Supersaturated Design:

A Review and Some Research Potentials

Dennis Lin
University Distinguished Professor
The Pennsylvania State University
DKL5@psu.edu

08 September, 2006
University of Southampton

Screening: What for???

- Factor
- Group of Factors (Team)
- Methods
- Parameters (x_i/x_j vs $\beta_i/\beta_i/\beta_{ij}/\beta_i^2/\beta_j^2$)
- Graduate Students (Screening Exam)
- Faculty (Tenure Review)
- ...

Model Screening

- Given $f_i \in$ F

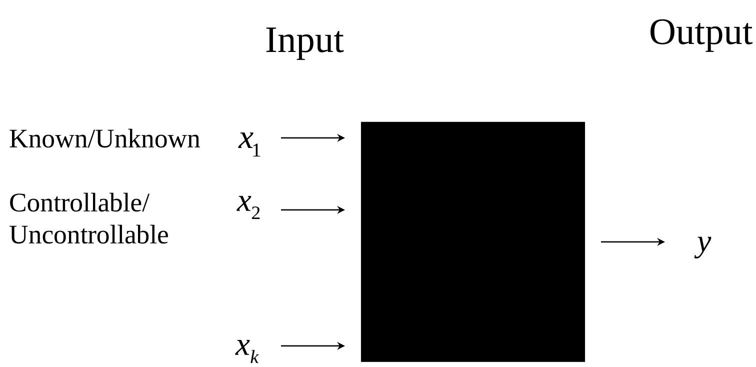
$$y=f_i(x)+\varepsilon$$

- Find $i=i^*$, such that

$$y=f_{i^*}(x)+\varepsilon$$

Which is most "appropriate".

Factor Screening



$$y = f(x_1, \ldots x_i, x_{i+1}, \ldots x_k) + \varepsilon$$

Situations: Good for Factor Screening

- Blood Tests
- Pareto Analysis
- Genetic Study
- Industrial Investigations
- Other Scientific Investigations
  - Physics, Chemistry, Biology
  - Human Resource, Business study etc.
- Others

Situations: Not Good for Factor Screening

Artificial Neural Network (ANN)
Drug Discovery
Others

First-Order Screening: Saturated Design

Lin (Handbook of Statistics, 2003)

- 2^{k-p} //R=III Design
- Plackett & Burman Design
- Optimal Design
- p-efficient Design
- Simplex Design
- T-optimal Design
- Uniform Design

Table 7
A summarized comparison on selected screening designs

Section	Design	Run	Level	Remarks
2.0	2_{III}^{k-p}	2^r	2	Orthogonal & symmetry
2.1	P&B	4t	2	Orthogonal & symmetry
2.2	Optimal	any	many	Nonorthogonal & asymmetry
2.3	p-eff	$k+1$	2	Nonorthogonal & symmetry
2.4	Simplex	$k+1$	many	Orthogonal & asymmetry
2.5	T-opt	$k+1$	many	Orthogonal & symmetry
2.6	Uniform	any	any	Symmetry

From Saturated
to
Supersaturated

No degree of freedom for σ
to
Negative degree of freedom for σ

Recent Applications in SSD

- (Nano-) Manufacturing
- Computer Experiments
- Numerical Analysis
- e-Business
- Marketing

Recent Advances in SSD

- Data Analysis Method:
  Panelized Least Squares
  (Li and Lin, 2003)
- Criterion:
  D-optimal Supersaturated Design
  (Nachtsheim, Jones and Lin, 2006, JSPI)
  Search Probability of Correct Identification
  (Shirakura, Takahashi and Srivastava, 1996)
- Construction:
  Combinatorial Approach
  (Fang, Ge and Liu, 2002 & 2004)

A situation for using supersaturated design:

— *A Small number of run is desired*

— *The number of potential factors is large*

— *Only a few active factors*

Supersaturated Design —how to study *k* parameters with *n(<<k)* observations?

•*What for ?*

• *How to construct ?*

• *How to analyze ?*

• *Limitations ?*

• *Does it really work ?*

Life After Screening

- Follow-Up Experiment
- Projection Properties

A New Class of Supersaturated Designs

Dennis K. J. Lin
Department of Statistics
The University of Tennessee
Knoxville, TN 37996

Supersaturated designs are useful in situations in which the number of active factors is very small compared to the total number of factors being considered. In this article, a new class of supersaturated designs is constructed using half fractions of Hadamard matrices. When a Hadamard matrix of order N is used, such a design can investigate up to $N - 2$ factors in $N/2$ runs. Results are given for $N \leq 60$. Extension to larger N is straightforward. These designs are superior to other existing supersaturated designs and are easy to construct. An example with real data is used to illustrate the ideas.

