1

of/de

1

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISION^{SM} RESOLUTION TARGETS

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file   Votre référence

Our file   Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

F- OPTIMAL DESIGNS FOR BINARY RESPONSE EXPERIMENTS

by

ILANA FAINARU, B.Sc.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of

Master of Science

Department of Mathematics and Statistics

Carleton University

Ottawa, Ontario

August, 1994

© Copyright, 1994

Ilana Fainaru

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file   Votre référence

Our file   Notre référence

THE AUTHOR HAS GRANTED AN
IRREVOCABLE NON-EXCLUSIVE
LICENCE ALLOWING THE NATIONAL
LIBRARY OF CANADA TO
REPRODUCE, LOAN, DISTRIBUTE OR
SELL COPIES OF HIS/HER THESIS BY
ANY MEANS AND IN ANY FORM OR
FORMAT, MAKING THIS THESIS
AVAILABLE TO INTERESTED
PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE
IRREVOCABLE ET NON EXCLUSIVE
PERMETTANT A LA BIBLIOTHEQUE
NATIONALE DU CANADA DE
REPRODUIRE, PRETER, DISTRIBUER
OU VENDRE DES COPIES DE SA
THESE DE QUELQUE MANIERE ET
SOUS QUELQUE FORME QUE CE SOIT
POUR METTRE DES EXEMPLAIRES DE
CETTE THESE A LA DISPOSITION DES
PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP
OF THE COPYRIGHT IN HIS/HER
THESIS. NEITHER THE THESIS NOR
SUBSTANTIAL EXTRACTS FROM IT
MAY BE PRINTED OR OTHERWISE
REPRODUCED WITHOUT HIS/HER
PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE
DU DROIT D'AUTEUR QUI PROTEGE
SA THESE. NI LA THESE NI DES
EXTRAITS SUBSTANTIELS DE CELLE-
CI NE DOIVENT ETRE IMPRIMES OU
AUTREMENT REPRODUITS SANS SON
AUTORISATION.

ISBN 0-315-98632-8

Canada

Name Iana Fainaru

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Statistics

SUBJECT TERM

0463

U·M·I

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language 0679
General 0289
Ancient 0290
Linguistics 0291
Modern 0401
Literature 0294
General 0295
Classical 0297
Comparative 0298
Medieval 0316
Modern 0591
African 0305
Asian 0352
Canadian (English) 0355
Canadian (French) 0593
English 0311
Germanic 0312
Latin American 0315
Middle Eastern 0313
Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion 0318
General 0321
Biblical Studies 0319
Clergy 0320
History of 0322
Philosophy of 0469
Theology 0323

SOCIAL SCIENCES

American Studies 0323
Anthropology 0324
Archaeology 0326
Cultural 0327
Physical 0310
Business Administration 0272
General 0770
Accounting 0454
Banking 0338
Management 0385
Marketing 0501
Canadian Studies 0503
Economics 0505
General 0508
Agricultural 0509
Commerce-Business 0510
Finance 0511
History 0358
Labor 0366
Theory 0351
Folklore 0578
Geography 0460
Gerontology 0383
History 0386

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science 0615
General 0616
International Law and Relations 0617
Public Administration 0814
Recreation 0452
Social Work 0626
Sociology 0627
General 0938
Criminology and Penology 0631
Demography 0628
Ethnic and Racial Studies 0629
Individual and Family Studies 0630
Industrial and Labor Relations 0700
Public and Social Welfare 0344
Social Structure and Development 0709
Theory and Methods 0999
Transportation 0453
Urban and Regional Planning 0537
Women's Studies 0538

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture 0473
General 0285
Agronomy 0475
Animal Culture and Nutrition 0476
Animal Pathology 0359
Food Science and Technology 0478
Forestry and Wildlife 0479
Plant Culture 0480
Plant Pathology 0817
Plant Physiology 0777
Range Management 0746
Wood Technology 0306
Biology 0287
General 0308
Anatomy 0309
Biostatistics 0379
Botany 0329
Cell 0353
Ecology 0369
Entomology 0793
Genetics 0410
Limnology 0307
Microbiology 0317
Molecular 0416
Neuroscience 0433
Oceanography 0821
Physiology 0778
Radiation 0472
Veterinary Science 0786
Zoology 0760
Biophysics 0425
General 0996
Medical

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0985
Paleozoology 0427
Palynology 0368
Physical Geography 0415
Physical Oceanography

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences 0566
General 0300
Audiology 0992
Chemotherapy 0567
Dentistry 0350
Education 0769
Hospital Management 0758
Human Development 0982
Immunology 0564
Medicine and Surgery 0347
Mental Health 0569
Nursing 0570
Nutrition 0380
Obstetrics and Gynecology 0354
Occupational Health and Therapy 0381
Ophthalmology 0571
Pathology 0419
Pharmacology 0572
Pharmacy 0382
Physical Therapy 0573
Public Health 0574
Radiology 0575
Recreation

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences 0485
Chemistry 0749
General 0486
Agricultural 0487
Analytical 0488
Biochemistry 0738
Inorganic 0490
Nuclear 0491
Organic 0494
Pharmaceutical 0495
Physical 0754
Polymer 0405
Radiation 0605
Mathematics 0986
Physics 0606
General 0608
Acoustics 0748
Astronomy and Astrophysics 0607
Atmospheric Science 0798
Atomic 0759
Electronics and Electricity 0609
Elementary Particles and High Energy 0610
Fluid and Plasma 0752
Molecular 0756
Nuclear 0611
Optics 0463
Radiation 0346
Solid State 0984
Statistics 0984
Applied Sciences 0984
Applied Mechanics 0984
Computer Science

Engineering 0537
General 0538
Aerospace 0539
Agricultural 0540
Automotive 0541
Biomedical 0542
Chemical 0543
Civil 0544
Electronics and Electrical 0348
Heat and Thermodynamics 0545
Hydraulic 0546
Industrial 0547
Marine 0794
Materials Science 0548
Mechanical 0743
Metallurgy 0551
Mining 0552
Nuclear 0549
Packaging 0765
Petroleum 0554
Sanitary and Municipal 0790
System Science 0428
Geotechnology 0796
Operations Research 0795
Plastics Technology 0994
Textile Technology

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

**Thesis Acceptance Form
ILANA FAINARU**

The undersigned recommend to the Faculty of Graduate Studies
and Research acceptance of the thesis

"F- OPTIMAL DESIGNS FOR BINARY RESPONSE EXPERIMENTS"

submitted by Ilana Fainaru, B.Sc.

in partial fulfilment of the requirements for the degree of Master of Science

---

Thesis Supervisor

Chair, Department of Mathematics and Statistics

Carleton University

August, 1994

Abstract

Binary response experiments are widely used for estimating the $ED\xi$ (the dose causing a $\xi\%$ response). The efficiency of the estimated $ED\xi$ depends largely on how the experiment was designed. Optimal designs are designs that focus on maximizing the efficiency of such estimates. In this thesis, an F-optimal design is proposed for estimating the $ED\xi$. For a class of symmetric two parameter models which include the logit and probit models, we show that F-optimal designs have 2 support points. We numerically obtain these designs for different ξ , and various sample sizes n , and compare them to the 1- and 2- point c-optimal designs in Wu (1988). The results indicate that for very large n and for extreme percentiles even when n is moderate, the c- and F-optimal designs are similar. The most pronounced difference occurs for $8.3 < \xi < 91.7$, and moderate n . In this case, c-optimality gives a 1-point design with all of the weight at the $ED\xi$, whereas F-optimality gives a 2-point design with weights nearly evenly split between the two design points.

Acknowledgements

I would like to thank Dr. R.R. Sitter for his great supervision and encouragement in preparing this thesis. I would also like to thank Barbara E. Forbes for her help with the Nelder-Mead algorithm and Mathematica.

Table of Contents

	Page
Acceptance Sheet	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	vi
List of Figures	vii
List of Appendices	viii
 1 Introduction	 1
2 Binary Response Experiments	5
3 Interval Estimation for μ_ξ	8
3.1 Asymptotic Confidence Intervals and Fieller intervals for μ_ξ	9
3.2 A Theoretical Comparison	14
4 Optimal Designs for Binary Response Experiments	17
4.1 c-Optimal Designs	19
4.2 F-Optimal Designs for Estimating μ_ξ	22
4.2.1 Case 1: $\xi = 50$	22
4.2.2 Case 2: $0 < \xi < 100$	26
4.3 Numerical Results	32
5 Summary	35
References	65

LIST OF TABLES

Table	Description	Page
1	c-Optimal Designs for Estimating Percentiles	37
2	F-Optimal Designs for Estimating Percentiles	38

LIST OF FIGURES

Figure	Description	Page
4.1	A plot of $\mathcal{H}$, $-\mathcal{H}$, and $\mathcal{H}^*$ for the logit model	41
4.2	A plot of $\mathcal{CH}_2$ for the logit model	42
4.3	A plot of $\mathcal{CH}_3$ for the logit model	43

LIST OF APPENDICES

Appendix	Description	Page
I	Proof of $J(\underline{\theta})$ in (3.1)	44
	Proof of (3.5), (3.6), (3.7)	48
II	Derivation of (3.9) and (3.10)	50
III	Proof of (4.2) and (4.7)	53
IV	Proof of (4.9)	56
V	Derivation of (4.16)	57
VI	Proof of Result 4.3 (i) for F-Optimality	61
VII	Proof of Result 4.3 (ii) for the logit and probit models	62
	Proof of Result 4.3 (iii) for the logit model	63

Chapter 1

Introduction

A typical binary response experiment consists of applying a stimulus to a subject and observing the subject's response or non-response. Such an experiment is widely used in bioassay for estimating the potency of suspect carcinogens. Here, the stimulus is a drug, the subject an animal, and the response is the occurrence or non-occurrence of a malignant tumor. Due to limited resources in space, time and cost, the dose-levels used in such experiments must be large enough to induce an observable rate of response. Estimates of risk in humans can then be obtained by first extrapolating from results obtained at high dose levels to the lower levels anticipated in the environment, and then converting them from animal to man (Krewski & Ryzin, 1981).

Many quantitative measures of potency have been proposed in the literature. The most widely used measure is the ED_{ξ} , the dose causing a $\xi\%$ response rate. The ED_{50} , the dose causing a 50% response rate, is widely used in bioassay. Estimates of the ED_{50} have been computed for over 3700 experiments which are summarized in the Carcinogenic Potency Database (Gold et al., 1984,1989). Quantities such as the ED_{75} and ED_{95} are also important in the pharmaceutical industry. For example, a pharmaceutical company may have a drug that works 50% of the time but may be interested in finding a drug that does better than that (ie 75% response). Also, extremely small values of ξ yield quantities that are used in the estimation of *virtually safe doses* (VSDs).

Many mathematical models are available for estimating the $ED\xi$. A model consists of an equation that relates the dose levels of a chemical to the observed responses. Some examples are the logit, probit, extreme value, gamma multi-hit and multi-stage models (Krewski & Ryzin, 1981). The logit model is desired in practice for it is computationally simple and provides a satisfactory fit to binary data, as implied by a chi-squared goodness of fit test. All of the above models will give similar estimates of the ED_{50} with improvements of fit occurring in the tail areas. For example, in the evaluation of virtually safe doses, response rates as low as 10^{-6} or 10^{-8} may be involved (Morgan, 1992).

Accurate estimation of the $ED\xi$ is often desired from a binary response experiment. The degree of accuracy depends largely on how the experiment was designed. The problem of designing an experiment consists of choosing the number of dose levels, the actual dose levels, and the number of subjects to test at each dose level. Choosing these quantities is not easy, and often involves some compromise to obtain a good design. According to Sitter & Wu (1992), a good design is one that (i) provides accurate point and interval estimation of the unknown model parameter $\underline{\theta}$, (ii) allows for model checking; and (iii) satisfies certain cost constraints. In this thesis, we consider optimal designs, designs that focus on (i) above.

Optimal designs measure accuracy by some function of the expected Fisher information matrix, called the design criterion. Unfortunately, the design criterion depends on the unknown model parameter $\underline{\theta}$, so a best guess of $\underline{\theta}$ is often employed when constructing the design. Designs which are optimal for a best guess of $\underline{\theta}$ are sometimes called locally optimal designs. The choice of the design criterion depends on the goals of the experiment. In this thesis, we are interested in estimating the $ED\xi$ for any $0 < \xi < 100$. The most widely used criterion for this is the asymptotic variance of the maximum likelihood estimator of the $ED\xi$. This criterion is called c-optimality. Not much work has been done in the area of optimal design for estimating the $ED\xi$ when $\xi \neq 50$. Below, we summarize some recent work.

Krewski & Kovar (1982) derive c-optimal designs for estimating virtually safe levels of exposure to chemical carcinogens (VSDs), under the one-hit, extreme-value, probit and logit models with known slope parameter. Estimating the VSD is equivalent to estimating the $ED\xi$ for extremely small ξ ($=10^{-4}$ or 10^{-6}). They show that the optimal design under the one-hit/extreme-value model is a one-point design at the ED_{80} , and the optimal design under the probit/logit model is a one-point design at the ED_{50} . These optimal designs are shown to be robust in the sense that a loss of efficiency is minimal if the dose employed is close to the optimal dose.

Krewski et al. (1984) derive c-optimal designs for estimating the VSD under the probit, logit and extreme-value models, with unknown parameter θ . The case of zero background, which assumes that there are no responses at zero dose, is examined. They show that the optimal design for estimating the VSD is a two-point design, with more weight at the lower dose level. Table I in Krewski et al. (1984) gives c-optimal designs for estimating the $ED\xi$ for $\xi = 1, 10^{-3}$, and 10^{-6} , for the probit, logit and extreme-value models. For example, the c-optimal design for $\xi = 10^{-3}$ under the logit model is a two-point design with 76% of the observations at 0.083 and 24% at 0.917. These results are consistent with those in Wu (1988).

Wu (1988) derives c-optimal designs for estimating the $ED\xi$ for $0 < \xi < 100$, for a class of symmetric models that include the probit and logit models. He shows that for $100p_1 < \xi < 100p_2$, the c-optimal design is a one-point design at the $ED\xi$, while for ξ outside of this range, the optimal design is a two-point design with weights α at $F^{-1}(p_2)$ and $1-\alpha$ at $1-F^{-1}(p_2)$. Table I in Wu (1988) gives c-optimal designs for the probit and logit models. For example, the c-optimal design for the logit model is a one-point design at ξ for $0.083 < \xi < 0.917$, and a two-point design with weights at $p_1 = 0.083$ and $p_2 = 0.917$ for ξ outside of this range.

The c-optimality criterion is only valid for very large sample sizes. In practice, however, n is finite so we might want to consider an alternative criterion. Sitter & Wu

(1993) conducted a simulation study to compare the performance of the asymptotic confidence interval and a Fieller interval for the ED50. They found that a Fieller interval gave more accurate intervals for the ED50 than the asymptotic confidence interval, particularly when n was small. This led them to consider the use of a Fieller interval as an alternative design criterion for estimating the ED50. Sitter & Wu (1993) derive F-optimal designs for estimating the ED50. They completely characterize these designs by explicit formula, and then numerically obtain them for a class of symmetric models that includes the logit and probit models (see Table 1 in Sitter & Wu, 1993).

In this thesis, we are concerned with optimal designs for binary response experiments conducted in order to estimate the ED ξ . We derive c- and F-optimal designs for estimating the ED ξ , under the logit and probit models. In chapter 2, we define the ED ξ , and show how it can be estimated from our models. Interval estimates of the ED ξ are obtained in chapter 3. We derive the asymptotic confidence interval and a Fieller interval for estimating the ED ξ , thus extending some of the results in Sitter & Wu (1993). The derivations are similar, and the Sitter & Wu results remain as a special case. The two intervals are then theoretically compared.

