\right\}\\
=&\mathrm{P}\left\{\frac{|\boldsymbol{x}_{p_i}^{\intercal}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})|}{\sqrt{\boldsymbol{x}_{p_i}^{\intercal}\boldsymbol{J}^{-1}\boldsymbol{x}_{p_i}}}<\mathrm{c}\text{ for }i=1,2,3\right\}\\
=&\mathrm{P}\left\{\frac{\left|\left\{\boldsymbol{P}\begin{pmatrix}1\\x_{p_i}\end{pmatrix}\right\}^{\intercal}\boldsymbol{N}\right|}{\left\|\boldsymbol{P}\begin{pmatrix}1\\x_{p_i}\end{pmatrix}\right\|}<\mathrm{c}\text{ for }i=1,2,3\right\}\\
=&\mathrm{P}\{\boldsymbol{N}\in\mathbb{V}(x_p)\text{ for }i=1,2,3\}\\
=&\mathrm{P}\{\boldsymbol{N}\in\mathbb{V}_3\}
\end{aligned}$$

where $\boldsymbol{N}$ is a standard bivariate normal random vector, and

$$\mathbb{V}(x_p) = \left\{ \boldsymbol{N} : \frac{\left| \left\{ \boldsymbol{P} \begin{pmatrix} 1 \\ x_p \end{pmatrix} \right\}^{\intercal} \boldsymbol{N} \right|}{\left\| \boldsymbol{P} \begin{pmatrix} 1 \\ x_p \end{pmatrix} \right\|} < \mathrm{c} \right\}$$

is the region given in the N-plane by the stripe bounded by the two parallel lines that are perpendicular to the directional vector $\boldsymbol{P} \begin{pmatrix} 1 \\ x_p \end{pmatrix} = \boldsymbol{P}\boldsymbol{x}_p$, and c distance from the origin, and $\mathbb{V}_3 = \cap_{i=1}^{3} \mathbb{V}(x_{p_i})$. This is an analogous concept to the striped regions found in Sections 2.2.2 and 2.3 on the $\boldsymbol{T}$ plane, and as such $\mathbb{V}_3$ is a 6 sided polygonal region, with its shape determined by the three effective doses via the directions $\boldsymbol{P}\boldsymbol{x}_1$, $\boldsymbol{P}\boldsymbol{x}_2$ and $\boldsymbol{P}\boldsymbol{x}_3$, depicted in Figure 4.2.

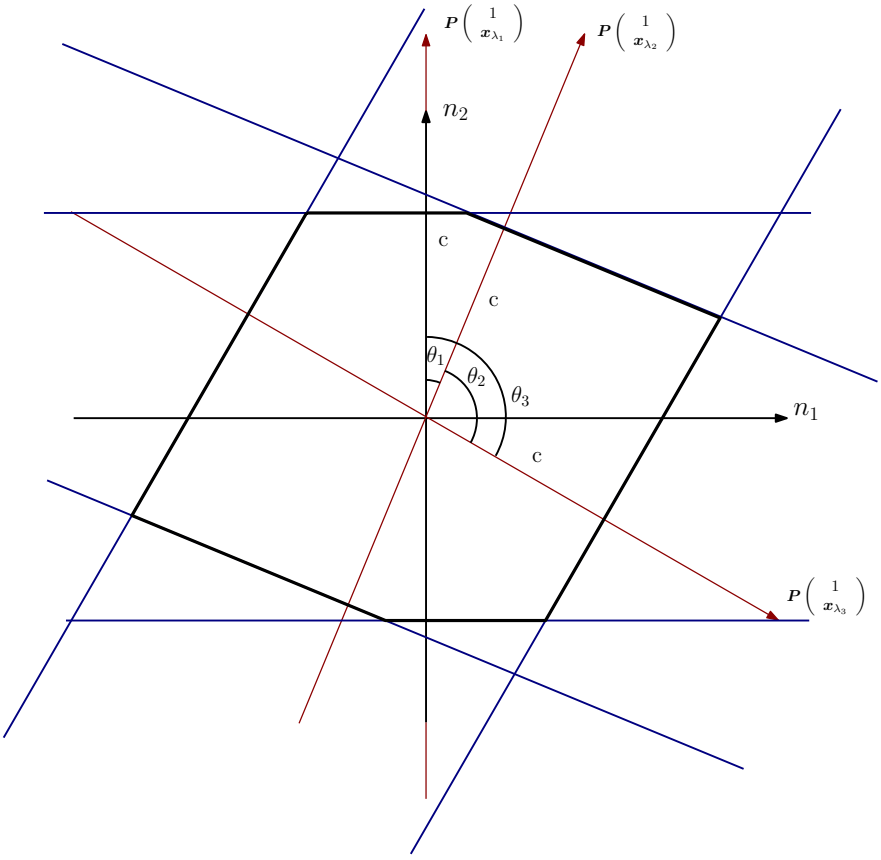


Figure 4.2: The region $\mathbb{V}_3$

We make the following observations on $\mathbb{V}_3$.

1. Just like the similar construct $R_{3,2}$, $\mathbb{V}_3$ is rotation invariant around the origin, furthermore the ordering of the directions $\boldsymbol{P}\boldsymbol{x}_i$ is arbitrary. Hence we can assume without loss of generality that $\mathbb{V}_3$ always takes shape as in Figure 4.2.

2. The probability of $\boldsymbol{N}$ in $\mathbb{V}_3$ depends on the effective doses through the angles between the three directions, θ_1 and θ_3 as in Figure 4.2. To minimise the expression

in (4.11), it is sufficient to minimise $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_3\}$ with respect to θ_1 and θ_3.

3. If $-\infty < x_{p_1}, x_{p_1}, x_{p_3} < \infty$, it is easy to see that the largest angle between any two directions is π, hence we have $0 < \theta_1 < \theta_3 \leq \pi$.

From Figure 4.3 below, it is clear $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_3\}$ can be expressed as the probability of N in the parallelogram region ABCD, which is clearly given by the intersection of the sets $\mathbb{V}(x_{p_1}) \cap \mathbb{V}(x_{p_3})$, less the probability of N lying in the grey shaded region. The grey shaded region is given by $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}(x_{p_1}) \cap \mathbb{V}(x_{p_3}) \cap \mathbb{V}^{\mathrm{c}}(x_{p_2})\}$ where

$$\mathbb{V}^{c}(x_p) = \left\{ \boldsymbol{N} : \frac{\left| \left\{ \boldsymbol{P} \begin{pmatrix} 1 \\ x_p \end{pmatrix} \right\}^{\mathsf{T}} \boldsymbol{N} \right|}{\left\| \boldsymbol{P} \begin{pmatrix} 1 \\ x_p \end{pmatrix} \right\|} > \mathrm{c} \right\}$$

is the compliment set of $\mathbb{V}(x_p)$.

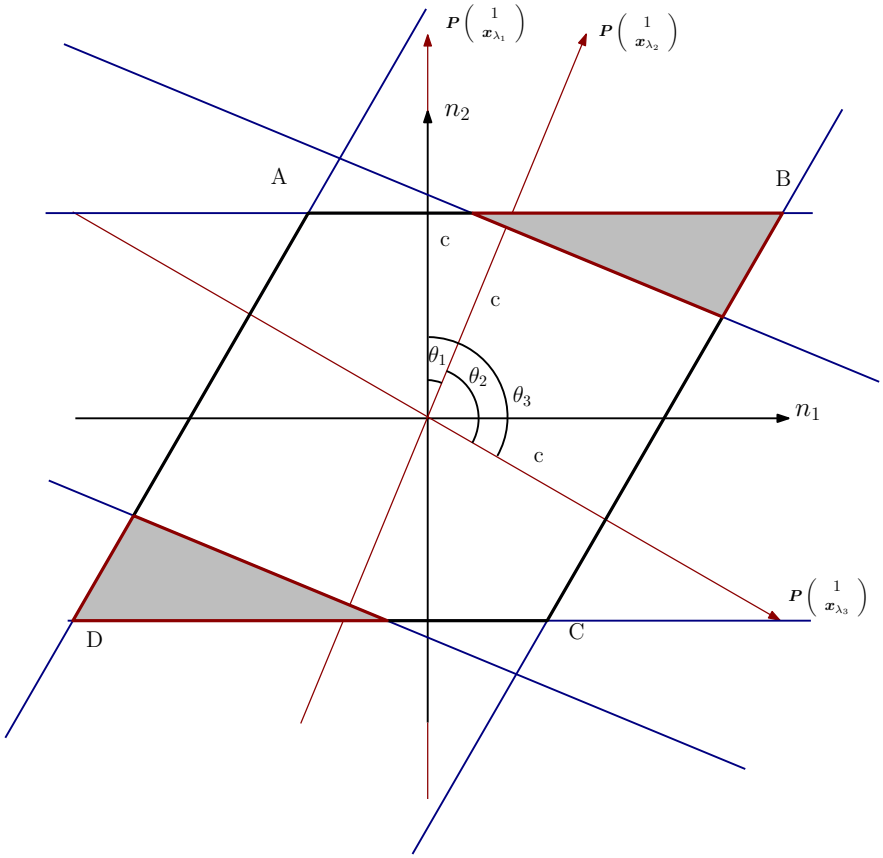


Figure 4.3: An alternative expression of the region $\mathbb{V}_3$

We first derive an expression for the probability of $\boldsymbol{N}$ in the parallelogram ABCD, in a way similar to Section 2.3 of Liu (2011), and section 2.3 of the thesis. The parallelogram ABCD is depicted in Figure 4.4 after rotation so that two sides of the parallelogram are perpendicular to the n_1 axis with additional angles of interest.

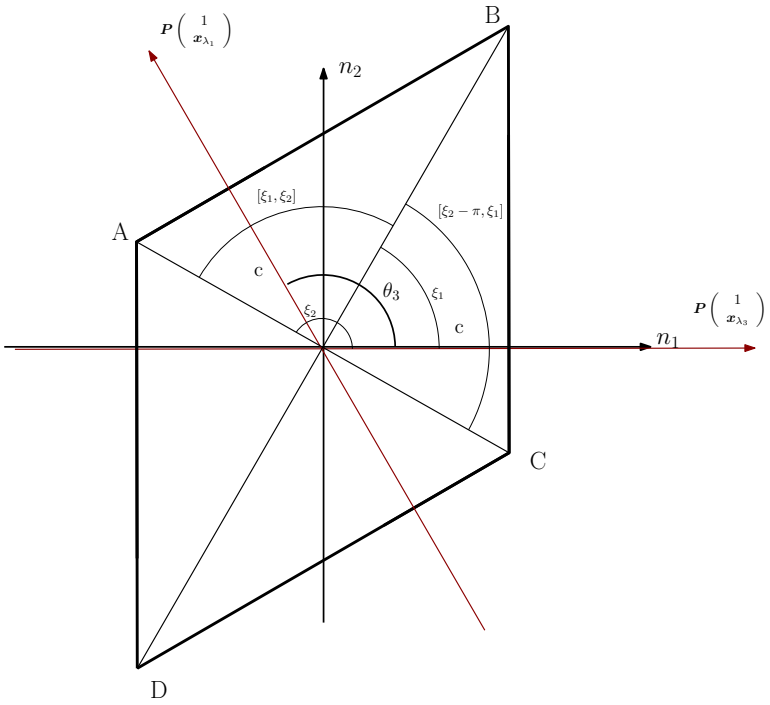


Figure 4.4: The parallelogram region with vertices ABCD.

It is clear that the problem is equivalent to the calculation of the probability of $\boldsymbol{T}$ in $R_{3,2}$ in Section 2.3, but on the $\boldsymbol{N}$ plane. The two additional angles ξ_1 and ξ_2, correspond to the angles ξ_1 and η_1 in Section 2.3, and are given as

$$\xi_1 = arcsin\left(\frac{c - ccos(\theta_3)}{\sqrt{c^2 + c^2 - 2c^2cos(\theta_3)}}\right)$$

$$\xi_2 = arccos\left(\frac{-csin(\theta_3)}{\sqrt{c^2 + c^2 + 2c^2cos(\theta_3)}}\right).$$

Note that the parallelogram ABCD is in fact a diamond, hence it is straightforward to show that

$$\xi_1 = \frac{\theta_3}{2} \text{ and } \xi_2 = \frac{\pi + \theta_3}{2}.$$

It then follows directly from the same method of section 2.3, and the preliminaries of section 2.1 in particular the distribution $\boldsymbol{F}_{\boldsymbol{R_N}}(x)$, that

$$\begin{aligned}
&\mathrm{P}(\boldsymbol{N} \in ABCD) \\
=&\frac{1}{\pi}\left\{\int_{\xi_2-\pi}^{\xi_1}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta + \int_{\xi_1-\theta_3}^{\xi_2-\theta_3}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\} \\
=&\frac{1}{\pi}\left\{\int_{\frac{\theta_3-\pi}{2}}^{\frac{\theta_3}{2}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta + \int_{-\frac{\theta_3}{2}}^{\frac{\pi-\theta_3}{2}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\}.
\end{aligned} \tag{4.12}$$

To calculate the probability of $\boldsymbol{N}$ in the grey shaded region, we refer to Figure 4.5, an enlargement of the top right section of Figure 4.3, with additional details on required

angles. The distances l_1 and l_2 are the distances in the n_1 axis from the origin and l_3 is the distance in the n_2 axis from the origin.

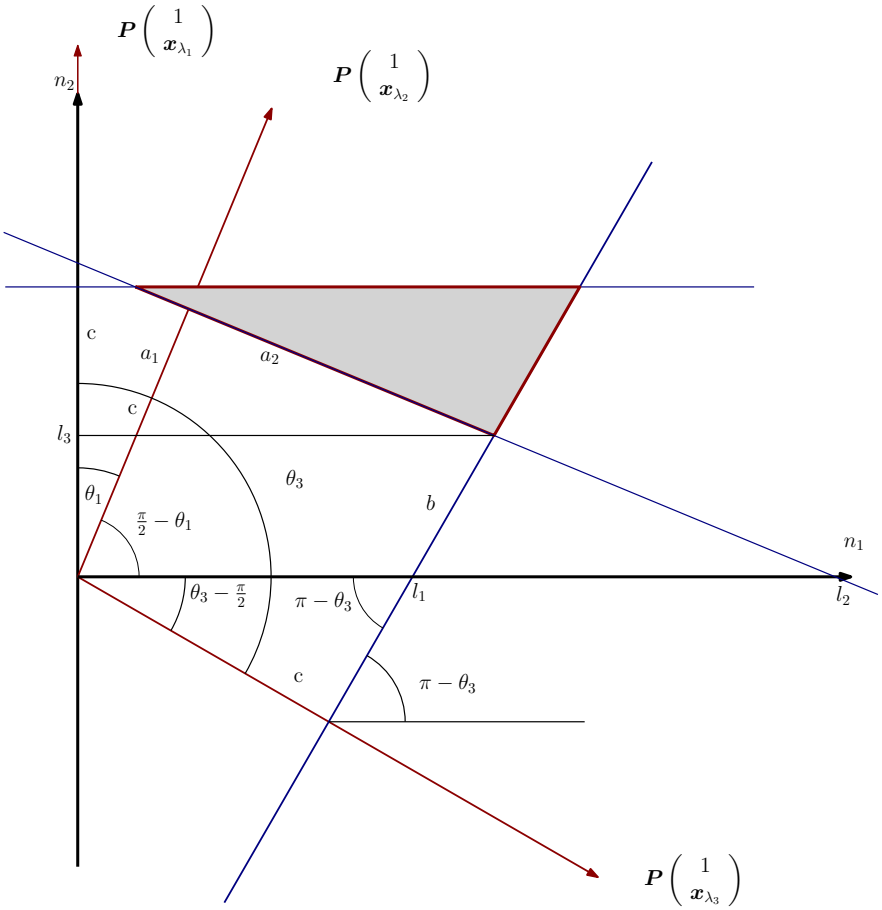


Figure 4.5: An enlargement of the shaded region of Figure 4.3.

It is clear that the upper grey region is bounded by three straight lines, line a_2, line b and line $n_2 = c$. Since line a_2 passes through the point $(ccos(\frac{\pi}{2}-\theta_1), csin(\frac{\pi}{2}-\theta_1))$ and is perpendicular to the vector $(cos(\frac{\pi}{2}-\theta_1), sin(\frac{\pi}{2}-\theta_1))$, its equation is given by

$$\Big[(n_1,n_2) - \Big(ccos(\frac{\pi}{2}-\theta_1), csin(\frac{\pi}{2}-\theta_1)\Big)\Big]\Big[cos(\frac{\pi}{2}-\theta_1), sin(\frac{\pi}{2}-\theta_1)\Big]^\intercal = 0,$$

which simplifies to

$$n_1 = -n_2 cot(\theta_1) + \frac{c}{sin(\theta_1)}.$$

Similarly, line b passes through the point $(ccos(\theta_3-\frac{\pi}{2}), -csin(\theta_3-\frac{\pi}{2}))$ and is perpendicular to $(cos(\theta_3-\frac{\pi}{2}), -sin(\theta_3-\frac{\pi}{2}))$ and so its equation is given by

$$n_1 = -n_2 cot(\theta_3) + \frac{c}{sin(\theta_3)}.$$

Now, l_3 is simply the n_2 coordinate of the intersection of a_2 and b, and can be solved from

$$-l_3 cot(\theta_3) + \frac{c}{sin(\theta_3)} = -l_3 cot(\theta_1) + \frac{c}{sin(\theta_1)}.$$

This gives

$$l_3 = \frac{c(sin(\theta_3) - \sin(\theta_1))}{sin(\theta_3 - \theta_1)}.$$

It follows therefore that

$$\begin{aligned}
&\mathrm{P}\{\boldsymbol{N} \in \text{Grey Shaded Region}\} \\
=&2\int_{\frac{c(sin(\theta_3)-\sin(\theta_1))}{sin(\theta_3-\theta_1)}}^{c} \int_{-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}}^{-n_2cot(\theta_3)+\frac{c}{sin(\theta_3)}} \phi(n_1)\phi(n_2)\mathrm{d}n_1\mathrm{d}n_2 \\
=&2\int_{l_3}^{c} \phi(n_2)\left[\Phi\left(-n_2cot(\theta_3)+\frac{c}{sin(\theta_3)}\right) - \Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]\mathrm{d}n_2. \quad (4.13)
\end{aligned}$$

We therefore must have

$$\begin{aligned}
&\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_3\} \\
&= \frac{1}{\pi}\left\{\int_{\xi_2-\pi}^{\xi_1}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta + \int_{\xi_1-\theta_3}^{\xi_2-\theta_3}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\} \\
&- 2\int_{l_3}^{c} \phi(n_2)\left[\Phi\left(-n_2cot(\theta_3)+\frac{c}{sin(\theta_3)}\right) - \Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]\mathrm{d}n_2. \quad (4.14)
\end{aligned}$$

Minimising the equation in (4.11) has therefore been reduced to minimising Equation (4.14) with respect to $0 < \theta_1 < \theta_3 \leq \pi$. This is given by the following theorem.

