Introduction to Optimal Design of Experiments

Weng Kee Wong

Department of Biostatistics **Fielding School of Public Health**

UCLA

An Overview of Optimal Design of Experiments
Biostatistics 279
Spring 2021

Outline

1.0 Model based optimal designs for a clinical study

Regression models:
- a statistical method for describing a 'response' or 'outcome' variable (Y) as a simple function of 'explanatory' or 'predictor' variables $(x_1, x_2, \ldots, x_k)$

A simple linear model has $k = 1$ and both Y and the predictor variable x are continuous variables.

Examples:

- total lung capacity (l) and height(m)

1.0 Model based optimal designs for a clinical study

Regression models:
- a statistical method for describing a 'response' or 'outcome' variable (Y) as a simple function of 'explanatory' or 'predictor' variables $(x_1, x_2, \ldots, x_k)$

A simple linear model has $k = 1$ and both Y and the predictor variable x are continuous variables.

Examples:

- total lung capacity (l) and height(m)
- birth weight (g) and gestational age (months)

1.0 Model based optimal designs for a clinical study

Regression models:
- a statistical method for describing a 'response' or 'outcome' variable (Y) as a simple function of 'explanatory' or 'predictor' variables $(x_1, x_2, \ldots, x_k)$

A simple linear model has $k = 1$ and both Y and the predictor variable x are continuous variables.

Examples:

- total lung capacity (l) and height(m)
- birth weight (g) and gestational age (months)
- systolic BP (mm Hg) and salt intake (mg)

1.0 Model based optimal designs for a clinical study

Regression models:
- a statistical method for describing a 'response' or 'outcome' variable (Y) as a simple function of 'explanatory' or 'predictor' variables $(x_1, x_2, \ldots, x_k)$

A simple linear model has $k = 1$ and both Y and the predictor variable x are continuous variables.

Examples:

- total lung capacity (l) and height(m)
- birth weight (g) and gestational age (months)
- systolic BP (mm Hg) and salt intake (mg)
- systolic BP (mmHg) and drug concentration (mg/kg)

1.1 Motivation using the Simple Linear Model

An Illustrative Dose Response Example

$$y = \theta_0 + \theta_1 x + error$$

E(error) = 0 var(error) = σ^2 N i.i.d. observations

On the dose range $X = [-1, 1]$, what is the optimal design for estimating

- (1) θ_0?

1.1 Motivationusing the Simple Linear Model

An Illustrative Dose Response Example

$$y = \theta_0 + \theta_1 x + error$$

E(error) = 0 var(error) = σ^2 N i.i.d. observations

On the dose range $X = [-1, 1]$, what is the optimal design for estimating

- (1) θ_0?
- (2) θ_1?

1.1 Motivation using the Simple Linear Model

An Illustrative Dose Response Example

$$y = \theta_0 + \theta_1 x + error$$

E(error) = 0 var(error) = σ^2 N i.i.d. observations

On the dose range $X = [-1, 1]$, what is the optimal design for estimating

- (1) θ_0?
- (2) θ_1?
- (3) both θ_0 and θ_1?

1.1 Motivationusing the Simple Linear Model

An Illustrative Dose Response Example

$$y = \theta_0 + \theta_1 x + error$$

E(error) = 0 var(error) = σ^2 N i.i.d. observations

On the dose range $X = [-1, 1]$, what is the optimal design for estimating

- (1) θ_0?
- (2) θ_1?
- (3) both θ_0 and θ_1?
- Answer for (2) comes from

$$var(\hat{\theta}_1) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \geq \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2} \geq \frac{\sigma^2}{N}.$$

1.1 Motivationusing the Simple Linear Model

An Illustrative Dose Response Example

$$y = \theta_0 + \theta_1 x + error$$

E(error) = 0 var(error) = σ^2 N i.i.d. observations

On the dose range $X = [-1, 1]$, what is the optimal design for estimating

- (1) θ_0?
- (2) θ_1?
- (3) both θ_0 and θ_1?
- Answer for (2) comes from

$$var(\hat{\theta}_1) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \geq \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2} \geq \frac{\sigma^2}{N}.$$

- For (2) and (3), take equal number of observations at ± 1.

1.1 Motivationusing the Simple Linear Model

An Illustrative Dose Response Example

$$y = \theta_0 + \theta_1 x + error$$

E(error) = 0 var(error) = σ^2 N i.i.d. observations

On the dose range $X = [-1, 1]$, what is the optimal design for estimating

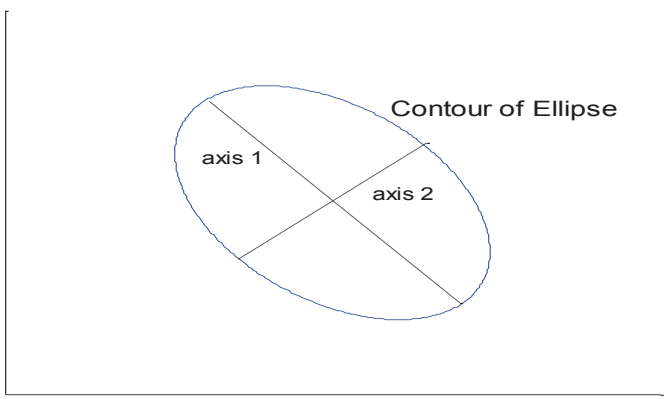
- (1) θ_0?
- (2) θ_1?
- (3) both θ_0 and θ_1?
- Answer for (2) comes from

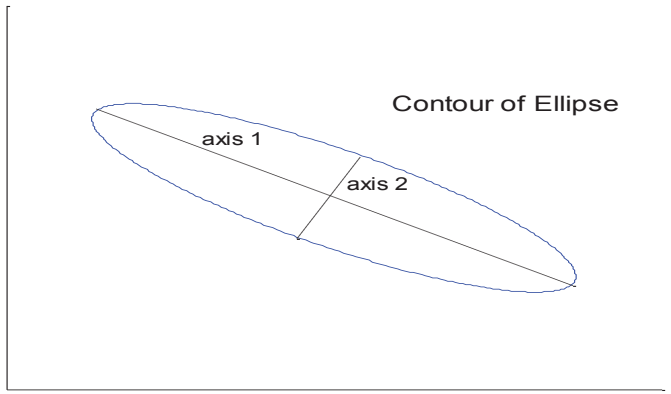
$$var(\hat{\theta}_1) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \geq \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2} \geq \frac{\sigma^2}{N}.$$

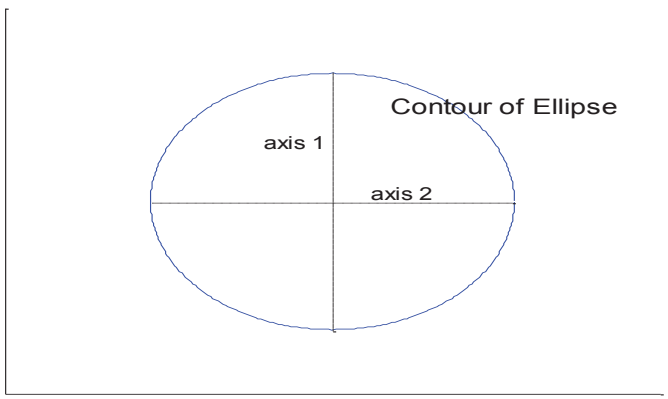
- For (2) and (3), take equal number of observations at ± 1.
- For (1) answer is any design with $\bar{x} = 0$.

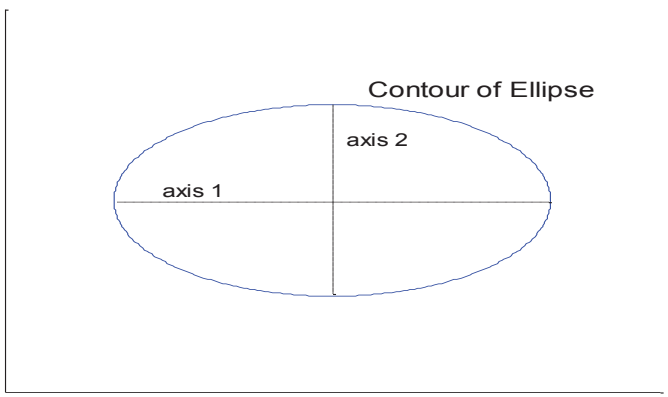
1.2 Design affects the quality of statistical inference

Simultaneous confidence ellipses for the two parameters in the simple linear model from different designs









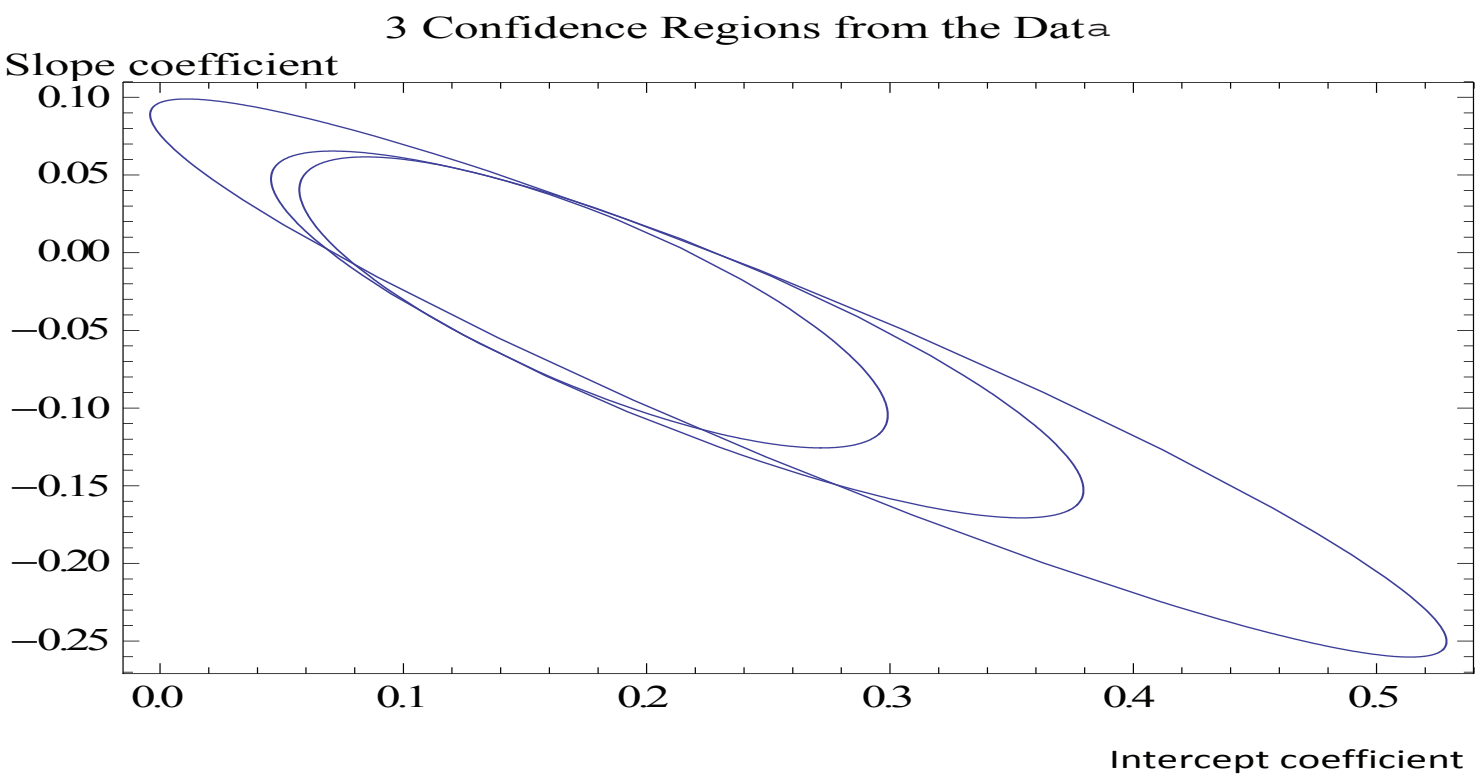
1.3 An illustrative Case with (x_i, y_i) from 3 data sets

```
data1 = {{1.0, .126}, {1.0, .219}, {2.0, .076}, {2.0, .126}, {0, .2}, {0, .13}};
data2 = {{1.0, .126}, {1.0, .219}, {2.0, .076}, {2.0, .126}, {0, .2}};
data3 = {{1.0, .126}, {1.0, .219}, {2.0, .076}, {2.0, .126}, {1, .2}};
```