In chapter 4, we derive c-optimal and F-optimal designs for estimating the ED ξ . Sitter & Wu (1993) derive F-optimal designs for estimating the ED50 for a class of symmetric models which include the logit and probit models. We present their derivation and discuss why their method does not work for $\xi \neq 50$. We then use a different approach to extend their results to any ξ . In the process, characterizations of F-optimal designs are obtained. We numerically obtain c- and F-optimal designs for the logit and probit models for various values of ξ , and compare the resulting designs for various values of n . We find that for extreme percentiles and large n , the c- and F-optimal designs give similar results. However, for ξ within a certain range, the two designs differ markedly: c-optimality yields a one-point design while F-optimality yields a two-point design.

Chapter 2

Binary Response Experiments

A binary response experiment involves administering a stimulus to a subject at some dose level x , and observing a response or non-response. These outcomes, response and non-response, occur with probabilities $p(x | \underline{\theta})$ and $q(x | \underline{\theta}) = 1 - p(x | \underline{\theta})$, where $\underline{\theta}$ is a vector of unknown parameters and $p(\cdot)$ is some specified function. Repeating the experiment k independent times with n_i subjects at dose level x_i , $i=1, \dots, k$, and observing the number of responses r_i at each dose level, gives us a set of mutually independent random variables $r_i \sim \text{bin}(n_i, p(x_i | \underline{\theta}))$ with log-likelihood

$$L(\underline{\theta}) \propto \sum_{i=1}^k \{ r_i \log p(x_i | \underline{\theta}) + (n_i - r_i) \log [1 - p(x_i | \underline{\theta})] \}. \quad (2.1)$$

The maximum likelihood estimator (M.L.E.) $\hat{\underline{\theta}}$ can be obtained by maximizing $L(\underline{\theta})$ over $\underline{\theta} \in \Theta$, where Θ is the parameter space.

Although many forms of $p(\cdot)$ are available, the most commonly used is

$$p(x | \underline{\theta}) = F[\beta(x - \mu)] \quad (2.2)$$

where $\underline{\theta} = (\mu, \beta)^T$ and $F(\cdot)$ is a cumulative distribution function with

$$F(-t) = 1 - F(t). \quad (2.3)$$

Two well known models of this type are the logit and probit models. Here, the parameter μ is the value of x that yields a 50% response rate, termed the ED50. The ED50 is a widely used potency measure in bioassay and therefore, it is often of interest to estimate it well. When this is so, the use of the form (2.2) is convenient, since estimation of the ED50 corresponds to estimation of μ .

In some cases, the primary interest is not in estimating the ED₅₀, but rather in estimating the value of x that yields a $\xi\%$ response rate, termed the ED ξ , where $0 < \xi < 100$. In order to obtain estimates of the ED ξ , it will be more convenient to use the following reparameterization,

$$p(x | \underline{\theta}) = F[\beta(x - \mu_\xi) + \gamma_\xi], \quad (2.4)$$

where $\underline{\theta} = (\mu_\xi, \beta)^T$, $F(\cdot)$ is defined as in (2.3), and $\gamma_\xi = F^{-1}(\xi/100)$. Model (2.2) is a special case of model (2.4) when $\xi = 50$, since (2.3) implies $F^{-1}(\frac{1}{2}) = 0$. We have simply reparameterized the model in (2.2), so that the parameter μ_ξ is now the value of x that causes a $\xi\%$ response rate. To see this, note that the ED ξ is the value of x which satisfies

$$p(x | \underline{\theta}) = \xi/100.$$

This implies

$$F[\beta(x - \mu_\xi) + \gamma_\xi] = \xi/100.$$

Applying F^{-1} to both sides of the above yields

$$\beta(x - \mu_\xi) + \gamma_\xi = F^{-1}(\xi/100) = \gamma_\xi$$

which implies

$$\mu_\xi = \text{ED}_\xi.$$

Hence, estimating the ED ξ corresponds to estimating μ_ξ . Throughout this thesis, we will restrict to the more general formulation in (2.4).

Below are examples of two types of models: symmetric (satisfies (2.3)) and asymmetric. Consider the models given below.

(1) Logit $F(t) = \exp(t)/[1 + \exp(t)]$

(2) Probit $F(t) = \int_{-\infty}^t \phi(y)dy$, where $\phi(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)$

(3) Power logistic $F(t) = \{\exp(t)/[1 + \exp(t)]\}^m, m > 0.$

The first two models are symmetric, and have been known to describe many types of data sets. This paper will deal primarily with these two models. The third model, the power logistic model (PLM), is an asymmetric model and may be used to describe skewed dose-

response curves. The term "power logistic" is used to indicate that the PLM is the m^{th} power of the logistic model. Model (1) is a special case of (3) when $m = 1$. A detailed discussion of the PLM may be found in Johnson and Kotz (1970).

We include the PLM for the following reason only. Gaudard et al. (1993) show that optimal designs are generally robust to misspecification of m . That is, with correct initial guesses of μ_{ξ} and β , the loss in efficiency is quite small if one uses a design which is optimal, assuming $m=1$ (logit model). Thus, the optimal designs which we will later derive for the logit model will remain good if the data is not too severely skewed, and follows a power logit model.

Chapter 3

Interval Estimation for μ_{ξ}

In chapter 2, we found that estimating the $\xi\%$ response dose, ED_{ξ} , is equivalent to estimating μ_{ξ} . Point or interval estimators of μ_{ξ} may be obtained. In this section, we consider methods of interval estimation.

The most commonly used interval estimator for μ_{ξ} is the *asymptotic confidence interval*, which is based on the asymptotic variance of $\hat{\mu}_{\xi}$, $AV(\hat{\mu}_{\xi})$. According to Finney (1971, p144), inferences made from this type of interval may lead to "gravely misleading conclusions", unless the sample size, n , is large enough to make use of the interval's asymptotic correctness. Thus, for small to moderate n , the asymptotic confidence interval should not be applied.

Finney (1971) proposed the use of what is sometimes termed a *Fieller interval*, which is based on Fieller's theorem. When estimating μ_{50} , Sitter & Wu (1993) show that the Fieller interval is asymptotically equivalent to the asymptotic confidence interval, but has superior performance for finite n . Abdelbasit & Plackett (1983) first compared the two intervals for estimating μ_{50} , but found a negligible difference in their performance. Sitter & Wu (1993) reinvestigated the results of Abdelbasit & Plackett, and extended the simulation study to include smaller values of n . They concluded that the previous authors had used too large a sample size, and showed the Fieller intervals to be indeed superior for small to moderate values of n .

In this part of the report, we theoretically compare the two competing intervals for estimating μ_ξ , for $0 < \xi < 100$, thus extending some of the results in Sitter & Wu (1993). The derivations are similar, and the Sitter & Wu results remain as a special case for $\xi = 50$. In section 3.1, we derive the asymptotic confidence interval and the Fieller interval for μ_ξ , for any $0 < \xi < 100$. In section 3.2, we compare the two intervals.

3.1 Asymptotic Confidence Intervals and Fieller Intervals for μ_ξ

The expected Fisher information matrix for $\underline{\theta} = (\mu_\xi, \beta)^T$ under the model implied by (2.1) and (2.4) is

$$J(\underline{\theta}) = n \begin{pmatrix} \beta^2 S_0 & -(S_1 - \gamma_\xi S_0) \\ -(S_1 - \gamma_\xi S_0) & \frac{1}{\beta^2} (S_2 - 2\gamma_\xi S_1 + \gamma_\xi^2 S_0) \end{pmatrix} = nI(\underline{\theta}), \quad (3.1)$$

where

$$S_j = \sum_{i=1}^k \lambda_i t_i^j \psi(t_i), \quad j = 0, 1, 2,$$

$$t_i = \beta(x_i - \mu_\xi) + \gamma_\xi, \quad 0 < \xi < 100,$$

$$\lambda_i = n_i / n, \quad n = \sum n_i,$$

and

$$\psi(t) = \frac{[F'(t)]^2}{F(t)[1 - F(t)]}$$

(See Appendix I.1). When $\xi = 50$, we have $\gamma_\xi = 0$, so (3.1) reduces to

$$J(\underline{\theta}) = n \begin{pmatrix} \beta^2 S_0 & -S_1 \\ -S_1 & \frac{1}{\beta^2} S_2 \end{pmatrix} = nI(\underline{\theta})$$

From (3.1), we obtain

$$\begin{aligned}
J^{-1}(\underline{\theta}) &= \frac{1}{nD} \begin{pmatrix} \frac{1}{\beta^2}(S_2 - 2\gamma_{\xi}S_1 + \gamma_{\xi}^2S_0) & (S_1 - \gamma_{\xi}S_0) \\ (S_1 - \gamma_{\xi}S_0) & \beta^2S_0 \end{pmatrix} \\
&= \frac{1}{n} I^{-1}(\underline{\theta}) \\
&= \frac{1}{n} \begin{pmatrix} I^{11} & I^{12} \\ I^{12} & I^{22} \end{pmatrix}, \text{ say,}
\end{aligned}$$

where

$$D = \frac{1}{n^2} \det[J(\underline{\theta})] = \det[I(\underline{\theta})] = S_0S_2 - S_1^2.$$

It is interesting to note that D is of the same form irrespective of the value of ξ .

Consider the following well known result, referred to below, as Result 3.1.

Result 3.1: Let $\hat{\underline{\theta}} = (\hat{\mu}_{\xi}, \hat{\beta})^T$ be the maximum likelihood estimator for $\underline{\theta}$. Then, under certain regularity conditions, we have

$$(i) \sqrt{n}(\hat{\underline{\theta}} - \underline{\theta}) \xrightarrow{d} N(\underline{0}, I^{-1}(\underline{\theta})) \text{ and}$$

$$(ii) \sqrt{n}(g(\hat{\underline{\theta}}) - g(\underline{\theta})) \xrightarrow{d} N(\underline{0}, \underline{c}^T I^{-1}(\underline{\theta}) \underline{c}), \text{ where } \underline{c} = \left(\frac{\partial g}{\partial \mu}, \frac{\partial g}{\partial \beta} \right)^T.$$

We can now use Result 3.1 to derive the two competing intervals. First, the asymptotic confidence interval. By letting $g(\hat{\underline{\theta}}) = \hat{\mu}_{\xi}$, Result 3.1 (ii) yields

$$\sqrt{n}(\hat{\mu}_{\xi} - \mu_{\xi}) \xrightarrow{d} N(0, I^{11})$$

where I^{11} is defined in (3.2). So for large n

$$\Pr(|\hat{\mu}_{\xi} - \mu_{\xi}| \leq z_{\alpha} \sqrt{AV(\hat{\mu}_{\xi})}) \doteq 1 - 2\alpha$$

or

$$\Pr(\hat{\mu}_{\xi} - z_{\alpha} \sqrt{AV(\hat{\mu}_{\xi})} \leq \mu_{\xi} \leq \hat{\mu}_{\xi} + z_{\alpha} \sqrt{AV(\hat{\mu}_{\xi})}) \doteq 1 - 2\alpha$$

where $AV(\hat{\mu}_\xi) = I^{11}/n$, $0 < \alpha < 1$ and z_α is the $1-\alpha$ point of the standard normal distribution. So an approximate $(1-2\alpha)100\%$ confidence interval for μ_ξ , (called the *asymptotic confidence interval*) is

$$\hat{\mu}_\xi \pm z_\alpha \sqrt{AV(\hat{\mu}_\xi)}. \quad (3.3)$$

When applying the above interval, $AV(\hat{\mu}_\xi)$ is estimated by $\hat{AV}(\hat{\mu}_\xi)$, that is, μ_ξ and β are replaced by $\hat{\mu}_\xi$ and $\hat{\beta}$ in I^{11} .

Now the Fieller interval. Suppose that we have the following simple linear regression model

$$y_i := F^{-1}(r_i/n_i) = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, k, \quad (3.4)$$

with $\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_k)^T$, $E(\underline{\varepsilon}) = 0$ and $V(\underline{\varepsilon}) = \sigma^2 \text{diag}(1/[n_1 \psi(t_1)], \dots, 1/[n_k \psi(t_k)])$. From (2.4) and (3.4), we have

$$\beta_0 + \beta_1 x_i = \beta(x_i - \mu) + \gamma$$

which implies that

$$\beta_0 = -\mu\beta + \gamma \quad \text{and} \quad \beta_1 = \beta.$$

Note that we have again suppressed the subscripts on μ and γ for simplicity. We assume approximate normality of r_i/n_i for large n_i , and therefore for y_i . Using the asymptotics of maximum likelihood estimators, we have

$$\begin{aligned} \hat{\beta}_0 + \hat{\beta}_1 \mu &\xrightarrow{d} N(\beta_0 + \mu\beta_1, V^*) \\ &= N(\gamma, V^*) \end{aligned}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the weighted least squares estimators and $V^* = AV(\hat{\beta}_0 + \hat{\beta}_1 \mu)$. This implies that

$$\hat{\beta}_0 + \hat{\beta}_1 \mu - \gamma \xrightarrow{d} N(0, V^*).$$

Now,

$$\begin{aligned} V^* &= AV(\hat{\beta}_0 + \hat{\beta}_1 \mu) \\ &= AV(\hat{\beta}_0) + 2\mu \text{ACov}(\hat{\beta}_0, \hat{\beta}_1) + \mu^2 AV(\hat{\beta}_1), \end{aligned}$$

where ACov denotes asymptotic covariance. So for large n ,

$$\Pr(|\hat{\beta}_0 + \hat{\beta}_1\mu - \gamma| \leq z_\alpha V^{*1/2}) \doteq 1 - 2\alpha$$

or

$$\Pr([\hat{\beta}_0 + \hat{\beta}_1\mu - \gamma]^2 - z_\alpha^2 V^* \leq 0) \doteq 1 - 2\alpha,$$

which can be re-written as

$$P((A\mu^2 + B\mu + C) \leq 0) \doteq 1 - 2\alpha,$$

where,

$$A = \hat{\beta}_1^2 - z_\alpha^2 AV(\hat{\beta}_1),$$

$$B = 2[\hat{\beta}_0\hat{\beta}_1 - \hat{\beta}_1\gamma - z_\alpha^2 ACov(\hat{\beta}_0, \hat{\beta}_1)],$$

and

$$C = \hat{\beta}_0^2 + \gamma^2 - 2\hat{\beta}_0\gamma - z_\alpha^2 AV(\hat{\beta}_0).$$

It is shown in Appendix I.2 that

$$AV(\hat{\beta}_1) = AV(\hat{\beta}) = I^{22}/n, \quad (3.5)$$

$$AV(\hat{\beta}_0) = AV(-\hat{\mu}\hat{\beta} + \gamma) = AV(\hat{\mu}\hat{\beta}) = (\beta^2 I^{11} + 2\mu\beta I^{12} + \mu^2 I^{22})/n, \quad (3.6)$$

and

$$ACov(\hat{\beta}_0, \hat{\beta}_1) = ACov(-\hat{\mu}\hat{\beta}, \hat{\beta}) = -ACov(\hat{\mu}\hat{\beta}, \hat{\beta}) = -(\beta I^{12} + \mu I^{22})/n. \quad (3.7)$$

Thus,

$$A = \hat{\beta}_1^2 - z_\alpha^2 I^{22}/n,$$

$$B = 2[z_\alpha^2(\hat{\beta}_1 I^{12} + \hat{\mu} I^{22})/n - \hat{\mu}\hat{\beta}_1^2],$$

and

$$C = \hat{\mu}^2\hat{\beta}_1^2 - z_\alpha^2(\hat{\beta}_1^2 I^{11} + 2\hat{\mu}\hat{\beta}_1 I^{12} + \hat{\mu}^2 I^{22})/n.$$

A $(1-2\alpha)$ 100% Fieller interval for μ is thus

$$\frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

which we write as

$$M \pm N, \quad (3.8)$$

with $M = -B/2A$ and $N = \sqrt{B^2 - 4AC}/2A$. When applying the interval in (3.8), I^{11} , I^{12} , and I^{22} are estimated by $\hat{I}^{11}$, $\hat{I}^{12}$, and $\hat{I}^{22}$. Substituting the values of $\hat{I}^{11}$, $\hat{I}^{12}$, and $\hat{I}^{22}$ into (3.8) and replacing the subscripts on μ and γ yields

$$\hat{\mu}_{\xi} - c_1 \frac{g}{1-g} \pm \frac{z_{\alpha}}{\hat{\beta}\sqrt{n}} \left\{ \frac{1}{(1-g)S_0} + \frac{(S_1 - \gamma_{\xi}S_0)^2}{(1-g)^2 S_0 D} \right\}^{1/2}, \quad (3.9)$$

which is equivalent to

$$\hat{\mu}_{\xi} - c_1 \frac{g}{1-g} \pm z_{\alpha} \sqrt{AV(\hat{\mu}_{\xi})} \left\{ \frac{(1-c_2g)^{1/2}}{1-g} \right\}, \quad (3.10)$$

where $c_1 = (S_1 - \gamma_{\xi}S_0)/\hat{\beta}S_0$, $c_2 = D/S_0(S_2 - 2\gamma_{\xi}S_1 + \gamma_{\xi}^2 S_0)$, and $g = z_{\alpha}^2 S_0/nD$. We should note that $\hat{\mu}_{\xi}$ and $\hat{\beta}$ are being used in S_0 , S_1 , and S_2 . Derivations of (3.9) and (3.10) are given in Appendix II.