KEY WORDS: Hadamard matrices; Plackett and Burman designs; Random balance designs.

UTK Technical Report, 1991

Most Popular Title

Optimal XYZ Supersaturated Designs

***XYZ* =**

- $E(s^2)$
- Weighing
- $E(f_{NOD})$
- Bayesian
- Minimum Moment Aberration
- Large
- Three-Level
- *s*-Level
- Multi-Level
- Discrepancy
- Uniform
- Cyclic BIBD
- k-circutant

Which Journals They Appeared?

- *Biometrika*
- *JRSS-B*
- *Statistical Science*
- *Annals of Statistics*
- *Communications in Statistics*
- *Journal of Statistical Planning & Inference*
- *Statistica Sinica*
- *Stat & Prob Letters*
- *Canadian Journal of Statistics*
- *Computational Statistics & Data Analysis*
- *Technometrics*
- *Journal of Quality Technology*
- *Quality and Reliability Engieerning, International*
- *Journal of Statistical Computation & Simulations*
- *Chinese Annals of Mathematics*
- *Metrika*
- *Discrete Mathematics*
- *Sceince in China*

In Random Order

Which Journals They Appeared?

- *Analytics Chimica Acta*
- *Trac-Trends in Analytical Chemistry*
- *Journal of AOAC International*
- *Journal of Analytical Atomic Spectrometry*
- *Journal of Environmental Monitoring*
- *Chemometrics & Intelligent Lab Systems*
- *Analytical Chemistry*
- *Journal of Pharmaceutical & Biomedical Analysis*
- *Computers & Industrial Engineering*

Supersaturated Design Example

Half Fraction of William's (1968) Data

	Factor																								
Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	y
1	+	+	+	-	-	-	+	+	+	+	+	-	+	-	-	+	+	-	-	+	-	-	-	+	133
2	+	-	-	-	-	-	+	+	+	-	-	-	+	+	+	+	-	+	-	-	+	+	-	-	62
3	+	+	-	+	+	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	-	+	+	-	45
4	+	+	-	+	-	+	-	-	-	+	+	-	+	-	+	+	-	+	+	+	-	-	-	-	52
5	-	-	+	+	+	+	-	+	+	-	-	-	+	-	+	+	+	-	-	+	-	+	+	+	56
6	-	-	+	+	+	+	+	-	+	+	+	-	-	+	+	-	+	+	+	+	+	+	-	-	47
7	-	-	-	-	+	-	-	+	-	+	-	+	+	+	-	+	+	+	+	+	+	-	-	+	88
8	-	+	+	-	-	+	-	+	-	+	-	-	-	-	-	-	-	-	+	-	+	+	+	-	193
9	-	-	-	-	-	+	+	-	-	-	+	+	-	-	+	-	+	+	-	-	-	-	+	+	32
10	+	+	+	+	-	+	+	+	-	-	-	+	-	+	+	-	+	-	+	-	+	-	-	+	53
11	-	+	-	+	+	-	-	+	+	-	+	-	-	+	-	-	-	+	+	-	-	-	+	+	276
12	+	-	-	-	+	+	+	-	+	+	+	+	+	-	-	+	-	-	+	-	+	+	+	+	145
13	+	+	+	+	+	-	+	-	+	-	-	+	-	-	-	-	-	+	-	+	+	-	+	-	130
14	-	-	+	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	+	-	+	-	-	127

Lin (1993, *Technometrics*)

Half Fraction Hadamard Matrix

$(n, k) = (2t, 4t - 2)$

$\updownarrow$

Balanced Incomplete Block Design

$$\begin{aligned} \upsilon &= 2t - 1 \\ b &= 4t - 2 \\ \gamma &= 2t - 2 \\ k &= t - 2 \end{aligned}$$

Hedayat & Wallis (1978)

- $ave(s^2) = n^2 / (2n - 3)$ proved to be $E(s^2)$-optimal!
- Non-isormorphic class exist!

