The Equivalence Theorem in Optimal Design

Rainer Schwabe[1] & Thomas Schmelter[1,2]

[1]Otto von Guericke University Magdeburg

[2]Bayer Schering Pharma, Berlin

rschwabe@ovgu.de

Outline

Prologue: Simple examples

1. A shortcut to linear models
2. Designs and design criteria
3. Approximate designs
4. The Equivalence Theorem
5. RCR models

Epilogue: Open problems

Prologue: Simple examples

- Comparison of two groups

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

observation
$j=1,\ldots,n_i$

group
$i=1,2$

group
mean

random
error

ε_{ij} i.i.d.

$\mathrm{Var}(\varepsilon_{ij}) = \sigma^2$

$$\delta = \mu_1 - \mu_2 = ?$$

Comparison of two groups

estimated difference $\hat{\delta} = \bar{y}_1 - \bar{y}_2$

$$\boxed{\mathrm{Var}(\hat{\delta}) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2}$$

$n_1 + n_2 = n$ fix

optimal choice

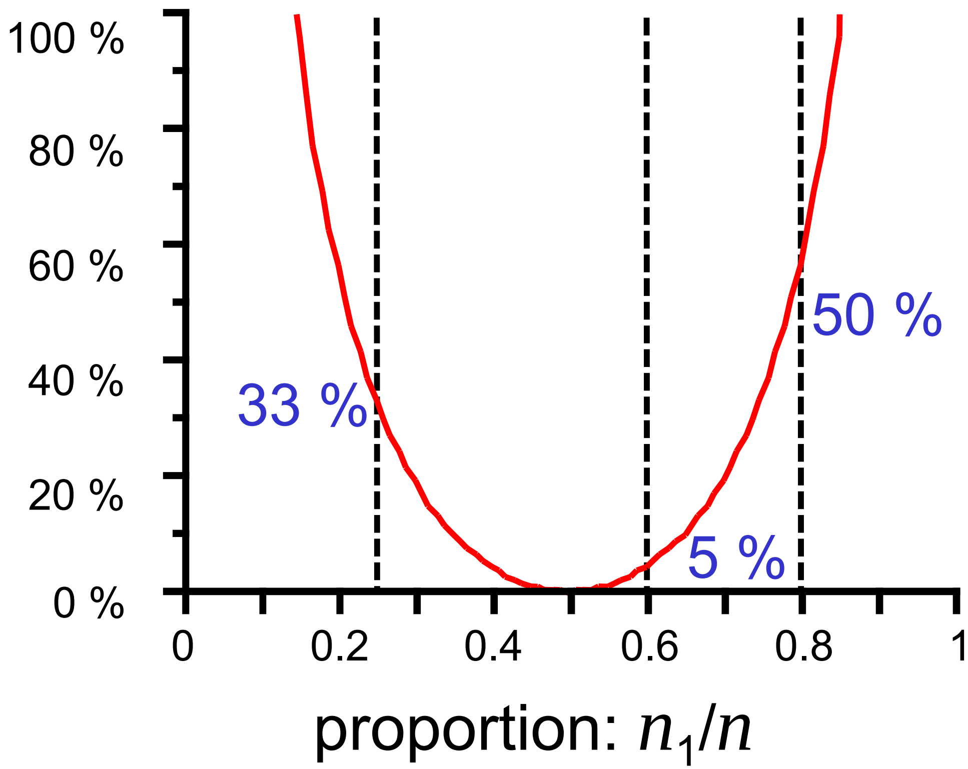
- n even: $n_1 = n_2 = n/2$
- n odd: $n_1 = (n \pm 1)/2$

Comparison of two groups

optimal choice $n_1 = n/2$

deficiency:

increase in sample size



Linear regression

$$Y_j(x_j) = \alpha + \beta \cdot x_j + \varepsilon_j$$

slope

random error

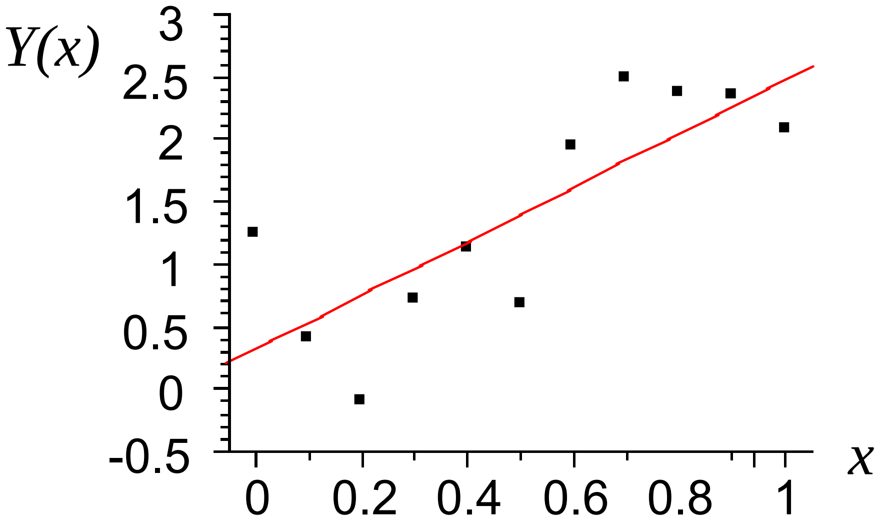
observation $j=1,\ldots,n_i$

intercept

explanatory variable

ε_j i.i.d.

$\mathrm{Var}(\varepsilon_j) = \sigma^2$



- <u>Linear regression</u>

  estimated slope $$\hat{\beta} = \frac{\sum (x_j - \bar{x})(y_j - \bar{y})}{\sum (x_j - \bar{x})^2}$$

$$\boxed{\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum (x_j - \bar{x})^2}}$$

design region: $0 \le x \le 1$

---

optimal choice

- n even: $x_j = 0$ or 1, each $n/2$ times

1. A shortcut to linear models

- general linear model

random error

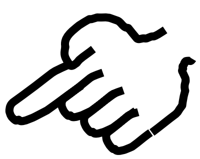
$$Y_j(x_j) = \mathbf{f}(x_j)^{\mathsf{T}} \boldsymbol{\beta} + \varepsilon_j$$

observation
j=1,...,n

explanatory variable

ε_j i.i.d.
$\mathrm{Var}(\varepsilon_j) = \sigma^2$

- regression functions $\mathbf{f} = (f_1, \ldots, f_p)$
- parameter $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$



Vector notation

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathrm{Cov}(\mathbf{Y}) =$$
$$\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$$

- “design” matrix ($n \times p$)

$$\mathbf{F} = \begin{pmatrix} \mathbf{f}(x_1)^\top \\ \vdots \\ \mathbf{f}(x_n)^\top \end{pmatrix}$$

Estimation

- <u>Gauss Markov Theorem</u>

least squares

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\mathbf{F}^{\mathsf{T}}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathsf{T}}\mathbf{Y}}$$

Is best linear unbiased esrimator for β

- covariance matrix

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2\left(\mathbf{F}^{\mathsf{T}}\mathbf{F}\right)^{-1}$$

- "information matrix"

≈ Fisher information

$$\mathbf{M} = \mathbf{F}^{\mathsf{T}}\mathbf{F} \quad = \sum_{j=1}^{n}\mathbf{f}(x_j)\mathbf{f}(x_j)^{\mathsf{T}}$$

- Linear regression

$$Y_j(x_j) = \beta_1 + \beta_2 \cdot x_j + \varepsilon_j$$

- dimension $p = 2$
- regression functions $f_1(x) \equiv 1,\ f_2(x) = x$

- information matrix (2×2)

$$\mathbf{F}^{\mathsf{T}}\mathbf{F} = \begin{pmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{pmatrix} = \sum \begin{pmatrix} 1 \\ x_j \end{pmatrix} \begin{pmatrix} 1 & x_j \end{pmatrix}$$

2. Designs and design criteria

- "exact design" $(x_1,\dots,x_n)$
- information matrix

$$\mathbf{M}(x_1,\dots,x_n)=\sum_{j=1}^{n}\mathbf{f}(x_j)\mathbf{f}(x_j)^{\mathsf{T}}$$

- aim: choose $x_1,\dots,x_m$ from design region X

to. minimise $\mathbf{M}^{-1}$ $\propto \mathrm{Cov}(\hat{\boldsymbol{\beta}})$

prediction variance

resp. $\mathrm{v}(x)=\mathbf{f}(x)^{\mathsf{T}}\mathbf{M}^{-1}\mathbf{f}(x)$

Design criteria

equivalently $-\ln\det\mathbf{M}$

- minimise

D: $\det \mathbf{M}^{-1}$

A: $\operatorname{trace} \mathbf{M}^{-1}$

IMSE: $\int \mathbf{f}(x)^{\mathsf{T}} \mathbf{M}^{-1} \mathbf{f}(x)\, dx$

G: $\max \mathbf{f}(x)^{\mathsf{T}} \mathbf{M}^{-1} \mathbf{f}(x)$

3. Approximate designs

$$\xi = \begin{pmatrix} x_1 & \cdots & x_m \\ \xi(x_1) & \cdots & \xi(x_m) \end{pmatrix} \begin{array}{l} \longleftarrow \text{design points} \\ \longleftarrow \text{weights} \end{array}$$