Theorem 4.1. *Expression (4.14), and consequently* $P\{\boldsymbol{N} \in \mathbb{V}_3\}$ *is minimised at* $\theta_1 = \frac{\pi}{3}$ *and* $\theta_3 = \frac{2\pi}{3}$

The proof of this Theorem hinges on the following Theorem below.

Theorem 4.2. *For given* $0 < \theta_3^* \leq \pi$*, the probability 4.14 is minimised with respect to* $\theta_1 \in (0, \theta_3)$*, when* $\theta_1 = \frac{\theta_3^*}{2}$*.*

Proof Of Theorem 4.2

Note that the probability of $\boldsymbol{N}$ lying in the parallelogram ABCD depends on θ_3, but not θ_1. As a result, with a fixed value of θ_3, Equation (4.14) varies only with the probability of lying in the grey shaded region. It is therefore clear that minimising 4.14 with respect to θ_1, given θ_3 fixed at some θ_3^*, is equivalent to maximising the probability of $\boldsymbol{N}$ lying in the grey shaded region, that is

$$\underset{0<\theta_1<\theta_3=\theta_3^*}{\text{Maximise}} \left\{\int_{l_3}^{c} \phi(n_2)\left[\Phi\left(-n_2cot(\theta_3^*)+\frac{c}{sin(\theta_3^*)}\right) - \Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]\mathrm{d}n_2\right\}. \quad (4.15)$$

Hence, Theorem 4.2 is proven if it can be established that $\theta_1 = \frac{\theta_3^*}{2}$ is a stationary point of of the expression in (4.15), and is a maximum, which is immediately given by the following two Sublemmas.

Sublemma 4.1. *We have*

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\theta_1}\left\{\int_{l_3}^{c}\phi(n_2)\left[\Phi\left(-n_2cot(\theta_3^*)+\frac{c}{sin(\theta_3^*)}\right)-\Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]\mathrm{d}n_2\right\}\\
=&\frac{1}{\sqrt{2\pi}}exp\left\{-\frac{c^2}{2}\right\}\left[\phi\left(\frac{c-ccos(\theta_1)}{sin(\theta_1)}\right)-\phi\left(\frac{\frac{c(sin(\theta_3^*)-\sin(\theta_1))}{sin(\theta_3^*-\theta_1)}-ccos(\theta_1)}{sin(\theta_1)}\right)\right]. \qquad (4.16)
\end{aligned}$$

Sublemma 4.2. *As a function of $\theta_1 \in (0,\theta_3^*)$ with $\theta_3^* < \pi$, the expression (4.16), denoted $D(\theta_1)$, has only one zero point at $\theta_1 = \frac{\theta_3^*}{2}$. Furthermore, $D(\theta_1) > 0$ for $\theta_1 \in (0, \frac{\theta_3^*}{2})$ and $D(\theta_1) < 0$ for $\theta_1 \in (\frac{\theta_3^*}{2}, \theta_3)$.*

These two results are proven below. We begin with Sublemma 4.1, label the expression inside the integral of Equation (4.15) as follows

$$\phi(n_2)\left[\Phi\left(-n_2cot(\theta_3^*)+\frac{c}{sin(\theta_3^*)}\right)-\Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]=g(n_2,\theta_1).$$

Noting that $l_3 = \frac{c(sin(\theta_3^*)-\sin(\theta_1))}{sin(\theta_3^*-\theta_1)}$ is now specifically a function of θ_1, we apply the Leibniz integral rule of the differential, given as

$$\frac{\mathrm{d}}{\mathrm{d}z}\int_{a(z)}^{b(z)}f(x,z)\mathrm{d}x=\int_{a(z)}^{b(z)}\frac{\mathrm{d}f(x,z)}{\mathrm{d}z}\mathrm{d}x+f(b(z),z)\frac{\mathrm{d}b}{\mathrm{d}z}-f(a(z),z)\frac{\mathrm{d}a}{\mathrm{d}z}.$$

Then the differential of $g(n_2,\theta_1)$ may be reduced to

$$\frac{\mathrm{d}}{\mathrm{d}\theta_1}\int_{l_3(\theta_1)}^{c}g(n_2,\theta_1)\mathrm{d}n_2=\int_{l_3(\theta_1)}^{c}\frac{\mathrm{d}g(n_2,\theta_1)}{\mathrm{d}\theta_1}\mathrm{d}n_2-g(l_3,\theta_1)\frac{\mathrm{d}l_3(\theta_1)}{\mathrm{d}\theta_1}.$$

Further noting that

$$g(l_3,\theta_1)=\phi(l_3)\left[\Phi\left(-l_3cot(\theta_3^*)+\frac{c}{sin(\theta_3^*)}\right)-\Phi\left(-l_3cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]=0$$

since $-l_3cot(\theta_3^*)+\frac{c}{sin(\theta_3^*)} = (-l_3cot(\theta_1)+\frac{c}{sin(\theta_1)}$, we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\theta_1}\left\{\int_{l_3}^{c}\phi(n_2)\left[\Phi\left(-n_2cot(\theta_3^*)+\frac{c}{sin(\theta_3^*)}\right)-\Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]\mathrm{d}n_2\right\}\\
=&\int_{l_3}^{c}\frac{\mathrm{d}}{\mathrm{d}\theta_1}\phi(n_2)\left[\Phi\left(-n_2cot(\theta_3^*)+\frac{c}{sin(\theta_3^*)}\right)-\Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]\mathrm{d}n_2\\
=&-\int_{l_3}^{c}\frac{\mathrm{d}}{\mathrm{d}\theta_1}\phi(n_2)\Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\mathrm{d}n_2\\
=&\int_{l_3}^{c}-\phi(n_2)\frac{\mathrm{d}}{\mathrm{d}\theta_1}\left[\Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]\mathrm{d}n_2\\
=&\int_{l_3}^{c}-\phi(n_2)\phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\frac{\mathrm{d}}{\mathrm{d}\theta_1}\left[-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right]\mathrm{d}n_2\\
=&\int_{l_3}^{c}\phi(n_2)\phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\left[\frac{c(cos(\theta_1))-n_2}{sin(\theta_1)^2}\right]\mathrm{d}n_2. \qquad (4.17)
\end{aligned}$$

We now simplify the term in Equation (4.17) given by

$$
\begin{aligned}
&\phi(n_2)\phi\left(-n_2 cot(\theta_1) + \frac{c}{sin(\theta_1)}\right) \\
=&\frac{1}{2\pi}\exp\left\{-\frac{n_2^2}{2} - \frac{(-n_2 cot(\theta_1) + \frac{c}{sin(\theta_1)})^2}{2}\right\} \\
=&\frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(n_2^2 + \frac{c^2}{sin(\theta_1)^2} - \frac{2cn_2 cos(\theta_1)}{sin(\theta_1)^2} + n_2^2 cot(\theta_1)^2\right)\right\} \\
=&\frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(n_2^2(1 + cot(\theta_1)^2) + \frac{c^2}{sin(\theta_1)^2} - \frac{2cn_2 cos(\theta_1)}{sin(\theta_1)^2}\right)\right\} \\
=&\frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(\frac{n_2^2}{sin(\theta_1)^2} + \frac{c^2}{sin(\theta_1)^2} - \frac{2cn_2 cos(\theta_1)}{sin(\theta_1)^2}\right)\right\} \\
=&\frac{1}{2\pi}\exp\left\{-\frac{1}{2sin(\theta_1)^2}\left[(n_2 - ccos(\theta_1))^2 + c^2 - (c^2 cos(\theta_1)^2)\right]\right\} \\
=&\frac{1}{2\pi}\exp\left\{-\frac{1}{2sin(\theta_1)^2}\left[(n_2 - ccos(\theta_1))^2 + c^2 sin(\theta_1)^2\right]\right\} \\
=&\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{c^2}{2}\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2sin(\theta_1)^2}\left[(n_2 - ccos(\theta_1))^2\right]\right\} \\
=&\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{c^2}{2}\right\}\phi(z)
\end{aligned}
$$

where $z = \frac{n_2 - ccos(\theta_1)}{sin(\theta_1)}$ and $\phi(z)$ is a standard normal pdf for z. Equation (4.17) now becomes

$$
\begin{aligned}
&\int_{l_3}^{c}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{c^2}{2}\right\}\left[\frac{c(cos(\theta_1)) - n_2}{sin(\theta_1)^2}\right]\phi(z)\mathrm{d}n_2 \\
=&\int_{l_3}^{c}\frac{ccos(\theta_1)}{sin(\theta_1)^2\sqrt{2\pi}}\exp\left\{-\frac{c^2}{2}\right\}\phi(z)\mathrm{d}n_2 - \int_{l_3}^{c}\frac{n_2}{sin(\theta_1)^2\sqrt{2\pi}}\exp\left\{-\frac{c^2}{2}\right\}\phi(z)\mathrm{d}n_2.
\end{aligned}
\tag{4.18}
$$

Equation (4.18) consists of two terms, which we evaluate individually. For the first term we change the variable of integration to z, noting that

$$
\frac{\mathrm{d}z}{\mathrm{d}n_2} = \frac{1}{sin(\theta_1)}
$$

which gives

$$
\begin{aligned}
&\int_{l_3}^{c}\frac{ccos(\theta_1)}{sin(\theta_1)^2\sqrt{2\pi}}\exp\left\{-\frac{c^2}{2}\right\}\phi(z)\mathrm{d}n_2 \\
=&\int_{\frac{l_3 - ccos(\theta_1)}{sin(\theta_1)}}^{\frac{c - ccos(\theta_1)}{sin(\theta_1)}}\frac{ccos(\theta_1)}{sin(\theta_1)\sqrt{2\pi}}\exp\left\{-\frac{c^2}{2}\right\}\phi(z)\mathrm{d}z \\
=&\left[\frac{ccos(\theta_1)}{sin(\theta_1)\sqrt{2\pi}}\exp\left\{-\frac{c^2}{2}\right\}\Phi(z)\right]_{t_2}^{t_1} \\
=&\frac{ccos(\theta_1)}{sin(\theta_1)\sqrt{2\pi}}\exp\left\{-\frac{c^2}{2}\right\}[\Phi(t_1) - \Phi(t_2)]
\end{aligned}
\tag{4.19}
$$

where $t_1 = \frac{c-ccos(\theta_1)}{sin(\theta_1)}$ and $t_2 = \frac{l_3-ccos(\theta_1)}{sin(\theta_1)}$. For the second term, we apply integration by parts to give

$$\begin{aligned}
&\int_{l_3}^{c} \frac{n_2}{sin(\theta_1)^2\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \phi(z) \mathrm{d}n_2 \\
=&\left[\frac{n_2}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(z)\right]_{l_3}^{c} - \int_{l_3}^{c} \frac{1}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(z) \mathrm{d}n_2 \\
=&\frac{1}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} [n_2\Phi(z)]_{l_3}^{c} - \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \int_{t_2}^{t_1} \Phi(z) \mathrm{d}z \\
=&\frac{1}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} [n_2\Phi(z)]_{l_3}^{c} - \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} [z\Phi(z) + \phi(z)]_{t_2}^{t_1} \qquad (4.20)
\end{aligned}$$

where the result $\int_{t_2}^{t_1} \Phi(z) \mathrm{d}z = [z\Phi(z) + \phi(z)]_{t_2}^{t_1}$ is an immediate consequence of integration by parts, given by Section 4.1, result 1, of Edward and Murray (1969). By using the expressions (4.19) and (4.20), (4.18) can be expressed as,

$$\begin{aligned}
&\frac{ccos(\theta_1)}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} [\Phi(z)]_{t_2}^{t_1} - \frac{1}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} [n_2\Phi(z)]_{l_3}^{c} \\
+&\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} [z\Phi(z) + \phi(z)]_{t_2}^{t_1} \\
=&\left[\left(\frac{ccos(\theta_1)}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_1)\right) - \left(\frac{ccos(\theta_1)}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_2)\right)\right] \\
-&\left[\left(\frac{c}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_1)\right) - \left(\frac{l_3}{sin(\theta_1)\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_2)\right)\right] \\
+&\left[\left(\frac{t_1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_1)\right) - \left(\frac{t_2}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_2)\right)\right] \\
+&\left[\left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \phi(t_1)\right) - \left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \phi(t_2)\right)\right] \\
=&-\left[\left(\frac{t_1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_1)\right) - \left(\frac{t_2}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_2)\right)\right] \\
+&\left[\left(\frac{t_1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_1)\right) - \left(\frac{t_2}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \Phi(t_2)\right)\right] \\
+&\left[\left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \phi(t_1)\right) - \left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \phi(t_2)\right)\right] \\
=&\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} [\phi(t_1) - \phi(t_2)] \\
=&\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{c^2}{2}\right\} \left[\phi\left(\frac{c - ccos(\theta_1)}{sin(\theta_1)}\right) - \phi\left(\frac{\frac{c(sin(\theta_3^*) - \sin(\theta_1))}{sin(\theta_3^* - \theta_1)} - ccos(\theta_1)}{sin(\theta_1)}\right)\right] \qquad (4.21)
\end{aligned}$$

which proves Sublemma 4.1.

We now prove Sublemma 4.2. Denote Equation (4.21) (also given as Equation (4.16)) as $D(\theta_1)$. This expression is the difference of two standard normal pdf values, multiplied by some constant term. Consequently, the behaviour of the differential with respect to θ_1 hinges on the relationship between the absolute values of the two terms inside the pdf's. Specifically, we are comparing

$$\left|\frac{c - ccos(\theta_1)}{sin(\theta_1)}\right| \text{ to } \left|\frac{\frac{c(sin(\theta_3^*) - \sin(\theta_1))}{sin(\theta_3^* - \theta_1)} - ccos(\theta_1)}{sin(\theta_1)}\right|.$$

Since $\phi(x) - \phi(y) \geq 0$ if and only if $|x| - |y| \leq 0$ we have

$$D(\theta_1) \leq 0 \Leftrightarrow d(\theta_1) \geq 0$$

$$\begin{aligned}\text{where } d(\theta_1) = &\left|\frac{c - ccos(\theta_1)}{sin(\theta_1)}\right| - \left|\frac{\frac{c(sin(\theta_3^*) - \sin(\theta_1))}{sin(\theta_3^* - \theta_1)} - ccos(\theta_1)}{sin(\theta_1)}\right| \\ =&\frac{c}{|sin(\theta_1)|}|1 - cos(\theta_1)| - \left|\frac{sin(\theta_3^*) - \sin(\theta_1)}{sin(\theta_3^* - \theta_1)} - cos(\theta_1)\right|.\end{aligned}$$

Since $c > 0$ and $\theta_1 \in (0, \pi)$, it is sufficient to focus on the sign of

$$\begin{aligned}g(\theta_1) = &|1 - cos(\theta_1)| - \left|\frac{sin(\theta_3^*) - \sin(\theta_1)}{sin(\theta_3^* - \theta_1)} - cos(\theta_1)\right| \\ =&(1 - cos(\theta_1)) - \left|\frac{sin(\theta_3^*) - \sin(\theta_1)}{sin(\theta_3^* - \theta_1)} - cos(\theta_1)\right|\end{aligned}$$

as $(1 - cos(\theta_1)$ must always be positive. We may also remove the absolute value on the second term of $g(\theta_1)$ by noting that

$$\begin{aligned}&\frac{sin(\theta_3^*) - \sin(\theta_1)}{sin(\theta_3^* - \theta_1)} - cos(\theta_1) \\ =&\frac{sin(\theta_3^*) - \sin(\theta_1) - cos(\theta_1)sin(\theta_3^* - \theta_1)}{sin(\theta_3^* - \theta_1)} \\ =&\frac{sin(\theta_3^*) - \sin(\theta_1) - [sin(\theta_3^*) - sin(\theta_1)cos(\theta_3^* - \theta_1)]}{sin(\theta_3^* - \theta_1)} \\ =&\frac{sin(\theta_1)}{sin(\theta_3^* - \theta_1)}(cos(\theta_3^* - \theta_1) - 1)\end{aligned}$$

which must clearly always be negative as $0 < \theta_3^* - \theta_1 \leq \pi$ and so

$$\begin{aligned}
&g(\theta_1) = 1 - cos(\theta_1) - \left(cos(\theta_1) + \frac{sin(\theta_1) - \sin(\theta_3^*)}{sin(\theta_3^* - \theta_1)} \right) \\
=&1 - 2cos(\theta_1) - \frac{sin(\theta_1) + \sin(\theta_3^*)}{sin(\theta_3^* - \theta_1)} \\
=&\frac{sin(\theta_3^* - \theta_1) - 2cos(\theta_1)sin(\theta_3^* - \theta_1) - sin(\theta_1) + \sin(\theta_3^*)}{sin(\theta_3^* - \theta_1)} \\
=&\frac{sin(\theta_3^* - \theta_1) - [sin(\theta_1 + (\theta_3^* - \theta_1)) - sin(\theta_1 - (\theta_3^* - \theta_1))] - sin(\theta_1) + \sin(\theta_3^*)}{sin(\theta_3^* - \theta_1)} \\
=&\frac{sin(\theta_3^* - \theta_1) - (sin(\theta_3^*) - sin(2\theta_1 - \theta_3^*)) - sin(\theta_1) + \sin(\theta_3^*)}{sin(\theta_3^* - \theta_1)} \\
=&\frac{sin(\theta_3^* - \theta_1) + sin(2\theta_1 - \theta_3^*) - sin(\theta_1)}{sin(\theta_3^* - \theta_1)}
\end{aligned}$$

using the product to sum formulae. It is clear that, at $\theta_1 = \frac{\theta_3 *}{2}$, $g(\theta_1) = 0$ and therefore $D(\theta_1) = 0$. Thus $\theta_1 = \frac{\theta_3 *}{2}$ is a stationary point of $D(\theta_1)$. For the sign of $g(\theta_1)$ over $0 < \theta_1 < \theta_3^* < \pi$, it suffices to focus on the numerator

$$h(\theta_1) = sin(\theta_3^* - \theta_1) + sin(2\theta_1 - \theta_3^*) - sin(\theta_1).$$

