A design criterion for symmetric model discrimination based on nominal confidence sets

Werner G. Müller
joint work with Radoslav Harman

Department of Applied Statistics, Johannes Kepler University of Linz, Austria

Biometrisches Kolloquium Wien, 9th April 2018

Setup

discrimination between a pair of (non-)linear regression models

$$y_i = \eta_0(\theta_0, x_i) + \varepsilon_i,\ i = 1, ..., n,$$
$$y_i = \eta_1(\theta_1, x_i) + \varepsilon_i,\ i = 1, ..., n,$$

with a finite design space $\mathfrak{X}$ and a design $\mathcal{D} = (x_1, ..., x_n)$ on $\mathfrak{X}$, independent, zero-mean and normal errors with the same variance $\sigma^2 \in (0, \infty)$ for all observations and both models.

We'll be mainly talking about **exact designs** today, but let us also introduce ξ as a probability measure on $\mathfrak{X}$ derived from $\mathcal{D}$ as

$$\xi(x) = n^{-1} \# \{i \in \{1, ..., n\} : x_i = x\}.$$

Brief review of discrimination designs

- Early ad-hoc approaches (Hunter & Reiner, 1965, Box & Hill, 1967) are reviewed in Hill (1978).
- Big step: T-optimality by Atkinson & Fedorov (1975):

$$T(\xi) = \inf_{\theta_1} \int_{\mathfrak{X}} (\eta_0(x) - \eta_1(\theta_1, x))^2 \, d\xi(x),$$

  where η_0 is assumed to be true and fixed (**asymmetry!**)
- For nested models T-optimal designs maximize the power of the LR-test against local alternatives.
- For nested linear models differing by one parameter T- and D_s-optimality (Stigler, 1971) coincide.
- Dette & Titoff (2009) conceive T-optimality as a general nonlinear approximation problem and reveal further connections between T- and D_s-optimality in the partially nonlinear case.
- Extension to nonnormal errors using KL-distance by López-Fidalgo et al. (2007).

Nonnested models

- We cannot constrain one of the models such that the parameter spaces coincide (Cox, 1961).
- An asymmetric design criterion is thus **not appropriate**.
- A natural decision rule is then whether likelihood ratio

$$L(\hat{\theta}_0)/L(\hat{\theta}_1) >< 1 \qquad (\pi_1/\pi_0 \text{ for Bayesians}).$$

- We will for this talk assume $m := m_0 = m_1$, but for $m_0 \neq m_1$ Cox (2013) recommends $L(\hat{\theta}_0)/L(\hat{\theta}_1)(e^{m_1}/e^{m_0})^{n/\tilde{n}}$ instead, which corresponds to the BIC.
- In the normal model the probability of a correct decision is then

$$P\left[\min_{\theta_0 \in \Theta_0} \sum_{i=1}^{n} (\eta_0(\theta_0, x_i) - y_i))^2 \leq \min_{\theta_1 \in \Theta_1} \sum_{i=1}^{n} (\eta_1(\theta_1, x_i) - y_i))^2\right].$$

“Symmetric” discrimination criteria

- “symmetrized” T-optimality:

$$T(\xi) = \inf_{\theta_0,\theta_1} \int_{\mathfrak{X}} (0 - \{\eta_0(\theta_0, x) - \eta_1(\theta_1, x)\})^2 \, d\xi(x),$$

- “weighted” T-optimality (Atkinson, 2008): maximize

$$Eff_{T_0}^{1-\kappa} Eff_{T_1}^{\kappa}$$

- D_s-optimality in an encompassing model (Atkinson, 1972):

$$\eta_2(\eta_0(\theta_0, x), \eta_1(\theta_1, x), \lambda)$$

- An algorithmic construction switching between assuming true η_0 and η_1 (Vajjah & Dufful, 2012).
- Bayesian approaches, eg. Felsenstein (1992), Nowak & Guthke (2016).
- Sequental design: Buzzi-Ferraris & Forzatti (1983), M.& Ponce de Leon (1996), Schwaab et al. (2006).
- Constrained “T-optimality”: Fedorov & Khabarov, 1986

A motivating example

Let $\eta_0(\theta_0, x) = \theta_0 x$ and $\eta_1(\theta_1, x) = e^{\theta_1 x}$. Two observations y_1, y_2 at fixed design points $x_1 = -1$ and $x_2 = 1$. Then $\hat{\theta}_0 = \frac{y_2 - y_1}{2}$ and $\hat{\theta}_1$ is the solution of $2e^{-\theta}\left(y_1 - e^{-\theta}\right) - 2e^{\theta}\left(y_2 - e^{\theta}\right) = 0$, which for $-2 \leq y_1 \leq 2$ is the log root of the polynomial $\theta^4 - \theta^3 y_2 + \theta y_1 - 1$.

