Penalized D-optimal design for dose finding

Luc Pronzato

Laboratoire I3S, CNRS-Univ. Nice Sophia Antipolis, France

Outline

1) Introduction

2) Penalized D-optimal design

3) Sequential design

4) Sequential penalized D-optimal design

5) Examples

6) Conclusions

1) Introduction

Bernoulli-type experiments:
$Y_i \in \{0,1\}$ (success or failure)
$\eta(x,\theta) = \mathrm{Prob}(Y_i = 1|x_i = x,\theta)$
$\hat{\theta}^n$ the maximum-likelihood estimator:

$$\hat{\theta}^n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \{ Y_i \log[\eta(x_i,\theta)] + (1 - Y_i) \log[1 - \eta(x_i,\theta)]\}$$

Fisher information matrix:

$$\mathbf{M}(\xi,\theta) = \int_{\mathcal{X}} \mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)\,\xi(dx)$$

with

$$\mathbf{f}_\theta(x) = \{\eta(x,\theta)[1-\eta(x,\theta)]\}^{-1/2} \frac{\partial\eta(x,\theta)}{\partial\theta}$$

1) Introduction

Bernoulli-type experiments:
$Y_i \in \{0{,}1\}$ (success or failure)
$\eta(x{,}\theta) = \mathrm{Prob}(Y_i = 1|x_i = x{,}\theta)$
$\hat{\theta}^n$ the maximum-likelihood estimator:

$$\hat{\theta}^n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \{ Y_i \log[\eta(x_i{,}\theta)] + (1 - Y_i) \log[1 - \eta(x_i{,}\theta)]\}$$

Fisher information matrix:

$$\mathbf{M}(\xi{,}\theta) = \int_{\mathcal{X}} \mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)\,\xi(dx)$$

with

$$\mathbf{f}_\theta(x) = \{\eta(x{,}\theta)[1 - \eta(x{,}\theta)]\}^{-1/2} \frac{\partial\eta(x{,}\theta)}{\partial\theta}$$

2) Penalized D-optimal design

Maximize $\log\det \mathbf{M}(\xi,\theta)$ under the contraint $\Phi(\xi,\theta) \leq C$
with $\Phi(\xi,\theta) = \int_{\mathcal{X}} \phi(x,\theta)\,\xi(dx)$
$(\phi(x,\theta)$ = cost of one observation at $x)$

NSC for the optimality of ξ^*:

$\Phi(\xi^*,\theta) \leq C$ and there exists $\lambda^* = \lambda^*(C,\theta) \geq 0$ such that

$$\begin{cases} \lambda^*[C - \Phi(\xi^*,\theta)] = 0 \\ \forall x \in \mathcal{X}\,,\; \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi^*,\theta)\mathbf{f}_\theta(x) \leq p + \lambda^*[\phi(x,\theta) - \Phi(\xi^*,\theta)] \end{cases}$$

☞ In practice:

maximize $\boxed{H_\theta(\xi,\lambda) = \log\det \mathbf{M}(\xi,\theta) - \lambda\,\Phi(\xi,\theta)}$

$\Leftrightarrow$ penalized D-optimal design

for an increasing sequence $\{\lambda_i\}$ of Lagrange multipliers, starting from $\lambda_0 = 0$ and stopping at the first λ_i for which $\xi^*(\lambda_i)$ satisfies $\Phi[\xi^*(\lambda_i),\theta] \leq C$ [Mikulecká 1983]

➠ not more complicated than the determination of (a sequence of) D-optimal design(s)!