Let $\theta_1 = \frac{\theta_3^*}{2} - \epsilon$ for $\epsilon \in (0, \frac{\theta_3^*}{2})$, then

$$\begin{aligned}
&h(\frac{\theta_3^*}{2} - \epsilon) \\
=&sin(\theta_3^* - \theta_1) + sin(2\theta_1 - \theta_3^*) - sin(\theta_1)|_{\theta_1 = \frac{\theta_3^*}{2} - \epsilon} \\
=&sin(\theta_3^* - (\frac{\theta_3^*}{2} - \epsilon)) + sin(2(\frac{\theta_3^*}{2} - \epsilon) - \theta_3^*) - sin(\frac{\theta_3^*}{2} - \epsilon) \\
=&sin(\frac{\theta_3^*}{2} + \epsilon) + sin(-2\epsilon) - sin(\frac{\theta_3^*}{2} - \epsilon) \\
=&sin(\frac{\theta_3^*}{2} + \epsilon) - sin(2\epsilon) - sin(\frac{\theta_3^*}{2} - \epsilon) \\
=&sin((\frac{\theta_3^*}{2} - \epsilon) + 2\epsilon) - sin(2\epsilon) - sin(\frac{\theta_3^*}{2} - \epsilon) \\
=&sin(\frac{\theta_3^*}{2} - \epsilon)cos(2\epsilon) + cos(\frac{\theta_3^*}{2} - \epsilon)sin(2\epsilon) - sin(2\epsilon) - sin(\frac{\theta_3^*}{2} - \epsilon) \\
=&sin(\frac{\theta_3^*}{2} - \epsilon)cos(2\epsilon) + sin(2\epsilon)(cos(\frac{\theta_3^*}{2} - \epsilon) - 1) - sin(\frac{\theta_3^*}{2} - \epsilon) \\
<&sin(\frac{\theta_3^*}{2} - \epsilon)cos(2\epsilon) - sin(\frac{\theta_3^*}{2} - \epsilon) < 0
\end{aligned}$$

as $(cos(\frac{\theta_3^*}{2} - \epsilon) - 1)$ must always be negative. Furthermore, from this manipulation, we immediately have

$$sin(\frac{\theta_3^*}{2} + \epsilon) - sin(2\epsilon) - sin(\frac{\theta_3^*}{2} - \epsilon) < 0.$$

It follows that

$$h(\frac{\theta_3^*}{2} + \epsilon) = -(sin(\frac{\theta_3^*}{2} + \epsilon) - sin(2\epsilon) - sin(\frac{\theta_3^*}{2} - \epsilon)) > 0.$$

This tells us that $d(\theta_1)$ is negative as the value of θ_1 approaches the stationary point, and hence $D(\theta_1)$ is positive, and that the reverse is true as we move away past the stationary point, which proves Sublemma 4.2. Theorem 4.2 follows now immediately, as Sublemma 4.2 states that the point $\theta_1 = \frac{\theta_3^*}{2}$, is a maximum point, and hence satisfies the Expression (4.15).

End Of Proof of Theorem 4.2

Proof Of Theorem 4.1

We can now prove Theorem 4.1. Due to the rotational invariance, and symmetry of $\mathbb{V}_3$, the probability of $\boldsymbol{N}$ lying in $\mathbb{V}_3$ can be expressed as in Equation (4.14) using any three adjacent directions from the six determined by the three directional vectors $\boldsymbol{P}\begin{pmatrix} 1 \\ \boldsymbol{x}_{p_1} \end{pmatrix}$, $\boldsymbol{P}\begin{pmatrix} 1 \\ \boldsymbol{x}_{p_2} \end{pmatrix}$, and $\boldsymbol{P}\begin{pmatrix} 1 \\ \boldsymbol{x}_{p_3} \end{pmatrix}$. This probability can always increase if one direction does not partition the angle formed between the other two directions in equal halves, as given by Theorem 4.2. Hence the minimum probability is attained only when all directions lie equally spaced apart over the region 2π, which means $\mathbb{V}_3$ is a regular hexagon and $\theta_1 = \frac{\pi}{3}$ and $\theta_3 = \frac{2\pi}{3}$.

End Of Proof Of Theorem 4.1

It is immediate from Theorem 4.1 that value of c that satisfies equation 4.11 must also satisfy the following condition,

$$\frac{1}{\pi}\left\{\int_{\xi_2-\pi}^{\xi_1}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta+\int_{\xi_1-\theta_3}^{\xi_2-\theta_3}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\}$$

$$-2\int_{l_3}^{c}\phi(n_2)\left[\Phi\left(-n_2cot(\theta_3)+\frac{c}{sin(\theta_3)}\right)-\Phi\left(-n_2cot(\theta_1)+\frac{c}{sin(\theta_1)}\right)\right]\mathrm{d}n_2=1-\alpha,$$

where

$$\theta_1=\frac{\pi}{3},\ \ \theta_3=\frac{2\pi}{3}\text{ and }l_3=\frac{c(sin(\theta_3)-\sin(\theta_1))}{sin(\theta_3-\theta_1)}.$$

There does not exist an analytical solution to this problem with respect to c, however since the condition involves only one dimensional integration, various numerical searches can be employed to find the correct value, for a specified value of α. Hence we find c with the bisection method using a custom code in "R", available on request.

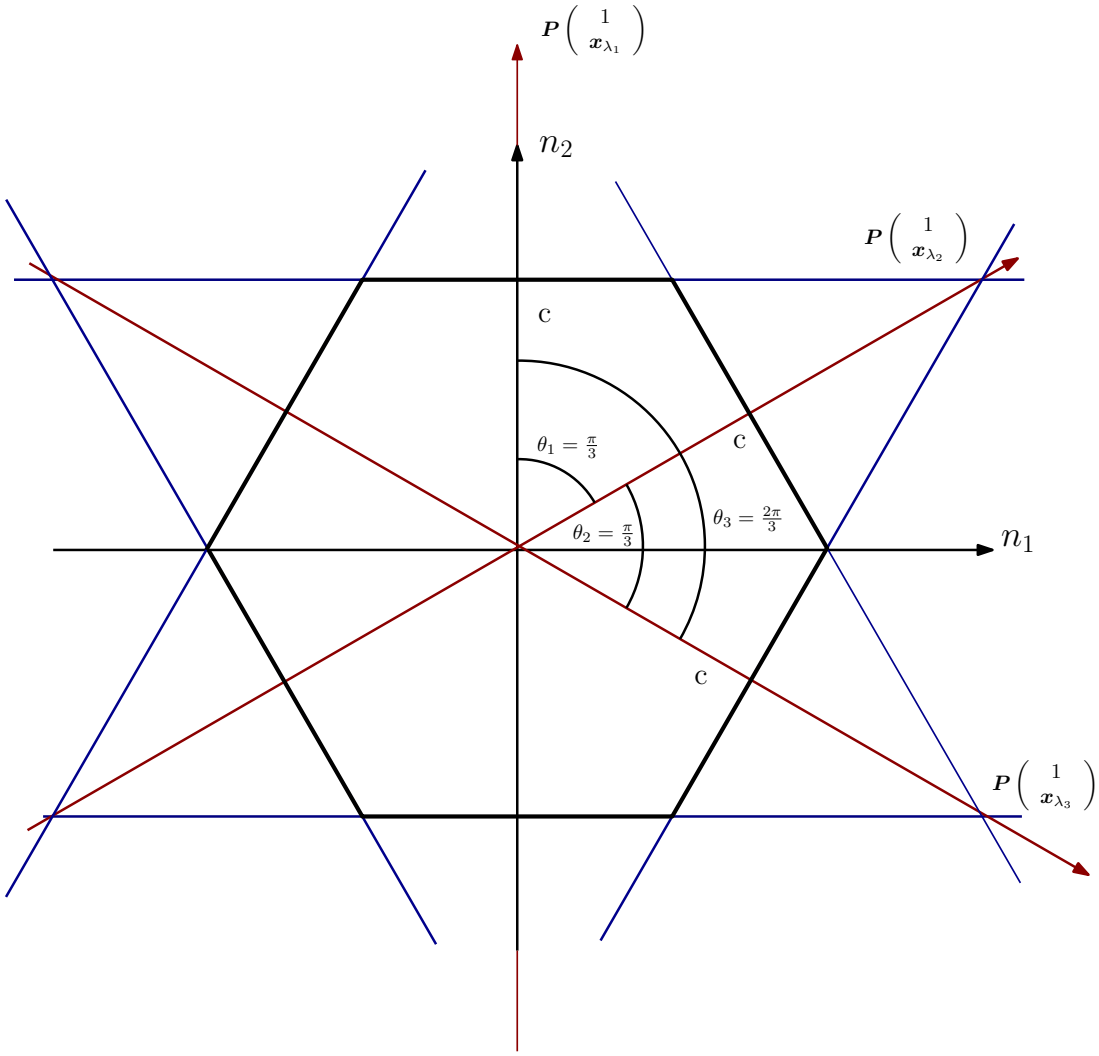


Figure 4.6: The region $\mathbb{V}_3$, as a regular hexagonal region.

4.2.3 Simultaneous Confidence Sets For More Than Three Effective Doses

We may generalise the methods to obtain simultaneous confidence sets for $k = 3$, to the case $k > 3$ for some specific k. Specifically we want to choose c such that for a particular $k > 3$

$$\min_{-\infty<x_{p_1},\ldots,x_{p_k}<\infty} \mathrm{P}\left\{x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1,\ldots,k\right\} = 1-\alpha. \tag{4.22}$$

It is immediate from the three effective dose case that we may write

$$\mathrm{P}\left\{\boldsymbol{x}_{p_i} \in \mathbf{C} \text{ for } \ i = 1,\ldots,k\right\} = \mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k\}$$

where $\mathbb{V}_k = \cap_{i=1}^{k} \mathbb{V}(x_p)$ is a $2k$ sided polygonal region on the $\boldsymbol{N}$ plane, with shaped determined by the k directions $\boldsymbol{P}\boldsymbol{x}_{p_i}$, depicted in figure 4.7 for $k = 4$.

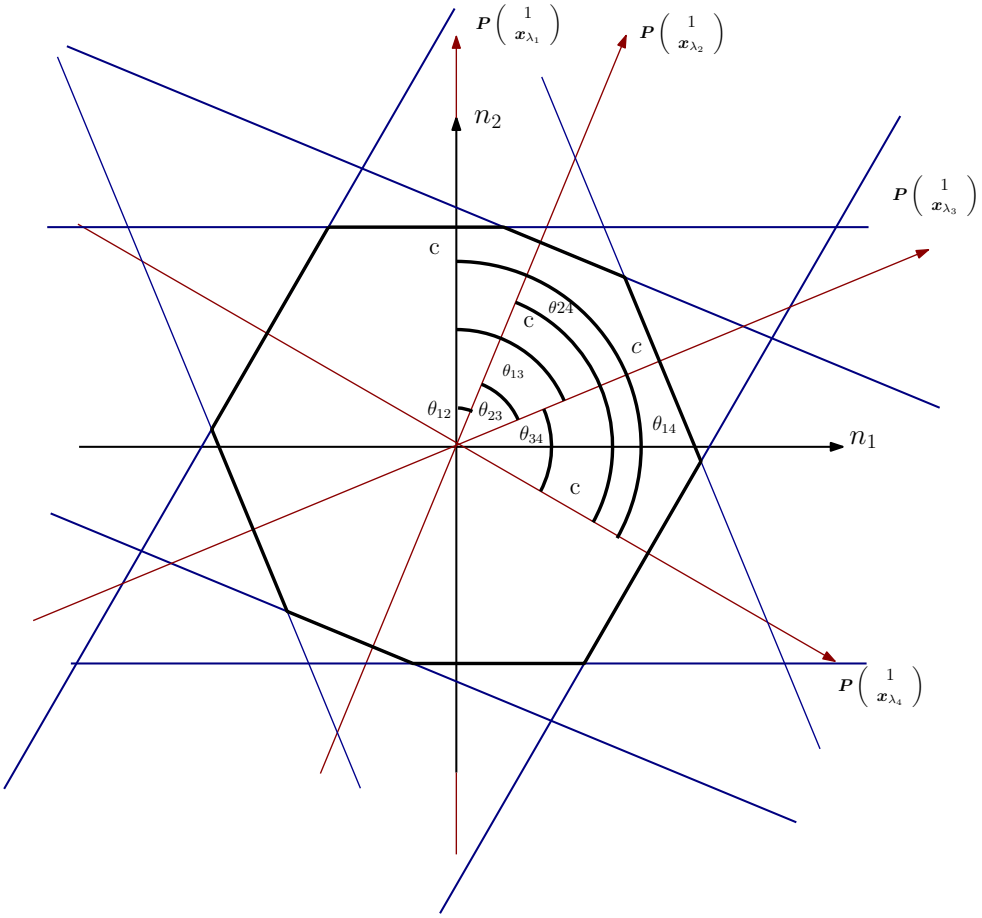


Figure 4.7: The region $\mathbb{V}_4$, an example of $\mathbb{V}_k$ for four effective doses.

As before, $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k\}$ is rotation invariant around the origin, and the ordering of the directions is arbitrary, hence we may assume without loss of generality that $\mathbb{V}_k$ is oriented as in Figure 4.7, with $\boldsymbol{P}\boldsymbol{x}_{p_i}$ lying on the n_2 axis. We denote the acute angle between any two directional vectors $\boldsymbol{P}\boldsymbol{x}_{p_i}$ and $\boldsymbol{P}\boldsymbol{x}_{p_j}$ as θ_{ij}. Hence the largest of all such angles is θ_{1k} which must be less than π. To establish c it is sufficient to minimise

$\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k\}$ with respect to $0 < \theta_{ij} \leq \pi$ $i, j = 1, \ldots, k$.

Referring to figure 4.8, we can form an expression for the probability of $\boldsymbol{N}$ in $\mathbb{V}_k$ as the probability of $\boldsymbol{N}$ in the parallelogram ABCD, which is formed by the intersection of the sets $\boldsymbol{N} \in \mathbb{V}(x_{p_1})$ and $\mathbb{V}(x_{p_k})$, less twice the probability of lying in the $k-2$ grey shaded regions of the upper right corner. These are given by the regions where the remaining $k-2$ directions $\boldsymbol{P}\boldsymbol{x}_{p_i}$ $i = 2, \ldots, k-1$ intersect the parallelogram ABCD and are given by

$$\mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}(x_{p_{i-1}}) \cap \mathbb{V}(x_{p_k}) \cap \mathbb{V}^{\mathrm{c}}(x_{p_i})\right\} \; i = 2, \ldots, k-1.$$

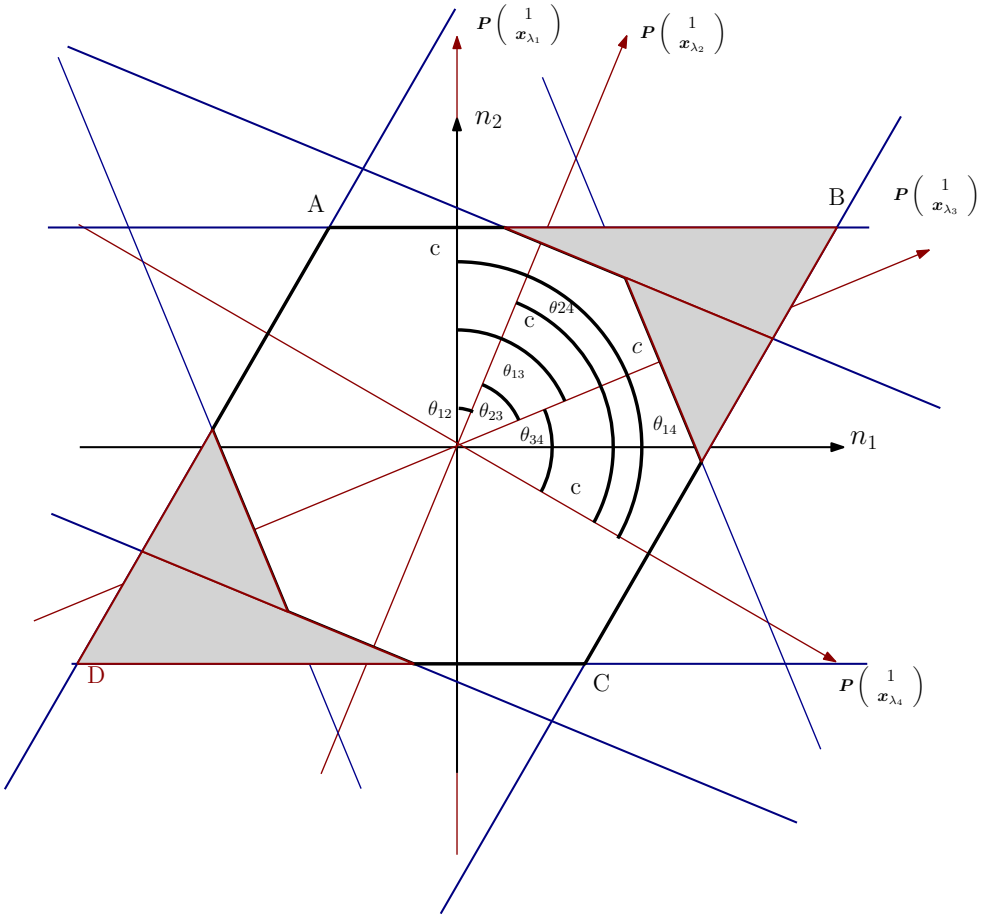


Figure 4.8: The region $\mathbb{V}_4$, as a parallelogram less the four grey shaded regions.