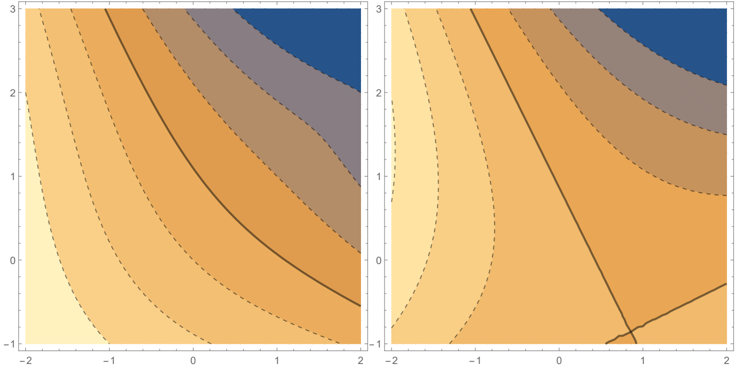


Figure: left panel: contour plot of $\log L(\hat{\theta}_0) - \log L(\hat{\theta}_1)$, solid line corresponds to 0; right panel: corresponding contour plot for the model η_1 linearized at $\theta_1 = 1$.

Nominal confidence sets

Traditional "localized" approach to the design of non-linear models (Chernoff 1953) requires a pair of "nominal parameter values":

$$\tilde{\theta}_0 \in \Theta_0, \;\; \tilde{\theta}_1 \in \Theta_1.$$

We extend this notion to "nominal confidence sets"

$$\tilde{\Theta}_0 \subseteq \Theta_0, \;\; \tilde{\Theta}_1 \subseteq \Theta_1,$$

such that $\tilde{\theta}_0 \in \tilde{\Theta}_0$ and $\tilde{\theta}_1 \in \tilde{\Theta}_1$.

We assume: if Model k is the correct one, $\tilde{\Theta}_k$ contains the true value $\bar{\theta}_k$ with a high degree of certainty (for both $k = 0, 1$).

Nominal confidence sets allow us to

1. modify the decision rule: maximize the likelihood functions over $\tilde{\Theta}_0$, $\tilde{\Theta}_1$;
2. use a "restricted" linear approximation of Models 0 and 1.

Linearisation over nominal confidence sets

Let $\mathcal{D} = (x_1, ..., x_n) \in \mathcal{D}_n$ (the set of permissible n-point exact designs on $\mathfrak{X}$). Let us perform the following linearisation of Models $k = 0, 1$ in $\tilde{\theta}_k$:

$$(y_i)_{i=1}^n \approx \mathbf{F}_k(\mathcal{D})\theta_k + \mathbf{a}_k(\mathcal{D}) + \varepsilon,$$

where $\mathbf{F}_k(\mathcal{D})$ in an $n \times m$ matrix given by

$$\mathbf{F}_k(\mathcal{D}) = \left(\nabla\eta_k(\tilde{\theta}_k, x_1), \ldots, \nabla\eta_k(\tilde{\theta}_k, x_n)\right)^T,$$

$\mathbf{a}_k(\mathcal{D})$ is an n-dimensional vector

$$\mathbf{a}_k(\mathcal{D}) = (\eta_k(\tilde{\theta}_k, x_i))_{i=1}^n - \mathbf{F}_k(\mathcal{D})\tilde{\theta}_k,$$

$\theta_k \in \tilde{\Theta}_k$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T \sim N_n(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$.

Note: here we do not subtract $\mathbf{a}_k(\mathcal{D})$ from the vector of observations, which is usual if we linearise a single non-linear regression model. (However, if η_k corresponds to the standard linear model then $\mathbf{a}_k(\mathcal{D}) = 0$ for any $\mathcal{D}$.)

