Bridge Designs for Modeling Systems with Small Error Variance

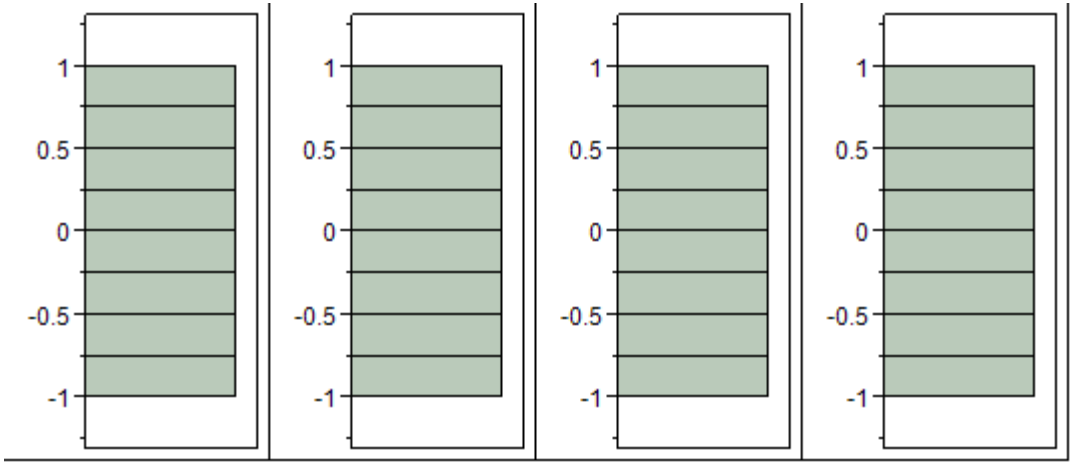
Bradley Jones
JMP Division of SAS

Motivation

Latin Hypercube designs (space filling) are the default for deterministic computer simulation studies.

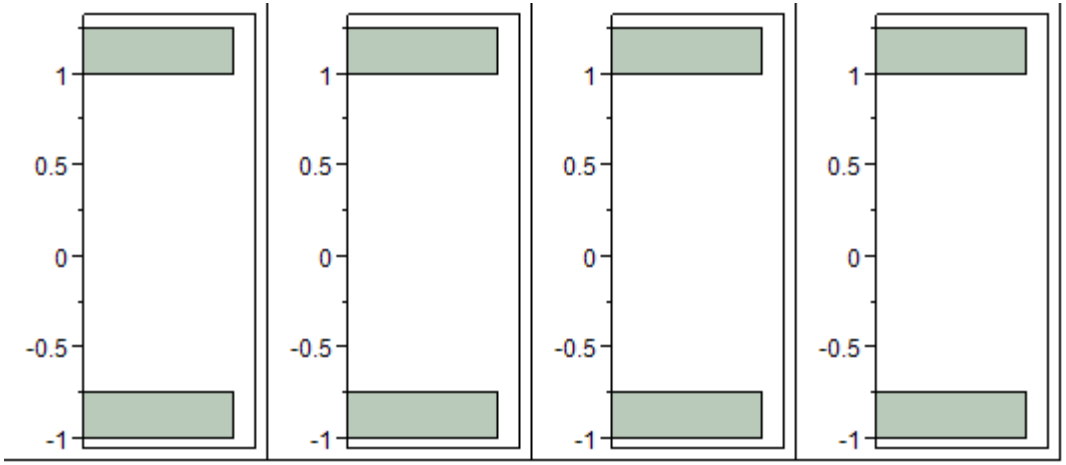
Factorial designs are standard for modeling systems with substantial noise.

Illustration of the Uniform 1d Projections of the LHC



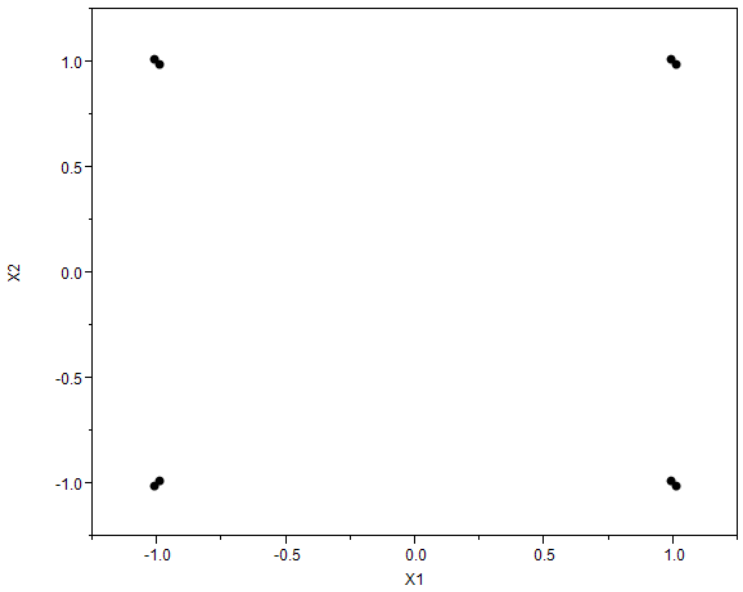
A four factor 40 run LHC – uniform spacing along each coordinate axis.

