Cost Effective Screening Experiments

Sue Lewis
Southampton Statistical Sciences Research Institute
University of Southampton

ENBIS 7 Dortmund
24 September 2007

www.doe.soton.ac.uk/screening.php

The Screening Process

Screening is the process of sifting through a large number of features
eg factors, genes, chemical compounds

- to discover the few important features
- using designed experiments and statistical analysis

Outline

- Broad areas of application
  – blood screening, drug discovery, genetic screening

- Screening in industry and business

- Strategies for screening experiments
  – traditional methods
  – supersaturated
  – two-stage group screening

- Examples and a case study

See, for example, articles in
"Screening", 2006, eds. Dean & Lewis, Springer

Application: Blood Screening

- Screen a large population for disease by testing blood samples
- Pooling experiments to reduce the amount of costly testing (Dorfman, 1943)
- Widely used today
  - HIV antibody screening
  - discovering synergies between drugs

Application: Drug Discovery

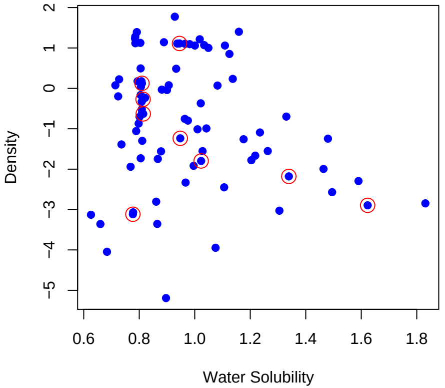
Screen huge numbers of compounds in chemical libraries to find

- novel, patentable compounds
- active against a particular disease (potency)
- good chemical properties, e.g. stable and low toxicity

– using a biological test or assay

Application: Drug Discovery

- Choose a subset of compounds for testing using descriptors
- Build a predictive model in the descriptors
  - to screen other compounds
- Reduces costs/time for finding new drugs



Application: Screening Human Genes

Genetic screening

- Aim: identify genes with a role in a biological process or disease
  – uses experiments in microarrays on tissue samples
  – measures the level to which genes are “in use” (expressed) in tissue cells

Example

- Compare samples of diseased and healthy tissue
- Identify genes with different expression levels
  – in the cells from the two tissue types

⇒ these genes may be involved in the disease process

Application: Industry and Business

- Aim: find the key factors that affect performance
- Experiment on factors at two levels +1, -1 (eg high, low)
- Response is approximated by a low order polynomial

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

  where X is the model matrix, $\boldsymbol{\beta}$ holds the unknown coefficients (parameters) and $\boldsymbol{\varepsilon}$ holds the independent $N(0, \sigma^2)$ errors

- Each factor is observed at each level the same number of times

Application: Industry and Business

Example: Cold start optimisation (Jaguar cars)

Vine, Lewis, Dean and Brunson (Technometrics, in press)

Factors Affecting Performance

Control factors - can be set by the engineers

Noise factors - cannot be controlled in use
eg ambient temperature

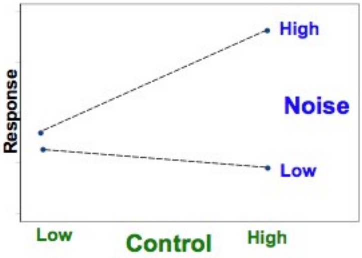
Response used to measure starting capability is "conditioned resistance"

Aim: to find control factor settings that

- Minimise the mean response
- Minimise variability

Interactions may be Important

- Control factors involved in control $\times$ noise interactions are key to reducing response variability



- Main effects and control $\times$ control interactions also of interest
- Interactions between three or more factors considered negligibly small

Application: Industry and Business

Example: Improvement in motor oil
Vermilya & Wilkinson; ASQ Fall Technical Conference, 2006

- Aim to find which factors have greatest effect on friction and corrosion of metal parts
- Scientists identified 40 factors – descriptors for different chemical constituents eg detergent, viscosity modifier
- Believed interactions between pairs of factors are likely to be small compared with main effects
- Could only afford a small number of runs