Half Fraction Hadamard Matrix

$(n, k) = (2t, 4t - 2)$

$\updownarrow$

Coding

Binary Code

length	$n = 2t$
weight	$\omega = t$
distance	d

$$\frac{n}{3} \le d \le \frac{2n}{3}$$

$$if\ |\gamma| \le 1/3$$

- Find $A[n, d, \omega]$: maximum number of codewords.

$$H_{n\times n} = \begin{bmatrix} 1 & 1 & H_1 \\ 1 & -1 & H_2 \end{bmatrix}_{n\times n}$$

H_1 and H_2 are isormophic?
$(n-2)\times n/2$

Lin (1991)

$$X = \left[H_m, \quad H_{Int}\right]$$

H_m : Hadamard Matrix N

H_{Int} : Interaction Columns $\binom{N}{2}$

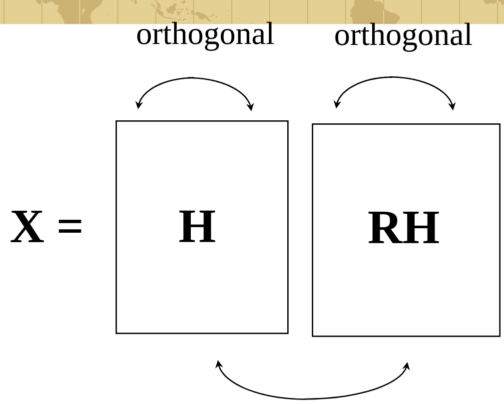
Wu (1993)

Table 34.1 Assignment Array

Class / Factor / Column — Experiment No.	Group I							Group II			
	A	B	C	D	E	F	G	H	J	L	M
	(9)	(10)	(12)	(13)	(5)	(1)	(2)	(3)	(5)	(1)	(2)
1—3	1	1	1	1	1	1	1	1	3	1	1
2—4	2	2	2	2	2	1	1	2	1	1	2
3—15	3	3	3	3	3	1	1	3	3	2	2
4—10	2	2	3	3	1	1	2	2	1	2	1
5—7	3	3	1	1	2	1	2	3	1	1	3
6—22	1	1	2	2	3	1	2	1	1	3	2
7—5	3	3	2	2	1	1	3	2	2	1	2
8—2	1	1	3	3	2	1	3	1	2	1	1
9—19	2	2	1	1	3	1	3	3	1	3	1
10—9	2	3	2	3	1	2	1	3	3	1	3
11—17	3	1	3	1	2	2	1	1	2	2	3
12—27	1	2	1	2	3	2	1	2	3	3	3
13—26	3	1	1	2	1	2	2	2	2	3	3
14—24	1	2	2	3	2	2	2	1	3	3	2
15—8	2	3	3	1	3	2	2	3	2	1	3
16—12	1	2	3	1	1	2	3	2	3	2	1
17—23	2	3	1	2	2	2	3	1	2	3	2
18—21	3	1	2	3	3	2	3	3	3	3	1
19—14	3	2	3	2	1	3	1	3	2	2	2
20—1	1	3	1	3	2	3	1	1	1	1	1
21—11	2	1	2	1	3	3	1	2	2	2	1
22—25	1	3	2	1	1	3	2	2	1	3	3
23—18	2	1	3	2	2	3	2	1	3	2	3
24—13	3	2	1	3	3	3	2	3	1	2	2
25—16	2	1	1	3	1	3	3	1	1	2	3
26—6	3	2	2	1	2	3	3	2	3	1	2
27—20	1	3	3	2	3	3	3	3	2	3	1

Random
Balance
Design

Taguchi (1986)



not orthogonal
Thus Permute rows of RH
to minimize $E(s^2)$, say.

$X = [H \quad RHC]$

↑ matrix for column selection to get rid of fully aliased columns

EXAMPLES:

(1) R = D(h_i) product — Wu (1993)

(2) R = P permute — Tang & Wu (1993)

(3) R = PD(h_i)

(4) R = $\frac{1}{n} H a H'$

↑ nonequivalent Hadamard mx

Design Criteria

Supersaturated Design

- Booth and Cox (1962): $E(s^2)$
- Wu (1993): Extension of classical optimalities (D_f, A_f etc)
- Deng and Lin (1994): 8 criteria
- Deng, Lin and Wang (1996): B-criterion
- Deng, Lin and Wang (1994): resolution rank
- Balkin and Lin (1997):
  Graphical Comparison (Harmonic mean of eigens)
- Fang, Lin and Liu (2002):

Recent Design Criteria

Supersaturated Design

- Uniformity
- Generalized Minimum Aberration
- Majorization
- E(f_{NOD})
- Projection Properties (D_f, A_f, G_f, etc)
- E(χ^2), as an extension of E(s^2)
- Minimax s_{ij}
- Model Robustness