The Fieller interval is only valid for $g < 1$. When $g \geq 1$, the resulting interval is either semi-infinite or infinite, and may not contain $\hat{\mu}_{\xi}$. When this occurs, making inferences about μ_{ξ} is non-sensical. To see this, if $g > 1$, then by definition, we have

$$z_{\alpha}^2 > \frac{nD}{S_0} = \frac{1}{(\frac{S_0}{nD})} = \frac{\hat{\beta}^2}{(\frac{\hat{\beta}^2 S_0}{nD})} = \frac{\hat{\beta}^2}{A\hat{V}(\hat{\beta})}$$

or

$$\frac{|\hat{\beta}|}{\sqrt{A\hat{V}(\hat{\beta})}} < z_{\alpha}, \quad (3.11)$$

where $AV(\hat{\beta})$ is given in (3.5). The statement in (3.11) says that the 2-tailed z-test for β is insignificant at level 2α . In terms of the experiment, this means that the data do not

provide sufficient evidence to indicate that a dose response relationship exists. Sitter & Wu (1993) advise either collecting more data at other dose levels or redefining the $ED\xi$.

3.2 A Theoretical Comparison

In this section, we compare the asymptotic confidence and Fieller intervals given in (3.3) and (3.10) respectively. The intervals to be compared are

$$\text{Confidence} \quad \hat{\mu}_\xi \pm z_\alpha \sqrt{AV(\hat{\mu}_\xi)}$$

$$\text{Fieller} \quad \hat{\mu}_\xi - c_1 \frac{g}{1-g} \pm z_\alpha \sqrt{AV(\hat{\mu}_\xi)} \left\{ \frac{(1-c_2g)^{1/2}}{1-g} \right\},$$

where $c_1 = (S_1 - \gamma_\xi S_0) / \hat{\beta} S_0$ and $c_2 = D / [S_0(S_2 - 2\gamma_\xi S_1 + \gamma_\xi^2 S_0)]$. We see that the Fieller interval differs from the confidence interval by a shift factor

$$K = c_1 \frac{g}{1-g}$$

and an inflation factor of

$$L = \left\{ \frac{(1-c_2g)^{1/2}}{1-g} \right\}.$$

First, note that since $g \rightarrow 0$ as $n \rightarrow \infty$, which implies $K \rightarrow 0$ and $L \rightarrow 1$, the Fieller interval and the asymptotic confidence interval are asymptotically equivalent.

Second, consider the inflation factor L which depends on ξ (through c_2) and the asymptotic variance of $\hat{\beta}$ (through g). The term c_2 may be rewritten as

$$c_2 = \frac{D}{D + (S_1 - \gamma_\xi S_0)^2}$$

which shows that $0 \leq c_2 \leq 1$. Recall that we would only use either interval if $g < 1$. For $g < 1$, we have

$$0 < 1-g \leq 1-c_2g \leq (1-c_2g)^{1/2} < 1$$

which implies that $L \geq 1$ and hence, the Fieller interval is always longer than the confidence interval.

Last, consider the shift factor K . The shift factor K can be rewritten as

$$\begin{aligned} K &= \frac{g}{1-g} c_1 \\ &= \frac{g}{1-g} \frac{(S_1 - \gamma_\xi S_0)}{\hat{\beta} S_0} \\ &= \frac{g}{1-g} \frac{\sum \lambda_i (t_i - \gamma_\xi) \psi(t_i)}{\hat{\beta} \sum \lambda_i \psi(t_i)} \\ &= \frac{g}{1-g} \frac{\sum \lambda_i (x_i - \hat{\mu}_\xi) \psi(t_i)}{\sum \lambda_i \psi(t_i)}. \end{aligned}$$

Sitter and Wu (1993) note that for the ED50 and $\psi(\cdot)$ symmetric, a symmetric about the ED50 design will imply this shift factor is zero. From the above equation, we see that for general ξ , this is not the case. Using the fact that $g = z_\alpha^2 S_0 / (nD) = O_p(n^{-1})$, it can easily be shown that $K = O_p(n^{-1})$ and $L = 1 + O_p(n^{-3/2})$. This implies that the shift factor K is the more important difference between the intervals (see Sitter & Wu, 1993 for discussion), and thus the two intervals will differ most when

$$|\sum \lambda_i (x_i - \hat{\mu}_\xi) \psi(t_i) / \sum \lambda_i \psi(t_i)|$$

is large.

To evaluate the performance of these intervals, one should conduct a simulation study that compares the coverage probability to nominal coverage for each interval. Sitter & Wu (1993) conducted such a study for Fieller and confidence intervals about μ_{50} . They used the logit model and a Hewlett data set (see Abdelbasit & Plackett, 1983) to simulate 10000 data sets. One and two-tailed coverage probabilities were compared to nominal coverages of 2.5% in each tail. Comparisons were made for various values of n : 180, 225, 270, 360, 450, 900, μ_{50} : .1, .2, .3, and β : 7, 14, using 9 fixed x_i values. They found that the

Fieller intervals perform better than the confidence intervals, particularly for small n . Comparing the coverage probabilities of the Fieller intervals with the confidence intervals, the Fieller intervals were found to have one- and two- tailed coverages that were closer to the nominal values of 2.5% and 5%.

If one were to conduct a similar study for intervals about μ_{ξ} , where $0 < \xi < 100$, then one would anticipate similar results. That is, the Fieller intervals should outperform the confidence intervals in terms of both one-tailed and two-tailed coverages. This suggests that one should use a Fieller interval for analysis over the Confidence interval. If one does, then it makes sense that they should also use it as the design criterion. This leads to the consideration of Fieller intervals as design criteria, in chapter 4.

Chapter 4

Optimal Designs for Binary Response Experiments

Designing a binary response experiment involves selecting k , the number of dose levels, $\{x_i\}_{i=1}^k$, the set of dose levels, and $\{n_i\}_{i=1}^k$, the number of test subjects at each dose level. Although many designs are possible, we want a good design. According to Sitter & Wu (1992), a good design is one that (i) provides accurate point and interval estimation of $\underline{\theta}$, (ii) allows for model checking; and (iii) satisfies certain cost constraints. Constructing such a design is often difficult so one settles for a compromise design.

In this paper, we consider optimal designs, designs that maximize the accuracy of point estimates of $\underline{\theta}$. That is, we will focus on (i) above. Accuracy is usually measured by $J^{-1}(\underline{\theta})$, which is equal to $AV(\hat{\underline{\theta}})$, the asymptotic variance of the maximum likelihood estimator, $\hat{\underline{\theta}}$ of $\underline{\theta}$. Thus, a good design will make $J(\underline{\theta})$ 'large' or $J^{-1}(\underline{\theta})$ 'small'.

One of the main concerns about optimal designs is that they assume that a good initial estimate of $\underline{\theta}$ is available. In practice, such estimates are unavailable. Secondly, these designs yield too few dose levels (usually 1 or 2) at which to gather responses. This does not allow for adequate model checking. However, these designs may still be used as a benchmark for heuristic compromise designs which take (i), (ii) and (iii) into account, or in choosing a second stage in a two-stage procedure (see Forbes, 1994).

In general, optimal designs are obtained by minimizing $G(J^{-1}(\underline{\theta}))$ over integer values of k , $\{x_i\}_{i=1}^k$, and $\{n_i\}_{i=1}^k$ for some function $G(\cdot)$. There are two problems with this. First of all, $G(J^{-1}(\underline{\theta}))$ depends on $\underline{\theta}$, which is unknown at the design stage. Thus, one must assume that a good initial estimate of $\underline{\theta}$ is available, and use this estimate as if it were the true value of $\underline{\theta}$. From (2.1) and (2.4), we see that this dependence on $\underline{\theta}$ comes only through the $t_i = \beta(x_i - \mu_\xi) + \gamma_\xi$, $i = 1, \dots, k$. The set of values $\{t_i\}_{i=1}^k$ that minimizes the design criterion is called the support of the design. From the support points, and a given initial estimate of $\underline{\theta}$, one may determine the set of dose levels $\{x_i\}_{i=1}^k$ required for the experiment.

And secondly, minimizing over the integers is difficult due to the discrete nature of the problem. If we use a real λ_i with $0 \leq \lambda_i \leq 1$ in place of n_i/n , $n = \sum n_i$ then the design problem reduces to finding optimal k , $\{t_i\}_{i=1}^k$ and $\{\lambda_i\}_{i=1}^k$. Note that we have reformulated the problem to a continuous setting. After an optimal continuous design is obtained, the values of n_i are obtained by rounding the value of $n\lambda_i$ to the nearest integer. It is hoped that the solution to the continuous problem is quite close to the discrete solution.

Different designs use different functions of $J^{-1}(\underline{\theta})$, $G(J^{-1}(\underline{\theta}))$, as criteria. One question which frequently arises is "Which design criterion should be used?". The choice of criterion depends on the goals of the experiment. For example, if one is equally interested in estimating both μ_ξ and β , then either D, A, or E-optimality criteria may be used. The corresponding optimal designs are obtained by minimizing the determinant, the trace, and the maximum eigenvalue of $J^{-1}(\underline{\theta})$ (Kalish & Rosenberger, 1978).

When interest lies in estimating μ_ξ , for $0 < \xi < 100$, the criterion called c-optimality may be used. The c-optimal design is obtained by minimizing the asymptotic variance of μ_ξ , which is equivalent to minimizing the length of the asymptotic confidence interval in (3.6). Wu (1988) has shown that the c-optimal design for estimating μ_ξ has either one point at μ_ξ or two points, one on each tail of the distribution function F . In chapter 3, we suggested that interval estimates of μ_ξ , based on the asymptotic variance, would likely be

less accurate than Fieller intervals. For this reason, we consider the length of the Fieller interval as a design criterion, termed F-optimality. The F-optimal design is obtained by minimizing this length.

In this chapter, we study both c- and F-optimal designs for estimating μ_ξ , $0 < \xi < 100$. In section 4.1, we summarize the results of Wu (1988) on c-optimal designs. Sitter & Wu (1993) derive F-optimal designs for estimating μ_{50} for a class of symmetric models which include the logit and probit models. In section 4.2, we present their derivation and discuss why their method does not work for $\xi \neq 50$. We then use a different approach to extend their results to any ξ . In the process, characterizations of the F-optimal designs are obtained. In section 4.3, we use the results in 4.1 and 4.2 to numerically obtain the c-optimal and F-optimal designs for the logit and probit models for various values of ξ and compare the resulting designs for various values of n .

4.1 c-Optimal Designs

In this section, we summarize Wu's results on c-optimality. The c-optimality criterion may be used when estimation of μ_ξ is of primary interest. c-optimality involves minimizing the asymptotic variance of $\hat{\mu}_\xi$, i.e.

$$\min \underline{c}^T J^{-1}(\underline{\theta}) \underline{c} \quad (4.1)$$

where, $\underline{c}^T = (1, 0)$, and J^{-1} is the inverse of J . The minimization is over all choices of t_i and λ_i , subject to the constraint $0 < \lambda_i \leq 1$, $\sum \lambda_i = 1$. In Appendix III, we show that (4.1) is equivalent to

$$\min(1, \gamma_\xi) \left\{ \sum_{i=1}^k \lambda_i \psi(t_i) (1, t_i) (1, t_i)^T \right\}^{-1} (1, \gamma_\xi)^T \quad (4.2)$$

subject to $0 < \lambda_i \leq 1$, $\sum \lambda_i = 1$.

The problem in (4.1) may be solved by a method proposed by Elfving (1952) which involves geometrical and other arguments. Let

$$\mathcal{H} = \{ \psi^{1/2}(t)(1,t) : -\infty < t < \infty \}$$

be a curve in $\mathbb{R}^2$, and let $-\mathcal{H}$ be its reflection on the origin. Let $\mathcal{H}^*$ be the boundary of the convex hull generated by $\mathcal{H}$ and $-\mathcal{H}$. Figure 4.1 shows a plot of $\mathcal{H}$, $-\mathcal{H}$ and $\mathcal{H}^*$ for the logit distribution. One important assumption made in Elfving (1952) is that the curve $\mathcal{H}$ is closed and bounded. The point where the vector $\underline{v}^T = (1, \mu_\xi)$, extended if necessary, pierces this convex hull is a convex combination of some of the extreme points in $\mathcal{H}$ or $-\mathcal{H}$. These extreme points are the support points of the design.

The slope of $\mathcal{H}$ is

$$h(t) = \frac{\frac{d}{dt} [\psi^{1/2}(t)t]}{\frac{d}{dt} \psi^{1/2}(t)} = t + 2\psi(t) / \psi'(t). \quad (4.3)$$

The following conditions, stated in Wu (1988) are required for finding c-optimal designs:

- (i) $\lim_{|t| \rightarrow \infty} |t^2 \psi(t)| = 0$
- (ii) $\mathcal{H}$ is convex. (4.4)

Condition (4.4)(i) implies that $\mathcal{H}$ is closed and bounded with the origin as its leftmost point. To show that $\mathcal{H}$ is convex, we need to show that $h'(t) > 0$ for $t > 0$ and $h'(t) < 0$ for $t < 0$. A sufficient condition for this is to show that $\psi(t)$ is log-concave, that

$$\frac{\psi'(t)}{\psi(t)} = \frac{d}{dt} \log(\psi(t)) < 0. \quad (4.5)$$

Wu shows that conditions (i) and (4.5) hold for a class of symmetric models which include the logit and probit models.

The following result appearing in Wu (1988) summarizes Elfving's results:

Result 4.1: *Let the line $\mathcal{H}$ satisfy the conditions (4.4). When $F'(t)$ is symmetric about the origin, there exists $p_1 = F(-t^*)$ and $p_2 = F(t^*) = 1 - p_1$, with $t^* \geq 0$ satisfying*

$$h(t^*) = t^* + 2\psi(t^*) / \psi'(t^*) = 0, \quad (4.6)$$

where $t^* \geq 0$. Then,

(i) if $p_1 < \xi/100 < p_2$, the c -optimal design for estimating μ_ξ is the 1-point design at μ_ξ ;

(ii) if $\xi/100 < p_1$ or $\xi/100 > p_2$, the c -optimal design for estimating μ_ξ is the 2-point design, with weights λ at $F^{-1}(p_2) = t^*$ and $1-\lambda$ at $F^{-1}(p_1) = -t^*$, where λ is determined by

$$\lambda = \frac{1}{2} + \frac{1}{2} \frac{t^*}{\gamma_\xi}. \quad (4.7)$$

See Appendix III for a proof of (4.7).

Wu (1988) shows that part (ii) of Result 4.1 may still hold, even when $\mathcal{H}$ is not convex but satisfies (4.4)(i). He gives this as a corollary, which is stated below.