LinearModelFit[data1,x,x]: $\hat{y} = 0.18 - 0.03\ x$

LinearModelFit[data2,x,x]: $\hat{y} = 0.21 - 0.05\ x$

LinearModelFit[data3,x,x]: $\hat{y} = 0.26 - 0.08\ x$



1.4 A Typical Setup for a Design Problem

- a given design interval X
- a parametric model with unknown parameters
- errors are normally and independently distributed
- observations have with constant variance
- a pre-determined sample size N

QUESTION

Given a fixed sample size N, how to select the N points from the design interval X to observe the response y in some optimal way?

1.5 Practical Design Issues

What about

1.5 Practical Design Issues

What about

- alternative optimality criterion?

1.5 Practical Design Issues

What about

- alternative optimality criterion?
- more complicated models, eg. heteroscedatic models
  nonlinear models
  correlated responses?

1.5 Practical Design Issues

What about

- alternative optimality criterion?
- more complicated models, eg. heteroscedatic models
  nonlinear models
  correlated responses?
- model assumptions - are they reasonable?

1.5 Practical Design Issues

What about

- alternative optimality criterion?
- more complicated models, eg. heteroscedatic models
  nonlinear models
  correlated responses?
- model assumptions - are they reasonable?
- the actual interests in the study?

1.6 More Complicated Objectives

Consider $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2$ on $X = [0, 3]$

1.6 More Complicated Objectives

Consider $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2$ on $X = [0, 3]$

- (i) What about just estimating only θ_2?

1.6 More Complicated Objectives

Consider $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2$ on $X = [0, 3]$

- (i) What about just estimating only θ_2?
- (ii) Extrapolate at z: estimate $g(\theta) = \theta_0 + \theta_1 z + \theta_2 z^2,\ z > 3$.

1.6 More Complicated Objectives

Consider $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2$ on $X = [0, 3]$

- (i) What about just estimating only θ_2?
- (ii) Extrapolate at z: estimate $g(\theta) = \theta_0 + \theta_1 z + \theta_2 z^2,\ z > 3$.
- (iii) Estimate the turning point:

$$g(\theta) = -\frac{\theta_1}{2\theta_2}, \qquad \theta_2 \neq 0.$$

1.6 More Complicated Objectives

Consider $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2$ on $X = [0, 3]$

- (i) What about just estimating only θ_2?
- (ii) Extrapolate at z: estimate $g(\theta) = \theta_0 + \theta_1 z + \theta_2 z^2,\ z > 3$.
- (iii) Estimate the turning point:

$$g(\theta) = -\frac{\theta_1}{2\theta_2}, \qquad \theta_2 \neq 0.$$

- (iv) Estimate response surface over a given set Z ,say $Z = [4, 5]$.

1.6 More Complicated Objectives

Consider $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2$ on $X = [0, 3]$

- (i) What about just estimating only θ_2?
- (ii) Extrapolate at z: estimate $g(\theta) = \theta_0 + \theta_1 z + \theta_2 z^2,\ z > 3$.
- (iii) Estimate the turning point:

$$g(\theta) = -\frac{\theta_1}{2\theta_2}, \qquad \theta_2 \neq 0.$$

- (iv) Estimate response surface over a given set Z ,say $Z = [4, 5]$.

- For (i), (ii) & (iii), require design techniques for estimating a given function of model parameters.

1.6 More Complicated Objectives

Consider $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2$ on $X = [0, 3]$

- (i) What about just estimating only θ_2?
- (ii) Extrapolate at z: estimate $g(\theta) = \theta_0 + \theta_1 z + \theta_2 z^2,\ z > 3$.
- (iii) Estimate the turning point:

$$g(\theta) = -\frac{\theta_1}{2\theta_2}, \qquad \theta_2 \neq 0.$$

- (iv) Estimate response surface over a given set Z ,say $Z = [4, 5]$.

- For (i), (ii) & (iii), require design techniques for estimating a given function of model parameters.

- For (iv), minimize the largest variance among all variances of the estimated responses from Z.

1.7 More Complicated Models:heteroscedastic models

Suppose in the simple linear model on $X = [-1, 1]$, var(y(x)) is proportional to a known function, say, $\frac{1}{\lambda(x)}$.

Question: Is the equally weighted design at ± 1 still optimal for estimating both the intercept and the slope?

1.7 More Complicated Models:heteroscedastic models

Suppose in the simple linear model on $X = [-1, 1]$, var(y(x)) is proportional to a known function, say, $\frac{1}{\lambda(x)}$.

Question: Is the equally weighted design at ± 1 still optimal for estimating both the intercept and the slope?

- Depends on $\lambda(x)$

1.7 More Complicated Models:heteroscedastic models

Suppose in the simple linear model on $X = [-1, 1]$, var(y(x)) is proportional to a known function, say, $\frac{1}{\lambda(x)}$.

Question: Is the equally weighted design at ± 1 still optimal for estimating both the intercept and the slope?

- Depends on $\lambda(x)$
- Suppose (i) $\lambda(x) = 2 - x^2$, (ii) $\lambda(x) = 4 - x^2$

1.7 More Complicated Models:heteroscedastic models

Suppose in the simple linear model on $X = [-1, 1]$, var(y(x)) is proportional to a known function, say, $\frac{1}{\lambda(x)}$.

Question: Is the equally weighted design at ± 1 still optimal for estimating both the intercept and the slope?

- Depends on $\lambda(x)$
- Suppose (i) $\lambda(x) = 2 - x^2$, (ii) $\lambda(x) = 4 - x^2$
- Answers: No for (i) and Yes for (ii).

1.7 More Complicated Models:heteroscedastic models

Suppose in the simple linear model on $X = [-1, 1]$, var(y(x)) is proportional to a known function, say, $\frac{1}{\lambda(x)}$.

Question: Is the equally weighted design at ± 1 still optimal for estimating both the intercept and the slope?

- Depends on $\lambda(x)$
- Suppose (i) $\lambda(x) = 2 - x^2$, (ii) $\lambda(x) = 4 - x^2$
- Answers: No for (i) and Yes for (ii).
- When $\lambda(x) = c - x^2$, optimal design is to take equal number of observations at ± 1 if and only if c > 3; otherwise take equal number of observations at $\pm\sqrt{c/3}$.

1.8 More Complicated Models: nonlinear models

- Michaelis-Menten model widely used in enzyme-kinetic studies

$$E(y) = \frac{\theta_0 x}{\theta_1 + x}$$

y is some response and x is substrate concentration
θ_0 is the saturation rate and θ_1 is one-half the maximum rate

1.8 More Complicated Models: nonlinear models

- Michaelis-Menten model widely used in enzyme-kinetic studies

$$E(y) = \frac{\theta_0 x}{\theta_1 + x}$$

  y is some response and x is substrate concentration
  θ_0 is the saturation rate and θ_1 is one-half the maximum rate
- Logistic model for binary outcome y:

$$ln\ \frac{P(y=1))}{1-P(y=1)} = ln\ \frac{E(y)}{1-E(y)} = \theta_1(x-\theta_0),\ \ \theta^T = (\theta_0, \theta_1).$$

1.8 More Complicated Models: nonlinear models

- Michaelis-Menten model widely used in enzyme-kinetic studies

$$E(y) = \frac{\theta_0 x}{\theta_1 + x}$$

  y is some response and x is substrate concentration
  θ_0 is the saturation rate and θ_1 is one-half the maximum rate
- Logistic model for binary outcome y:

$$ln\ \frac{P(y=1))}{1-P(y=1)} = ln\ \frac{E(y)}{1-E(y)} = \theta_1(x-\theta_0),\ \ \theta^T = (\theta_0, \theta_1).$$

- Find the dose that yields a user-selected $100\pi\%$ response rate:

$$g(\theta) = ED100\pi = \theta_0 + \frac{ln(\pi/(1-\pi))}{\theta_1}.$$

1.8 More Complicated Models: nonlinear models

- Michaelis-Menten model widely used in enzyme-kinetic studies

$$E(y) = \frac{\theta_0 x}{\theta_1 + x}$$

  y is some response and x is substrate concentration
  θ_0 is the saturation rate and θ_1 is one-half the maximum rate
- Logistic model for binary outcome y:

$$ln\ \frac{P(y=1))}{1-P(y=1)} = ln\ \frac{E(y)}{1-E(y)} = \theta_1(x-\theta_0),\ \ \theta^T = (\theta_0, \theta_1).$$

- Find the dose that yields a user-selected $100\pi\%$ response rate:

$$g(\theta) = ED100\pi = \theta_0 + \frac{ln(\pi/(1-\pi))}{\theta_1}.$$

- Design techniques for estimating $g(\theta)$ first compute its gradient before minimizing the asymptotic variance of the estimated $g(\theta)$; see Zhu & Wong (Statistics in Medicine, 2001).

1.9 Further Design Issues: Model Uncertainty and Different Interests

- Consider simple versus quadratic polynomial regression models: How do you design under model uncertainty?