The δ-criterion and exact δ-optimal designs

Consider the following criterion on the set of all exact designs $\mathcal{D} \in \mathcal{D}_n$:

$$\begin{aligned}\delta(\mathcal{D}) &= \inf_{\theta_0 \in \tilde{\Theta}_0, \theta_1 \in \tilde{\Theta}_1} \delta(\mathcal{D}|\theta_0, \theta_1), \text{ where} \\ \delta^2(\mathcal{D}|\theta_0, \theta_1) &= \left\| \mathbf{F}_0(\mathcal{D})\theta_0 + \mathbf{a}_0(\mathcal{D}) - \{\mathbf{F}_1(\mathcal{D})\theta_1 + \mathbf{a}_1(\mathcal{D})\} \right\|^2 .\end{aligned}$$

The criterion δ can be viewed as an approximation of the square of the nearest distance of the mean-value surfaces of the models, in the neighbourhoods of the vectors $(\eta_0(\tilde{\theta}_0, x_i))_{i=1}^n$ and $(\eta_1(\tilde{\theta}_1, x_i))_{i=1}^n$.

Value of $\delta(\mathcal{D})$ is always well defined, and if $\tilde{\Theta}_0, \tilde{\Theta}_1$ are both compact (or if $\tilde{\Theta}_0 = \tilde{\Theta}_1 = \mathbb{R}^m$), the infimum is attained.

The design $\mathcal{D}^* \in \mathcal{D}_n$ maximizing $\delta(\mathcal{D})$ will be called δ-optimal:

$$\mathcal{D}^* \in \operatorname{argmax}_{\mathcal{D} \in \mathcal{D}_n} \delta(\mathcal{D}).$$

Note that $\mathcal{D}^*$ depends on n, $\tilde{\theta}_0$, $\tilde{\theta}_1$, as well as on $\tilde{\Theta}_0$ and $\tilde{\Theta}_1$.

The δ-criterion and exact δ-optimal designs

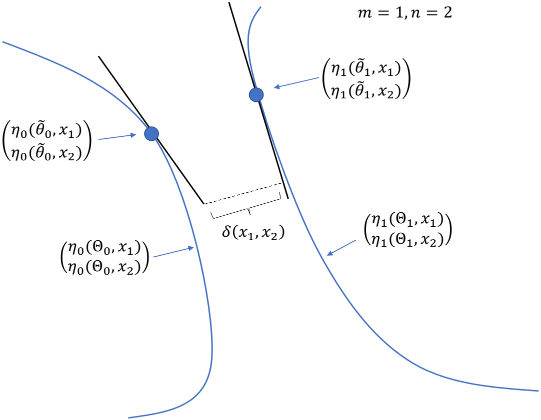


Figure: Illustrative graph for the definition of $\delta(\mathcal{D})$ for a one-parametric model ($\Theta_0, \Theta_1 \subset \mathbb{R}$) and a two-point design ($\mathcal{D} = (x_1, x_2)$). The line segments correspond to the sets $\{\mathbf{F}_0(\mathcal{D})\theta_0 + \mathbf{a}_0(\mathcal{D}) : \theta_0 \in \tilde{\Theta}_0\}$ and $\{\mathbf{F}_1(\mathcal{D})\theta_1 + \mathbf{a}_1(\mathcal{D}) : \theta_1 \in \tilde{\Theta}_1\}$ for some nominal confidence sets $\tilde{\Theta}_0 \subseteq \Theta_0$ and $\tilde{\Theta}_1 \subseteq \Theta_1$.

Computation of the δ-criterion

For a fixed design $\mathcal{D}$, $\delta^2(\mathcal{D}|\theta)$ is a quadratic and convex function of $\theta = (\theta_0^T, \theta_1^T)^T$.
Thus we can view the computation of $\delta(\mathcal{D}|\theta)$ as follows. Since

$$\delta^2(\mathcal{D}|\theta_0, \theta_1) = \left\| \{\mathbf{a}_0(\mathcal{D}) - \mathbf{a}_1(\mathcal{D})\} - [-\mathbf{F}_0(\mathcal{D}), \mathbf{F}_1(\mathcal{D})]\theta \right\|^2,$$

its minimization is equivalent to computing the minimum sum of squares for a least squares estimate of θ restricted to $\tilde{\Theta} := \tilde{\Theta}_0 \times \tilde{\Theta}_1$ in the response difference model with artificial observations

$$\tilde{z}_i = \{\mathbf{a}_0(\mathcal{D}) - \mathbf{a}_1(\mathcal{D})\}_i, \; i = 1, \ldots, n.$$

If the nominal confidence sets are compact cuboids, this can be evaluated by the very rapid and stable method by Stark & Parker, 1995.

