

**Randomisation, Replication,
Response Surfaces
and
Rosemary**



A.C. Atkinson
a.c.atkinson@lse.ac.uk
Department of Statistics
London School of Economics
London WC2A 2AE, UK

RAB AND ME

- One joint publication
- “One hundred years of the design of experiments on and off the pages of *Biometrika*”. Bka **88**, 53–97
- Perhaps this talk should be
- “Randomization, Replication, Response Surfaces and Rosemary”
- We found that the subject of the “Design of Experiments” pretty much fell into two disjoint halves, corresponding to our own interests
- Here are some examples:

AGRICULTURAL FIELD TRIAL

- **Objective**

Compare 113 varieties of wheat

- **Experimental Variables**

4 replicates of each variety divided between blocks

(fields, farms)

- **Response Variables**

Yield ↑

Rust resistance ↑

Straw ?

Resistance to flattening ↑

PHOTOGRAPHIC INDUSTRY

- **Objective**

Optimise sensitisation conditions for an emulsion (i.e. speed)

- **Experimental Variables**

Sensitisor I, Sensitisor II

Dye

Reaction Time

- **Response Variables**

Sensitivity ↑

Contrast ↑

Fog ↓

Graininess ↓

CLINICAL TRIALS

- **Objective**

Compare several treatments (drugs, therapies)

- **Experimental Variables**

Treatment allocation - A, B, C, ...

Concomitant variables/ prognostic factors

History of
Measurements on } patient

- **Response Variables**

Physical or chemical measurements

“Quality of life”

CONSUMER DURABLE: FRIDGE, CAR, CAMERA,..

- **Objective**

Make a “better” product

- **Experimental Variables**

Conditions of use have also to be included

Need “robustness” to abuse

TWO KINDS OF EXPERIMENT

'Agricultural' experiment:

- Physically distinct units;
- Complicated error structure associated with units, plots and blocks;
- Discrete treatments with no or little structure.

Industrial experiments are often also of this type.

'Industrial' experiment:

- Unit, for example, just another portion of chemical reagents, with no specific properties;
- Simple error structure;
- Complicated treatment structure - maybe the setting of several continuous variables.

FISHERIAN PRINCIPLES

‘**Agricultural**’ experiment (Fisher was at Rothamsted)

- Replication
- Randomization
- Local control
- Ideas of
 - Treatment Structure
 - Block Structure

‘**Industrial**’ experiment

- Gossett (1917) and later. Days as blocks in laboratory testing
- Box (1974 in BHH and earlier). “Block what you can and randomize what you cannot”

DESIGNING AN EXPERIMENT

At least three phases:

- (a) choice of treatments;
- (b) choice of experimental units; and
- (c) deciding which treatment to apply to which experimental unit.

The relative importance of the three phases depends on the application.

Agriculture and medicine: insufficient experimental units which are all alike, so some sort of **blocking** must be considered: divide experimental units into groups likely to be homogeneous.

In an industrial experiment a **block** might be a batch of raw material.

As soon as there is structure, phase (c) becomes important. In this context the ‘design’ is the function allocating treatments to units.

DESIGNING AN EXPERIMENT 2

(c). Allocation of treatments to units.

Agricultural Experiment. 113 treatments \times 4 replicates.

- With large blocks replicate 113 treatments at each of four sites. These may be very different - clay, sand, wet, marshy,....
- Analyse with additive block effects
- Blocks are usually too small for complete replication
- Design so that, if possible, each treatment occurs in the same number of blocks - balance
- Allocate treatments at random, subject to design. If design has treatments A, B, C, ... and you have treatments 1, 2, 3, ... Randomize mapping from numbers to alphabet.
- Again analyse with additive block effects.

