How Bayesian Thinking Can Help in Designing Experiments

Bradley Jones
Senior Manager: Statistical R&D
SAS Institute Inc.
June 7, 2006

Outline

- Why Go Bayesian?
- Bayesian Screening
- Bayesian Design Augmentation
- Bayesian Design for Nonlinear Models
- Summary

Why Go Bayesian?

Screening

Answer: Model Uncertainty

Augmentation

Answer: Intelligent Use of Previously Acquired Data

Nonlinear

Answer: Avoid Local Designs

Outline

- Why Go Bayesian?
- → Bayesian Screening
- Bayesian Design Augmentation
- Bayesian Design for Nonlinear Models
- Summary

D-optimal Design

Given the usual linear regression model

$$y = X\beta + \varepsilon$$

find a design matrix, X, to maximize

$$\left| X^T X \right|$$

Problem

D-optimal designs depend on the choice of the *a priori* model.

Solution: Bayesian D-optimality

Consider two kinds of effects:

Primary effects are ones you are sure you want to estimate. There are p_1 of these.

Potential effects are ones you are afraid to ignore. There are p_2 of these.

For sample size, n

$$p_1 < n < p_1 + p_2$$

Defining the K matrix

$$K = \begin{bmatrix} 0_{p_1 x\, p_1} & 0_{p_1 x\, p_2} \\ 0_{p_2 x\, p_1} & I_{p_2 x\, p_2} \end{bmatrix}$$

Bayesian D-optimal designs

find a design matrix, X, to maximize

$$D_{Bayes} = \left| X^T X + K / \tau^2 \right|$$

where τ is a tuning constant.

Example – Bayesian D-optimal = Res IV FF

2^{6-2} Fractional Factorial Resolution IV design

intercept and main effects are primary

2-factor interactions are potential

$p_1 < n < p_1 + p_2$

$p_1 = 7$ $p_2 = 15$ $n = 16$ $(7 < 16 < 22)$

Reference DuMouchel W and Jones, B. (1994) "A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model," *Technometrics 36*, 37-47.

Outline

- Why Go Bayesian?
- Bayesian Screening
- → Bayesian Design Augmentation
- Bayesian Design for Nonlinear Models
- Summary

How do you use the information you already have?

Foldover design does not use the data.

Reference:

Meyers, R. D., Steinberg, D. M and Box, G. “Follow-up designs to resolve confounding in multifactor experiments,” (1996) *Technometrics 38, No 4* , 303-332

Technical Details…

Model

$$y \sim N(X_i\beta_i, \sigma^2 I)$$

Posterior Mean

$$\hat{\beta}_i = (\Gamma_i + X_i' X_i)^{-1} X_i' Y$$

Posterior Variance

$$V_i = (\Gamma_i + X_i' X_i)^{-1}$$

More Technical Details…

Posterior Probability of the ith Model

$$p_i = \pi^{f_i}(1-\pi)^{k-f_i}\gamma^{-t_i}|\Gamma_i + X_i'X_i|^{-1/2}S_i^{-(n-1)/2}$$

Prior Parameters and Variance Matrix for the Augmented Design

$$\beta_0 = \sum_i p_i \beta_i^*$$

$$V_0 = \sum_i p_i V_i^* + \sum_i p_i(\beta_i^* - \beta_0)(\beta_i^* - \beta_0)'$$

Whew! Finally something to optimize…

$$F = \left|V_0^{-1} + Z'Z\right|$$

Z is the design matrix for augmentation

First term is the information contained in the original design.

Outline

- Why Go Bayesian?
- Bayesian Screening
- Bayesian Design Augmentation
- → Bayesian Design for Nonlinear Models
- Summary

Example

$y = \exp(-\boldsymbol{a}x) + \varepsilon$

Procedure:

Make a guess for ***a***.

Find the information matrix given this guess.

Find the design that maximizes the determinant of this matrix.

Problem:

This locally “optimal” design puts all the points at 1/***a***

Bayesian Approach

Characterize your uncertainty about the parameter, ***a***, using a prior distribution.

Summary

- Bayesian Paradigm deals with model uncertainty.
- It uses previous information to improve the next step.

  (both for analysis and sequential design)
- It avoids the silliness of locally optimal designs.

Contact info

Bradley.Jones@jmp.com

sas

The Power to Know.