2) Penalized D-optimal design

Maximize $\log\det \mathbf{M}(\xi,\theta)$ under the contraint $\Phi(\xi,\theta) \leq C$
with $\Phi(\xi,\theta) = \int_{\mathcal{X}} \phi(x,\theta)\,\xi(dx)$
($\phi(x,\theta)$ = cost of one observation at x)

NSC for the optimality of ξ^*:

$\Phi(\xi^*,\theta) \leq C$ and there exists $\lambda^* = \lambda^*(C,\theta) \geq 0$ such that

$$\begin{cases} \lambda^*[C - \Phi(\xi^*,\theta)] = 0 \\ \forall x \in \mathcal{X}\,,\ \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi^*,\theta)\mathbf{f}_\theta(x) \leq p + \lambda^*[\phi(x,\theta) - \Phi(\xi^*,\theta)] \end{cases}$$

☞ In practice:

maximize $H_\theta(\xi,\lambda) = \log\det \mathbf{M}(\xi,\theta) - \lambda\,\Phi(\xi,\theta)$

$\Leftrightarrow$ penalized D-optimal design

for an increasing sequence $\{\lambda_i\}$ of Lagrange multipliers, starting from $\lambda_0 = 0$ and stopping at the first λ_i for which $\xi^*(\lambda_i)$ satisfies $\Phi[\xi^*(\lambda_i),\theta] \leq C$ [Mikulecká 1983]

➠ not more complicated than the determination of (a sequence of) D-optimal design(s)!

2) Penalized D-optimal design

Maximize $\log\det \mathbf{M}(\xi,\theta)$ under the contraint $\Phi(\xi,\theta) \leq C$
with $\Phi(\xi,\theta) = \int_{\mathcal{X}} \phi(x,\theta)\,\xi(dx)$
($\phi(x,\theta)$ = cost of one observation at x)

NSC for the optimality of ξ^*:

$\Phi(\xi^*,\theta) \leq C$ and there exists $\lambda^* = \lambda^*(C,\theta) \geq 0$ such that

$$\begin{cases} \lambda^*[C - \Phi(\xi^*,\theta)] = 0 \\ \forall x \in \mathcal{X}\,,\ \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi^*,\theta)\mathbf{f}_\theta(x) \leq p + \lambda^*[\phi(x,\theta) - \Phi(\xi^*,\theta)] \end{cases}$$

☞ In practice:

maximize $\boxed{H_\theta(\xi,\lambda) = \log\det \mathbf{M}(\xi,\theta) - \lambda\,\Phi(\xi,\theta)}$

$\Leftrightarrow$ penalized D-optimal design

for an increasing sequence $\{\lambda_i\}$ of Lagrange multipliers, starting from $\lambda_0 = 0$ and stopping at the first λ_i for which $\xi^*(\lambda_i)$ satisfies $\Phi[\xi^*(\lambda_i),\theta] \leq C$ [Mikulecká 1983]

➟ not more complicated than the determination of (a sequence of) D-optimal design(s)!

Dose-response problem in clinical trials:

Define $\phi(x,\theta)$ from the success probability (efficacy, no toxicity)

[Dragalin & Fedorov 2006; Dragalin, Fedorov & Wu 2008]

☞ λ in $\boxed{H_\theta(\xi,\lambda) = \log\det \mathbf{M}(\xi,\theta) - \lambda\,\Phi(\xi,\theta)}$ sets a compromise between

- the information gain (a pb. of collective ethics) and
- the rejection of non efficient or toxic doses (a pb. of individual ethics — for the enroled patients)

$\xi^*(\lambda)$ maximizes $H_\theta(\xi,\lambda) = \log\det \mathbf{M}(\xi,\theta) - \lambda\, \Phi(\xi,\theta)$, with
$\Phi(\xi,\theta) = \int_{\mathcal{X}} \phi(x,\theta)\, \xi(dx)$

3 remarks:

1. If $\Phi(\xi,\theta) = \frac{1}{p} \log \Psi(\xi,\theta)$
$\Leftrightarrow$ maximize $\log\det \left[\frac{\mathbf{M}(\xi,\theta)}{\Psi^\lambda(\xi,\theta)}\right]$
$\Leftrightarrow$ maximize $\log\det[N\mathbf{M}(\xi,\theta)]$
with the total cost constraint $N\Psi^\lambda(\xi,\theta) \leq C$, any $C > 0$
[Dragalin & Fedorov 2006]

