Some aspects of Design of Experiments

Nancy Reid

University of Toronto

June 28, 2007

Statistician's view

- intervention applied to experimental units
- interventions conventionally called treatments
- treatments normally randomized to units, sometimes with restraints
- response under various treatments to be compared
- intervention provides a basis for stronger conclusions on how treatment affects response

agriculture	types of fertilizer	plots of land	yield
'technology'	reaction time, concentration	samples subject to biochemical reaction	percent contamination
computer experiments	settings for systematics	simulation runs	output (climate model epidemic model,)

Factorial experiments

- treatments are combinations of levels of several factors
- time, concentration, pressure, temperature, ...
- very common to combine each factor at each of 2 levels $\rightarrow 2^k$ designs
- e.g. 10 systematic parameters; several runs at 'mean' value; several runs with each systematic at $\pm 1\sigma$
- "OFAT", one factor at a time
- full factorial provides better estimation of mean effects with same resources
- Example 2^4 factorial design

Four factors at each of 2 levels

run	A	B	C	D
1	−1	−1	−1	−1
2	−1	−1	−1	+1
3	−1	−1	+1	−1
4	−1	−1	+1	+1
5	−1	+1	−1	−1
6	−1	+1	−1	+1
7	−1	+1	+1	−1
8	−1	+1	+1	+1
9	+1	−1	−1	−1
10	+1	−1	−1	+1
11	+1	−1	+1	−1
12	+1	−1	+1	+1
13	+1	+1	−1	−1
14	+1	+1	−1	+1
15	+1	+1	+1	−1
16	+1	+1	+1	+1

... 2^4 factorial

- estimation of average effect of changing *A*, *B*, *C*, *D* are each based on 8 observations at each level
- efficiency increased by a factor of 4 over OFAT

- 11 further estimates available (need 1 for overall mean)
- can estimate all possible interactions: *AB*, *AC*, ... *CD*, *ABC*, *ABD*, *ACD*, *BCD*, *ABCD*
- many of these will be 'noise': use for internal replication

... 2^4 factorial

run	A	B	C	D
1	-1	-1	-1	-1
2	-1	-1	-1	$+1$
3	-1	-1	$+1$	-1
4	-1	-1	$+1$	$+1$
5	-1	$+1$	-1	-1
6	-1	$+1$	-1	$+1$
7	-1	$+1$	$+1$	-1
8	-1	$+1$	$+1$	$+1$
9	$+1$	-1	-1	-1
10	$+1$	-1	-1	$+1$
11	$+1$	-1	$+1$	-1
12	$+1$	-1	$+1$	$+1$
13	$+1$	$+1$	-1	-1
14	$+1$	$+1$	-1	$+1$
15	$+1$	$+1$	$+1$	-1
16	$+1$	$+1$	$+1$	$+1$

... 2^4 factorial

run	A	B	C	D	response
1	-1	-1	-1	-1	$y_{(1)}$
2	-1	-1	-1	$+1$	y_d
3	-1	-1	$+1$	-1	y_c
4	-1	-1	$+1$	$+1$	y_{cd}
5	-1	$+1$	-1	-1	y_b
6	-1	$+1$	-1	$+1$	y_{bd}
7	-1	$+1$	$+1$	-1	y_{bc}
8	-1	$+1$	$+1$	$+1$	y_{bcd}
9	$+1$	-1	-1	-1	y_a
10	$+1$	-1	-1	$+1$	y_{ad}
11	$+1$	-1	$+1$	-1	y_{ac}
12	$+1$	-1	$+1$	$+1$	y_{acd}
13	$+1$	$+1$	-1	-1	y_{ab}
14	$+1$	$+1$	-1	$+1$	y_{abd}
15	$+1$	$+1$	$+1$	-1	y_{abc}
16	$+1$	$+1$	$+1$	$+1$	y_{abcd}

