Minimax efficient random designs with application to model-robust design for prediction

Tim Waite

timothy.waite@manchester.ac.uk

School of Mathematics

University of Manchester, UK

Joint work with Dave Woods

S3RI, University of Southampton, UK

Supported by the UK Engineering and Physical Sciences Research Council

8 Aug 2017,
Banff, AB, Canada

Outline

Randomized decisions and experimental design

Random designs for prediction - correct model

- Extension of G-optimality

Model-robust random designs for prediction

- Theoretical results - tractable classes
- Algorithms for optimization
- Examples: illustration of bias-variance tradeoff

Randomized decisions

A well known fact in statistical decision theory and game theory:

- Under minimax expected loss, random decisions beat deterministic ones.

Experimental design can be viewed as a game played by the Statistician against nature (Wu, 1981; Berger, 1985).

Therefore a random design strategy should often be beneficial.

Despite this, consideration of minimax efficient random design strategies is relatively unusual.

Game theory

Consider a two-person zero-sum game.

Player I takes action $\theta \in \Theta$ and Player II takes action $\xi \in \Xi$.

Player II experiences a loss $L(\theta, \xi)$, to be minimized.

A random strategy for Player II is a probability measure π on Ξ. Deterministic actions are a special case (point mass distribution).

Strategy π_1 is preferred to π_2 ($\pi_1 \succ \pi_2$) iff

$$\mathsf{E}_{\pi_1} L(\theta, \xi) < \mathsf{E}_{\pi_2} L(\theta, \xi)\,.$$

However, Player I's choice of θ is unknown to Player II.

To account for uncertainty about θ, the standard choice is to play (if it exists) a minimax strategy, π^*, such that

$$\max_{\theta \in \Theta} \mathsf{E}_{\pi^*} L(\theta, \xi) = \inf_{\pi} \max_{\theta \in \Theta} \mathsf{E}_{\pi} L(\theta, \xi) \,.$$

If both action spaces Θ and Ξ are finite (and not too large), minimax random strategies can be computed easily by solving a related linear programming problem.

Example: paper-rock-scissors, $\Theta = \Xi = \{P, R, S\}$, with loss matrix $L(\theta, \xi)$ below

		ξ		
		P	R	S
	P	0	1	-1
θ	R	-1	0	1
	S	1	-1	0

Let δ be any deterministic strategy and $\pi = U(\{P, R, S\})$, then

$$\mathsf{E}_\pi\, L(\theta, \xi) = \frac{1}{3} \times (-1) + \frac{1}{3} \times 0 + \frac{1}{3} \times 1 = 0\,, \quad \forall \theta \in \Theta\,.$$

Hence $\max_\theta \mathsf{E}_\pi\, L(\theta, \xi) = 0$ and $\max_\theta \mathsf{E}_\delta\, L(\theta, \xi) = 1$.

Thus π is preferable to any deterministic design. Indeed π is optimal.

Frequentist decision-theoretic experimental design

In optimal (exact) design, attention is usually restricted to a deterministic choice of design, $\boldsymbol{\xi} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \Xi = \mathcal{X}^n$, a set of n points in design space $\mathcal{X}$.

We prefer a design with the lowest possible value of the risk

$$R(\boldsymbol{\theta}, \boldsymbol{\xi}) = \mathrm{E}_{\mathbf{y}|\boldsymbol{\xi},\boldsymbol{\theta}}\, L(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{y})$$

However the risk often depends on a vector $\boldsymbol{\theta} \in \Theta$ of fixed unknowns (e.g. model parameters in a nonlinear model).

Hence, it is unknown which designs have minimum risk.

[Design selection can be viewed as a game with loss $L(\boldsymbol{\theta}, \boldsymbol{\xi}) = R(\boldsymbol{\theta}, \boldsymbol{\xi})$.
Player I: Nature, chooses $\boldsymbol{\theta}$; Player II: the Statistician, chooses $\boldsymbol{\xi}$.]

Minimax design

There is thus a need to account for uncertainty about $\boldsymbol{\theta}$ when choosing $\boldsymbol{\xi}$.

Many frequentists are reluctant to use prior distributions. In this case, typically a deterministic minimax design is sought, i.e. a $\boldsymbol{\xi}^* \in \Xi$ that minimizes $\max_{\boldsymbol{\theta} \in \Theta} R(\boldsymbol{\theta}, \boldsymbol{\xi})$.

Random designs

Considerations from game theory and statistical decision theory would suggest that we also allow a random design, i.e. a probability measure π on Ξ.

Interpretation: choose the realized design $\boldsymbol{\xi}$ at random by sampling from π.

(A deterministic design corresponds to a point mass distribution.)

Expected loss for random designs

If L is truly the loss function, utility theory implies that the performance of π is to be measured via

$$R(\boldsymbol{\theta}, \pi) = \mathsf{E}_{\mathbf{y}, \boldsymbol{\xi} | \boldsymbol{\theta}} \, L(\boldsymbol{\theta}, \boldsymbol{\xi}, y) \, .$$

This makes intuitive sense:

- For a deterministic design we considered the repeated sampling distribution for L over hypothetical replications of the entire experiment.
- For a random design we do the same, but now for a different hypothetical replication, a different $\boldsymbol{\xi}$ will be sampled from π.