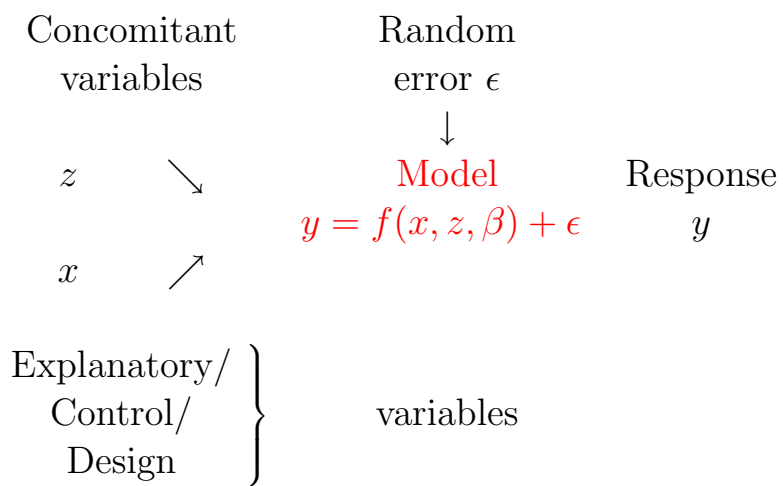
DESIGNING AN EXPERIMENT 3

(c). Allocation of treatments to units.

Industrial Experiment. Treatments are sensitiser, dye, reaction time

- Units are just another batch of chemical
- But there may be technical analytical variables: development conditions, time between manufacture and photographic testing, batches of material, shift of operators, ...
- Brien and Bailey (2006) “Multiple Randomizations”
- Should be allowed for in design as extra factors
- Randomize order of experiments if possible to avoid confounding with ‘lurking’ variables

COMMON STRUCTURE



- **Experimental Design** Choose x

ANALYSIS

- Second-order assumptions:

$$y_i = f(x_i, \beta) + \epsilon_i$$

with

$$E(\epsilon_i) = 0 \quad \text{var}(\epsilon_i) = \sigma^2$$

lead to least squares

- Additional normality of ϵ_i gives distribution of estimates and test statistics
- Randomization Analysis
 - BHH have yields of 11 tomato plants (?) six on treatment A, five on B. Calculate $\bar{y}_A - \bar{y}_B$ for the data. Also for all $11!/5!6!$ permutations and compare observed value with this reference distribution
 - Not much used for linear models, but may be for clinical trials.

SMITH 1918

A remarkable paper, 35 years ahead of its time.

K. Smith. “On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations”. Bka 12, 1–85.

- Polynomials up to order six in one variable
- Design region $\mathcal{X} : -1 \leq x \leq 1$
- Additive errors of constant variance, so least squares giving predictions $\hat{y}(x)$
- “G-optimum” designs minimizing maximum, for $x \in \mathcal{X}$, of $\text{var}\{\hat{y}(x)\}$
- Also considered non-constant variance
- Showed the designs were optimum

SMITH, PEARSON AND FISHER

- Kirstine Smith (b. 1878) was a Danish pupil of Hald who worked with K. Pearson
- Her 1916 paper “On the “best” values of the constants in frequency distributions” upset Fisher - he thought parameter estimates shouldn't depend on grouping: ML rather than minimum chi-squared
- Pearson twice refused Fisher's paper “I must keep the little space I have in *Biometrika* free from controversy”
- Fisher never again (?) published in *Biometrika*
- In 1924 Smith became a school teacher in Denmark - “because she felt a need to work more closely with people”

OPTIMUM EXPERIMENTAL DESIGN 1

- **Unified Theory** providing
- **Numerical Methods** for design construction and comparison
- Modern theory developed by Kiefer

JACK KIEFER

- Kiefer (1959). “Optimum experimental designs (with discussion)” JRSSB
- Proposer of the vote of thanks (Tocher) “I think Dr Kiefer has shown conclusively just how useful mathematics is in the theory of design...”
- No comment from Box!
- Kiefer in reply “I must admit being somewhat disappointed to see that such a large proportion of the comments have been engendered by a careless reading of my paper...”
- David Cox arranged a conference in 1974 at Imperial College at which Kiefer was the main speaker
- Papers published in *Biometrika* 1975