Parametrization of nominal confidence sets

For simplicity, we will focus on cuboid nominal confidence sets centered at the nominal parameter values. Specifically, we will employ the homogeneous dilations

$$\tilde{\Theta}_k^{(r)} := r\left(\tilde{\Theta}_k^{(1)} - \tilde{\theta}_k\right) + \tilde{\theta}_k, \qquad r \in [0,\infty),\ k = 0,1,$$

$\tilde{\Theta}_0^{(\infty)} := \mathbb{R}^m$, $\tilde{\Theta}_1^{(\infty)} := \mathbb{R}^m$, such that r can be considered a tuning parameter governing the size of the nominal confidence sets.

For any design $\mathcal{D}$ and $r \in [0,\infty]$, we can now define

$$\delta_r(\mathcal{D}) := \inf_{\theta_0 \in \tilde{\Theta}_0^{(r)}, \theta_1 \in \tilde{\Theta}_1^{(r)}} \delta(\mathcal{D}|\theta_0,\theta_1).$$

There exists a finite interval $[0, r^*]$ beyond which the set of δ_r-optimal designs are unchanged, which can be easily determined.

The motivating example continued

Consider the two models from the previous example, both with $m = 1$ parameter and mean value functions

$$\eta_0(\theta_0, x) = \theta_0 x, \quad \eta_1(\theta_1, x) = e^{\theta_1 x},$$

where $x \in \mathfrak{X} = \{1.00, 1.01, 1.02, \ldots, 2.00\}$. Let

$$\tilde{\theta}_0 = e^{-1}, \quad \tilde{\theta}_1 = 1.$$

We used an adaptation of the KL exchange heuristic to compute δ-optimal designs by selecting

$$\tilde{\Theta}_0 = [e^{-1} - r, e^{-1} + r], \;\; \tilde{\Theta}_1 = [1 - r, 1 + r]$$

for $r = 0.01, 0.1, 0.2, \ldots, 1.0$. The size is $n = 6$ trials.

Note: If the $\tilde{\Theta}$'s are very narrow, the δ-optimal design is concentrated in the design point $x = 2$ maximizing the difference between $\eta_0(\tilde{\theta}_0, x)$ and $\eta_1(\tilde{\theta}_1, x)$. For large values of r, the δ-optimal design has a 3-point support.

The motivating example continued

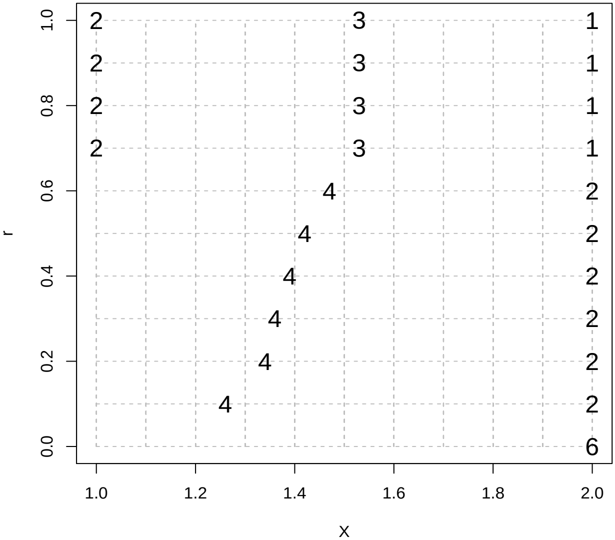


Figure: δ-optimal exact designs of size $n = 6$ for different r's in Illustrative Example.

An application in enzyme kinetics

Taken from Bogacka et al. (2011) and used in Atkinson (2012) to illustrate model discrimination designs. Two types of enzyme kinetic reactions, where the velocity y is modeled by competitive and noncompetitive inhibition:

$$y = \frac{\theta_{01} x_1}{\theta_{02}\left(1 + \frac{x_2}{\theta_{03}}\right) + x_1},$$

and respectively

$$y = \frac{\theta_{11} x_1}{(\theta_{12} + x_1)\left(1 + \frac{x_2}{\theta_{13}}\right)}.$$

	estimate	st.err.			estimate	st.err.
θ_{01}	7.298	0.114		θ_{11}	8.696	0.222
θ_{02}	4.386	0.233		θ_{12}	8.066	0.488
θ_{03}	2.582	0.145		θ_{13}	12.057	0.671

Table: Parameter estimates and corresponding standard errors from data on Dextrometorphan-Sertraline provided by B.Bogacka.

