D-optimal Split-split-plot Designs

Bradley Jones
SAS Institute Inc.
Peter Goos
University of Antwerp

Outline

- Motivation
- Model
- Algorithm
- Advice concerning minimizing the number of whole plots
- Counterintuitive Example
- Effect of Changing Variance Ratios
- Cheese Processing Example

Motivation – Why would you want to compute optimal SSPDs?

- Experiments on multi-step processes
- Processes having factors with varying degrees of difficulty to change

Example Application

Cheese Production has 3 stages

1. Store milk in large tanks
2. Divide milk from tanks among curds processors and make curds.
3. Further process the curds to make individual cheeses.

Outline

- Motivation
- Model
- Algorithm
- Advice concerning minimizing the number of whole plots
- Counterintuitive Example
- Effect of Changing Variance Ratios
- Cheese Processing Example

Model

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{Z}_2\boldsymbol{\gamma}_2 + \boldsymbol{\varepsilon}$$

$$\mathbf{Z}_1 = \mathbf{I}_{b_1} \otimes \mathbf{1}_{b_2 k}$$

$$\mathbf{Z}_2 = \mathbf{I}_{b_1} \otimes \mathbf{I}_{b_2} \otimes \mathbf{1}_k = \mathbf{I}_{b_1 b_2} \otimes \mathbf{1}_k$$

$$\begin{aligned} \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}_n \text{ and } \mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_n, \\ \mathrm{E}(\boldsymbol{\gamma}_1) = \mathbf{0}_{b_1} \text{ and } \mathrm{cov}(\boldsymbol{\gamma}_1) = \sigma_{\gamma_1}^2 \mathbf{I}_{b_1}, \\ \mathrm{E}(\boldsymbol{\gamma}_2) = \mathbf{0}_{b_1 b_2} \text{ and } \mathrm{cov}(\boldsymbol{\gamma}_2) = \sigma_{\gamma_2}^2 \mathbf{I}_{b_1 b_2}, \end{aligned}$$

Where X is the design matrix for the fixed effects, Z_1 is an indicator matrix for the whole plots, Z_2 is an indicator matrix for the subplots, β is a vector of fixed effects, γ_1 is a vector of the whole plot random effects, γ_2 is a vector of the subplot random effects, ε is the vector random errors, b_1 is the number of whole plots, b_2 is the number of subplots per whole plot and k is the number of runs per subplot.

Variance of Y

$$\mathbf{V} = \sigma_{\varepsilon}^{2}\mathbf{I}_{n} + \sigma_{\gamma_1}^{2}\mathbf{Z}_1\mathbf{Z}_1' + \sigma_{\gamma_2}^{2}\mathbf{Z}_2\mathbf{Z}_2'$$

Information Matrix

$$\mathbf{M} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$$

the D-optimal design maximizes the determinant of M

Determinant depends on unknown variance ratios.

$$\eta_1 = \sigma_{\gamma_1} / \sigma_{\varepsilon}$$

$$\eta_2 = \sigma_{\gamma_2} / \sigma_{\varepsilon}$$

You don't have to calculate the inverse of V!

Theorem 1 *The inverse of the covariance matrix* $\mathbf{V}$ *is equal to*

$$\mathbf{V}^{-1} = \sigma_{\varepsilon}^{-2}\mathbf{I}_n - c_1\mathbf{Z}_1\mathbf{Z}_1' - c_2\mathbf{Z}_2\mathbf{Z}_2',$$

where

$$c_1 = \sigma_{\varepsilon}^{-2}\frac{\eta_1 - \frac{\eta_1\eta_2 k}{1+\eta_2 k}}{1 + \eta_1 b_2 k + \eta_2 k}$$

and

$$c_2 = \sigma_{\varepsilon}^{-2}\frac{\eta_2}{1 + \eta_2 k}.$$

Algorithm

Inspired by coordinate exchange

Meyer & Nachtsheim *Technometrics* 1995

Starting Design

WP	SP	X1	X2	X3
1	1	0.25	0.37	-0.66
1	1	0.25	0.37	0.05
1	2	0.25	-0.69	-0.87
1	2	0.25	-0.69	-0.72
2	3	0.57	0.44	-0.59
2	3	0.57	0.44	0.49
2	4	0.57	-0.87	-0.74
2	4	0.57	-0.87	-0.74

Determinant = 0.026

After First Row

WP	SP	X1	X2	X3
1	1	-1.00	1.00	-1.00
1	1	-1.00	1.00	0.05
1	2	-1.00	-0.69	-0.87
1	2	-1.00	-0.69	-0.72
2	3	0.57	0.44	-0.59
2	3	0.57	0.44	0.49
2	4	0.57	-0.87	-0.74
2	4	0.57	-0.87	-0.74

