Experimental Design for Simulation

[Law, Ch. 12][Sanchez et al.[1]]

Peter J. Haas

CS 590M: Simulation
Spring Semester 2020

[1] S. M. Sanchez, P. J. Sanchez, and H. Wan. "Work smarter, not harder: a tutorial on designing and conducting simulation experiments". *Proc. Winter Simulation Conf.*, 2018, p. 237–251.

Overview

Goal: Understand the behavior of your simulation model

- Gain general understanding (today's focus)
- What factors are important?
- What choices of controllable factors are robust to uncontrollable factors?
- Which choice of controllable factors optimizes some performance measure?

Overview, Continued

Challenge: Exploring the parameter space

- Ex: 100 parameters, each "high" or "low"
- Number of combinations to simulate: $2^{100} \approx 10^{30}$
- Say each simulation consists of one floating point operation(!)
- Use world's fastest computer: Summit (148.6 petaflops)
- Required time for simulation: approximately 271,000 years

Basic Concepts: Factors

Factors (simulation inputs)

- Have impact on responses (simulation outputs)
- Levels: Values of a factor used in experiments
- Factor taxonomy:
  - Quantitative vs qualitative (can encode qualitative)
  - Discrete vs continuous
  - Binary or not
  - Controllable vs uncontrollable
- Factors must be carefully defined
  - Ex: (s, S)-inventory model
  - Use (s, S) or $(s, S - s)$ as the factors?

Factor type	Example
quantitative (cont.)	Poisson arrival rate
quantitative (discr.)	# of machines
qualitative	service policy (FIFO, LIFO, ...)
binary	(open,closed), (high,low),...
controllable	# of servers
uncontrollable	weather (sun, rain, fog)

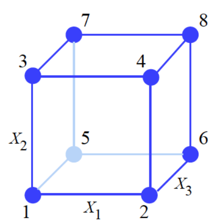
Basic Concepts: Designs

Design matrix

- One column per factor
- Each row is a design point
  - Contains a level for each factor
  - Level values determined by a domain expert
  - Natural or coded design levels
- Can have multiple replications of the design
  - Especially in simulation!

Design point	Factor settings x_1	x_2	x_3
1	−1	−1	−1
2	+1	−1	−1
3	−1	+1	−1
4	+1	+1	−1
5	−1	−1	+1
6	+1	−1	+1
7	−1	+1	+1
8	+1	+1	+1

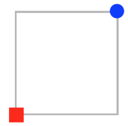
2^3 factorial design



Some Bad Designs: Capture the Flag

Confounded effects

Speed	Stealth	Success?
Low	Low	No
High	High	Yes

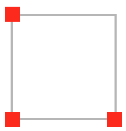


- Claim: Speed is the most important
- Claim: Stealth is the most important
- Claim: Both are equally important
- There is no way to determine who is right without more data
- Moral: haphazardly choosing design points can use up a lot of time while not providing insight

One-factor-at-a-time (OFAT) sampling

- Claim: Neither speed nor stealth is important
- Problem: an interaction between two factors is being missed

Speed	Stealth	Success?
Low	Low	No
High	Low	No
Low	High	No



Understanding Simulation Behavior: Metamodels

Simulation metamodels approximate true response

- Simplified representation for greater insight
- Allows "simulation on demand"
- Allows factor screening and optimization

Main-effects metamodel (quantitative factors)

$$R(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

Metamodel with second-order interaction effects

$$R(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \sum_i \sum_j \beta_{ij} x_i x_j + \epsilon$$

- R = simulation model output (i.e., response)
- Factors $x = (x_1, \ldots, x_k)$
- ϵ = mean-zero noise term, often assumed to be $N(0, \sigma^2)$

A Classical Design: 2^k Factorial Design

Basic setup: k factors with two levels each $(-1, +1)$

- Metamodel for $k = 2$: $R(x) = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$
- So $r(x) = E[R(x)] = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$

Estimating "main effects"

- Avg. change in r when x_1 goes from -1 to $+1$ (x_2 fixed):
  - $\frac{(r_3 - r_1) + (r_4 - r_2)}{2} = \frac{-r_1 - r_2 + r_3 + r_4}{2} = \frac{\mathbf{r} \cdot \mathbf{x_1}}{2} = 2\beta_1$
- Similarly, $\frac{\mathbf{r} \cdot \mathbf{x_2}}{2} = 2\beta_2$
- Method-of-moments estimators: $2\hat{\beta}_1 = \frac{\mathbf{R} \cdot \mathbf{x_1}}{2}$ and $2\hat{\beta}_2 = \frac{\mathbf{R} \cdot \mathbf{x_2}}{2}$