Illustration of 1d Projections of 2-Level Factorial Designs

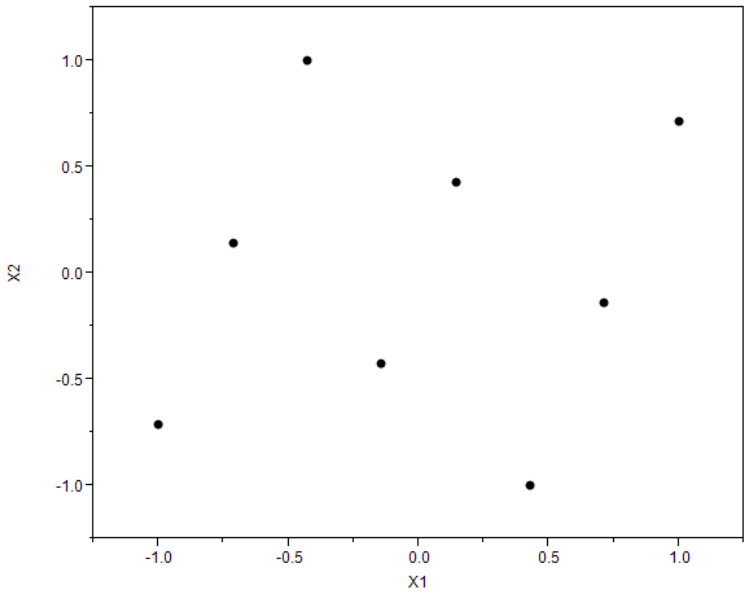


A 16 run 2x2x2x2 Full Factorial

Factorial vs. Latin Hypercube



Factorial



Latin Hypercube

Jitter added to reveal replication

Pros and Cons

Factorial (Two Level)

Pros

1. Good for precise parameter estimation
2. Good prediction for 1st order polynomials

Cons

1. Cannot detect curvature
2. Replication is useless if there is no noise

Latin Hypercube

Pros

1. Good for detecting and modeling curvature
2. Good for integral approximation

Cons

1. Inefficient estimation if there is noise
2. Less powerful for modeling interactions

Latin Hypercube Variance Efficiency

Number of Factors	Number of Runs	Relative Efficiency of LHC (%)
2	16	60.3
2	32	57.6
3	16	54.3
3	32	52.2
4	16	51.1
4	32	48.7
5	16	45.7
5	32	45.2

Average relative prediction variance for a main effects model comparing LHC and 2-level factorial and fractional factorial.

Coefficient Variance Comparison LHC vs FF

Effect	Variance	Variance
Intercept	0.036	0.031
X1	0.106	0.031
X2	0.103	0.031
X3	0.108	0.031
X4	0.112	0.031
X5	0.109	0.031
X6	0.104	0.031
X1*X2	0.384	0.031
X1*X3	0.397	0.031
X1*X4	0.398	0.031
X1*X5	0.341	0.031
X1*X6	0.482	0.031
X2*X3	0.353	0.031
X2*X4	0.302	0.031
X2*X5	0.411	0.031
X2*X6	0.422	0.031
X3*X4	0.441	0.031
X3*X5	0.464	0.031
X3*X6	0.368	0.031
X4*X5	0.407	0.031
X4*X6	0.38	0.031
X5*X6	0.291	0.031

6 Factors 32 Runs
Interactions model

Motivation…

When the response function is deterministic you are only concerned about the model bias induced by using a simpler surrogate model.

Prediction variance is a primary concern when the noise is larger than the signal.

When the noise is smaller than the signal, then it makes sense to consider both variance and bias in design construction.

Bridge Design

Goals

1. High D-efficiency for low order polynomial models
2. No replication in projection for GASP model fitting

Implementation Approach

Algorithmic – find a design that maximizes D-efficiency subject to a minimum projection distance between coordinates for each point pair.

Bridge Design Optimization Problem

$$\max \ |\mathbf{X}'\mathbf{X}|, \ \text{subject to} \ |x_{ir} - x_{is}| \geq \delta, \ i=1,\ldots,k, \ r,s=1,\ldots,n, \ r \neq s.$$

Where **X** is the design matrix for the specified polynomial model and x_{ij} is an element in the ith row of X corresponding to the jth factor setting.