Recent Design Criteria

Supersaturated Design

- Minimum Moment Aberration
- E(d_2)
- G_2-Aberration
- Asymptotic Power Properties
- Orthogonal-Based
- Factor-Covering
- Marginally Over-saturated

Example: $E(f_{NOD})$

$$E(f_{NOD}) = \sum_{1 \le i < j \le m} f_{NOD}^{ij} \Big/ \binom{m}{2}$$

where $$f_{NOD}^{ij} = \sum_{u=1}^{q_i} \sum_{v=1}^{q_j} \left(n_{uv}^{(ij)} - \frac{n}{q_i q_j} \right)^2$$

Lower Bound of $E(f_{NOD})$

Theorem 1. *For any design* $\boldsymbol{X} \in \mathscr{U}(n; q_1, \ldots, q_m)$,

$$\begin{aligned} E(f_{NOD}) &= \frac{\sum_{k,l=1,k\neq l}^{n} \lambda_{kl}^2}{m(m-1)} + C(n, q_1, \ldots, q_m) \\ &\ge \frac{n(\sum_{j=1}^{m} n/q_j - m)^2}{m(m-1)(n-1)} + C(n, q_1, \ldots, q_m), \end{aligned}$$

where $C(n, q_1, \ldots, q_m) = \frac{nm}{m-1} - \frac{1}{m(m-1)} \left(\sum_{i=1}^{m} \frac{n^2}{q_i} + \sum_{i,j=1,j\neq i}^{m} \frac{n^2}{q_i q_j} \right)$

Connection with Previous Criteria

Corollary 1. *For any design* $\boldsymbol{X} \in \mathcal{U}(n; q^m)$,

$$E(f_{NOD}) \geq \frac{mn}{(m-1)(n-1)}\left(\frac{n}{q}-1\right)^2 + \frac{n}{m-1}\left(m-\frac{n}{q}\right) - \left(\frac{n}{q}\right)^2,$$

$$E(f_{NOD}) = \frac{n}{9}\text{ave } \chi^2, \ \text{when } q_i = 3$$

$$E(f_{NOD}) = \frac{1}{4}E(s^2), \ \text{when } q_i = 2,$$

Data Analysis Methods

Supersaturated Design

- Pick-the-Winner
- Graphical Approach
- "PARC" (Practical Accumulation Record Computation)
- Compact Two-Sample Test
- Forward Selection
- Ridge Regression
- Normal Plot

Data Analysis Methods

Supersaturated Design

- *Satterthwaite (1959)*
- *Lin (1993): Forward Selection*
- *Westfall, Young and Lin (1998): Adjusted p-value*
- *Chen and Lin (1998): Identifiability*
- *Ryan and Lin (1997): Half Effect*
- *Contrasts-Based*
- *Staged Dimension Reduction*
- *Ye (1995): Generalized degree of freedom*

Design Analysis: Advances

Supersaturated Design

- *Sequential Analysis*
- *All Subsets Models*
- *Adjusted p-value*
  (Westfall, Young & Lin, Statistica Sinica, 1998)
- *Bayesian Approach*
  (Beattie, Fong & Lin, Technometrics, 2002)
- *Penalized Least Squares*
  (Li & Lin, 2002)

Algorithmic

Supersaturated Design

Lin (1991, 1995): Pair-wise Optimality

Nguyen (1996): Exchange Algorithm

Li and Wu (1997): Column-wise and Pair-wise Algorithm

Church (1993): Projection Properties

Jones (2000): JMP Product

Other Construction Methods

Table 1. Supersaturated designs derived from $L_{16}(4^5)$ (using **1** as the branching column)

$S(12;3^1 4^4)$	$S(8;2^1 4^4)$	Row	**1**	**2**	**3**	**4**	**5**
1		1	1	1	1	1	1
2		2	1	2	2	2	2
3		3	1	3	3	3	3
4		4	1	4	4	4	4
5	1	5	2	1	2	3	4
6	2	6	2	2	1	4	3
7	3	7	2	3	4	1	2
8	4	8	2	4	3	2	1
	5	9	3	1	3	4	2
	6	10	3	2	4	3	1
	7	11	3	3	1	2	4
	8	12	3	4	2	1	3
9		13	4	1	4	2	3
10		14	4	2	3	1	4
11		15	4	3	2	4	1
12		16	4	4	1	3	2