2. $\Phi[\xi^*(\lambda),\theta] \leq \min_x \phi(x,\theta) + \frac{\dim(\theta)}{\lambda}$

$\xi^*(\lambda)$ maximizes $H_\theta(\xi,\lambda) = \log\det \mathbf{M}(\xi,\theta) - \lambda\,\Phi(\xi,\theta)$, with $\Phi(\xi,\theta) = \int_{\mathcal{X}} \phi(x,\theta)\,\xi(dx)$

3 remarks:

1. If $\Phi(\xi,\theta) = \frac{1}{p}\log\Psi(\xi,\theta)$

$\Leftrightarrow$ maximize $\log\det\left[\frac{\mathbf{M}(\xi,\theta)}{\Psi^\lambda(\xi,\theta)}\right]$

$\Leftrightarrow$ maximize $\log\det[N\mathbf{M}(\xi,\theta)]$

with the total cost constraint $N\Psi^\lambda(\xi,\theta) \leq C$, any $C > 0$

[Dragalin & Fedorov 2006]

2. $\Phi[\xi^*(\lambda),\theta] \leq \min_x \phi(x,\theta) + \frac{\dim(\theta)}{\lambda}$

$\xi^*(\lambda)$ maximizes $H_\theta(\xi,\lambda) = \log\det \mathbf{M}(\xi,\theta) - \lambda\,\Phi(\xi,\theta)$, with $\Phi(\xi,\theta) = \int_{\mathcal{X}} \phi(x,\theta)\,\xi(dx)$

3 remarks:

1. If $\Phi(\xi,\theta) = \frac{1}{p}\log\Psi(\xi,\theta)$

$\Leftrightarrow$ maximize $\log\det\left[\frac{\mathbf{M}(\xi,\theta)}{\Psi^\lambda(\xi,\theta)}\right]$

$\Leftrightarrow$ maximize $\log\det[N\mathbf{M}(\xi,\theta)]$

with the total cost constraint $N\Psi^\lambda(\xi,\theta) \leq C$, any $C > 0$ [Dragalin & Fedorov 2006]

2. $\Phi[\xi^*(\lambda),\theta] \leq \min_x \phi(x,\theta) + \frac{\dim(\theta)}{\lambda}$

3. If $\phi(x,\theta)$ has a unique minimum in $\mathcal{X}$ at x^* and is flat enough around x^*, the support points of $\xi^*(\lambda)$ minimizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\, \Phi(\xi,\theta)$, tend to gather around x^* when $\lambda \to \infty$ [LP, JSPI, 2009]

Suppose $\mathcal{X}$ finite with
$\underbrace{\phi(x^{(1)},\theta) \leq \ldots \leq \phi(x^{(m)},\theta)} \leq \ldots \leq \phi(x^{(K)},\theta)$

Define $\mathcal{X}_m = \{x^{(1)},\ldots,x^{(m)}\}$, $\delta = \phi(x^{(m+1)},\theta) - \phi(x^{(m)},\theta)$ and suppose that $\mathbf{f}_\theta(x^{(1)}),\ldots,\mathbf{f}_\theta(x^{(m)})$ span $\mathbb{R}^p$

Define $\tilde{\phi}(x,\theta) = [\phi(x,\theta) - \phi(x^{(1)},\theta)]$ and consider $\Phi_q(\xi,\theta) = \int_{\mathcal{X}} \tilde{\phi}^q(x,\theta)\,\xi(dx)$

Then $\xi^*(\lambda)$ maximizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\, \Phi_q(\xi,\theta)$ is supported on $\mathcal{X}_m$ when λ and q large enough