Corollary: Suppose that the line $\mathcal{H}$ satisfies (4.4)(i), and that $F'(t)$ is symmetric. If there exists a $t^* \geq 0$ satisfying (4.6) and

$$\psi^{1/2}(t^*)t^* = \max \{ \psi^{1/2}(t)t : -\infty < t < \infty \}, \quad (4.8)$$

then for $\xi < 100(1-p^*)$ or $\xi > 100p^*$, $p^* = F(t^*)$, the c -optimal design for estimating μ_ξ is a 2-point design, with weights λ on $F^{-1}(p^*)$ and $1-\lambda$ on $F^{-1}(1-p^*)$, λ determined by (4.7).

Wu(1988) uses Result 4.1 to obtain c -optimal designs for a class of symmetric models that includes the logit and probit models. In section 4.3, we numerically obtain these designs for the logit model for various values of ξ . The c -optimal design for estimating μ_ξ has either one point at μ_ξ or two points, one on each tail of the distribution function F . When using c -optimality as a benchmark for percentiles within a certain range, the one point design seems counterintuitive. The reason for this is that taking all observations near a single percentile, μ_ξ , may not give an accurate estimate of μ_ξ , particularly if the initial guesses of μ_ξ and β are incorrect.

4.2 F-Optimal Designs for Estimating μ_ξ

An alternative criterion for estimating μ_ξ , based on the length of a Fieller interval, is termed F-optimality. The criterion is given by

$$\begin{aligned} I &= \frac{1}{1 - aS_0/D} + \frac{(S_1 - \gamma_\xi S_0)^2}{(1-g)^2 S_0 D} \\ &= \frac{1}{(1 - a/S_2)S_0} + (S_1 - \gamma_\xi S_0)^2 \left[\frac{aS_1^2}{D(1-g)(1 - a/S_2)S_0 S_2 (S_1 - \gamma_\xi S_0)^2} + \frac{1}{S_0 D(1-g)^2} \right] \\ &= I_1 + I_2, \end{aligned} \quad (4.9)$$

where $a = z_\alpha^2/n$, $g = aS_0/D$ and the design is chosen to minimize I (see Appendix IV).

Designs minimizing (4.9) are called F-optimal designs. When $\xi = 50$, (4.9) reduces to

$$\begin{aligned} I &= \frac{1}{(1 - a/S_2)S_0} + S_1^2 \left[\frac{a}{D(1-g)(1 - a/S_2)S_0 S_2} + \frac{1}{S_0 D(1-g)^2} \right] \\ &= I_1 + I_2, \end{aligned} \quad (4.10)$$

and the problem reduces to minimizing (4.10). Sitter & Wu (1993) derive F-optimal designs for estimating μ_{50} . Their method is simple, but only works for $\xi = 50$. We summarize their results below.

4.2.1 Case 1: $\xi = 50$

The F-optimality criterion in (4.10) depends on the matrix

$$M = \begin{pmatrix} S_0 & S_1 \\ S_1 & S_2 \end{pmatrix} \quad (4.11)$$

which can be identified with the point (S_0, S_1, S_2) in R^3 . The symmetry property of $F'(t)$ about zero may be used to simplify the design problem. Consider any design $d = \{ (t_i, \lambda_i) \}$. Let $\bar{d} = \{ (-t_i, \lambda_i) \}$; then the design given by $d^* = (d + \bar{d})/2$ is symmetric. Since $F'(t)$ is symmetric about zero then $\psi(t) = \psi(-t)$ which implies that

$$S_0(d^*) = S_0(d), \quad S_1(d^*) = 0, \quad \text{and} \quad S_2(d^*) = S_2(d).$$

Since

$$g = aS_0/D = aS_0/(S_0S_2 - S_1^2) \geq aS_0/S_0S_2 = a/S_2,$$

then $g < 1$ implies that $a/S_2 < 1$. Hence, the term $I_2 \geq 0$ in (4.10). Setting $S_1 = 0$ yields $I_2 = 0$. This shows that the design d^* improves over d for F-optimality. Thus given any design d , there exists a symmetric design which is at least as good.

Thus, in finding F-optimal designs, only designs that are symmetric about zero need to be considered. So any optimal design will have $S_1 = 0$, which means that we only have to consider designs over $t \geq 0$ and

$$M = \begin{pmatrix} S_0 & 0 \\ 0 & S_2 \end{pmatrix}. \quad (4.12)$$

This simplifies the design problem in two ways. First of all, with $S_1 = 0$, (4.10) reduces to

$$I = \frac{1}{(1 - a/S_2)S_0}$$

over $t_i \geq 0$, and $0 < \lambda_i \leq 1$, noting that the resulting design will split the λ_i mass at t_i equally between $\pm t_i$.

Secondly, we can use Caratheodory's theorem (see Federov (1972) p.66) to determine the maximum number of support points for the optimal design. To see this, consider the matrix in (4.12), which corresponds to the point (S_0, S_2) in R^2 . Let $\mathcal{M}$ be the convex hull of all such points. Let

$$\mathcal{H}_2 = \{ \psi(t) (1, t^2) : 0 < t < \infty \}$$

be a curve in R^2 , and let $\mathcal{CH}_2$ be the convex hull of $\mathcal{H}_2$. A plot of $\mathcal{CH}_2$ for the logit distribution is shown in Figure 4.2. The upper line of $\mathcal{H}_2$ is traced by moving from $t = 0$ to $t = \infty$. Take any point $(x, y) \in \mathcal{CH}_2$. By definition, any such point can be expressed as a linear combination of the points $t_1, t_2, \dots, t_k$ in $\mathcal{H}_2$. That is,

$$\begin{aligned} (x, y) &= \left(\sum_{i=1}^k \lambda_i \psi(t_i), \sum_{i=1}^k \lambda_i t_i^2 \psi(t_i) \right) \\ &= (S_0, S_2). \end{aligned}$$

This shows that any point in the $\mathcal{CH}_2$ corresponds to a matrix of the form in (4.12). Thus, the set of all points in $\mathcal{CH}_2$ corresponds to the set of all matrices M in $\mathcal{M}$. Caratheodory's theorem implies that for any $M \in \mathcal{M}$, there exists a design with at most 3 support points. But each support point represents 2 points: t and $-t$, so $3 t_i \geq 0$ imply a total of 6 points. If M is a boundary point of $\mathcal{M}$, then we get at most 4 points in total.

For the F-criterion, the function $x(1-a/y)$ to be maximized in the (x,y) -plane has a convex contour that increases in the outward direction when $a = z_\alpha^2/n$ is small. This implies that the optimal solution must lie on the upper boundary of $\mathcal{CH}_2$. So once we characterize this upper boundary, solutions to the optimal design problem may be obtained in terms of $\mathcal{H}_2$. Sitter & Wu (1993) consider three possible characterizations for the upper boundary of $\mathcal{CH}_2$. We discuss two of these, although only one of them is of primary interest to us. Result 4.2 below summarizes their results on F-optimality.

Result 4.2: (a) Suppose that the curve $\mathcal{H}_2$ is concave. Then, the boundary point of $\mathcal{CH}_2$ is $\psi(t)(1, t^2)$ for some t . This means that over $(0, \infty)$, the optimal design is the 1-point design at t . Recalling that each support point in our solution represents 2 points: half the weight at t and half at $-t$. So over $(-\infty, \infty)$, the F-optimal design is the 2-point design with weights $1/2$ at t and $-t$. Here, t must satisfy

$$\psi'(t) + \frac{2a}{t^3} = 0,$$

where $a = z_\alpha^2 / n$.

(b) Suppose that the curve $\mathcal{H}_2$ is convex over $[0, t^*]$, for some $0 \leq t^* \leq t_{\max}$. Then, the boundary point is on the line segment joining $(\psi(0), 0)$ and $(\psi(t), \psi(t)t^2)$ for some t . So by symmetry, the F-optimal design is the 3-point design at $-t, 0, t$ with weights $\lambda/2, 1-\lambda$, and $\lambda/2$. Here, t must satisfy

$$t + 2\left\{1 - \frac{\psi(t)}{\psi(0)}\right\} \frac{\psi(t)}{\psi'(t)} = 0,$$

and λ is given by

$$\lambda = \left[\frac{a\psi(0)}{t^2\psi(t)\{\psi(0) - \psi(t)\}} \right]^{1/2},$$

where $a = z_\alpha^2 / n$.

Showing that $\mathcal{H}_2$ is concave is equivalent to showing that

$$h_2'(t) > 0 \text{ for } t > 0,$$

where

$$h_2(t) = \frac{\frac{d}{dt}[t^2\psi(t)]}{\frac{d}{dt}\psi(t)} = th(t),$$

where

$$h(t) = t + 2\psi(t)/\psi'(t)$$

appears in equation (4.3) in section 4.1. It will be shown that $\mathcal{H}_2$ is concave for the logit model. A plot of $\mathcal{H}_2$ for the probit indicates that it is concave, although no rigorous proof has been found.

A sufficient condition for the concavity of $\mathcal{H}_2$ is that

$$h''(t) > 0, \text{ for } t > 0.$$

Sitter & Wu (1993) have shown that this condition holds for some well known symmetric models (the double exponential and double reciprocal models).

Unfortunately, the above method of Sitter & Wu (1993) does not work for values of $\xi \neq 50$. For the F-optimality criterion in (4.9), the multiplier in I_2 is

$$(S_1 - \gamma_\xi S_0)^2 = S_1^2 - 2\gamma_\xi S_0 S_1 + \gamma_\xi^2 S_0^2. \quad (4.13)$$

Setting $S_1 = 0$ gives us

$$(S_1 - \gamma_\xi S_0)^2 = \gamma_\xi^2 S_0^2. \quad (4.14)$$

For certain values of $\gamma_\xi \neq 0$, the right-hand side of (4.13) may actually be less than the right-hand side of (4.14). Therefore, there will be cases where design d^* will not improve over d . Thus, the above method does not work for $\xi \neq 50$. In the next section, we consider a different approach which works for any ξ .

4.2.2 Case 2: $0 < \xi < 100$

In this section, we extend the results of section 4.2.1 and derive F-optimal designs for estimating μ_ξ for any ξ . From section (4.2.1), we saw that the F-optimality criterion depends on the matrix in (4.11),

$$M = \begin{pmatrix} S_0 & S_1 \\ S_1 & S_2 \end{pmatrix},$$

which can be identified with the point (S_0, S_1, S_2) in R^3 . Let $\mathcal{M}_2$ be the convex hull of all such points. Let $\mathcal{H}_3$ be the line in R^3 , denoted by

$$\mathcal{H}_3 = \{ \psi(t)(1, t, t^2) : -\infty < t < \infty \}$$

and let $\mathcal{CH}_3$ be the convex hull of $\mathcal{H}_3$. A plot of $\mathcal{CH}_3$ for the logit distribution is given in Figure 4.3. The upper and lower lines of $\mathcal{H}_3$ are traced by moving from $t = 0$ to $t = \infty$ and from $t = -\infty$ to $t = 0$ respectively. The shaded area is called the vertical boundary of the convex hull, $\mathcal{CH}_3$. Any point $(x, y, z) \in \mathcal{CH}_3$ can be written as a linear combination of the points $t_1, \dots, t_k$ in $\mathcal{H}_3$. That is,

$$\begin{aligned} (x, y, z) &= \left(\sum_{i=1}^k \lambda_i \psi(t_i), \sum_{i=1}^k \lambda_i t_i \psi(t_i), \sum_{i=1}^k \lambda_i t_i^2 \psi(t_i) \right) \\ &= (S_0, S_1, S_2). \end{aligned}$$

Therefore, the set of all points in $\mathcal{CH}_3$ corresponds to the set of all points in $\mathcal{M}_2$. Caratheodory's theorem implies that for any $M \in \mathcal{M}_2$, there exists a design with at most 4 support points. If M is a boundary point of $\mathcal{M}_2$, then there will be at most 3 points.

The following result shows that under certain conditions, the optimal solution to the design problem must lie on the vertical boundary of $\mathcal{CH}_3$. From Figure 4.3, we see that any such point can be written as a convex combination of the $\pm t$ values directly above and below it on the line $\mathcal{H}_3$.

Result 4.3: *Let (u,v,w) represent the three axes in Figure 4.3. Suppose that all of the following conditions hold:*

- (i) the optimality function, as a function of (u,v,w) is monotonically improving in w for fixed u and v ;*
- (ii) the curve $(\psi(t), t\psi(t))$ is concave for $t > 0$,*
- (iii) the curve $(\psi(t), t^2\psi(t))$ is concave for $t > 0$.*

Then the optimal design has two points at $-t$ and t with weights λ and $1-\lambda$ for some $0 \leq \lambda \leq 1$.

Each of the conditions in Result 4.3 can be viewed geometrically. For example, condition (iii) implies that the curved vertical boundary of the logit, shown in Figure 4.3 as the shaded curved part of the plane, has a convex shape. Condition (ii) implies that if we observe the line $\mathcal{H}_3$ from the (u,v) - plane, then for $t > 0$ it is concave and for $t < 0$ it is convex. This condition is related to condition (ii) in (4.4). The proof of Result 4.2 consists of showing that the optimal solution lies on the vertical boundary of $\mathcal{CH}_3$. The argument goes as follows:

In Figure 4.3, we see that $\mathcal{CH}_3$ may be thought of to consist of 3 pieces: a top lid, a bottom lid, and the vertical boundary, which is indicated by the shaded region. Any point on the top or bottom lid can always be improved by moving it horizontally, in the w direction, until it reaches the vertical boundary. This is true because: 1) Conditions (ii) and (iii) imply that any point moving in the w direction remains in $\mathcal{CH}_3$; and 2) from condition (i), moving a point in the w direction will always decrease (or improve) the optimality function. Similarly, any point in the interior of $\mathcal{CH}_3$ can also be improved by moving it in the w direction until it reaches the vertical boundary.

In Appendix VI, we show that the F-optimality criterion satisfies condition (i). Appendix VII shows that condition (ii) is satisfied by the logit and probit functions and that the logit satisfies (iii). It is clear from a plot of $(\psi(t), t^2\psi(t))$ for the probit that (iii) holds, but neither Sitter & Wu (1993) nor this author has yet to prove it. With this result,

the design problem has been reduced to a two variable search for the logit and probit models.

We can now characterize F-optimal designs for the logit and probit models (and any model satisfying the conditions of Result 4.3). Recall from section 4.2 that the F-optimal design is obtained by minimizing

$$I_{\xi}(\lambda, t) = \frac{1}{(1 - aS_0/D)S_0} + \frac{(S_1 - \gamma_{\xi}S_0)^2}{(1 - g)^2 S_0 D}$$

over $0 < \lambda < 1$ and $-\infty < t < \infty$. We assume that a is small to guarantee the existence of a positive maximum. Using Result 4.3, the problem reduces to minimizing

$$I_{\xi}(\lambda, t) = \frac{1}{AS_0} + \frac{B}{A^2 S_0 D} \quad (4.15)$$

over $0 < \lambda < 1$ and $t > 0$, where,

$$\begin{aligned} A &= 1 - \frac{a}{4t^2 \psi(t) \lambda (1 - \lambda)}, \\ B &= \psi^2(t) [(2\lambda - 1)t - \gamma_{\xi}]^2, \\ S_0 &= \psi(t), \end{aligned}$$

and

$$D = 4t^2 \psi^2(t) \lambda (1 - \lambda).$$

From $\partial I / \partial \lambda = 0$ and $\partial I / \partial t = 0$, we obtain

$$a'\lambda^2 + b'\lambda + c' = 0 \quad (4.16)$$

where

$$\begin{aligned} a' &= -4t^2 \psi'(t) \psi(t) \gamma_{\xi}, \\ b' &= 4t^2 \psi'(t) \psi(t) \gamma_{\xi} - 4t^2 \psi^2(t) - 2at \psi'(t), \end{aligned}$$

and

$$c' = 2t \psi^2(t) \gamma_{\xi} + 2t^2 \psi^2(t) + at \psi'(t).$$

Refer to Appendix V for details.

