Designing Computer Experiments to Determine Robust Control Variables

Thomas Santner
Department of Statistics
Ohio State University

Los Alamos National Laboratory
December 2003

(Joint with Jeff Lehman and William Notz)
OUTLINE

I. Introduction

II. Problem formulation (what is robustness?)

III. Modeling output from computer experiments

IV. Sequential computer designs to find a robust engineering design

V. An example

VI. Take home points
I. Introduction

- Three Types of Experiments
  
  1. **Physical Experiments:**
     - Gold standard for establishing cause and effect relationships
     - Mainstay of Agricultural, Industry, Medicine
     - Principles of randomization, blocking, choice of sample size, and stochastic modeling of response variables all developed in response to needs of physical experiments
  
  2. **Simulation Experiments:** Complex physical system each of whose parts interact in a known stochastic manner but whose ensemble is not understood analytically
     - Heavily used in Industrial Engineering–compare job shop set ups
  
  3. **Computer Experiments:** relatively new
  
  4. **Combinations** of the above
I. Introduction (cont)

- In some situations performing a physical experiment is not feasible
  1. Physical process is technically too difficult to study
  2. Number of variables is too large
  3. Too expensive to study directly

- When physical experiments are not possible, it may still be feasible to conduct a computer experiment, i.e., if the physical process relating the inputs to the response(s)
  1. Can be described by a mathematical model relating the response to inputs $\mathbf{x}$
  2. Numerical methods exist for solving the mathematical model
  3. The numerical methods can be implemented with computer code (in finite time!)

then one can run the computer code to produce one or more “responses” $y(\mathbf{x})$ at any input $\mathbf{x}$, i.e., one can conduct a computer experiment
I. Introduction (cont)

- **Setup:**
  1. **Inputs:** \( \mathbf{x} = (\mathbf{x}_c, \mathbf{x}_e) \) where
     - \( \mathbf{x}_c = \text{control} \) (manufacturing) variables
     - \( \mathbf{x}_e = \text{noise} \) (field, environmental) variables
       (model variables)
  2. \( \mathbf{X}_e \sim F(\cdot) \) ("target field conditions")
  3. \( F(\cdot) \) not completely known
  4. \( y(\cdot) \) real-valued
  5. Summary of \( y(\mathbf{x}_c, \mathbf{X}_e) \) distribution
     \( \mu_F(\mathbf{x}_c) = E_F\{y(\mathbf{x}_c, \mathbf{X}_e)\} \)
     (but not only one)

- **Engineering design** \( \equiv \) choice of \( \mathbf{x}_c \)

- **Design of computer experiment** \( \equiv \) choice of \( \mathbf{x}_1, \ldots, \mathbf{x}_n \) at which to evaluate computer code
II. Problem formulation

What is a "robust" choice of $x_c$?

Example: Suppose $x_c, x_e$ are real-valued and $y(\cdot)$ is

Consider four $F(\cdot)$

\begin{itemize}
  \item Dist. 1
  \item Dist. 2
  \item Dist. 3
  \item Dist. 4
\end{itemize}
Problem formulation (cont)
Corresponding $\mu_F(x_c)$ are:

$$x_c = \arg \min \mu_F(x_c)$$

- **Philosophy**: robust $x_c^*$ should make $y(x_c^*, x_e)$ as “flat” as possible
  $$(\Rightarrow \mu_F(x_c^*) \text{ insensitive to choice of } F(\cdot))$$

- **Taguchi-like**: robust product design not sensitive to choice of noise variables

- Does not require specification of a set of alternative models for $F(\cdot)$
Problem formulation (cont)

- **Alternative 1**: Assume $F \in \mathcal{G}$ (known). Call $x^G_c$ $\mathcal{G}$-robust if

\[
\max_{G \in \mathcal{G}} \mu_G(x^G_c) = \min_{x_c \in x_c} \max_{G \in \mathcal{G}} \mu_G(x_c)
\]

**Examples of $\mathcal{G}$**

1. $\{\Phi(\frac{x-\mu}{\sigma}) : \mu_L \leq \mu \leq \mu_U, \sigma^2_L \leq \sigma^2 \leq \sigma^2_U\}$
   where $\mu_L$, $\mu_U$, $\sigma^2_L$ and $\sigma^2_U$ are known.

2. $\{(1 - \varepsilon)\Phi(\frac{x-\mu}{\sigma}) + \varepsilon \text{H}(x - \mu) :$ \begin{align*}
\mu, \sigma^2 \text{ above, H(\cdot) symmetric}
\end{align*}

- **Alternative 2**: Assume $F \in \mathcal{G}$ (known) and $\pi(\cdot)$ is a distribution on $\mathcal{G}$. $x^{\pi}_c$ is $\pi$-robust if

\[
\int \mu_G(x^{\pi}_c) d\Pi(G) = \min_{x_c} \int \mu_G(x_c) d\Pi(G)
\]

(parametric)
Problem formulation (cont)

- **This talk:** Given $G(\cdot)$ a $X_c$ distribution let

$$\sigma^2_G(x_c) = \text{Var}_G[y(x_c, X_e)]$$

($G(\cdot)$ uniform?)

