Optimal Design of Experiments

presentation for Stanford University Statistics 252 class

Data Mining and Electronic Business

July 7, 2004

Selden B. Crary, Ph.D.
Crary Group, Palo Alto
selden@crary.cc
Design of Experiments Makes a Difference

Weighing two apples

Simple method:

\[ W_1 = 0.63 \text{ kg} \pm 2\sigma \]
\[ W_2 = 0.57 \text{ kg} \pm 2\sigma \]

\( \sigma \) = std. dev. of measurement

Is there a better method?
\[ R_1 = W_1 + W_2 \]
\[ R_2 = W_1 - W_2 \]

Solve for \( W_1 \) and \( W_2 \):
\[ W_1 = \frac{1}{2} (R_1 + R_2) \]
\[ W_2 = \frac{1}{2} (R_1 - R_2) \]

\[
Var(W_1) = Var(W_2) = \frac{1}{2^2} [Var(R_1) + Var(R_2)] = \frac{\sigma^2}{2}
\]

\[
Std. dev. = \frac{\sigma}{\sqrt{2}} \quad \text{Errors in weights reduced by 29%}
\]
Many industrial products and processes benefit from designed experiments

Spaghetti Sauce

Semiconductors
Statistical Design of Experiments:
Response-Surface Methodology (RSM)

Application Areas:
Improvement of products and processes

Three Examples:
Quadratic response: one factor
Sensor calibration: two factors
Direct mail marketing campaign: multiple factors
Classical Design of Experiments

Examples:
- Agriculture (static designs)
- Pharma (sequential designs)
- Quality control in manufacturing

Features:
- Easy to find
- Sample space ~ uniformly
- Not optimal

Full-factorial design

Fractional-factorial design
Problem: The size of factorial and fractional-factorial designs grows exponentially with the number of variables.
Methods used to design experiments depended upon the level of computation available:

- Classical designs (Fisher)
- Orthogonal arrays (Taguchi)
- D-optimal designs
- I-optimal, G-optimal, and application-specific optimal designs
**Optimal Statistical Design of Experiments**

**Application Areas**

- Further improvement of products and processes
  
  (50% smaller confidence bounds on prediction are typical)

- Increased accuracy of standard-cell libraries

- Optimization of hospital staffing

- Optimization of marketing campaigns
Theory
Model

Polynomial response model:

\[ e.g., \ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 \]

or \[ Y = f^T \beta \] (linear in the \( \beta_i \)'s)

Measurements:

\[ e.g., \ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon \]

\[ \{Y_i\} \ i=1, \ldots, N \]

or \[ Y = X \beta + \epsilon, \]

\[ \epsilon \sim \text{i.i.d. with mean zero and variance } \sigma^2 \]
Example of $X$, the design matrix

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon$$

or, $Y = X\beta + \epsilon$

$$X = \begin{bmatrix}
1 & (1) x_1 & (1) x_2 & (1) x_1^2 & (1) x_2^2 & (1) x_1 & (1) x_2 \\
1 & (2) x_1 & (2) x_2 & (2) x_1^2 & (2) x_2^2 & (2) x_1 & (2) x_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & (N) x_1 & (N) x_2 & (N) x_1^2 & (N) x_2^2 & (N) x_1 & (N) x_2
\end{bmatrix}$$
Optimal Designs for Parameter Estimation

Parameters $\beta$

\[
\hat{\beta} = \left( X^T X \right)^{-1} X^T Y
\]

\[
Var(\hat{\beta}) = \sigma^2 \left( X^T X \right)^{-1}
\]

$D$-optimality:

\[
\min_{\omega_N} \| \left( X^T X \right)^{-1} \|,
\]

where $\omega_N$ is the set of $N$-point designs
Example 1:
D-optimal design for fitting a parabola with N=5 points

\[ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \]

Var of prediction =

\[ \text{Var}(\hat{Y}) = E\left[ (\hat{Y} - Y)^2 \right] \]

*IV = Integrated Variance: defined on next slide
Optimal Designs for Prediction

$G$ – optimality: minimize expected worst-case prediction error over a region $\chi$

$$\min_{\omega_N} \max_{\chi} \mathbb{E}\left[ (\hat{Y}(x) - Y(x))^2 \right] = \min_{\omega_N} \max_{\chi} f^T \left( X^T X \right)^{-1} f$$

$I$ – optimality: minimize average prediction error over $\chi$

$$\min_{\omega_N} \int_{\chi} \mathbb{E}\left[ (\hat{Y}(x) - Y(x))^2 \right] \ dx = \min_{\omega_N} \int_{\chi} f^T \left( X^T X \right)^{-1} f \ dx$$

$$IV = \frac{\int_{\chi} \mathbb{E}\left[ (\hat{Y}(x) - Y(x))^2 \right] \ dx}{\int_{\chi} dx} = \text{Integrated Variance}$$
D-, G-, and I-optimal designs for fitting a parabola with N=5 points

\[ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \]
Interesting Observations Re I-optimal Designs

- **Asymmetric** optimal designs can result from **symmetric** problem statements.