1.9 Further Design Issues: Model Uncertainty and Different Interests

- Consider simple versus quadratic polynomial regression models: How do you design under model uncertainty?

- For the Michaelis-Menten Model: $E(y) = \frac{\theta_0 x}{\theta_1 + x}$, more interest to estimate θ_1 than θ_0; find a design that can estimate θ_1 with 90% efficiency and then estimate θ_0 optimally (Lopez-Fidalgo & Wong, Journal of Theoretical Biology, 2004)

1.9 Further Design Issues: Model Uncertainty and Different Interests

- Consider simple versus quadratic polynomial regression models: How do you design under model uncertainty?

- For the Michaelis-Menten Model: $E(y) = \frac{\theta_0 x}{\theta_1 + x}$, more interest to estimate θ_1 than θ_0; find a design that can estimate θ_1 with 90% efficiency and then estimate θ_0 optimally (Lopez-Fidalgo & Wong, Journal of Theoretical Biology, 2004)

- Is the Emax Model: $E(y) = \frac{\theta_0 x^h}{\theta_1 + x^h}$ a better fitting model? Find an optimal design for estimating θ_0 and θ_1 subject to our estimate for h is sufficiently accurate (Dette, Melas & Wong, JASA, 2005)

1.9 Further Design Issues: Model Uncertainty and Different Interests

- Consider simple versus quadratic polynomial regression models: How do you design under model uncertainty?
- For the Michaelis-Menten Model: $E(y) = \frac{\theta_0 x}{\theta_1 + x}$, more interest to estimate θ_1 than θ_0; find a design that can estimate θ_1 with 90% efficiency and then estimate θ_0 optimally (Lopez-Fidalgo & Wong, Journal of Theoretical Biology, 2004)
- Is the Emax Model: $E(y) = \frac{\theta_0 x^h}{\theta_1 + x^h}$ a better fitting model? Find an optimal design for estimating θ_0 and θ_1 subject to our estimate for h is sufficiently accurate (Dette, Melas & Wong, JASA, 2005)
- 2 or 3-parameter logistic model? Which efficiency function?

Animal experiments under fire for poor design

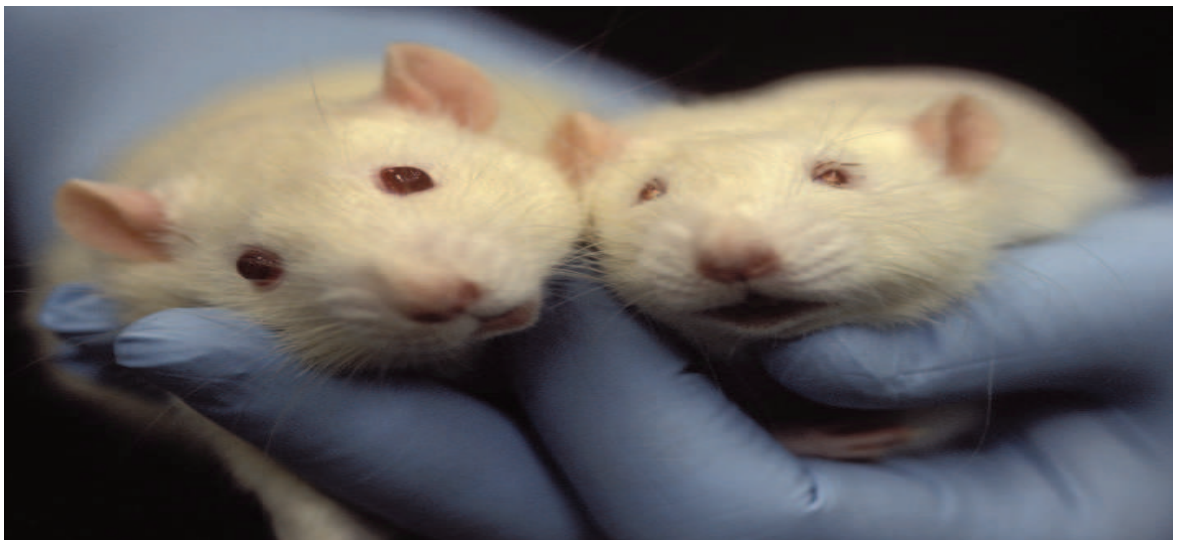
(from NATURE|Vol 444|21/28 December 2006)

In the contentious world of animal research, one question surfaces time and again: how useful are animal experiments as a way to prepare for trials of medical treatments in humans? The issue is crucial, as public opinion is behind animal research only if it helps develop better drugs. Consequently, scientists defending animal experiments insist they are essential for safe clinical trials, whereas animal-rights activists vehemently maintain that they are useless.

Now a British team has made the first attempt to answer the question in a scientific way, and the result suggests that animal researchers need to raise their game. The team claims that animal experiments are often poorly designed, and so fail to lay the ground properly for subsequent human studies.

The study looked at six treatments that have been evaluated in detail in human trials. The researchers assessed whether animal studies had accurately predicted the outcome of the human work, a task that involved reviewing more than 200 papers. In three of the six cases, the answer was no (P. Perel *et al. Br. Med. J.* doi:10.1136/bmj.39048.407928.BE; 2006).

Ian Roberts, an epidemiologist at the London School of Hygiene and Tropical Medicine and one of the study's authors, says further analysis revealed that poor-quality methodologies and inappropriate models caused the failures. Some studies didn't randomize the allocation of animals to control and treatment groups properly, whereas others showed evidence for publication bias — the tendency not to publish negative results. A third set involved inappropriate models: for example in studies of head injury, treatments were administered just five minutes after rodents were deliberately injured, but humans are typically



K. KASMAUSKI/CORBIS

Are animals being wasted in badly thought through experiments?

to happen in animal experiments," he says.

Some scientists familiar with both animal and human studies have welcomed the paper, saying it highlights the need for researchers from the two arenas to work more closely on the design of animal experiments. But others have questioned whether a crude comparison of trials says anything meaningful about the usefulness of animal work.

"Small-scale studies are pointless if they do not produce results that people have confidence in."

In many cases, say critics, animal studies are designed to address specific questions not related to a drug's effectiveness, such as how it distributes

College London, and don't expect the results to map straight over to human work.

Roberts counters that his paper looked only at animal studies that were specifically designed to model human disease. And he adds that despite the low cost, small-scale studies are pointless if they do not produce results that people can have confidence in, even if researchers are aware of the limitations. "If you have something cheap and cheerful you have a high probability that the result is wrong," he says. "Just think of the money that is wasted down the line if you follow up that lead."

The results are likely to feature in future campaigns by animal-rights groups as evi-

2 Design Problems for Dose Response Studies and General Principles of Experimental Design

Dose response studies are widely used and applied to various types of problems in biomedicine and also outside of biomedicine (eg. crop yield in agriculture, stress testing in manufacturing and the defence industry)

References:

- Berger, M. P. F. and Wong, W. K. (2005). Applied Optimal Designs. John Wiley Sons.

- Berger, M. P. F and Wong, W. K. (2009). An Introduction to Optimal Designs with Applications to Social and Biomedical Research. John Wiley Sons.

Now two specific examples of dose response studies in biomedicine:

2.1 A 4-parameter Heteroscedastic Hill Model

$$y_i = \frac{(E_{con} - b)(\frac{D_i}{IC_{50}})^m}{1 + (\frac{D_i}{IC_{50}})^m} + b + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma(Ey_i)^{2\lambda})$$

D_i = dose of a drug assigned to subject i

y_i = drug effect of subject i

E_{con} = the control effect at zero drug concentration

b = background effect at infinite drug concentration

IC_{50} = inflection point on the curve (a measure of the drug potency)
= drug concentration that induces a 50% decrease in the maximal effect ($Econ - b$)

m = slope parameter, which is negative for an inhibitory drug.

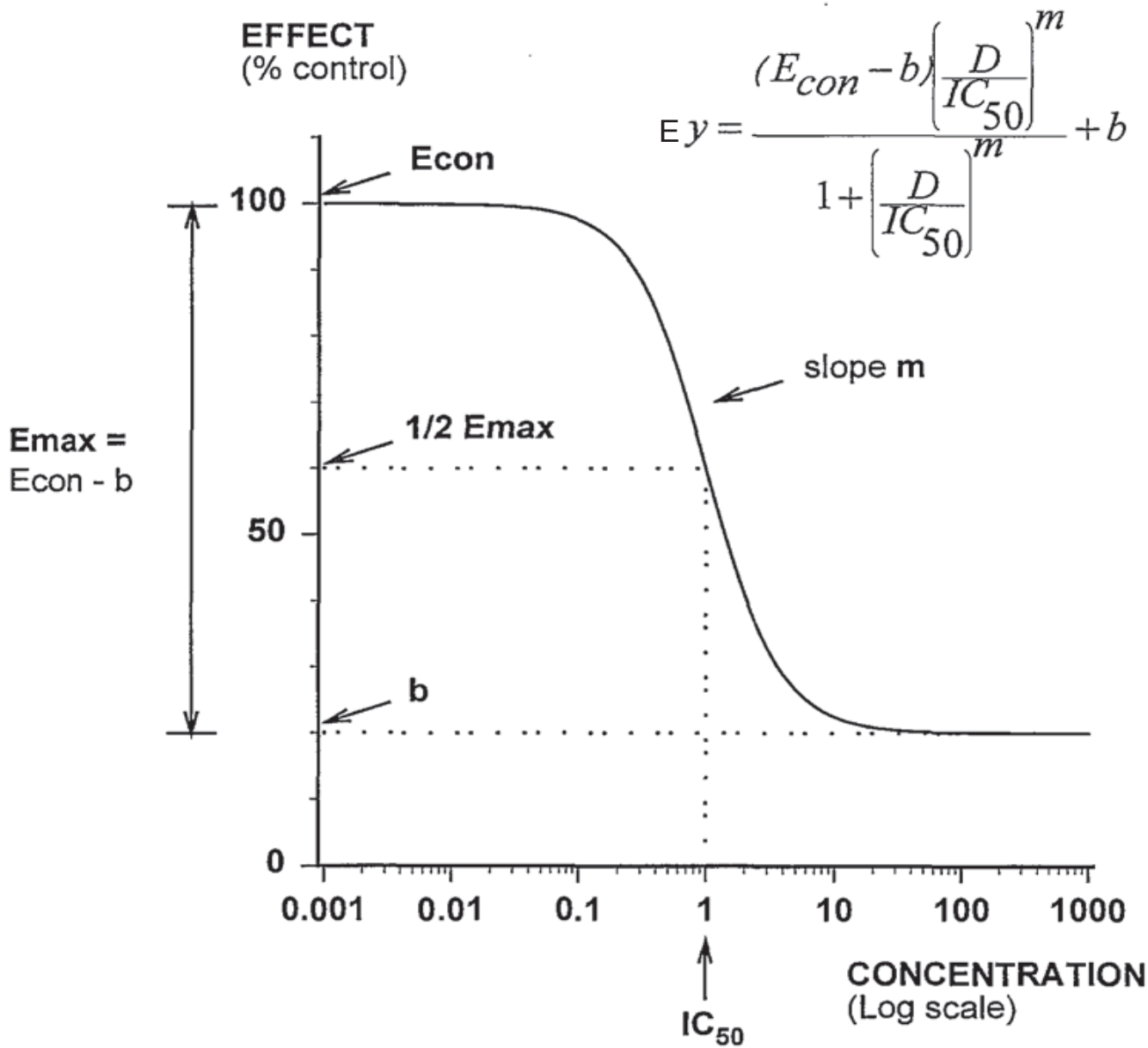


Figure 1. Graph of the 4-parameter Hill model. The following parameter values have been assumed: $E_{con} = 100$, $b = 20$, $IC_{50} = 1$, and $m = -1.5$.