Design point	Factor settings x_1	x_2	$x_1 x_2$	Observed response (R)	Predicted expected value (r)
1	-1	-1	$+1$	R_1	$r_1 = -\beta_1 - \beta_2 + \beta_{12}$
2	-1	$+1$	-1	R_2	$r_2 = -\beta_1 + \beta_2 - \beta_{12}$
3	$+1$	-1	-1	R_3	$r_3 = \beta_1 - \beta_2 - \beta_{12}$
4	$+1$	$+1$	$+1$	R_4	$r_4 = \beta_1 + \beta_2 + \beta_{12}$

2^k Factorial Design, Continued

Estimating "interaction effect"

- (Effect of $\uparrow x_1$ with x_2 high minus effect with x_2 low) / 2
  - $\frac{(r_4 - r_2) - (r_3 - r_1)}{2} = \frac{\mathbf{r} \cdot (\mathbf{x_1 x_2})}{2} = 2\beta_{12}$
- Method of moments estimator: $2\hat{\beta}_{12} = \frac{\mathbf{R} \cdot (\mathbf{x_1 x_2})}{2}$

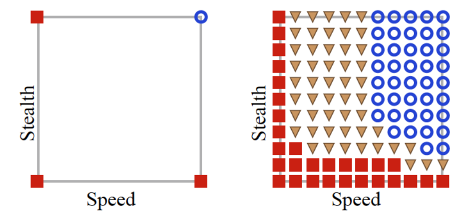
Observations:

- Can replicate design to get (Student-t) CI's for coefficients
- Estimating effects $\Leftrightarrow$ estimating regression coefficients
- Above analysis generalizes to more factors, e.g.,

$$R(x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

Design point	Factor settings x_1	x_2	$x_1 x_2$	Observed response (R)	Predicted expected value (r)
1	-1	-1	$+1$	R_1	$r_1 = -\beta_1 - \beta_2 + \beta_{12}$
2	-1	$+1$	-1	R_2	$r_2 = -\beta_1 + \beta_2 - \beta_{12}$
3	$+1$	-1	-1	R_3	$r_3 = \beta_1 - \beta_2 - \beta_{12}$
4	$+1$	$+1$	$+1$	R_4	$r_4 = \beta_1 + \beta_2 + \beta_{12}$

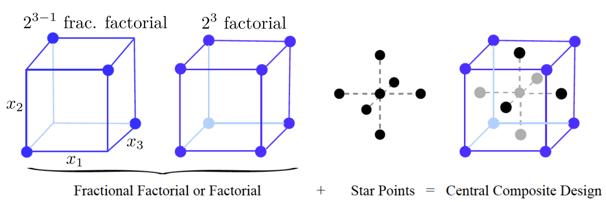
m^k Designs



Using more than two levels gives more detail

- E.g., capture the flag with 2^2 versus 11^2 designs
  - After achieving a minimal level of stealth, speed is more important
- Only possible for very small number of factors

2^{k-p} Fractional Factorial and Central Composite Designs

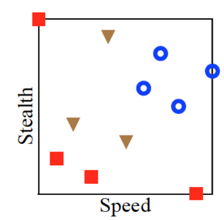


2^{k-p} fractional factorial designs

- Fewer design points, carefully chosen (see Law, Table 12.17)
  - E.g., 2^{3-1} design with 4 design points
  - Left/right faces: 1 val. of x_2 at each level, 1 val. of x_3 at each level (can isolate x_1 effect)
  - Similarly for other face pairs
- The degree of confounding is specified by the resolution
  - No m-way and n-way effect are confounded if $m + n <$ resolution
  - So for Resolution V design, no main effect or 2-way interaction are confounded

Space-Filling Designs

DP	Speed	Stealth	DP	Speed	Stealth
1	1	11	7	10	1
2	3	5	8	4	2
3	7	7	9	11	8
4	2	3	10	8	9
5	5	10	11	9	6
6	6	4			



Random Latin Hypercube design

- Based on random permutations of levels for each factor
- Good coverage of param. space w. relatively few design points
- Carefully crafted LH designs are needed in practice

Experimental Design for Simulation

Overview

Basic Concepts and Terminology

Pitfalls

Regression Metamodels and Classical Designs

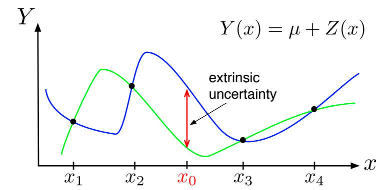
Other Metamodels

Data Farming

Gaussian Metamodeling (Kriging)

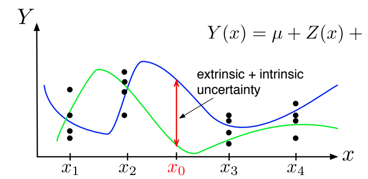
Ordinary kriging (deterministic simulations)