... 2^4 factorial

run	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC...
1	−1	−1	−1	−1	+1	+1	+1	+1	+1	+1	−1
2	−1	−1	−1	+1	+1	+1	−1	+1	−1	−1	−1
3	−1	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1
4	−1	−1	+1	+1	+1	−1	−1	−1	−1	+1	+1
5	−1	+1	−1	−1	−1	+1	+1	−1	−1	+1	+1
6	−1	+1	−1	+1	−1	+1	−1	−1	+1	−1	+1
7	−1	+1	+1	−1	−1	−1	+1	+1	−1	−1	−1
8	−1	+1	+1	+1	−1	−1	−1	+1	+1	+1	−1
9	+1	−1	−1	−1	−1	−1	−1	+1	+1	+1	+1
10	+1	−1	−1	+1	−1	−1	+1	+1	−1	−1	+1
11	+1	−1	+1	−1	−1	+1	−1	−1	+1	−1	−1
12	+1	−1	+1	+1	−1	+1	+1	−1	−1	+1	−1
13	+1	+1	−1	−1	+1	−1	−1	−1	−1	+1	−1
14	+1	+1	−1	+1	+1	−1	+1	−1	+1	−1	−1
15	+1	+1	+1	−1	+1	+1	−1	+1	−1	−1	+1
16	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1

... 2^4 factorial

- pool five 3-factor interactions and one 4-factor interaction to estimate error
- or, assign higher order interactions to new factors $\rightarrow$ fractional factorial
- e.g., assign new factor E to 4-factor interaction ABCD
- obtain information on 5 main effects from 16 runs
- every 2-factor interaction aliased with a 3-factor interaction
- continuing, assign new factor F to, for example, ABC (=DE); now some 2-factor interactions are aliased with each other
- 6 factors, 16 runs (instead of 2^6)

8 run screening design for 7 factors

run	A	B	C	D	E	F	G
1	−1	−1	−1	+1	+1	+1	−1
2	−1	−1	+1	−1	−1	+1	+1
3	−1	+1	−1	−1	+1	−1	+1
4	−1	+1	+1	+1	−1	−1	−1
5	+1	−1	−1	+1	−1	−1	+1
6	+1	−1	+1	−1	+1	−1	−1
7	+1	+1	−1	−1	−1	+1	−1
8	+1	+1	+1	+1	+1	+1	+1

... fractional factorial designs

- screening a large number of factors in few runs;
  most factors expected to be inactive
- inactive factors provide replication
- alternatively, investigating a smaller number of factors and interactions
- somewhat more complicated to run
- not suitable if factor levels are difficult to change
- if one or more runs are lost, considerable information is lost
- may need to block runs to ensure homogeneity
- for example if all runs cannot be completed in one day and there is concern about drift of conditions over time

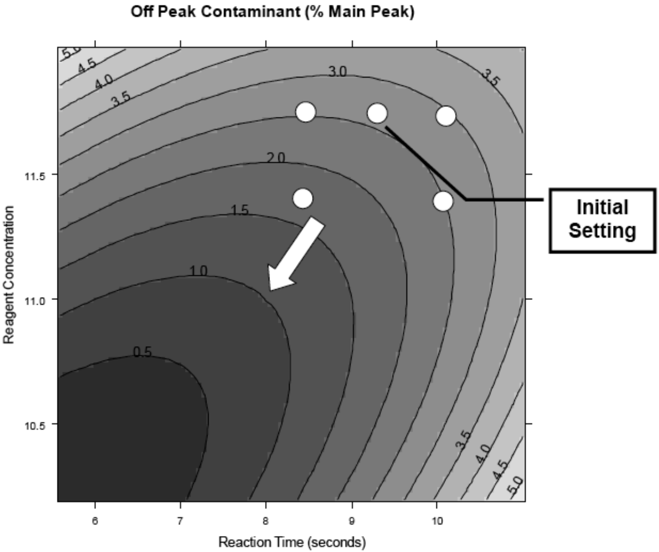
Analysis of the data

- **very easy** if we use a linear model with Gaussian error: $y = Z\beta + \epsilon$
- Z has columns with entries ± 1, plus a column of 1s
- in fact in this case nearly everything can be quickly computed by hand
- not difficult to generalize to non-Gaussian and non-linear (in β) models, either using likelihood methods or some transformation of the response
- standard regression software usually fits both Gaussian and at least a selection of nonGaussian models
- in `R`, `lm` for linear models and `glm` for generalized linear models