2.3 Adaptive Optimal Design for the E_{max} Model (Leonov & Miller (2009). J. of Biopharm. Stat.)

Let $\theta = (\gamma, ED_{50}, E_{max}, \sigma_A^2, \sigma_M^2)$.

$$y_i = f(d_i, \theta) + e_i = \frac{E_{max} d_i^{\gamma}}{ED_{50}^{\gamma} + d_i^{\gamma}} + e_i = \frac{E_{max}}{1 + (ED_{50}/d_i)^{\gamma}} + e_i,$$

$$var(e_i) = \sigma_A^2 + \sigma_M^2 f(d_i, \theta) \text{ or } var(ei) = \sigma_A^2 + \sigma_M^2 f(d_i, \theta)(E_{max} - f(d_i, \theta))$$

GOAL 1: Estimate ED_{90}, the dose that attains 90% of the maximal response. A direct calculation shows

$$ED_{90} = ED_{50} 9^{1/\gamma}$$

GOAL 2: Estimate D_{90}, the dose that attains the absolute response of 90%, not related to percentages of E_{max}. A direct calculation shows

$$D_{90} = ED_{50} \left(\frac{90}{E_{max} - 90}\right)^{1/\gamma}.$$

2.4a General Principles of Experimental Designs

(Taken from Box and Draper, 1975, Biometrika)

2.4a General Principles of Experimental Designs

(Taken from Box and Draper, 1975, Biometrika)

- Generate a satisfactory distribution of information throughout the region of interest, which may not coincide with the design region

2.4a General Principles of Experimental Designs

(Taken from Box and Draper, 1975, Biometrika)

- Generate a satisfactory distribution of information throughout the region of interest, which may not coincide with the design region
- Ensure that the fitted value is close to the true value of the response

2.4a General Principles of Experimental Designs

(Taken from Box and Draper, 1975, Biometrika)

- Generate a satisfactory distribution of information throughout the region of interest, which may not coincide with the design region
- Ensure that the fitted value is close to the true value of the response
- Possible to detect a lack of fit

2.4a General Principles of Experimental Designs

(Taken from Box and Draper, 1975, Biometrika)

- Generate a satisfactory distribution of information throughout the region of interest, which may not coincide with the design region
- Ensure that the fitted value is close to the true value of the response
- Possible to detect a lack of fit
- Allow estimation of transformation of both the response and the quantitative experiment factors

2.4a General Principles of Experimental Designs

(Taken from Box and Draper, 1975, Biometrika)

- Generate a satisfactory distribution of information throughout the region of interest, which may not coincide with the design region
- Ensure that the fitted value is close to the true value of the response
- Possible to detect a lack of fit
- Allow estimation of transformation of both the response and the quantitative experiment factors
- Allow designs of increasing order to be built sequentially, like second-order design to follow first order design

2.4a General Principles of Experimental Designs

(Taken from Box and Draper, 1975, Biometrika)

- Generate a satisfactory distribution of information throughout the region of interest, which may not coincide with the design region
- Ensure that the fitted value is close to the true value of the response
- Possible to detect a lack of fit
- Allow estimation of transformation of both the response and the quantitative experiment factors
- Allow designs of increasing order to be built sequentially, like second-order design to follow first order design
- Provide an internal estimate of error from replication

2.4a General Principles of Experimental Designs

(Taken from Box and Draper, 1975, Biometrika)

- Generate a satisfactory distribution of information throughout the region of interest, which may not coincide with the design region
- Ensure that the fitted value is close to the true value of the response
- Possible to detect a lack of fit
- Allow estimation of transformation of both the response and the quantitative experiment factors
- Allow designs of increasing order to be built sequentially, like second-order design to follow first order design
- Provide an internal estimate of error from replication
- Be insensitive to wild observations and to violation of the usual normal theory assumptions

2.4b General Principles of Experimental Designs (cont'd)

- Require a minimum number of experiment runs

2.4b General Principles of Experimental Designs (cont'd)

- Require a minimum number of experiment runs
- Allow experiments to be performed in blocks

2.4b General Principles of Experimental Designs (cont'd)

- Require a minimum number of experiment runs
- Allow experiments to be performed in blocks
- Provide simple data patterns that allow ready visual appreciation

2.4b General Principles of Experimental Designs (cont'd)

- Require a minimum number of experiment runs
- Allow experiments to be performed in blocks
- Provide simple data patterns that allow ready visual appreciation
- Ensure simplicity of calculation/analysis

2.4b General Principles of Experimental Designs (cont'd)

- Require a minimum number of experiment runs
- Allow experiments to be performed in blocks
- Provide simple data patterns that allow ready visual appreciation
- Ensure simplicity of calculation/analysis
- Behave well when errors occur in the settings of the experimental variables

2.4b General Principles of Experimental Designs (cont'd)

- Require a minimum number of experiment runs
- Allow experiments to be performed in blocks
- Provide simple data patterns that allow ready visual appreciation
- Ensure simplicity of calculation/analysis
- Behave well when errors occur in the settings of the experimental variables
- Not require an impractically large number of levels of the experimental factors

2.4b General Principles of Experimental Designs (cont'd)

- Require a minimum number of experiment runs
- Allow experiments to be performed in blocks
- Provide simple data patterns that allow ready visual appreciation
- Ensure simplicity of calculation/analysis
- Behave well when errors occur in the settings of the experimental variables
- Not require an impractically large number of levels of the experimental factors
- Provide a check on the "constancy of variance" assumption

Exact vs Approximate Optimal Designs with a fixed sample size of N

Exact vs Approximate Optimal Designs with a fixed sample size of N

- An exact optimal design tells you how many doses you need, where these dose levels are and how many subjects to assign to each of these doses in some optimal way.

Exact vs Approximate Optimal Designs with a fixed sample size of N

- An exact optimal design tells you how many doses you need, where these dose levels are and how many subjects to assign to each of these doses in some optimal way.

- Approximate optimal design tells you how many doses you need, where these dose levels are and roughly how many subjects to assign to each of these doses in some optimal way. If $N = 25, X = [0, 3]$, and the approximate design takes equal proportions of observations at 0 and 3, the implemented design assigns 12 subjects dose 0 and 13 subjects to the dose at 3 or assign 13 subjects to dose 0 and 12 subjects to the dose at 3.

Exact vs Approximate Optimal Designs with a fixed sample size of N

- An exact optimal design tells you how many doses you need, where these dose levels are and how many subjects to assign to each of these doses in some optimal way.

- Approximate optimal design tells you how many doses you need, where these dose levels are and roughly how many subjects to assign to each of these doses in some optimal way. If $N = 25, X = [0, 3]$, and the approximate design takes equal proportions of observations at 0 and 3, the implemented design assigns 12 subjects dose 0 and 13 subjects to the dose at 3 or assign 13 subjects to dose 0 and 12 subjects to the dose at 3.

- Approximate designs were proposed and developed by Kiefer from 1958-1980ś.

3.1 Approximate designs (Kiefer, 1958-1982)

Suppose X, $f(x)$, N and an optimality criterion ϕ are given.

Formulation for Optimal Approximate Design Problem:
How many points are needed to optimize the criterion? Find k
Where are the optimal design (or support) points? Find
$x_1, x_2, \cdots, x_k \in X$
What is the optimal proportion of the total observations to take at each of these points? Find $w_1, w_2, , w_k$ such that

$$0 < w_i < 1, \ i = 1, 2, \cdots, k$$

$$w_1 + w_{2+} \cdots + w_k = 1.$$

The implemented design takes $n_i = [Nw_i]$ observations at design points $x_i, i = 1, 2, \cdots, k$ such that $n_1 + n_2 + \cdots + n_k = N$.

3.1 Approximate designs (Kiefer, 1958-1982)

Suppose X, $f(x)$, N and an optimality criterion ϕ are given.

Formulation for Optimal Approximate Design Problem:
How many points are needed to optimize the criterion? Find k
Where are the optimal design (or support) points? Find
$x_1, x_2, \cdots, x_k \in X$
What is the optimal proportion of the total observations to take at each of these points? Find $w_1, w_2, , w_k$ such that

$$0 < w_i < 1, \ i = 1, 2, \cdots, k$$

$$w_1 + w_{2+} \cdots + w_k = 1.$$

The implemented design takes $n_i = [Nw_i]$ observations at design points $x_i, i = 1, 2, \cdots, k$ such that $n_1 + n_2 + \cdots + n_k = N$.

- Optimal Exact Design Problem finds positive integers n_i (instead of w_i) subject to the constraint $n_1 + n_2 + \cdots + n_k = N$.

3.2 Exact Optimal Design for the Quadratic Model

J. R. Statist. Soc. B (1982),
44, *No.* 3, *pp.* 394–397

Exact *D*-optimum Designs for Quadratic Regression

By N. GAFFKE and O. KRAFFT

Rheinisch Westfälische Technische Hochschule Aachen, West Germany

[Received November 1980. Revised November 1981]

SUMMARY

It is shown that for quadratic regression on an interval a *D*-optimum design has its spectrum in the end- and mid-points of the interval with frequencies as equal as possible.

Keywords: QUADRATIC REGRESSION; *D*-OPTIMUM DESIGN

FOR $x = (x_1, \ldots, x_n) \in [a, b]^n$ let

$$M_n(x) = \begin{pmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{pmatrix},$$

where all sums run from one to n. A vector $x^* \in [a, b]^n$ is called an exact n-point *D*-optimum design for the quadratic regression

$$y = a_0 + a_1 \xi + a_2 \xi^2$$

(here errors are assumed to be uncorrelated and to have the same variance) iff

$$\det M_n(x^*) \geqslant \det M_n(x) \quad \text{for all } x \in [a, b]^n,$$

cf. Kiefer (1959). Since by the Binet, Cauchy and Vandermonde formulas we have

$$\det M_n(x) = \sum_{1 \leqslant i < j < k \leqslant n} \{(x_k - x_j)(x_k - x_i)(x_j - x_i)\}^2$$

without loss of generality we can assume that $[a, b] = [-1, 1]$. We are going to prove the following theorem.