An encompassing model

Atkinson (2012) combines those models into

$$y = \frac{\theta_{21} x_1}{\theta_{22}\left(1 + \frac{x_2}{\theta_{23}}\right) + x_1\left(1 + \frac{(1-\lambda)x_2}{\theta_{23}}\right)},$$

with nominal values $\tilde{\theta}_{21} = 10$, $\tilde{\theta}_{22} = 4.36$, $\tilde{\theta}_{23} = 2.58$, and $\tilde{\lambda} = 0.8$.

	estimate	st.err.
θ_{21}	7.425	0.130
θ_{22}	4.681	0.272
θ_{23}	3.058	0.281
λ	0.964	0.019

Table: Parameter estimates and standard errors for the encompassing model.

Anthony's designs

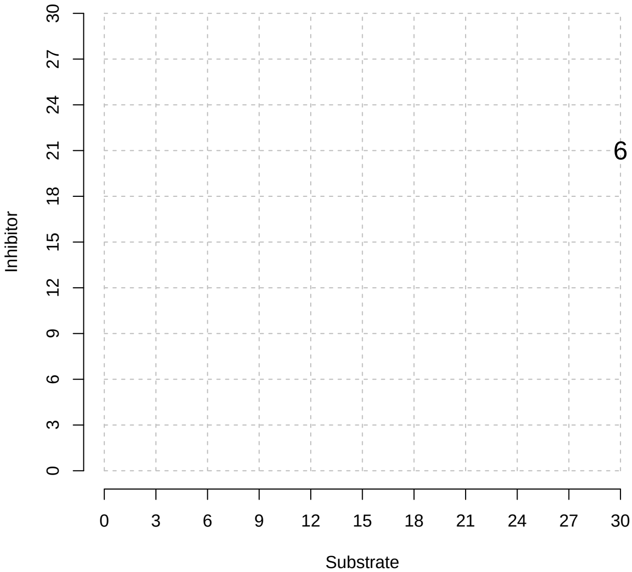
Experimental Designs for Enzyme Inhibition 2295

Table 2
Some T-optimum, compound T-optimum ($\kappa = 0.5$) and Ds-optimum ($\lambda = 0.8$) designs and their T-efficiencies

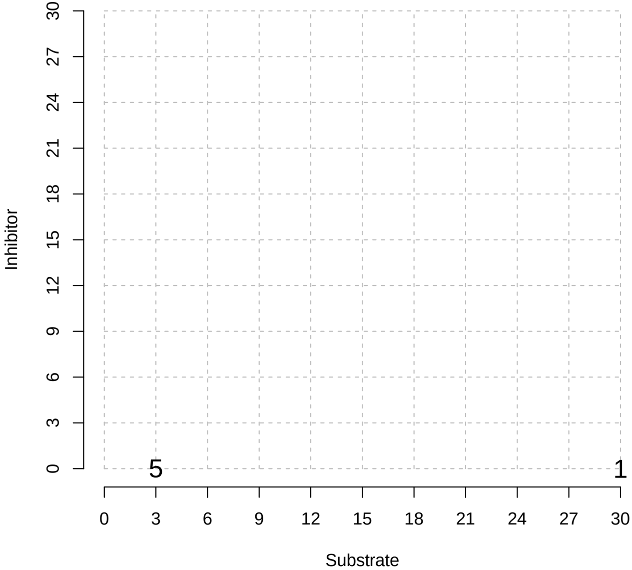
				Efficiencies (%) for	
Design	$[S]$	$[I]$	w	$\kappa = 0$	$\kappa = 1$
$\kappa = 0$	30.000	0.000	0.067	100	13.47
	1.828	0.000	0.046		
	30.000	10.154	0.337		
	4.107	4.153	0.550		
$\kappa = 0.5$	30.000	0.000	0.056	73.45	83.84
	3.269	0.000	0.216		
	30.000	13.281	0.234		
	4.815	6.934	0.494		
$\lambda = 0.8$	30.000	0.000	0.082	70.41	79.12
	2.484	0.000	0.204		
	30.000	14.492	0.266		
	4.666	7.103	0.448		
$\kappa = 1$	30.000	0.000	0.059	44.61	100
	3.072	0.000	0.250		
	30.000	22.613	0.250		
	5.453	11.614	0.441		

Figure: Four designs from Atkinson (2012); suitably rounded exact designs referred to as A1-A4 from top to bottom.

