Multi-Objective Optimal Design of Experiments

Steven Gilmour

The Open University Statistics Seminar 2021

Outline

1. Response surface designs
2. Optimal designs, classical designs or MOODs?
3. Response surface designs for predicting differences in response
4. Current and future work

Response surface experiments

Common features:

- one (or more) response of interest, Y;
- several (continuous) input variables, $X_1, \ldots, X_q$, changing whose levels might affect Y;
- the opportunity to do a restricted number of experimental runs n, to study the relationships.

Response surface experiments

Common features:

- one (or more) response of interest, Y;
- several (continuous) input variables, $X_1, \ldots, X_q$, changing whose levels might affect Y;
- the opportunity to do a restricted number of experimental runs n, to study the relationships.

Note: code $-1 \leq x_{ri} \leq 1$, $r = 1, \ldots, q$, $i = 1, \ldots, n$.

The Ultra-Clean Instrument Project

Steel instruments used in (human) surgery are reused many times on different patients.

They are washed in automatic washing machines, similar to domestic dishwashers.

To prevent cross-contamination, it is essential that all protein residues are cleaned completely, especially given concerns about vCJD (the “human form of mad cow disease”).

How can ultra-clean instruments be achieved?

Guidelines exist on how to carry out cleaning; these must be adhered to.

However, it is not clear that these guidelines are sufficient, or necessary.

Experiments were to be performed to determine which aspects could be improved and which aspects could be relaxed.

Study Objectives

To define a baseline performance for utilisation of a washer disinfector for decontamination of reusable surgical instruments.

Seeking at least a routinely achievable 2 orders of magnitude reduction in protein residues over current practice; to inform future standards.

Simplicity before complexity at a practical threshold is essential.

The outcome should not compromise current safety standards of toxicity, general cleaning performance and ability to sterilize subsequently.

The Washer

The Washer

Factor levels can be programmed.

The Washer

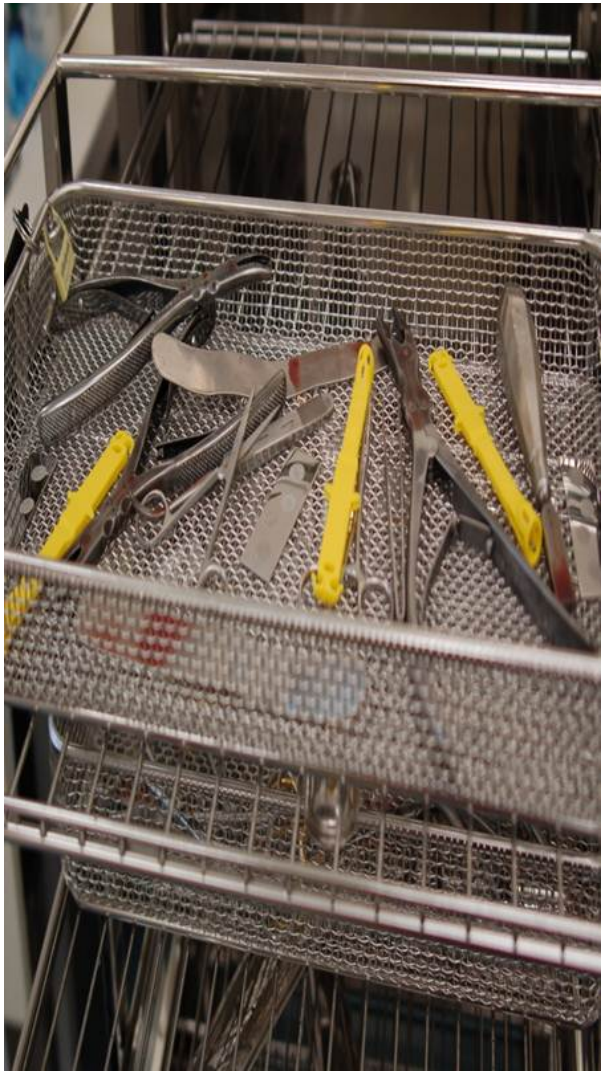
The Washer

Test Wash Loads

Instruments smeared with protein (from pig's brains).