Theorem. The exact n-point *D*-optimum designs $x^* = (x_1^*, \ldots, x_n^*)$ on $[-1, 1]$, with $n = 3p + t$ and $t = 1$ or $t = 2$, are obtained by choosing the $x_i^* \in \{-1, 0, 1\}$ in such a way that $3 - t$ of the values $-1, 0$ and 1 occur p times and the remaining t of them $p + 1$ times.

3.3 Advantages of Approximate Designs

Denote a generic approximate design by ξ as

$$\xi = \{x_1, x_2, \ldots, x_k; w_1, w_2, \ldots, w_k\},$$

$$x_i \in X, \;\; 0 < w_i < 1, i = 1, 2, \ldots, k \;\; \& \;\; \sum_{i=1}^{k} w_i = 1.$$

3.3 Advantages of Approximate Designs

Denote a generic approximate design by ξ as

$$\xi = \{x_1, x_2, \ldots, x_k; w_1, w_2, \ldots, w_k\},$$

$$x_i \in X, \;\; 0 < w_i < 1, i = 1, 2, \ldots, k \;\; \& \;\; \sum_{i=1}^{k} w_i = 1.$$

- For fixed N, the implemented design assigns Nw_i subjects to $x_i, i = 1, \ldots, k$, subject to $Nw_1 + Nw_2 + \cdots + Nw_k = N$.

3.3 Advantages of Approximate Designs

Denote a generic approximate design by ξ as

$$\xi = \{x_1, x_2, \ldots, x_k; w_1, w_2, \ldots, w_k\},$$

$$x_i \in X, \;\; 0 < w_i < 1, i = 1, 2, \ldots, k \;\; \& \;\; \sum_{i=1}^{k} w_i = 1.$$

- For fixed N, the implemented design assigns Nw_i subjects to $x_i, i = 1, \ldots, k$, subject to $Nw_1 + Nw_2 + \cdots + Nw_k = N$.
- Computer algorithms available for generating many types of optimal designs.

3.3 Advantages of Approximate Designs

Denote a generic approximate design by ξ as

$$\xi = \{x_1, x_2, \ldots, x_k; w_1, w_2, \ldots, w_k\},$$

$$x_i \in X, \;\; 0 < w_i < 1, i = 1, 2, \ldots, k \;\; \& \;\; \sum_{i=1}^{k} w_i = 1.$$

- For fixed N, the implemented design assigns Nw_i subjects to $x_i, i = 1, \ldots, k$, subject to $Nw_1 + Nw_2 + \cdots + Nw_k = N$.
- Computer algorithms available for generating many types of optimal designs.
- Can evaluate efficiency of any design without knowing what the optimum is.

3.3 Advantages of Approximate Designs

Denote a generic approximate design by ξ as

$$\xi = \{x_1, x_2, \ldots, x_k; w_1, w_2, \ldots, w_k\},$$

$$x_i \in X, \;\; 0 < w_i < 1, i = 1, 2, \ldots, k \;\; \& \;\; \sum_{i=1}^{k} w_i = 1.$$

- For fixed N, the implemented design assigns Nw_i subjects to $x_i, i = 1, \ldots, k$, subject to $Nw_1 + Nw_2 + \cdots + Nw_k = N$.
- Computer algorithms available for generating many types of optimal designs.
- Can evaluate efficiency of any design without knowing what the optimum is.
- Does not require an endless list of tables describing optimal design for each model, each N and each type of criterion

3.3 Advantages of Approximate Designs

Denote a generic approximate design by ξ as

$$\xi = \{x_1, x_2, \ldots, x_k; w_1, w_2, \ldots, w_k\},$$

$$x_i \in X, \;\; 0 < w_i < 1, i = 1, 2, \ldots, k \;\; \& \;\; \sum_{i=1}^{k} w_i = 1.$$

- For fixed N, the implemented design assigns Nw_i subjects to $x_i, i = 1, \ldots, k$, subject to $Nw_1 + Nw_2 + \cdots + Nw_k = N$.
- Computer algorithms available for generating many types of optimal designs.
- Can evaluate efficiency of any design without knowing what the optimum is.
- Does not require an endless list of tables describing optimal design for each model, each N and each type of criterion
- A simple way to verify whether you have the optimal design among all designs on X**********draw pictures!

3.4 Optimal Approximate Designs on $X = [-1, 1]$

Examples of D-optimal designs for estimating model parameters and making inference on the mean response at a selected dose level.

design criterion		linear model		quadratic model		
D-optimality	x_i	-1	1	-1	0	1
	w_i	$1/2$	$1/2$	$1/3$	$1/3$	$1/3$
Extrapolation	x_i	-1	1	-1	0	1
at dose level $z = 2$	w_i	$1/4$	$3/4$	$1/7$	$3/7$	$3/7$

3.4 Optimal Approximate Designs on $X = [-1, 1]$

Examples of D-optimal designs for estimating model parameters and making inference on the mean response at a selected dose level.

design criterion		linear model		quadratic model		
D-optimality	x_i	-1	1	-1	0	1
	w_i	$1/2$	$1/2$	$1/3$	$1/3$	$1/3$
Extrapolation	x_i	-1	1	-1	0	1
at dose level $z = 2$	w_i	$1/4$	$3/4$	$1/7$	$3/7$	$3/7$

- For nonlinear models, the information matrix depends on the unknown parameters and so nominal values are typically assumed to be available.

3.4 Optimal Approximate Designs on $X = [-1, 1]$

Examples of D-optimal designs for estimating model parameters and making inference on the mean response at a selected dose level.

design criterion		linear model		quadratic model		
D-optimality	x_i	-1	1	-1	0	1
	w_i	$1/2$	$1/2$	$1/3$	$1/3$	$1/3$
Extrapolation	x_i	-1	1	-1	0	1
at dose level $z = 2$	w_i	$1/4$	$3/4$	$1/7$	$3/7$	$3/7$

- For nonlinear models, the information matrix depends on the unknown parameters and so nominal values are typically assumed to be available.
- Resulting optimal designs are termed locally optimal because they depend on the nominal values.

3.5 Locally D-optimal Designs for the Logistic Model on $X = [-1, 1]$ (Ford's PhD thesis, 1972)

$$log\ \frac{\pi(x)}{1-\pi(x)} = \theta_1 + \theta_2 x, \quad \theta^T = (\theta_1, \theta_2), \ \ \theta_1 > 0 \ \ \& \ \ \theta_2 > 0.$$

3.5 Locally D-optimal Designs for the Logistic Model on $X = [-1, 1]$ (Ford's PhD thesis, 1972)

$$log\ \frac{\pi(x)}{1-\pi(x)} = \theta_1 + \theta_2 x, \quad \theta^T = (\theta_1, \theta_2), \quad \theta_1 > 0 \ \ \& \ \ \theta_2 > 0.$$

- Let a solve $exp(z) = (z+1)/(z-1)$, i.e. $a = 1.54$ and let u^* solve

$$exp(\theta_1 + \theta_2 u) = \frac{2 + (u+1)\theta_2}{-2 + (u+1)\theta_2}.$$

3.5 Locally D-optimal Designs for the Logistic Model on $X = [-1, 1]$ (Ford's PhD thesis, 1972)

$$log\ \frac{\pi(x)}{1-\pi(x)} = \theta_1 + \theta_2 x, \quad \theta^T = (\theta_1, \theta_2), \quad \theta_1 > 0 \ \ \& \ \ \theta_2 > 0.$$

- Let a solve $exp(z) = (z+1)/(z-1)$, i.e. $a = 1.54$ and let u^* solve

$$exp(\theta_1 + \theta_2 u) = \frac{2 + (u+1)\theta_2}{-2 + (u+1)\theta_2}.$$

condition	locally D-optimal design
$\{\theta : \theta_2 - \theta_1 \geq a\}$	$\{\frac{a-\theta_1}{\theta_2}, \frac{-a-\theta_1}{\theta_2}; \frac{1}{2}, \frac{1}{2}\}$
$\{\theta : \theta_2 - \theta_1 < a, exp(\theta_1 + \theta_2) \leq \frac{\theta_2+1}{\theta_2-1}\}$	$\{-1, u^*; \frac{1}{2}, \frac{1}{2}\}$
$\{\theta : exp(\theta_1 + \theta_2) > \frac{\theta_2+1}{\theta_2-1}\}$	$\{-1, 1; \frac{1}{2}, \frac{1}{2}\}$

3.5 Locally D-optimal Designs for the Logistic Model on $X = [-1, 1]$ (Ford's PhD thesis, 1972)

$$log\ \frac{\pi(x)}{1-\pi(x)} = \theta_1 + \theta_2 x, \quad \theta^T = (\theta_1, \theta_2), \quad \theta_1 > 0 \ \ \& \ \ \theta_2 > 0.$$

- Let a solve $exp(z) = (z+1)/(z-1)$, i.e. $a = 1.54$ and let u^* solve

$$exp(\theta_1 + \theta_2 u) = \frac{2 + (u+1)\theta_2}{-2 + (u+1)\theta_2}.$$

condition	locally D-optimal design
$\{\theta : \theta_2 - \theta_1 \geq a\}$	$\{\frac{a-\theta_1}{\theta_2}, \frac{-a-\theta_1}{\theta_2}; \frac{1}{2}, \frac{1}{2}\}$
$\{\theta : \theta_2 - \theta_1 < a, exp(\theta_1 + \theta_2) \leq \frac{\theta_2+1}{\theta_2-1}\}$	$\{-1, u^*; \frac{1}{2}, \frac{1}{2}\}$
$\{\theta : exp(\theta_1 + \theta_2) > \frac{\theta_2+1}{\theta_2-1}\}$	$\{-1, 1; \frac{1}{2}, \frac{1}{2}\}$

- Corrected results in Sebastiani and Settimi (JSPI, 1997)

3.6 Information Matrices for linear models

$$y(x) = f^T(x)\theta + e(x)/\sqrt{\lambda(x)}, \qquad x \in X.$$

- $f(x)$ = a given $d \times 1$ vector of regression functions
- $Ee(x) = 0$; $var(e(x))= \sigma^2$; $\lambda(x)$ = known positive function

If errors are normally and independently distributed, Fisher information matrix for a k-point design ξ is proportional to

$$M(\xi) = \sum_{i=1}^{k} \lambda(x_i) w_i f(x_i) f^T(x_i), \qquad k \geq d$$

and $cov(\widehat{\theta}) = M(\xi)^{-1}$(apart from a multiplicative constant).