OPTIMUM EXPERIMENTAL DESIGN 2

Requires:

- **A model**

Many models are **linear**

$$y = \beta^T f(x) + \epsilon$$

$f(x)$ is a $p \times 1$ vector of powers and products of x

The errors ϵ are independent, with variance σ^2

In matrix form

$$E(y) = F\beta$$

Some models are **nonlinear** in the **parameters**

$$y = 1 - e^{-\theta x} + \epsilon$$

OPTIMUM EXPERIMENTAL DESIGN 3

Also requires:

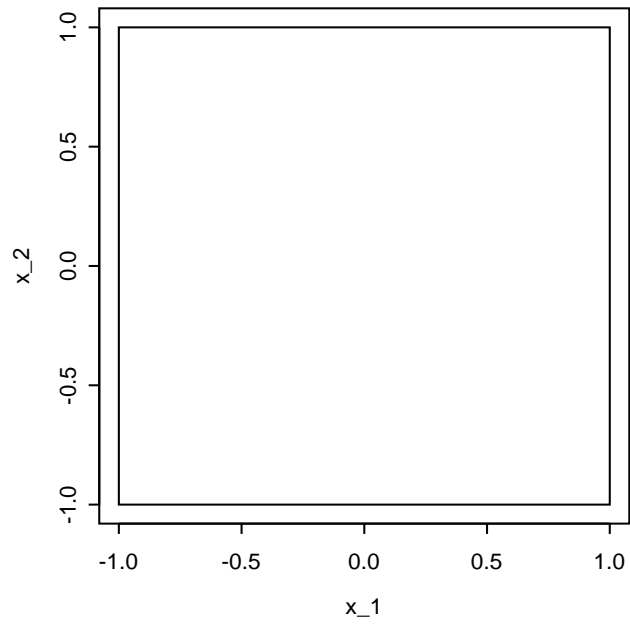
- **A design region \mathcal{X}**

With two factors x_1 and x_2 the square

$$-1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1$$

- The scaling is for convenience: for temperature could be:
 $-1; 20^\circ C ; 1; 40^\circ C$

Figure 1: A square design region



OPTIMUM EXPERIMENTAL DESIGN 4

An experimental design is a set of n points in \mathcal{X}

- **Design Measure**

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_m \\ w_1 & w_2 & \dots & w_m \end{array} \right\}$$

- The design has

- Support points $x_1 \dots x_m$
- design weights $w_1 \dots w_m$

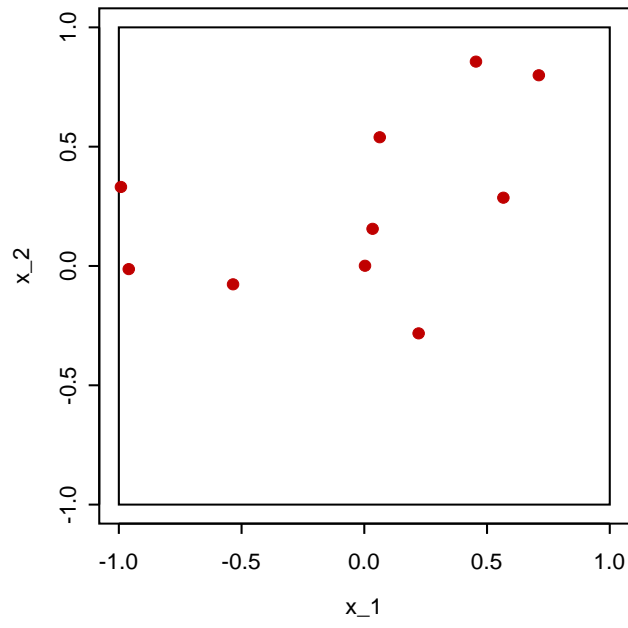
- **Exact design**

$$w_i = r_i/n,$$

r_i # replicates at x_i

- **Continuous/ Approximate Design** ξ is a measure over the design points. Removes dependence of design on n .