The Factor Screening Problem

Large number of potentially important factors

- Factors may or may not interact
- Factor sparsity – Box & Meyer (1986)
  – expect only a few factors have a large effect on response
- Need to discover by experiment the active or important factors
  – those that have substantial effects

Knowledge gained $\Rightarrow$ further experiments to build a detailed model for the response

Strategies and Designs for Screening Experiments

Well established approach

1. Run a small experiment to estimate only main effects
   - “main effects screening”
   - eg fractional factorials or Plackett and Burman designs
2. Select factors with large main effects
3. Use follow-up experimentation to allow estimation of interactions

See Box, Hunter & Hunter (2005) and references therein

Example of a Design for Main Effects Screening

Plackett and Burman design for up to 11 factors in 12 runs

Factors										
A	B	C	D	E	F	G	H	I	J	K
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1	-1	1	-1	-1	-1	1	1	1	-1	1
1	1	-1	1	-1	-1	-1	1	1	1	-1
-1	1	1	-1	1	-1	-1	-1	1	1	1
1	-1	1	1	-1	1	-1	-1	-1	1	1
1	1	-1	1	1	-1	1	-1	-1	-1	1
1	1	1	-1	1	1	-1	1	-1	-1	-1
-1	1	1	1	-1	1	1	-1	1	-1	-1
-1	-1	1	1	1	-1	1	1	-1	1	-1
-1	-1	-1	1	1	1	-1	1	1	-1	1
1	-1	-1	-1	1	1	1	-1	1	1	-1
-1	1	-1	-1	-1	1	1	1	-1	1	1

- Estimated main effects are uncorrelated when interactions are negligible
- Correlation between any pair of columns is zero

Lubrizol Example: 40 factors, main effects only

Can we use this type of design for the Lubrizol experiment?

- Need at least 41 runs to estimate all effects
- Smallest Plackett and Burman design for 40 factors has 44 runs
- Far too expensive
  – around 24 runs was acceptable

Supersaturated Designs

- Number of runs $<$ number of effects to be estimated
  $\Rightarrow$ cannot estimate all effects simultaneously
- First systematic development: Booth & Cox (1962)
  – interest revived by Lin (1993), Wu (1993)
- Mostly for main effects screening
- Case studies just beginning to emerge
  - validation and robustness testing in chemistry (Dejaegher et al., 2007)
  - need published experiences

Supersaturated Design for Main Effects Screening

Example: 10 factors in 6 runs (Lin, 1993)

Factors									
A	B	C	D	E	F	G	H	I	J
1	-1	1	-1	-1	-1	1	1	1	-1
-1	1	1	-1	1	-1	-1	-1	1	1
1	-1	1	1	-1	1	-1	-1	-1	1
1	1	-1	1	1	-1	1	-1	-1	-1
-1	-1	-1	1	1	1	-1	1	1	-1
-1	1	-1	-1	-1	1	1	1	-1	1

- Sum of cross-products between the columns are either -2 or 2
- Columns no longer uncorrelated
- Largest absolute correlation between pairs of columns is 0.33

$\Rightarrow$ Estimated main effects are correlated

Criteria for Good Supersaturated Designs

What makes one design better than another?

- Ideally, want to estimate main effects independently
- Not possible, as we cannot estimate all effects simultaneously
- Instead, designs have been sought with column correlations as small as possible

A Formal Criterion for Design Choice

- $E(s^2)$ criterion:

  Minimise the average of the squares of the cross-products, s_{ij}, between pairs of columns for f factors

  $$\text{minimise } E(s^2) = \frac{2}{f(f-1)} \sum_{i<j} s_{ij}^2$$

  Booth & Cox (1962)

- Example: Lin's design is $E(s^2)$-optimal
  – from the bound of Nguyen (1996)