For a nonlinear model, we have $E(y) = f(x, \theta)$; replace above $f(x)$ by gradient of $f(x, \theta)$ (with respect to θ).

3.7 Optimality Design Criteria

Formulate objective as a function of the information matrix:

- D-optimality: $\Phi(\xi) = ln\,|M(\xi)^{-1}|$

3.7 Optimality Design Criteria

Formulate objective as a function of the information matrix:

- D-optimality: $\Phi(\xi) = ln\,|M(\xi)^{-1}|$
- L-optimality: $\Phi(\xi) = tr\ L\ M(\xi)^{-1}$ for a user-selected matrix L

If L = identity matrix, we have $\Phi(\xi) = tr\ M(\xi)^{-1}$ (A-optimality)

If L = $f(z)f(z)^T$, $\Phi(\xi) = tr\ f(z)f(z)^T M(\xi)^{-1} = f(z)^T M(\xi)^{-1} f(z)$,
=var(z, ξ)
(the variance of y at z).

If L = $\frac{\partial g(\theta)}{\partial \theta}\frac{\partial g^T(\theta)}{\partial \theta^T}$ for a given function $g(\theta)$, $\Phi(\xi)$ is proportional to the asymptotic variance of the estimated $g(\theta)$.
(use a first order Taylor's expansion)

3.8 Convex Analysis

Each of the above Φ is convex; so find ξ^* that minimizes $\Phi(\xi)$ over ALL designs ξ on X.

3.8 Convex Analysis

Each of the above Φ is convex; so find ξ^* that minimizes $\Phi(\xi)$ over ALL designs ξ on X.

- The Frechet derivative of Φ at $M(\xi^*)$ in the direction of $M(\xi)$ is

$$F_\Phi(M(\xi^*), M(\xi)) = \lim_{\alpha \to 0} \frac{\Phi((1-\alpha)M(\xi^*) + \alpha M(\xi)) - \Phi(M(\xi^*))}{\alpha}.$$

- If Φ is convex on the set of information matrices and differentiable at $M(\xi^*)$, then ξ^* is Φ-optimal if and only if

$$F_\Phi(M(\xi^*), f(x)f^T(x))) \geq 0$$

for all $x \in$ X with equality at the design points of ξ^*.

3.8 Convex Analysis

Each of the above Φ is convex; so find ξ^* that minimizes $\Phi(\xi)$ over ALL designs ξ on X.

- The Frechet derivative of Φ at $M(\xi^*)$ in the direction of $M(\xi)$ is

$$F_\Phi(M(\xi^*), M(\xi)) = \lim_{\alpha \to 0} \frac{\Phi((1-\alpha)M(\xi^*) + \alpha M(\xi)) - \Phi(M(\xi^*))}{\alpha}.$$

- If Φ is convex on the set of information matrices and differentiable at $M(\xi^*)$, then ξ^* is Φ-optimal if and only if

$$F_\Phi(M(\xi^*), f(x)f^T(x))) \geq 0$$

  for all $x \in$ X with equality at the design points of ξ^*.
- If Φ is non-differentiable, work with sub-gradient; see Wong (Biometrika, 1992), Wong and Cook (JRSSB, 1993) and Wong (JSPI, 1994), Berger, King & Wong (Psychometrika, 2001).

3.9 Checking Conditions or Equivalence Theorems

Aim: To verify optimality of a design among ALL designs on X.

Let $(X, f(x), \lambda(x))$ be given. The design ξ^* is D-optimal if and only if

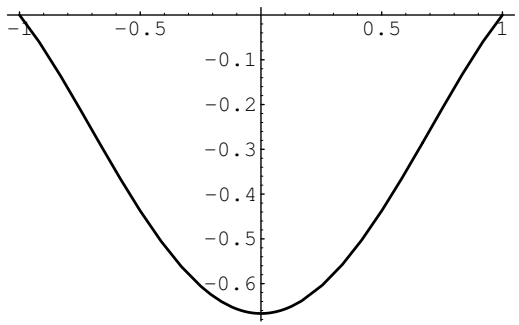
$$-F_\Phi(M(\xi^*), f(x)f^T(x))) = \lambda(x)f^T(x)M(\xi^*)^{-1}f(x) - d \leq 0 \quad for\ \forall\ x \in X,$$

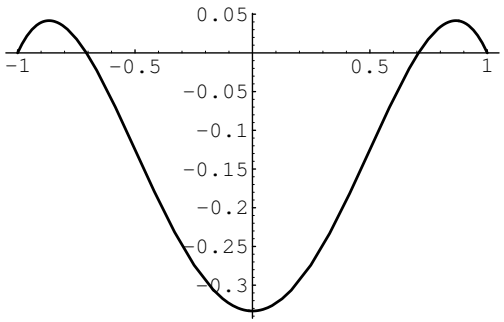
with equality at the support points of ξ^*. Here d is the dimension of f(x). Draw a picture to verify optimality!

An Example:

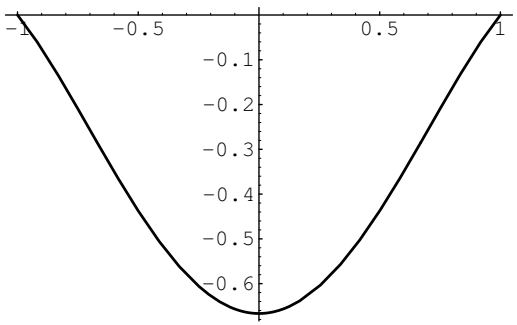
Suppose $f^T(x) = (1, x), X = [-1, 1]$ and $\lambda(x) = c - x^2$, $c > 1$. Is the design equally supported at ± 1 D-optimal for all $c > 1$?

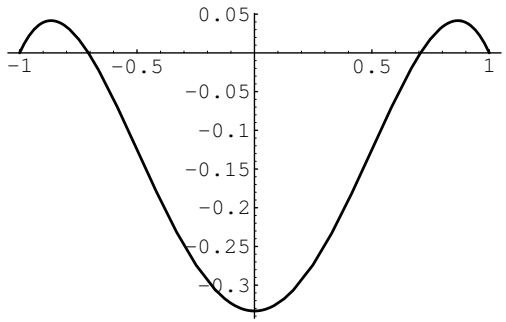
3.10: Plots of the directional derivative for the equally weighted design at ± 1 when $\lambda(x) = 4 - x^2$ (left) and when $\lambda(x) = 2.5 - x^2$ on $X = [-1, 1]$.





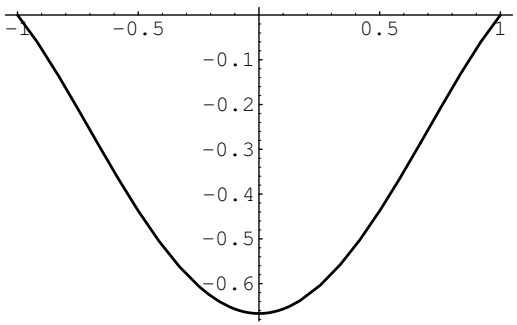
3.10: Plots of the directional derivative for the equally weighted design at ± 1 when $\lambda(x) = 4 - x^2$ (left) and when $\lambda(x) = 2.5 - x^2$ on $X = [-1, 1]$.

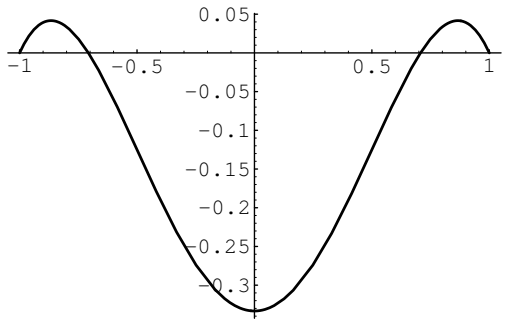




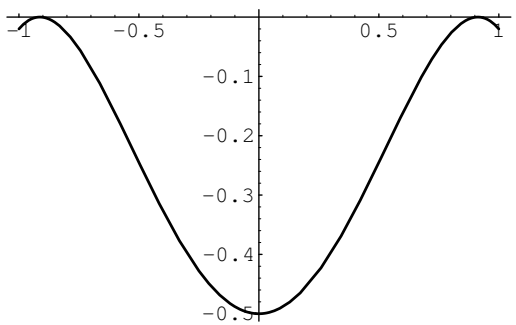
- The D-optimal design is equally supported at ± 1 if $c \geq 3$; otherwise it is equally supported at $\pm\sqrt{c/3}$ ($c = 2.5$ below).

3.10: Plots of the directional derivative for the equally weighted design at ± 1 when $\lambda(x) = 4 - x^2$ (left) and when $\lambda(x) = 2.5 - x^2$ on $X = [-1, 1]$.





- The D-optimal design is equally supported at ± 1 if $c \geq 3$; otherwise it is equally supported at $\pm\sqrt{c/3}$ ($c = 2.5$ below).



3.11 Basic Design Strategy for Multiple Objectives

- Constrained Optimal Designs
  i.e. design that satisfies a set of user-specified efficiency requirements; eg. minimize $\phi_2(\xi)$ subject to $\phi_1(\xi) \leq c$.

- Compound Optimal Designs
  i.e. design that minimizes a fixed convex combination of convex functionals: $\phi(\xi|\lambda) = \lambda\phi_1(\xi) + (1-\lambda)\phi_2(\xi)$.

- Compound Optimal Designs are equivalent to Constrained Optimal Designs: Plot efficiencies of each compound optimal versus λ, $\lambda \in [0, 1]$.

- In practice, prioritize the importance of the objectives and apply theory for single-objective study, Cook & Wong , JASA (1994), Wong, Statistica Neerlandica (1999), Huang & Wong, Drug Information Journal (2004)

3.12 Efficiency Plots for Dual-Objective Optimal Designs

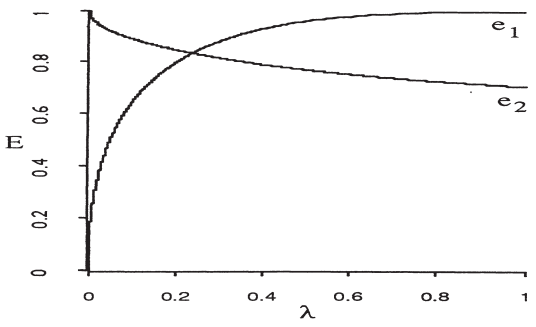


Figure 1. Efficiencies E Versus λ for Example 1.

Further, using (9), we obtain the values given in the first two lines of table 1 in Studden (1982).