A minimax random design π^* satisfies

$$\max_{\boldsymbol{\theta}} R(\boldsymbol{\theta}, \pi^*) = \inf_{\pi} \max_{\boldsymbol{\theta}} R(\boldsymbol{\theta}, \pi) \, .$$

Example: Fisherian randomization

Consider a linear model contaminated by fixed unknown additive unit effects, $\mathbf{u} = (u_1, \ldots, u_n)^{\mathrm{T}} \in \mathcal{U}$,

$$y_i = \mathbf{f}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta} + u_i + \epsilon_i\,, \quad \epsilon_i \sim N(0, \sigma^2)\,.$$

It was shown in many cases that the minimax random design strategy is Fisherian randomization of a standard design.

- π minimizes $\max_{\mathbf{u}\in\mathcal{U}} R(\mathbf{u}, \pi)$.
- Assumptions about the structure of the experimental units, e.g. exchangeability/blocks, described by a permutation group G.
- Different loss functions considered, e.g. A, L-optimality.

[Wu, 1981; Li, 1983; Hooper, 1989; Bhaumik and Mathew, 1995].

Fisherian randomization 'is one of the greatest contributions of R. A. Fisher to science and statistics' (Wu, 1981).

It seems to us a weakness of standard optimal design theory that Fisherian randomization does not arise as a necessary mathematical consequence.

The preceding slide shows that it does arise directly from random designs and the minimax principle.

This seems to be a big hint that minimax random designs are worth considering more widely.

Design for point prediction

Suppose we have a normal theory linear model,

$$y_i = \mathbf{f}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta} + \epsilon_i\,, \quad \epsilon_i \sim N(0, \sigma^2)\,,$$

with design points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X} \subseteq [-1, 1]^q$.

Suppose the goal is prediction at an unknown point $\mathbf{x}$, with squared error loss

$$L(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{y}) = [\mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta} - \mathbf{f}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}}]^2\,,$$

depending on $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{x})$.

Above, $\hat{\boldsymbol{\beta}} = (\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}})^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{y}$ is the least squares estimator, $\mathbf{F}_{\boldsymbol{\xi}}$ is the model matrix.

Given a design $\boldsymbol{\xi}$ and σ^2, the risk is

$$R(\mathbf{x}, \boldsymbol{\xi}) = \sigma^2 \mathbf{f}^{\mathrm{T}}(\mathbf{x})(\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\xi}})^{-1} \mathbf{f}(\mathbf{x}) \,.$$

Thus the minimax deterministic design minimizes

$$\Phi(\boldsymbol{\xi}) = \max_{\mathbf{x} \in \mathcal{X}} \mathbf{f}^{\mathrm{T}}(\mathbf{x})(\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\xi}})^{-1} \mathbf{f}(\mathbf{x}) \,,$$

i.e. it is the classic G-optimal design.

However, this may be beaten by a minimax random design π^*, which minimizes

$$\Phi(\pi) = \max_{\mathbf{x} \in \mathcal{X}} \mathbf{f}^{\mathrm{T}}(\mathbf{x}) \, \mathrm{E}_{\pi}\{(\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\xi}})^{-1}\} \mathbf{f}(\mathbf{x}) \,.$$

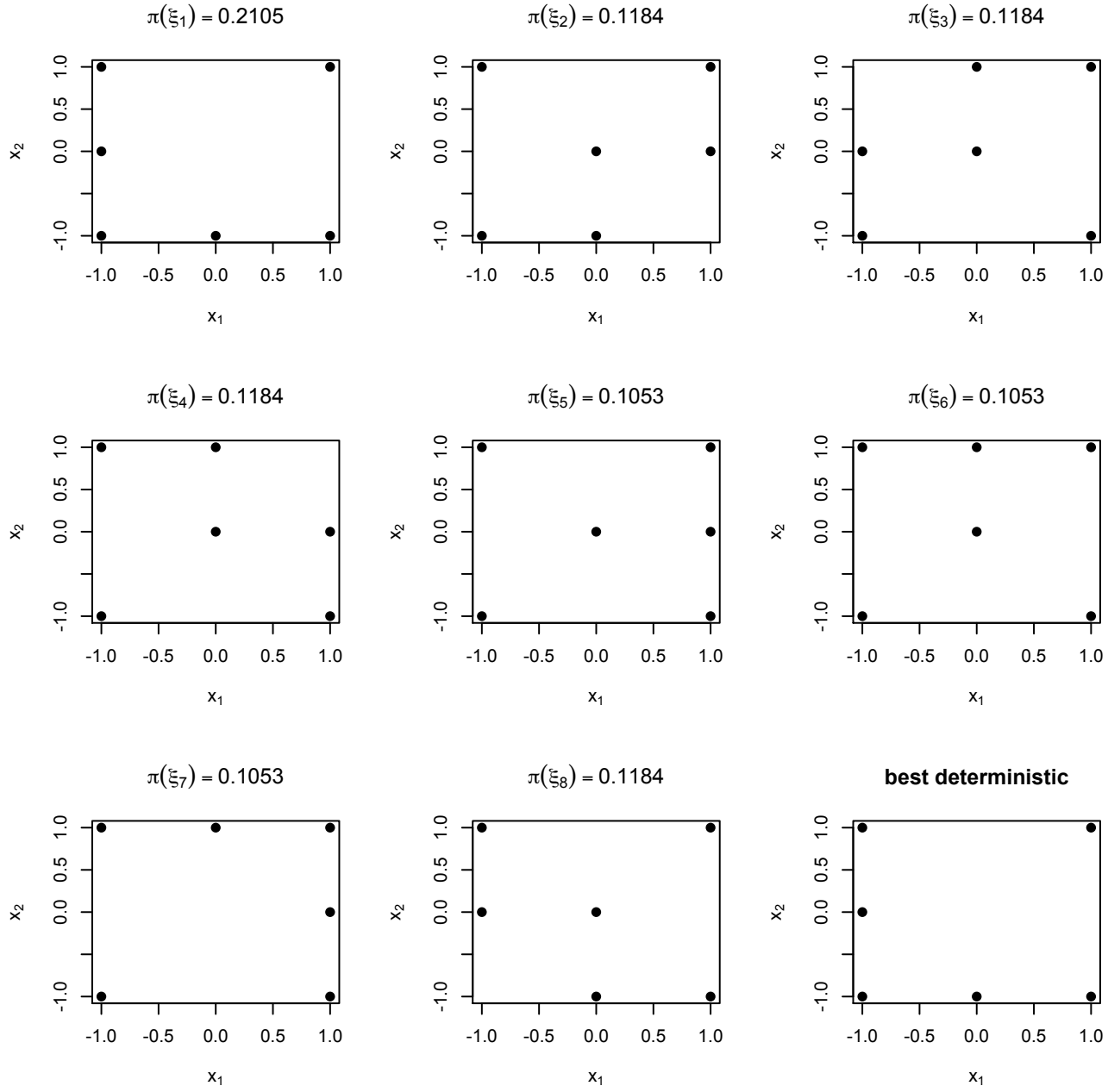
Example: quadratic model, 2 factors, $\mathcal{X} = \{-1, 0, 1\}^2$, $n = 6$.