When $\gamma_{\xi} = 0$, $a' = 0$ and so (4.16) reduces to

$$b' \lambda + c' = 0$$

which implies that $\lambda = 1/2$. If we substitute $\gamma_{\xi} = 0$ and $\lambda = 1/2$ into (4.17), then we get

$$\psi'(t) + \frac{2a}{t^3} = 0,$$

as in Result 4.2(a). These results agree with those obtained in Sitter & Wu (1993).

When $\gamma_{\xi} \neq 0$, $a' \neq 0$ and the situation does not as readily simplify, so we must solve for λ using the quadratic formula. We get

$$\lambda_1(t) = \begin{cases} 1, & \text{if } g_1(t) \geq 1 \\ g_1(t), & \text{if } 0 < g_1(t) < 1 \\ 0, & \text{if } g_1(t) \leq 0 \end{cases} \quad \text{and} \quad \lambda_2(t) = \begin{cases} 1, & \text{if } g_2(t) \geq 1 \\ g_2(t), & \text{if } 0 < g_2(t) < 1, \\ 0, & \text{if } g_2(t) \leq 0 \end{cases} \quad (4.17)$$

where

$$g_1(t) = \frac{-b' - \sqrt{b'^2 - 4a'c'}}{2a'} \quad \text{and} \quad g_2(t) = \frac{-b' + \sqrt{b'^2 - 4a'c'}}{2a'}.$$

Solutions to (4.16) may not always be real. Without loss of generality, consider the case $t > 0$. When $t > 0$, $\psi(t) > 0$ and $\psi'(t) < 0$ for the logit model. Letting $k = at\psi'(t) + 2t^2\psi^2(t)$, we get

$$a' = -4t^2\psi'(t)\psi(t)\gamma,$$

$$b' = -a' - 2k,$$

and

$$c' = 2t\psi^2(t)\gamma + k.$$

So then,

$$\begin{aligned} b'^2 - 4a'c' &= (-a' - 2k)^2 - 4a'[2t\psi^2(t)\gamma + k] \\ &= a'^2 + 4k^2 + 4a'k - 8a't\psi^2(t)\gamma - 4a'k \\ &= 16t^4[\psi'(t)]^2\psi^2(t)\gamma^2 + 4k^2 + 32t^3\psi^3(t)\psi'(t)\gamma^2 \\ &= 16t^3\psi^2(t)\psi'(t)\gamma^2 [t\psi'(t) + 2\psi(t)] + 4k^2. \end{aligned}$$

Now, letting $r(t) = t\psi'(t) + 2\psi(t)$, we have $r(0) = 2\psi(0) = 0.5 > 0$. Since r is continuous, then by definition, we can find a $t^* > 0$, such that $r(t^*) > 0$. Using this t^* then implies the first term is negative for any ξ , while the second term is positive and independent of ξ . As $0 \leq \xi < \infty$, $-\infty \leq \gamma \leq \infty$. So if we use t^* we can choose ξ to make the first term as large a negative as we wish and thus drive $b'^2 - 4a'c'$ negative. Thus, solutions to (4.16) may not always be real. However, for moderate values of a and γ_ξ , as in Table 2, we can show numerically that $b'^2 - 4a'c' \geq 0$, and hence both solutions to (4.16) will be real.

Sitter & Wu (1993) also came up with a quadratic in λ when deriving formulae for A-optimality. They were able to show that both solutions to this quadratic were real, with only one of the solutions being in the unit interval. Here, the situation is more difficult as neither choice of λ_1 or λ_2 can be excluded. That is, there will be some cases where λ_1 will be optimal and others where λ_2 will be optimal.

Thus, the minimization in (4.15) reduces to a one variable minimization in t . To determine the F-optimal design, we minimize the functions $I_\xi(\lambda_1(t), t)$ and $I_\xi(\lambda_2(t), t)$ over $0 < t < \infty$, and then take the smaller of the two functions. So if $I_\xi(\lambda_1(t), t)$ is smaller, then the optimal design is $(t, \lambda_1(t))$, otherwise it is $(t, \lambda_2(t))$. That is, the optimal design is obtained by

$$\min \left\{ \min_t (I_\xi(\lambda_1(t), t), \min_t I_\xi(\lambda_2(t), t)) \right\}. \quad (4.18)$$

Notice that we have greatly reduced the problem to a one variable search for an optimal t . The minimization in (4.18) is usually performed by some computer algorithm. The downhill simplex algorithm, proposed by Nelder & Mead (1965) is one commonly used multidimensional minimization algorithm. We discuss this algorithm in section 4.3.

It is interesting to note that since $F(\cdot)$ is symmetric, the F-optimal design for estimating μ_ξ is the same as the F-optimal design for $\mu_{(100-\xi)}$, but with weights exchanged. To see this, note that by definition

$$\gamma_\xi = F^{-1}(\xi/100),$$

so by symmetry of F and applying F to both sides of the above, we get

$$1 - F(-\gamma_\xi) = F(\gamma_\xi) = \xi/100$$

or

$$F(-\gamma_\xi) = 1 - \xi/100$$

or

$$-\gamma_\xi = F^{-1}(1 - \xi/100) = \gamma_{(100-\xi)}. \quad (4.19)$$

From (4.15) and (4.19), we obtain

$$\begin{aligned} I_{100-\xi}(\lambda, t) &= \frac{1}{AS_0} + \frac{\psi^2(t)[(2\lambda-1)t - \gamma_{100-\xi}]^2}{A^2S_0D} \\ &= \frac{1}{AS_0} + \frac{\psi^2(t)[(2\lambda-1)t + \gamma_\xi]^2}{A^2S_0D} \\ &= \frac{1}{AS_0} + \frac{\psi^2(t)[-(2\lambda-1)t - \gamma_\xi]^2}{A^2S_0D} \\ &= \frac{1}{AS_0} + \frac{\psi^2(t)[(2(1-\lambda)-1)t - \gamma_\xi]^2}{A^2S_0D} \\ &= I_\xi(1-\lambda, t). \end{aligned}$$

Therefore,

$$\min_{0 < \lambda < 1, 0 < t < \infty} I_{100-\xi}(\lambda, t) = \min_{0 < \lambda < 1, 0 < t < \infty} I_\xi(1-\lambda, t).$$

That is, minimizing $I_{100-\xi}(\lambda, t)$ is equivalent to minimizing $I_\xi(1-\lambda, t)$. Thus, in obtaining the F -optimal designs numerically, we will only need to state designs for $\xi \geq 50$.

4.3 Numerical Results

In practice, optimal designs are obtained by numerical optimization algorithms. The most widely used algorithm is the downhill simplex method of Nelder and Mead (1965), which minimizes any multidimensional function. The Nelder-Mead algorithm may be applied to both minimization and maximization problems, since maximizing any given function $f(\cdot)$ is equivalent to minimizing $-f(\cdot)$.

The most attractive feature of the Nelder-Mead algorithm is that no derivative calculations are required. This fact alone reduces the number of assumptions made on the function being minimized. The method works best for functions of dimension less than about 10. The actual code is very concise, written in under 100 lines. The main drawback to the Nelder-Mead algorithm is that it requires a large number of function evaluations making it slower than other multidimensional optimization algorithms that involve derivative calculations. However, speed is not a problem for the size of problem we are considering and the speed of computers available.

We used the C code written by Press et al. (1988) in finding F-optimal designs for unbounded design variables. Using the characterization in section 4.2.2, we obtained F-optimal designs for the logit model, for various values of ξ , and we compared them to the c-optimal designs obtained by Wu (1988) for various values of n . To use the algorithm in the minimization of (4.18), we need only specify 1 design point, t where $-\infty < t < \infty$. If $t^* > 0$, then the optimal design will be either $\{(t^*, \lambda_1(t)), (-t^*, 1-\lambda_1(t))\}$ or $\{(t^*, \lambda_2(t)), (-t^*, 1-\lambda_2(t))\}$, where $\lambda_1(t)$ and $\lambda_2(t)$ are given in (4.17). A similar design is obtained when $t^* < 0$.

The resulting optimal designs are described by their response probabilities $p_i = F(t_i)$ and corresponding weights λ_i , $i=1,2$ for ease of interpretation. Hence, a 2-point design is written in the form $[p_1, p_2]$ with weights $(\lambda, 1-\lambda)$. Tables 1 and 2 give c- and F-optimal designs for the logit model. From Table 1, we see that for $8.3 < \xi < 91.7$, the c-optimal

design is a 1-point design at μ_{ξ} , and for ξ outside of this range, it is a 2-point design with mass λ at p_1 and $1-\lambda$ at p_2 . The F-optimal design in Table 2 are all 2-point designs for different values of ξ and $a = z_{\alpha}^2 / n$. The maximum allowable value of a that satisfies $g < 1$ for the logit model is $a = 0.40$. Below we compare the c- and F-optimal designs for different values of ξ and n .

When $\xi = 50$, the c-optimal design is a 1-point design, with all of its weight at μ_{ξ} . On the other hand, the F-optimal design is a 2-point design, with weights equally split between the two design points. These 2-point designs are exactly the same as those obtained in Sitter & Wu (1993) for the ED50. Notice that as n increases, the 2-point designs have design points that are nearer to μ_{ξ} , but with weights still evenly split.

For $91.7 < \xi \leq 99.8$, the c-optimal design is a 2-point design that has fixed design points with more weight at p_1 . As the percentiles become more extreme, the weights distribute more evenly around p_1 and p_2 . F-optimal designs are more similar to c-optimal designs for percentiles in this range. When n is large ($a=0.01, 0.05, 0.10, 0.20$), we see that the F-optimal design tends toward the c-optimal design. However for moderate n , it is only the design points that tend toward the c-optimal design points. For small n (large a), the weights become more evenly split between the design points. When $a = 0.40$, the F-optimal design has fixed design points, with weights about evenly split. In this case, the design is more similar to a D-optimal than a c-optimal design.

The most important differences between the c- and F-optimality criteria are evident when $8.3 < \xi < 91.7$, for different values of n . For percentiles in this range, c-optimality yields a 1-point design, with all of the weight at μ_{ξ} . This is different from the F-optimality criterion, which gives a 2-point design with more weight at p_1 . Notice that as n increases, the 2-point designs have design points nearer to μ_{ξ} , with increasing weight at p_1 . For moderate n , we see that the weights are nearly evenly split between the two design points.

When using either of the designs in Tables 1 and 2 as benchmarks, it is important to keep in mind that each gives a different impression on how to design the experiment. For

moderate n and percentiles between 8.3 and 91.7, the differences are more prominent. The c -optimality criterion suggests taking points near μ_ξ , while the F -optimality criterion suggests a 2-point design with points symmetrically placed about the ED50.

Chapter 5

Summary

When estimation of the $ED\xi$, the dose causing a $\xi\%$ response, is of particular interest, either the c- or the F-optimality criterion may be used for optimal design. We found that estimating the $ED\xi$ is equivalent to estimating μ_ξ , where μ_ξ is defined in model (2.4).

In chapter 3, we derived two competing intervals for estimating μ_ξ , the asymptotic confidence interval and the Fieller interval. From the results in Sitter & Wu (1993), it is reasonable to anticipate that a Fieller interval would give more efficient estimates of μ_ξ than the asymptotic confidence interval, particularly for small n . This is important since many experiments use a finite sample size n . Thus, making inferences from the asymptotic confidence interval which is only valid for very large n , may lead to misleading conclusions when n is small to moderate. This led us to consider the Fieller interval as a design criteria.

In chapter 4, we derived F- and c-optimal designs for estimating μ_ξ . The method of Sitter & Wu (1993) could not be used for $\xi \neq 50$. So instead, we used a geometric approach that led to Result 4.3. Using this result, we showed that the F-optimal design for estimating μ_ξ , for any $0 < \xi < 100$, would be a 2-point design for the logit and probit models. This simplified the design problem to a 2 variable search in t and λ . We were able to use our characterizations of the F-optimal design to further reduce the search to a one variable search in t .

The Nelder-Mead downhill simplex algorithm was used to numerically obtain the F-optimal designs in Table 4 for the logit model, for various values of ξ and n . These designs were compared to the c-optimal designs in Table 3. The results indicate that for large n , and for extreme percentiles with moderate n , the c- and F-optimal designs are similar. The most pronounced difference occurred for moderate n when $8.3 < \xi < 91.7$. In this case, c-optimality gave a 1-point design with all of the weight at μ_ξ , whereas F-optimality gave a 2-point design with weights nearly evenly distributed between the two points.

Optimal designs have two main disadvantages. First of all, they are only optimal for a best guess of the parameters μ_ξ and β . This is a problem because designs that are optimal for a best guess may not be efficient if these guesses are incorrect. Thus optimal designs are not robust to poor initial estimates of μ_ξ and β . And secondly, these designs yield too few design points to allow for model checking. Despite these drawbacks, optimal designs are still important when used as benchmarks for more heuristically chosen designs or to choose a second stage design in a two-stage procedure. Our results suggest that, though c- and F-optimality are criterion for choosing designs which have the same basic goal in mind, when c- and F-optimal designs are used as a benchmark to guide the choice of a heuristically chosen design, they can give very different intuition.

The area of optimal design for non-linear models is an expanding field. Further research still needs to be done. We end this chapter with two open problems in optimal design:

- (1) F-optimal designs for a bounded design space;

and

- (2) Finding F- and c-optimal designs for asymmetric models like the power logit and power probit models and comparing the efficiencies of these with those based on symmetric models.

Table 1. c-Optimal designs for estimating percentiles, $F = \text{logit}$.

For $0.083 < \xi/100 < 0.917$, the c-optimal design is a 1-point design at μ_ξ .