- $x_c^V$ is $V$-robust if

$$\sigma^2_G(x_c^V) = \min \sigma^2_G(x_c)$$

s.t.

$$\mu_F(x_c) \leq a \times \min x_c^* \mu_F(x_c^*) + c$$

**Special Cases**

1. $(a, c) = (1, 0)$ : $\mu(x_c) \leq a \times \min x_c^* \mu_F(x_c^*)$
2. $(a, c) = (0, c)$ : $\mu(x_c) \leq c$

- $x_c^M$ is $M$-robust if

$$\mu_F(x_c^M) = \min \mu(x_c)$$

s.t.

$$\sigma^2_G(x_c) \leq a \times \min x_c^* \sigma^2_G(x_c^*)$$
III. Modeling output from computer experiments

- **Assume**: distribution of the environmental variables, \( F(\cdot) \) of \( X_e \) is discrete on \( \{x_{e,j}\}_{j=1}^{n_e} \) with \( w_j \equiv P_F[X_e = x_{e,j}] \).

- **Bayesian Formulation**: prior on \( y(\cdot) \)

\[
Y(x) = \sum_{i=1}^{k} \beta_i f_i(x) + Z(x) = \beta^T f(x) + Z(x)
\]

where \( f_1(\cdot), \ldots, f_k(\cdot) \) are known, \( \beta \) is unknown, and \( Z(x) \) is a zero mean stationary GSP with unknown variance \( \sigma^2_Z \) and (known) positive definite correlation function \( R(\cdot) \).

- \([\beta, \sigma^2_Z] \propto 1/\sigma^2_Z \)

- \( Y(\cdot) \) prior induces a prior distribution on

\[
\mu_F(x_c) = E_F[y(x_c, X_e)] = \sum_{j=1}^{n_e} w_j y(x_c, x_{e,j})
\]

\[
M_F(x_c) = \sum_{j=1}^{n_e} w_j Y(x_c, x_{e,j})
\]
**Modeling output (cont)**

and on

\[ \sigma_F(x_c) = \text{Var}(y(x_c, X_e)) = y^{n_e}(x_c)^\top A y^{n_e}(x_c) \]

where \( y^{n_e}(x_c) = (y(x_c, x_e, 1, \ldots, y(x_c, x_e, n_e)) \)

\[ (V_F(x_c) = Y^{n_e}(x_c)^\top A Y^{n_e}(x_c)) \]

- In practice \( R(\cdot) \equiv R(\cdot|\theta) \) with \( \theta \) unknown: \( [\beta, \sigma^2, \theta] \)

- **Idea**: prior should be capable of producing a wide range of possible \( y(\cdot) \) (with different # of max/min, path smoothness)

- Sample path properties depend on smoothness of \( R(\cdot) \) at the origin

- **Example**:

\[ R(h) = \prod_{i=1}^{k} \exp \{-\theta_i |h_i|^{p_i}\} \]
Modeling output (cont)

Example: Four draws from a zero mean, unit variance GSP with the exponential correlation function

$$R(h) = \exp \left\{ -\theta h^2 \right\}$$

for $\theta = 0.5$ (solid lines),
$\theta = 1.0$ (dotted lines), and
$\theta = 10.0$ (dashed lines)
IV. Sequential computer designs to find a robust engineering design

\[
\chi^M_c = \arg\min \mu_F(\chi_c) \quad \text{s.t.} \quad \sigma^2_G(\chi_c) \ll "small"
\]

- **Idea**: Initial sample to gives rough idea of 
  \(y(\chi_c, x_e)\) (hence \(\mu_F(\chi_c)\) and \(\sigma^2_G(\chi_c)\)); additional observations improve estimation of \(\chi^M_c (\chi^V_c)\) and identification of the feasible region

- **Notation**
  - \(S_n = \{x^{tr}_1, \ldots, x^{tr}_n\} \quad n \text{ sites at which to evaluate } y(\cdot, \cdot)\)
  - \(x^{tr}_i = (\chi^{tr}_{c,i}, x^{tr}_{e,i})\)
  - \(Y^n = (Y(x^{tr}_1), \ldots, Y(x^{tr}_n))^\top, \quad y^n\)

- **Update Algorithm**
  - Estimate \(\theta\) in \(R(\cdot|\theta)\) [REML/MLE]
  - Choose \(x^{tr}_{c,n+1}\) to maximize “improvement” criterion (\(\chi^M_c, \chi^V_c\) specific)
  - Choose \(x^{tr}_{e,n+1} = \arg\max x_e \rho((x^{tr}_{c,n+1}, x_e), S_n)\) \((\rho(\cdot, \cdot) = \text{point-by-point minimum})\)
  - Compute \(y(x^{tr}_{c,n+1}, x^{tr}_{e,n+1})\)
  - Check if algorithm should be stopped
Sequential computer design (cont)

