General Stratum Orthogonal Arrays (GSOAs): constructions and properties

Ulrike Grömping

17 December 2022, CMStatistics London

BHT Berliner Hochschule für Technik

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Thanks

A big **Thank You** to Rob Carnell!

He provides

- Galois field arithmetics in his R package **lhs**
- and a home on his Github repository (https://github.com/bertcarnell/SOAs)

for my R package **SOAs**.

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Preliminaries

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Goal

- accommodate quantitative experimental variables, especially for computer experiments
- good space-filling properties
- good low-dimensional stratification properties

Terminology

- He and Tang (2013) proposed "**S**trong **O**rthogonal **A**rray**s**"
- Better name for these: "**S**tratum **O**rthogonal **A**rray**s**"
- "**G**" added for generalization by Tian and Xu (2022)

(G)SOAs

- starting point: an OA with s-level columns
- outcome: a (G)SOA with columns in s^ℓ levels
- method: expand the levels of the s-level OA

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Three SOAs with three 8-level columns in 16 runs ($s = 2, \ell = 3$)

OA			Shi Tang			He Tang			Liu Liu		
0	0	0	1	0	0	1	1	0	0	1	0
0	0	0	3	2	2	3	3	2	2	1	3
0	0	1	2	1	6	1	0	5	0	3	5
0	0	1	0	3	4	3	2	7	2	3	6
0	1	0	0	6	3	0	5	1	1	7	2
0	1	0	2	4	1	2	7	3	3	7	1
0	1	1	3	7	5	0	4	4	1	5	7
0	1	1	1	5	7	2	6	6	3	5	4
1	0	0	7	3	1	5	1	1	4	2	3
1	0	0	5	1	3	7	3	3	6	2	0
1	0	1	4	2	7	5	0	4	4	0	6
1	0	1	6	0	5	7	2	6	6	0	5
1	1	0	6	5	2	4	5	0	5	4	1
1	1	0	4	7	0	6	7	2	7	4	2
1	1	1	5	4	4	4	4	5	5	6	4
1	1	1	7	6	6	6	6	7	7	6	7

OA strength

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

An OA has strength t

$\Longleftrightarrow$

any subset of t columns is a (possibly replicated) full factorial.

strength + 1 = resolution

The above 2-level OA has OA strength 3,

the 8-level SOAs have OA strength 1.

Coarsening and expanding the levels of the He Tang SOA

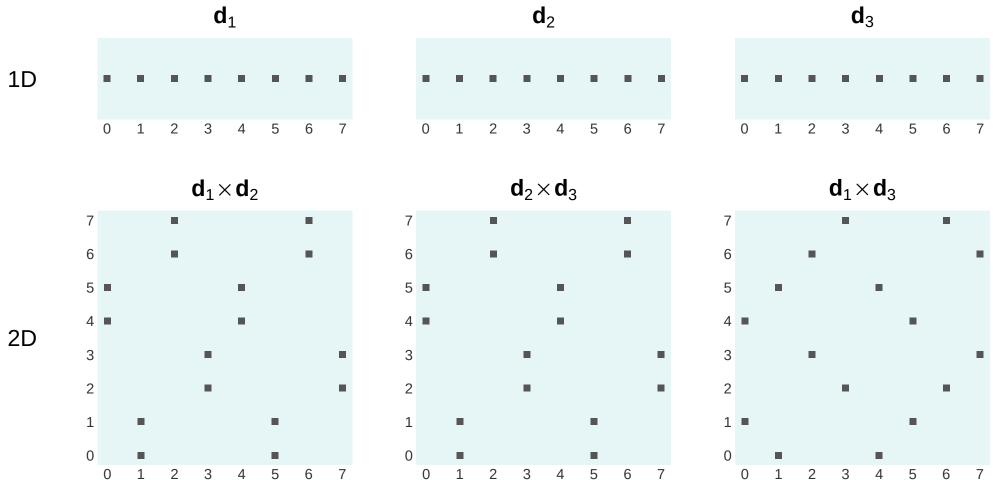
GSOA			He Tang			He Tang			He Tang		
2	3	0	1	1	0	0	0	0	0	0	0
6	6	4	3	3	2	1	1	1	0	0	0
3	1	10	1	0	5	0	0	2	0	0	1
7	5	15	3	2	7	1	1	3	0	0	1
1	11	3	0	5	1	0	2	0	0	1	0
4	14	7	2	7	3	1	3	1	0	1	0
0	9	8	0	4	4	0	2	2	0	1	1
5	12	12	2	6	6	1	3	3	0	1	1
10	2	2	5	1	1	2	0	0	1	0	0
15	7	6	7	3	3	3	1	1	1	0	0
11	0	9	5	0	4	2	0	2	1	0	1
14	4	13	7	2	6	3	1	3	1	0	1
8	10	1	4	5	0	2	2	0	1	1	0
12	15	5	6	7	2	3	3	1	1	1	0
9	8	11	4	4	5	2	2	2	1	1	1
13	13	14	6	6	7	3	3	3	1	1	1