It is clear that $\mathrm{P}(\boldsymbol{N} \in ABCD)$ is the same problem as lying in the parallelogram region for the case $k = 3$, with the angle equal to θ_{ik} instead of θ_3, hence

$$\begin{aligned}&\mathrm{P}(\boldsymbol{N} \in ABCD) = \mathrm{P}\{\boldsymbol{N} \in \mathbb{V}(x_{p_1}) \cap \mathbb{V}(x_{p_k})\} \\ =&\frac{1}{\pi}\left\{\int_{\frac{\theta_{1k}-\pi}{2}}^{\frac{\theta_{1k}}{2}} \left(1 - \exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right) \mathrm{d}\theta + \int_{-\frac{\theta_{1k}}{2}}^{\frac{\pi-\theta_{1k}}{2}} \left(1 - \exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right) \mathrm{d}\theta\right\}.\end{aligned}$$

By rotational invariance we may place $\boldsymbol{P}\boldsymbol{x}_{p_{i-1}}$ parallel with the n_1 axis on the positive

n_2 plane, and then $\mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}(x_{p_{i-1}}) \cap \mathbb{V}(x_{p_k}) \cap \mathbb{V}^{\mathrm{c}}(x_{p_i})\right\}$ is the same problem of lying in the grey shaded region as that for $k = 3$, with $\theta_{(i-1)i}$ and θ_{1k} taking the places of θ_1 and θ_3 respectively. Hence

$$\mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}(x_{p_{i-1}}) \cap \mathbb{V}(x_{p_k}) \cap \mathbb{V}^{\mathrm{c}}(x_{p_i})\right\}$$

$$= 2\int_{l_3}^{c} \phi(n_2) \left[\Phi\left(-n_2 cot(\theta_{(i-1)k}) + \tfrac{c}{sin(\theta_{(i-1)k})}\right) - \Phi\left(-n_2 cot(\theta_{(i-1)i}) + \tfrac{c}{sin(\theta_{(i-1)i})}\right)\right] \mathrm{d}n_2 \tag{4.23}$$

where

$$l_3(k) = \frac{c(sin(\theta_{(i-1)k}) - \sin(\theta_{(i-1)i}))}{sin(\theta_{(i-1)k} - \theta_{(i-1)i})}.$$

As, by definition, none of these regions overlap, the probability of $\boldsymbol{N}$ lying in all $k-2$ of these regions must the sum of the individual probabilities, therefore

$$\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k\} = \mathrm{P}\{\boldsymbol{N} \in \mathbb{V}(x_{p_1}) \cap \mathbb{V}(x_{p_k})\} - \sum_{i=2}^{k-1} \mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}(x_{p_{i-1}}) \cap \mathbb{V}(x_{p_k}) \cap \mathbb{V}^{\mathrm{c}}(x_{p_i})\right\}$$

$$= \frac{1}{\pi}\left\{\int_{\xi_2-\pi}^{\xi_1} \left(1 - \exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right) \mathrm{d}\theta + \int_{\xi_1-\theta_{1k}}^{\xi_2-\theta_{1k}} \left(1 - \exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right) \mathrm{d}\theta\right\}$$

$$-2\sum_{i=2}^{k-1} \int_{l_3(k)}^{c} \phi(n_2) \left[\Phi\left(-n_2 cot(\theta_{(i-1)k}) + \tfrac{c}{sin(\theta_{(i-1)k})}\right) - \Phi\left(-n_2 cot(\theta_{(i-1)i}) + \tfrac{c}{sin(\theta_{(i-1)i})}\right)\right] \mathrm{d}n_2. \tag{4.24}$$

Thus it is sufficient to minimise equation 4.24 with respect to the angles $0 \leq \theta_{ij} \leq \pi$. The following theorem gives the proposed minimum of 4.24.

Theorem 4.3. *The probability $P\{\boldsymbol{N} \in \mathbb{V}_k\}$ is minimised with respect to the θ_{ij} when $\mathbb{V}_k$ takes a regular $2k$ sided polygonal shape, that is when for all $1 \leq i < j \leq k$ $\theta_{ij} = \frac{(j-i)\pi}{k}$.*

This relies on the following Lemma.

Lemma 4.1. *Suppose that $k-1$ directional vectors $\boldsymbol{Px}_{p_i}$ are fixed, and only one $\boldsymbol{Px}_{p_j}$ is allowed to vary between adjacent $\boldsymbol{Px}_{p_{j-1}}$ and $\boldsymbol{Px}_{p_{j+1}}$. Then $P\{\boldsymbol{N} \in \mathbb{V}_k\}$ is minimised when $\boldsymbol{Px}_{p_j}$ lies halfway between $\boldsymbol{Px}_{p_{j-1}}$ and $\boldsymbol{Px}_{p_{j+1}}$.*

Proof of Lemma 4.1

The region $\mathbb{V}_k$ may always be represented as the parallelogram formed by the intersection of any two $\mathbb{V}(x_{p_i})$, less the grey shaded regions formed by the intersection of this parallelogram and the remaining $k-2$ $\mathbb{V}(x_{p_i})$. Hence form $\mathbb{V}_k$ as the intersection of $\mathbb{V}(x_{p_{j-1}})$ and $\mathbb{V}(x_{p_{j+1}})$, less the remaining $k-2$ regions of the form $\mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}(x_{p_{j-1}}) \cap \mathbb{V}(x_{p_{j+1}}) \cap \mathbb{V}^{\mathrm{c}}(x_{p_i})\right\}$ $i \neq j-1, j+1$. As demonstrated in Figure 4.9, if only $\boldsymbol{Px}_{p_j}$ can vary then the probability of $\boldsymbol{N}$ in $\mathbb{V}_k$ depends only on the grey shaded region $\mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}(x_{p_{j-1}}) \cap \mathbb{V}(x_{p_{j+1}}) \cap \mathbb{V}^{\mathrm{c}}(x_{p_j})\right\}$. Due to rotational invariance of $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k\}$, The proof is then immediate by applying the proof of Theorem 4.2 to $\mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}(x_{p_{j-1}}) \cap \mathbb{V}(x_{p_{j+1}}) \cap \mathbb{V}^{\mathrm{c}}(x_{p_j})\right\}$.

End of Proof of Lemma 4.1

Proof of Theorem 4.3

The proof of Theorem 4.3 is then immediate from repeated application of Lemma 4.1. Since $\mathbb{V}_k$ is formed from any k of the available $2k$ adjacent directions, the minimum arrangement must occur when they are all equally spaced from one another over a region of 2π, and the result follows immediately.

End of Proof of Theorem 4.3

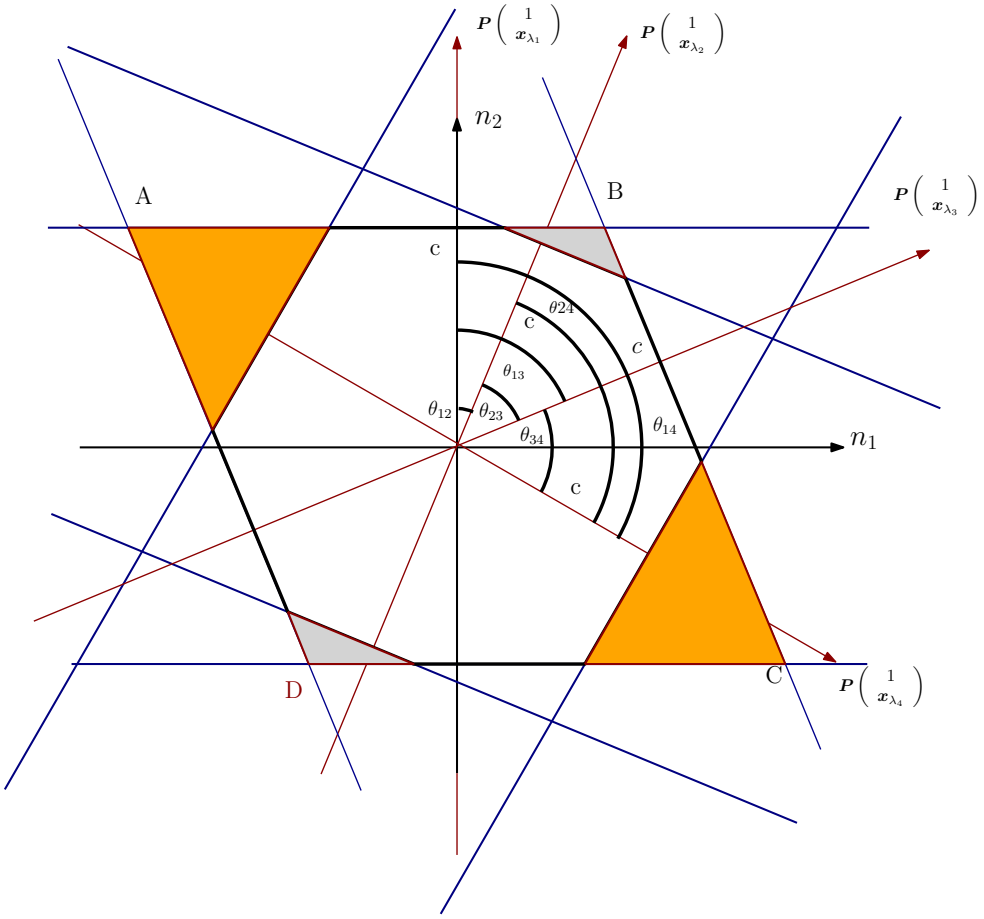


Figure 4.9: An alternative definition of the region $\mathbb{V}_4$, where the grey region is independent of the parallelogram ABCD and the orange region.

Therefore c set the following equation to $1-\alpha$

$$\frac{1}{\pi}\left\{\int_{\xi_2-\pi}^{\xi_1}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta+\int_{\xi_1-\theta_{1k}}^{\xi_2-\theta_{1k}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\}$$

$$-2\sum_{i=2}^{k-1}\int_{l_3(k)}^{c}\phi(n_2)\left[\Phi\left(-n_2cot(\theta_{(i-1)k})+\frac{c}{sin(\theta_{(i-1)k})}\right)-\Phi\left(-n_2cot(\theta_{(i-1)i})+\frac{c}{sin(\theta_{(i-1)i})}\right)\right]\mathrm{d}n_2,$$

where $\theta_{ij}=\frac{(j-i)\pi}{k}$ $\forall 1\leq i<j\leq k$. The relevant value of c can then again be generated for any k via a simple numeric search, for which a code is available for $k=3$ and 4.

4.3 One sided Simultaneous Confidence Sets For Several Effective Doses

Two sided confidence sets are primarily sought to establish a region of candidate values for the effective dose. However, there also exists a motivation to obtain a worst, or best case scenario. To obtain information on the smallest or largest plausible dose needed to elicit a specific response p. In this case, one sided sets offer less conservative information, one such example can be seen in Deutsch and Piegorsch (2012), which constructed lower one sided confidence sets to obtain the mimimum level of exposure, the bench mark dose (BMD), to induce a certain bench mark response (BMR). This was obtained in the intuitive way, by inverting the bounds of a **one sided** simultaneous confidence band. We hence adapt the methodology in this thesis to establish simultaneous one sided confidence sets for specific numbers of effective doses.

4.3.1 One Sided Simultaneous Confidence Sets For Two Effective Doses

We want to obtain simultaneous confidence sets that are a one sided equivalent of Equation (4.8). This means inverting the bounds of a one sided confidence band with some critical constant c. Hence, a lower one sided confidence set is obtained by inverting an **upper** one sided confidence band, that is

$$\mathbf{C}_p^- = \left\{ \boldsymbol{x} : \frac{\boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} - \pi(p)}{\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}}} < \mathrm{c} \right\}. \tag{4.25}$$

Using the same methodology as in the two sided case, for $k = 2$, c is chosen such that

$$\mathrm{P}\left\{ x_{p_1} \in \mathbf{C}_{p_1}^- \cap x_{p_2} \in \mathbf{C}_{p_2}^- \right\} \geq 1 - \alpha, \tag{4.26}$$

and therefore sets

$$\min_{\boldsymbol{x}_{p_1}, \boldsymbol{x}_{p_2}} \mathrm{P}\left\{ \boldsymbol{x}_{p_i} \in \mathbf{C}_{p_i}^- \text{ for } i = 1, 2 \right\} = 1 - \alpha, \tag{4.27}$$

$$\forall \boldsymbol{\beta} \in \mathcal{R}^{p+1}.$$

We focus on the case $q = 1$. It is immediate from the two sided case that

$$\mathrm{P}\left\{ \boldsymbol{x}_p \in \mathbf{C}_p^- \right\} = \mathrm{P}\{ \boldsymbol{N} \in \mathbb{V}(x_p)^- \},$$

where

$$\mathbb{V}(x_p)^- = \left\{ \boldsymbol{N} : \frac{\left\{ \boldsymbol{P} \begin{pmatrix} 1 \\ \boldsymbol{x}_p \end{pmatrix} \right\}^\intercal \boldsymbol{N}}{\left\| \boldsymbol{P} \begin{pmatrix} 1 \\ \boldsymbol{x}_p \end{pmatrix} \right\|} < \mathrm{c} \right\}$$

is the one sided equivalent of $\mathbb{V}(x_p)$, that is the open region on the N-plane, which includes the origin, and bounded by the line perpendicular to $\boldsymbol{P}\begin{pmatrix} 1 \\ \boldsymbol{x}_p \end{pmatrix}$ with distance c from the origin in the direction of $\boldsymbol{P}\begin{pmatrix} 1 \\ \boldsymbol{x}_p \end{pmatrix}$. Then it is clear that

$$\mathrm{P}\left\{\boldsymbol{x}_{p_i} \in \mathbf{C}_{p_i}^{-} \text{ for } i = 1, 2\right\} = \mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_2^{-}\}.$$

with $\mathbb{V}_2^{-} \equiv \mathbb{V}(x_{p_1})^{-} \cap \mathbb{V}(x_{p_2})^{-}$ as show in Figure 4.10.

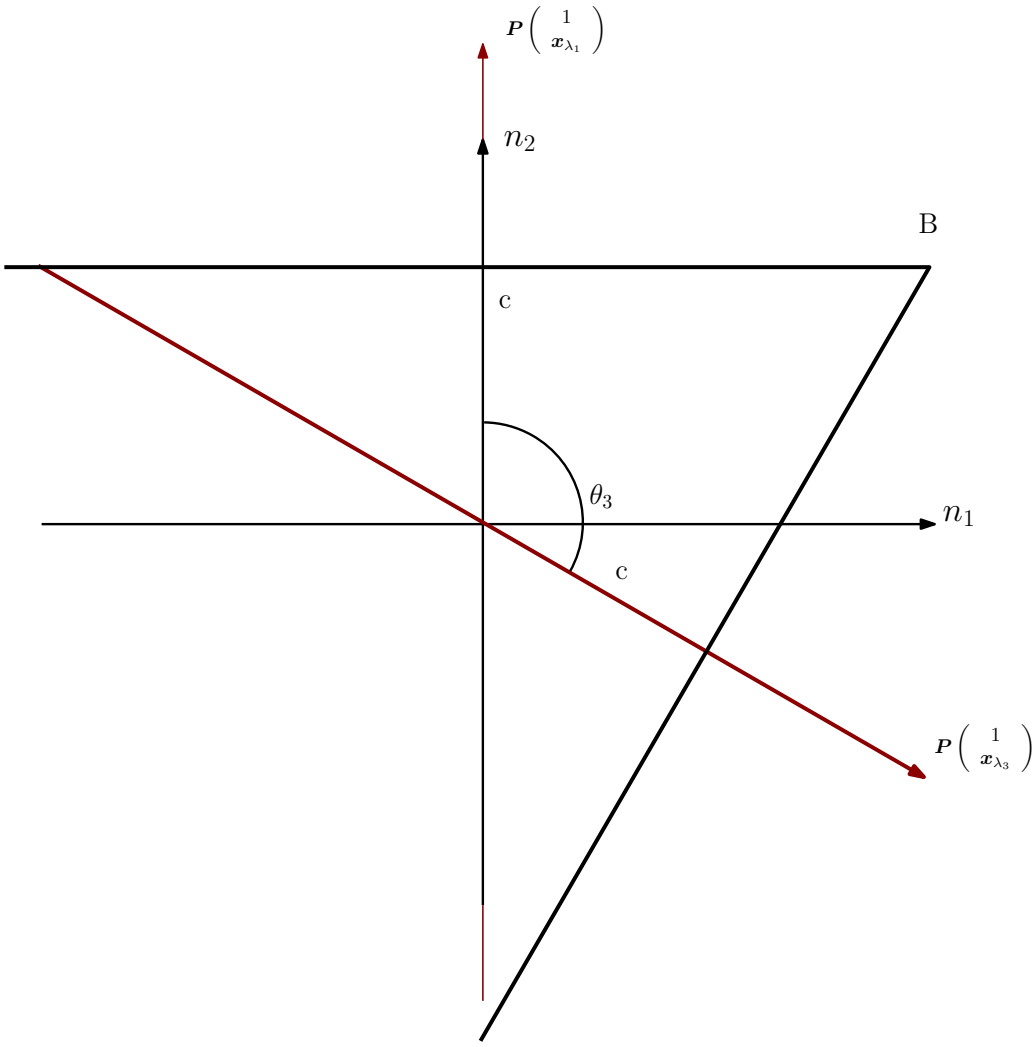


Figure 4.10: The region $\mathbb{V}_2^{-}$

Calculation of $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_2^{-}\}$ is an equivalent problem to that of establishing the probability of lying in a one sided constant width band over a finite covariate range, as seen in pages $41 - 44$ of Liu (2011) but for a bivariate $\boldsymbol{N}$, instead of $\boldsymbol{T}$. As $\mathbb{V}_2^{-}$ is rotation invariant around the origin, it may be represented as in Figure 4.11.