Percentiles	Design points [$p_1=F(-t)$, $p_2=F(t)$]	Weights (λ , $1-\lambda$)
0.92	(0.917, 0.083)	(0.99, 0.01)
0.93	(0.917, 0.083)	(0.96, 0.04)
0.94	(0.917, 0.083)	(0.94, 0.06)
0.95	(0.917, 0.083)	(0.91, 0.09)
0.96	(0.917, 0.083)	(0.88, 0.12)
0.97	(0.917, 0.083)	(0.85, 0.15)
0.98	(0.917, 0.083)	(0.81, 0.19)
0.99	(0.917, 0.083)	(0.76, 0.24)
0.994	(0.917, 0.083)	(0.73, 0.27)
0.998	(0.917, 0.083)	(0.69, 0.31)

Table 2. F-optimal designs for estimating μ_ξ , $F = \text{logit}$

$a = z_\alpha^2/n$	Percentiles $\xi/100$	Design points	Weights
		$[p_1=F(-t), p_2=F(t)]$	$(\lambda, 1-\lambda)$
0.01	0.50	(0.658, 0.342)	(0.50, 0.50)
	0.60	(0.683, 0.323)	(0.74, 0.26)
	0.70	(0.739, 0.261)	(0.86, 0.14)
	0.80	(0.812, 0.189)	(0.91, 0.09)
	0.90	(0.884, 0.115)	(0.91, 0.09)
	0.92	(0.896, 0.104)	(0.90, 0.10)
	0.93	(0.901, 0.099)	(0.89, 0.11)
	0.94	(0.906, 0.094)	(0.88, 0.12)
	0.95	(0.909, 0.091)	(0.86, 0.14)
	0.96	(0.912, 0.089)	(0.84, 0.16)
	0.97	(0.913, 0.087)	(0.82, 0.18)
	0.98	(0.915, 0.086)	(0.79, 0.21)
	0.99	(0.915, 0.084)	(0.75, 0.25)
	0.994	(0.916, 0.084)	(0.72, 0.28)
	0.998	(0.916, 0.084)	(0.68, 0.32)
0.05	0.50	(0.737, 0.264)	(0.50, 0.50)
	0.60	(0.749, 0.250)	(0.65, 0.35)
	0.70	(0.785, 0.215)	(0.75, 0.25)
	0.80	(0.840, 0.163)	(0.81, 0.19)
	0.90	(0.884, 0.116)	(0.82, 0.18)
	0.92	(0.890, 0.092)	(0.80, 0.20)
	0.93	(0.897, 0.103)	(0.80, 0.20)
	0.94	(0.900, 0.100)	(0.79, 0.21)
	0.95	(0.903, 0.097)	(0.78, 0.22)
	0.96	(0.906, 0.093)	(0.77, 0.23)
	0.97	(0.909, 0.091)	(0.75, 0.25)
	0.98	(0.911, 0.089)	(0.73, 0.27)
	0.99	(0.913, 0.087)	(0.70, 0.30)
	0.994	(0.914, 0.086)	(0.68, 0.32)
	0.998	(0.915, 0.085)	(0.65, 0.35)
0.10	0.50	(0.782, 0.218)	(0.50, 0.50)
	0.60	(0.791, 0.209)	(0.61, 0.39)
	0.70	(0.817, 0.183)	(0.70, 0.30)
	0.80	(0.853, 0.147)	(0.75, 0.25)
	0.90	(0.891, 0.109)	(0.75, 0.25)
	0.92	(0.897, 0.103)	(0.74, 0.26)

	0.93	(0.900, 0.100)	(0.74, 0.26)
	0.94	(0.902, 0.098)	(0.73, 0.27)
	0.95	(0.905, 0.095)	(0.72, 0.28)
	0.96	(0.907, 0.093)	(0.71, 0.29)
	0.97	(0.909, 0.091)	(0.70, 0.30)
	0.98	(0.911, 0.089)	(0.68, 0.32)
	0.99	(0.913, 0.087)	(0.66, 0.24)
	0.994	(0.914, 0.086)	(0.64, 0.36)
	0.998	(0.915, 0.085)	(0.62, 0.38)
0.20	0.50	(0.836, 0.163)	(0.50, 0.50)
	0.60	(0.842, 0.157)	(0.58, 0.42)
	0.70	(0.860, 0.141)	(0.64, 0.36)
	0.80	(0.883, 0.117)	(0.67, 0.33)
	0.90	(0.903, 0.097)	(0.66, 0.34)
	0.92	(0.906, 0.094)	(0.65, 0.35)
	0.93	(0.908, 0.092)	(0.65, 0.35)
	0.94	(0.909, 0.091)	(0.64, 0.36)
	0.95	(0.910, 0.090)	(0.64, 0.36)
	0.96	(0.911, 0.089)	(0.63, 0.37)
	0.97	(0.912, 0.088)	(0.62, 0.38)
	0.98	(0.913, 0.087)	(0.61, 0.39)
	0.99	(0.914, 0.086)	(0.60, 0.40)
	0.994	(0.915, 0.085)	(0.60, 0.40)
	0.998	(0.916, 0.084)	(0.60, 0.40)
0.30	0.50	(0.874, 0.126)	(0.50, 0.50)
	0.60	(0.879, 0.121)	(0.56, 0.44)
	0.70	(0.890, 0.109)	(0.60, 0.40)
	0.80	(0.904, 0.096)	(0.61, 0.39)
	0.90	(0.912, 0.088)	(0.59, 0.41)
	0.92	(0.913, 0.087)	(0.58, 0.42)
	0.93	(0.914, 0.086)	(0.58, 0.42)
	0.94	(0.914, 0.086)	(0.58, 0.42)
	0.95	(0.914, 0.085)	(0.57, 0.43)
	0.96	(0.915, 0.085)	(0.56, 0.44)
	0.97	(0.915, 0.085)	(0.57, 0.43)
	0.98	(0.916, 0.084)	(0.56, 0.44)
	0.99	(0.916, 0.084)	(0.55, 0.45)
	0.994	(0.916, 0.084)	(0.54, 0.46)
	0.998	(0.916, 0.084)	(0.54, 0.46)
0.40	0.50	(0.905, 0.095)	(0.50, 0.50)
	0.60	(0.909, 0.091)	(0.54, 0.46)
	0.70	(0.909, 0.091)	(0.54, 0.46)

0.80	(0.916, 0.084)	(0.54, 0.46)
0.90	(0.916, 0.084)	(0.52, 0.48)
0.92	(0.917, 0.083)	(0.52, 0.48)
0.93	(0.917, 0.083)	(0.52, 0.48)
0.94	(0.917, 0.083)	(0.52, 0.48)
0.95	(0.917, 0.083)	(0.52, 0.48)
0.96	(0.917, 0.083)	(0.52, 0.48)
0.97	(0.917, 0.083)	(0.52, 0.48)
0.98	(0.917, 0.083)	(0.51, 0.49)
0.99	(0.917, 0.083)	(0.51, 0.49)
0.994	(0.917, 0.083)	(0.51, 0.49)
0.998	(0.917, 0.083)	(0.51, 0.49)

Figure 4.1: A plot of K , $-K$, and K^* for the logit model

Figure 4.2: A plot of $\mathcal{H}_2$ for the logit model. The dashed line completes the convex hull.

Figure 4.3: A plot of $\mathcal{CH}_3$ for the logit model. The line $\mathcal{H}_3$ is $(u,v,w) = (\psi(t), t\psi(t), t^2\psi(t))$.

Appendix I

1. Proof of $J(\underline{\theta})$ in (3.1)

2. Proof of (3.5), (3.6) and (3.7)

1. Proof of $J(\underline{\theta})$ in (3.1)

The Expected Fisher information matrix for $\underline{\theta} = (\mu_x, \beta)^T$ is defined by

$$J(\underline{\theta}) = \begin{pmatrix} E\left(-\frac{\partial^2}{\partial \mu_x^2} L(\underline{t})\right) & E\left(-\frac{\partial^2}{\partial \beta \partial \mu_x} L(\underline{t})\right) \\ E\left(-\frac{\partial^2}{\partial \beta \partial \mu_x} L(\underline{t})\right) & E\left(-\frac{\partial^2}{\partial \beta^2} L(\underline{t})\right) \end{pmatrix} = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}, \text{ say,} \quad (I.1)$$

where $\underline{t} = (t_1, \dots, t_k)$, $t_i = \beta(x_i - \mu_x) + \gamma_x$ and

$$L(\underline{t}) = L(\underline{\theta} | \underline{t}) = \sum_{i=1}^k \left\{ r_i \log F[\beta(x_i - \mu_x) + \gamma_x] + (n_i - r_i) \log [1 - F[\beta(x_i - \mu_x) + \gamma_x]] \right\}$$

From here on, we will suppress the subscripts on μ and γ for simplicity. For a single observation t , we have

$$L(\underline{\theta} | t) = r \log F(t) + (n-r) \log [1 - F(t)]. \quad (I.2)$$

From $t = \beta(x - \mu) + \gamma$, we obtain

$$\frac{\partial t}{\partial \mu} = -\beta \quad \text{and} \quad \frac{\partial t}{\partial \beta} = x - \mu. \quad (I.3)$$

Partial derivatives in (I.1) are now calculated using the chain rule. We have,

$$\begin{aligned}
\frac{\partial^2}{\partial \mu^2} L(t) &= \frac{\partial}{\partial \mu} \left(\frac{\partial}{\partial \mu} L(t) \right) \\
&= \frac{\partial}{\partial \mu} \left(\frac{\partial}{\partial t} L(t) \cdot \frac{\partial t}{\partial \mu} \right) \\
&= \frac{\partial}{\partial t} \left(\frac{\partial}{\partial t} L(t) \cdot \frac{\partial t}{\partial \mu} \right) \cdot \frac{\partial t}{\partial \mu} \\
&= \frac{\partial^2}{\partial t^2} L(t) \cdot \left(\frac{\partial t}{\partial \mu} \right)^2 \\
&= \beta^2 \frac{\partial^2}{\partial t^2} L(t),
\end{aligned} \tag{I.4}$$

by (I.3). Similarly,

$$\begin{aligned}
\frac{\partial^2}{\partial \beta \partial \mu} L(t) &= \frac{\partial^2}{\partial t^2} L(t) \cdot \left(\frac{\partial t}{\partial \mu} \right) \left(\frac{\partial t}{\partial \beta} \right) \\
&= -\beta(x-\mu) \frac{\partial^2}{\partial t^2} L(t) = \frac{\partial^2}{\partial \mu \partial \beta} L(t),
\end{aligned} \tag{I.5}$$

by (I.3) and continuity of derivatives, and

$$\begin{aligned}
\frac{\partial^2}{\partial \beta^2} L(t) &= \frac{\partial^2}{\partial t^2} L(t) \cdot \left(\frac{\partial t}{\partial \beta} \right)^2 \\
&= (x-\mu)^2 \frac{\partial^2}{\partial t^2} L(t).
\end{aligned} \tag{I.6}$$

From (I.2), we obtain

$$\begin{aligned}
\frac{\partial}{\partial t} L(t) &= \frac{rf(t)}{F(t)} + \frac{(r-n)f(t)}{1-F(t)} \\
&= \frac{rf(t)[1-F(t)] + (r-n)f(t)F(t)}{F(t)[1-F(t)]} \\
&= \frac{rf(t) - rf(t)F(t) - nf(t)F(t) + rf(t)F(t)}{F(t)[1-F(t)]} \\
&= \frac{f(t)[r - nF(t)]}{F(t)[1-F(t)]}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2}{\partial t^2} L(t) &= \frac{[f'(t)(r - nF(t)) - nf^2(t)]F(t)[1-F(t)] - f(t)(r - nF(t))(f(t)[1-F(t)] - f(t)F(t))}{F(t)[1-F(t)]^2} \\
&= \frac{[r - nF(t)][f'(t)F(t)[1-F(t)] - f^2(t)[1-F(t)] + f^2(t)F(t)] - nf^2(t)F(t)[1-F(t)]}{(F(t)[1-F(t)])^2},
\end{aligned}$$

where $f(t) = \frac{\partial}{\partial t} F(t)$. Since $x \sim \text{bin}(n, F(t))$, then

$$\begin{aligned}
E[r - nF(t)] &= E[r] - nF(t) \\
&= nF(t) - nF(t) \\
&= 0
\end{aligned}$$

and so

$$\begin{aligned}
E\left(-\frac{\partial^2}{\partial t^2} L(t)\right) &= -E\left(\frac{\partial^2}{\partial t^2} L(t)\right) \\
&= \frac{nf^2(t)}{F(t)[1-F(t)]} \\
&= n\psi(t)
\end{aligned} \tag{1.7}$$

where

$$\psi(t) = \frac{f^2(t)}{F(t)[1-F(t)]}. \tag{1.8}$$

Now, let $S_j = \sum_{i=1}^k \lambda_i t_i^j \psi(t_i)$, $j = 0, 1, 2$, $\lambda_i = n_i/n$, and $n = \sum n_i$. Since

$$\frac{\partial^2}{\partial \mu^2} L(t) = \sum_{i=1}^k \frac{\partial^2}{\partial \mu^2} L(t_i) \quad \text{and} \quad \frac{\partial^2}{\partial \beta^2} L(t) = \sum_{i=1}^k \frac{\partial^2}{\partial \beta^2} L(t_i), \quad (I.9)$$

it follows from (I.4)-(I.9) that

$$\begin{aligned} J_{11} &= \beta^2 \sum_{i=1}^k n_i \psi(t_i) \\ &= n \beta^2 \sum_{i=1}^k \lambda_i \psi(t_i) \\ &= n \beta^2 S_0, \end{aligned} \quad (I.10)$$

$$\begin{aligned} J_{12} &= -\sum_{i=1}^k n_i \beta (x_i - \mu) \psi(t_i) \\ &= -n \sum_{i=1}^k \lambda_i (t_i - \gamma) \psi(t_i) \\ &= -n \left[\sum_{i=1}^k \lambda_i t_i \psi(t_i) - \gamma \sum_{i=1}^k \lambda_i \psi(t_i) \right] \\ &= -n(S_1 - \gamma S_0), \end{aligned} \quad (I.11)$$

and

$$\begin{aligned} J_{22} &= \sum_{i=1}^k n_i (x_i - \mu)^2 \psi(t_i) \\ &= \frac{n}{\beta^2} \sum_{i=1}^k \lambda_i (\beta(x_i - \mu))^2 \psi(t_i) \\ &= \frac{n}{\beta^2} \sum_{i=1}^k \lambda_i (t_i - \gamma)^2 \psi(t_i) \\ &= \frac{n}{\beta^2} \left(\sum_{i=1}^k \lambda_i t_i^2 \psi(t_i) - 2\gamma \sum_{i=1}^k \lambda_i t_i \psi(t_i) + \gamma^2 \sum_{i=1}^k \lambda_i \psi(t_i) \right) \\ &= \frac{n}{\beta^2} (S_2 - 2\gamma S_1 + \gamma^2 S_0). \end{aligned} \quad (I.12)$$

On substituting (I.10)-(I.12) into (I.1), we obtain (3.1).

2. Proof of (3.5), (3.6) and (3.7)

Let

$$g_1(\hat{\theta}) = \hat{\beta}, \quad g_2(\hat{\theta}) = \hat{\mu}_\xi \hat{\beta}, \quad \text{and} \quad g_3(\hat{\theta}) = \hat{\mu}_\xi \hat{\beta} + \hat{\beta}.$$

Then,

$$g_1(\underline{\theta}) = \beta, \quad g_2(\underline{\theta}) = \mu_\xi \beta, \quad \text{and} \quad g_3(\underline{\theta}) = \mu_\xi \beta + \beta.$$

Letting $\underline{c}_i = (\partial g_i / \partial \mu, \partial g_i / \partial \beta)^T$, $i = 1, \dots, 3$, we get

$$\underline{c}_1 = (0, 1)^T, \quad \underline{c}_2 = (\beta, \mu_\xi)^T, \quad \underline{c}_3 = (\beta, \mu_\xi + 1)^T.$$

From Result 3.1 (ii) and (3.2), we obtain

$$\begin{aligned} AV(\hat{\beta}) &= \underline{c}_1^T J^{-1}(\underline{\theta}) \underline{c}_1 \\ &= \frac{1}{n} (0, 1) \begin{pmatrix} I^{11} & I^{12} \\ I^{12} & I^{22} \end{pmatrix} (0, 1)^T \\ &= \frac{I^{22}}{n} \\ &= \frac{\beta^2}{nD} S_0, \end{aligned}$$

and thus (3.5) holds;

$$\begin{aligned} AV(\hat{\mu}_\xi \hat{\beta}) &= \underline{c}_2^T J^{-1}(\underline{\theta}) \underline{c}_2 \\ &= \frac{1}{n} (\beta, \mu_\xi) \begin{pmatrix} I^{11} & I^{12} \\ I^{12} & I^{22} \end{pmatrix} (\beta, \mu_\xi)^T \\ &= \frac{1}{n} (\beta^2 I^{11} + 2\mu_\xi \beta I^{12} + \mu_\xi^2 I^{22}) \\ &= (3.6), \end{aligned}$$

and

$$\begin{aligned} AV(\hat{\mu}_\xi \hat{\beta} + \hat{\beta}) &= \underline{c}_3^T J^{-1}(\underline{\theta}) \underline{c}_3 \\ &= \frac{1}{n} (\beta^2 I^{11} + 2\beta(\hat{\mu}_\xi + 1)I^{12} + (\hat{\mu}_\xi + 1)^2 I^{22}) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} ((\beta^2 I^{11} + 2\hat{\mu}_\xi \beta I^{12} + \mu_\xi^2 I^{22}) + 2\beta I^{12} + I^{22} + 2\hat{\mu}_\xi I^{22}) \\
&= AV(\hat{\mu}_\xi \hat{\beta}) + \frac{1}{n} (2\beta I^{12} + I^{22} + 2\hat{\mu}_\xi I^{22}).
\end{aligned}$$

It is well established that, for any two random variables X and Y,

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).$$

(I.13)

Letting $X = \hat{\mu}_\xi \hat{\beta}$ and $Y = \hat{\beta}$ in the third term of (I.13), and using asymptotic variances and covariances yields

$$\begin{aligned}
-ACov(\hat{\beta} \hat{\mu}_\xi, \hat{\beta}) &= -\frac{1}{2} (AV(\hat{\mu}_\xi \hat{\beta} + \hat{\beta}) - AV(\hat{\mu}_\xi \hat{\beta}) - AV(\hat{\beta})) \\
&= -\frac{1}{2} (AV(\hat{\mu}_\xi \hat{\beta}) + \frac{1}{n} (2\beta I^{12} + I^{22} + 2\hat{\mu}_\xi I^{22}) - AV(\hat{\mu}_\xi \hat{\beta}) - I^{22}) \\
&= -\frac{1}{n} (\beta I^{12} + \hat{\mu}_\xi I^{22}),
\end{aligned}$$

and thus (3.7) holds.