Lin's Method from Plackett & Burman 12-run Design

- Choose any "branching column"
- Delete all rows corresponding to -1

				Factors						b.c.	
A	B	C	D	E	F	G	H	I	J	(K)	
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	—
1	-1	1	-1	-1	-1	1	1	1	-1	1	
1	1	-1	1	-1	-1	-1	1	1	1	-1	—
-1	1	1	-1	1	-1	-1	-1	1	1	1	
1	-1	1	1	-1	1	-1	-1	-1	1	1	
1	1	-1	1	1	-1	1	-1	-1	-1	1	
1	1	1	-1	1	1	-1	1	-1	-1	-1	—
-1	1	1	1	-1	1	1	-1	1	-1	-1	—
-1	-1	1	1	1	-1	1	1	-1	1	-1	—
-1	-1	-1	1	1	1	-1	1	1	-1	1	
1	-1	-1	-1	1	1	1	-1	1	1	-1	—
-1	1	-1	-1	-1	1	1	1	-1	1	1	

Supersaturated designs — Lin's method

- More generally, designs can be constructed in this way from Hadamard matrices of size $2n \times 2n$

- Can examine up to $2n - 2$ factors in n runs

  eg up to 46 factors in $n = 24$ runs

  up to 22 factors in $n = 12$ runs

- Hadamard matrices are available at www.research.att.com/$\sim$njas/hadamard

Lubrizol Experiment

For 40 factors in 24 runs

- Used Lin's method, starting with a Plackett and Burman design for 47 factors and 48 runs

- Obtained a design with 46 factors in 24 runs

- For 40 factors, 6 columns must be deleted
  – the choice of columns makes little difference to $E(s^2)$

Other Criteria for Selecting and Comparing Designs

- Minimise the maximum absolute correlation ($r_{\max}$) between columns

$$\text{i.e. minimise } \max_{i<j} |s_{ij}|$$

- Bayesian *D*-optimality: maximise the determinant of the posterior variance-covariance matrix of model parameters (Jones et al., 2007)

- Maximise the percentage of models that can be estimated using the design
  – Estimation Capacity (Sun, 1993)

An Even More Ambitious Example

Risk management in financial services
William Li (Minnesota); personal communication, 2007

- Find factors that most influence credit-worthiness
- 473 factors at 2 levels, eg average amount of debt
- Chose 120 of the 2^{473} possible combinations
  – using $E(s^2)$
  – and estimation capacity
- Data are currently being analysed

Techniques for Finding Supersaturated Designs

Mathematical construction methods include

- From Hadamard matrices: Wu (1993), Tang & Wu (1997)
- From block designs or cyclic structures: Nguyen (1996), Liu & Zhang (2000), Butler et al. (2001), Liu & Dean (2004), Eskridge et al. (2004), Liu et al. (2007)

Computer search methods include

- $E(s^2)$-optimality: Lin (1995), Nguyen (1996)
  – see http://designcomputing.net/gendex/noa/
- Bayesian D-optimality: Jones et al. (2007)
  – algorithm available in the JMP package

Analysis of Supersaturated Designs

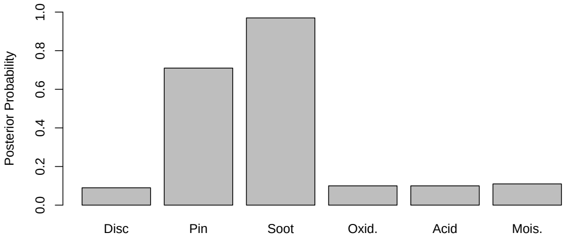
- Difficult: trying to extract a lot of information from a few observations
- Mistakes can be made
  - failing to detect one or more active factors
  - selecting a factor as active when it is not
- The use of “good” designs reduces this risk

Analysis of Supersaturated Designs

- Frequentist methods
  - Simple linear regression (Holcomb et al., 2003)
  - All subsets, forward selection and stepwise regression (Westfall et al., 1998; Abraham et al., 1999; Kelly & Voelkel, 2000)
  - Ridge regression, SCAD – penalised least squares (Lin, 1995; Li & Lin, 2002)
- Bayesian methods
  - Model selection – Markov Chain Monte Carlo (Chipman et al., 1997; Beattie et al., 2002)
  - Graphical variable assessment (Box & Meyer, 1986; Meyer & Wilkinson, 1998)