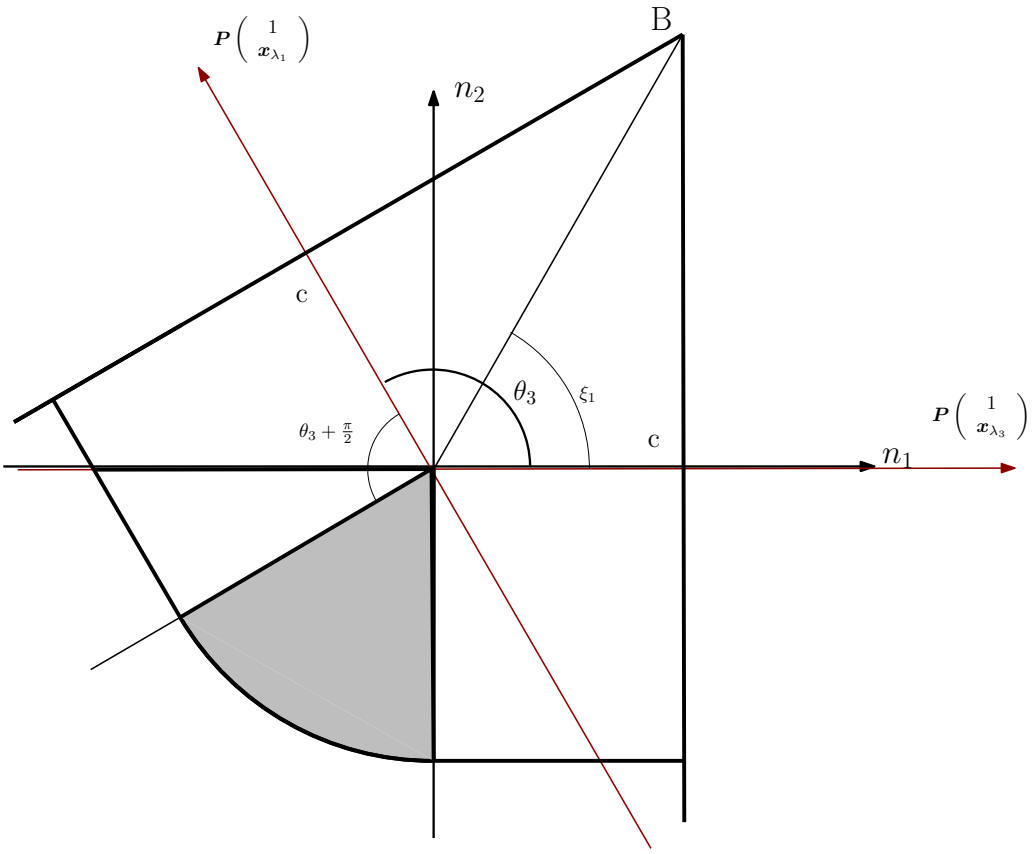


Figure 4.11: A rotation of $\mathbb{V}_2^-$

The relevant equation is then

$$\mathrm{P}\{\boldsymbol{N}\in\mathbb{V}_2^-\}=\tfrac{1}{2\pi}\left\{\int_{\xi_1-\theta_3}^{\frac{\pi}{2}}\mathrm{P}\left\{\boldsymbol{R_N}cos(\theta)\leq c\right\}\mathrm{d}\theta+\int_{-\frac{\pi}{2}}^{\xi_1}\mathrm{P}\left\{\boldsymbol{R_N}cos(\theta)\leq c\right\}\mathrm{d}\theta\right\}+\tfrac{\pi-\theta_3}{2\pi}$$

$$=\frac{1}{2\pi}\left\{\int_{-\frac{\theta_3}{2}}^{\frac{\pi}{2}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta+\int_{-\frac{\pi}{2}}^{\frac{\theta_3}{2}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\}+\frac{\pi-\theta_3}{2\pi}\tag{4.28}$$

where $\xi_1=\frac{\theta_3}{2}$. As in the two sided case, Equation (4.28) depends on the effective doses through the angle θ_3, therefore c sets the minimum of Equation (4.28) with respect to $0<\theta_3\leq\pi$ to $1-\alpha$. This minimum of Equation (4.28) is then given by the following Theorem.

Theorem 4.4. *Equation (4.28) is minimised with respect to* $0<\theta_3\leq\pi$ *at* $\theta_3=\pi$.

Proof of Theorem 4.4

We establish the differential of Equation (4.28) with respect to θ_3 as follows

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\theta_3}\left\{\frac{1}{2\pi}\left\{\int_{-\frac{\theta_3}{2}}^{\frac{\pi}{2}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta+\int_{-\frac{\pi}{2}}^{\frac{\theta_3}{2}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\}\right\}\\
&+\frac{\mathrm{d}}{\mathrm{d}\theta_3}\left\{\frac{\pi-\theta_3}{2\pi}\right\}\\
&=\frac{1}{2\pi}\left\{\int_{-\frac{\theta_3}{2}}^{\frac{\pi}{2}}\frac{\mathrm{d}}{\mathrm{d}\theta_3}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta-(-\frac{1}{2})\left(1-\exp\left\{-\frac{c^2}{2cos(-\frac{\theta_3}{2})^2}\right\}\right)\right\}\\
&+\frac{1}{2\pi}\left\{\int_{-\frac{\pi}{2}}^{\frac{\theta_3}{2}}\frac{\mathrm{d}}{\mathrm{d}\theta_3}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta+\frac{1}{2}\left(1-\exp\left\{-\frac{c^2}{2cos(\frac{\theta_3}{2})^2}\right\}\right)\right\}-\frac{1}{2\pi}\\
&=\frac{1}{2\pi}\left\{1-\exp\left\{-\frac{c^2}{2cos(\frac{\theta_3}{2})^2}\right\}-1\right\}\\
&=-\frac{1}{2\pi}\exp\left\{-\frac{c^2}{2cos(\frac{\theta_3}{2})^2}\right\} \qquad (4.29)
\end{aligned}$$

as a result of cosine being an even function and the Leibniz integral rule. It is clear from (4.29) that the derivative is negative for all θ_3, thus Equation (4.28) is a monotonically decreasing function of $0<\theta_3\leq\pi$. Furthermore, the stationary point of (4.29) satisfies,

$$-\frac{1}{2\pi}\exp\left\{-\frac{c^2}{2cos(\frac{\theta_3}{2})^2}\right\}=0\Leftrightarrow\lim_{0<\theta_3\leq\pi\to u}cos\left(\frac{u}{2}\right)=0\Leftrightarrow u=\pi.$$

Therefore $\theta_3=\pi$ is the only stationary point of (4.28) between 0 and π and this must be the minimum as required.

End of Proof Of Theorem 4.4

Hence the critical constant sets Equation (4.28) at $\theta_3=\pi$ to $1-\alpha$. Note that $\mathrm{P}\{\boldsymbol{N}\in\mathbb{V}_2^-\}$ at $\theta_3=\pi$ is a special case, equal to the probability that $\boldsymbol{N}$ lies in the region bound by the two lines parallel to the n_2 axis, with distance c from the origin in the direction of the n_1 axis. Therefore we may write

$$\min_{\boldsymbol{x}_{p_1},\boldsymbol{x}_{p_2}}\mathrm{P}\{\boldsymbol{N}\in\mathbb{V}_2^-\}=\mathrm{P}\{-c\leq n_1\leq c,-\infty\leq n_2\leq\infty\}$$

$$=\mathrm{P}\{-c\leq n_1\leq c\}\mathrm{P}\{-\infty\leq n_2\leq\infty\}=\mathrm{P}\{-c\leq n_1\leq c\}$$

by the independence of n_1 and n_2. Since by definition n_1 takes a univariate normal distribution, it is immediately obvious that the relevant c value must be a regular $z^{\frac{\alpha}{2}}$ value, the same as is used for establishing two sided confidence sets for a single effective dose.

4.3.2 One Sided Simultaneous Confidence Sets For Three Or More Effective Doses

For values of $k \geq 3$, we look to establish confidence sets of the form of Equation (4.25) such that

$$\mathrm{P}\left\{x_{p_i} \in \mathbf{C}_{p_i}^{-} \text{ for } i = 1, \ldots, k\right\} \geq 1 - \alpha, \tag{4.30}$$

for some $k \geq 3$. Once again c sets

$$\min_{\boldsymbol{x}_{p_1}, \ldots, \boldsymbol{x}_{p_k}} \mathrm{P}\left\{\boldsymbol{x}_{p_i} \in \mathbf{C}_{p_i}^{-} \text{ for } i = 1, \ldots, k\right\} = 1 - \alpha, \tag{4.31}$$

$$\forall \boldsymbol{\beta} \in \mathcal{R}^{p+1}.$$

It is immediate that we may write

$$\mathrm{P}\left\{\boldsymbol{x}_{p_i} \in \mathbf{C}_{p_i}^{-} \text{ for } i = 1, \ldots, k\right\} = \mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k^{-}\}.$$

Where $\mathbb{V}_k^{-} = \cap_{i=1}^{k} \mathbb{V}(x_{p_1})^{-}$. As before, denote the angle between any two directional vectors $\boldsymbol{P}\boldsymbol{x}_{p_i}$ and $\boldsymbol{P}\boldsymbol{x}_{p_j}$ as θ_{ij}. In the same way as in the two sided case, represent $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k^{-}\}$ as that of $\boldsymbol{N}$ lying in the intersection $\boldsymbol{N} \in \mathbb{V}(x_{p_1})^{-} \cap \mathbb{V}(x_{p_k})^{-}$, less the sum of probabilities of lying in the $k-2$ grey shaded regions $\mathbb{V}(x_{p_i-1})^{-} \cap \mathbb{V}(x_{p_k})^{-} \cap \mathbb{V}^{\mathrm{c}}(x_{p_i})^{-}$ as shown in Figure (4.12).

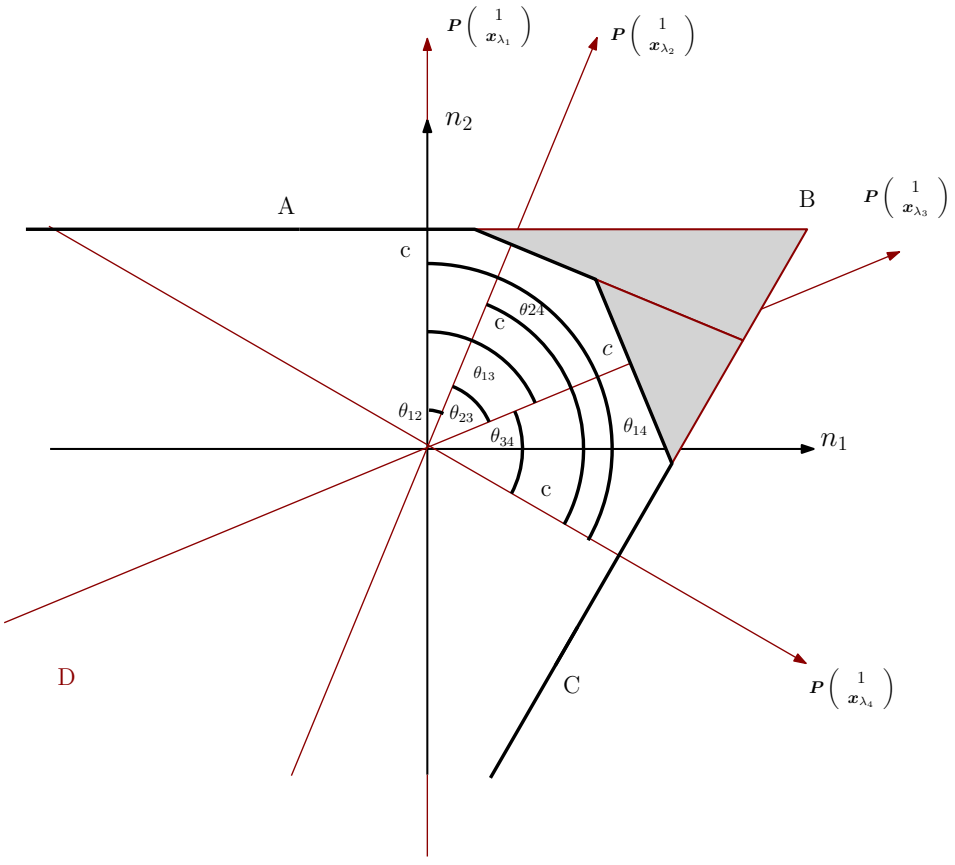


Figure 4.12: The region $\mathbb{V}_4^{-}$

Since $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}(x_{p_1})^- \cap \mathbb{V}(x_{p_k})^-\}$ is equivalent to $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_2^-\}$ with angle θ_{1k}, and the additional terms are identical to the single upper grey shaded regions given in the two sided case, we must then have

$$\begin{aligned}
&\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k^-\} \\
=&\frac{1}{2\pi}\left\{\int_{-\frac{\theta_{1k}}{2}}^{\frac{\pi}{2}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta+\int_{-\frac{\pi}{2}}^{\frac{\theta_{1k}}{2}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\} \\
+&\frac{\pi-\theta_{1k}}{2\pi}-\sum_{i=2}^{k-1}\int_{l_3(k)}^{c}\phi(n_2)\left[\Phi\left(-n_2cot(\theta_{(i-1)k})+\frac{c}{sin(\theta_{(i-1)k})}\right)\right]\mathrm{d}n_2 \\
-&\sum_{i=2}^{k-1}\int_{l_3(k)}^{c}\phi(n_2)\left[-\Phi\left(-n_2cot(\theta_{(i-1)i})+\frac{c}{sin(\theta_{(i-1)i})}\right)\right]\mathrm{d}n_2 \qquad (4.32)
\end{aligned}$$

with $l_3(k)$ as in the two sided case. We need to minimise Equation (4.32) with respect to $0 \leq \theta_{ij} \leq \pi, 1 \leq i < j \leq k$, hence we turn this into a one variable minimisation problem through the following Theorem

Theorem 4.5. *Let θ_{ik} be the largest angle as depicted in Figure (6.3). Suppose that $\boldsymbol{P}\boldsymbol{x}_{p_1}$, $\boldsymbol{P}\boldsymbol{x}_{p_k}$ and subsequently θ_{ik} are fixed. Then Equation (4.32) is minimised with respect to the remaining angles θ_{ij} when the remaining $\boldsymbol{P}\boldsymbol{x}_{p_i}$ lie equally spaced between $\boldsymbol{P}\boldsymbol{x}_{p_1}$ and $\boldsymbol{P}\boldsymbol{x}_{p_k}$, that is $\theta_{ij} = \frac{(j-i)\theta_{1k}}{k}$.*

Proof Of Theorem 4.5

We note that if θ_{1k} is fixed and say only $\boldsymbol{P}\boldsymbol{x}_{p_2}$ is allowed to change between the two adjacent $\boldsymbol{P}\boldsymbol{x}_{p_1}$ and $\boldsymbol{P}\boldsymbol{x}_{p_3}$, then $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k^-\}$ can be represented in the same way as shown in Figure 4.9 and Lemma 4.1 in the equivalent two sided case. Then, when $\boldsymbol{P}\boldsymbol{x}_{p_2}$ changes between $\boldsymbol{P}\boldsymbol{x}_{p_1}$ and $\boldsymbol{P}\boldsymbol{x}_{p_3}$, the probability of $\boldsymbol{N}$ in the parallelogram $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}(x_{p_1})^- \cap \mathbb{V}(x_{p_3})^-\}$ does not change, only the probability of lying in the specific grey shaded region, $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}(x_{p_1})^- \cap \mathbb{V}(x_{p_3})^- \cap \mathbb{V}^{\mathrm{c}}(x_{p_2})^-\}$. As this is exactly half the value of the equivalent two sided region, Lemma 4.1 applies and $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k^-\}$ is minimised when $\boldsymbol{P}\boldsymbol{x}_{p_2}$ is halfway between $\boldsymbol{P}\boldsymbol{x}_{p_1}$ $\boldsymbol{P}\boldsymbol{x}_{p_3}$. The Theorem then follows from repeated application of Lemma 4.1 on $\boldsymbol{P}\boldsymbol{x}_{p_2} \ldots \boldsymbol{P}\boldsymbol{x}_{p_{k-1}}$.

End of Proof of Theorem 4.5

We still need to minimise $\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k\}$ as given in Equation (4.32) with respect to $\theta_{1k} \in (0, \pi]$, with $\theta_{ij} = \frac{(j-i)\theta_{1k}}{k}$ for all $1 \leq i < j \leq k-1$. While we are unable to establish any analytical result, this minimisation may be easily done numerically since only one variable is involved. Hence the minimum probability in Equation (4.31) for a given c can be easily computed using Theorem 4.4, and a one variable numeric search. By using the bisection method or other search algorithm on c, combined with a numeric search for the minimum at each step, the c which sets the minimum probability in Equation (4.32) to $1-\alpha$ can be computed quickly and accurately, using the following algorithm.

1. Choose two constants, u and l for which the solution c is such $l \leq c \leq u$ and set $c_1 = \frac{(l+u)}{2}$.

2. Minimise the equation in (4.32) with respect to θ_{1k} under the constraints, $\theta_{ij} = \frac{(j-i)\theta_{1k}}{k}$ for all $1 \leq i < j \leq k-1$ and $0 < \theta_{1k} \leq \pi$ and $c = c_1$.

3. Calculate the value of the equation in (4.32), if the value is greater than $1-\alpha$, set $u = c_1$, else set $l = c_1$.

4. Return to step 1 and repeat with until the value of Equation (4.32) converges to $1-\alpha$ within tolerance level.

5. If we denote c_n as the value of c in the n'th iteration of the algorithm and c_n is the value at convergence then we set $c = c_n$ and we have found the solution.

4.3.3 A Note On Upper One Sided Confidence Sets

To construct an upper one sided confidence set, we invert the bounds of a lower one sided confidence band, that is

$$\mathbf{C}_p^+ = \left\{ \boldsymbol{x} : \frac{-\left(\boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} - \pi(p)\right)}{\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}}} < \mathrm{c} \right\}, \tag{4.33}$$

where the relevant c sets

$$\min_{\boldsymbol{x}_{p_1},\ldots,\boldsymbol{x}_{p_k}} \mathrm{P}\left\{\boldsymbol{x}_{p_i} \in \mathbf{C}_{p_i}^+ \text{ for } i = 1,\ldots,k\right\} = 1-\alpha, \tag{4.34}$$

$$\forall \boldsymbol{\beta} \in \mathcal{R}^{p+1}.$$

We immediately have

$$\mathrm{P}\left\{\boldsymbol{x}_{p_i} \in \mathbf{C}_{p_i}^+ \text{ for } i = 1,\ldots,k\right\} = \mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k^+\}$$

where $\mathbb{V}_k^+ = \cap_{i=1}^k \mathbb{V}(x_{p_1})^+$ and

$$\mathbb{V}(x_p)^+ = \left\{ \boldsymbol{N} : \frac{\left\{-\boldsymbol{P}\begin{pmatrix} 1 \\ \boldsymbol{x}_p \end{pmatrix}\right\}^\intercal \boldsymbol{N}}{\left\|\boldsymbol{P}\begin{pmatrix} 1 \\ \boldsymbol{x}_p \end{pmatrix}\right\|} < \mathrm{c} \right\}.$$

It is then clear to see that

$$\mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k^+\} = \mathrm{P}\{\boldsymbol{N} \in \mathbb{V}_k^-\}$$

by rotational invariance. Constructing upper one sided simultaneous confidence sets is therefore as simple as using the same c as the corresponding lower one sided problem.