Figure 2: A random design. Any good?



INFERENCE ABOUT BETA

- The **Least Squares Estimator** of β from the design ξ is

$$\hat{\beta} = (F^T W F)^{-1} F^T W Y,$$

where

$$W = \text{diag } w_i$$

- The $100(1 - \alpha)\%$ **Confidence Region** for β is the set of values for which

$$(\beta - \hat{\beta})^T F^T W F (\beta - \hat{\beta}) \leq p s^2 F_{p, \nu, 1-\alpha}$$

- Confidence regions are **ellipsoids**
- **Volume** of confidence **ellipsoids**

$$\propto |F^T W F|^{-1/2}$$

THE INFORMATION MATRIX

- The **Information Matrix** $F^T W F$ is often written

$$F^T W F = M(\xi)$$

The information matrix depends on the design

- **D-optimum** designs

$$\max |F^T W F| = \det M(\xi^*)$$

They minimise the volumes of the confidence region

- Other **alphabetical** criteria maximize other functions of $M(\xi)$
- Often **maximise ξ numerically** to find ξ^*

THE PREDICTED RESPONSE

- **Prediction** at x is

$$\hat{y}(x) = \hat{\beta}^T f(x)$$

- The **variance of prediction** is

$$\text{var}\{\hat{y}(x)\} = n\sigma^2 f^T(x)M^{-1}(\xi)f(x)$$

- The **standardised variance** is

$$d(x, \xi) = f^T(x)M^{-1}(\xi)f(x)$$

GENERAL EQUIVALENCE THEOREM 1

– **Links**

$$\begin{array}{cc} \text{var}(\hat{\beta}) & \text{var}\{\hat{y}(x)\} \\ \text{parameters} & \text{predictions} \end{array}$$

– Provides **Algorithm**

Kiefer and Wolfowitz (1959)
(with a check on convergence)

GENERAL EQUIVALENCE THEOREM 2

- **Definition**

$$\xi^* \text{ maximises } \det M(\xi)$$

- **Minimax Equivalence** ξ^* minimises

$$\max_{x \in \mathcal{X}} d(x, \xi)$$

-

$$\max_{x \in \mathcal{X}} d(x, \xi^*) = p$$

So we can **check** any purported optimum design

- **Maximum** is at points of support of ξ^*

SECOND-ORDER POLYNOMIAL

- **Model**

$$E(y) = \beta_0 + \beta_1x + \beta_2x^2$$

- **Design Region**

$$\mathcal{X} \quad x \in [-1, 1]$$

- **Compare Two Designs**

Remember - designs are standardized for n

Figure 3: Two designs. Which is better?