Bayesian Data Analysis Methods

Example: Experiment on surface wear between a pin and a disc
Tribology group, Southampton



- Dominant factors
  – addition of soot to the lubricant
  – type of pin material

Analysis of Supersaturated Designs

- Important effects can be identified provided they are very large and there are very few of them
- There is no "best" method of analysis
  – mistakes can occur with all methods
  – methods may select different sets of active effects
- Root of the problem is the correlation between the observed factor levels
- All subsets – computationally infeasible
- Forward selection – rough guide: $r_{\text{max}} < 0.33$ (Liu et al., 2007)
- Bayesian approach worked well in practice at Lubrizol

Interactions

Interactions can be ignored when

- main effects are much larger than interactions
- interactions occur only between factors with large main effects

Otherwise, interaction screening is possible using

- Some supersaturated design methods
  Wu (1993); Jones et al. (2007); Liu et al. (2007)
- Two-stage interaction group screening
  Lewis & Dean (2001); Vine et al. (2005)

Two-stage Interaction Group Screening

- Label factor levels

  high – larger response anticipated
  low – smaller response anticipated

- Arrange factors in groups

- For each group define a new grouped factor with two levels

  high – all factors in group high
  low – all factors in group low

- Experiment on the grouped factors

Example: 6 individual factors assigned to 3 groups

$$\begin{array}{lccc} & \{A, B\} & \{C, D\} & \{E, F\} \\ & \Downarrow & \Downarrow & \Downarrow \\ \text{Grouped Factors} & G_1 & G_2 & G_3 \end{array}$$

- Main effect of a grouped factor is the sum of main effects of factors in its group eg $G_1 = A + B$

- Similar expressions for interactions

  eg $G_1 G_2 = AC + AD + BC + BD$

- Issue of cancellation of effects
  – chance is small (factor sparsity)
  – investigate via simulation (Vine et al., 2005)

Two-stage Interaction Group Screening

Stage 1: perform an experiment on the grouped factors

- Estimate main effects, control$\times$control and control$\times$noise
- Decide which groups of factors are important

Stage 2: dismantle those groups found to be important and experiment on their individual factors

- Estimate the main effects and interactions of interest

Opinions Elicited from Experts

- Factors that might be included in the experiment
  - and their levels
- The likely importance of each factor
- The direction of each main effect
- Any insight/experiences on interactions or nonlinear trends

Local brainstorming – but experts often at different sites

Web-based System (GiSEL)

- Gathers opinions/suggestions on factors and their levels
  - via a dynamic questionnaire with free form comments
- Keeps a record of opinions, experiments and results
- Guides factor groupings via software
  - explores the resources needed for various strategies and factor groupings
  - estimates via simulation the probability of failing to detect
    – active main effects
    – active control$\times$control
    – active control$\times$noise

Factors Under Consideration

Factor List

40 factors

- [x] Afr
- [x] Air Assisted Injection
- [x] Altitude
- [x] Ambient Temperature
- [x] Back Pressure
- [x] Calibration
- [x] Carbon Deposits
- [x] Combustion Chamber Contamination
- [x] Combustion/start History
- [x] Cranking Speed
- [x] Early Entry Into Fuel Cutoff
- [x] Effective Compression Ratio
- [x] Engine Age

Factor Information

Afr

Preset Attributes
Description: No description
Type: Design
Submitted by: David Dupplaw
Submitted on: 11:04 Fri 11th October 2002

You previously submitted an opinion at 14:04 Wed 09th January 2002.

How important do you believe Afr is in influencing spark plug foul resistance?

( ) Not Important ( ) Slightly Important ( ) Quite Important (•) Very Important

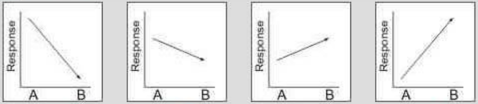
How confident are you of your above view?