There are 76 possible non-singular designs up to permutations of run order.

The minimax deterministic design has maximum expected loss $2.75\sigma^2$.

Using linear programming, an optimal random design can be obtained; it has maximum expected loss $1.55\sigma^2$.

The efficiency of the deterministic design is just 56%.


$\pi(\xi_1) = 0.2105$
$\pi(\xi_2) = 0.1184$
$\pi(\xi_3) = 0.1184$
$\pi(\xi_4) = 0.1184$
$\pi(\xi_5) = 0.1053$
$\pi(\xi_6) = 0.1053$
$\pi(\xi_7) = 0.1053$
$\pi(\xi_8) = 0.1184$
best deterministic
x_1
x_2
-1.0
-0.5
0.0
0.5
1.0

Model-robust design

Variance-based optimality criteria

Classic D, A, E-optimality etc. all assume that a particular parametric model is correct.

Bias

Box & Draper (1959) - polynomials of uncertain degree, found if model incorrect better off choosing the design to minimize bias, ignoring variance.

Their investigation focussed only on polynomials.

More sophisticated treatments of model-robust design exist (e.g. Wiens, 2015).

Suppose that $\mathbf{x} \in \mathcal{X}$, with $\mathcal{X} \subseteq [-1, 1]^q$, the design space. We assume

$$y \sim N[\mu(\mathbf{x}), \sigma^2] \,,$$

and $\lambda(\mathcal{X}) > 0$, where λ is Lebesgue measure.

Standard linear model approach

Find an optimal design assuming that there exists a true parameter vector, $\boldsymbol{\beta}_{\text{true}}$, such that

$$\mu(\mathbf{x}) = \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}_{\text{true}} \,.$$

Above, $\mathbf{f} : \mathcal{X} \to \mathbb{R}^p$ a vector of regressor functions, e.g. $\mathbf{f}(x) = (1, x, x^2)^{\mathrm{T}}$.

Not very robust.

Model-robust approach (e.g. Wiens, 2015)

Assume explicitly that the linear regression function is an approximation to μ, i.e.

$$\mu(\mathbf{x}) = \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}_{\mathsf{ba}} + \psi(\mathbf{x})\,,$$

where $\boldsymbol{\beta}_{\mathsf{ba}}$ minimizes the L_2-norm of the approximation error, i.e.

$$\boldsymbol{\beta}_{\mathsf{ba}} \in \arg\min_{\boldsymbol{\beta}} \int_{\mathcal{X}} [\mu(\mathbf{x}) - \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}]^2 d\lambda(\mathbf{x})\,.$$

In this case, the discrepancy ψ is orthogonal to the regressors

$$\langle \psi, \mathbf{f} \rangle = \int_{\mathcal{X}} \psi(\mathbf{x})\mathbf{f}(\mathbf{x}) d\lambda(\mathbf{x}) = \mathbf{0}_q\,.$$

[cf. L_2-calibration of computer models (Tuo and Wu, 2015; Plumlee, 2016).]

Discrepancy classes

An approximately linear model is specified as

$$\mu(\mathbf{x}) = \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}_{\mathsf{ba}} + \psi(\mathbf{x})\,, \quad \boldsymbol{\beta}_{\mathsf{ba}} \in B,\, \psi \in \mathcal{H}\,,$$

where $\mathcal{H}$ is a set containing all discrepancy functions considered possible.

The choice for $\mathcal{H}$ that has received the most attention is

$$\mathcal{H} = \left\{ \psi : \langle \psi, \mathbf{f} \rangle = \mathbf{0}, \int \psi(\mathbf{x})^2 d\lambda(\mathbf{x}) \leq \tau^2 \right\}.$$

(cf. Huber, 1981; Dette and Wiens, 2009; Wiens, 2015).

[Alternatives: Box and Draper (1959), ψ a polynomial; Li and Notz (1980), $|\psi| \leq \tau_\infty$; Yue and Hickernell (1999), ψ belongs to a smoothness class.]

Loss function

We suppose that the loss is the integrated squared prediction error (ISPE)

$$L(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{y}) = \int_{\mathcal{X}} [\mu(\mathbf{x}) - \mathbf{f}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}}]^2 d\lambda(\mathbf{x}) \,,$$

with $\boldsymbol{\theta} = (\psi, \boldsymbol{\beta}_{\mathsf{ba}})$ and $\hat{\boldsymbol{\beta}} = (\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}})^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{y}$ the usual least squares estimator.