$[q > \log(2\hat{t}_m)/\log\{1 + \delta/[\phi(x^{(m)},\theta) - \phi(x^{(1)},\theta)]\}$
$\lambda > \lambda_{m,q} = p/\Phi_q(\hat{\xi}_m,\theta)$
with $\hat{t}_m = \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\hat{\xi}_m,\theta)\mathbf{f}_\theta(x)$
and $\hat{\xi}_m = \arg\min_{\xi \text{ supp. on } \mathcal{X}_m} \max_{x\in\mathcal{X}\setminus\mathcal{X}_m} \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi,\theta)\mathbf{f}_\theta(x)]$

3. If $\phi(x,\theta)$ has a unique minimum in $\mathcal{X}$ at x^* and is flat enough around x^*, the support points of $\xi^*(\lambda)$ minimizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\, \Phi(\xi,\theta)$, tend to gather around x^* when $\lambda \to \infty$ [LP, JSPI, 2009]

Suppose $\mathcal{X}$ finite with
$\underbrace{\phi(x^{(1)},\theta) \leq \ldots \leq \phi(x^{(m)},\theta)} \leq \ldots \leq \phi(x^{(K)},\theta)$

Define $\mathcal{X}_m = \{x^{(1)},\ldots,x^{(m)}\}$, $\delta = \phi(x^{(m+1)},\theta) - \phi(x^{(m)},\theta)$ and suppose that $\mathbf{f}_\theta(x^{(1)}),\ldots,\mathbf{f}_\theta(x^{(m)})$ span $\mathbb{R}^p$

Define $\tilde{\phi}(x,\theta) = [\phi(x,\theta) - \phi(x^{(1)},\theta)]$
and consider $\Phi_q(\xi,\theta) = \int_{\mathcal{X}} \tilde{\phi}^q(x,\theta)\,\xi(dx)$

Then $\xi^*(\lambda)$ maximizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\, \Phi_q(\xi,\theta)$ is supported on $\mathcal{X}_m$ when λ and q large enough

[$q > \log(2\hat{t}_m)/\log\{1 + \delta/[\phi(x^{(m)},\theta) - \phi(x^{(1)},\theta)]\}$
$\lambda > \lambda_{m,q} = p/\Phi_q(\hat{\xi}_m,\theta)$
with $\hat{t}_m = \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\hat{\xi}_m,\theta)\mathbf{f}_\theta(x)$
and $\hat{\xi}_m = \arg\min_\xi$ supp. on $\mathcal{X}_m$ $\max_{x\in\mathcal{X}\setminus\mathcal{X}_m} \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi,\theta)\mathbf{f}_\theta(x)$]

3. If $\phi(x,\theta)$ has a unique minimum in $\mathcal{X}$ at x^* and is flat enough around x^*, the support points of $\xi^*(\lambda)$ minimizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\,\Phi(\xi,\theta)$, tend to gather around x^* when $\lambda \to \infty$ [LP, JSPI, 2009]

Suppose $\mathcal{X}$ finite with
$\underbrace{\phi(x^{(1)},\theta) \leq \ldots \leq \phi(x^{(m)},\theta)} \leq \ldots \leq \phi(x^{(K)},\theta)$
Define $\mathcal{X}_m = \{x^{(1)},\ldots,x^{(m)}\}$, $\delta = \phi(x^{(m+1)},\theta) - \phi(x^{(m)},\theta)$
and suppose that $\mathbf{f}_\theta(x^{(1)}),\ldots,\mathbf{f}_\theta(x^{(m)})$ span $\mathbb{R}^p$

Define $\tilde{\phi}(x,\theta) = [\phi(x,\theta) - \phi(x^{(1)},\theta)]$
and consider $\Phi_q(\xi,\theta) = \int_{\mathcal{X}} \tilde{\phi}^q(x,\theta)\,\xi(dx)$

Then $\xi^*(\lambda)$ maximizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\,\Phi_q(\xi,\theta)$ is supported on $\mathcal{X}_m$ when λ and q large enough