UD OD SSD

$$U \oplus L = \begin{bmatrix} 1 & 1 \\ 2 & 7 \\ 3 & 3 \\ 4 & 9 \\ 5 & 5 \\ 6 & 6 \\ 7 & 2 \\ 8 & 8 \\ 9 & 4 \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 2 & 0 \\ 1 & 2 & 0 & 1 \\ 2 & 0 & 2 & 1 \\ 2 & 1 & 0 & 2 \\ 2 & 2 & 1 & 0 \end{bmatrix} = X = \left[\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 2 & 0 & 2 & 1 \\ 0 & 2 & 2 & 2 & 0 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 & 2 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 & 1 & 1 & 2 & 0 \\ 1 & 2 & 0 & 1 & 1 & 2 & 0 & 1 \\ 2 & 0 & 2 & 1 & 0 & 1 & 1 & 1 \\ 2 & 1 & 0 & 2 & 2 & 1 & 0 & 2 \\ 2 & 2 & 1 & 0 & 1 & 0 & 1 & 2 \end{array}\right]$$

Fang, Lin & Ma (2000)

k-circulant Supersaturated Designs

Liu and Dean

Technometrics, 2004

Recent Advances in SSD

- Data Analysis Method:
  Panelized Least Squares
  (Li and Lin, 2003)
- Criterion:
  Search Probability of Correct Identification
  (Shirakura, Takahashi and Srivastava, 1996)
- Construction:
  Combinatorial Approach
  (Fang, Ge and Liu, 2002 & 2004)

Penalized Least Squares *(Li and Lin, 2003)*

Model

$$Y_i = \mathbf{x}_i^T \beta + \varepsilon_i$$

Penalized Likelihood (Fan and Li, 1999)

$$\frac{1}{n}\sum_{i=1}^{n} \log f(y_i, \mathbf{x}_i^T\beta) - \sum_{j=1}^{d} p_\lambda(|\beta_j|)$$

Becomes

$$Q(\beta) \equiv \frac{1}{2n}\sum_{i=1}^{n}(Y_i - \mathbf{x}_i^T\beta)^2 + \sum_{j=1}^{d} p_\lambda(|\beta_j|)$$

Choices of p_λ

$$Q(\beta) \equiv \frac{1}{2n}\sum_{i=1}^{n}(Y_i - \mathbf{x}_i^T\beta)^2 + \sum_{j=1}^{d} p_\lambda(|\beta_j|)$$

- L_1 penalty *(Donoho and Johnstone, 1994)*

$$p_\lambda(|\beta|) = \lambda|\beta|$$

- L_2 penalty & L_q penalty (Frank and Friedman, 1993)

$$p_\lambda(|\beta|) = \lambda\frac{|\beta|^2}{2}$$

- Hard Thresholding penalty

$$p_\lambda(|\beta|) = \lambda^2 - (|\beta| - \lambda)^2 I(|\beta| < \lambda)$$

Choices of p_λ

$$Q(\boldsymbol{\beta}) \equiv \frac{1}{2n}\sum_{i=1}^{n}(Y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2 + \sum_{j=1}^{d} p_\lambda(|\beta_j|)$$

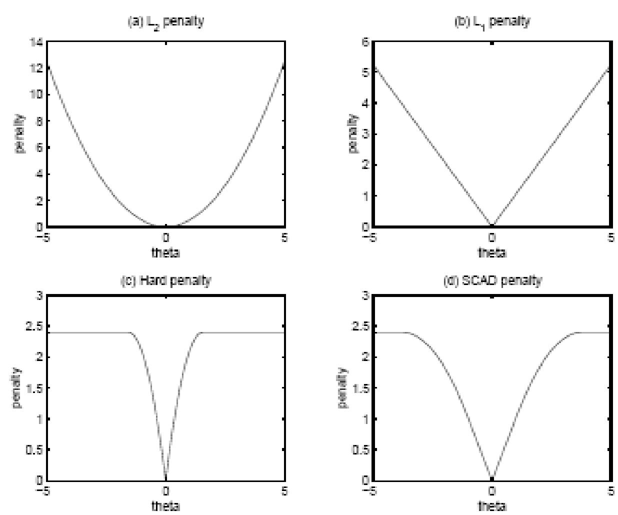
- SCAD (*Smoothly Clipped Absolute Deviation*, Fan, 1997)

$$p'_\lambda(\beta) = \lambda\left\{I(\beta \le \lambda) + \frac{(a\lambda - \beta)_+}{(a-1)\lambda} I(\beta > \lambda)\right\}$$