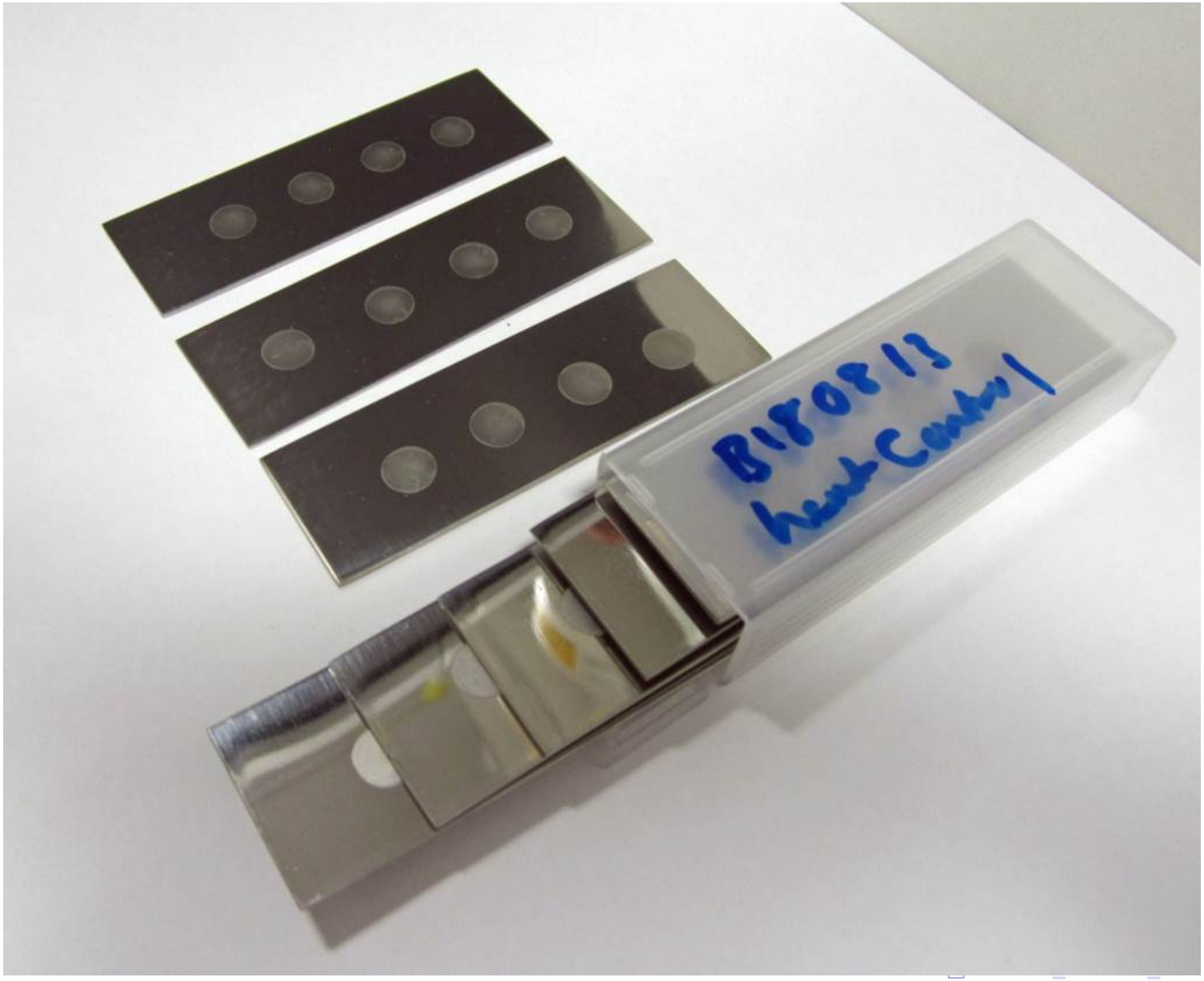
Test Wash Loads

Test tags included to measure wash performance.