( ) Not Confident ( ) Slightly Confident ( ) Quite Confident (•) Very Confident

Do you have an opinion on the nature of the effect? (see following question)

( ) Yes ( ) No

What effect do you think Afr will have on spark plug foul resistance?
A is Lean, B is Rich.



(•) Greatly decreases spark plug foul resistance	( ) Slightly decreases spark plug foul resistance	( ) Slightly increases spark plug foul resistance	( ) Greatly increases spark plug foul resistance

How confident are you of your above view?

( ) Not Confident ( ) Slightly Confident (•) Quite Confident ( ) Very Confident

Ten Factors

Control – very likely

- Plug type*
- Plug gap*
- Air fuel ratio
- Injection timing

Control – less likely

- Spark during crank
- Spark time during run-up
- Higher idle speed
- Cranking fuel

*hard-to-change; group together

Noise

- Temperature
- Injector tip-leakage

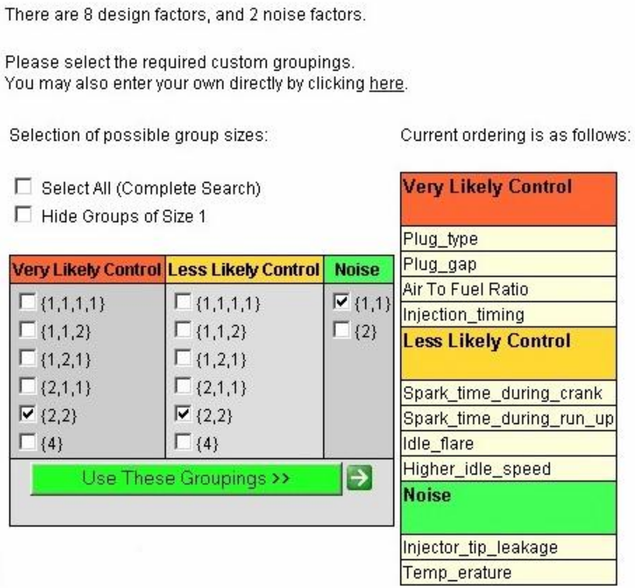
Criteria for Choice of Group Sizes and Design

Criteria include:

1. Minimise the expected total number of effects to be estimated
    - across both stages
2. Minimise the probability of exceeding a target number of effects
3. Minimise the probability of failing to detect
    - active main effects
    - active control x noise interactions
    - active control x control interactions

Criteria 1 & 2 act in opposite direction from criterion 3

Investigation of Different Groupings



Selected Groups and Design

Control:

Group 1: Plug type* & Plug gap*

Group 2: Air to fuel ratio & Injection timing

Group 3: Spark time during crank & During run-up

Group 4: Higher idle speed & Cranking fuel

Noise:

Group 5: Injector tip leakage; Group 6 Temperature

Design:

- Half-replicate ($I = G_1 G_2 G_3 G_4 G_5 G_6$)
- Arranged in 4 sessions of 8 runs
- Split-plot design: G_1, $G_5 G_6$, $G_1 G_5 G_6$ confounded with sessions (eg Bingham et al., 2004)