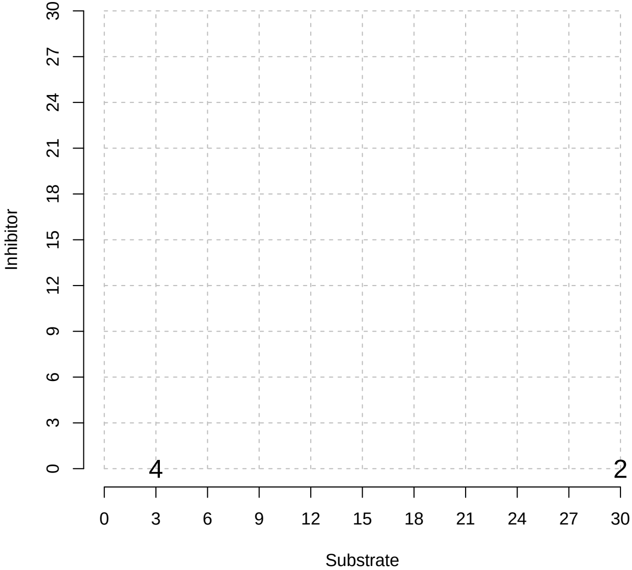
Confirmatory experiment $n = 6, r = [0.01, 0.16]$



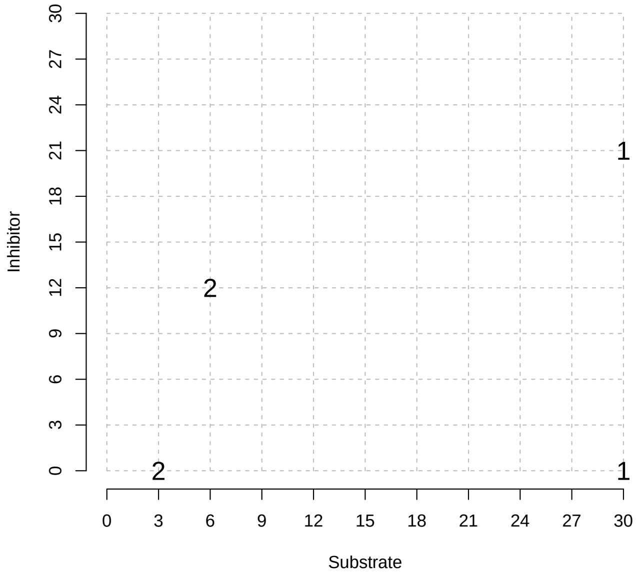
Confirmatory experiment $n = 6, r = [0.17, 0.23]$



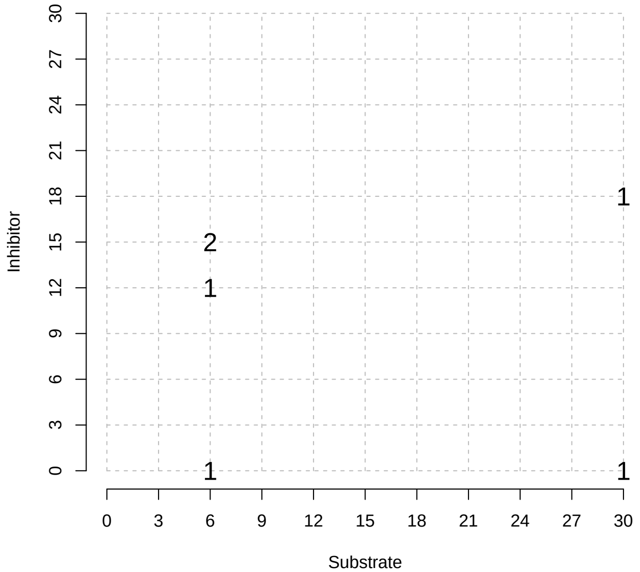
Confirmatory experiment $n = 6, r = [0.24, 0.49]$



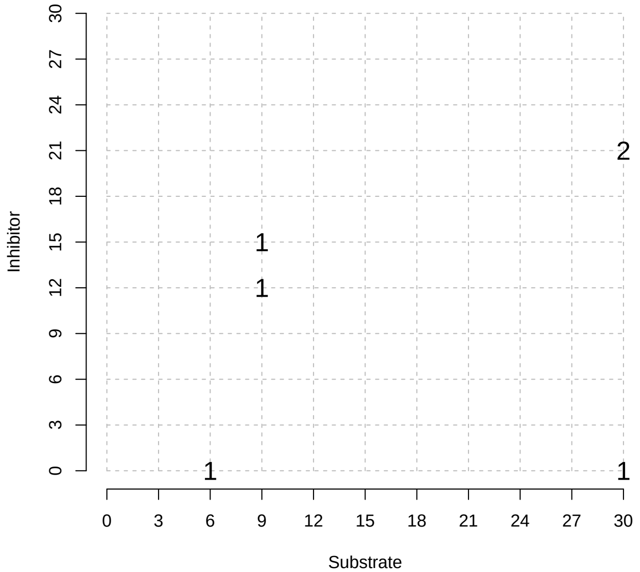
Confirmatory experiment $n = 6, r = [0.50, 4.56]$