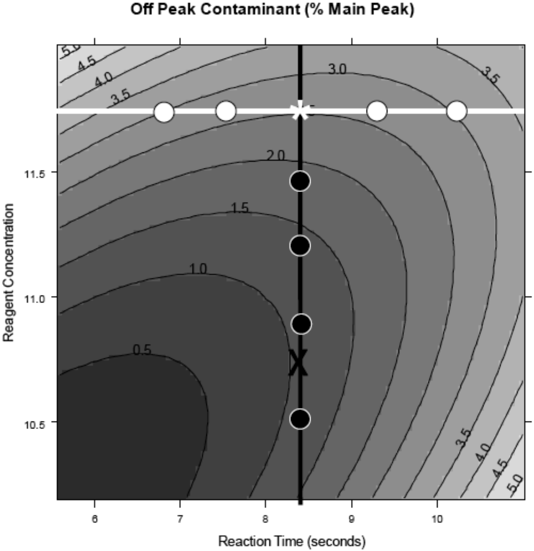
... analysis of data

- quantitative factors (temperature, pressure etc.), goal might be to maximize (or minimize) response
- sequential experimentation in relevant ranges starts with screening design
- points added in direction of response increase
- near the maximum additional levels added, to model curvature in response surface
- goal might be to see which values of systematics produce simulated data consistent with observations: derived response to be minimized
- goal might be to see which systematic parameters affect simulation output

from Gunter (2007), a 2^2 design

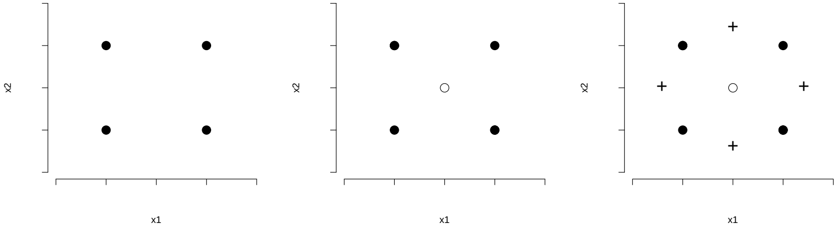


... and an OFAT design



curvature in response

- if settings correspond to quantitative factors, x_1, x_2, etc., then interaction corresponds to $x_1 x_2$
- other quadratic terms x_1^2 etc. can only be measured by adding further points
- usually added at center $(0, \ldots, 0)$ and on radius of a circle
- central composite designs



Orthogonal arrays

run	A	B	C	D	E	F	G
1	-1	-1	-1	$+1$	$+1$	$+1$	-1
2	-1	-1	$+1$	-1	-1	$+1$	$+1$
3	-1	$+1$	-1	-1	$+1$	-1	$+1$
4	-1	$+1$	$+1$	$+1$	-1	-1	-1
5	$+1$	-1	-1	$+1$	-1	-1	$+1$
6	$+1$	-1	$+1$	-1	$+1$	-1	-1
7	$+1$	$+1$	-1	-1	-1	$+1$	-1
8	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$

- an eight-run design for seven factors; in the jargon a 2^{7-4} fractional factorial; also called a Plackett-Burman design
- a two-symbol array of size $n \times n - 1$ can be generated by an $n \times n$ Hadamard matrix
- can be generalized to more than two levels (symbols), giving an extension of fractional factorials

Example: a 3-level OA

run	A	B	C	D	E	F
1	−1	−1	−1	−1	−1	−1
2	−1	0	0	0	0	0
3	−1	+1	+1	+1	+1	+1
4	0	−1	−1	0	0	+1
5	0	0	0	+1	+1	−1
6	0	+1	+1	−1	−1	0
7	+1	−1	0	−1	+1	0
8	+1	+1	−1	+1	0	−1
9	+1	+1	−1	+1	0	+1
10	−1	−1	+1	+1	0	0
11	−1	0	−1	−1	+1	+1
12	−1	+1	0	0	−1	−1
13	0	−1	0	+1	−1	+1
14	0	0	+1	−1	0	−1
15	0	+1	−1	0	+1	0
16	+1	−1	+1	0	+1	−1
17	+1	0	−1	+1	−1	0
18	+1	+1	0	−1	0	+1

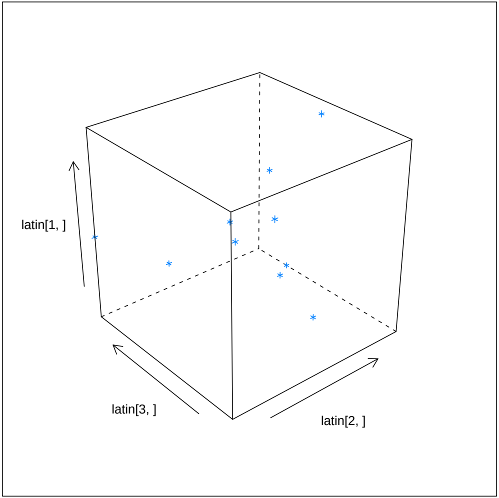
Space-filling designs

- exploration of possibly complex response surface
- examples: computer experiments, epidemic modelling, simulations
- Latin hypercube, uniformly scattered, ...
- often used in numerical integration; quasi Monte Carlo