4.4 Table Of Critical Constants and Examples

4.4.1 Values Of c

In this section, we compare the simultaneous confidence sets obtained by the original Scheffé band method (S), to the two sided confidence sets (AS2) and one sided confidence sets (AS1) of the improved methodology. We have computed the critical constants c for AS2, AS1 and S sets for $k = 2, 3$ and 4 in Table 4.1 below.

Table 4.1: Values of c for AS2, AS1 and S simultaneous confidence sets

		AS2	AS1	S for $q = 1$	S for $q = 2$
k	$1-\alpha$	c		$\sqrt{\chi^{\alpha}_2}$	$\sqrt{\chi^{\alpha}_3}$
2	0.99	2.806225	2.575829	3.034854	3.368214
	0.95	2.236477	1.960000	2.447747	2.795483
	0.90	1.948822	1.644854	2.145966	2.500278
3	0.99	2.913494	2.712313	3.034854	3.368214
	0.95	2.343701	2.123498	2.447747	2.795483
	0.90	2.052293	1.823565	2.145966	2.500278
4	0.99	2.962385	2.787521	3.034854	3.368214
	0.95	2.387280	2.195720	2.447747	2.795483
	0.90	2.092173	1.89069	2.145966	2.500278

The relative size of each confidence set may be directly compared by the size of c. In each case a smaller c indicates a smaller set or bound. Thus the relative improvement over S type confidence sets at level $1-\alpha$ is

$$\left| \frac{\sqrt{\chi^{\alpha}_{q+1}} - c}{\sqrt{\chi^{\alpha}_{q+1}}} \right| * 100.$$

For S type sets, $c = 2.447747$ for $q = 1$ and 2.795483 for $q = 2$. There is a clear reduction therefore in the size of c, and by extension the set for all k. For $\alpha = 0.05$ and $q = 1$, AS2 sets show a relative improvement of 8.6%, 4.25% and 2.5% for $k = 2, 3$ and 4 respectively. For AS1 sets, the equivalent improvements are 19.9%, 13.2% and 10.3%. AS2 sets for $k = 2$ can be used for multiple covariate models. For $q = 2$, there is a significant jump in improvement, to approximately 23%.

It is clear that the improved methodology is at its most useful with AS2 type sets for two effective doses, where the percentage improvement will only become more significant with increased q. All other methods are limited in that they may only apply to a univariate model, where S type sets are at their smallest. This is most noticeable for AS2 sets, where the percentage improvement beyond $k = 3$ becomes marginal, though less so for AS1 sets where the improvement remains beyond 10% at $k = 4$. Establishing methodology for more than two effective doses that applies to a multiple covariate model is therefore of significant interest.

4.4.2 Examples

Example 1

In Walter (1983), the Ontario Exercise Heart Collaborative Study recorded data from 341 cases, and measured the chance of a Mycardial Infarction (MI) over a four year period. Logistic analysis is performed on the odds of MI with respect to smoking status (x_1) and serum triglyceride level (x_2), the model gave results,

$$\hat{\boldsymbol{\beta}} = (-2.2791, 0.7682, 0.001952)^\intercal \quad \boldsymbol{J}^{-1} = \begin{pmatrix} 0.06511 & & \\ -0.04828 & 0.09839 & \\ -0.0001915 & -0.00003572 & 0.000002586 \end{pmatrix}.$$

We construct simultaneous confidence sets for two CED's of Triglyceride level for Non smokers ($x_1 = 0$), at $\alpha = 0.05$ corresponding to $p = 0.4, 0.5,$ and 0.6.. We apply the the S and AS2 methods for $k = 2$. For a confidence set for the ED_p or CED_p, with critical constant c, the lower bound is the value of x which solves,

$$p - \left(\frac{\exp\left(\boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} + c\sqrt{\boldsymbol{x}^\intercal\boldsymbol{J}^{-1}\boldsymbol{x}}\right)}{1 + \exp\left(\boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} + c\sqrt{\boldsymbol{x}^\intercal\boldsymbol{J}^{-1}\boldsymbol{x}}\right)} \right) = 0,$$

with the upper bound given as,

$$p - \left(\frac{\exp\left(\boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} - c\sqrt{\boldsymbol{x}^\intercal\boldsymbol{J}^{-1}\boldsymbol{x}}\right)}{1 + \exp\left(\boldsymbol{x}^\intercal\hat{\boldsymbol{\beta}} - c\sqrt{\boldsymbol{x}^\intercal\boldsymbol{J}^{-1}\boldsymbol{x}}\right)} \right) = 0.$$

In this case we set $x = (x_1, x_2) = (1, x_2)$ in order to establish the appropriate CED. We calculate the implicit 95% confidence bands and confidence sets for $p = 0.4, 0.5$ and 0.6 in Figure 4.12 and Table 4.2 below.

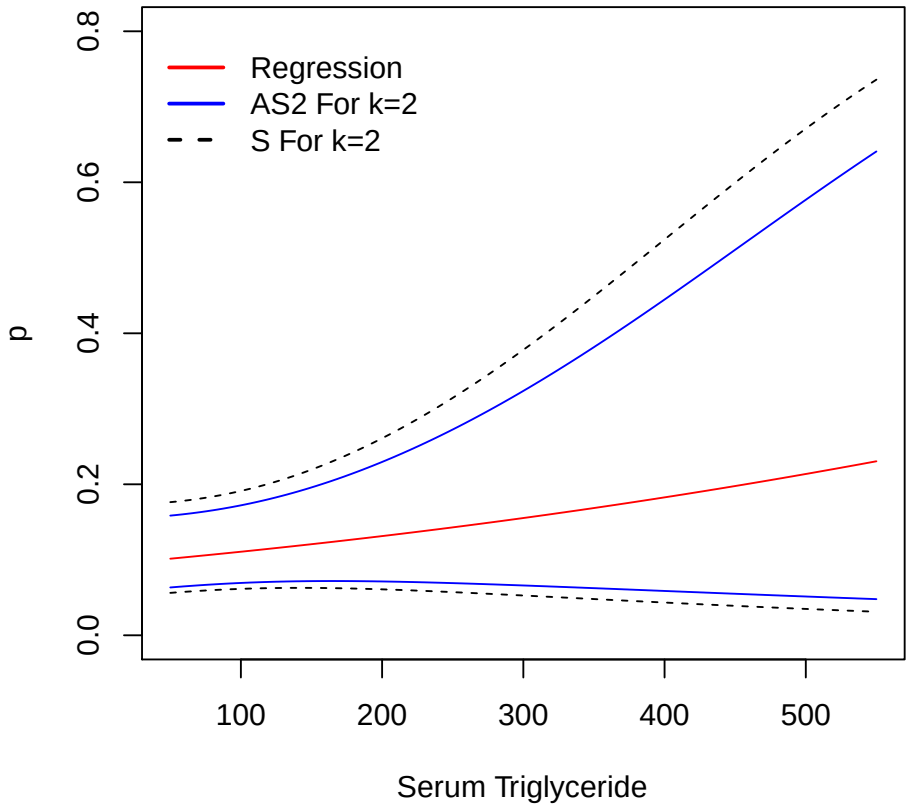


Figure 4.13: The implicit 95% confidence bands for the Scheffé band method and two sided adapted Scheffé band method for k=2, for the Heart Exercise Data

Table 4.2: $\boldsymbol{C}_p$ at $\alpha = 0.05$ for the S type, and AS2 type sets for k=2

	AS2 For $k = 2$		S for $q = 2$	
p	Lower	*Upper*	*Lower*	*Upper*
0.4	364.9	∞	315.9	∞
0.5	442.0	∞	384.0	∞
0.6	517.8	∞	450.4	∞

It is clear that the AS2 method demonstrates a noticeable improvement over the original S method. Figure 4.13 shows the confidence band that generates $\boldsymbol{C}_p$, for AS2 sets, clearly has a significantly smaller average width over the whole range than the simultaneous confidence band of the S method. This results in smaller confidence sets, as demonstrated by Table 4.2, in which the sets constructed from the AS2 method have an improved lower bound of around 50-70 units. There is no upper bound as a result of the shape of the logistic curve, in fact we note that $\boldsymbol{C}_p$ is not guaranteed to be a closed or even a single interval.

Example 2

The first example illustrates the improvement in the size of simultaneous confidence sets in a real world dataset. Here, we use the example dataset "mtcars" to compare simultaneous confidence sets at different values of k. This model has parameter estimates

$$\hat{\boldsymbol{\beta}} = (-2.2791, 0.7682, 0.001952)^{\intercal} \quad \boldsymbol{J}^{-1} = \begin{pmatrix} 5.5292778 & -0.26495522 \\ -0.2649552 & 0.01318859 \end{pmatrix}.$$

We construct and compare S, AS2 and AS1 confidence sets for $k = 2$ and 3 for $\alpha = 0.05$ below.

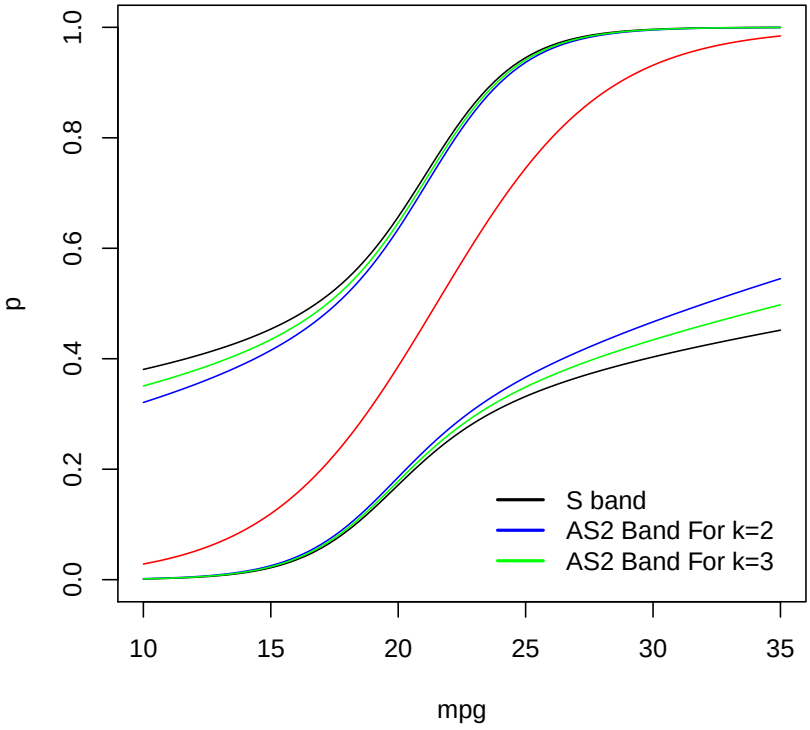


Figure 4.14: The implicit 95% confidence bands for the Scheffé band method and two sided adapted Scheffé band method for k=2 and 3, for the "mtcars" dataset.

Table 4.3: $\boldsymbol{C}_p$ at $\alpha = 0.05$ for the S type, and AS2 type sets for k=2 and 3

	AS2 For $k = 2$		AS2 For $k = 3$		S for $q = 1$	
p	Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
0.4	14.39	26.46	13.29	27.73	11.67	29.71
0.2	$-\infty$	20.31	$-\infty$	20.48	$-\infty$	20.65
0.1	$-\infty$	18.07	$-\infty$	18.2	$-\infty$	18.32

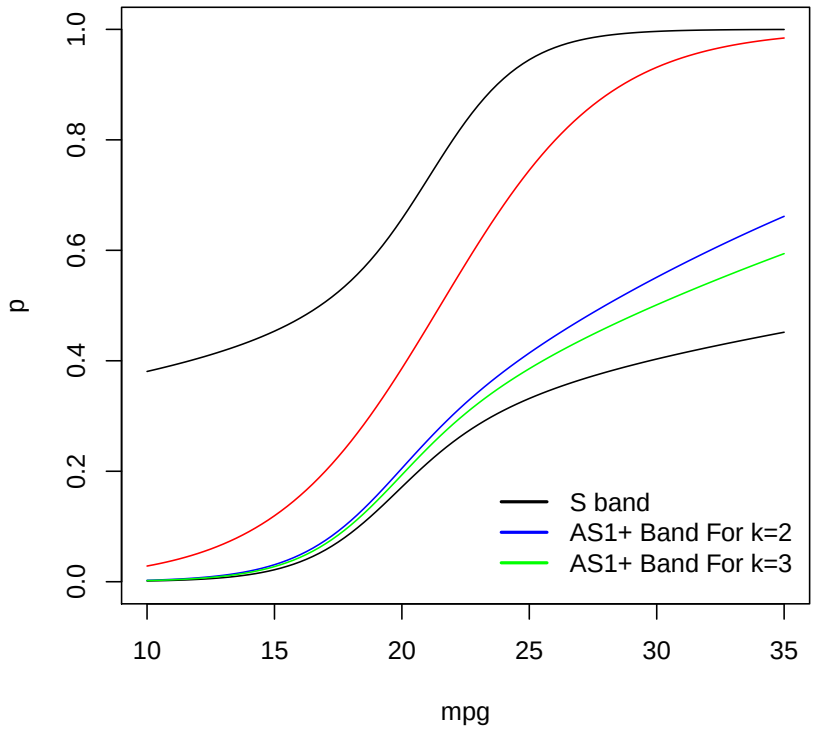


Figure 4.15: The implicit 95% confidence bands for the Scheffé band method and upper one sided adapted Scheffé band method for k=2 and 3, for the "mtcars" dataset.

Table 4.4: $\boldsymbol{C}_p$ at $\alpha = 0.05$ for the S type, and AS1 type sets for k=2 and 3

	AS1 For $k = 2$		AS1+ For $k = 3$		S for $q = 2$	
p	Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
0.4	N/A	24.57	N/A	25.53	11.67	29.71
0.2	N/A	19.90	N/A	20.14	N/A	20.65
0.1	N/A	17.73	N/A	17.93	N/A	18.32

From Tables 4.3 and 4.4 we note a noticeable reduction in the size of two sided sets at $p = 0.4$, from approximately 18 units, to around 12 and 14.5 for $k = 2$ and 3, with the upper bound of one sided sets improved by 5 and 4 units. The extent of the improvement is much less for $p = 0.2$ and 0.1, where the implicit bands are much narrower. Indeed, in the case that the variance covariance matrix leads to particularly narrow confidence bands, the extent of "real" improvement in the size of each may not be substantial. However, as we need only change c to adapt methods of constructing S type sets to constructing AS1 and AS2 sets, there is no reason to use S sets over the improved methodology when confidence sets for a specific number of ED's are in mind.

Chapter 5

Simultaneous Confidence Sets For Several Effective Doses When Constrained Over Some Finite Region

The results of the previous Chapter make the assumption that the k effective doses could take any value over the whole real line. However, particularly in clinical trials, it can sometimes be assumed that the effective doses lie over some particular range of interest. At the very least, most clinical trials may consider some general sensible range for the ED, for example non negativity, or an upper bound on the sensible maximum dosage. Hence one may consider the methods established in the previous chapter to still be overly conservative. In this chapter, we fully define simultaneous confidence sets for k effective doses at $q = 1$, using the methodology of Chapter 4, under the additional constraint that the effective doses of the model x_{p_i} $i = 1, \ldots, k$ all lie within some finite range $[a, b]$. In particular, the Chapter will address how the additional constraint affects the current work established in the previous chapter, if at all, and seek to modify existing work in order to account for this.