- Assumption: predictions are made from the fitted linear model, ignoring discrepancy.
- Alternatively, one could attempt to model ψ nonparametrically (e.g. Plumlee, 2016).
- For small τ^2, our 'shrinkage' approach may be more efficient in terms of expected ISPE.

For a random design π, given σ^2 the risk satisfies

$$R(\boldsymbol{\theta}, \pi) = R(\psi, \pi)\,.$$

A minimax design is found by minimizing

$$\sup_{\psi \in \mathcal{H}} R(\psi, \pi)\,.$$

A fundamental problem

For any finite and deterministic design, $\sup_{\psi \in \mathcal{H}} R(\psi, \boldsymbol{\xi}) = \infty$. (Wiens, 1992)

Thus, minimax MISPE cannot be used to select a finite deterministic design.

Several authors have considered infinitely supported deterministic designs (defined via a pdf). We argue that it is more coherent to use a finite but random design.

Random translation designs

Definition

A random design $\pi = \pi^{\mathsf{RT}}(\{\mathbf{c}_i,\}_{i=1}^n, \mathcal{T})$ is a random translation design if there exists

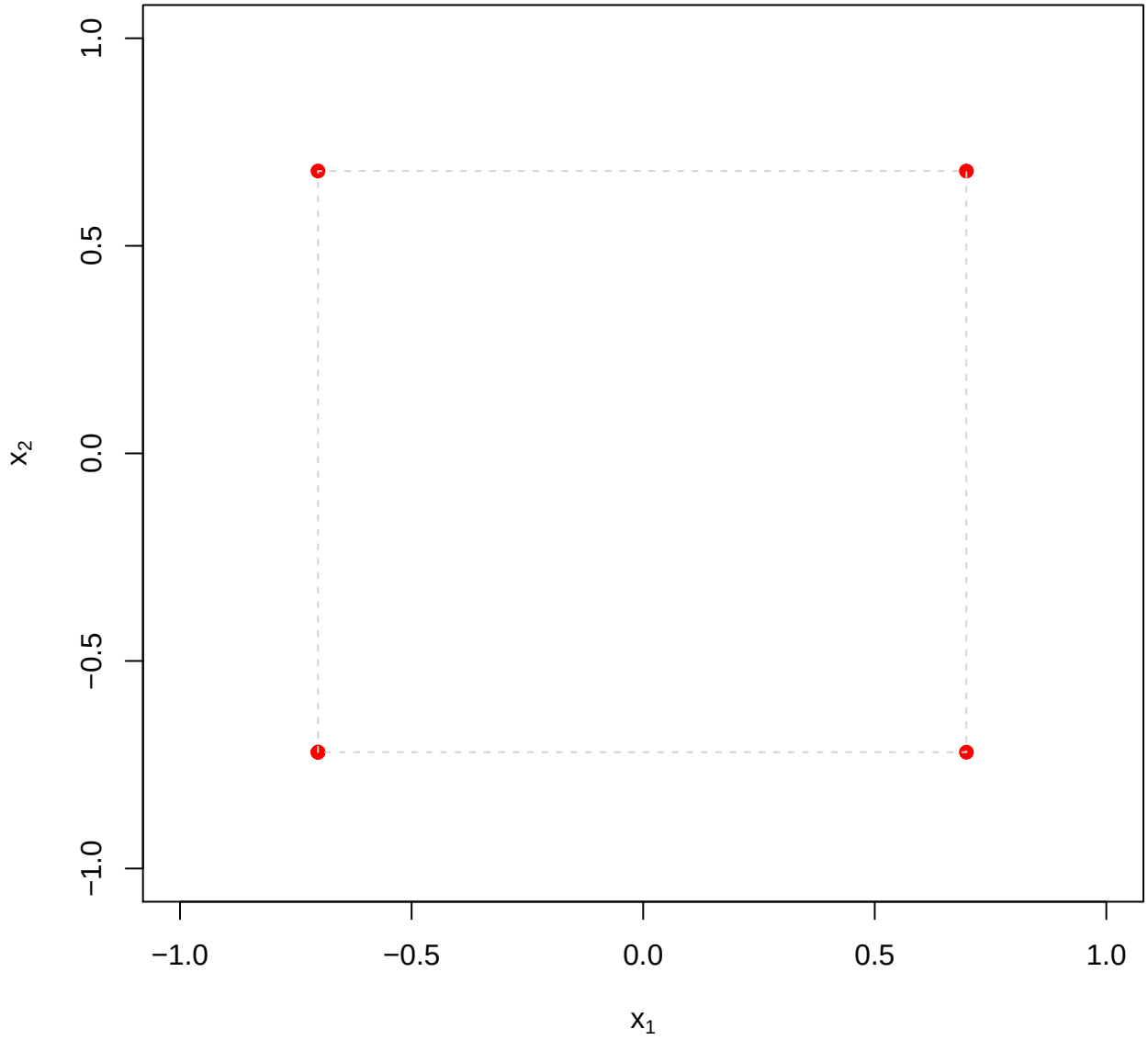
- $\mathbf{c}_i \in \mathcal{X}$, $i = 1, \ldots, n$,
- a closed convex measurable set $\mathcal{T} \subseteq \mathbb{R}^q$,

such that:

(i) the design can be written as $\boldsymbol{\xi} = \boldsymbol{\xi}(\mathbf{t}) = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ with

$$\mathbf{x}_i = \mathbf{c}_i + \mathbf{t}\,, \quad \mathbf{t} \sim U(\mathcal{T})\,, \quad \mathsf{E}[\mathbf{t}] = \mathbf{0}_q\,,$$