- $Z(x)$ is a Gaussian process
- Models uncertainty due to interpolation
- $(Z(v_1), Z(v_2), \ldots, Z(v_n)) \sim N(\mathbf{0}, \mathbf{R}(\theta))$
- $r(v_i, v_j) = e^{-\theta(v_i - v_j)^2}$
- $\hat{Y}(x_0) = \hat{\mu} + \mathbf{r}^\top(x_0)\mathbf{R}(\hat{\theta})^{-1}(\mathbf{Y} - \mathbf{1}\hat{\mu})$
  - $\hat{\mu}$ and $\hat{\theta}$ are MLE estimates
  - $\mathbf{Y} = (Y_1, \ldots, Y_m)$ and $\mathbf{1} = (1, 1, \ldots, 1)$
  - $\mathbf{r} = \big(r(x_0, x_1), r(x_0, x_2), \ldots, r(x_0, x_m)\big)$

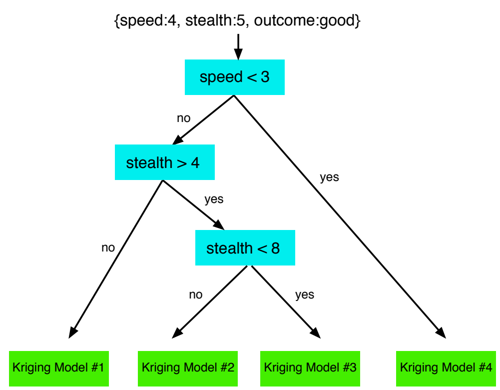


Stochastic kriging (stochastic simulations)

- ϵ is $N(0, \sigma^2)$ ("the nugget")
- Captures simulation variability
- Many other variants
  - Fitted derivatives
  - Varying σ^2
  - Non-constant mean function



Kriging + Trees



Idea: Build multiple models on subsets of homogeneous data

- Recursively split data to
  - Maximize heterogeneity (e.g., Gini index)
  - Maximize goodness of fit statistic (e.g., R^2)
- Build model on each subset

Experimental Design for Simulation

Overview

Basic Concepts and Terminology

Pitfalls

Regression Metamodels and Classical Designs

Other Metamodels

Data Farming

Data Farming

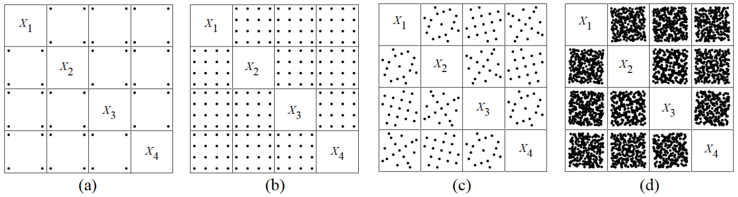


Figure 4: Scatterplot matrices for selected factorial and nearly orthogonal Latin hypercube (NOLH) designs: (a) 2^4 factorial with 16 design points, (b) 4^4 factorial with 256 design points, (c) NOLH with 17 design points, and (d) NOLH with 257 design points.

Modern "big data" approach

- Unlike real-world experiments, easier to generate a lot of simulation data
- Most effort usually spent building model, so work it hard!
- Use analytical, graphical, and data mining techniques on generated data

Graphical Methods

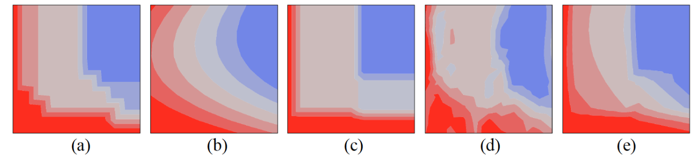
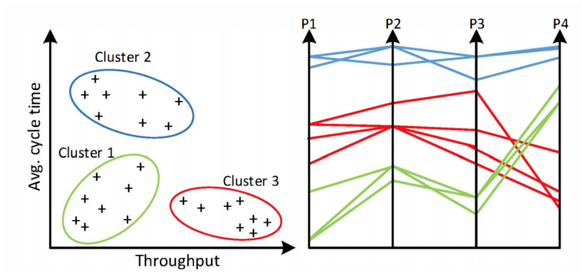


Figure 7: Capture-the-flag contour plots. (a) displays the actual response based on an 11×11 grid. The remaining contours are metamodels following a 65-dp NOLH: (b) 2nd-order regression metamodel, (c) partition tree with five splits, (d) Gaussian process metamodel, and (e) regression/partition metamodel.

Gaining insight through visualizations

- More sophisticated methods than simple regression
- Analyze flat areas (robustness)
- Other characteristics of interest

Data Mining and Visual Analytics



Visual analytics

- Experiments are clustered based on system performance
- Parallel-coordinate plot relates performance to factor levels
- Ex: Manufacturing model with parameters P1, P2, P3, P4

N. Feldkamp, S. Bergmann, and S. Strassburger. Visual analytics of manufacturing simulation data. *Proc. Winter Simulation Conference*, 2015, pp. 779–790.