Bridge Design Required Inputs

1. Number of design points
2. Polynomial model form
3. Minimum distance between point coordinates (tuning parameter)

Tuning parameter notes…

1. LHC has coordinate distance of 2/(n-1)

   e.g. 21 points implies closest coordinate is 0.1

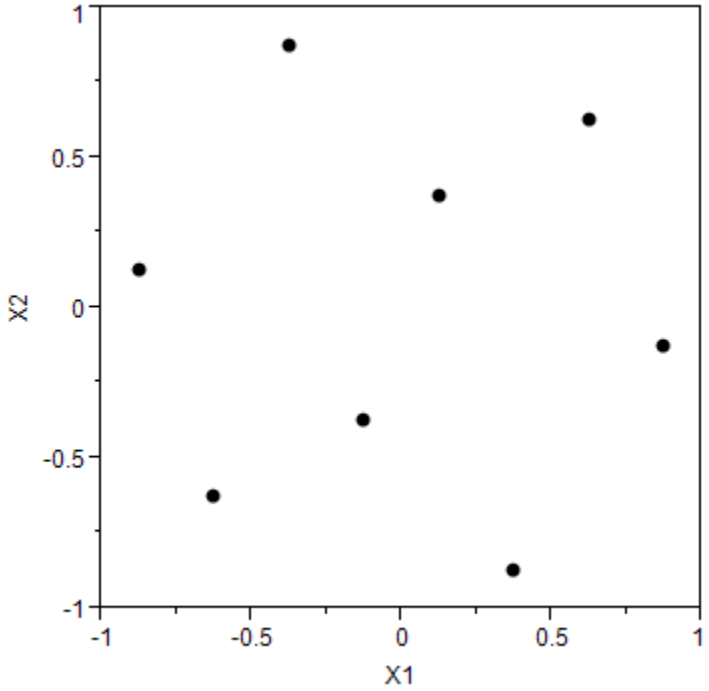
2. Factorial coordinate distance is zero.

3. “Half way” is 1/(n-1) i.e. coordinates must be at least 0.05 apart.

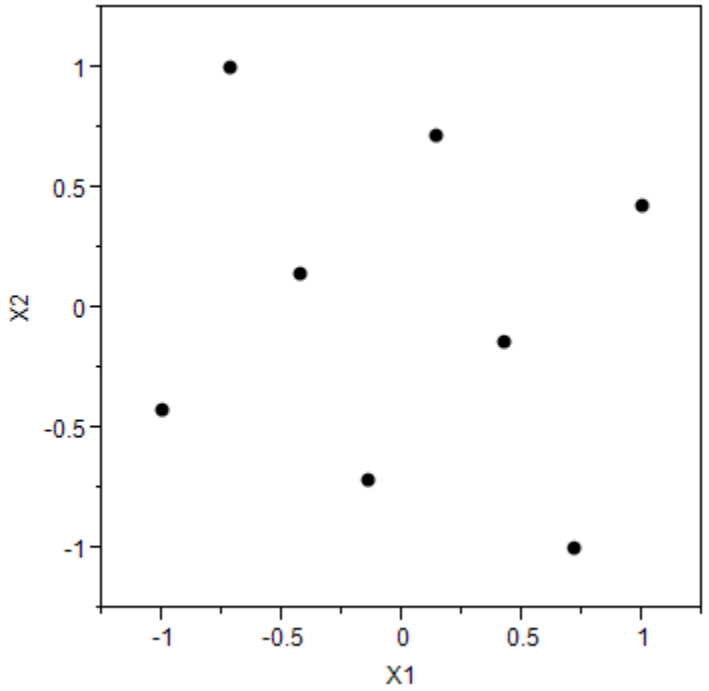
Polynomial model notes…

1. Main-effect models with small minimum distance requirements result in “X” shaped 2-d projections
2. Adding quadratic terms enforces coordinates near the middle of every factor’s range.