5.1 Introduction

In the same way as in Chapter 4, we look to establish confidence regions of the form,

$$\mathbf{C}_p = \left\{ \boldsymbol{x} : \frac{|\boldsymbol{x}^\intercal \hat{\boldsymbol{\beta}} - \pi(p)|}{\sqrt{\boldsymbol{x}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x}}} < \mathrm{c} \right\} \tag{5.1}$$

where now we want the simultaneous confidence level to satisfy

$$\mathrm{P}\left\{ x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, \ldots, k | \boldsymbol{x}_{p_i} \in [a, b] \; \forall \; i = 1, \ldots, k \right\} \geq 1 - \alpha, \tag{5.2}$$

for finite constants $a < b$. Hence we want to find the c in 5.2 such that

$$\min_{x_{p_i}, \ldots, x_{p_k}} \mathrm{P}\left\{ x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, \ldots, k | \boldsymbol{x}_{p_i} \in [a, b] \text{ for } i = 1, \ldots, k \right\} = 1 - \alpha. \tag{5.3}$$

It is immediate from Chapter 4 that we still have

$$\mathrm{P}\left\{ x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, \ldots, k \right\} = \mathrm{P}\left\{ \boldsymbol{N} \in \mathbb{V}_k \right\}.$$

Recall that $\mathrm{P}\left\{ \boldsymbol{N} \in \mathbb{V}_k \right\}$ depends on the choice of the x_{p_i} through the directions $\boldsymbol{P}\boldsymbol{x}_{p_i}$ and subsequently the angles θ_{ij}. Hence if all x_{p_i}'s lie within $[a, b]$, subsequently every possible directional vector, $\boldsymbol{P}\boldsymbol{x}_{p_i}$ must also lie between $\boldsymbol{P}\begin{pmatrix} 1 \\ a \end{pmatrix}$ and $\boldsymbol{P}\begin{pmatrix} 1 \\ b \end{pmatrix}$, that is when the effective doses are chosen as the endpoints of the interval. Therefore we denote the angle between $\boldsymbol{P}\begin{pmatrix} 1 \\ a \end{pmatrix}$ and $\boldsymbol{P}\begin{pmatrix} 1 \\ b \end{pmatrix}$ as

$$\theta_{ab} = acos\left(\frac{(1, a)\boldsymbol{J}^{-1}\begin{pmatrix} 1 \\ b \end{pmatrix}}{\sqrt{v(1, a)v(1, b)}} \right).$$

Under this restriction, we must have that $\theta_{ij} \leq \theta_{ab} \leq \pi \;\; \forall \, 1 \leq i < j \leq k$, and therefore

$$\begin{aligned} &\mathrm{P}\left\{ x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1, \ldots, k | \boldsymbol{x}_{p_i} \in [a, b] \text{ for } i = 1, \ldots, k \right\} \\ =&\mathrm{P}\left\{ \boldsymbol{N} \in \mathbb{V}_k | 0 < \theta_{ij} \leq \theta_{ab} \leq \pi \text{ for } \forall \; 1 \leq i < j \leq k \right\}. \end{aligned}$$

To establish the critical constant, we then solve

$$\min_{0 < \theta_{ij} \leq \theta_{ab} \leq \pi} \mathrm{P}\left\{ \boldsymbol{N} \in \mathbb{V}_k \right\} = 1 - \alpha. \tag{5.4}$$

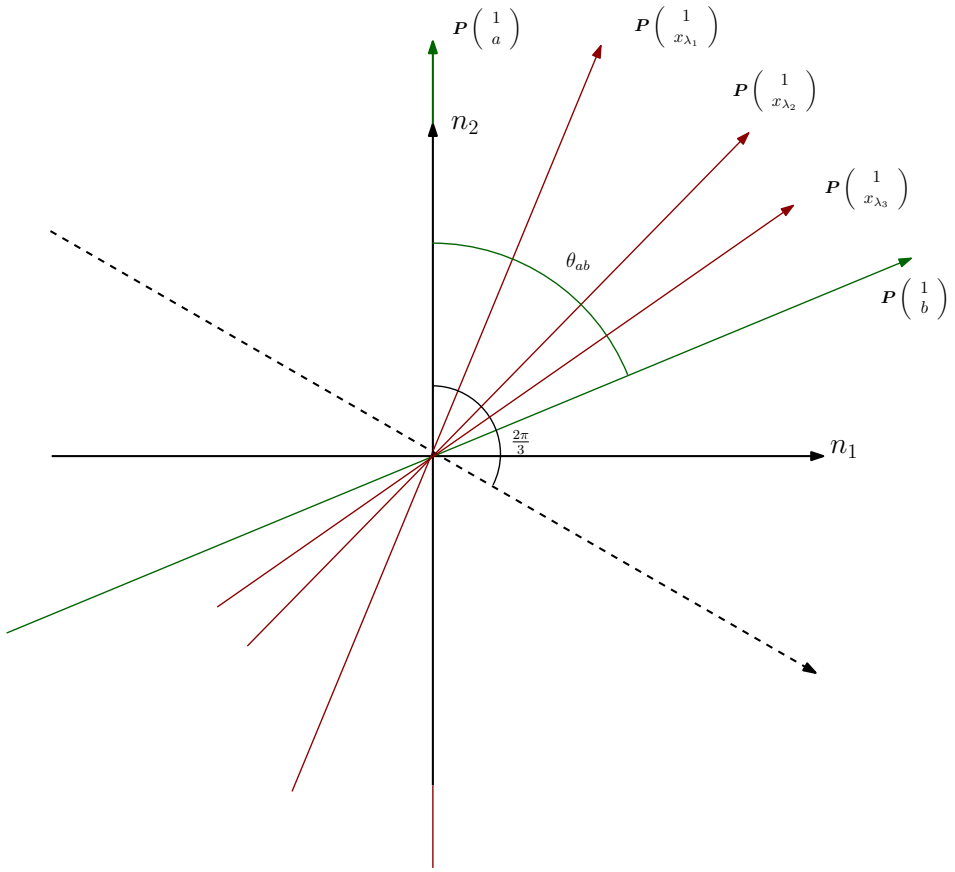


Figure 5.1: An illustration of the effect of the constraint on the angles θ_{ij} for three effective doses.

If we recall that in Chapter four, with unconstrained effective doses, the optimal arrangement of $\mathbb{V}_k$ took the largest angle as $\theta_{1k} = \frac{(k-1)\pi}{k}$. We therefore have two distinct cases under the new constraint.

1. In the case that $\theta_{ab} \geq \frac{(k-1)\pi}{k}$, the minimum arrangement established in Chapter four is still achievable and the values of c can hence be taken from the appropriate case in Chapter four. Intuitively this suggests that the finite region of interest is not a tight enough bound to offer significant additional information.

2. In the case that $\theta_{ab} < \frac{(k-1)\pi}{k}$ this global minimum cannot be reached. A new minimum arrangement under the constraint needs to be developed which will generate a smaller c, and a more tightly bounded set.

It is the latter situation that is the focus of this Chapter, and we restrict work to the case $q = 1$.

5.2 Two Sided Simultaneous Confidence Sets For Two Effective Doses

We need to establish the c such that

$$\min_{\boldsymbol{x}_{p_1},\boldsymbol{x}_{p_2}} \mathrm{P}\left\{x_{p_i} \in \mathbf{C}_{p_i} \text{ for } i = 1,2 |\ \boldsymbol{x}_{p_1}, \boldsymbol{x}_{p_2} \in [a,b]\right\} = 1-\alpha. \tag{5.5}$$

The method of Sidaks inequality from Chapter four does not offer information on the minimum of Equation (5.5) when the lower bound cannot be reached. Hence in the univariate case, we may represent Equation 5.5 as

$$\min_{\theta_1} \mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}_2 | 0 < \theta_1 \leq \theta_{ab}\right\} = 1-\alpha \tag{5.6}$$

where $\mathbb{V}_2$ is the parallelogram region formed by $\mathbb{V}(x_{p_1}) \cap \mathbb{V}(x_{p_2})$, and θ_1 is the angle between the two directional vectors $\boldsymbol{P}\begin{pmatrix}1\\ \boldsymbol{x}_{p_1}\end{pmatrix}$ and $\boldsymbol{P}\begin{pmatrix}1\\ \boldsymbol{x}_{p_2}\end{pmatrix}$. This probability is immediately given by Equation () with angle θ_1. Also noting that, given the $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ from the case $k=2$ in Chapter four, and the calculation of the angles θ_{ij} given in Chapter 2

$$Cov(\boldsymbol{Z}_1, \boldsymbol{Z}_2) = \frac{\boldsymbol{x_{pi_1}}^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{pi_2}}}{\sqrt{(\boldsymbol{x_{pi_1}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{pi_1}}}\sqrt{(\boldsymbol{x_{pi_2}})^\intercal \boldsymbol{J}^{-1} \boldsymbol{x_{pi_2}}}} = \cos(\theta_1)$$

hence

$$Cov(\boldsymbol{Z}_1, \boldsymbol{Z}_2) = 0 \ \leftrightarrow\ \theta_1 = \frac{\pi}{2}.$$

This immediately implies that Sidaks lower bound may be reached when $\theta_{ab} \geq \frac{\pi}{2}$ and we may reduce the minimisation problem to

$$\min_{0<\theta_1\leq\theta_{ab}<\frac{\pi}{2}}$$

$$\frac{1}{\pi}\left\{\int_{\frac{\theta_1-\pi}{2}}^{\frac{\theta_1}{2}} \left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta + \int_{-\frac{\theta_1}{2}}^{\frac{\pi-\theta_1}{2}} \left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\}. \tag{5.7}$$

We may evaluate the differential of the term inside Equation (5.7) as

$$\frac{1}{\pi}\left\{\frac{\mathrm{d}}{\mathrm{d}\theta_1}\int_{\frac{\theta_1-\pi}{2}}^{\frac{\theta_1}{2}} \left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta + \frac{\mathrm{d}}{\mathrm{d}\theta_1}\int_{-\frac{\theta_1}{2}}^{\frac{\pi-\theta_1}{2}} \left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\}.$$

Let $h(\theta) = 1 - \exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}$, we have

$$
\begin{aligned}
=&\frac{1}{\pi}\left\{h\left(\frac{\theta_1}{2}\right)\frac{\mathrm{d}}{\mathrm{d}\theta_1}\left(\frac{\theta_1}{2}\right) - h\left(\frac{\theta_1-\pi}{2}\right)\frac{\mathrm{d}}{\mathrm{d}\theta_1}\left(\frac{\theta_1-\pi}{2}\right)\right\}\\
+&\frac{1}{\pi}\left\{h\left(\frac{\pi-\theta_1}{2}\right)\frac{\mathrm{d}}{\mathrm{d}\theta_1}\left(\frac{\pi-\theta_1}{2}\right) - h\left(-\frac{\theta_1}{2}\right)\frac{\mathrm{d}}{\mathrm{d}\theta_1}\left(-\frac{\theta_1}{2}\right)\right\}\\
=&\frac{1}{2\pi}\left\{h\left(\frac{\theta_1}{2}\right) - h\left(\frac{\theta_1-\pi}{2}\right) - h\left(\frac{\pi-\theta_1}{2}\right) + h\left(\frac{-\theta_1}{2}\right)\right\}\\
=&\frac{1}{\pi}\left\{h\left(\frac{\theta_1}{2}\right) - h\left(\frac{\theta_1-\pi}{2}\right)\right\}\\
=&\frac{1}{\pi}\left\{\exp\left\{-\frac{c^2}{2sin(\frac{\theta_1}{2})^2}\right\} - \exp\left\{-\frac{c^2}{2cos(\frac{\theta_1}{2})^2}\right\}\right\}\\
=&\frac{1}{\pi}\left\{h\left(\frac{\theta_1}{2}\right) - h\left(\frac{\theta_1-\pi}{2}\right)\right\}\\
=&\frac{1}{\pi}\left\{\exp\left\{-\frac{c^2}{2sin(\frac{\theta_1}{2})^2}\right\} - \exp\left\{-\frac{c^2}{2cos(\frac{\theta_1}{2})^2}\right\}\right\}
\end{aligned}
$$

by applying the Leibniz integral rule to the differential, and noting that $cos(-x) = cos(x) \Rightarrow h(-x) = h(x)$ and $cos(x - \frac{\pi}{2}) = sin(x)$. Now since $0 \leq sin(x) < cos(x)\ \forall x \in [0, \frac{\pi}{4})$, we have $0 \leq sin(\frac{\theta_1}{2}) < cos(\frac{\theta_1}{2})\ \forall \theta_1 \in [0, \frac{\pi}{2})$. This immediately implies that $-\frac{c^2}{2sin(\frac{\theta_1}{2})^2} < -\frac{c^2}{2cos(\frac{\theta_1}{2})^2}\ \forall \theta_1 \in [0, \frac{\pi}{2})$ and hence

$$\frac{1}{\pi}\left\{\exp\left\{-\frac{c^2}{2sin(\frac{\theta_1}{2})^2}\right\} - \exp\left\{-\frac{c^2}{2cos(\frac{\theta_1}{2})^2}\right\}\right\} < 0 \quad \forall \theta_1 \in [0, \frac{\pi}{2}).$$

This tells us that the expression in Equation (5.7) is a monotonically decreasing function of θ_1 over $[0, \frac{\pi}{2})$, and therefore the minimum must occur at $\theta_1 = \theta_{ab}$. Hence, to establish the value of c we calculate θ_{ab}, if this is a least $\frac{\pi}{2}$ we use the method in Chapter 4, otherwise apply a search algorithm to find the c that sets the expression in Equation (5.7) to $1 - \alpha$ at $\theta_1 = \theta_{ab}$.

5.3 Two Sided Simultaneous Confidence Sets For Three Or More Effective Doses

We now look to establish confidence sets of the form of Equation (5.1) for specific values of $k \geq 3$, in the case where the constraint has an effect on the value of c, that is when $\theta_{ab} < \frac{(k-1)\pi}{k}$. It is immediate from Section 5.1 and the expressions in Section 4.2.3, in particular Equation (4.24), that c sets

$$\min_{\theta_{ij}} \mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}_k | 0 < \theta_{ij} \leq \theta_{ab} < \frac{(k-1)\pi}{k} \text{ for } 1 \leq i < j \leq k\right\} = 1 - \alpha.$$

$$\min_{0<\theta_{ij}\leq\theta ab<\frac{(k-1)\pi}{k}}$$

$$\frac{1}{\pi}\left\{\int_{\xi_2-\pi}^{\xi_1}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta+\int_{\xi_1-\theta_{1k}}^{\xi_2-\theta_{1k}}\left(1-\exp\left\{-\frac{c^2}{2cos(\theta)^2}\right\}\right)\mathrm{d}\theta\right\}$$
$$-2\sum_{i=2}^{k-1}\int_{l_3(k)}^{c}\phi(n_2)\left[\Phi\left(-n_2cot(\theta_{(i-1)k})+\frac{c}{sin(\theta_{(i-1)k})}\right)\right]\mathrm{d}n_2$$
$$-2\sum_{i=2}^{k-1}\int_{l_3(k)}^{c}\phi(n_2)\left[-\Phi\left(-n_2cot(\theta_{(i-1)i})+\frac{c}{sin(\theta_{(i-1)i})}\right)\right]\mathrm{d}n_2. \tag{5.8}$$

We may reduce this to a one dimensional minimisation problem in terms of the largest angle θ_{1k} using the following theorem.

Theorem 5.1. *Let θ_{ik} be the largest angle between all the $\boldsymbol{Px}_{p_i}$'s which is spanned by $\boldsymbol{Px}_{p_1}$ and $\boldsymbol{Px}_{p_k}$. Furthermore, allow these two vectors, and therefore θ_{ik}, to be fixed. Then the expression of $P\{\boldsymbol{N}\in\mathbb{V}_k\}$ in (5.8) is minimised , when $\boldsymbol{Px}_{p_2},\ldots,\boldsymbol{Px}_{p_{k-1}}$ are allowed to vary freely between $\boldsymbol{Px}_{p_1}$ and $\boldsymbol{Px}_{p_k}$ at the configuration that the angle between $\boldsymbol{Px}_{p_j}$ and $\boldsymbol{Px}_{p_{j+1}}$ is equal to $\frac{\theta_{1k}}{k}$ for all $j=1,\ldots,k-1$.*

We note that Theorem 4.3 still applies, and hence the proof is immediate from repeated application of Theorem 4.3 on the vectors $\boldsymbol{Px}_{p_2},\ldots,\boldsymbol{Px}_{p_{k-1}}$. We need then only find the minimum probability in (5.8), with respect to $\theta_{1k}\in(0,\theta_{ab}]$ with $\theta_{ij}=\frac{(j-i)\theta_{1k}}{k}$ for all $1\leq i<j\leq k-1$ which may be done using the following algorithm.

1. Choose two constants, u and l for which the solution c is such $l\leq c\leq u$ and set $c_1=\frac{(l+u)}{2}$.

2. Minimise the equation in 5.8 with respect to θ_{1k} under the constraints $\theta_{ij}=\frac{(j-i)\theta_{1k}}{k}$ for all $1\leq i<j\leq k-1$ and $0<\theta_{1k}\leq\theta_{ab}<\frac{(k-1)\pi}{k}$ and $c=c_1$.

3. Calculate the value of the Equation in (5.8), if the value is greater than $1-\alpha$, set $u=c_1$, else set $l=c_1$.

4. Return to step 1 and repeat with until the value of Equation (5.8) converges to $1-\alpha$ within tolerance level.

5. If we denote c_n as the value of c in the i'th iteration of the algorithm and c_n is the value at convergence then we set $c=c_n$ and we have found the solution.

We can now fully specify confidence sets of the form of Equation 5.1 for any $k\geq 3$ using the following method.

1. Calculate $\boldsymbol{J}^{-1}$ and choose a finite region $[a,b]$.

2. Calculate θ_{ab} using Section 5.1. If $\theta_{ab}\geq\frac{(k-1)\pi}{k}$ then calculate c using the methods described in Chapter 4 for the appropriate k, otherwise, use the methods as described in this section.

5.4 One Sided Simultaneous Confidence Sets For Several Constrained Effective

We now look to establish the equivalent methodology in the one sided case, that is we construct confidence sets of the form of Equation (4.25) such that c sets

$$\min_{x_{p_i},\ldots,x_{p_k}} \mathrm{P}\left\{x_{p_i} \in \mathbf{C}^-_{p_i} \text{ for } i = 1,\ldots,k | \boldsymbol{x}_{p_i} \in [a,b] \text{ for } i = 1,\ldots,k\right\} = 1-\alpha. \tag{5.9}$$

It is immediate from the theory of Sections 5.1 and 4.3 that

$$\begin{aligned}&\mathrm{P}\left\{x_{p_i} \in \mathbf{C}^-_{p_i} \text{ for } i = 1,\ldots,k | \boldsymbol{x}_{p_i} \in [a,b] \text{ for } i = 1,\ldots,k\right\}\\ =&\mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}^-_k | 0 < \theta_{ij} \leq \theta_{ab} \leq \pi \text{ for } \forall\ 1 \leq i < j \leq k\right\}\end{aligned}$$

where θ_{ab} is as described previously. Once again, it is therefore sufficient for c to solve

$$\min_{0<\theta_{ij}\leq\theta_{ab}\leq\pi} \mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}^-_k\right\} = 1-\alpha$$

with $\mathrm{P}\left\{\boldsymbol{N} \in \mathbb{V}^-_k\right\}$ given by Equation (4.28) for $k = 2$, and Equation (4.32) for $k \geq 3$. For k=2, the value of the largest angle, θ_3 that minimises Equation (4.28) is at its largest possible value, π, hence in theory there will always be a reduction in the value of c for any θ_{ab}. For $k = 3$ there is no reasonable way to determine at what value of θ_{ab} the minimum arrangement of Equation (4.32) changes and c decreases.

However, noting that for $k \geq 3$, Theorem 4.5 still applies, minimising both of these equations may be viewed as a one dimensional search in terms of the largest angles, that is θ_3 for $k = 2$, and θ_{1k} for $k \geq 3$. Hence c may be found using a small modification to the algorithm used in Section 4.3.2. For $k = 2$, we apply steps 2 and 3 to Equation (4.28), with the constraint in step 2 given as $0 < \theta_3 \leq \theta_{ab}$. For $k \geq 3$, steps 2 and 3 apply to Equation (4.32), with the constraints in step 2 as $\theta_{ij} = \frac{(j-i)\theta_{1k}}{k}$ for all $1 \leq i < j \leq k-1$ and $0 < \theta_{1k} \leq \theta_{ab}$.

5.5 Values of c and Discussion

We apply thr algorithms of the previous section to obtain some example values of c for 3, 4 and 5 effective doses, with specific values of θ_{ab} for one and two sided simultaneous confidence sets. These are shown in the following two tables.