Coarsening and expanding the levels of the He Tang SOA

GSOA			He Tang			He Tang			He Tang		
			D			$\lfloor D/2 \rfloor$			$\lfloor D/4 \rfloor$		
2	3	0	1	1	0	0	0	0	0	0	0
6	6	4	3	3	2	1	1	1	0	0	0
3	1	10	1	0	5	0	0	2	0	0	1
7	5	15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1	11	3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	14	7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0	9	8	0	4	4	0	2	2	0	1	1
5	12	12	2	6	6	1	3	3	0	1	1
10	2	2	5	1	1	2	0	0	1	0	0
15	7	6	7	3	3	3	1	1	1	0	0
11	0	9	5	0	4	2	0	2	1	0	1
14	4	13	7	2	6	3	1	3	1	0	1
8	10	1	4	5	0	2	2	0	1	1	0
12	15	5	6	7	2	3	3	1	1	1	0
9	8	11	4	4	5	2	2	2	1	1	1
13	13	14	6	6	7	3	3	3	1	1	1

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Any subset of d columns of a (G)SOA is called a d-dimensional (dD) projection.



Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

(G)SOAs: Equation-based level expansion

- start from an s-level OA $\mathbf{A}_1\left((\lambda \cdot s^\ell) \times m\right)$, expand to m columns in s^ℓ levels
- implemented via the ℓ-summand equation

$$\mathbf{D} = s^{\ell-1}\mathbf{A}_1 + s^{\ell-2}\mathbf{A}_2 + \cdots + s^0\mathbf{A}_\ell,$$

  with $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_\ell$ s-level OAs (some of which may have strength 1)
- All SOA constructions can be cast as choices for $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_\ell$ (see Grömping 2022a for a review)
- OA strength of $\mathbf{A}_1$ is upper bound for SOA strength

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Equation for $\ell \leq 3$

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

$\ell \leq 3$: avoid subscripts by using distinct letters, e.g.

$$\mathbf{D} = s^2 \cdot \mathbf{A} + s \cdot \mathbf{B} + \mathbf{C}$$

with **A**, **B** and **C** *s*-level OAs.

- OA strength of **A** is upper bound for SOA strength
- Derive a SOA construction for s^3 level columns:
    - derive suitable **A**, **B**, **C** for creating **D** with good properties
    - literature on relation of matrix properties to properties of the SOA (reviewed in Grömping 2022a)

General criteria for computer experiments

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Conventional space-filling

- important, but not a key topic of this talk
- implementation for **SOAs**:
  - use criterion ϕ_p (small=good) as a proxy for maximin distance (large=good)
  - optimize using parsimonious algorithm proposed by Weng (2014)

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Column Orthogonality

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

- column orthogonality $\iff$ pairwise correlation zero
- some SOA constructions enforce orthogonality
- Liu and Liu even enforces **3-orthogonality** (Bingham, Sitter and Tang 2009): correlation zero also with second order model matrix columns
- enforcing orthogonality may imply worsening space-filling