(ii) the sets $\mathbf{c}_i + \mathcal{T}$ are distinct, disjoint subsets of $\mathcal{X}$


1.0
0.5
0.0
−0.5
−1.0
x_2
−1.0
−0.5
0.0
0.5
1.0
x_1

Theorem

For a random translation design,

$$\sup_{\psi \in \mathcal{H}} R(\psi, \pi) = \underbrace{\sigma^2 \, \mathsf{E}_\pi \, \mathrm{tr}(\mathbf{A}\mathbf{M}_{\boldsymbol{\xi}}^{-1})}_{\text{variance}} + \underbrace{\tau^2 + \frac{\tau^2}{\lambda(\mathcal{T})} \cdot \max_{\mathbf{t} \in \mathcal{T}} \lambda_{\max}(\mathbf{K}_{\boldsymbol{\xi}(\mathbf{t})})}_{\text{bias}^2} .$$

$$\mathbf{M}_{\boldsymbol{\xi}} = \sum_{i=1}^{n} \mathbf{f}(\mathbf{x}_i)\mathbf{f}^{\mathrm{T}}(\mathbf{x}_i) , \qquad \mathbf{F}_{\boldsymbol{\xi}} = [\mathbf{f}(\mathbf{x}_1) \ldots \mathbf{f}(\mathbf{x}_n)]^{\mathrm{T}} ,$$

$$\mathbf{A} = \int_{\mathcal{X}} \mathbf{f}(\mathbf{x})\mathbf{f}^{\mathrm{T}}(\mathbf{x}) d\lambda(\mathbf{x}) , \quad \mathbf{K}_{\boldsymbol{\xi}} = \mathbf{F}_{\boldsymbol{\xi}} \mathbf{M}_{\boldsymbol{\xi}}^{-1} \mathbf{A} \mathbf{M}_{\boldsymbol{\xi}}^{-1} \mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}} .$$

$\lambda(\mathcal{T})$ denotes Lebesgue measure of $\mathcal{T}$.
$\lambda_{\max}(K)$ denotes the maximal eigenvalue of matrix K.

A random hypercube translation design (RHTD), $\pi = \pi^{\mathsf{RHT}}(\{\mathbf{c}_i\}_{i=1}^n, \delta)$, has

$$\mathcal{T} = [-\delta/2, \delta/2]^q \,,$$

with $\delta \geq 0$ controlling the degree of randomness.

Given $\mathbf{c}_i$ and δ we may compute the maximum risk via

1. Monte Carlo/quadrature evaluation of $\mathsf{E}_\pi \,\mathrm{tr}(\mathbf{A}\mathbf{M}_{\boldsymbol{\xi}}^{-1})$
2. Numerical search for $\mathbf{t}^* \in [-\delta/2, \delta/2]^q$ maximizing $\lambda_{\max}(\mathbf{K}_{\boldsymbol{\xi}(\mathbf{t})})$
   - $\mathbf{t}^*$ determines the 'most bias-sensitive' potential design realization

Extension

- The design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ can be replicated $r_1, \ldots, r_n$ times respectively.
- The expression in the theorem needs revising in this case.

What if a common translation is not used?

- E.g. if the design points are sampled from independent uniform distributions.
- In this case the expression in the Theorem is an upper bound.
- However, the upper bound is not sharp.
- Thus in this case the maximum risk for the random design becomes unknown.

Example

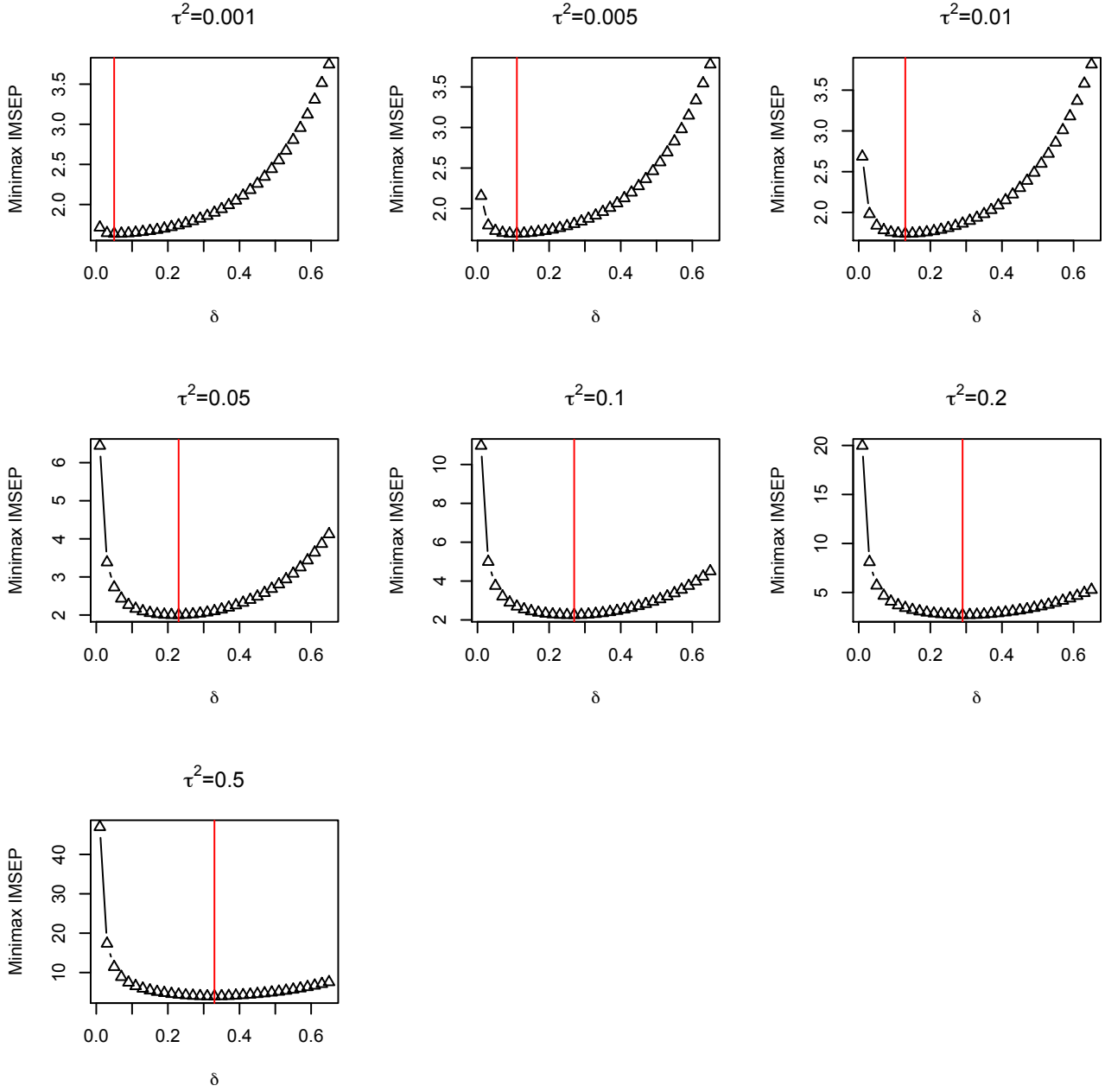
Approximately quadratic model. Assume $n = 3$, $q = 1$, $\mathcal{X} = [-1, 1]$, $\mathbf{f}^{\mathrm{T}}(x) = (1, x, x^2)$.