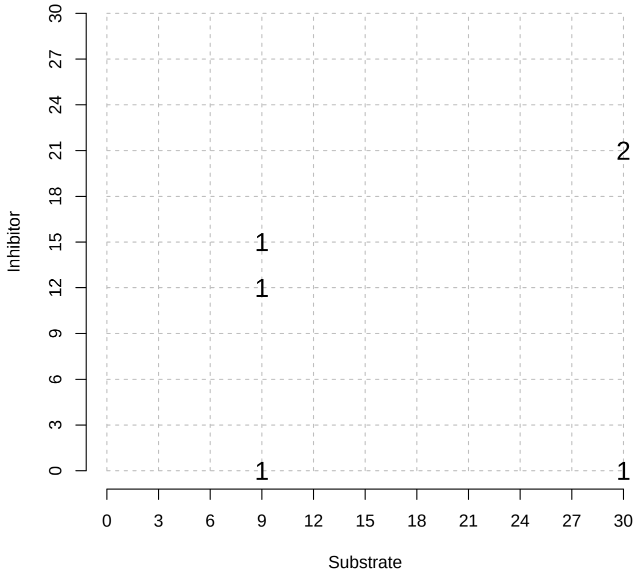
Confirmatory experiment $n = 6, r = [4.57, 9.89]$



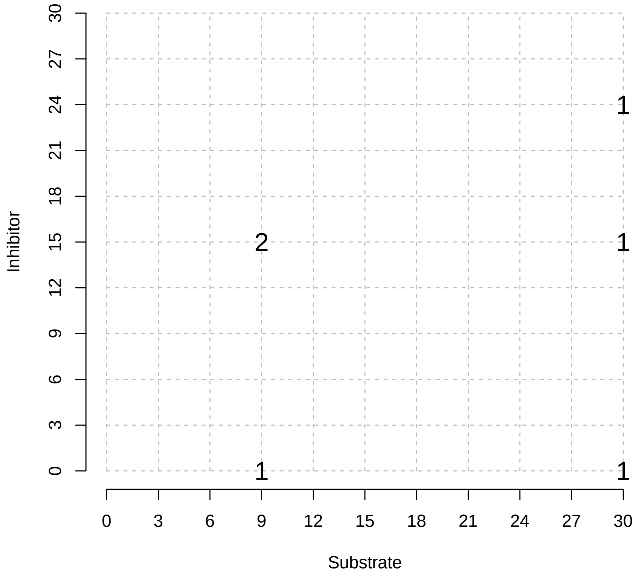
Confirmatory experiment $n = 6, r = [9.90, 11.94]$



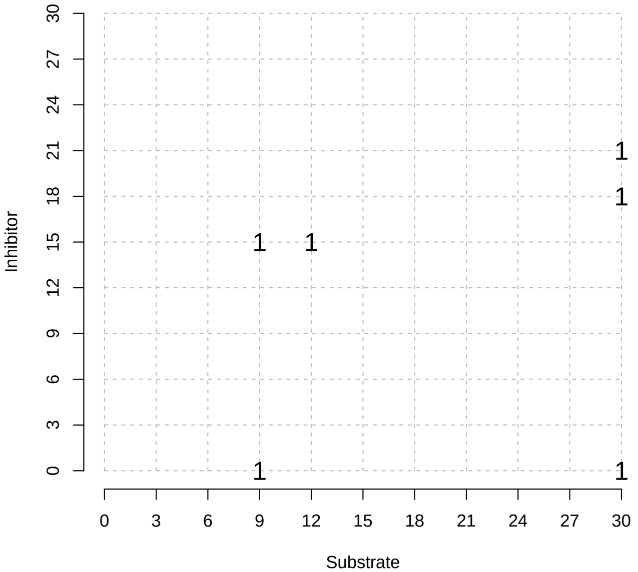
Confirmatory experiment $n = 6, r = [11.95, 14.84]$