Determinant = 1.456

After 2nd Row

WP	SP	X1	X2	X3
1	1	-1.00	1.00	-1.00
1	1	-1.00	1.00	1.00
1	2	-1.00	-0.69	-0.87
1	2	-1.00	-0.69	-0.72
2	3	0.57	0.44	-0.59
2	3	0.57	0.44	0.49
2	4	0.57	-0.87	-0.74
2	4	0.57	-0.87	-0.74

Determinant = 3.182

After 3rd Row

WP	SP	X1	X2	X3
1	1	-1.00	1.00	-1.00
1	1	-1.00	1.00	1.00
1	2	-1.00	-1.00	1.00
1	2	-1.00	-1.00	-0.72
2	3	0.57	0.44	-0.59
2	3	0.57	0.44	0.49
2	4	0.57	-0.87	-0.74
2	4	0.57	-0.87	-0.74

Determinant = 6.46

After 4th Row

WP	SP	X1	X2	X3
1	1	-1.00	1.00	-1.00
1	1	-1.00	1.00	1.00
1	2	-1.00	-1.00	1.00
1	2	-1.00	-1.00	-1.00
2	3	0.57	0.44	-0.59
2	3	0.57	0.44	0.49
2	4	0.57	-0.87	-0.74
2	4	0.57	-0.87	-0.74

Determinant = 7.20

After 5th Row

WP	SP	X1	X2	X3
1	1	-1.00	1.00	-1.00
1	1	-1.00	1.00	1.00
1	2	-1.00	-1.00	1.00
1	2	-1.00	-1.00	-1.00
2	3	1.00	1.00	-1.00
2	3	1.00	1.00	0.49
2	4	1.00	-0.87	-0.74
2	4	1.00	-0.87	-0.74

Determinant = 16.777

After 6th Row

WP	SP	X1	X2	X3
1	1	-1.00	1.00	-1.00
1	1	-1.00	1.00	1.00
1	2	-1.00	-1.00	1.00
1	2	-1.00	-1.00	-1.00
2	3	1.00	1.00	-1.00
2	3	1.00	1.00	1.00
2	4	1.00	-0.87	-0.74
2	4	1.00	-0.87	-0.74

Determinant = 19.86

After 7th Row

WP	SP	X1	X2	X3
1	1	-1	1	-1
1	1	-1	1	1
1	2	-1	-1	1
1	2	-1	-1	-1
2	3	1	1	-1
2	3	1	1	1
2	4	1	-1	1
2	4	1	-1	-0.74

Determinant = 26.19

Optimal Design

WP	SP	X1	X2	X3
1	1	-1	1	-1
1	1	-1	1	1
1	2	-1	-1	1
1	2	-1	-1	-1
2	3	1	1	-1
2	3	1	1	1
2	4	1	-1	1
2	4	1	-1	-1

Determinant = 27.86

Graphical Kinetic View

Bubble Plot Demonstration

Outline

- Motivation
- Model
- Algorithm
- Advice concerning minimizing the number of whole plots
- Counterintuitive Example
- Effect of Changing Variance Ratios
- Cheese Processing Example

What if you can only do 2 whole plots?

i.e. the whole plot factor is *really* hard to change

Recommendation

Make sure that you include two-factor interactions involving the whole plot factor in the model

Example

- One whole plot factor – 2 whole plots
- One subplot factor – 4 subplots
- Three sub-subplot factors – 24 runs

Optimal Design

Whole plot	Subplot	w	s	t_1	t_2	t_3	Whole plot	Subplot	w	s	t_1	t_2	t_3
1	1	-1	1	-1	-1	-1	2	3	1	-1	-1	-1	1
1	1	-1	1	-1	-1	1	2	3	1	-1	-1	1	-1
1	1	-1	1	-1	1	-1	2	3	1	-1	1	1	1
1	1	-1	1	-1	1	1	2	3	1	-1	1	-1	-1
1	1	-1	1	1	-1	-1	2	3	1	-1	-1	1	1
1	1	-1	1	1	1	1	2	3	1	-1	1	1	-1
1	2	-1	-1	1	-1	1	2	4	1	1	1	-1	1
1	2	-1	-1	-1	1	-1	2	4	1	1	1	1	-1
1	2	-1	-1	-1	-1	-1	2	4	1	1	1	1	1
1	2	-1	-1	-1	1	1	2	4	1	1	-1	1	1
1	2	-1	-1	1	1	-1	2	4	1	1	-1	-1	-1
1	2	-1	-1	1	1	1	2	4	1	1	-1	-1	1