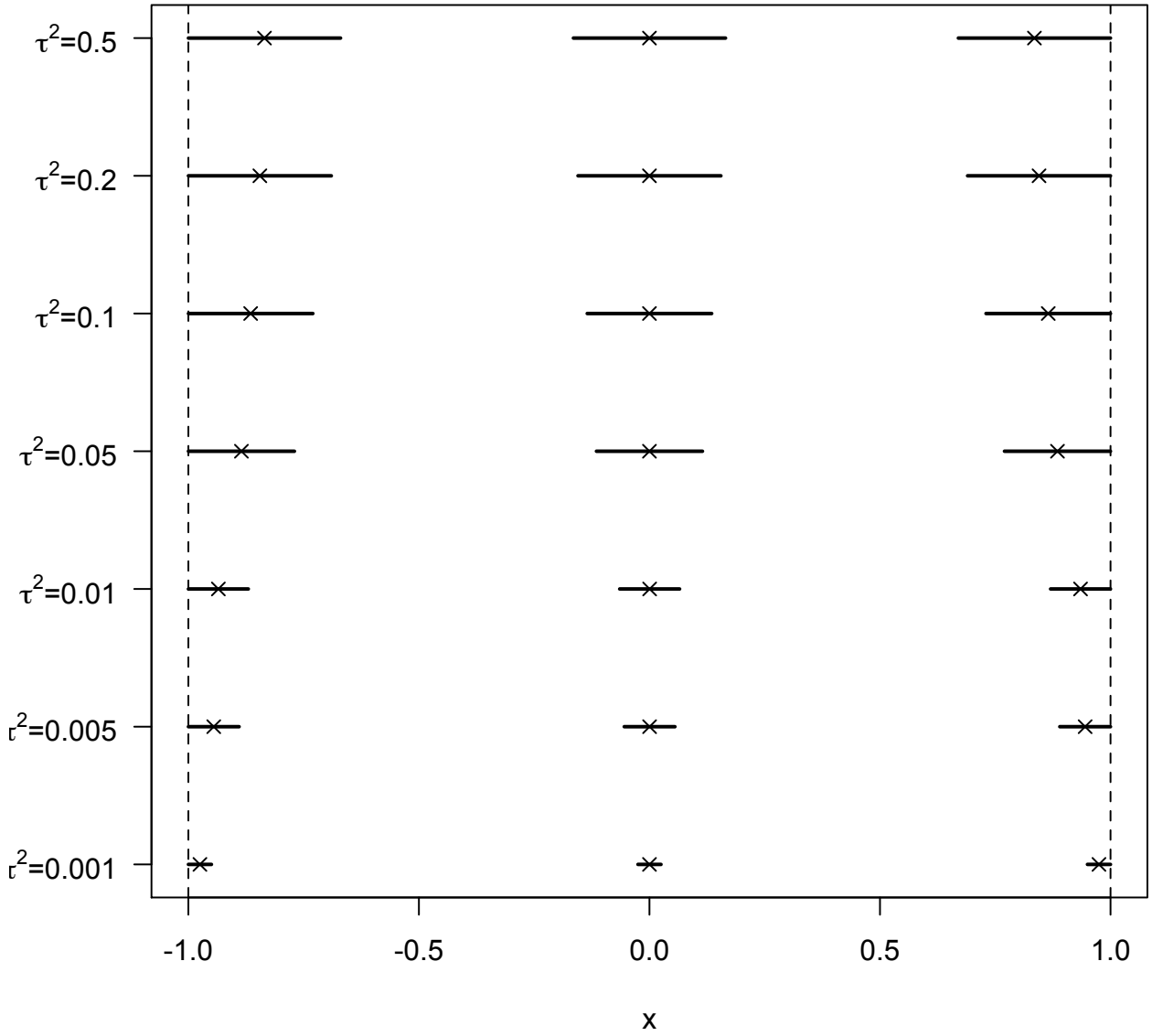
What is the optimal value of δ?