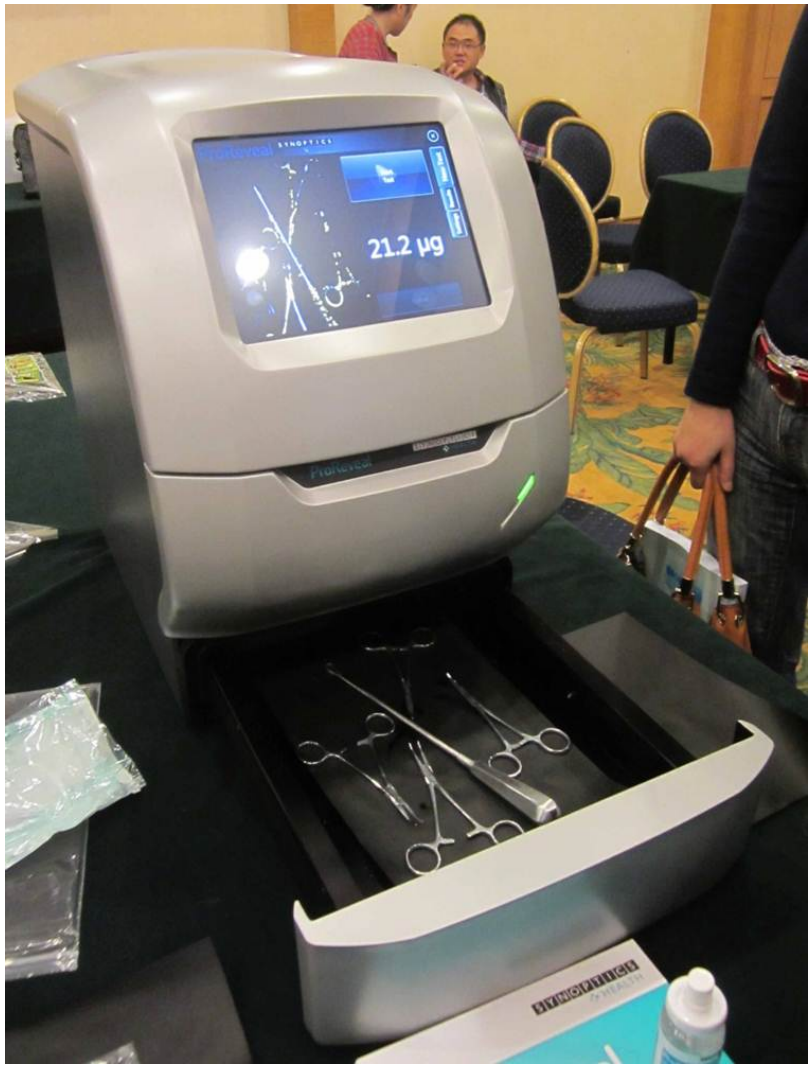
Test Tags

Test Tags



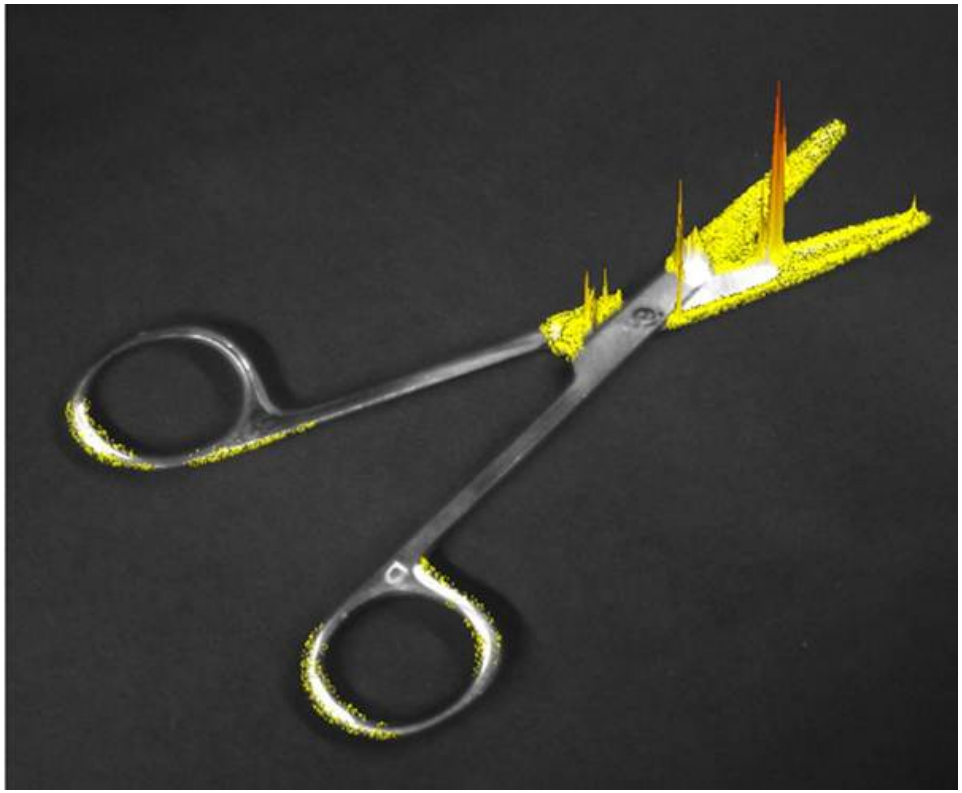
Response Measurements

Residual protein measured.



Response Measurements

Software illustrates spatial layout of protein, as well as total amount.



First Experiment

Aim to understand which wash settings lead to better performance and which settings do not matter.

Treatment factors:
Chemistry: 0 = water; 1 = alkili; 2 = enzymatic; 3 = non-ionic
Second wash: 0 = no; 1 = yes
Initial rinse time; Initial rinse temperature; Main wash dose; main wash time; main wash temperature; Second wash dose; Second wash time; Second wash temperature; Intermediate rinse time; Intermediate rinse temperature; Final rinse time; Final rinse temperature: all quantitative.

Polynomial response surface models

In response surface methodology (RSM), low degree polynomial regression models are very widely used as empirical approximations:

$$Y_i = \beta_0 + \sum_{r=1}^{q} \beta_r x_{ri} + \epsilon_i,$$

or

$$Y_i = \beta_0 + \sum_{r=1}^{q} \beta_r x_{ri} + \sum_{i=1}^{q} \beta_{rr} x_{ri}^2 + \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} \beta_{rs} x_{ri} x_{si} + \epsilon_i,$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ (normality not crucial).

Polynomial response surface models

In response surface methodology (RSM), low degree polynomial regression models are very widely used as empirical approximations:

$$Y_i = \beta_0 + \sum_{r=1}^{q} \beta_r x_{ri} + \epsilon_i,$$

or

$$Y_i = \beta_0 + \sum_{r=1}^{q} \beta_r x_{ri} + \sum_{i=1}^{q} \beta_{rr} x_{ri}^2 + \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} \beta_{rs} x_{ri} x_{si} + \epsilon_i,$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ (normality not crucial).

All the usual regression techniques are useful.

Response surface analysis

However, these are experiments, so not just regression.

Response surface analysis

However, these are experiments, so not just regression.

- Correlation is causation.
- Simple additivity assumption of treatment and unit effects justifies model for variance structure.
- We can have replicates of treatment combinations.
- Nuisance variables can be designed out using an appropriate blocking structure.

Response surface analysis

However, these are experiments, so not just regression.

- Correlation is causation.
- Simple additivity assumption of treatment and unit effects justifies model for variance structure.
- We can have replicates of treatment combinations.
- Nuisance variables can be designed out using an appropriate blocking structure.

All of this depends on doing a valid randomisation

Response surface analysis

However, these are experiments, so not just regression.

- Correlation is causation.
- Simple additivity assumption of treatment and unit effects justifies model for variance structure.
- We can have replicates of treatment combinations.
- Nuisance variables can be designed out using an appropriate blocking structure.

All of this depends on doing a valid randomisation (the only thing which raises us above the level of the beasts).

Response surface designs

Historically, "classical" designs with symmetric structures and some replication have been widely used.