Table 5.1: Example Values of c For Two Sided Simultaneous Confidence Sets At Specific Values of θ_{ab}

		c		
k	$1-\alpha$	π	$\frac{\pi}{2}$	$\frac{\pi}{3}$
2	0.99	2.806255	2.806255	2.794273
	0.95	2.236477	2.236477	2.212128
	0.90	1.948822	1.948822	1.916271
3	0.99	2.913494	2.892699	2.836824
	0.95	2.343701	2.317184	2.245666
	0.90	2.052293	2.023910	1.944533
4	0.99	2.962385	2.916309	2.846085
	0.95	2.387280	2.335878	2.252412
	0.90	2.092173	2.039958	1.950018

Table 5.2: Example Values of c For One Sided Simultaneous Confidence Sets At Specific Values of θ_{ab}

		c		
k	$1-\alpha$	π	$\frac{\pi}{2}$	$\frac{\pi}{3}$
2	0.99	2.575829	2.574961	2.557816
	0.95	1.960000	1.954508	1.916332
	0.90	1.644854	1.632219	1.576989
3	0.99	2.712313	2.659548	2.596961
	0.95	2.123498	2.028012	1.944581
	0.90	1.823565	1.695619	1.598824
4	0.99	2.787521	2.681167	2.605202
	0.95	2.195720	2.043793	1.950064
	0.90	1.890690	1.708132	1.602907

It is notable from this table that it is possible to obtain a far more than marginal reduction in the value of c, compared to the methods in Chapter 4, but may require a particularly small value of θ_{ab}, that may be difficult to obtain in practical situations. For two sided simultaneous confidence sets, given in Table 5.1, as expected there is no improvement in the value of c at $\theta_{ab} = \pi$, and $\theta_{ab} = \frac{\pi}{2}$ for $k = 2$, these are the same values of c as given in Table 4.1. We note that for $k = 2$ and $\theta_{ab} = \frac{\pi}{3}$ there is only a very minor reduction of about 0.025 units in c compared to the AS2 method at $\alpha = 0.05$. This is a particularly small angle to obtain practically. Improvement is however more

noticeable for higher k and can then be obtained at $\theta_{ab} = \frac{\pi}{2}$. As one would expect, the size of c, and by extension the size of the confidence set increases as k increases, however the extent of the improvement also increases.

The values of c are lower in all cases for the one sided sets of Table 5.2 compared to those in Table 5.1 as expected. Furthermore, we note there exist more significant reductions in the values of c for AS1 sets, compared to the methods of Chapter 4, than in the AS2 case. This is particularly noticeable for $\theta_{ab} = \frac{\pi}{2}$, when $k = 3$ and 4, and in all cases for $\theta_{ab} = \frac{\pi}{3}$. That we begin to see a smaller c at $\theta_{ab} = \frac{\pi}{2}$ for all cases would suggest that the analytical minimal values of θ_{1k}, for AS1 sets with an unconstrained effective dose, are at least larger than $\frac{\pi}{2}$. As expected, the improvements are more noticeable at higher$k$ and also at higher values of $1 - \alpha$. Overall the extent of improvement is better for one sided sets than two.

The main limitation of the methods of this chapter is that noticeable improvement can seemingly only be obtained when θ_{ab} is around $\frac{\pi}{3}$, and rarely at $\frac{\pi}{2}$. It may be very difficult in practice to assume a small enough sensible range of the effective doses to achieve these values. For the "mtcars" dataset, it is sensible to assume a range the same as the range of observed covariates, at $[10-35]$, which gives $\theta_{10,35} = 2.508589$ and $c = 2.236477$ and 1.96 respectively for AS2 and AS1 confidence sets, which represents a negligible improvement. The values of c were almost identical over$[15 - 25]$ with $\theta_{15,25} = 1.802701$, and eventually with a very small sensible range of $[20 - 25]$, we get $\theta_{20,25} = 0.915296$ with AS1 and AS2 c values as 2.197 and 1.898. However the confidence set in this case, at $p = 0.4$ for AS2 sets gave an interval $[14.68, 26.1]$ which is much wider than the assumed range. As such, this small a sensible range was not acceptable. More work in this area is needed to show that a sensible range can give a small enough θ_{ab} to achieve a significant, if at all, reduction in the value of c, over the methods of Chapter 4.

Chapter 6

Conclusions and Further Work

The thesis conducts research into optimal designs for one covariate simultaneous confidence bands constrained over a finite interval. Unlike previous work, the thesis has focused on optimal designs under a continuous framework, and formulated the Average Width and Minimum Area Confidence Set criterion functions for continuous designs. In doing so, a method of numeric analysis has been developed where designs may be defined over a two dimensional surface that is constrained within intervals. This addresses the limitation of the numeric analysis conducted in Ah-Kine (2010), and permits a numeric search algorithm for the optimal. We showed numerically that for a Scheffé type or constant width type band over $[-1, 1]$, the best design was D optimal for both criteria for a given constant N and α. We also showed strong evidence that this is the case for all possible N and α. An analytical proof was also attempted for the first time by applying the methods of Silvey (1980) to the dual objective function.

A method is also introduced to construct simultaneous confidence sets on the effective dose by inverting Scheffé type confidence bands, that are less conservative than currently established methodology. The method hinges on establishing specialised values of the critical constant c, involving one dimensional integration, to obtain simultaneous confidence sets for a specific number of effective doses. Both one and two sided simultaneous confidence sets are fully defined for two, three and four effective doses for one covariate models, with two sided sets for two effective doses available for multiple covariate models. The same sets are also constructed under the stipulation that the effective doses lie over some sensible range. These are shown to be an improvement over the previous methods with examples.

Furthermore, the thesis identifies a number of problems that may be of interest in further research. An analytical proof of the optimal designs for Scheffé type and constant width bands, constrained over a finite interval remains to be proven. The thesis manipulates the criterion function into an appropriate function ψ, but ran into difficulty applying Theorem 3.2 due to the complicated nature of the derivatives. Proving the required properties of these derivatives to apply this Theorem warrants further investigation.

The numeric analysis is flawed in that bounding the variables m_1 and m_2 over a rectangular interval is not entirely appropriate, as they are not independent of one

another. This limited the utility of numeric methods. In practice the investigation would be of greater utility if one could establish the surface area in which the points (m_1, m_2) correspond only to real designs, of which is contained within the rectangular bound. For $[-1, 1]$ this area is likely parabolic, however the surface will likely change with the number of design points, and the support.

For Chapter four, with the exception of two sided sets for $k = 2$, the main limitation is that the improved methodology for simultaneous confidence sets on the effective dose only applies to a univariate linear model. Trying to generalise the methods to obtain the relevant c for a multiple covariate model represents a significant challenge. The thesis considered a geometric approach. In order to extend this approach to a multiple covariate problem, one would have to minimise, with respect to k effective doses, the probability of $\boldsymbol{N}$ lying in a $2k$ sided (or k in the one sided case) $q+1$ dimensional region. As many interesting problems in real life involve multiple covariate models, this is an area of further research of particular interest. An algebraic, or numeric approach could be considered an alternative when more than one covariate is considered.

Finally we note that the critical constant for the confidence bands, used to generate the confidence sets of the Scheffé band method do **not** depend on the design, the precise difficulty with the research conducted in Chapter three. Hence it is realistically possible to obtain the optimal design of experiments that minimise the size of the simultaneous confidence sets on the effective dose, including the original, and the adapted methodology established in this thesis.

The most important contribution of the thesis is the method to construct improved simultaneous confidence sets on the effective dose. These methods are immediately applicable to univariate linear GLM, and are guaranteed to be at least as informative as the current Scheffé band method. It can be easily used in replacement, provided the number of effective doses is specified, by simply changing the value of c.

Appendices

A: Proof Of Theorem 3.1

If the support of the design is $[a, A]$ then we may write $x_i = A - e_i$ where $e_i \in [0, A-a]$. We may then write

$$\begin{aligned} m_1 =& \sum_{i=1}^{n} p_i x_i \\ =& \sum_{i=1}^{n} p_i (A - e_i) \\ =& A \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} p_i e_i \\ =& A - \sum_{i=1}^{n} p_i e_i \leq A \end{aligned}$$

since both p_i and e_i are always at least zero. By instead setting $x_i = a + e_i$, $e_i \in [0, A-a]$ we immediately have

$$m_1 = a + \sum_{i=1}^{n} p_i e_i \geq a.$$

Hence $m_1 \in [a, A]$ as required. It is immediately obvious that m_2 is always at least greater than zero, and may attain the value of zero when $0 \in [a, A]$, when the design is a special case that has one design point $x_1 = 0$ with $p_1 = 1$. Let $x_i = A - e_i$, then

$$\begin{aligned}
m_2 =& \sum_{i=1}^{n} p_i x_i^2 \\
=& \sum_{i=1}^{n} p_i (A - e_i)^2 \\
=& \sum_{i=1}^{n} p_i (A^2 - 2Ae_i + e_i^2)^2 \\
=& A^2 \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} (-2Ap_i e_i) + \sum_{i=1}^{n} p_i e_i^2 \\
=& A^2 + \sum_{i=1}^{n} p_i (e_i^2 - 2Ae_i) \\
=& A^2 + \sum_{i=1}^{n} p_i (e_i (e_i - 2A)).
\end{aligned}$$

Therefore $m_1 \leq A^2$ when $\sum_{i=1}^{n} p_i (e_i (e_i - 2A)) \leq 0$. This is guaranteed if $e_i - 2A \leq 0 \; \forall i = 1, \ldots, n$. As $e_i \in [0, A - a]$, its maxmum value is $A - a$ and hence this condition is guaranteed when $A - a \leq 2A$, or $A + a \geq 0$. With $x_i = a + e_i$ a similar argument gives

$$m_2 = a^2 + \sum_{i=1}^{n} p_i (e_i (e_i + 2a)).$$

This is less than a^2 when $e_i + 2a \leq 0$ with e_i at its maximum at $A - a$, that is hence when $A - a + 2a \leq 0$ or when $A + a \leq 0$ as required. When $A = -a$ it is clear that either case is correct. These conditions also clearly cover all possible $a < A$ and the proof is finished.

B: List Of "R" Codes For Numeric Analyses

R Codes For Datasets

- Linear reg bands.R" Simultaneous confidence bands for the "bloodfat" dataset.
- "Plotting confidence bands for logistic- final.R" Simultaneous confidence bands and confidence sets for the heart exercise dataset.
- "Logistic Confidence bands.R" Simultaneous confidence bands and confidence sets for the "mtcars" dataset.

Chapter 3

- "Constant aw complete code.R" Response surface investigation for the Average Width of a constant width band.
- "Constant Macs complete code.R" Response surface investigation for the MACS criterion of a constant width band.
- "Hyperbolic aw complete code.R" Response surface investigation for the Average Width of a Scheffé type band.
- "Hyperbolic Macs complete code.R" Response surface investigation for the MACS criterion of a Scheffé band.
- "optimCAW.R" Optimisation algorithm for the Average Width of a constant width band.
- "optimCMACS.R" Optimisation algorithm for the MACS criterion of a constant width band.
- "optimHYAW.R" Optimisation algorithm for the Average Width of a Scheffé type band.
- "optimHYMACS.R" Optimisation algorithm for the MACS criterion of a Scheffé band.
- "optimCAWe.R" End point optimisation algorithm for the Average Width of a constant width band.
- "optimCMACSe.R" End point optimisation algorithm for the MACS criterion of a constant width band.
- "optimHYAWe.R" End point optimisation algorithm for the Average Width of a Scheffé type band.
- "optimHYMACSe.R" End point optimisation algorithm for the MACS criterion of a Scheffé band.

Chapter 4

- "c values.R" Critical constants for two sided simultaneous confidence sets.
- "c values one sided.R" Critical constants for one sided simultaneous confidence sets.

Chapter 5

- "c values for finite effective dose.R" Critical constants for two sided simultaneous confidence sets for constrained effective doses.
- "finite c one sided.R" Critical constants for one sided simultaneous confidence sets for constrained effective doses.

References

Ah-Kine, P. (2010.). Simultaneous confidence bands in linear regression analysis. Southampton University E-Prints.

Atkinson, A., Donev, A. and Tobias, R. (2007.). *Optimal Experimental Designs With SAS*. Oxford University Press.

Bowden, D. and Graybill, F. (1966.). Confidence bands of uniform and proportional width for linear models. *Journal of The American Statistical Association*, 61:182–198.

Brand, R. J., Pinnock, D. E. and Jackson, K. (1973.). Large sample confidence bands for the logistic response curve and its inverse. *The American Statistician*, 27:157–160.

Buckley, B. and Piegorsch, W. (2008.). Simultaneous confidence bands for abbott adjusted quantal response models. *Statistical Methodology*, 5:209–219.

Carter, W., Chinchilli, V., Wilson, J., Campbell, E., Kessler, F. and Carchman, R. (1986.). An asymtotic confidence region for the ed_{100p} from the logistic response surface for a combination of agents. *The American Statistician*, 40:124–128.

Cox, C. (1990.). Fiellers theorem, the likelihood and the delta method. *Biometrics*, 46:709–718.

Deutsch, R. and Piegorsch, W. (2012.). Benchmark dose profiles for joint action quantal data in quantitative risk assessment. *Biometrics*, 68:1313–1322.

Edward, W. and Murray, G. (1969.). A table of integrals of the error functions. *Journal of Research of The National Bureau of Standards (B)*, 73.

Elfving, G. (1952.). Optimum allocation in linear regression theory. *Annals of Mathematical Statistics*, 23:255–262.

Fedorov, V. (1972.). *Theory of Optimal Experiments*. New York: Academic Press.

Fieller, E. (1954.). Some problems in interval estimation. *Journal of The Royal Statistical Society (B)*, 16:175–185.

Finney, D. (1971.). *Probit Analysis, Second Edition*. Cambridge University Press.

Finney, D. (1978.). *Statistical Method In Biological Assay*. Oxford University Press.

Florida State University (2011). Regression: Linear regression datasets. http://people.sc.fsu.edu/∼jburkardt/datasets/regression/regression.html. Last accessed on Mar 25, 2016.

Gafarian, A. (1964.). Confidence bands in straight line regression. *Journal of the American Statistical Association*, 59:182–213.

Graybill, F. and Bowden, D. (1967.). Linear segment confidence bands for simple linear regression models. *Journal of The American Statistical Association*, 62:403–408.

Hochberg, Y. and Quade, D. (1975.). One-sided simultaneous confidence bounds on regression surfaces with intercepts. *Journal of The American Statistical Association*, 70:889–891.

Hollander, M., McKeague, I. and Yang, J. (1997.). Likeihood ratio-based confidence bands for the survival function. *Journal of The American Statistical Association*, 92:215–226.

Hsu, J. (1996.). *Multiple Comparisons, Theory and Methods*. Chapman and Hall.

Kiefer, J. (1958.). On the non-randomized optimality and randomized non-optimality of symmetrical designs. *Annals of Mathematical Statistics*, 28:675–699.

Kiefer, J. (1959.). Optimum experimental designs (with discussion). *Journal of the Royal Statistical Society (B)*, 21:272–319.

Kiefer, J. and Wolfowitz, J. (1960.). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366.

Kotz, S. and Johnson, N. (1992.). *Breakthroughs In Statistics Volume I: Foundations and Basic Theory*. Springer-Verlag.

Li, J., Nordheim, E., Zhang, C. and Lehner, C. (2008.). Estimation and confidence regions for multi-dimensional effective dose. *Biometrical Journal*, 50:110–122.

Lin, D. (1994.). Confidence bands for survival curves under the proportional hazards model. *Biometrika*, 81:73–81.

Lin, H. and Chang, F. (2007.). On minimally-supported d-optimal designs for polynomial regression with log-concave weight function. *Metrika*, 65:227–233.

Liu, W. (2011.). *Simultaneous Inference In Regression*. CRC Press.

Liu, W. and Ah-Kine, P. (2010.). Optimal simultaneous confidence bands in simple linear regression. *Journal of Statistical Planning and Inference*, 140:1225–1235.

Liu, W. and Hayter, A. (2007.). Minimum area confidence set optimality for confidence bands in simple linear regression. *Journal of The American Statistical Association*, 102:181–190.

Liu, W., Jamshidian, M., Zhang, Y. and Bretz, F. (2005.a). Constant width simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals. *Journal of Statistical Computation and Simulation*, 75(6):425–436.

Liu, W., Jamshidian, M., Zhang, Y. and Donnerlly, J. (2005.b). Simulation based simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals. *Journal of Computational and Graphical Statistics*, 14:459–484.

Liu, W., Lin, S. and Piegorsch, W. (2008.). Construction of exact simultaneous confidence bands for a simple linear regression model. *International Statistical Review*, 76:39–57.

Naiman, D. (1983.). Comparing scheffé type to constant width confidence bounds in regression. *Journal of the American Statistical Association*, 78:906–912.

Naiman, D. (1984a). Average width optimality of simultaneous confidence bands. *Annals of Statistics*, 12:1199–1214.

Pan, W., Piegorsch, W. and West, R. (1975.). Exact one-sided simultaneous confidence bands via uusipaikkas method. *Annals of The Institute of Statistical Mathematics*, 55:243–250.

Piegorsch, W. (1985b.). Average with optimality for confidence bands in simple linear regression. *Journal of the American Statistical Association*, 80:692–697.

Piegorsch, W. and Casella, G. (1988.). Confidence bands for logistic regression with restricted predictor variables. *Biometrics*, 44:739–750.

Piegorsch, W. and West, R. (2005.). Low dose risk estimation via simultaneous statistical inferences. *Journal of the Royal Statistical Society (C)*, 54:245–258.

Ross, S. (1988.). *A First Course In Probability*. Macmillan Publishing Company.

Scheffé, H. (1953.). A method for judging all contrasts in analysis of variance. *Biometrika*, 40:87–104.

Silvey, S. (1980.). *Optimal Design:An Introduction to the Theory For Parameter Estimation*. Chapman and Hall.

Smith, K. (1918.). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12:1–85.

Uusipaikka, E. (1983.). Exact confidence bands for linear regression over intervals. *Journal of the American Statistical Association*, 78:638–644.

Walter, W. (1983.). A note on confidence bands for the logistic response curve. *The American Statistician*, 37:158–160.

Working, H. and Hotelling, H. (1929.). Applications of the theory of error to the interpretation of trends. *Journal of The American Statistical Association*, 24:73–85.

Wynn, H. and Bloomfield, P. (1971.). Simultaneous confidence bands in regression analysis. *Journal of The Royal Statistical Society (B)*, 33:202,217.