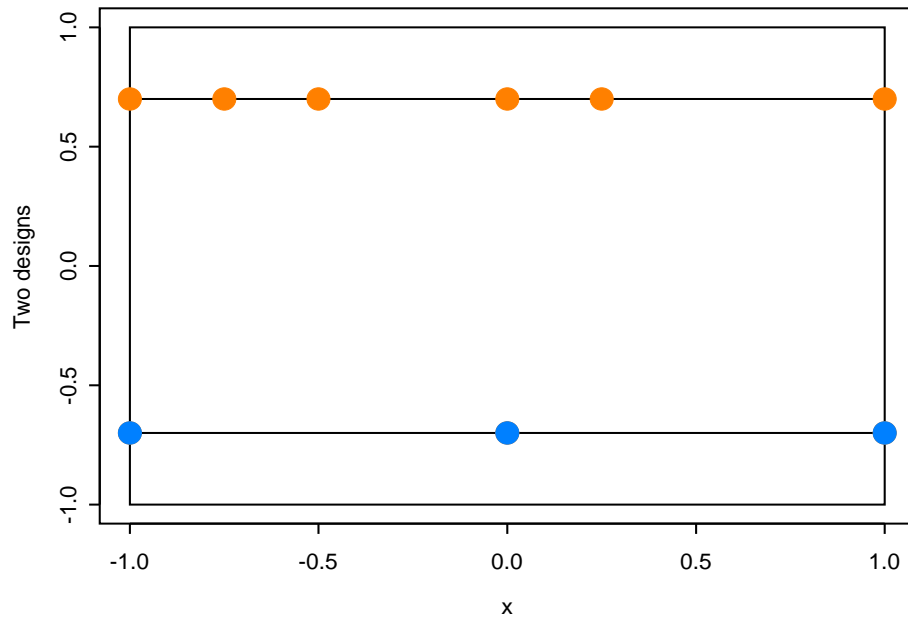
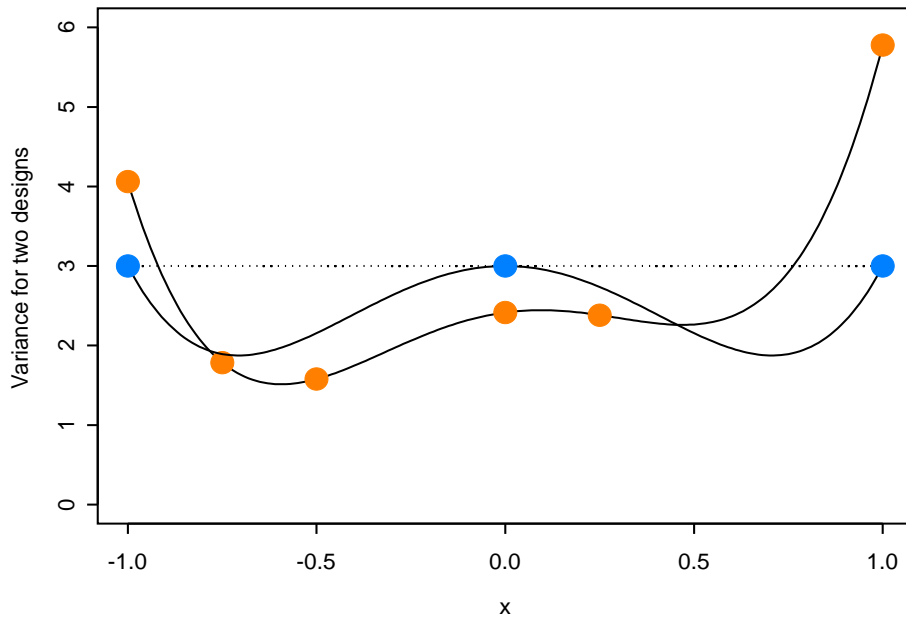


Figure 4: Variances for the two designs



COMPARISON OF DESIGNS

- **Three-Point Design**

Clearly D-optimum

$$\max_{x \in \mathcal{X}} d(x, \xi) = 3$$

and occurs at design points

- As found by Smith

- **Six-Point Design**

Not D-optimum

– $d(x, \xi) > 3$ at $x = -1, 1$

– < 3 at other 4 points

OPTIMUM EXPERIMENTAL DESIGN 5

Dependence on n

- Can always (?) maximize $\det M(\xi)$
- **Exact** design - i.e. for particular n may not satisfy General Equivalence Theorem
- What about $n = 4$ for quadratic model?
 - D-optimum design has three points of support, $-1, 0$ and 1 with one replicated.
 - G-optimum design is symmetric with four points of support