- Depends on $\frac{\tau^2}{\sigma^2}$ (wlog $\sigma^2 = 1$)
- We plot the 'profiled' minimax risk as a function of δ,

$$\tilde{R}^*(\delta) = \min_{c_i} \sup_{\psi \in \mathcal{H}} R\{\psi; \pi^{\mathsf{RHT}}(\{c_i\}_{i=1}^3, \delta)\}.$$

- For each δ, a co-ordinate algorithm is used to find the optimal c_i.


τ²=0.001
τ²=0.005
τ²=0.01
τ²=0.05
τ²=0.1
τ²=0.2
τ²=0.5
Minimax IMSEP
δ


τ²=0.5
τ²=0.2
τ²=0.1
τ²=0.05
τ²=0.01
τ²=0.005
τ²=0.001
-1.0
-0.5
0.0
0.5
1.0
x

How much variance efficiency do we need to sacrifice to guard against model discrepancy?

V-optimality

The deterministic design $\boldsymbol{\xi}_V^*$ is V-optimal for the approximate model if it minimizes

$$R(0, \boldsymbol{\xi}) = \sigma^2 \operatorname{tr}(\mathbf{A}\mathbf{M}_{\boldsymbol{\xi}}^{-1}),$$

computed using the assumption $\psi \equiv 0$ (i.e. the approximate model is correct).

The V-efficiency of a design realization $\boldsymbol{\xi}$ is

$$V\text{-eff}(\boldsymbol{\xi}) = \frac{R(0, \boldsymbol{\xi}_V^*)}{R(0, \boldsymbol{\xi})}.$$

For a random design, the V-efficiency is a random variable.

τ^2	δ^*	$(\mathbf{c}^*)^{\mathrm{T}}$	V-efficiency of $\boldsymbol{\xi}$ (%)
0.001	0.05	(-0.975, 0.000, 0.975)	99.1 – 99.3
0.005	0.11	(-0.945, 0.000, 0.945)	97.2 – 98.0
0.01	0.13	(-0.935, 0.000, 0.935)	96.4 – 97.5
0.05	0.23	(-0.885, 0.000, 0.885)	89.8 – 94.0
0.1	0.27	(-0.865, 0.000, 0.865)	86.0 – 92.2
0.2	0.31	(-0.845, 0.000, 0.845)	81.6 – 90.1
0.5	0.33	(-0.835, 0.000, 0.835)	79.1 – 88.9

Table: Approximately optimal random translation designers for several values of τ^2 in Example 1.

Example 2

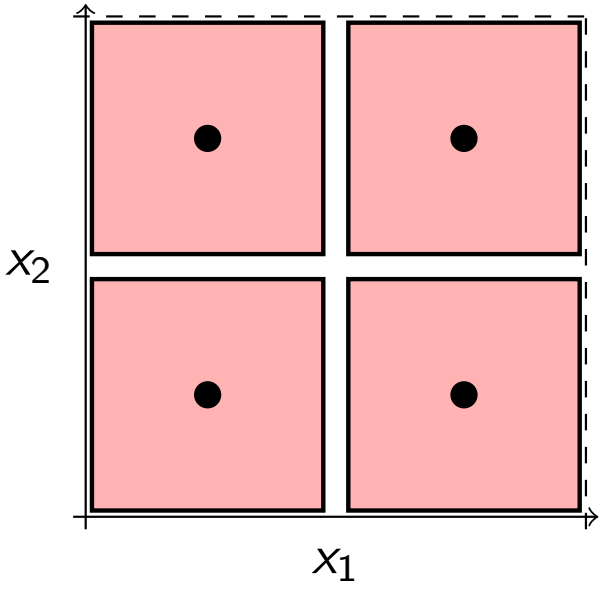
Two factors, approximately first-order model. Assume $n = 4$, $q = 2$, $\mathcal{X} = [-1, 1]^2$, $\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}) = (1, x_1, x_2)$.

Range of values for $\frac{\tau^2}{\sigma^2}$ tried.

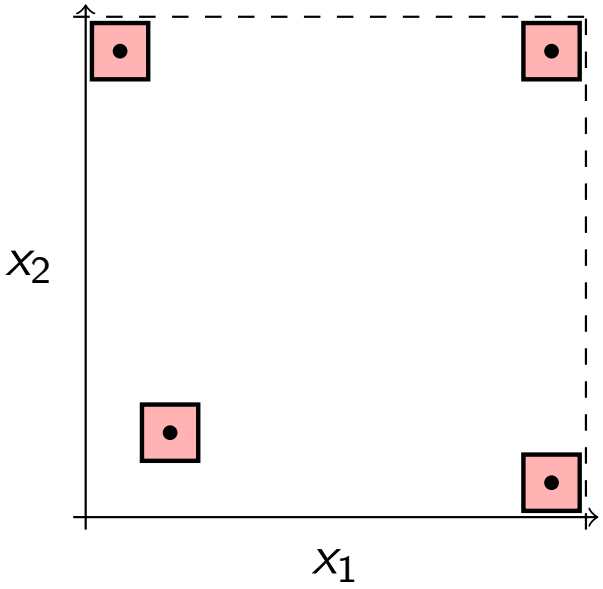
Simulated annealing algorithm used to performed constrained optimization of $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n, \delta$ simultaneously.

The constraints arise due to condition (ii) of the definition of random translation designs, i.e. the sets $\mathbf{c}_i + \mathcal{T}$ must be disjoint subsets of $\mathcal{X}$.