Bridge Design vs. Latin Hypercube



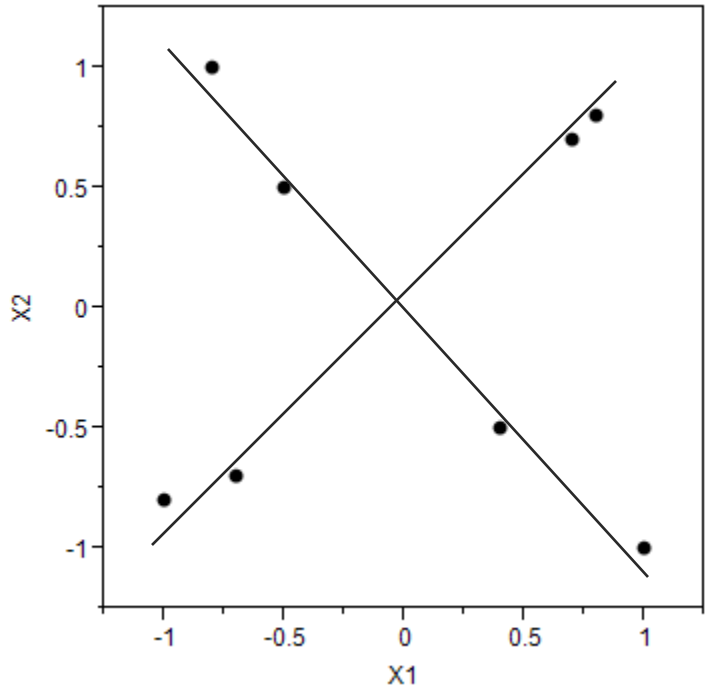
Bridge Design



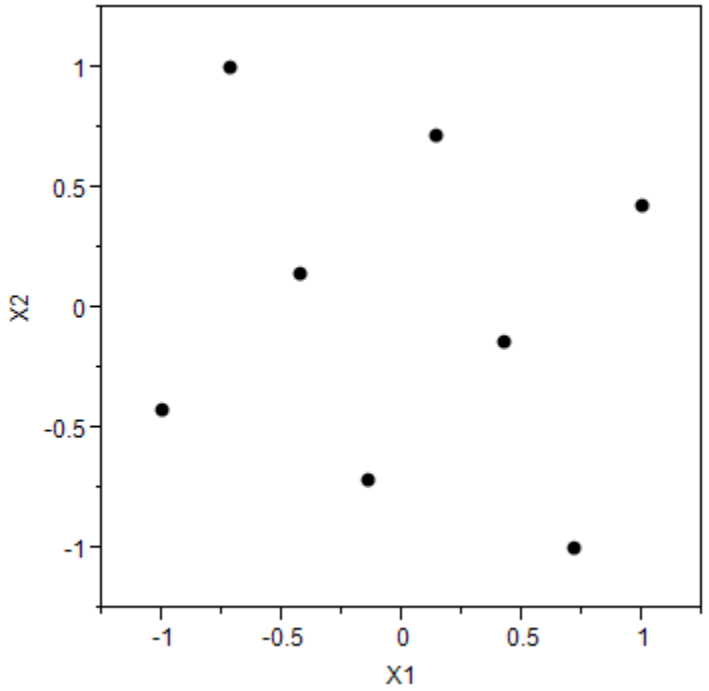
Latin Hypercube

Main-effects model
Minimum distance = 0.25

Bridge Design vs. Latin Hypercube



Bridge Design



Latin Hypercube

Main-effects model
Minimum distance = 0.1

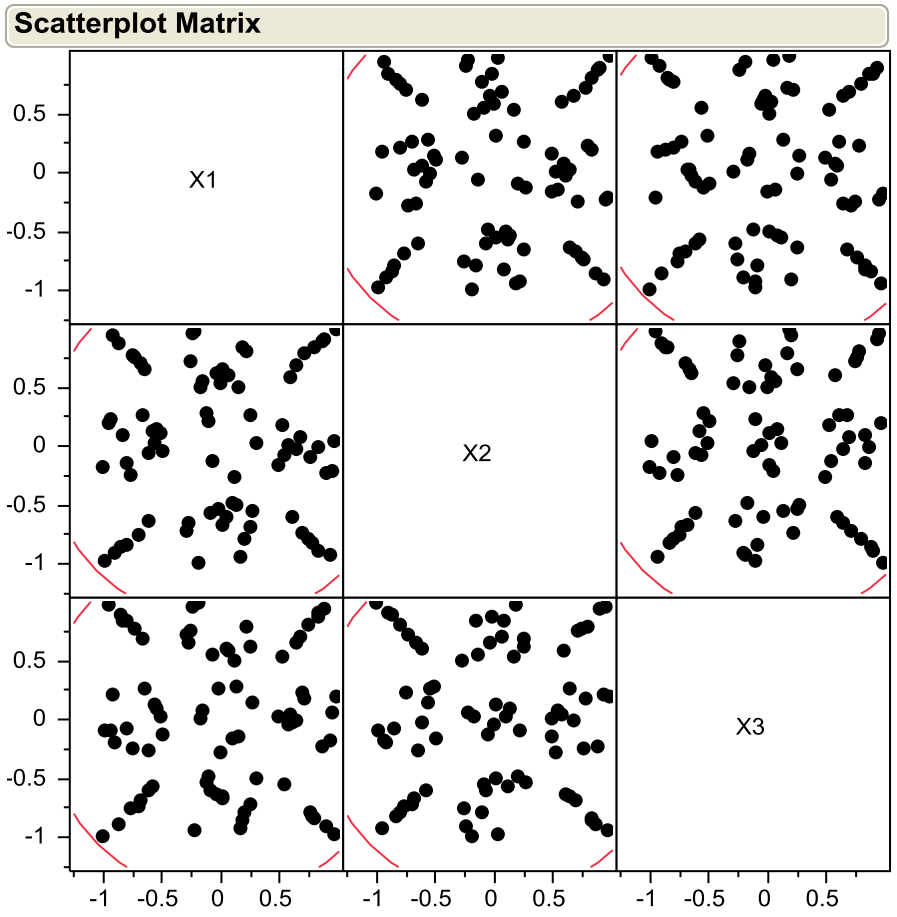
Avoiding the X Shape – Use Quadratic Terms