Appendix II

Derivation of (3.9) and (3.10)

We derive (3.10) and then (3.9). We will suppress the subscripts on μ and γ for simplicity, and note that $\hat{\mu}_\xi$ and $\hat{\beta}$ are being used in the definitions of S_0 , S_1 , and S_2 . From (3.8), we have

$$\begin{aligned}
 M &= \frac{B}{2A} \\
 M &= \frac{\hat{\mu}\hat{\beta}^2 - z_\alpha^2(\hat{\beta}\hat{I}^{12} + \hat{\mu}\hat{I}^{22})/n}{\hat{\beta}^2 - z_\alpha^2\hat{I}^{22}/n} \\
 &= \frac{\hat{\mu}(\hat{\beta}^2 - z_\alpha^2\hat{I}^{22}/n) - \hat{\beta}z_\alpha^2\hat{I}^{12}/n}{\hat{\beta}^2 - z_\alpha^2\hat{I}^{22}/n} \\
 &= \hat{\mu} - \frac{\hat{\beta}z_\alpha^2\hat{I}^{12}/n}{\hat{\beta}^2 - z_\alpha^2\hat{I}^{22}/n} \\
 &= \hat{\mu} - \frac{\hat{\beta}z_\alpha^2(S_1 - \gamma S_0)/(nD)}{\hat{\beta}^2 - z_\alpha^2\hat{\beta}^2 S_0/(nD)} \\
 &= \hat{\mu} - \frac{(S_1 - \gamma S_0)}{\hat{\beta}S_0(1-g)} \frac{z_\alpha^2 S_0}{nD} \\
 &= \hat{\mu} - c_1 \frac{g}{1-g}, \tag{II.1}
 \end{aligned}$$

where

$$c_1 = \frac{(S_1 - \gamma S_0)}{\hat{\beta}S_0},$$

and

$$g = z_\alpha^2 S_0 / nD.$$

Also,

$$\begin{aligned}
N &= \frac{\sqrt{B^2 - 4AC}}{2A} \\
&= \frac{(\hat{\beta}^2 z_\alpha^4 (\hat{I}^{12} / n)^2 + \hat{\beta}^4 z_\alpha^2 \hat{I}^{11} / n - \hat{\beta}^2 z_\alpha^4 \hat{I}^{11} \hat{I}^{22} / n^2)^{1/2}}{\hat{\beta}^2 - z_\alpha^2 \hat{I}^{22} / n} \\
&= \frac{(\hat{\beta}^2 z_\alpha^4 (\hat{I}^{12} / n)^2 + \hat{\beta}^4 z_\alpha^2 \hat{I}^{11} / n - \hat{\beta}^2 z_\alpha^4 \hat{I}^{11} \hat{I}^{22} / n^2)^{1/2}}{\hat{\beta}^2 (1-g)} \\
&= \frac{1}{1-g} \left(\frac{1}{\hat{\beta}^4} (\hat{\beta}^2 z_\alpha^4 (\hat{I}^{12} / n)^2 + \hat{\beta}^4 z_\alpha^2 \hat{I}^{11} / n - \hat{\beta}^2 z_\alpha^4 \hat{I}^{11} \hat{I}^{22} / n^2) \right)^{1/2} \\
&= \frac{z_\alpha}{1-g} \left(\frac{z_\alpha^2 (S_1 - \gamma S_o)^2}{\hat{\beta}^2 n^2 D^2} + \frac{(S_2 - 2\gamma S_1 + \gamma^2 S_o)}{\hat{\beta}^2 n D} - \frac{z_\alpha^2 S_o (S_2 - 2\gamma S_1 + \gamma^2 S_o)}{\hat{\beta}^2 n^2 D^2} \right)^{1/2} \\
&= \frac{z_\alpha}{1-g} \sqrt{A \hat{V}(\hat{\mu})} \left(1 - \frac{z_\alpha^2 S_o}{n D} + \frac{z_\alpha^2 (S_1 - \gamma S_o)^2}{n D (S_2 - 2\gamma S_1 + \gamma^2 S_o)} \right)^{1/2} \\
&= \frac{z_\alpha}{1-g} \sqrt{A \hat{V}(\hat{\mu})} \left(1 - \frac{D}{S_o (S_2 - 2\gamma S_1 + \gamma^2 S_o)} \frac{z_\alpha^2 S_o}{n D} \right)^{1/2} \\
&= \frac{z_\alpha}{1-g} \sqrt{A \hat{V}(\hat{\mu})} (1 - c_2 g)^{1/2}, \tag{II.2}
\end{aligned}$$

where

$$c_2 = \frac{D}{S_o (S_2 - 2\gamma S_1 + \gamma^2 S_o)}.$$

Substituting (II.1) and (II.2) into (3.8) yields the Fieller interval in (3.10).

The interval in (3.9) is obtained by algebraically manipulating (II.2). We can rewrite (II.2) as

$$\begin{aligned}
 N &= \frac{z_\alpha}{1-g} \sqrt{A \hat{V}(\hat{\mu})} \left(1 - \frac{gD}{S_o(S_2 - 2\gamma S_1 + \gamma^2 S_o)} \right)^{1/2} \\
 &= \frac{z_\alpha}{1-g} \sqrt{A \hat{V}(\hat{\mu})} \left(\frac{S_o(S_2 - 2\gamma S_1 + \gamma^2 S_o) - g(S_o S_2 - S_1^2)}{S_o(S_2 - 2\gamma S_1 + \gamma^2 S_o)} \right)^{1/2} \\
 &= \frac{z_\alpha}{1-g} \sqrt{A \hat{V}(\hat{\mu})} \left(\frac{S_o S_2(1-g) - 2\gamma S_o S_1 + \gamma^2 S_o^2 + g S_1^2}{S_o(S_2 - 2\gamma S_1 + \gamma^2 S_o)} \right)^{1/2} \\
 &= \frac{z_\alpha}{1-g} \sqrt{A \hat{V}(\hat{\mu})} \left(\frac{D(1-g) + (S_1 - \gamma S_o)^2}{S_o(S_2 - 2\gamma S_1 + \gamma^2 S_o)} \right)^{1/2} \\
 &= z_\alpha \sqrt{\frac{(S_2 - 2\gamma S_1 + \gamma^2 S_o)}{\beta^2 n D}} \left(\frac{D}{S_o(1-g)(S_2 - 2\gamma S_1 + \gamma^2 S_o)} + \frac{(S_1 - \gamma S_o)^2}{S_o(1-g)^2(S_2 - 2\gamma S_1 + \gamma^2 S_o)} \right)^{1/2} \\
 &= \frac{z_\alpha}{\beta \sqrt{n}} \left(\frac{1}{S_o(1-g)} + \frac{(S_1 - \gamma S_o)^2}{S_o D(1-g)^2} \right)^{1/2} \\
 &= \frac{z_\alpha}{\beta \sqrt{n}} \left(\frac{1}{S_o(1-aS_o/D)} + \frac{(S_1 - \gamma S_o)^2}{S_o D(1-g)^2} \right)^{1/2}, \tag{II.3}
 \end{aligned}$$

where $a = z_\alpha^2 / n$. Substitution of (II.1) and (II.3) into (3.8) gives the Fieller interval in (3.9).

Appendix III

Proof of (4.2) and (4.7)

1. Proof of (4.2)

In (4.1), we minimize

$$\underline{c}^T J^{-1}(\underline{\theta}) \underline{c} = \frac{1}{nD\beta^2} (S_2 - 2\gamma_\xi S_1 + \gamma_\xi^2 S_0)$$

over all choices of $0 < \lambda_i \leq 1$ and $-\infty < t_i < \infty$. This is equivalent to minimizing

$$\begin{aligned} Q &= \frac{1}{D} (S_2 - 2\gamma_\xi S_1 + \gamma_\xi^2 S_0) \\ &= (1, \gamma_\xi) \begin{pmatrix} S_0 & S_1 \\ S_1 & S_2 \end{pmatrix}^{-1} (1, \gamma_\xi)^T \\ &= (1, \gamma_\xi) \left\{ \sum_{i=1}^k \lambda_i \psi(t_i) \begin{pmatrix} 1 & t_i \\ t_i & t_i^2 \end{pmatrix} \right\}^{-1} (1, \gamma_\xi)^T \\ &= (1, \gamma_\xi) \left\{ \sum_{i=1}^k \lambda_i \psi(t_i) (1, t_i)(1, t_i)^T \right\}^{-1} (1, \gamma_\xi)^T. \end{aligned}$$

2. Proof of (4.7)

To determine the weights λ and $1-\lambda$ for the two-point design, the problem reduces to minimizing

$$Q = \frac{1}{D} (S_2 - 2\gamma_\xi S_1 + \gamma_\xi^2 S_0)$$

where

$$S_0 = \sum_{i=1}^k \lambda_i \psi(t_i) = \psi(t), \quad (\text{III.1a})$$

$$S_1 = \sum_{i=1}^k \lambda_i t_i \psi(t_i) = t\psi(t)(2\lambda - 1), \quad (\text{III.1b})$$

$$S_2 = \sum_{i=1}^k \lambda_i t_i^2 \psi(t_i) = t^2 \psi(t), \quad (\text{III.1c})$$

and

$$D = S_0 S_2 - S_1^2 = t^2 \psi^2(t) 4\lambda(1-\lambda). \quad (\text{III.1d})$$

From (III.1a-d), we obtain

$$\frac{\partial S_0}{\partial \lambda} = 0, \quad \frac{\partial S_1}{\partial \lambda} = 2t\psi(t), \quad \frac{\partial S_2}{\partial \lambda} = 0,$$

and
$$\frac{\partial D}{\partial \lambda} = 4t^2 \psi^2(t)(1-2\lambda).$$

From $\partial Q / \partial \lambda = 0$, and the above, we obtain

$$-\frac{\partial D}{\partial \lambda}(S_2 - 2\gamma_\xi S_1 + \gamma_\xi^2 S_0) = \frac{\partial S_1}{\partial \lambda}(2\gamma_\xi D)$$

or

$$\lambda^2(-4t\gamma_\xi) + 2\lambda(t + \gamma_\xi)^2 - (t + \gamma_\xi)^2 = 0. \quad (\text{III.2})$$

We can rewrite (III.2) as

$$a\lambda^2 + b\lambda + c = 0, \quad (\text{III.3})$$

where $a = -4t\gamma_\xi$, $b = 2(t + \gamma_\xi)^2$, and $c = -(t + \gamma_\xi)^2$. So

$$\begin{aligned} \lambda &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\ &= \frac{-2(t + \gamma_\xi)^2 \pm \sqrt{4(t + \gamma_\xi)^4 - 16t\gamma_\xi(t + \gamma_\xi)^2}}{-8t\gamma_\xi} \\ &= \frac{(t + \gamma_\xi)^2}{4t\gamma_\xi} \pm \frac{(t + \gamma_\xi)}{-4t\gamma_\xi} \sqrt{(t - \gamma_\xi)^2} \\ &= \frac{(t + \gamma_\xi)^2}{4t\gamma_\xi} \pm \frac{(t + \gamma_\xi)(t - \gamma_\xi)}{-4t\gamma_\xi}. \end{aligned}$$

Thus,

$$\begin{aligned} \lambda_1 &= \frac{t^2 + 2t\gamma_\xi + \gamma_\xi^2 - (t^2 - \gamma_\xi^2)}{4t\gamma_\xi} \\ &= \frac{1}{2} + \frac{1}{2} \frac{\gamma_\xi}{t}, \end{aligned}$$

and

$$\lambda_2 = \frac{t^2 + 2t\gamma_\xi + \gamma_\xi^2 + (t^2 - \gamma_\xi^2)}{4t\gamma_\xi}$$

$$= \frac{1}{2} + \frac{1}{2} \frac{t}{\gamma_\xi}.$$

Since, $p > p_2$ and $F(t)$ is monotonic, then $\gamma_\xi = F^{-1}(p) > F^{-1}(p_2) = t^*$. Therefore $\lambda_1 \notin (0,1)$

and so the only solution to (III.3) is

$$\lambda_2 = \frac{1}{2} + \frac{1}{2} \frac{t}{\gamma_\xi}.$$

Appendix IV

Proof of (4.9)

Proof of (4.9)

The problem reduces to showing that

$$\frac{1}{(1-aS_0/D)S_0} = \frac{1}{(1-a/S_2)S_0} + \frac{aS_1^2}{D(1-g)(1-a/S_2)S_0S_2}. \quad (\text{IV.1})$$

The righthand side of (IV.1) becomes

$$\begin{aligned} & \frac{1}{(1-a/S_2)S_0} + \frac{1}{(1-aS_0/D)S_0} \cdot \frac{aS_1^2}{S_2D(1-a/S_2)} \\ &= \frac{1}{(1-aS_0/D)S_0} \left[\frac{aS_1^2}{S_2D(1-a/S_2)} + \frac{(1-aS_0/D)}{(1-a/S_2)} \right] \\ &= \frac{1}{(1-aS_0/D)S_0} \left[\frac{aS_1^2}{S_2D(1-a/S_2)} + \frac{1}{(1-a/S_2)} \frac{S_2(D-aS_0)}{S_2D} \right] \\ &= \frac{1}{(1-aS_0/D)S_0} \left[\frac{aS_1^2 + S_2(D-aS_0)}{S_2D(1-a/S_2)} \right] \\ &= \frac{1}{(1-aS_0/D)S_0} \left[\frac{a(S_0S_2-D) + S_2D - aS_0S_2}{S_2D(1-a/S_2)} \right] \\ &= \frac{1}{(1-aS_0/D)S_0} \left[\frac{-aD + S_2D}{S_2D(1-a/S_2)} \right] \\ &= \frac{1}{(1-aS_0/D)S_0} \left[\frac{S_2D(1-a/S_2)}{S_2D(1-a/S_2)} \right] \\ &= \frac{1}{(1-aS_0/D)S_0} \end{aligned}$$

which is equal to the left hand side of (IV.1). Thus, we get (4.9).