SOME APPLICATIONS

RESPONSE SURFACE DESIGNS 1

- Two Factors

- **Model**

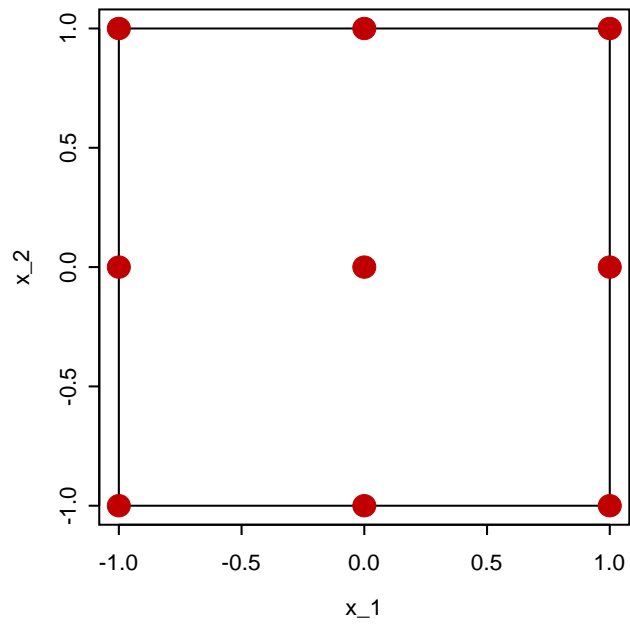
$$E(y) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{12}x_1x_2$$

- **Design Region**

$$\mathcal{X} \quad x \in [-1, 1] \times [-1, 1]$$

- There are 6 parameters
- The 3^2 factorial is often used

Figure 5: The 3^2 factorial



RESPONSE SURFACE DESIGNS 2

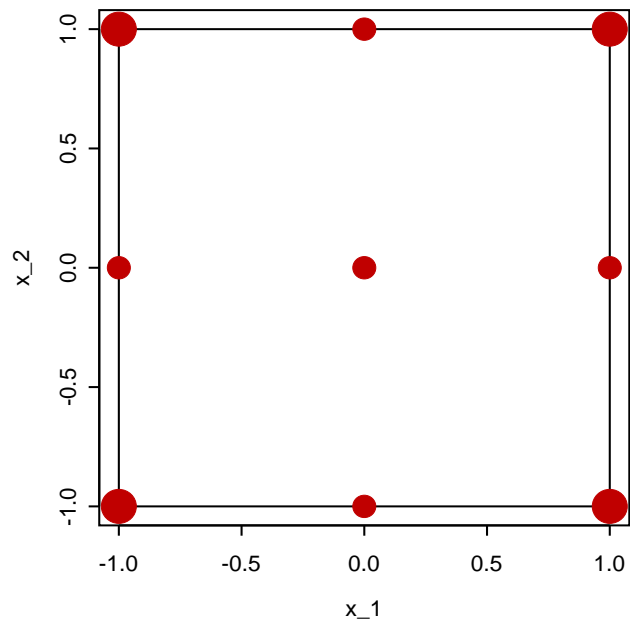
- 3^2 factorial has D-efficiency of 97.4%
- **D-efficiency**

$$D_{\text{eff}} = \{\det M(\xi) / \det M(\xi^*)\}^{1/p},$$

equivalent to number of trials

- **13-point design** with replicated corner points:

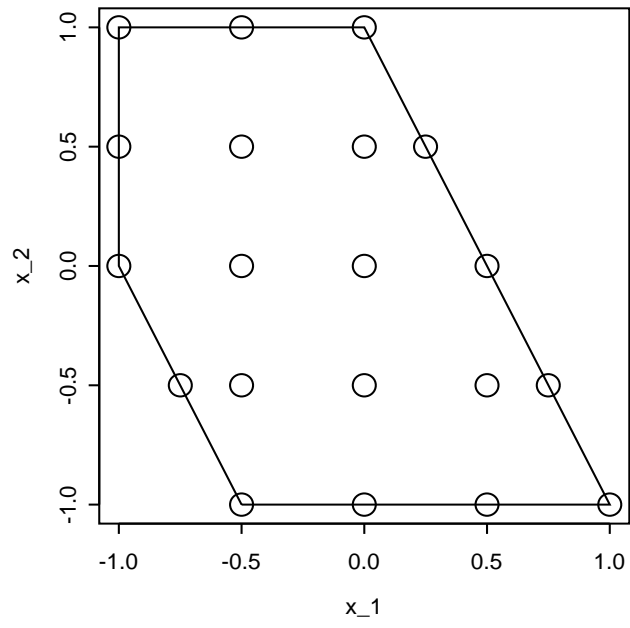
Figure 6: An almost optimum 13-trial design, with a D-efficiency of virtually 100%