- λ is to be estimated (e.g., via GCV of Wahba, 1977)

$$\mathrm{GCV}(\lambda) = \frac{1}{n}\frac{\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\lambda)\|^2}{\{1 - e(\lambda)/n\}^2}$$

and $\widehat{\lambda} = \mathrm{argmin}_\lambda\{\mathrm{GCV}(\lambda)\}$



Various Examples: Real-life & Simulative

Li and Lin (2001)

Supersaturated Designs with High Searching Probability *(Chatterjee, Bhavana and Lin, 2006)*

$$\mathbf{y} = 1\mu + T_1(\zeta_0)\zeta_0 + \epsilon, \qquad V(\epsilon) = \sigma^2 I$$

Theorem

For any $\zeta_0 \subset \xi_2$ and $\zeta (\neq \zeta_0)$, we have

$$P(h(\zeta_0, \mathbf{y}) > h(\zeta, \mathbf{y})) = G_d(x, \rho)$$
$$= 1 - \Phi_d(\rho\sqrt{(n-x)/2}\,) - \Phi_d(\rho\sqrt{(n+x)/2}\,) + 2\Phi_d(\rho\sqrt{(n-x)/2}\,)\Phi_d(\rho\sqrt{(n+x)/2}\,)$$

where $x = t(\zeta_0)'t(\zeta)$, $\rho = \zeta_0/\sigma$ and Φ is the distribution function of $N(0,1)$.

Supersaturated Designs with High Searching Probability

(Chatterjee, Bhavana and Lin, 2006)

Searching Probabilities for Selected SSD

n	m	ρ = 1	1.2	1.4	1.6	1.8	2	Reference
6	10	0.9022	0.9477	0.9737	0.9875	0.9944	0.9976	L&D (2004)
8	21	0.9153	0.9537	0.9759	0.9881	0.9945	0.9977	L&D (2004)
10	15	0.9194	0.9548	0.9761	0.9882	0.9945	0.9977	B&C (2004)
12	16	0.975	0.9915	0.9974	0.9993	0.9998	0.9999	B&C (1962)
12	22	0.9207	0.9551	0.9761	0.9882	0.9945	0.9977	Lin(1993)
12	24	0.9207	0.9551	0.9761	0.9882	0.9945	0.9977	B&C (1962)
12	66	0.975	0.9915	0.9974	0.9993	0.9998	0.9999	Lin (1995)
14	19	0.9765	0.9917	0.9974	0.9993	0.9998	0.9999	B&C (2004)

Supersaturated Design via Combinatorial Construction

Fang, Ge and Liu, 2002 & 2004

- Constructing symmetrical SSDs from

  a. Resolvable balanced incomplete block designs, see Fang, Ge and Liu (2002b), Fang, Ge, Liu and Qin (2003);

  b. Room squares, see Fang, Ge and Liu (2002a);

  c. Resolvable packings and coverings, see Fang, Ge and Liu (2004) and Fang, Lu, Tang and Yin (2004);

  d. Super-simple resolvable t-designs, see Fang, Ge, Liu and Qin (2004b).

Summary: Combinatorial Construction

- The combinatorial approaches can be regarded as extensions of the methods of Nguyen (1996) and Liu and Zhang (2000) for 2-level SSDs.
- For $\lambda \geq 2$, the properties of the resulting optimal SSDs need further studies.
- How to construct optimal SSDs when $\lambda_{kl} = \lambda$ for all $k \neq l$ can not be satisfied?
- Data-analysis procedures and modelling methods for these SSDs need to be investigated.

SSD: Looking Ahead

There will be more and more factors & parameters in the future experimental investigations!!!

SSD: Looking Ahead

Supersaturated Design

- SSD is much more mature than ever
- Nano-Manufacturing Applications
- Micro-Array Design and Analysis
- Marketing & e-Business Applications
- Computer Experiment: Model Building (using SSD)
- Spotlight Interaction Effects (Lin, 1998, QE)
- Combination Designs: Rotated FFD & SSD

Thanks Sue Lewis & Dave Woods and the Organization Committee for your kind invitation

Thanks also for those helped me in the past (and perhaps the future).

Send $500 to

- Dennis Lin

  University Distinguished Professor

  483 Business Building

  Department of Supply Chain & Information Systems

  Penn State University

- +1 814 865-0377 (phone)
- +1 814 863-7076 (fax)
- DKL5@psu.edu

(Customer Satisfaction or your money back!)