- Phase transitions exist between designs of various symmetries, as weights assigned to different regions are changed continuously.
Extensions of I-optimal Designs (1)

Region need not be square, cube, or hyper-cube, i.e., inequality constraints can be handled.

Penalty for variance can be weighted.

Equality constraints can be handled, e.g., mixtures.

Variance of responses can be taken into account, i.e., heteroscedasticity can be treated.
Extensions of I-optimal Designs (2)

Points can be forced to be in the design.

Optimal augmented designs are possible.

Prior information on the distribution of $\beta$’s, e.g., from previous measurements, can be taken into account (Bayesian approach).

I-optimality can be extended to deterministic error models (computer experiments).
Example 2: Optimal Sensor Calibration

MEMS pressure sensor in an automotive setting

Sensor capacitance $C=C(P,T)$

Find an efficient design for determining a compensation formula:

$$C = C(P,T) = \beta_0 + \beta_1 T + \beta_2 P + \beta_3 P^2 + \beta_4 P^3$$

Design must have at least $N_{\text{min}} = 5$ points

Changing $P$ is less expensive than changing $T$. 
I-optimal Designs for

\[ C(P,T) = \beta_0 + \beta_1 T + \beta_2 P + \beta_3 P^2 + \beta_4 P^3 \]
Variance Contours of I-optimal Designs for
\[ C(P,T) = \beta_0 + \beta_1 T + \beta_2 P + \beta_3 P^2 + \beta_4 P^3 \]

N=5  
N=40
Integrated Variance vs. Number of Points

\[ C(P,T) = \beta_0 + \beta_1 T + \beta_2 P + \beta_3 P^2 + \beta_4 P^3 \]

Compensated Pressure = \( P(C,T) + \Delta \)

- Slope \( \sim N^{-1} \)
If some regions are more important for prediction than others, use weighted I-optimal designs.

\[
\min_{\omega_N} \int_0^1 \int_0^1 f^T \left( X^T X \right)^{-1} f e^{-\left( C^2 + T^2 \right)/2\sigma^2} dCdT 
\]

Example: \( C(P,T) = \beta_0 + \beta_1 T + \beta_2 P + \beta_3 P^2 + \beta_4 P^3 \), N=7

\( N=7, \sigma = \infty \)

\( N=7, \sigma = 0.1 \)

© Crary Group, 2004
Example 3: Similar to On-Line Marketing:
Direct Mail Marketing of Credit-Cards* (1)

Challenges:
- Response rates down: 2% → 0.6%
- Balance surfers: People move accounts to 0% APR
- Adverse selection: bankruptcy
- Solicitation volumes up: 6,000,000,000/yr
- Costs per account activation up

*Source: Gerald Fahner, Fair Isaac seminar at Stanford, Jan. 31, 2003
Direct Mail Marketing of Credit-Cards (2)

Segmentation factors: Risk score, income, credit usage

Treatment factors: APR’s, fees, promotional incentives, credit line, communications (words, fonts, colors, layout, envelopes, etc.)

Strategy: Should an offer be made, and, if so, what treatment should be applied?:

\{\text{segmentation factors } X}\rightarrow \{\text{treatment factors } T\}

Statistical question: How to make splits on decision tree?
Direct Mail Marketing of Credit-Cards (3)

Segmentation factors: \{X\}

Treatment factors: \{T\}

\[ \text{Profit} = f(T,X) \]

Goal: Find: \( T^* = \arg\max_{\{T\}} f(T, \text{given } X) \)

A possible optimal-design strategy: Find the design that minimizes the average expected squared error of prediction (i.e., IV) of the profit function \( f(T,X) \).

N.B.: The form of the profit function would need to be determined in order to use the approach given here.
Aside:

- Assume cost of $0.50 per piece mailed
- Fact: 10% of mailings are designed experiments
- $300,000,000/yr is expended on designed experiments?
- If 50% is saved using optimal design of experiments, then savings of $150,000,000 are possible.
Summary

Optimal designs use 50%* fewer points and/or are 50%* more accurate for prediction

Desirable features:
- I- and G-optimal designs focus on accurate prediction
- Region of greatest interest can be weighted
- Complex cost functionals can be envisioned
- Adaptive, sequential approach possible
  (Computer experiments are accommodated)

* Typical
Sources

Books: www.WebDOE.com/Resources

Classical designs: WebDOE (registration req’d.)

Optimal designs: ditto

Training: WebDOE home page

Questions: selden@crary.cc