Constraints

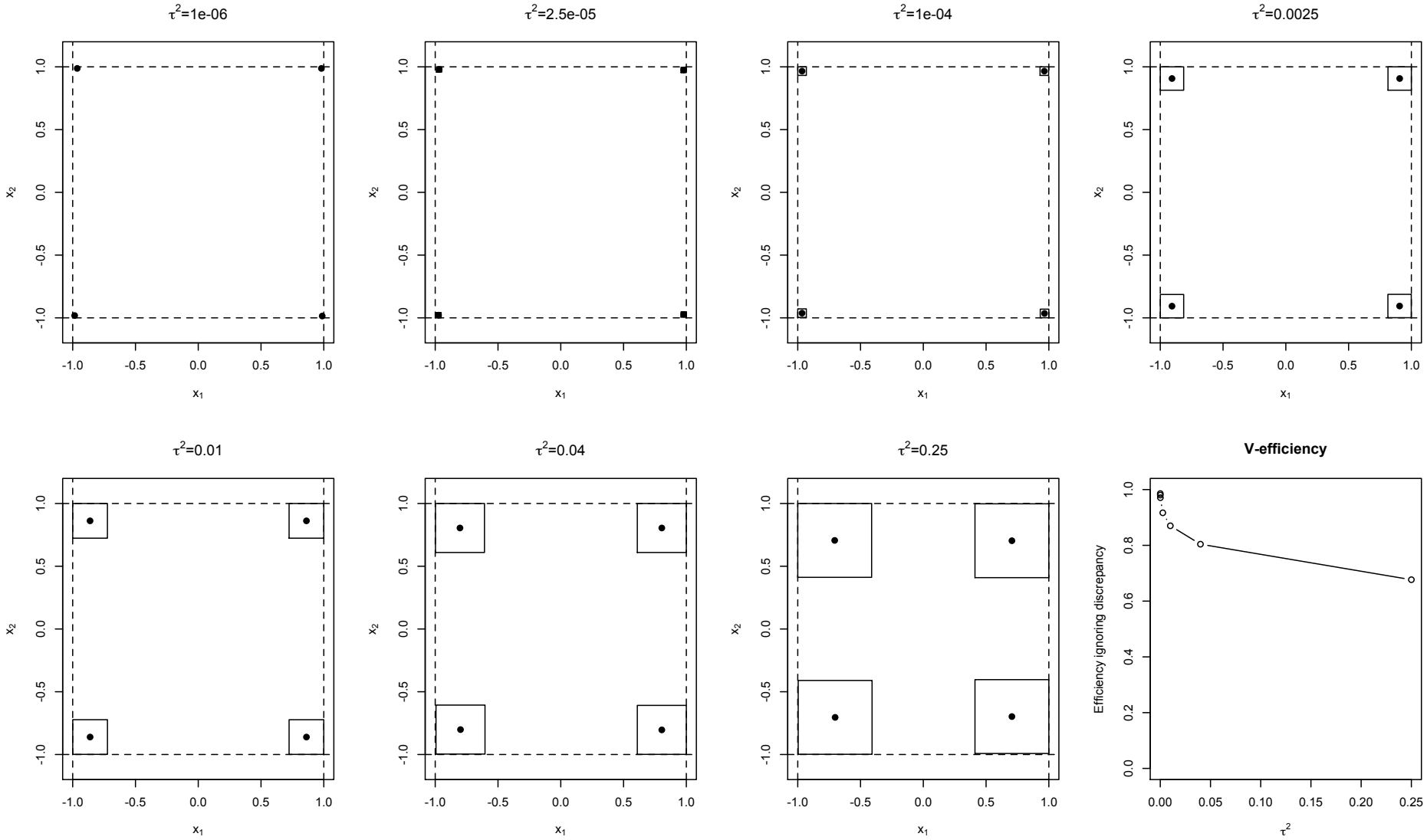


$\delta = 0.925$ - can not move centres very much



$\delta = 0.225$ - centres can be moved easily

Example 2 - results



Concluding remarks

- Random designs appear to have untapped potential to improve minimax efficiency in many problems.
- For model-robust design, random translation designs have several attractive properties:
  - They yield finitely-supported designs with finite and computable IMSEP.
  - We recover classic variance optimal designs as $\tau^2 \to 0$.

Related and future work

- Random designs for a wider range of design problems.
- More powerful optimization algorithms needed.

References

- Box, G. and Draper, N. (1959), *JASA*, **54**, 622–654
- Bhaumik, D. and Mathew, T. (1995), Sankhya B, **57**, 122-127
- Berger, J. (1985) *Statistical Decision Theory and Bayesian Analysis*, Springer
- Dette, H. and Wiens, D. (2009), *Stat. Sinica*, **19**, 83-102
- Hooper, P. (1989), *Ann. Stat.*, **17**, 1315–1324
- Huber, P. (1981) *Robust statistics*, Wiley
- Li, K. (1983) *Ann. Stat.*, **11**, 225-239
- Li, K. and Notz, W. (1982), *JSPI*, **6**, 135–151
- Plumlee, M. (2016), *JASA*, to appear
- Tuo, R. and Wu, C.-F. (2015), *Ann. Stat.*, **43**, 2331-2352
- Wiens, D. (2015), *Handbook of Design and Analysis of Experiments*, CRC. *Chapter 20*.
- Wu, C.-F. (1981), *Ann. Stat.*, **9**, 1168–1177
- Yue, R.-X. and Hickernell, F. (1999), *Stat. Sinica*, **9**, 1053–1069