Design & Observations for First Stage Experiment

Session	Run	Response	Session	Run	Response
1	1 0 1 0 0 0	6.57	3	0 0 1 1 1 1	28.09
	1 0 1 0 1 1	2.74		0 1 0 1 0 0	-3.86
	1 0 0 1 0 0	5.15		0 0 0 0 1 1	-5.12
	1 0 0 1 1 1	10.19		0 0 0 0 0 0	-12.36
	1 1 1 1 0 0	8.17		0 0 1 1 0 0	-11.36
	1 1 0 0 0 0	41.98		0 1 0 1 1 1	-2.61
	1 1 0 0 1 1	10.27		0 1 1 0 0 0	-6.52
	1 1 1 1 1 1	9.50		0 1 1 0 1 1	-5.69
2	1 0 0 0 1 0	-11.59	4	0 1 0 0 0 1	-7.43
	1 0 0 0 0 1	-6.04		0 0 1 0 1 0	-11.23
	1 1 1 0 0 1	-5.57		0 1 1 1 0 1	-7.25
	1 0 1 1 1 0	-1.75		0 1 1 1 1 0	-2.73
	1 0 1 1 0 1	-3.67		0 1 0 0 1 0	-0.82
	1 1 0 1 0 1	-2.48		0 0 1 0 0 1	-6.16
	1 1 0 1 1 0	3.63		0 0 0 1 1 0	-1.50
	1 1 1 0 1 0	-7.71		0 0 0 1 0 1	-2.82

Results of First Stage Experiment

The two largest effects were

- {Higher idle speed & Cranking fuel} $\times$ Injector tip leakage
- {AFR & Injection timing} $\times$ Temperature

– both grouped control $\times$ noise interactions

6 factors to investigate at the second stage
– no need to investigate all interactions

Second Stage Experiment

Six individual factors

- A AFR
- B Injection timing
- C Higher idle speed
- D Cranking fuel
- E Injector tip leakage
- F Temperature

Other factors held at nominal values
Decided to estimate 6 main effects and 14 interactions

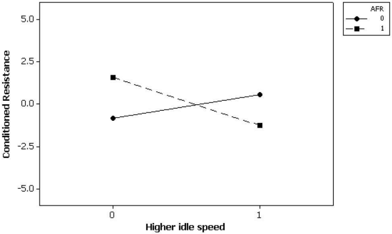
Design

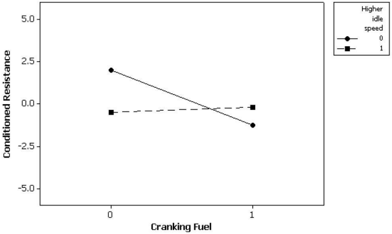
- Half-replicate in 32 runs (I = ABCDEF)

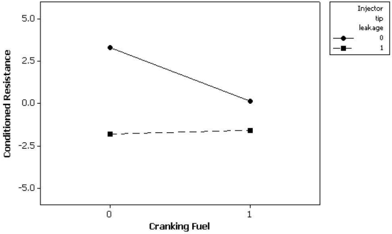
  - could have been smaller

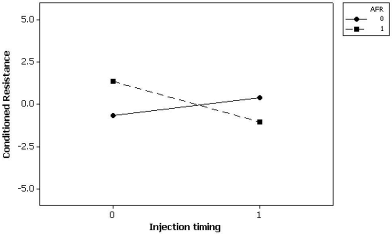
Temperature main effect and the four largest interactions were the most important

Most Important Interactions









Interaction Screening – Summing Up

- Ignoring interactions in an initial screening experiment may lead to important factors being overlooked
- Group screening can be effective in practice
  - worked well in the case study
  - no evidence of cancellation of effects
  - results used to plan experiments to fit a complex response surface
- Tools developed for elicitation, choosing group sizes and experiment simulation are freely available www.doe.soton.ac.uk/screening.php

Conclusions

- Screening large numbers of factors is important due to increasing complexity of products and processes
- All methods require effect sparsity and can benefit from elicited information, where available
- Follow-up experiments are needed
- Both supersaturated designs and group screening can work in practice
  – need more applications, especially published

Estimates of Grouped and Individual Interactions

Grouped effect	1st Stage Estimate	2nd Stage Individual effect estimates			
$G_2 G_6$	-6.17	AF: -0.27	BF: -1.19		
$G_4 G_5$	5.91	CE: 1.46	DE: 1.70		
$G_2 G_4$	-5.10	AC: -2.13	AD: -0.18	BC: -1.39	BD: 0.99
	—	AB: -1.72	CD: 1.77		

- Signs of all large individual interactions agree with signs of the corresponding grouped interactions