Strength

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

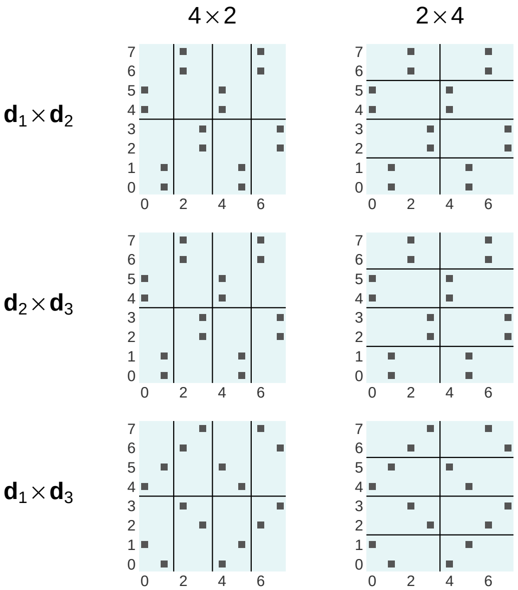
Original SOA strength

- strength $t \implies s^t$ levels
- all dD projections ($d \leq t$) stratify into s^t equally-sized strata, regardless of coarsening combination
- all three example SOAs ($s = 2, t = 3$) have strength 3:
  - 8 for 1D (trivial),
  - 4×2 and 2×4 for 2D,
  - $2 \times 2 \times 2$ for 3D (two points in each orthant)

    obvious from OA strength of matrix **A**
- qualifiers modify the narrow original definition (2^*, $2+$, $3-$, $3+$, ...)
  - The Shi and Tang example SOA has strength $3+$ (see below).

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Balance for 2^3 strata: 2D projections for He Tang SOA

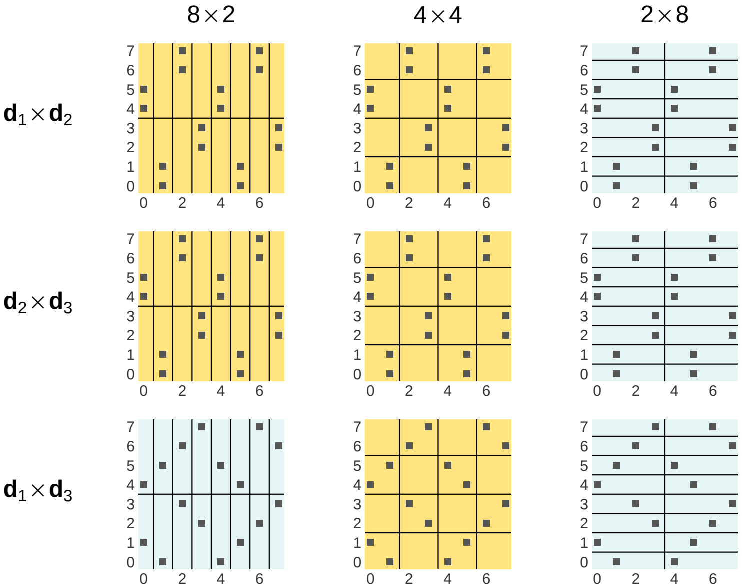


Variants of seeing s^ℓ strata in dD

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Each composition $\ell = \ell_1 + \cdots + \ell_d$ of d positive integers implies a specific set of dD 2^ℓ strata with coarsening combination $s^{\ell_1} \times \cdots \times s^{\ell_d}$.

Imbalances for 2^4 strata: 2D projections for He and Tang



(G)SOA decouple strength t and number of levels s^ℓ:

- $\ell > t$ (e.g. strength 2*) or $\ell < t$ (e.g. strength 3—) per construction
- change number of levels by expanding (or rarely also collapsing) levels of an existing s^t level SOA

(G)SOA strength t: all *possible* stratifications in dD projections ($d \leq t$) into *at most* s^t strata are balanced

- in mathematical terms:
  strength $t \iff$ entries $S_1, \ldots, S_t$ of *stratification pattern* are zero
    - Non-existence of stratification does not count as imbalanced, only imbalance in existing stratifications does.
    - The Shi Tang SOA (3 columns at 2^3 levels in 16 runs) has strength 4 now, although 1D and 4D projections with 16 strata do not exist.

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Stratification pattern

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Stratification pattern

Tian and Xu (2022) defined a “space filling pattern”, called “stratification pattern” here:

$$S_1, \ldots, S_{m \cdot \ell},$$

where m is the number of columns and s^ℓ is the number of levels.

- The j-th entry measures imbalance for stratification into s^j strata.
- $S_j = 0$: there are no unbalanced stratifications into s^j strata
- Entry S_j can be decomposed into contributions from dD projections, $d = 1, \ldots, j$.
- The S_j are closely related to the GWLP by Xu and Wu (2001):
  - they also consist of (differently grouped) sums of squared column sums of normalized orthogonal model matrix columns
  - their computation can be resource-intensive
  - see also Grömping (2022b)

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Stratification patterns for the example SOAs

	1	2	3	4	5	6	7	8	9
Shi Tang	0	0	0	0	12	3	6	6	4
He Tang	0	0	0	5	3	6	7	6	4
Liu Liu	0	0	0	5	3	6	7	6	4

- Shi and Tang has the best stratification pattern: no imbalance in stratifications into $2^4 = 16$ strata (strength 4).
- The other two SOAs have the same stratification patterns.