- information matrix

$$\mathbf{M}(\xi) = \sum \xi(x_j)\mathbf{f}(x_j)\mathbf{f}(x_j)^{\mathsf{T}}$$

- ξ^* D-optimal $\Leftrightarrow$ ξ^* minimises $\det \mathbf{M}(\xi)^{-1}$

etc. …

Embedding of exact designs

$$(x_1, \ldots, x_n) \to \xi = \begin{pmatrix} x_1 & \cdots & x_n \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

- information matrix

$$\mathbf{M}(\xi) = \frac{1}{n} \sum \mathbf{f}(x_j) \mathbf{f}(x_j)^\mathsf{T}$$

$$= \frac{1}{n} \mathbf{M}$$

standardised

Convexity

- set Ξ of approximate designs is convex

  $\alpha \cdot \xi_1 + (1-\alpha) \cdot \xi_2$ is a design

- standard criteria $\Phi : \Xi \to (-\infty, \infty]$ are convex

$$\Phi(\alpha \cdot \xi_1 + (1-\alpha) \cdot \xi_2) \leq \alpha \cdot \Phi(\xi_1) + (1-\alpha) \cdot \Phi(\xi_2)$$

D-criterion: $\Phi(\xi) = -\ln \det \mathbf{M}(\xi)$

4. The Equivalence Theorem

- directional derivative

$$F_{\Phi}(\xi;\eta) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha}[\Phi((1-\alpha)\cdot\xi + \alpha\cdot\eta) - \Phi(\xi)]$$

at ξ in the direction of η

- note: $F_{\Phi}(\xi;\xi) = 0$

- regularity condition:

$\{\mathbf{f}(x); x \in X\}$ is compact

The General Equivalence Theorem

ξ^* minimises $\Phi(\xi)$ (ξ^* Φ-opt.)

if and only if

$F_\Phi(\xi^*;\eta) \geq 0$ for **all** η

$\Leftrightarrow$

ξ^* maximises $\min\limits_{\eta} F_\Phi(\xi^*;\eta)$

Sensitivity

- Φ diffferentiable

$$F_\Phi(\xi;\eta) = \sum_x \eta(x) F_\Phi(\xi;\delta_x) \qquad \delta_x = \begin{pmatrix} x \\ 1 \end{pmatrix}$$

- sensitivity function $\qquad \varphi(x;\xi) = -F_\Phi(\xi;\delta_x)$

ξ^* minimises $\Phi(\xi)$

$\Leftrightarrow \qquad \varphi(x;\xi^*) \le 0 \qquad$ for **all** x

$\Leftrightarrow \qquad \xi^*$ minimises $\max_x \varphi(x;\xi^*)$

Sensitivity

- sensitivity function $\varphi(x;\xi) = -F_{\Phi}(\xi;\delta_x)$

ξ^* minimises $\Phi(\xi)$

$\Leftrightarrow$ $\varphi(x;\xi^*) \le 0$ for **all** x

$\Leftrightarrow$ ξ^* minimises $\max_x \varphi(x;\xi^*)$

- note: $\varphi(x;\xi^*) = 0$ for $\xi^*(x) > 0$

D-criterion

$$\Phi(\xi) = -\ln\det \mathbf{M}(\xi)$$

$$\mathrm{F}_\Phi(\xi;\eta) = p - \mathrm{tr}(\mathbf{M}(\eta)\mathbf{M}(\xi)^{-1})$$

$$\varphi(x;\xi) = \mathbf{f}(x)^\mathsf{T}\mathbf{M}(\xi)^{-1}\mathbf{f}(x) - p$$

Kiefer, Wolfowitz (1960)

ξ^* D-optimal $\Leftrightarrow$

$\mathrm{v}(x;\xi^*) = \mathbf{f}(x)^\mathsf{T}\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x) \le p$ for **all** x