- Two-level factorials (Fisher, 1925) and regular fractions (Finney, 1945), plus centre points.
- Central composite designs (Box and Wilson, 1951): add axial points at $(\pm\alpha, 0, \ldots, 0)$, etc. to the above.
- Box-Behnken designs (BBDs) (Box and Behnken, 1960): three-level designs with all points an equal distance from the centre.
- Subset designs (Gilmour, 2006): combinations of points at different distances.
- Fractional BBDs (Edwards and Mee, 2011; Lin *et al.*, 2012).

Response surface designs

Historically, "classical" designs with symmetric structures and some replication have been widely used.

- Two-level factorials (Fisher, 1925) and regular fractions (Finney, 1945), plus centre points.
- Central composite designs (Box and Wilson, 1951): add axial points at $(\pm\alpha, 0, \ldots, 0)$, etc. to the above.
- Box-Behnken designs (BBDs) (Box and Behnken, 1960): three-level designs with all points an equal distance from the centre.
- Subset designs (Gilmour, 2006): combinations of points at different distances.
- Fractional BBDs (Edwards and Mee, 2011; Lin *et al.*, 2012).

Designs are not chosen explicitly to have good statistical properties.

Optimal design criteria

Write polynomial model as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

Then $Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

In practice, designs are increasingly chosen, using widely available software, to optimise:

- $1/|\mathbf{X}'\mathbf{X}|$, D-optimality, to (asymptotically) minimise volume of joint confidence region on parameters;
- $tr\{(\mathbf{X}'\mathbf{X})^{-1}\}$, A-optimality, to minimise average variance of parameters;
- $tr\{\mathbf{W}(\mathbf{X}'\mathbf{X})^{-1}\}$, L-optimality, for linear functions of parameters;
- $\max_{\mathbf{x}\in[-1,1]^q} \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}$, G-optimality, for prediction;
- $\int_{\mathbf{x}\in[-1,1]^q} \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}d\mathbf{x}$, I- (or V-)optimality, for prediction.

Optimal versus Classical?

Leaving aside computational restrictions, I see no contradiction.

Classical criteria help to decide the appropriate class of designs over which to optimize, the optimality criterion and how to implement the design.

Optimality criteria help to find a design which obeys the classical principles.

Any conflict is due to computational limitations (or poor choices).

It's the Design that Matters

The 2nd order response surface design in 4 three-level factors $S_3 + 4S_0$ was recommended by:

- Gilmour(2006) as D- and A-optimal within the class of subset designs
- Edmondson (1994) by aliasing canonical contrasts
- Todd (1988) who dreamt it up

For an experimenter, the method of construction is irrelevant.
Linear models analysis is conditional on $\mathbf{X}$, not the method of choosing $\mathbf{X}$.

A Unified Approach

How to choose a design:

- Objectives lead to choice of treatment structure
- Practical restrictions lead to definition of class of designs and randomization scheme
- When we know what we want, we choose a method of design construction which will best meet our needs

Design for the experiment, don't experiment for the design

A Unified Approach

How to choose a design:

- Objectives lead to choice of treatment structure
- Practical restrictions lead to definition of class of designs and randomization scheme
- When we know what we want, we choose a method of design construction which will best meet our needs

Design for the experiment, don't experiment for the design and choose a program for the design, don't choose a design for the program.

How do we analyse data from multi-factor experiments?

Many experimenters still prefer to use standard designs (regular fractional factorial, central composite, Box-Behnken, subset), which include replicate points, especially centre points.

Analysis:

- Fit one or more polynomial models.
- Use F tests to compare models of different orders, test for lack of fit and test whether the model is better than a null model.
- Use t tests to test individual parameters.
- Estimate the parameters.

All tests require an estimate of σ^2.

A simple design

A central composite design (CCD) for $q = 2$:

X_1	X_2
-1	-1
-1	1
1	-1
1	1
-1	0
1	0
0	-1
0	1
0	0
0	0
0	0
0	0

Analysis of variance

Which is "correct"?

Source	df	MS
Regression	5	
Residual:	6	S_p^2
Lack of fit	3	
Pure error	3	S^2
Total	11	

Source	df	MS
Treatments:	8	
Regression	5	
Lack of fit	3	
Pure error	3	S^2
Total	11	

Analysis of variance

Which is "correct"?

Source	df	MS
Regression	5	
Residual:	6	S_p^2
Lack of fit	3	
Pure error	3	S^2
Total	11	

Source	df	MS
Treatments:	8	
Regression	5	
Lack of fit	3	
Pure error	3	S^2
Total	11	

The right hand table emphasises the design structure and clearly identifies the unbiased estimator of σ^2.

Estimating Error

How should we estimate σ^2?

- Regression viewpoint: MS_E from the model fitted.
- Anova viewpoint: "pure error" MS from fitting full treatment model.

The latter is unbiased, justified by a valid randomization, objective and gives unambiguous inferences whose properties are understood.

Estimating Error

How should we estimate σ^2?

- Regression viewpoint: MS_E from the model fitted.
- Anova viewpoint: "pure error" MS from fitting full treatment model.

The latter is unbiased, justified by a valid randomization, objective and gives unambiguous inferences whose properties are understood.

Then optimum design criteria do not have the statistical interpretations claimed for them, since the sizes of confidence intervals and regions depend on the degrees of freedom for pure error.

Which estimate of σ^2?