- $S_n \equiv$ Maximin LHD (product of design in $x_c$ and Niederreiter sequence in $x_e$?)
- **Improvement ?? Idea**

\[
\left\{ \begin{array}{ll}
\mu_n^{\min} - \mu(x_c), & \mu(x_c) < \mu_n^{\min} \\
0, & \mu(x_c) \geq \mu_n^{\min}
\end{array} \right.
\]

where $\mu_n^{\min} = \min\{\mu(x_{c,i}^{tr}) : 1 \leq i \leq n\}$
- **In practice** choose $x_{c,n+1}$ to maximize

\[
I(x_c) = \mathbb{E}\{\max\{0, M_n^{\min} - M(x_c)\} \mid \mathbf{Y}_n, \gamma}\times P\{\text{constraint} \mid \mathbf{Y}_n, \gamma\},
\]

where

- **constraint** is $V(x_c) \leq a \times \nu_n^{\min} + c. =$
  posterior variance estimate $\mid \mathbf{Y}_n^n$
- $\nu_n^{\min} = \min\{\mathbb{E}[V(x_{c,i}^{tr})] \mid \mathbf{Y}_n^n, \gamma : 1 \leq i \leq n\}$
- $M_n^{\min} = \min\{M(x_{c,i}^{tr}) : 1 \leq i \leq n\}$
  = posterior estimate of $\mu_n^{\min}$

- $x_{e,n+1}^{tr}$ attempts to “spread” points to better estimate $\mu_F(x_c)$
**2-d Example:** Find $x_c^V$ s.t.

$$
\sigma^2_F(x_c^V) = \min_{x_c} \sigma^2_F(x_c) \\
\text{subject to} \\
\mu_F(x_c^V) \leq -0.08
$$

where $y(\cdot)$ is

and $F(\cdot)$ is uniform on $\{0.025, 0.075, \ldots, 0.975, 1\}$
\[ \sigma^2_F(x_c) \]

\[ \mu_F(x_c) \]
• **Algorithm**
  
  – **Initial Design**: 20 pt MmLHD  
  – \( R(\cdot) \equiv \text{power exponential} \)  
  – \( \rho(\cdot, \cdot) \equiv \text{Euclidean distance} \)  
  – **Stop** if five step MA of Exp Impr. < \( 10^{-5} \)

• **Results**
  
  – Initial estimated \( x^V_c = 0.811 \) (feasible region not correct)  
  – Final estimated \( x^V_c = 0.176 \) (3 place accuracy; 19 additional inputs)
4-d Example: Find $x_c^M = \arg\min_{x_c} \mu_F(x_c)$ when

$$y(x_c^1, x_c^2, x_e^1, x_e^2) = \text{“Branin function”}$$

i.e., minimize $\mu_F(x_c)$ where

subject to
• **True minimizer**: $(\pi, 2, 275)$

• $\sigma_F^2(\chi_c^M) \leq 10,000$

• **Algorithm**
  
  – **Initial Design**: 40 pt MmLHD
  – $R(\cdot) \equiv$ power exponential
  – $\rho(\cdot, \cdot) \equiv$ Euclidean distance
  – **Stop** if five step MA of Exp Impr. $< 10^{-5}$

• **Results**

<table>
<thead>
<tr>
<th># Points +ed</th>
<th>Improvement</th>
<th>Pred Minimizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.904673</td>
<td>$(2.65, 0.00)$</td>
</tr>
<tr>
<td>40</td>
<td>0.236627</td>
<td>$(3.31, 2.02)$</td>
</tr>
<tr>
<td>60</td>
<td>0.020322</td>
<td>$(2.92, 1.98)$</td>
</tr>
<tr>
<td>75</td>
<td>0.024939</td>
<td>$(3.15, 1.97)$</td>
</tr>
<tr>
<td><strong>80</strong></td>
<td><strong>0.019231</strong></td>
<td><strong>(3.15, 2.25)</strong></td>
</tr>
</tbody>
</table>

• % Rel Error of $\chi_{c,1}^M$ is 0.32% and of $\chi_{c,2}^M$ is 1.1%
VI. Take home points

- Initial allocation? 10/variable too many...
- ρ??
- Number of variables for which this works well?
- Many alternatives to $\mu_F(x_c): \xi_F(x_c, \alpha)$—given $\alpha$
  \[ P_F\{y(x_c, X_e) \leq \xi_F(x_c, \alpha)\} \equiv \alpha \]
  (minimize $\xi_F(x_c, 0.95)$)
- Robustness for multiple outputs? Suppose $y^1(\cdot)$, $y^2(\cdot)$ and $\mu^i_F(x_c) \equiv E_F\{y^i(x_c, X_e)\}$ ($i = 1, 2$)
  \[ \min \mu^1_F(x_c) \quad \text{subject to} \quad \mu^2_F(x_c) \geq B \]
  What $F(\cdot)$ not completely known?
- Alternative stochastic process models for $Y(\cdot)$?
- Contrast with setting in which there are multiple outputs. Suppose $y^1(\cdot)$, $y^2(\cdot)$ and $\mu^i_F(x_c) \equiv E_F\{y^i(x_c, X_e)\}$ ($i = 1, 2$)
  \[ \min \mu^1_F(x_c) \\
  \text{subject to} \\
  \mu^2_F(x_c) \geq B \]
- Validation experiments to confirm determination of robust $x_c$ ?