$[q > \log(2\hat{t}_m)/\log\{1+\delta/[\phi(x^{(m)},\theta) - \phi(x^{(1)},\theta)]\}$
$\lambda > \lambda_{m,q} = p/\Phi_q(\hat{\xi}_m,\theta)$
with $\hat{t}_m = \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\hat{\xi}_m,\theta)\mathbf{f}_\theta(x)$
and $\hat{\xi}_m = \arg\min_\xi$ supp. on $\mathcal{X}_m$ $\max_{x\in\mathcal{X}\setminus\mathcal{X}_m} \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi,\theta)\mathbf{f}_\theta(x)]$

3. If $\phi(x,\theta)$ has a unique minimum in $\mathcal{X}$ at x^* and is flat enough around x^*, the support points of $\xi^*(\lambda)$ minimizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\, \Phi(\xi,\theta)$, tend to gather around x^* when $\lambda \to \infty$ [LP, JSPI, 2009]

Suppose $\mathcal{X}$ finite with
$\underbrace{\phi(x^{(1)},\theta) \leq \ldots \leq \phi(x^{(m)},\theta)} \leq \ldots \leq \phi(x^{(K)},\theta)$

Define $\mathcal{X}_m = \{x^{(1)},\ldots,x^{(m)}\}$, $\delta = \phi(x^{(m+1)},\theta) - \phi(x^{(m)},\theta)$ and suppose that $\mathbf{f}_\theta(x^{(1)}),\ldots,\mathbf{f}_\theta(x^{(m)})$ span $\mathbb{R}^p$

Define $\tilde{\phi}(x,\theta) = [\phi(x,\theta) - \phi(x^{(1)},\theta)]$
and consider $\Phi_q(\xi,\theta) = \int_{\mathcal{X}} \tilde{\phi}^q(x,\theta)\, \xi(dx)$

Then $\xi^*(\lambda)$ maximizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\, \Phi_q(\xi,\theta)$ is supported on $\mathcal{X}_m$ when λ and q large enough

$[q > \log(2\hat{t}_m)/\log\{1 + \delta/[\phi(x^{(m)},\theta) - \phi(x^{(1)},\theta)]\}$
$\lambda > \lambda_{m,q} = p/\Phi_q(\hat{\xi}_m,\theta)$
with $\hat{t}_m = \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\hat{\xi}_m,\theta)\mathbf{f}_\theta(x)$
and $\hat{\xi}_m = \arg\min_\xi$ supp. on $\mathcal{X}_m$ $\max_{x\in\mathcal{X}\setminus\mathcal{X}_m} \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi,\theta)\mathbf{f}_\theta(x)]$

3. If $\phi(x,\theta)$ has a unique minimum in $\mathcal{X}$ at x^* and is flat enough around x^*, the support points of $\xi^*(\lambda)$ minimizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\, \Phi(\xi,\theta)$, tend to gather around x^* when $\lambda \to \infty$ [LP, JSPI, 2009]

Suppose $\mathcal{X}$ finite with
$\underbrace{\phi(x^{(1)},\theta) \leq \ldots \leq \phi(x^{(m)},\theta)} \leq \ldots \leq \phi(x^{(K)},\theta)$

Define $\mathcal{X}_m = \{x^{(1)},\ldots,x^{(m)}\}$, $\delta = \phi(x^{(m+1)},\theta) - \phi(x^{(m)},\theta)$ and suppose that $\mathbf{f}_\theta(x^{(1)}),\ldots,\mathbf{f}_\theta(x^{(m)})$ span $\mathbb{R}^p$

Define $\tilde{\phi}(x,\theta) = [\phi(x,\theta) - \phi(x^{(1)},\theta)]$
and consider $\Phi_q(\xi,\theta) = \int_{\mathcal{X}} \tilde{\phi}^q(x,\theta)\,\xi(dx)$

Then $\xi^*(\lambda)$ maximizing $\log\det \mathbf{M}(\xi,\theta) - \lambda\, \Phi_q(\xi,\theta)$ is supported on $\mathcal{X}_m$ when λ and q large enough