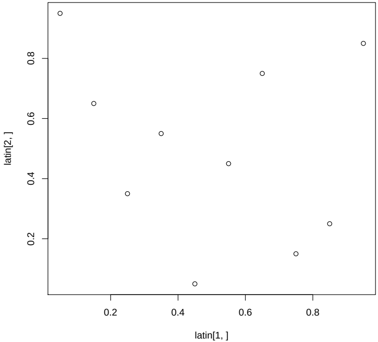
$$\int_{R^k} f(x)dx \approx \frac{1}{n}\sum f(X_i)$$

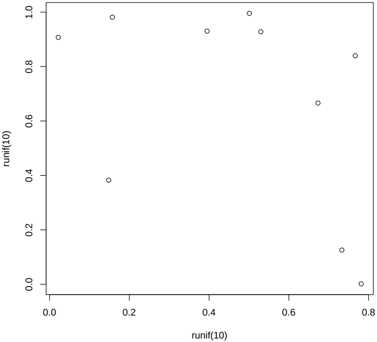
- X_i evaluated at 'space-filling' points instead of sampled uniformly

... space-filling designs



... space-filling designs





trading mean against variance

- work by Zi Jin, following suggestion by Radford Neal
- a simulation run generates M background events
- and mistakenly identifies some number y of these as signal
- with some small probability p, say $p = 0.001$ or $p = 0.0001$
- model y as Poisson with parameter p
- p depends on various settings for systematics
- approximate this dependence by a Gamma distribution with mean 0.001 or 0.0001
- find out how (Fisher) information in full set of N simulation runs depends on trade-off between sampling K different values of the parameters and increasing the number of events generated at each parameter value

... mini-Boone

- Example: mean $p = 0.0001$, allow approximately $\pm 3\sigma$
- $N = 2,000,000$ simulations
- optimal split between number of events and number of samples is approximately $M = 160,000$ and $K = 13$
- for larger mean $p = .001$, optimal split is $M = 16,000$ and $K = 130$
- 10-fold increase in information over the (arbitrary) choice $M = 100,000$ events and $K = 20$ parameter values
- these values are randomly chosen from a fairly ad-hoc distribution
- suggests that it will be worthwhile to investigate information in orthogonal array designs

... mini-Boone

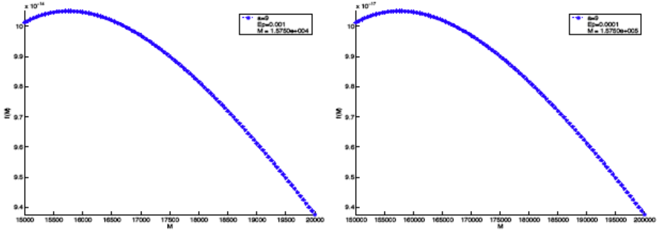


Figure 1: $a = 9, Ep = 0.001, \hat{M} = 15750$ and $a = 9, Ep = 0.0001, \hat{M} = 157500$.

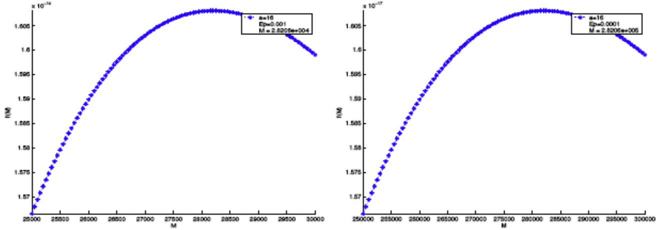


Figure 2: $a = 16, Ep = 0.001, \hat{M} = 28206$ and $a = 16, Ep = 0.0001, \hat{M} = 282060$.

References

1. Gunter, B.H. (1993) *Computers in Physics* **7**, 262–272.
2. Gunter, B.H. (2007?) *Sequential Experimentation*.
3. Cox, D.R. and Reid, N. (2000) *The Theory of Design of Experiments.* Chapman & Hall/CRC, London.
4. Owen, A. (1992). Orthogonal arrays for computer experiments, integration and visualisation. *Statistica Sinica* **2**, 459–462.
5. Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989). Design and analysis of computer experiments. *Statistical Science* **4** 409–423.
6. Welsh, J.P., Koenig, G.G. and Bruce, D. (1997). Screening design for model sensitivity studies. *J. Geophys. Res.* **102** D14, 16,499–16,505.