To connect explicitly the constrained design problems and the compound design problems, the relationship between a constraint expressed as $E_1(\xi) \geq e_1$ (or $E_2(\xi) \geq e_2$) and the corresponding value of λ must be established. Using results from Fedorov (1980, thm. 1), it can be shown that ξ_λ maximizes $\phi(\xi|\lambda)$ if and only if

$$2(1-\lambda)d_1(x,\xi_\lambda) + \lambda\{b^T M_2^{-1}(\xi_\lambda)f_2(x)\}^2 - 4(1-\lambda) - \lambda b^T M_2^{-1}(\xi_\lambda)b \leq 0 \quad (11)$$

for all x in $[-1, 1]$. Further (11) becomes an equality at the support points for ξ_λ. In that case, substituting (9) into (11) yields

$$\lambda = \frac{e_1^2}{4 + 4(1-e_1)^{1/2} - 4e_1 + e_1^2}. \quad (12)$$

Figure 1, constructed using (10) and (12), shows the relationship between optimal designs with efficiency constraints and compound optimal designs. For example, a design that maximizes ϕ_2 subject to the constraint $E_1(\xi) \geq .6$ can be found by maximizing $\phi(\xi|\lambda)$ with $\lambda \approx .1$. Figure 1 contains useful information on the interpretation of λ as well. In particular, it might be felt that setting $\lambda = .5$ would yield a design in which equal interest is placed on the two criteria. But from Figure 1, the compound design problem with $\lambda = .5$ is equivalent to the constrained problem in which we maximize ϕ_2 subject to the constraint that $E_1(\xi) \geq .96$. The resulting constrained design has $E_2(\xi_{\lambda=.5}) \approx .78$. In terms of the efficiencies, placing equal interest on the two criteria would seem to require $\lambda = .25$, because at that point $E_1(\xi_\lambda) \approx E_2(\xi_\lambda) \approx .84$. Finally, reconstructing the plot in Figure 1 so that the horizontal axis is $1 - \lambda$ rather than λ provides the corresponding plot for maximizing ϕ_1 subject to a constraint on ϕ_2.

Example 2. For the simple linear regression model $f_1^T(x) = (1, x)$ on $\mathcal{X} = [-1, 1]$, consider balancing A optimality with precise estimation of the response at the point $z = .75$:

$$\phi_1(\xi) = -d_1(z,\xi)/d_1(z,\xi^1) = -[E_1(\xi)]^{-1}$$

and

$$\phi_2(\xi) = -\text{tr}\, M_1^{-1}(\xi)/\text{tr}\, M_1^{-1}(\xi^2) = -[E_2(\xi)]^{-1}.$$

The design ξ^1 is optimal for ϕ_1 and has the minimum variance possible for a fitted value at the point z. This minimum variance is the same as that obtained under the design that places mass 1 at z. The design ξ^2 is the A-optimal design. Both ξ^1 and ξ^2 are supported at ± 1, with the masses at 1 being $\frac{1}{8}$ and $\frac{1}{2}$. Thus $d_1(z,\xi^1) = 1$ and $\text{tr}[M_1^{-1}(\xi^2)] = 2$.

Define $w(x) = (9x^2 - 25x + 16)^{1/2}$ and $g(x) = (3x + 4 - w(x))/8$ for $0 \leq x \leq 1$. Then

$$\xi_\lambda(1) = g(E_1(\xi_\lambda)), \quad (13)$$

which is the analog of (9) for this example. Next, again from Fedorov (1980, thm. 1) ξ_λ satisfies

$$\lambda(f_1^T(x)M_1^{-1}(\xi_\lambda)f_1(z))^2 + (1-\lambda)f_1(x)M_1^{-2}(\xi_\lambda)f_1(x) = \lambda d_1(z,\xi_\lambda) + (1-\lambda)\text{tr}\, M_1^{-1}(\xi_\lambda)$$

at the support points $x = \pm 1$. Substituting (13) into this equation yields

$$\lambda = \frac{8(w(e_1) - 3e_1)}{34g(e_1) - 17 - 48g^2(e_1)},$$

which is the analog of (12) for $e_1 \geq .64$. From this we constructed Figure 2. The relationships in Figure 2 are, of course, qualitatively similar to those in Figure 1. But note that the value of λ at which the efficiencies are equal is much larger than that for Figure 1. Generally, the interpretation of λ depends heavily on the functionals involved. Useful inter-

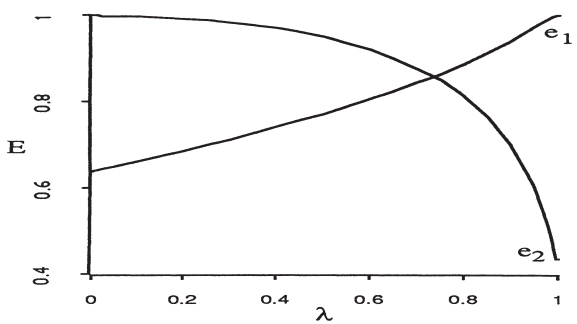


Figure 2. Efficiencies E Versus λ for Example 2.

3.13 Design Efficiency - what it means exactly?

Efficiency of a design ξ is defined relative to the optimal design ξ^*.

For estimating θ_1 in the simple linear model: $eff(\xi) = \frac{var_{\xi^*}(\hat{\theta}_1)}{var_{\xi}(\hat{\theta}_1)}$

Interpretation: If the efficiency of ξ is 0.5, ξ needs to be replicated twice to do as well as ξ^*.

3.13 Design Efficiency - what it means exactly?

Efficiency of a design ξ is defined relative to the optimal design ξ^*.

For estimating θ_1 in the simple linear model: $eff(\xi) = \frac{var_{\xi^*}(\hat{\theta}_1)}{var_{\xi}(\hat{\theta}_1)}$

- For estimating both parameters, define efficiency by

Interpretation: If the efficiency of ξ is 0.5, ξ needs to be replicated twice to do as well as ξ^*.

3.13 Design Efficiency - what it means exactly?

Efficiency of a design ξ is defined relative to the optimal design ξ^*.

For estimating θ_1 in the simple linear model: $eff(\xi) = \frac{var_{\xi^*}(\hat{\theta}_1)}{var_{\xi}(\hat{\theta}_1)}$

- For estimating both parameters, define efficiency by
- A-efficiency = $\frac{var_{\xi^*}(\hat{\theta}_0) + var_{\xi^*}(\hat{\theta}_1)}{var_{\xi}(\hat{\theta}_0) + var_{\xi}(\hat{\theta}_1)}$

Interpretation: If the efficiency of ξ is 0.5, ξ needs to be replicated twice to do as well as ξ^*.

3.13 Design Efficiency - what it means exactly?

Efficiency of a design ξ is defined relative to the optimal design ξ^*.

For estimating θ_1 in the simple linear model: $eff(\xi) = \frac{var_{\xi^*}(\hat{\theta}_1)}{var_{\xi}(\hat{\theta}_1)}$

- For estimating both parameters, define efficiency by
- A-efficiency = $\frac{var_{\xi^*}(\hat{\theta}_0) + var_{\xi^*}(\hat{\theta}_1)}{var_{\xi}(\hat{\theta}_0) + var_{\xi}(\hat{\theta}_1)}$
- D-efficiency = some ratio of the areas of the two confidence regions from the two designs

Interpretation: If the efficiency of ξ is 0.5, ξ needs to be replicated twice to do as well as ξ^*.

4 Methods for Searching Optimal Designs

- Brute force and guess work
- A purely theoretical approach
- Algorithms

4.1 By brute force: find a minimax D-optimal design

Consider the Logistic Model on a given interval X. Let $\theta^T = (\theta_1, \theta_2) \in \Theta$, Θ known and let Ξ be the set of all designs on X.

Design Criterion: Find ξ^* = arg $\min_{\xi \in \Xi} \max_{\theta \in \Theta}$ *log* $|M(\xi, \theta)|^{-1}$.

4.1 By brute force: find a minimax D-optimal design

Consider the Logistic Model on a given interval X. Let $\theta^T = (\theta_1, \theta_2) \in \Theta$, Θ known and let Ξ be the set of all designs on X.

Design Criterion: Find ξ^* = arg $\min_{\xi \in \Xi} \max_{\theta \in \Theta}$ *log* $|M(\xi, \theta)|^{-1}$.

- No known algorithms that will generate a minimax optimal design.

4.1 By brute force: find a minimax D-optimal design

Consider the Logistic Model on a given interval X. Let $\theta^T = (\theta_1, \theta_2) \in \Theta$, Θ known and let Ξ be the set of all designs on X.

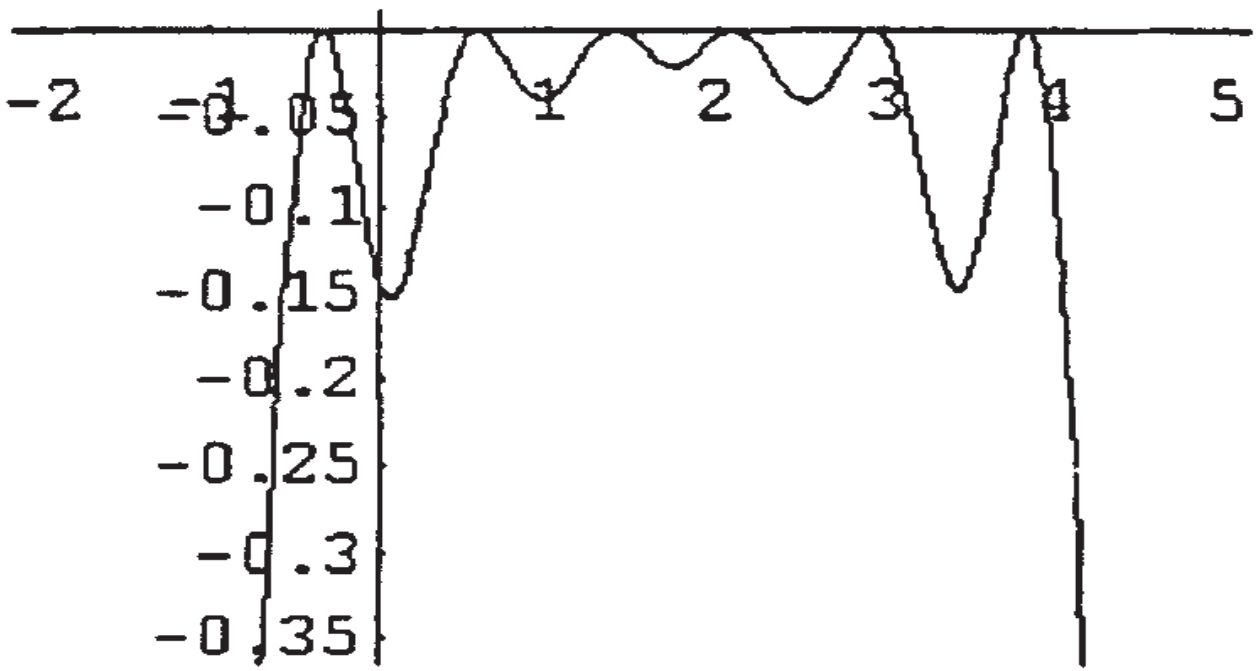
Design Criterion: Find ξ^* = arg $\min_{\xi \in \Xi} \max_{\theta \in \Theta}$ *log* $|M(\xi, \theta)|^{-1}$.