Inference should be carried out using the unbiased pure-error estimator of σ^2, since the biases induced by s_p^2 are unmeasureable and the inferences are therefore difficult to interpret.

If carrying out inference is important, then the design should be chosen in order to make that inference as informative as possible.

Note, however, statistical inference (beyond point estimation) is often not the most important part of the analysis and interpretation of experimental data.

D-optimality

Assume that the aim is to obtain unbiased confidence intervals or regions of minimal length or volume (standard hypothesis tests are essentially the same).

D-optimality minimises $1/|\mathbf{X}'\mathbf{X}|$, which is intended to minimise the volume of the joint confidence region for the parameters, since this is proportional to $|\mathbf{X}'\mathbf{X}|^{-1/2}$, where $\mathbf{X}$ is the polynomial model matrix, given the treatment design, with ith row $\mathbf{f}(\mathbf{x}_i)'$.

This is correct, "with σ^2 known or else (pure error degrees of freedom) the same for all designs" (Kiefer, 1959).

DP-optimality

However, the confidence region is actually proportional to

$$(F_{p,d;1-\alpha})^{p/2} \, |\mathbf{X}'\mathbf{X}|^{-1/2},$$

where p is the number of parameters in the model, d is the number of pure error degrees of freedom and $F_{p,d;1-\alpha}$ is the $1-\alpha$ quantile of the F distribution with p numerator and d denominator degrees of freedom.

Thus the *DP*(α) criterion is to minimise

$$\frac{(F_{p,d;1-\alpha})^{p}}{|\mathbf{X}'\mathbf{X}|}.$$

DP_S-optimality

Similarly, to minimise the volume of a joint confidence region for a subset of p_2 of the parameters, the $(DP)_S$ criterion is to minimise

$$(F_{p_2,d;1-\alpha})^{p_2} \, |(\mathbf{M}^{-1})_{22}|,$$

where $(\mathbf{M}^{-1})_{22}$ is the portion of $\mathbf{M}^{-1} = (\mathbf{X}'\mathbf{X})^{-1}$ corresponding to the p_2 parameters of interest.

Note that if the nuisance parameters are the intercept or the intercept plus block effects, then standard D_S-optimality reduces to D-optimality, but $(DP)_S$ does not reduce to DP. Hence, $(DP)_S$ is usually more useful than DP.

$(DP)_S$ with the subset of p_2 parameters being those of second order should be used when a major objective of the experiment is to compare the first order model with the second order model.

Other criteria

- *LP*-optimality for confidence intervals on single parameters, or functions of parameters;
- *GP*-optimality for confidence intervals on predicted responses;
- *IP*-optimality for confidence intervals on predicted responses;
- *TP*-optimality for comparing non-nested models.

Other criteria

- *LP*-optimality for confidence intervals on single parameters, or functions of parameters;
- *GP*-optimality for confidence intervals on predicted responses;
- *IP*-optimality for confidence intervals on predicted responses;
- *TP*-optimality for comparing non-nested models.

Usually, the inference actually done will be related to $(DP)_S$ or LP.

Comments

1. The general idea appears in Fisher (1966, p.242-245) in the context of sample size calculations, but based on fiducial probability: it "is unintelligible only to those who over a long period resisted the cogency of the fiducial argument".
2. With unstructured treatments, the number of distinct treatments is constant, so d depends only on the total number of experimental units and hence the new criteria are identical to the standard criteria.
3. As $d \to \infty$, the new criteria converge to the standard criteria. Hence in very large experiments, the designs chosen will be the same.
4. The concept of continuous design is not meaningful with the new criteria, since the quantiles of the F distributions are not proportional to n.
5. The standard versions of most criteria are meaningful in terms of point estimation, so the choice depend on the proposed analysis.

Multiple analyses

If a joint confidence region or a global F-test of the treatment parameters will be the only relevant analysis, then a $(DP)_S$-optimum design should be chosen.

However, this is not how experimental data are usually analysed.

Multiple criteria

In practice, several types of data analysis are important, not all of them requiring an estimate of error, e.g.:

1. a global F-test of the treatment parameters, for which we should use $(DP)_S$-optimality;
2. t-tests of the individual treatment parameters, for which we should use weighted-AP-optimality;
3. point estimation of the individual treatment parameters, for which we should use D- or weighted-A-optimality;
4. checking for lack of fit of the assumed treatment model and, if appropriate, fitting a few higher order terms, for which we use degree-of-freedom efficiency (for now).

Efficiencies

Define the following efficiencies, for the design with treatment model matrix $\mathbf{X}$ which has d pure error degrees of freedom:

- $(DP)_S$-efficiency:

$$E_1 = \frac{|\mathbf{X}'\mathbf{Q}_0\mathbf{X}|^{\frac{1}{p-1}} F_{p-1,d_D;1-\alpha_1}}{F_{p-1,d;1-\alpha_1} |(\mathbf{X}'_{DP}\mathbf{Q}_0\mathbf{X}_{DP})|^{\frac{1}{p-1}}},$$

where $\mathbf{X}_{DP}$ is the model matrix for the $(DP)_S$-optimum design, which has d_D degrees of freedom for pure error, and the global F-test will be performed at the $100\alpha_1\%$ level of significance.