Correlations

	X1	X2	X3
X1	1.0000	0.0029	-0.0145
X2	0.0029	1.0000	0.0016
X3	-0.0145	0.0016	1.0000

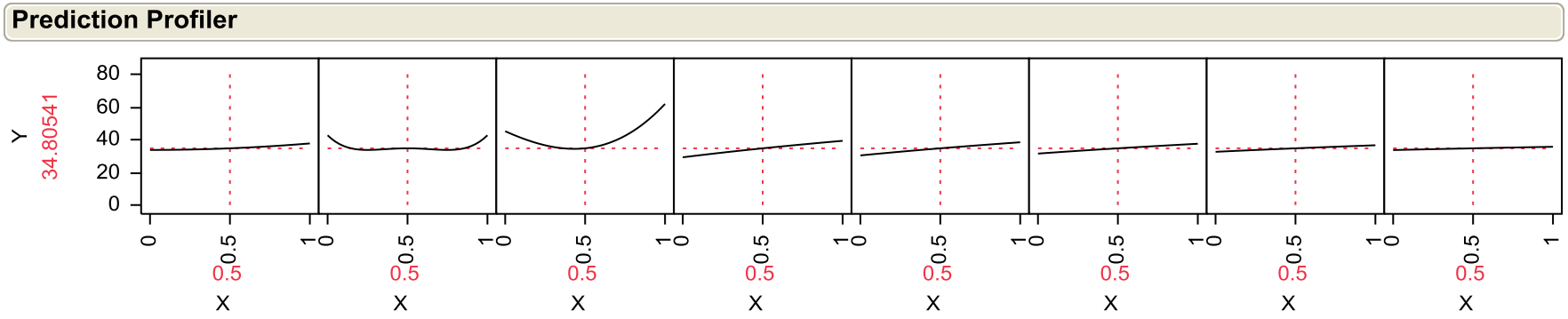
Bridge Design

Pure Quadratic model
Minimum distance = 0.0125

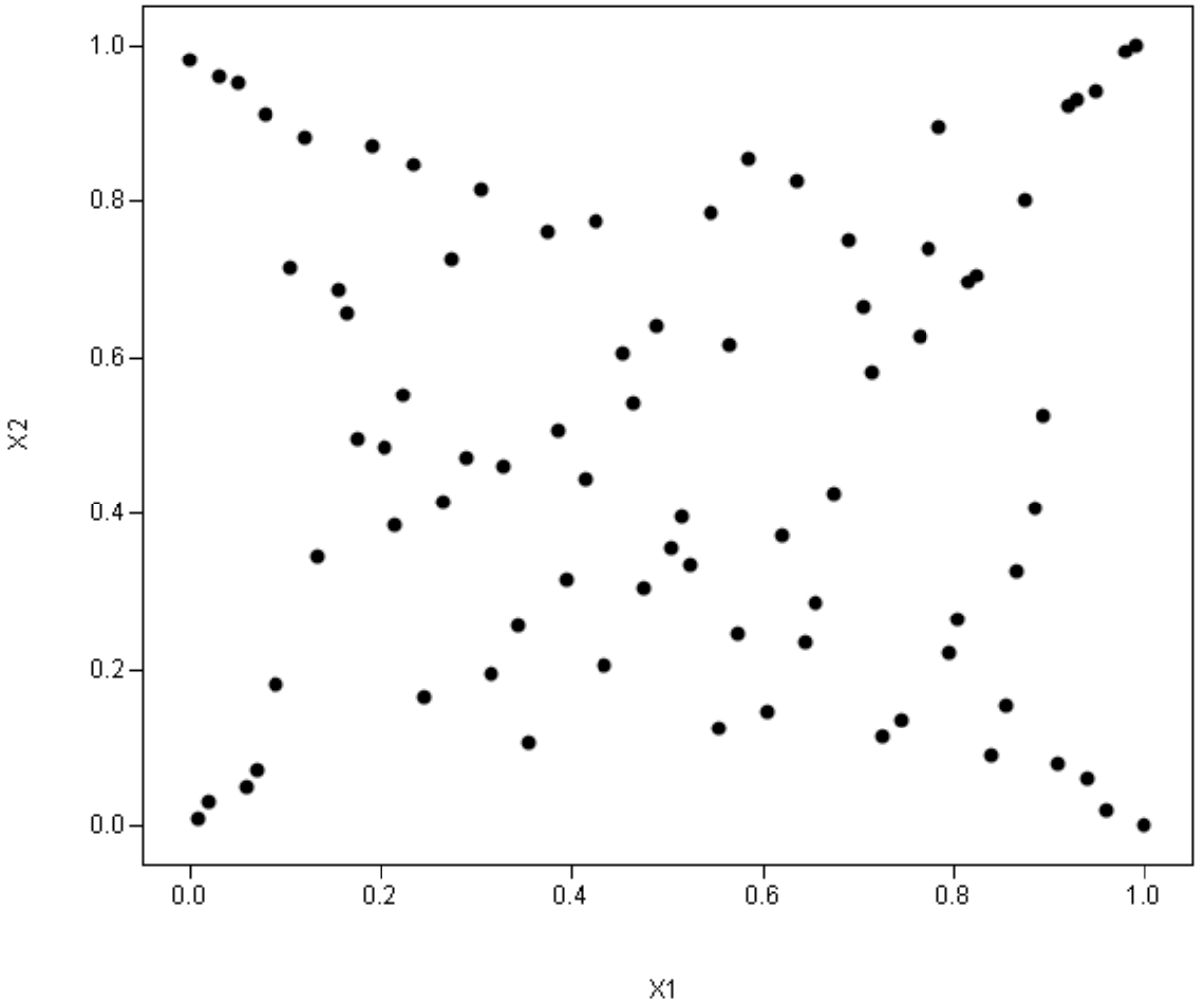


Sample Test Function – Dette & Pepelyshev (2010) Equation 6

$$
\begin{aligned}
&4 * \left(\left(\left(X1 - 2 \right) + 8 * X2 \right) - 8 * X2^{2} \right)^{2} \\
&+ \left(3 - 4 * X2 \right)^{2} \\
&+ 16 * \sqrt{X3 + 1} * \left(2 * X3 - 1 \right)^{2} \\