$\Leftrightarrow$ ξ^* G-optimal

Linear criteria

$$\Phi(\xi) = \mathrm{tr}(\mathbf{A}\mathbf{M}(\xi)^{-1})$$

$$F_{\Phi}(\xi;\eta) = \mathrm{tr}(\mathbf{A}\mathbf{M}(\xi)^{-1}) - \mathrm{tr}(\mathbf{A}\mathbf{M}(\xi)^{-1}\mathbf{M}(\eta)\mathbf{M}(\xi)^{-1})$$

$$\varphi(x;\xi) = \mathbf{f}(x)^{\mathsf{T}}\mathbf{M}(\xi)^{-1}\mathbf{A}\mathbf{M}(\xi)^{-1}\mathbf{f}(x) - \mathrm{tr}(\mathbf{A}\mathbf{M}(\xi)^{-1})$$

Fedorov (1971): $\mathbf{A}=\mathbf{I}$

ξ^* A-optimal $\Leftrightarrow$

$$\mathbf{f}(x)^{\mathsf{T}}\mathbf{M}(\xi^*)^{-2}\mathbf{f}(x) \leq \mathrm{tr}(\mathbf{M}(\xi^*)^{-1}) \quad \text{for } \textbf{all } x$$

ξ^* c-optimal $\Leftrightarrow$ $(\mathbf{f}(x)^{\mathsf{T}}\mathbf{M}(\xi)^{-1}\mathbf{c})^2 \leq \mathbf{c}^{\mathsf{T}}\mathbf{M}(\xi)^{-1}\mathbf{c}$

Pros and cons

- usually not constructive
  exception:polynomial regression
- efficiency bounds
- algorithms

Polynomial regression

$$\boxed{Y(x) = \beta_1 + \beta_2 x +,,, + \beta_p x^{p-1} + \varepsilon} \quad 0 \le x \le 1$$

$\varphi(x;\xi)$ polynomial of degree 2(p-1)

- minimal support

$$\xi^* = \begin{pmatrix} 0 & x_2 & \dots & x_{p-1} & 1 \\ \xi^*(0) & \xi^*(x_2) & \dots & \xi^*(x_{p-1}) & \xi^*(1) \end{pmatrix}$$

D-optimal: $\xi^*(x_j) = 1/p$

Efficiency bounds

$$F_{\Phi}(\xi;\eta) \le \Phi(\eta) - \Phi(\xi)$$ convexity

$$\Phi(\xi) \le \Phi(\xi^*) + \max_{x\in X} \varphi(x;\xi)$$

- *D*-criterion

$$\mathrm{eff}_{\mathrm{D}}(\xi) = \left(\frac{\det \mathbf{M}(\xi^*)^{-1}}{\det \mathbf{M}(\xi)^{-1}}\right)^{\frac{1}{p}} \ge e^{1-\frac{1}{p}\max_x \mathrm{v}(x;\xi)}$$

- *A*-criterion

$$\mathrm{eff}_{\mathrm{A}}(\xi) = \frac{\Phi(\xi^*)}{\Phi(\xi)} \ge 1 - \frac{\max_x \mathbf{f}(x)^{\mathrm{T}} \mathbf{M}(\xi)^{-2} \mathbf{f}(x)}{\mathrm{tr}(\mathbf{M}(\xi)^{-1})}$$

Algorithms

"steepest descent"

$$x_n = \arg\max_{x \in X} \varphi(x; \xi_{n-1})$$

$$\xi_n = (1 - \alpha_n)\xi_{n-1} + \alpha_n \delta_{x_n}$$

- "Fedorov"

"add one point"

$$\alpha_n = \frac{1}{n}$$

- "Wynn"

$$\alpha_n = \arg\min_{x \in X} \Phi((1-\alpha)\xi_{n-1} + \alpha\delta_{x_n})$$

$$\xi_n \to \xi^*$$

σ^2 unknown

- information matrix (for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$)

$$\mathbf{M}_{\boldsymbol{\theta}}(\xi) = \begin{pmatrix} \frac{1}{\sigma^2}\mathbf{M}(\xi) & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

$$\mathbf{M}_{\boldsymbol{\theta}}(\xi)^{-1} = \begin{pmatrix} \sigma^2\mathbf{M}(\xi)^{-1} & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$$

constant

$\rightarrow$ same story as before

5. Random Coefficient Regression

- individual curves are given by a common linear model

$$\boxed{Y_{ij} = \mathbf{f}(x_{ij})^{\mathsf{T}}\mathbf{b}_i + \varepsilon_{ij}}$$

individual
$i=1,\ldots,n$

replication
$j=1,\ldots,m_i$

explanatory variable

error

$\varepsilon_{ij} \sim N(0,\sigma^2)$

independent

- individual parameters: $\mathbf{b}_i \sim N(\boldsymbol{\beta}, \sigma^2\mathbf{D})$

population parameter

Design

- individual design

$$\xi_i = \begin{pmatrix} x_{i1} & \cdots & x_{i\mu} \\ \xi(x_{i1}) & \cdots & \xi(x_{i\mu}) \end{pmatrix}$$