Efficiencies

- Weighted-AP-efficiency,

$$E_2 = \frac{tr\{\mathbf{W}(\mathbf{X}'_{AP}\mathbf{X}_{AP})^{-1}\}F_{1,d_A;1-\alpha_2}}{tr\{\mathbf{W}(\mathbf{X}'\mathbf{X})^{-1}\}F_{1,d;1-\alpha_2}},$$

  where $\mathbf{X}_{AP}$ is the model matrix for the weighted-AP optimum design, which has d_A degrees of freedom for pure error and the individual t-tests will be calculated at the $100\alpha_2\%$ level of significance.

Efficiencies

- Weighted-A-efficiency,

$$E_3 = \frac{tr\{\mathbf{W}(\mathbf{X}'_A\mathbf{X}_A)^{-1}\}}{tr\{\mathbf{W}(\mathbf{X}'\mathbf{X})^{-1}\}},$$

where $\mathbf{X}_A$ is the model matrix for the weighted-A optimum design.

Efficiencies

- Weighted-A-efficiency,

$$E_3 = \frac{tr\{\mathbf{W}(\mathbf{X}'_A\mathbf{X}_A)^{-1}\}}{tr\{\mathbf{W}(\mathbf{X}'\mathbf{X})^{-1}\}},$$

  where $\mathbf{X}_A$ is the model matrix for the weighted-A optimum design.

- Degree-of-freedom efficiency,

$$E_4 = \frac{n-d}{n}.$$

Compound criterion

Use weights $\kappa_1, \ldots, \kappa_4$ respectively to get $E = E_1^{\kappa_1} E_2^{\kappa_2} E_3^{\kappa_3} E_4^{\kappa_4}$.

Choose a design to maximise

$$\frac{|\mathbf{X}'\mathbf{Q}_0\mathbf{X}|^{\frac{\kappa_1}{p-1}}(n-d)^{\kappa_4}}{(F_{p-1,d;1-\alpha_1})^{\kappa_1}(F_{1,d;1-\alpha_2})^{\kappa_2}[tr\{\mathbf{W}(\mathbf{X}'\mathbf{X})^{-1}\}]^{\kappa_2+\kappa_3}}. \tag{5}$$

The weights $\boldsymbol{\kappa}$ should be chosen to reflect the relative importance of different aspects of the analysis.

Compound criterion

Use weights $\kappa_1, \ldots, \kappa_4$ respectively to get $E = E_1^{\kappa_1} E_2^{\kappa_2} E_3^{\kappa_3} E_4^{\kappa_4}$.

Choose a design to maximise

$$\frac{|\mathbf{X}'\mathbf{Q}_0\mathbf{X}|^{\frac{\kappa_1}{p-1}}(n-d)^{\kappa_4}}{(F_{p-1,d;1-\alpha_1})^{\kappa_1}(F_{1,d;1-\alpha_2})^{\kappa_2}[tr\{\mathbf{W}(\mathbf{X}'\mathbf{X})^{-1}\}]^{\kappa_2+\kappa_3}}. \tag{5}$$

The weights $\boldsymbol{\kappa}$ should be chosen to reflect the relative importance of different aspects of the analysis.

What about prediction?

Prediction of Responses

For any point $\mathbf{x} \in \mathcal{X}$,

$$\text{var}(\hat{y}(\mathbf{x})) = \sigma^2 \mathbf{f}(\mathbf{x})'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}).$$

An I-optimum design minimizes the average variance of predictions over $\mathcal{X}$. Let $\Psi = \int_{\mathbf{x}\in\mathcal{X}} d\mathbf{x}$ be the volume of $\mathcal{X}$. Then

$$\text{average variance} = \Psi^{-1} \int_{\mathbf{x}\in\mathcal{X}} \text{var}(\hat{y}(\mathbf{x}))d\mathbf{x} \propto \int_{\mathbf{x}\in\mathcal{X}} \mathbf{f}(\mathbf{x})'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})d\mathbf{x}.$$

It can be shown that

$$\text{average variance} \propto \text{trace}\left[\mathcal{M}(\mathbf{X}'\mathbf{X})^{-1}\right],$$

where $\mathcal{M} = \int_{\mathbf{x}\in\mathcal{X}} \mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})'d\mathbf{x}$ is the moment matrix of the region.

For spherical and cubic regions and the second order model, $\mathcal{M}$ is given explicitly in Hardin and Sloane (1991a,1991b).

Differences in Response

We are often interested in differences between the response at the expected optimum or standard operating conditions and the response at other locations, i.e. $y(\mathbf{x}) - y(\mathbf{x}_0)$.

If $\mathbf{x}_0 = \mathbf{0}$ the focus should be on estimating $y(\mathbf{x}) - \beta_0$.

Several arguments for this:

- Randomization ensures unbiased estimation of β_j for $j \neq 0$, but not for β_0, which requires the experimental units to be a random sample from a population of units.
- Estimating the location of the optimum, identifying ridges and canonical analysis do not depend on β_0.
- If $\mathbf{x}_0 = \mathbf{0}$ are standard operating conditions, then we already know β_0 and the best prediction is $\tilde{y}(\mathbf{x}) = \beta_0 + \hat{y}(\mathbf{x}) - \hat{\beta}_0$.

Designs for Differences in Response

Since

$$\text{var}[\tilde{y}(\mathbf{x})] = \text{var}[\hat{y}(\mathbf{x}) - \hat{\beta}_0] = \text{var}[\hat{y}(\mathbf{x}) - \hat{y}(\mathbf{x}_0)],$$

even if we want to predict the response, we should optimize the prediction of differences in response.