[$q > \log(2\hat{t}_m)/\log\{1 + \delta/[\phi(x^{(m)},\theta) - \phi(x^{(1)},\theta)]\}$
$\lambda > \lambda_{m,q} = p/\Phi_q(\hat{\xi}_m,\theta)$
with $\hat{t}_m = \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\hat{\xi}_m,\theta)\mathbf{f}_\theta(x)$
and $\hat{\xi}_m = \arg\min_\xi$ supp. on $\mathcal{X}_m$ $\max_{x\in\mathcal{X}\setminus\mathcal{X}_m} \mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi,\theta)\mathbf{f}_\theta(x)$]

Example 1 [Dragalin & Fedorov 2006]

Cox model for bivariate responses: $Y \to$ efficacy, $Z \to$ toxicity
11 doses available, evenly spread in $[-3,3]$,
$\mathcal{X} = \{x^{(1)}, \ldots, x^{(11)}\}$, $x^{(i)} < x^{(i+1)}$, $i = 1, \ldots, 10$
$\mathrm{Prob}\{Y = y, Z = z | x, \theta\} = \pi_{yz}(x,\theta)$, $Y, y, Z, z \in \{0,1\}$
$\theta = (a_{11}, b_{11}, a_{10}, b_{10}, a_{01}, b_{01})^\top = (3,3,4,2,0,2)^\top \in \mathbb{R}^6$

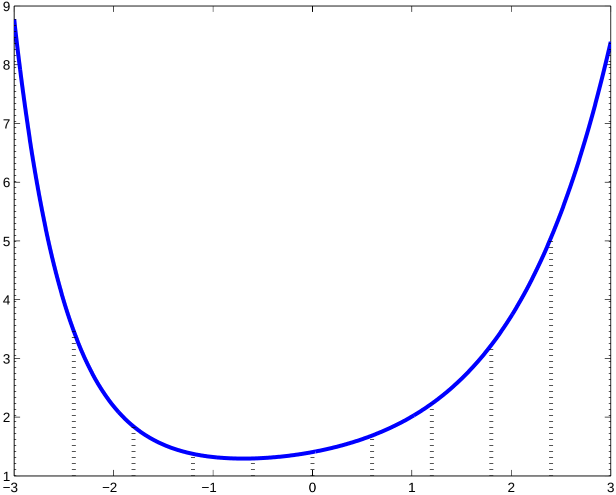
$$\pi_{11}(x,\theta) = \frac{e^{a_{11}+b_{11}x}}{1+e^{a_{01}+b_{01}x}+e^{a_{10}+b_{10}x}+e^{a_{11}+b_{11}x}}$$
$$\pi_{10}(x,\theta) = \frac{e^{a_{10}+b_{10}x}}{1+e^{a_{01}+b_{01}x}+e^{a_{10}+b_{10}x}+e^{a_{11}+b_{11}x}}$$
$$\pi_{01}(x,\theta) = \frac{e^{a_{01}+b_{01}x}}{1+e^{a_{01}+b_{01}x}+e^{a_{10}+b_{10}x}+e^{a_{11}+b_{11}x}}$$
$$\pi_{00}(x,\theta) = \left(1+e^{a_{01}+b_{01}x}+e^{a_{10}+b_{10}x}+e^{a_{11}+b_{11}x}\right)^{-1}$$

Example 1 [Dragalin & Fedorov 2006]