&+ 4 * \mathrm{Log}\left(1 + X3 + X4 \right) \\
&+ 5 * \mathrm{Log}\left(1 + X3 + X4 + X5 \right) \\
&+ 6 * \mathrm{Log}\left(1 + X3 + X4 + X5 + X6 \right) \\
&+ 7 * \mathrm{Log}\left(1 + X3 + X4 + X5 + X6 + X7 \right) \\
&+ 8 * \mathrm{Log}\left(1 + X3 + X4 + X5 + X6 + X7 + X8 \right)
\end{aligned}
$$



Bridge Design Bivariate View



Comparison with Dette & Pepelyshev

Design	RMSE
MLHD	3.08
GMLHD	1.47
Bridge	1.36

Deterministic Response

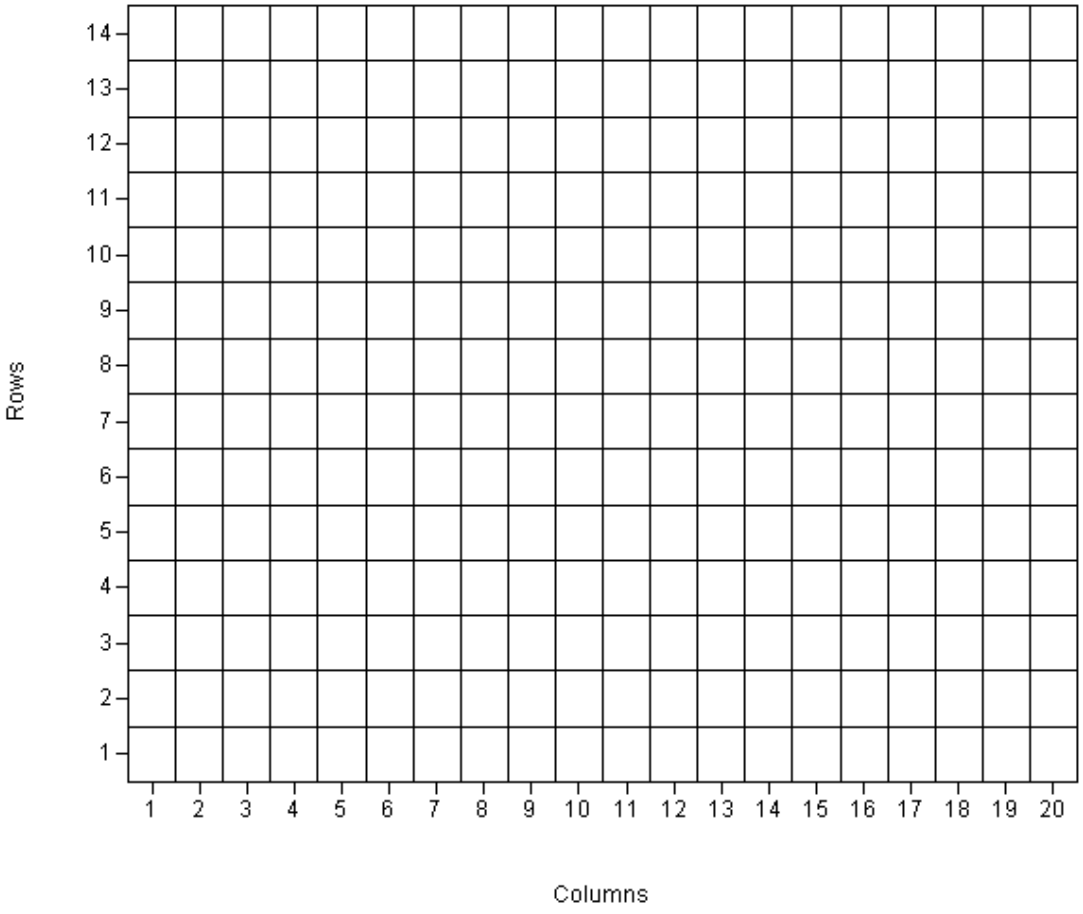
Design	RMSE
MLHD	5.07
GMLHD	2.88
Bridge	2.94