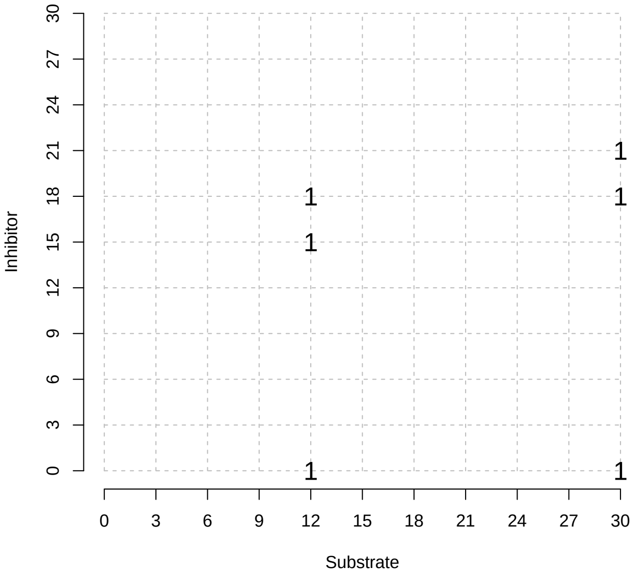
Confirmatory experiment $n = 6, r = [14.85, 16.90]$



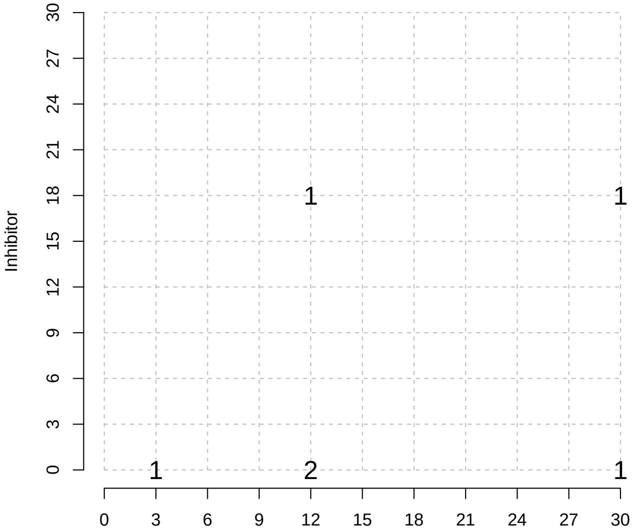
Confirmatory experiment $n = 6, r = [16.91, 17.30]$



Confirmatory experiment $n = 6, r = [17.31, 18.63]$



Confirmatory experiment $n = 6, r = [18.64, 20.00]$, normal errors



The simulation

We generated 10000 sets of observations with $n = 6$ from each of the two models at the nominal values given in Atkinson (2012) and the error st.dev. estimated from the data $\hat{\sigma} = 0.1526$, and recorded the total correct discrimination (hit) rates from the LR-rule.

c	0		1		5	
true model	η_0	η_1	η_0	η_1	η_0	η_1
A1	91.11	94.45	91.35	93.95	90.44	93.24
A2	97.11	96.75	97.47	96.64	96.74	96.27
A3	96.60	96.51	96.47	96.40	95.69	96.06
A4	**97.94**	96.57	97.73	96.29	97.62	96.07
$\delta 1$	97.59	95.11	97.43	94.90	**97.71**	94.56
$\delta 2$	97.93	**97.03**	**97.77**	**96.67**	97.20	**96.54**
$\delta 3$	96.50	95.29	96.42	95.36	96.19	95.64

Table: Total hit rates for $N = 10000$ under each model.

A second large experiment $n = 60$, lognormal errors

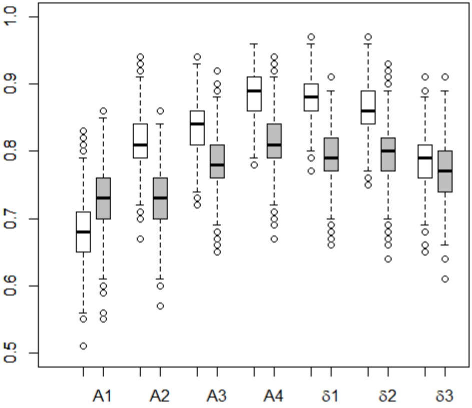


Figure: Boxplot for the total correct classification rates for all designs using nominal values and error standard deviations of $5 \times \hat{\sigma}$; white under η_0, grey under η_1.

Thank you for your attention!