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

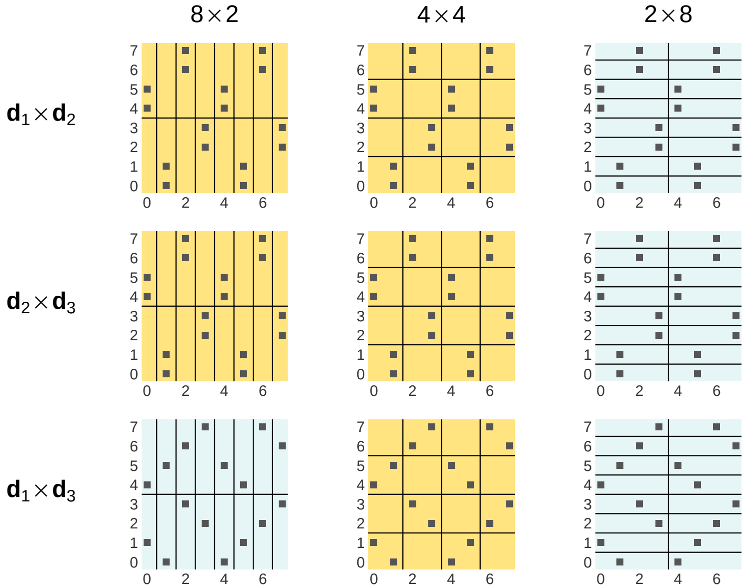
Stratification patterns decomposed by dimension

	1	2	3	4	5	6	7	8	9	GWLP
He Tang										
1D	0	0	0	**.**	.	.	.	.	.	0
2D	.	0	0	**5**	2	2	.	.	.	9
3D	.	.	0	**0**	1	4	7	6	4	22
Sum	0	0	0	**5**	3	6	7	6	4	31
Liu Liu										
1D	0	0	0	**.**	.	.	.	.	.	0
2D	.	0	0	**4**	2	3	.	.	.	9
3D	.	.	0	**1**	1	3	7	6	4	22
Sum	0	0	0	**5**	3	6	7	6	4	31

- He Tang has five 16-strata imbalances in 2D (we saw them!) and none in 3D,
- Liu Liu has one 3D 16-strata subdivision that is not balanced ($2 \times 4 \times 2$) and 4 in 2D.

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Imbalances for 2^4 strata: 2D projections for He and Tang



Constructions: Overview (see also Grömping 2022a)

- **He and Tang** (2013, 2014): strength t SOAs with s^t level columns
- **Liu and Liu** (2015): strength t SOAs with (fewer) orthogonal columns (OSOAs), 3-orthogonal for strength at least 3
- $t = 2$, Li, Liu and Yang (2021): many *orthogonal* columns with s^3 levels in strength 2 with 1D and 2D balance of strength 3 (2^*)
- $t = 2$, He, Cheng and Tang (2018) and Zhou and Tang (2019): many columns with s^2 levels and 2D balance of strength 3 ($2+$); He et al. more columns, Zhou and Tang *orthogonal*
- $s = 2, t = 3$: **Shi and Tang** (2020) obtained 8-level columns with additional strength 4 balance properties
- Tian and Xu (2022) discussed to expand the levels of (G)SOAs to the maximum possible number (to ideally obtain LHDs).