Coefficient Variances

Stratum	Effect	Variance
WP	Intercept	0.796875
WP	w	0.796875
SP	s	0.296875
SP	ws	0.296875
SSP	t_1	0.046875
SSP	t_2	0.046875
SSP	t_3	0.046875
SSP	wt_1	0.046875
SSP	wt_2	0.046875
SSP	wt_3	0.046875
SSP	st_1	0.046875
SSP	st_2	0.046875
SSP	st_3	0.046875
SSP	t_1t_2	0.046875
SSP	t_1t_3	0.046875
SSP	t_2t_3	0.046875

Outline

- Motivation
- Model
- Algorithm
- Advice concerning minimizing the number of whole plots
- Counterintuitive Example
- Effect of Changing Variance Ratios
- Cheese Processing Example

Diagonal information matrix = Optimal Design?

For 2-level completely randomized designs orthogonality equates to optimality.

e.g. all 2-level orthogonal designs are also globally optimal.

But, this may not be true for split-split-plot designs!

Example:

- Two whole plot factors with eight whole plots
- One subplot factor with 16 subplots
- Three sub-subplot factors with 32 runs.

Design

Whole plot	Subplot	w_1	w_2	s	t_1	t_2	t_3	Whole plot	Subplot	w_1	w_2	s	t_1	t_2	t_3
1	1	1	1	1	-1	-1	1	5	9	-1	-1	1	1	1	-1
1	1	1	1	1	1	1	-1	5	9	-1	-1	1	-1	-1	-1
1	2	1	1	-1	1	-1	-1	5	10	-1	-1	-1	1	-1	1
1	2	1	1	-1	-1	1	1	5	10	-1	-1	-1	-1	1	1
2	3	-1	1	-1	-1	-1	1	6	11	1	-1	-1	1	1	-1
2	3	-1	1	-1	1	1	-1	6	11	1	-1	-1	-1	-1	1
2	4	-1	1	1	1	-1	-1	6	12	1	-1	1	1	-1	1
2	4	-1	1	1	-1	1	1	6	12	1	-1	1	-1	1	-1
3	5	1	-1	-1	-1	-1	-1	7	13	-1	1	1	1	-1	1
3	5	1	-1	-1	1	1	1	7	13	-1	1	1	-1	1	-1
3	6	1	-1	1	1	-1	-1	7	14	-1	1	-1	-1	-1	-1
3	6	1	-1	1	-1	1	1	7	14	-1	1	-1	1	1	1
4	7	-1	-1	-1	-1	1	-1	8	15	1	1	-1	-1	1	-1
4	7	-1	-1	-1	1	-1	-1	8	15	1	1	-1	1	-1	1
4	8	-1	-1	1	1	1	1	8	16	1	1	1	1	1	1
4	8	-1	-1	1	-1	-1	1	8	16	1	1	1	-1	-1	-1

Features of Design

The information matrix is not diagonal.

There are three off-diagonal elemennts.

Fractional Factorial Alternatives

- There are very many designs with diagonal information matrices.
- Construction method of the best we could find.
  1. $t_2 = w_1 w_2 s t_1$
  2. Use contrast columns w_1, w_2 and $w_2 t_1 t_3$ to partition the 8 whole plots.

Coefficient Variances

Stratum	Effect	D-optimal	Alternative
WP	Intercept	0.21875	0.21875
WP	w_1	0.21875	0.21875
WP	w_2	0.21875	0.21875
WP	w_1w_2	0.21875	0.21875
SP	s	0.09375	0.09375
SP	w_1s	0.09375	0.09375
SP	w_2s	0.09375	0.09375
SSP	t_1	0.03125	0.03125
SSP	t_2	0.03125	0.03125
SSP	t_3	0.04167	0.03125
SSP	w_1t_1	0.03125	0.03125
SSP	w_1t_2	0.03125	0.03125
SSP	w_1t_3	0.04167	0.03125
SSP	w_2t_1	0.03125	0.03125
SSP	w_2t_2	0.03125	0.03125
SSP	w_2t_3	0.04167	0.03125
SSP	st_1	0.03125	0.03125
SSP	st_2	0.03125	0.03125
SSP	st_3	0.03977	0.03125
SSP	t_1t_2	0.09375	0.09375
SSP	t_1t_3	0.07721	0.21875
SSP	t_2t_3	0.06908	0.09375

Outline

- Motivation
- Model
- Algorithm
- Advice concerning minimizing the number of whole plots
- Counterintuitive Example
- Effect of Changing Variance Ratios
- Cheese Processing Example

Determinant depends on variance ratios

How much difference does this make?