- No known algorithms that will generate a minimax optimal design.
- King & Wong (Biometrics, 2002) found minimax D-optimal designs when $\Theta = [0, 3.5] \times [1, 3.5]$ and X is unrestricted:

x_i	-0.35	0.62	1.39	2.11	2.88	3.85
w_i	0.18	0.21	0.11	0.11	0.21	0.18

4.1 Equivalence Plot confirms the Minimax D-Optimal Design for the Logistic Model

1266 *Biom*



4.2 A Purely Theoretical Approach (is limiting)

D-optimal designs for mixture experiments using cubic polynomial models without 3-way effect (Mikaeili, JSPI, 1989) and with 3-way effect (Mikaeili, JSPI, 1993) on the regular q-simplex:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{1 \le i < j \le q} \beta_{ij} x_i x_j + \sum_{1 \le i < j \le q+1} \gamma_{ij} x_i x_j (x_i - x_j). \quad (1)$$

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{1 \le i < j \le q} \beta_{ij} x_i x_j + \sum_{1 \le i < j \le q} \gamma_{ij} x_i x_j (x_i - x_j) + \sum_{1 \le i < j < k \le q+1} \beta_{ijk} x_i x_j x_k. \quad (2)$$

4.2 A Purely Theoretical Approach (is limiting)

D-optimal designs for mixture experiments using cubic polynomial models without 3-way effect (Mikaeili, JSPI, 1989) and with 3-way effect (Mikaeili, JSPI, 1993) on the regular q-simplex:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{1 \le i < j \le q} \beta_{ij} x_i x_j + \sum_{1 \le i < j \le q+1} \gamma_{ij} x_i x_j (x_i - x_j). \quad (1)$$

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{1 \le i < j \le q} \beta_{ij} x_i x_j + \sum_{1 \le i < j \le q} \gamma_{ij} x_i x_j (x_i - x_j) + \sum_{1 \le i < j < k \le q+1} \beta_{ijk} x_i x_j x_k. \quad (2)$$

- Proofs involve lengthy tedious algebra that exploit symmetry of the design problem.

4.2 A Purely Theoretical Approach (is limiting)

D-optimal designs for mixture experiments using cubic polynomial models without 3-way effect (Mikaeili, JSPI, 1989) and with 3-way effect (Mikaeili, JSPI, 1993) on the regular q-simplex:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{1 \le i < j \le q} \beta_{ij} x_i x_j + \sum_{1 \le i < j \le q+1} \gamma_{ij} x_i x_j (x_i - x_j). \quad (1)$$

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{1 \le i < j \le q} \beta_{ij} x_i x_j + \sum_{1 \le i < j \le q} \gamma_{ij} x_i x_j (x_i - x_j) + \sum_{1 \le i < j < k \le q+1} \beta_{ijk} x_i x_j x_k. \quad (2)$$

- Proofs involve lengthy tedious algebra that exploit symmetry of the design problem.
- Method fails when certain terms are deleted from the model.

4.3 Need for Algorithms To Search for Optimal Designs

- Derivation of optimal designs for nonlinear models is tedious, difficult and method for one model does not usually generalize to another

4.3 Need for Algorithms To Search for Optimal Designs

- Derivation of optimal designs for nonlinear models is tedious, difficult and method for one model does not usually generalize to another
- Formulae for optimal designs rarely exist and if they do, they are complicated and frequently unhelpful to the practitioners

4.3 Need for Algorithms To Search for Optimal Designs

- Derivation of optimal designs for nonlinear models is tedious, difficult and method for one model does not usually generalize to another
- Formulae for optimal designs rarely exist and if they do, they are complicated and frequently unhelpful to the practitioners
- Algorithms are very helpful - available only for some types of optimal designs

4.4 Fedorov's Algorithm for D-optimality (1972)

- Let $k = 0$ and let $\triangle$ be a user-selected precision level, say 0.001.
- Start with a design ξ_k with n support points and $n \geq d$ (dim of f(x)).
- Find information matrix of ξ_k : $M(\xi_k) = \sum_{i=1}^{n} w_i \lambda(x_i) f(x_i) f^T(x_i)$ (*)
- Find $x_k \in X$ that satisfies

$$\lambda(x_k) var(x_k, \xi_k) = max_{x \in X} \lambda(x) var(x, \xi_k) = m_k.$$

- Let $\delta_k = m_k - d$. Stop if $\delta_k < \triangle$ and declare ξ_k is D-optimal (numerically). Otherwise, let ξ^{x_k} be the point mass design at x_k and form a new design ξ_{k+1}:

$$\xi_{k+1} = (1 - p_k)\xi_k + p_k \xi^{x_k}, \qquad p_k = \frac{\delta_k}{(\delta_k + (d-1))d} \in (0, 1).$$

- Set k=k+1, go to (*) and repeat the process until $\delta_k < \triangle$.

4.5 Algorithms: Issues

4.5 Algorithms: Issues

- Some algorithms can be proved to converge to the optimum

4.5 Algorithms: Issues

- Some algorithms can be proved to converge to the optimum
- Examples are Fedorov's algorithm and related algorithms for finding c, L-optimal designs

4.5 Algorithms: Issues

- Some algorithms can be proved to converge to the optimum
- Examples are Fedorov's algorithm and related algorithms for finding c, L-optimal designs
- Convergence is usually slow and may not work for more complicated models

4.5 Algorithms: Issues

- Some algorithms can be proved to converge to the optimum
- Examples are Fedorov's algorithm and related algorithms for finding c, L-optimal designs
- Convergence is usually slow and may not work for more complicated models
- Periodic collapsing of cluster of points into fewer points is required

4.5 Algorithms: Issues

- Some algorithms can be proved to converge to the optimum
- Examples are Fedorov's algorithm and related algorithms for finding c, L-optimal designs
- Convergence is usually slow and may not work for more complicated models
- Periodic collapsing of cluster of points into fewer points is required
- Algorithms may not work for singular optimal designs, designs with fewer than d points

4.5 Algorithms: Issues

- Some algorithms can be proved to converge to the optimum
- Examples are Fedorov's algorithm and related algorithms for finding c, L-optimal designs
- Convergence is usually slow and may not work for more complicated models
- Periodic collapsing of cluster of points into fewer points is required
- Algorithms may not work for singular optimal designs, designs with fewer than d points
- Is there an easy-to-use and efficient method for finding optimal designs for different types of optimal designs for any given model?

4.6 Basic Tools to Construct and Study Optimal Design

Knowledge in matrix algebra and matrix derivatives are helpful. Exemplary tools include results from the simultaneous diagonalization theorem and typical results from matrix derivatives:

Suppose the elements of each of the matrices A and B are functions of a scalar t. Assume A is non-singular.

(i) $\frac{dAB}{dt} = \frac{dA}{dt} + A\frac{dB}{dt}$

(ii) $\frac{dA^{-1}}{dt} = -A^{-1}\frac{dA}{dt}A^{-1}$

(iii) $\frac{d\ ln\ |A|}{dt} = -trA^{-1}\frac{dA}{dt}$

For proofs of these results and many other interesting related results, see the monograph by Rogers, G. S, 1980: Matrix Derivatives. Lecture Notes in Statistics, Vol. 2, Dekker.

In Praise of Simplicity not Mathematistry! Ten Simple Powerful Ideas for the Statistical Scientist

Roderick J. LITTLE

Ronald Fisher was by all accounts a first-rate mathematician, but he saw himself as a scientist, not a mathematician, and he railed against what George Box called (in his Fisher lecture) "mathematistry." Mathematics is the indispensable foundation of statistics, but for me the real excitement and value of our subject lies in its application to other disciplines. We should not view statistics as another branch of mathematics and favor mathematical complexity over clarifying, formulating, and solving real-world problems. Valuing simplicity, I describe 10 simple and powerful ideas that have influenced my thinking about statistics, in my areas of research interest: missing data, causal inference, survey sampling, and statistical modeling in general. The overarching theme is that statistics is a missing data problem and the goal is to predict unknowns with appropriate measures of uncertainty.

KEY WORDS: Calibrated Bayes; Causal inference; Measurement error; Missing data; Penalized spline of propensity.

1. INTRODUCTION: THE UNEASY RELATIONSHIP BETWEEN STATISTICS AND MATHEMATICS

American Statistical Association President, Sastry Pantula, recently proposed renaming the Division of Mathematical Sciences at the U.S. National Science Foundation as the Division of Mathematical and Statistical Sciences. Opponents, who viewed statistics as a branch of mathematics, questioned why statistics should be singled out for special treatment.

Data can be assembled in support of the argument that statistics *is* different—for example, the substantial number of academic departments of statistics and biostatistics, the rise of the statistics advanced placement examination, and the substantial number of undergraduate statistics majors. But the most important factor for me is that statistics is not just a branch of mathematics. It is an inductive method, defined by its applications to the sciences and other areas of human endeavor, where we try to glean information from data.

The relationship between mathematics and statistics is somewhat uneasy. Since the mathematics of statistics is often viewed as basically rather pedestrian, statistics is rather low on the totem pole of mathematical subdisciplines. Statistics needs its mathematical parent, since it is the indispensable underpinning of the subject. On the other hand, unruly statistics has ambitions to reach beyond the mathematics fold: it comes alive in applications and medicine, and with increasing influence recently on the hard sciences such as astronomy, geology and physics.

The scientific theme of modern statistics fits the character of its most influential developer, the great geneticist, R. A. Fisher, who seemed to revolutionize the field of statistics in his spare time! Fisher's momentous move to Rothampsted Experimental Station rather than academia underlined his dedication to science. Though an excellent mathematician, Fisher viewed himself primarily as a scientist, and disparaged rivals like Neyman and Pearson as mere "mathematicians."

George Box's engaging Fisher lecture focused on the links between statistics and science (Box 1976). He wrote:

> My theme then will be first to show the part that [Fisher] being a good scientist played in his astonishing ingenuity, originality, inventiveness, and productivity as a statistician, and second to consider what message that has for us now.

Box attributed Fisher's hostility to mathematicians to distaste for what he called "mathematistry," which he defined as

> [. . .] the development of theory for theory's sake, which, since it seldom touches down with practice, has a tendency to redefine the problem rather than solve it. Typically, there has once been a statistical problem with scientific relevance but this has long since been lost sight of. (Box 1976)