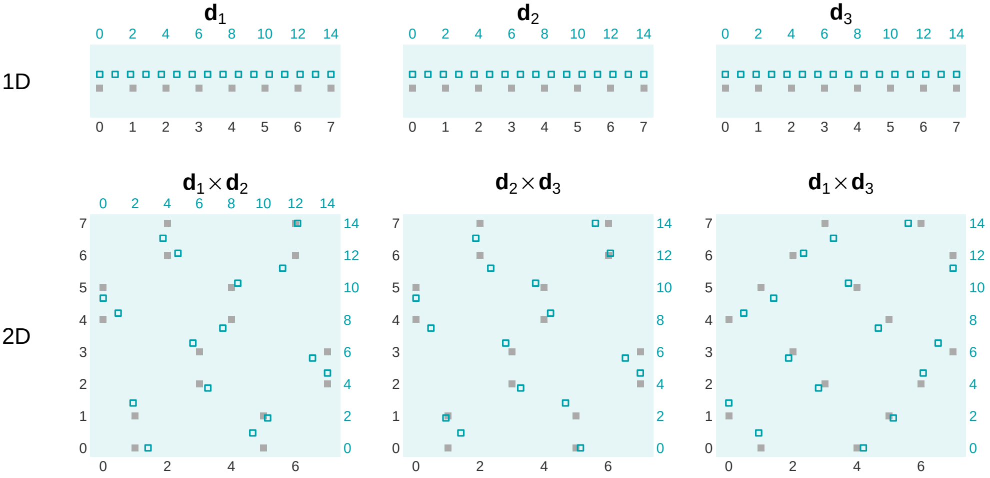
Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Coarsening and expanding the levels of the He Tang SOA

GSOA			He Tang			He Tang			He Tang		
2	3	0	1	1	0	0	0	0	0	0	0
6	6	4	3	3	2	1	1	1	0	0	0
3	1	10	1	0	5	0	0	2	0	0	1
7	5	15	3	2	7	1	1	3	0	0	1
1	11	3	0	5	1	0	2	0	0	1	0
4	14	7	2	7	3	1	3	1	0	1	0
0	9	8	0	4	4	0	2	2	0	1	1
5	12	12	2	6	6	1	3	3	0	1	1
10	2	2	5	1	1	2	0	0	1	0	0
15	7	6	7	3	3	3	1	1	1	0	0
11	0	9	5	0	4	2	0	2	1	0	1
14	4	13	7	2	6	3	1	3	1	0	1
8	10	1	4	5	0	2	2	0	1	1	0
12	15	5	6	7	2	3	3	1	1	1	0
9	8	11	4	4	5	2	2	2	1	1	1
13	13	14	6	6	7	3	3	3	1	1	1

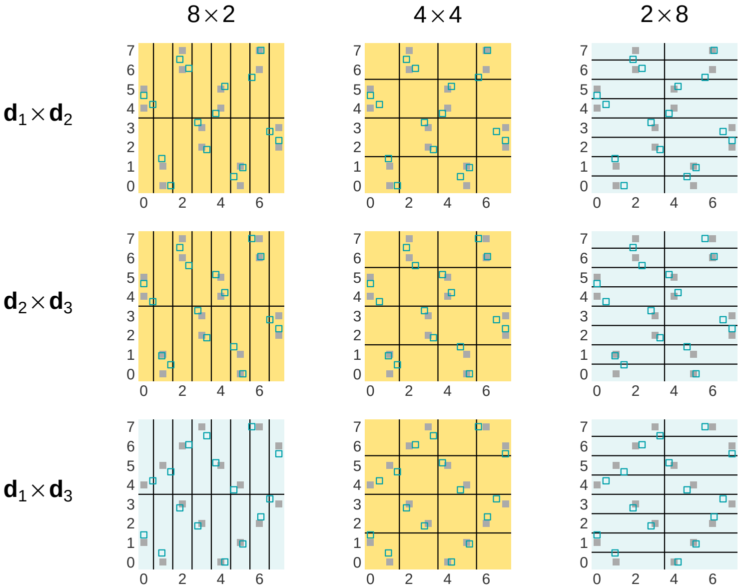
Filled Gray: SOA

Open Green: GSOA



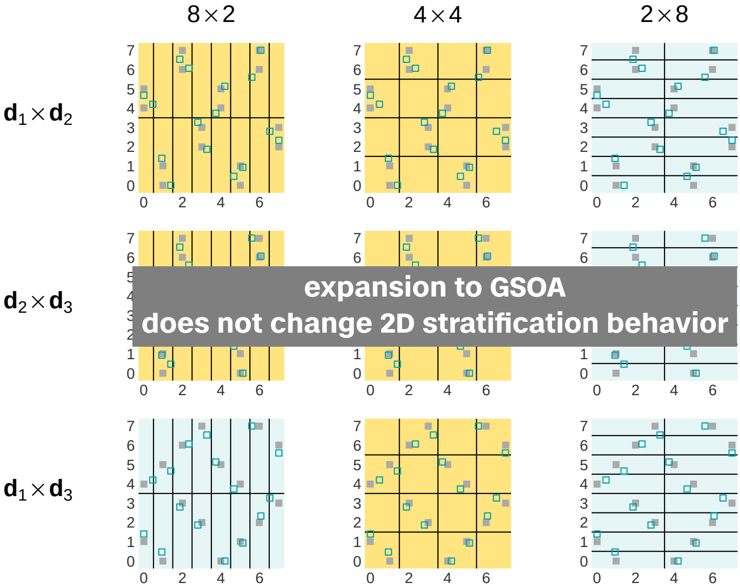
Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Imbalances for 2^4 strata: 2D projections for He and Tang GSOA



Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

Imbalances for 2^4 strata: 2D projections for He and Tang GSOA



expansion to GSOA
does not change 2D stratification behavior

- (G)SOAs are an interesting class of arrays
- they have a strong relation to OAs
- only discussed single array (G)SOAs without slicing or other modifications
- limited experience with SOA usage
  - simulations by Tian and Xu with the borehole function (Fang et al. 2006) indicate good performance of high strength SOA
- Stratification patterns and their decomposition into dimensional contributions provide insights – but are computer-intensive for larger arrays.

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

References (1)

Bingham, D., Sitter, R. and Tang, B. (2009). Orthogonal and nearly orthogonal designs for computer experiments. *Biometrika* **96**, 51-65.

Carnell, R. (2022). R package **lhs**, version 1.1.5. In: R Core Team (2022). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Fang, K.-T., Li, R. & Sudjianto, A. (2006). *Design and Modeling for Computer ExperimentsÜ. Boca Raton, Florida: Chapman and Hall/CRC.

Grömping, U. (2022a). A unifying implementation of stratum (aka strong) orthogonal arrays. Report 2/2022, Reports in Mathematics, Physics and Chemistry, Department II, BHT.

Grömping, U. (2022b). Implementation of the stratification pattern by Tian and Xu via power coding. *Report 3/2022*, Reports in Mathematics, Physics and Chemistry, Department II, BHT.

Grömping, U. (2022c). R package **SOAs**: Creation of Stratum Orthogonal Arrays, version 1.3. https://github.com/bertcarnell/SOAs. He, Y., Cheng, C.S. and Tang, B. (2018). Strong orthogonal arrays of strength two plus. *The Annals of Statistics* **46**, 457-468.

He, Y. and Tang, B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika* **100**, 254-260.

He, Y. and Tang, B. (2014). A characterization of strong orthogonal arrays of strength three. *The Annals of Statistics* **42**, 1347-1360.

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References

References (2)

Li, W., Liu, M.-Q. and Yang, J.-F. (2021). Construction of column-orthogonal strong orthogonal arrays. *Statistical Papers* https://doi.org/10.1007/s00362-021-01249-w.

Liu, H. and Liu, M.-Q. (2015). Column-orthogonal strong orthogonal arrays and sliced strong orthogonal arrays. *Statistica Sinica* **25**, 1713-1734.

Shi, L. and Tang, B. (2020). Construction results for strong orthogonal arrays of strength three. *Bernoulli* **26**, 418-431. https://doi.org/10.3150/19-BEJ1130

Tang, B. (1993). Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association* **88**, 1392-1397.

Tian, Y. and Xu, H. (2022). A minimum aberration-type criterion for selecting space-filling designs. *Biometrika* **109**, 489-501. https://doi.org/10.1093/biomet/asab021

Weng, J. (2014). Maximin Strong Orthogonal Arrays. *Master's thesis* at Simon Fraser University under supervision of Boxin Tang and Jiguo Cao.

Xu, H. and Wu, C.F.J. (2001). Generalized minimum aberration for factorial designs. *The Annals of Statistics* **29**, 1066-1077.

Ye, K.-Q. (1998). Orthogonal column Latin hypercubes and their application in computer experiments. *Journal of the American Statistical Association* **93**, 1430-1439.

Zhou, Y.D. and Tang, B. (2019). Column-orthogonal strong orthogonal arrays of strength two plus and three minus. *Biometrika* **106**, 997-1004. https://doi.org/10.1093/biomet/asz043

Introduction
Thanks
Preliminaries
Constructions, Part 1
Properties
General criteria for computer experiments
Strength
Stratification pattern
Constructions, Part 2
Final Remarks
References