- population design

$$\zeta = \begin{pmatrix} \xi_{i1} & \cdots & \xi_{i\nu} \\ \zeta(\xi_{i1}) & \cdots & \zeta(\xi_{i\nu}) \end{pmatrix}$$

- problem: set of information matrices is not convex
- $m_i \equiv m$

Single group designs

$$\zeta = \begin{pmatrix} \xi \\ 1 \end{pmatrix}$$

- information for population parameter

$$\mathbf{M}_{\boldsymbol{\beta}}(\xi) = \left(\frac{1}{m}\mathbf{M}(\xi)^{-1} + \mathbf{D}\right)^{-1}$$

fixed model information

Linear criteria

A, IMSE, c

- minimise

$$\boxed{\mathrm{tr}\left(\mathbf{A}\left(\frac{1}{m}\mathbf{M}(\xi)^{-1}+\mathbf{D}\right)\right)}$$

$$=\frac{1}{m}\mathrm{tr}\left(\mathbf{AM}(\xi)^{-1}\right)+\overbrace{\mathrm{tr}(\mathbf{AD})}^{\text{constant !}}$$

- result

Luoma (2000), Liski et al. (2002)

$$\boxed{\begin{array}{ll} & \xi \quad \text{optimal in reduced model} \\ \Rightarrow & \\ & \xi \quad \text{optimal in RCR model} \end{array}}$$

An Equivalence Theorem

- convexity

  Φ convex

  $\Rightarrow \Phi$ convex in $\mathbf{M}_{\boldsymbol{\beta}}(\xi) = (\frac{1}{m}\mathbf{M}(\xi)^{-1} + \mathbf{D})^{-1}$

ξ^* D-optimal for β $\Leftrightarrow$

$$\mathbf{f}(x)^{\mathsf{T}}\mathbf{M}(\xi^*)^{-1}\mathbf{M}_{\boldsymbol{\beta}}(\xi^*)\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x)$$

$$\leq \mathrm{tr}(\mathbf{M}_{\boldsymbol{\beta}}(\xi^*)\mathbf{M}(\xi^*)^{-1}) \quad \text{for } \textbf{all } x$$

Fedorov, Hackl (1997)

locally optimal at $\mathbf{D}$

Variance components

- convexity fails (within individuals)
  even for single parameters (e.g. d_{11})

- "equivalence theorems" (which?)
  give only necessary conditions

A Multivariate Equivalence Theorem

- generalised multivariate designs

$$\tilde{\xi} = \begin{pmatrix} (x_{11},\ldots,x_{1m}) & \cdots & (x_{l1},\ldots,x_{lm}) \\ \tilde{w}_1 & \cdots & \tilde{w}_l \end{pmatrix} \begin{array}{l} \longleftarrow \text{settings} \\ \longleftarrow \text{weights} \end{array}$$

- generalised multivariate information

$$\tilde{\mathbf{M}}(\tilde{\xi}) = \sum_{i=1}^{l} \tilde{w}_i \mathbf{F}_i (\mathbf{I}_m + \mathbf{F}_i^{\mathsf{T}} \mathbf{D} \mathbf{F}_i)^{-1} \mathbf{F}_i^{\mathsf{T}}$$

- $\tilde{\xi}^*$ maximises $\det \tilde{\mathbf{M}}(\tilde{\xi})$ ($\tilde{\xi}^*$ D-opt.)

if and only if

$$\boxed{\sup_{\mathbf{x}=(x_1,\ldots,x_m)} \operatorname{tr} \mathbf{V}(\mathbf{x})^{-1} \mathbf{F}(\mathbf{x}) \tilde{\mathbf{M}}(\tilde{\xi}^*)^{-1} \mathbf{F}(\mathbf{x})^{\mathsf{T}} \le p}$$

Fedorov (1972)

Epilogue: Open Problems

- unbalanced designs
  (e.g. individual designs singular)
- single group designs not applicable for treatment comparisons
- non-linear models
  (linear approximation: appropriate?)

Need for a more general approach !!!