Appendix V

Derivation of (4.16)

Using (4.15) and result 4.3, we obtain

$$S_0 = \psi(t), \quad (V.1a)$$

$$S_1 = t \psi(t)(2\lambda - 1), \quad (V.1b)$$

$$S_2 = t^2 \psi(t), \quad (V.1c)$$

and hence

$$D = 4t^2 \psi^2(t) \lambda(1 - \lambda), \quad (V.2)$$

$$A = 1 - \frac{a}{4t^2 \psi(t) \lambda(1 - \lambda)}, \quad (V.3)$$

and

$$B = \psi^2(t) [(2\lambda - 1)t - \gamma]^2. \quad (V.4)$$

From (V.2) – (V.4), we obtain

$$S_0' = \begin{pmatrix} \frac{\partial S_0}{\partial \lambda} \\ \frac{\partial S_0}{\partial t} \end{pmatrix} = \begin{pmatrix} 0 \\ \psi'(t) \end{pmatrix} \quad (V.5)$$

$$D' = \begin{pmatrix} \frac{\partial D}{\partial \lambda} \\ \frac{\partial D}{\partial t} \end{pmatrix} = \begin{pmatrix} 4t^2 \psi^2(t)(1 - 2\lambda) \\ 8t \psi(t) [\psi(t) + t \psi'(t)] \lambda(1 - \lambda) \end{pmatrix}, \quad (V.6)$$

$$A' = \begin{pmatrix} \frac{\partial A}{\partial \lambda} \\ \frac{\partial A}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{a(1 - 2\lambda)}{4t^2 \psi(t) \lambda^2 (1 - \lambda)^2} \\ \frac{a[2\psi(t) + t \psi'(t)]}{4t^3 \psi^2(t) \lambda(1 - \lambda)} \end{pmatrix}, \quad (V.7)$$

and

$$B' = \begin{pmatrix} \frac{\partial B}{\partial \lambda} \\ \frac{\partial B}{\partial t} \end{pmatrix} = \begin{pmatrix} 4t\psi^2(t)[(2\lambda-1)t-\gamma] \\ 2\psi(t)[(2\lambda-1)t-\gamma][(2\lambda-1)\psi(t) + [(2\lambda-1)t-\gamma]\psi'(t)] \end{pmatrix} \quad (V.8)$$

From $\partial I / \partial \lambda = 0$ and $\partial I / \partial t = 0$, we obtain

$$\frac{\partial A}{\partial \lambda} (AD^2 + 2BD) = A \left(D \frac{\partial B}{\partial \lambda} - B \frac{\partial D}{\partial \lambda} \right) \quad (V.9)$$

and

$$\frac{\partial A}{\partial t} (AD^2 + 2BD) = A \left(D \frac{\partial B}{\partial t} - B \frac{\partial D}{\partial t} - \frac{D(AD+B)}{S_0} \frac{\partial S_0}{\partial t} \right) \quad (V.10)$$

Dividing (V.10) through by $(\partial A / \partial t)$ and substituting the result into (V.9) gives

$$\frac{\frac{\partial A}{\partial t} \left(D \frac{\partial B}{\partial t} - B \frac{\partial D}{\partial t} - \frac{D(AD+B)}{S_0} \frac{\partial S_0}{\partial t} \right)}{\frac{\partial A}{\partial t}} = D \frac{\partial B}{\partial t} - B \frac{\partial D}{\partial t}. \quad (V.11)$$

From (V.1a), (V.2)–(V.8), we obtain the following

$$\frac{\frac{\partial A}{\partial \lambda}}{\frac{\partial A}{\partial t}} = \frac{(1-2\lambda)t\psi(t)}{\lambda(1-\lambda)[2\psi(t) + t\psi'(t)]}, \quad (V.12a)$$

$$D \frac{\partial B}{\partial t} = 8t^2\psi^3(t)\lambda(1-\lambda)[(2\lambda-1)t-\gamma][(2\lambda-1)\psi(t) + [(2\lambda-1)t-\gamma]\psi'(t)] \quad (V.12b)$$

$$B \frac{\partial D}{\partial t} = 8t\psi^3(t)\lambda(1-\lambda)[(2\lambda-1)t-\gamma]^2[\psi(t) + t\psi'(t)] \quad (V.12c)$$

$$\frac{D(AD+B)}{S_0} \frac{\partial S_0}{\partial t} = 4t^2\psi^2(t)\psi'(t)\lambda(1-\lambda)[-a + \psi(t)[t^2 - 2(2\lambda-1)t\gamma + \gamma^2]] \quad (V.12d)$$

and so

$$D \frac{\partial B}{\partial t} - B \frac{\partial D}{\partial t} - \frac{D(AD+B)}{S_0} \frac{\partial S_0}{\partial t} = 4t\psi^2(t)\lambda(1-\lambda)[2t\psi(t)q - 2\psi(t)r - t\psi'(t)s] \quad (V.13)$$

where,

$$q = [(2\lambda-1)t - \gamma][(2\lambda-1)\psi(t) + (2\lambda-1)t - \gamma)\psi'(t)],$$

$$r = [(2\lambda-1)t - \gamma]^2[\psi(t) + t\psi'(t)],$$

and

$$s = -a + \psi(t)[t^2 - 2(2\lambda-1)t\gamma + \gamma^2].$$

Also,

$$D \frac{\partial B}{\partial \lambda} - B \frac{\partial D}{\partial \lambda} = [(2\lambda-1)t - \gamma][t - (2\lambda-1)\gamma]. \quad (V.14)$$

Therefore, on substituting (V.12a), (V.13) and (V.14) into (V.11), we obtain

$$\frac{4t^2\psi^3(t)(1-2\lambda)}{2\psi(t) + t\psi'(t)} [2t\psi(t)q - 2\psi(t)r - 2\psi(t)s] = 4t^2\psi^4(t)[(2\lambda-1)t - \gamma][t - (2\lambda-1)\gamma]$$

or

$$(1-2\lambda)[2ty(t)q - 2y(t)r - 2y(t)s] = \psi(t)[2\psi(t) + t\psi'(t)][(2\lambda-1)t - \gamma][t - (2\lambda-1)\gamma]$$

or

$$\begin{aligned} & \lambda^2 \{-8t\psi^2(t)\gamma - 8t^2\psi'(t)\psi(t)\gamma\} + \\ & \lambda \{8t\psi^2(t)\gamma + 8t^2\psi'(t)\psi(t)\gamma + 4\psi^2(t)\gamma^2 + 2t^3\psi'(t)\psi(t) + 2t\psi'(t)\psi(t)\gamma^2 - 2at\psi'(t)\} - \\ & \{2t\psi^2(t)\gamma + 2\psi^2(t)\gamma^2 + t^3\psi'(t)\psi(t) + 2t^2\psi'(t)\psi(t)\gamma + t\psi'(t)\psi(t)\gamma^2 - at\psi'(t)\} \\ & = \lambda^2 \{-8t\psi^2(t)\gamma - 4t^2\psi'(t)\psi(t)\gamma\} + \\ & \lambda \{8t\psi^2(t)\gamma + 4t^2\psi'(t)\psi(t)\gamma + 4\psi^2(t)\gamma^2 + 2t^3\psi'(t)\psi(t) + 2t\psi'(t)\psi(t)\gamma^2 + 4t^2\psi^2(t)\} - \\ & \{2t^2\psi^2(t) + t^3\psi'(t)\psi(t) + 4t\psi^2(t)\gamma + 2t^2\psi'(t)\psi(t)\gamma + 2\psi^2(t)\gamma^2 + t\psi'(t)\psi(t)\gamma^2\} \end{aligned}$$

or

$$\lambda^2 \{-4t^2\psi'(t)\psi(t)\gamma\} + \lambda \{4t^2\psi'(t)\psi(t)\gamma - 4t^2\psi^2(t) - 2at\psi'(t)\} + 2t\psi^2(t)\gamma + at\psi'(t) + 2t^2\psi^2(t) = 0$$

or

$$a'\lambda^2 + b'\lambda + c' = 0, \quad (\text{V.15})$$

where

$$a' = -4t^2\psi'(t)\psi(t)\gamma,$$

$$b' = 4t^2\psi'(t)\psi(t)\gamma - 4t^2\psi^2(t) - 2at\psi'(t),$$

and

$$c' = 2t\psi^2(t)\gamma + at\psi'(t) + 2t^2\psi^2(t).$$

Appendix VI

Proof of Result 4.3 (i) for F-Optimality

Letting $(u, v, w) = (S_0, S_1, S_2)$ in (4.4) we obtain

$$I = \frac{1}{(1-au/D)u} + \frac{(v-\gamma_\xi u)^2}{(1-g)^2 uD} = F + G, \text{ say} \quad (\text{VI.1})$$

where $D = uw - v^2$, $a = z_\alpha^2/n$, and $g = au/D$. Note that $u, w \geq 0$. We will show that (VI.1) is monotonically decreasing in w , for fixed u and v . This is done by showing that $\partial I / \partial w < 0$.

We have

$$\frac{\partial I}{\partial w} = \frac{\partial F}{\partial w} + \frac{\partial G}{\partial w} \quad (\text{VI.2})$$

where

$$\frac{\partial F}{\partial w} = \frac{\partial A}{\partial w} = -\frac{au}{D^2(1-g)^2}, \quad (\text{VI.3})$$

and

$$\begin{aligned} \frac{\partial G}{\partial w} &= (v-\gamma_\xi u)^2 \frac{\partial}{\partial w} \left(\frac{1}{(1-g)^2 uD} \right) \\ &= -(v-\gamma_\xi u)^2 \frac{(1-g^2)}{(1-g)^4 D^2} \end{aligned} \quad (\text{VI.4})$$

so that on substituting (VI.3) and (VI.4) into (VI.2), we get

$$\begin{aligned} \frac{\partial I}{\partial w} &= -\frac{au}{D^2(1-g)^2} - (v-\gamma_\xi u)^2 \frac{(1-g^2)}{(1-g)^4 D^2} \\ &= -\left(\frac{au}{D^2(1-g)^2} + (v-\gamma_\xi u)^2 \frac{(1-g^2)}{(1-g)^4 D^2} \right) \\ &< 0 \end{aligned}$$

for $g < 1$. Thus, I is monotonic decreasing in w for fixed u and v . Therefore, we have shown that the F-optimality criterion is decreasing in w for fixed u and v .

Appendix VII

1. Proof of Result 4.3 (ii) for the logit and probit models

2. Proof of Result 4.3 (iii) for the logit model

1. Proof of Result 4.3 (ii) for the logit and probit models

Let $\mathcal{H}_1$ be the line denoted by

$$\mathcal{H}_1 = \{ \psi(t)(1,t): t > 0 \}.$$

Then, the slope of $\mathcal{H}_1$ is

$$h_1(t) = \frac{\frac{d}{dt}\{t\psi(t)\}}{\frac{d}{dt}\psi(t)} = t + \psi(t)/\psi'(t).$$

We must show that $\mathcal{H}_1$ is concave, ie $h_1'(t) > 0$ for $t > 0$. Wu (1988) shows that for the logit and probit models, $h'(t) > 0$ for $t > 0$, where $h(t) = t + 2\psi(t)/\psi'(t)$. Now,

$$h_1'(t) = 1 + 2\left\{ 1 - \frac{\psi(t)\psi''(t)}{[\psi'(t)]^2} \right\} > 0,$$

where $t > 0$, which implies that

$$k(t) = \frac{1}{2} + \left\{ 1 - \frac{\psi(t)\psi''(t)}{[\psi'(t)]^2} \right\} > 0,$$

where $t > 0$. But

$$\begin{aligned} h_1'(t) &= 1 + \left\{ 1 - \frac{\psi(t)\psi''(t)}{[\psi'(t)]^2} \right\} \\ &= \frac{1}{2} + k(t) \end{aligned}$$

so then $h_1'(t) > 0$ for $t > 0$ as required. Hence the line $\mathcal{H}_1$ is concave.

2. Proof of Result 4.3 (iii) for the logit model

Let $\mathcal{H}_2$ be the line denoted by

$$\mathcal{H}_2 = \{ \psi(t)(1, t^2) : t > 0 \}.$$

Then, the slope of $\mathcal{H}_2$ is

$$h_2(t) = \frac{\frac{d}{dt} \{t^2 \psi(t)\}}{\frac{d}{dt} \psi(t)} = t r(t),$$

where

$$r(t) = t + 2 \frac{\psi(t)}{\psi'(t)}.$$

Then

$$r'(t) = \frac{(1+e^t)^2}{(1-e^t)^2}$$

and thus

$$\begin{aligned} h_2'(t) &= r(t) + t r'(t) \\ &= 2 \frac{t + te^{2t} + 1 - e^{2t}}{(1 - e^t)^2} \\ &\geq 0 \end{aligned}$$

if and only if

$$t + te^{2t} + (1 - e^{2t}) \geq 0.$$

Now, let

$$g(t) = t + te^{2t} + (1 - e^{2t}).$$

Since $g(0) = 0$, then, to show that $g(t) \geq 0$ for $t \geq 0$, we need only show that $g'(t) > 0$ for $t > 0$. But,

$$\begin{aligned} g'(t) &= 1 + e^{2t} + 2te^{2t} - 2e^{2t} \\ &= 1 - e^{2t} + 2te^{2t}. \end{aligned}$$

Since $g'(0) = 0$, then $g'(t) > 0$ for $t > 0$ if and only if $g''(t) > 0$ for $t > 0$. We have

$$g''(t) = 4te^{2t} > 0$$

for $t > 0$. Thus, $g(t) > 0$ for $t > 0$, which implies that $h_2'(t) > 0$, for $t > 0$. Hence, the line $\mathcal{H}_2$ is concave.

References

- Abdelbasit, K.M., and Plackett, R.L. (1983). Experimental Design for Binary Data. *Journal Of American Statistical Association*, **78**, 90-98.
- Elfving, G. (1952). Optimum allocation in linear regression theory. *Annals of Mathematical Statistics*, **23**, 255-262.
- Federov, V.V (1972). *Theory of Optimal Experiments*. Academic Press, New York.
- Finney, D.J. (1971). *Probit analysis*, 3rd Edition. Cambridge University Press, Cambridge.
- Forbes, B.E. (1994). Two-stage designs for binary response data. *MSc. thesis*, Department of Mathematics and Statistics, Carleton University.
- Gaudard, M.A., Karson, M.J., Linder, E. and Tse, S.K. (1990). The power logistic model for quantal response analysis. *Technical report*, Department of Mathematics, University of New Hampshire.
- Gaudard, M.A., Karson, M.J., Linder, E. and Tse, S.K. (1993). Efficient designs for estimation of a quantal response curve. *Statistica Sinica* **3**, 233-243.
- Gold L.S., Sawyer C.B., Magaw R., Backman G.M., de Veciana M., Levinson R., Hooper N.K., Havender W.R., Bernstein L., Peto R., Pike M.C., Ames B.N. (1984). A carcinogenic potency database of the standardized results of animal bioassays. *Environmental Health Perspectives* **58**, 9-322.
- Gold L.S., Slone T.H., Bernstein L. (1989). Summary of carcinogenic potency and positivity for 492 rodent carcinogens in the carcinogenic potency database. *Environmental Health Perspectives* **79**, 259-272.
- Johnson, N.L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, 2, Houghton Mifflin, New York.
- Kalish, L.A. & Rosenberger, J.L. (1978). Optimal designs for the estimation of

the logistic function. *Technical Report 33*, Department of Statistics, Pennsylvania State University.

- Krewski, D. & Ryzin, J.L. (1981). Dose response models for quantal Response Toxicity Data. *Statistics and Related Topics* (editors M. Csergo, D.A. Dawson, J.N.K. Rao, A.K.Md. Saleh).
- Krewski, D., Bickis, M., Kovar, J., Arnold, D.L. (1984). Optimal experimental designs for low dose extrapolation. *Utilitas Mathematica*.
- Morgan, B.J.T. (1985). The cubic logistic model for quantal assay data. *Applied Statistics* **34**, 105-113.
- Morgan, B.J.T. (1992). Analysis of quantal response data (1st ed.). Chapman & Hall.
- Nelder, J.A. and Mead, R. (1965). *Computer Journal* **7**, 308.
- Press, W., Flannery, B., Teukolsky, S., Vetterling, W. (1988). *Numerical Recipes in C*. Cambridge University Press, New York, 305-309.
- Sitter, R.R. & Wu, C.F.J. (1993). On the accuracy of Fieller intervals for binary response data. *Journal of American Statistical Association* **88**, 1021-1025.
- Sitter, R.R. & Wu, C.F.J. (1993). Optimal designs for binary response experiments: Fieller, D, and A criteria. *Scandinavian Journal of Statistics* **20**, 329-341.
- Wu, C.F.J. (1988). Optimal design for percentile estimation of a quantal response curve. *Optimal Design and Analysis of Experiments* (editors Y. Dodge, V. Federov & H.P. Wynn), Elsevier, North-Holland, 213-223.

END

0 1 - 0 6 - 9 5

FIN