RESPONSE SURFACE DESIGNS 4

- **Constrained regions** One real problem is that the design region may be irregular.
 - **Photographic experiment**
 - x_1 : sensitisor
 - x_2 : speed
 - high and low values of both variables may not be possible

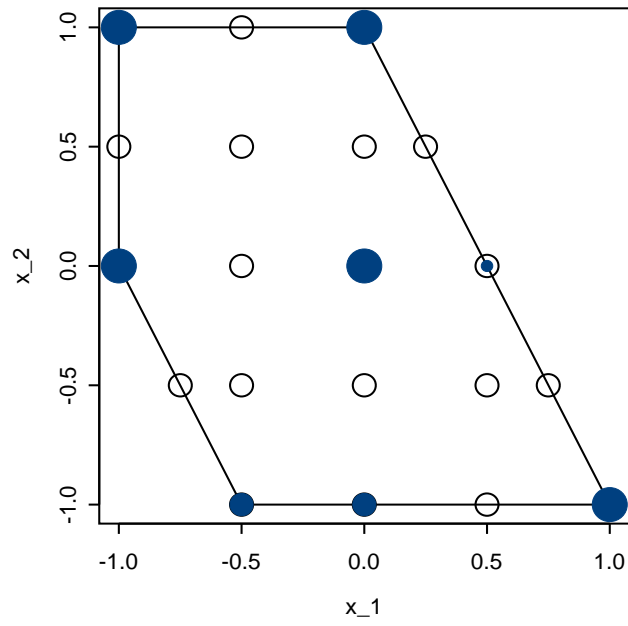
Figure 7: Irregular design region with 20 candidate points



IRREGULAR DESIGN REGION 1

- **Model** Full second-order
- **Support points** There are 20 possible sets of conditions for an experiment
- Let $n = 20$
- Which points are best? Use all 20?

Figure 8: Irregular design region with 20-point design



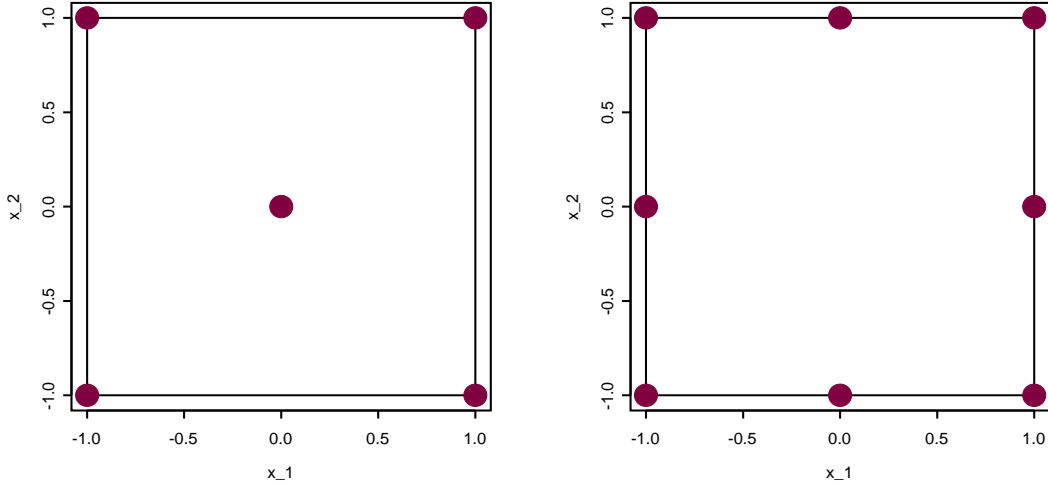
IRREGULAR DESIGN REGION 2

- $p = 6$
- $n = 20$
- Only 8 support points
- **Concentration** Designs are typically concentrated on a few sets of conditions
- Very difficult (impossible ?) to find good design without algorithm