Define the I_D criterion to be to minimize

$$\begin{aligned} \text{average difference variance} &= \Psi^{-1} \int_{\mathbf{x}\in\mathcal{X}} \text{var}[\hat{y}(\mathbf{x}) - \hat{y}(\mathbf{x}_0)] d\mathbf{x} \\ &\propto \int_{\mathbf{x}\in\mathcal{X}} [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)]'(\mathbf{X}'\mathbf{X})^{-1}[\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)] d\mathbf{x}. \end{aligned}$$

For $\mathbf{x}_0 = \mathbf{0}$ we have

$$\text{average difference variance} \propto \text{trace}\left[\mathcal{M}_0(\mathbf{X}'\mathbf{X})^{-1}\right],$$

where $\mathcal{M}_0 = \int_{\mathbf{x}\in\mathcal{X}} [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})][\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})]' d\mathbf{x}$ and $\mathcal{M}_0$ is the $\mathcal{M}$ matrix with first row and first column set to zero.

Example 1: Cassava Bread

Three factors in 26 runs:

- X_1 amount of powder albumen;
- X_2 amount of yeast;
- X_3 amount of cassava flour.

Cubic region of interest, several response variables.

Experimenters used CCD; we consider alternatives.

Properties of Designs

Criterion	df(PE, LoF)†	Efficiency							
		D_S	$(DP)_S$	A_S	$(AP)_S$	I	(IP)	I_D	(I_DP)
D_S, A_S	(9, 7)	**100.00**	86.77	**100.00**	95.50	75.80	72.32	91.93	87.00
$(DP)_S$	(15, 1)	93.81	**100.00**	87.12	93.72	69.62	74.82	83.47	88.98
$(AP)_S$	(12, 4)	98.79	97.45	97.13	**100.00**	72.30	74.36	89.23	91.02
I	(5, 11)	90.71	52.42	87.71	64.87	**100.00**	73.88	99.87	73.19
(IP)	(12, 4)	79.79	78.70	72.80	74.95	97.23	**100.00**	87.47	89.23
I_D	(5, 11)	93.36	53.96	90.67	67.06	97.22	71.83	**100.00**	73.28
(I_DP)	(12, 4)	95.29	93.99	92.11	94.82	92.00	94.63	98.03	**100.00**
CCD	(11, 5)	90.89	86.56	82.43	83.16	84.70	85.37	87.99	87.95
BBD	(13, 3)	78.71	79.99	70.79	74.13	68.64	71.81	58.70	60.90

†df(PE, LoF): degrees of freedom for pure error, degrees of freedom for lack of fit.

Graphical Comparison

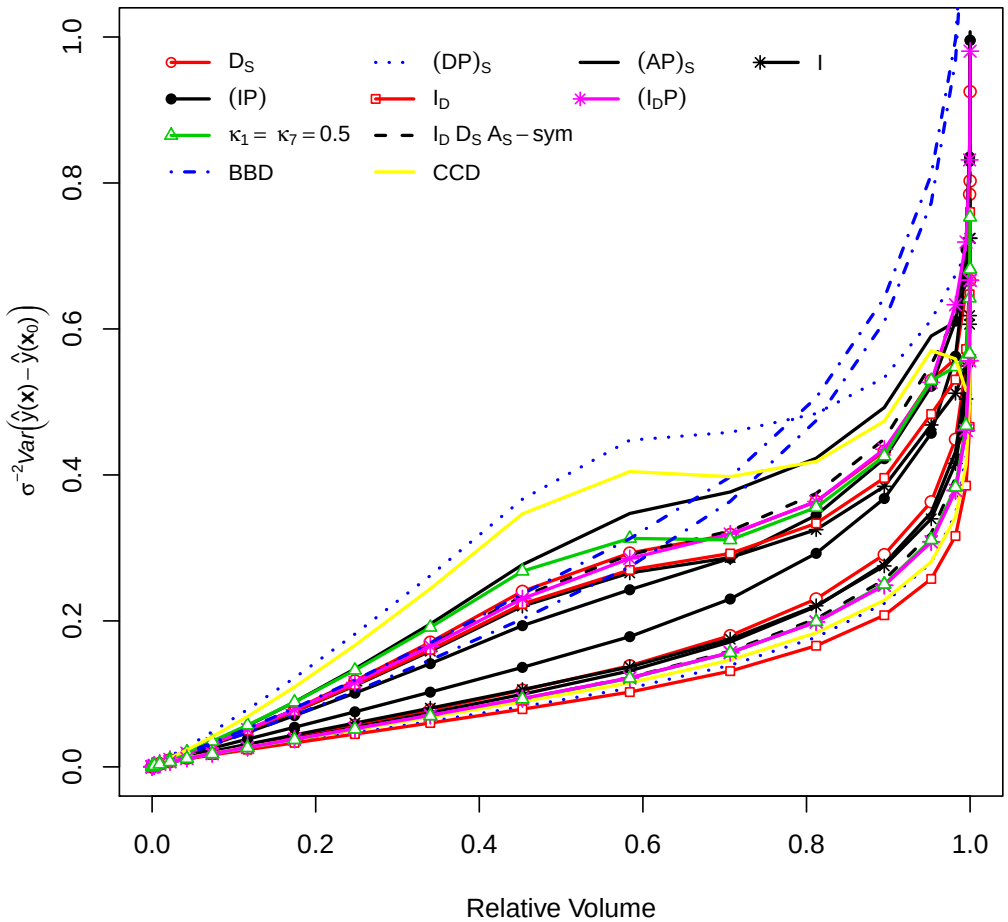


Figure: DVDG for point predictions vs. relative volume of the region of designs in Example 1

Example 2

Five factors in a spherical region, $n = 30$.