Example

- One whole plot factor – six whole plots
- One subplot factor – 12 subplots
- Three easy-to-change factors – 24 runs
- Model with main effects and all two-factor interactions
- Consider all combinations of $log10(\eta_1)$ and $log10(\eta_2)$ each with three levels -1, 0 and 1.

There were six different designs, but…

	η_1		
η_2	0.1	1.0	10
0.1	98.5%	99.3%	100%
1.0	98.7%	100%	100%
10	95.5%	95.6%	95.7%

Assuming that the true variance ratios were both 1, here are the relative efficiencies of the designs. There is little practical difference.

Outline

- Motivation
- Model
- Algorithm
- Advice concerning minimizing the number of whole plots
- Counterintuitive Example
- Effect of Changing Variance Ratios
- Cheese Processing Example

Cheese Processing Experiment

- Two milk storage factors (8 whole plots)
- Five curds production factors (32 subplots)
- Three cheese making factors – one at 4 levels with 128 total runs

Study Design - Resolution IV

But, using our algorithm, we found a design that could estimate all the two-factor interactions and was orthogonal for the main effects.

We also found a design with only one quarter of the runs that was orthogonal for all the main effects.

Design – 32 runs

Whole plot	Subplot	w_1	w_2	s_1	s_2	s_3	s_4	s_5	t_1	t_2	t_3
1	1	1	-1	-1	-1	-1	1	1	1	1	D
1	1	1	-1	-1	-1	-1	1	1	-1	-1	B
1	2	1	-1	1	1	1	-1	-1	1	-1	A
1	2	1	-1	1	1	1	-1	-1	-1	1	C
2	3	1	1	1	1	-1	1	-1	1	1	B
2	3	1	1	1	1	-1	1	-1	-1	-1	D
2	4	1	1	-1	-1	1	-1	1	-1	-1	A
2	4	1	1	-1	-1	1	-1	1	1	1	C
3	5	-1	1	1	1	1	1	1	1	-1	C
3	5	-1	1	1	1	1	1	1	-1	1	D
3	6	-1	1	-1	-1	-1	-1	-1	-1	1	A
3	6	-1	1	-1	-1	-1	-1	-1	1	-1	B
4	7	-1	-1	-1	1	-1	1	-1	-1	-1	C
4	7	-1	-1	-1	1	-1	1	-1	1	1	A
4	8	-1	-1	1	-1	1	-1	1	-1	1	D
4	8	-1	-1	1	-1	1	-1	1	1	-1	B
5	9	1	-1	-1	1	-1	-1	1	-1	1	C
5	9	1	-1	-1	1	-1	-1	1	1	-1	D
5	10	1	-1	1	-1	1	1	-1	1	1	A
5	10	1	-1	1	-1	1	1	-1	-1	-1	B
6	11	-1	-1	-1	1	1	-1	-1	1	-1	D
6	11	-1	-1	-1	1	1	-1	-1	-1	1	B
6	12	-1	-1	1	-1	-1	1	1	-1	1	A
6	12	-1	-1	1	-1	-1	1	1	1	-1	C
7	13	-1	1	1	1	-1	-1	1	-1	-1	A
7	13	-1	1	1	1	-1	-1	1	1	1	B
7	14	-1	1	-1	-1	1	1	-1	1	1	C
7	14	-1	1	-1	-1	1	1	-1	-1	-1	D
8	15	1	1	1	-1	-1	-1	-1	-1	-1	C
8	15	1	1	1	-1	-1	-1	-1	1	1	D
8	16	1	1	-1	1	1	1	1	-1	1	B
8	16	1	1	-1	1	1	1	1	1	-1	A

Coefficient Variances

Stratum	Effect	Variance
WP	Intercept	7/32
WP	w_1	7/32
WP	w_2	7/32
SP	s_1	3/32
SP	s_2	3/32
SP	s_3	3/32
SP	s_4	3/32
SP	s_5	3/32
SSP	t_1	1/32
SSP	t_2	1/32
SSP	$t_3[1]$	3/64
SSP	$t_3[2]$	3/64
SSP	$t_3[3]$	1/32

Summary

1. We have supplied an algorithmic approach for computing SSPDs.
2. The approach is useful for either screening or RSM.
3. We discussed the problem of confounding of whole plot fixed effects and variances and proposed a practical way of proceeding.
4. We introduced a case where diagonal information matrices appear not to be optimal.
5. We considered the effect of unknown variance ratios on the design – more work to do here.
6. We applied our method to a previously run experiment with useful results.

Contact Information

Bradley.Jones@jmp.com

Statistical Discovery.™ From SAS.