BLOCKING

- Sometimes cannot perform all runs at once, or under same conditions
 - Different days, machines, raw material, ...
 - Industry as well as agriculture
- **Second-order model** $n=13$
 - Divide into two blocks
 - $n_1 = 5, n_2 = 8$, for example

Figure 9: The 13-trial design in blocks of 5 and 8



RELATED PROBLEMS

- **Qualitative and Quantitative Factors**
 - Quantitative factors, eg. 13-trial second-order design
 - Could repeat at each level of qualitative factor
 - Number of trials rapidly becomes excessive
 - Find optimum design splitting trials over qualitative levels
- **Mixture designs**
 - Constraints $\sum x_j = 1, 0 \leq x_j \leq 1$, give unintuitive models
 - Simplest design region \mathcal{X} is a simplex
 - Design region may be irregular

TREATMENT DESIGNS

- Optimum design theory has been illustrated for response surface designs
- Also applies to treatment designs
- Not just D-optimality, but A-, E-, G-, V-, functions of the eigenvalues of $M(\xi)$.
- Interest is in exact designs - Latin Squares, for example.
- “Universal Optimality”. Can’t do better than this whatever function of the parameters is of interest. Tends to be restrictive.

PHASE III CLINICAL TRIAL

- Patients arrive sequentially
- Each is immediately given one of t treatments
- Patient i also has a vector of prognostic factors x_i
- Main purpose is to find the best treatment
- Subsidiary purpose is to estimate the treatment effect
- Responses on patients before the i th may be available

REQUIREMENTS

Just consider the simplest case of the allocation of two treatments in the absence of prognostic factors

- The “design” allocates patient i to treatment A or B
- We want to allocate sequentially
- Balance over time - want n_A near n_B whenever we stop
- Blindedness
- Randomization: “objectivity”, avoidance of conscious or unconscious biases (Lanarkshire)
- The effect of randomization is slightly to unbalance the trial and increase the variance of the estimated treatment difference

DESIGNS

- Efron's biased coin. The probabilities of allocating treatment A depend on n_A :

$$\begin{array}{ll} n_A < n_B & p \\ n_A = n_B & 0.5 \\ n_A > n_B & 1 - p, \end{array}$$

for $0.5 < p \leq 1$

- An alternative is to use a permuted randomized block design of length 8 or 12
- Perhaps only **some** permutations of block designs should be considered. (Nelson and Bailey (2003) - Hadamard randomization)
- How to compare designs?

DESIGNS WITH COVARIATES

- Allocation probability depends both on previous allocations and on covariate vector x_{n+1} .
- Sequential construction of **optimum** design depends on variances $d(j, x_{n+1}, \xi_n), j = 1, \dots, t$
- ACA JRSSA (2002) suggests a randomized allocation with probability $\propto d(j, x_{n+1}, \xi_n)$.
- The ordering of the treatments by $d(\cdot)$ makes generalized Efron schemes possible.
- Also Bayes' rules.
- What about permuted randomized blocks?

THIS TALK

- History
- Agricultural and industrial experimentation
- Randomization and randomization analysis
- Smith, Kiefer and optimum design
- Response surfaces and extensions
- No mention of GLM's
- Nor of nonlinear models (Box and Lucas, 1959 Bka). Pharmaceutical industry
- Clinical trials and the comparison of designs

BIRTHDAYS

Application of Olkin's algorithm yields

$$60 \rightarrow 16$$