Deterministic Response
Plus Added Noise

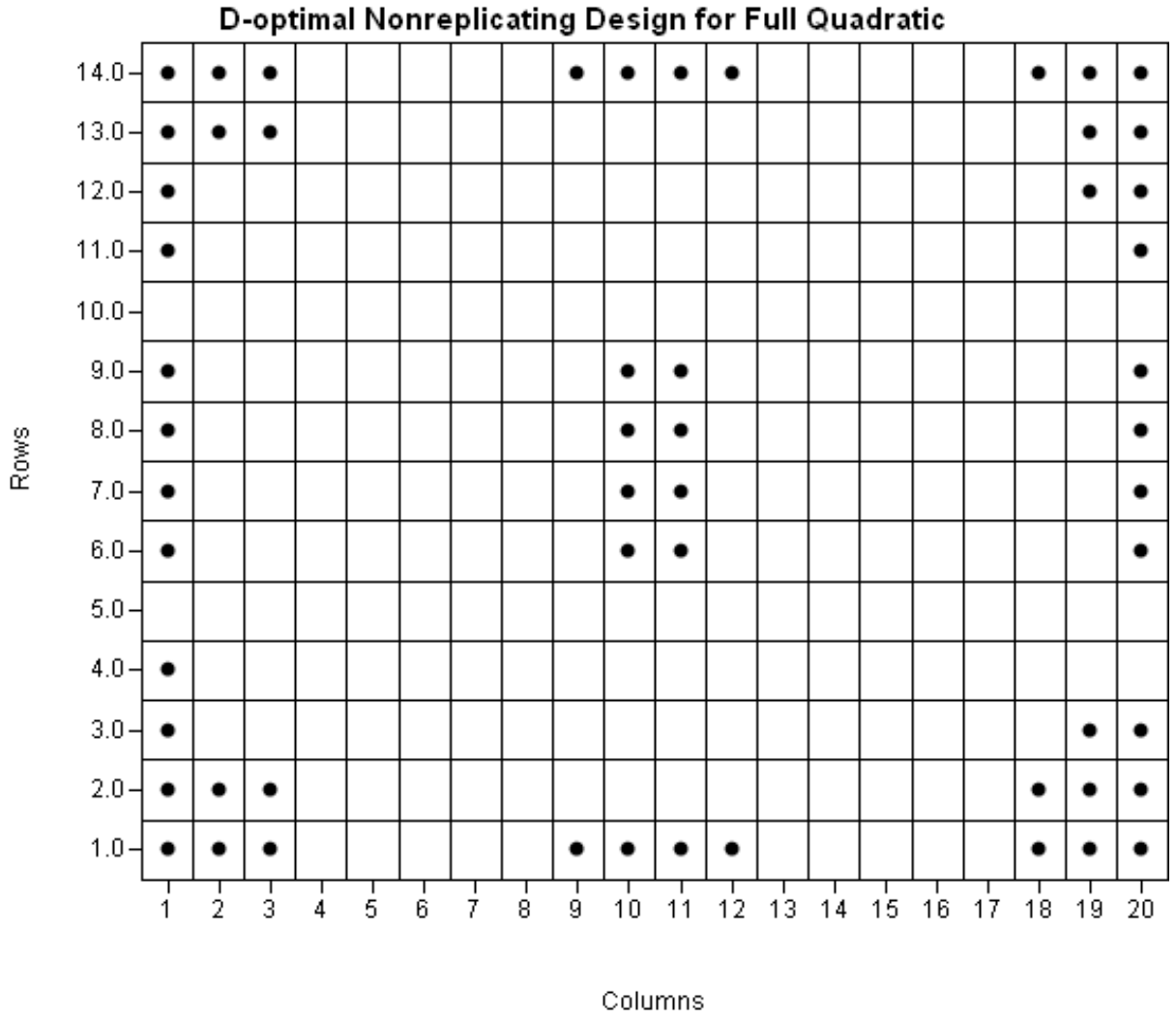
Summary

1. Bridge designs provide a tunable family of designs between Latin Hypercube and variance optimal designs.
2. Bridge designs allow for efficient polynomial estimation if there is noise in the response.
3. Bridge designs also allow for fitting GASP models by avoiding replication in one-dimensional projections.
4. Using higher order polynomial models can make bridge designs as “space filling” as desired.