I_D-optimal design turns out to be a CCD.

Criterion	df(PE, LoF)†	Efficiency							
		D_S	$(DP)_S$	A_S	$(AP)_S$	I	(IP)	I_D	(I_DP)
D_S, I	(0, 9)	**100.00**	0.00	94.02	0.00	**100.00**	0.00	60.31	0.00
$(DP)_S$	(9, 0)	86.30	**100.00**	74.33	90.36	74.73	97.81	52.80	63.56
A_S	(1, 8)	98.16	1.35	**100.00**	3.85	92.86	3.85	81.20	3.10
$(AP)_S$	(8, 1)	87.39	94.39	85.48	**100.00**	74.34	93.64	84.84	98.28
(IP)	(8, 1)	88.84	95.95	79.04	92.47	79.39	**100.00**	54.37	62.99
CCD, I_D	(3, 6)	96.96	38.09	95.25	58.51	91.82	60.73	**100.00**	60.82
(I_DP)	(8, 1)	85.37	92.20	83.63	97.83	72.21	90.95	86.32	**100.00**

†df(PE, LoF): degrees of freedom for pure error, degrees of freedom for lack of fit.

Graphical Comparison

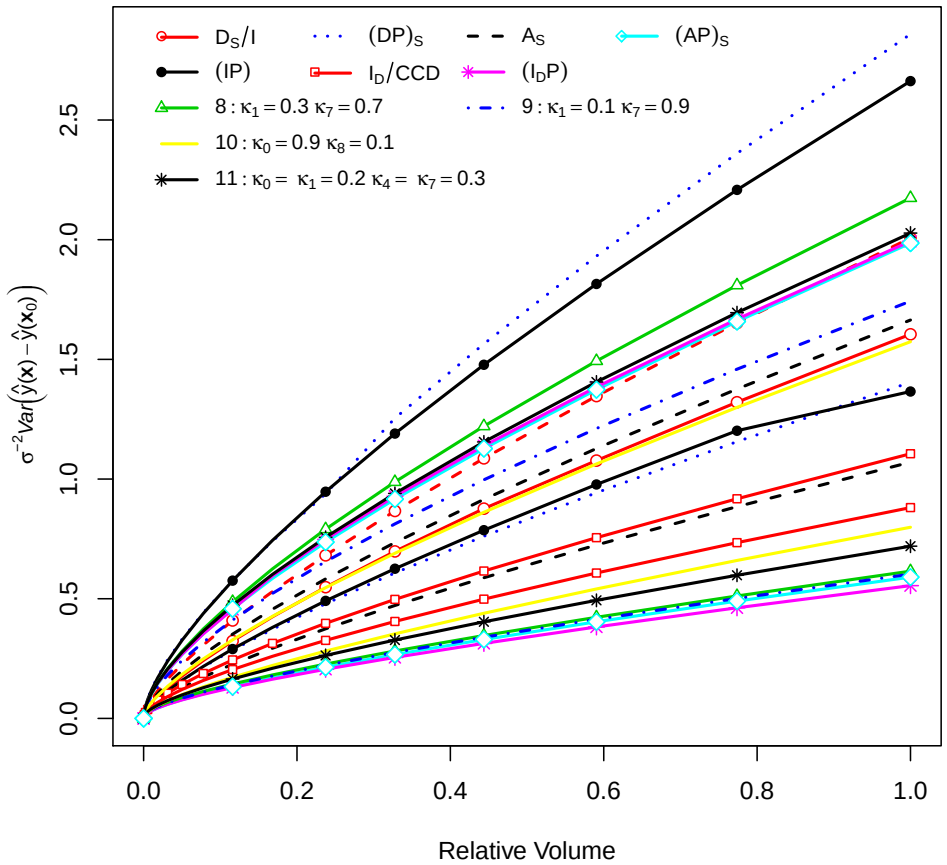


Figure: DVDG for point predictions vs. relative volume of the region of designs in Example 2

CCDs which are I_D-optimal

The I_D-optimality in a spherical region of the CCD is not unique to this run size.

- For three factors, the CCD is I_D-optimal for $17 \leq n \leq 20$, i.e. 3 to 6 centre points.
- For four factors, the CCD is I_D-optimal for $28 \leq n \leq 32$, i.e. 4 to 8 centre points.
- For five factors, the CCD, with a half-replicate of the factorial points, is I_D-optimal for $30 \leq n \leq 33$, i.e. 4 to 7 centre points.
- For six factors, the CCD, with a half-replicate of the factorial points, is I_D-optimal for $50 \leq n \leq 55$, i.e. 6 to 11 centre points.

Current and Further Work

- Improved criteria for testing lack of fit.
- Designs for interval prediction - (IP) and $(I_D P)$ criteria.
- Theoretical results on compound criteria - do we really need them all?
- Not just polynomial response surfaces.

EPSRC grant Multi-Objective Optimal Design of Experiments.

Current and Further Work

- Improved criteria for testing lack of fit.
- Designs for interval prediction - (IP) and $(I_D P)$ criteria.
- Theoretical results on compound criteria - do we really need them all?
- Not just polynomial response surfaces.

EPSRC grant Multi-Objective Optimal Design of Experiments.
Thank You!

Questions?