Estimates of Grouped and Individual Interactions

Grouped effect	1st Stage Estimate	2nd Stage Individual effect estimates			
G_2G_6	-6.17	AF: -0.27	BF: -1.19		
G_4G_5	5.91	CE: 1.46	DE: 1.70		
G_2G_4	-5.10	AC: -2.13	AD: -0.18	BC: -1.39	BD: 0.99
	—	AB: -1.72	CD: 1.77		
G_4G_6	3.41	CF: -1.65	DF: -0.97		
G_2G_5	-3.36	AE: -1.22	BE: 0.47		

- Unlikely that smaller size of grouped effects was caused by cancellation
- Inconsistencies in sign – from larger random error at Stage 2?

Frequentist Data Analysis Methods

Fitting submodels of the full main effects model via least squares

- Simple linear regression on each factor
  Satterthwaite (1959); Holcomb et al. (2003)

- All subsets regression
  Wu (1993); Abraham et al. (1999)

- Stepwise and forward selection
  Lin (1993); Wu (1993); Kelly and Voelkel (2000)

Frequentist Data Analysis Methods

Fitting submodels of the full main effects model via least squares

- Simple linear regression on each factor
  Satterthwaite (1959); Holcomb et al. (2003)
- All subsets regression
  Wu (1993); Abraham et al. (1999)
- Stepwise and forward selection
  Lin (1993); Wu (1993); Kelly and Voelkel (2000)

General findings

- Different methods may identify different numbers of effects
- All subsets regression compares well but is usually impractical
- Rough guide – designs with $r_{\text{max}} < 0.33$

Frequentist Data Analysis Methods

Fitting a single model via penalised least squares

Choose β to

$$\text{minimise } \{(\boldsymbol{Y} - X\boldsymbol{\beta})'(\boldsymbol{Y} - X\boldsymbol{\beta}) + \text{penalty function}\}$$

- Ridge regression – penalty $\propto \sum \beta_i^2$
  Lin (1995)
- SCAD (Smoothly Clipped Absolute Deviation)
  R. Li and Lin (2002)

Frequentist Data Analysis Methods

Fitting a single model via penalised least squares

Choose $\boldsymbol{\beta}$ to

$$\text{minimise } \{(\boldsymbol{Y} - X\boldsymbol{\beta})'(\boldsymbol{Y} - X\boldsymbol{\beta}) + \text{penalty function}\}$$

- Ridge regression – penalty $\propto \sum \beta_i^2$
  Lin (1995)
- SCAD (Smoothly Clipped Absolute Deviation)
  R. Li and Lin (2002)

General findings

- Ridge regression is poor if f is much bigger than n
- SCAD is better than stepwise regression

Bayesian Data Analysis Methods

Compare models based on their posterior model probabilities

$$P(m_i|Y) = \frac{P(Y|m_i)P(m_i)}{\sum_{\text{all models}} P(Y|m_j)P(m_j)} \qquad \text{(Bayes Theorem)}$$

- Prior model probability $P(m_i)$ is
  - chosen to reflect effect sparsity
    i.e. a model composed of only a few effects has higher prior probability
  - calculated from the probabilities of each individual effect being active
    – ideally elicited from experts
- Problem: too many models to evaluate

Bayesian Data Analysis Methods

Solution: search the model space using Markov Chain Monte Carlo techniques to find a set of models that have high posterior probability

- MCMC generates a sample from $P(m|Y)$ using only $P(Y|m)P(m)$
- From this sample, approximate the posterior model probabilities by counting the number of times each model occurs in the sample
- Chipman et al. (1997), Meyer & Wilkinson (1998), Beattie et al. (2002)

Bayesian Data Analysis Methods

Extract practical information from the set of models

- Factors in the top few models selected as active
- Use formal Bayes model comparison methods to compare top models eg intrinsic Bayes factors
  Beatie et al. (2002)
- Assess importance of each factor by summing the posterior probabilities for each model in which it features
  Box & Meyer (1986); Meyer & Wilkinson (1998)