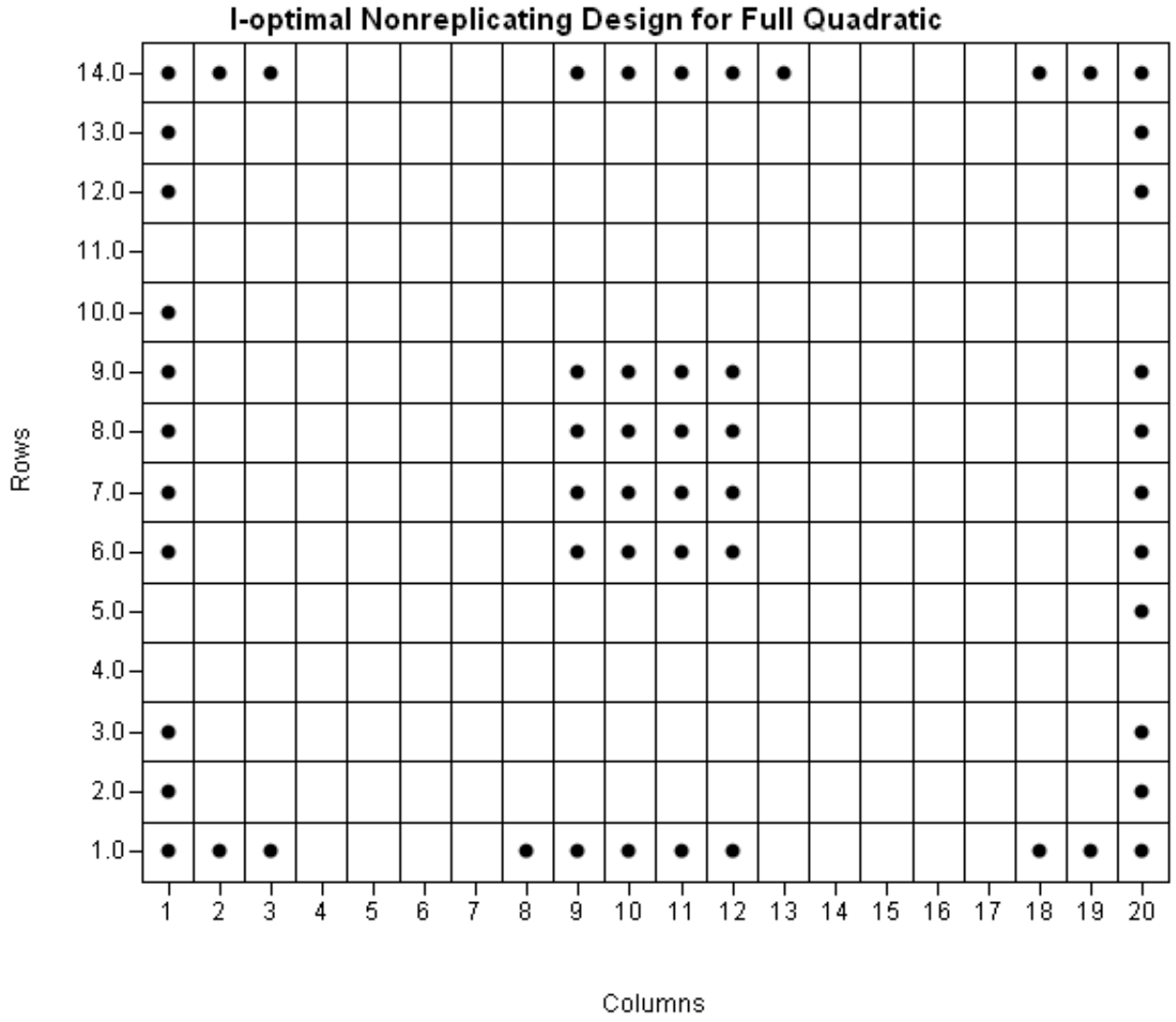
Rosemary's Problem from Last Week



We have a grid of plots with possible fertility differences over the grid. Choose 56 plots to model these differences.



D-optimal Design is 87.5% I-efficient



I-optimal Design is 96.5% D-efficient

Statistical Discovery.™ From SAS.

SEEING IS BELIEVING