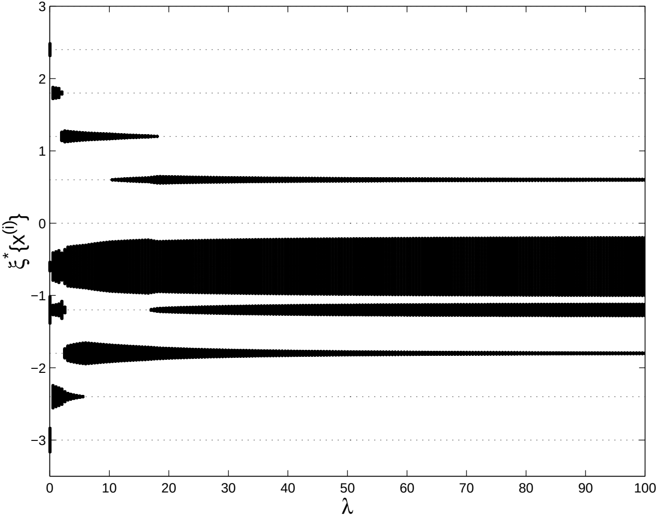
Cox model for bivariate responses: $Y \rightarrow$ efficacy, $Z \rightarrow$ toxicity
11 doses available, evenly spread in $[-3,3]$,
$\mathcal{X} = \{x^{(1)},\ldots,x^{(11)}\}$, $x^{(i)} < x^{(i+1)}$, $i = 1,\ldots,10$
$\mathrm{Prob}\{Y = y, Z = z | x, \theta\} = \pi_{yz}(x,\theta)$, $Y,y,Z,z \in \{0,1\}$
$\theta = (a_{11}, b_{11}, a_{10}, b_{10}, a_{01}, b_{01})^\top = (3,3,4,2,0,2)^\top \in \mathbb{R}^6$

$$
\begin{aligned}
\pi_{11}(x,\theta) &= \frac{e^{a_{11}+b_{11}x}}{1+e^{a_{01}+b_{01}x}+e^{a_{10}+b_{10}x}+e^{a_{11}+b_{11}x}} \\
\pi_{10}(x,\theta) &= \frac{e^{a_{10}+b_{10}x}}{1+e^{a_{01}+b_{01}x}+e^{a_{10}+b_{10}x}+e^{a_{11}+b_{11}x}} \\
\pi_{01}(x,\theta) &= \frac{e^{a_{01}+b_{01}x}}{1+e^{a_{01}+b_{01}x}+e^{a_{10}+b_{10}x}+e^{a_{11}+b_{11}x}} \\
\pi_{00}(x,\theta) &= \left(1+e^{a_{01}+b_{01}x}+e^{a_{10}+b_{10}x}+e^{a_{11}+b_{11}x}\right)^{-1}
\end{aligned}
$$

$\phi_1(x,\theta) = \pi_{10}^{-1}(x,\theta)$ (π_{10} = efficacy and no toxicity)
Optimal dose x^*: minimizes $\phi_1(x,\theta) \rightarrow x^* = x^{(5)} = -0.6$



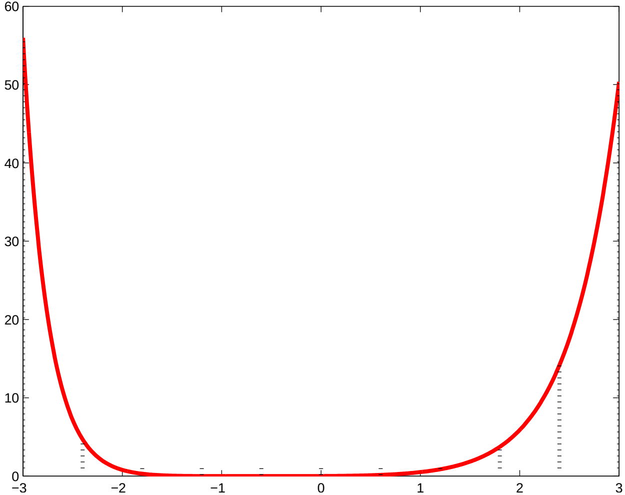
$\phi_1(x,\theta) = \pi_{10}^{-1}(x,\theta)$ (π_{10} = efficacy and no toxicity)
Optimal dose x^*: minimizes $\phi_1(x,\theta) \to x^* = x^{(5)} = -0.6$



Other cost function:

$\phi_2(x,\theta) = \{\pi_{10}^{-1}(x,\theta) - [\max_x \pi_{10}(x,\theta)]^{-1}\}^2$

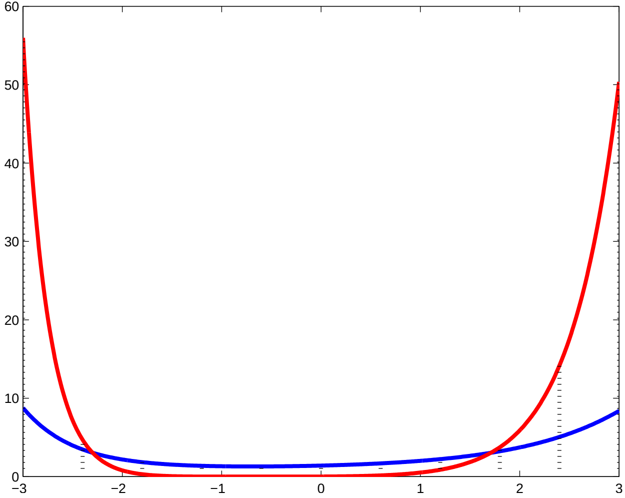
(more flat than $\phi_1(x,\theta)$ around x^*)



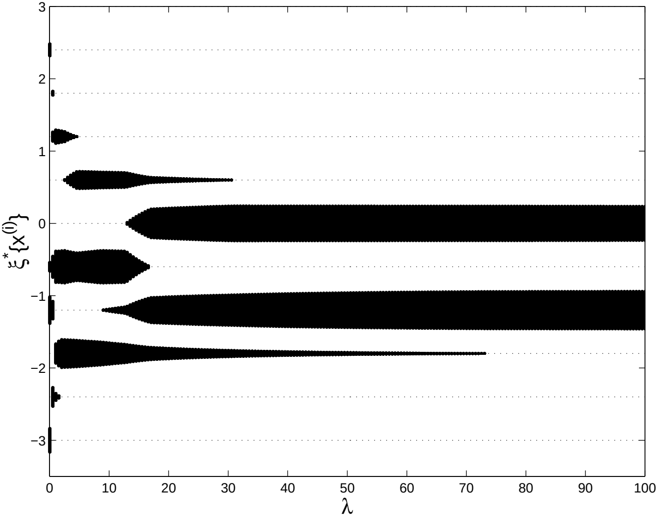
Other cost function:

$\phi_2(x,\theta) = \{\pi_{10}^{-1}(x,\theta) - [\max_x \pi_{10}(x,\theta)]^{-1}\}^2$

(more flat than $\phi_1(x,\theta)$ around x^*)

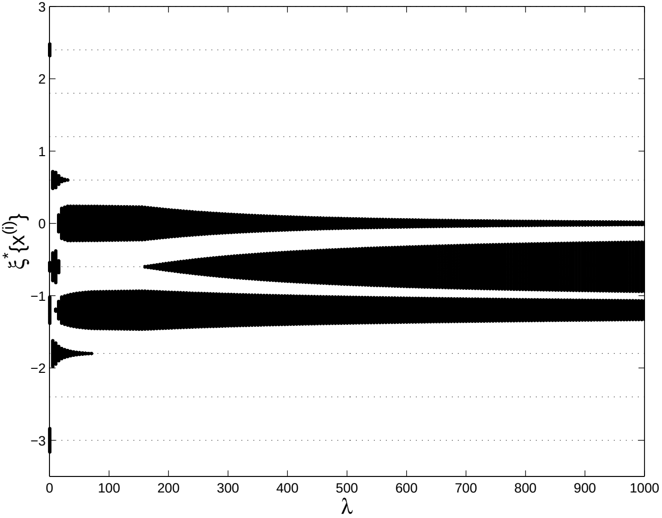


$\phi_2(x,\theta) = \{\pi_{10}^{-1}(x,\theta) - [\max_x \pi_{10}(x,\theta)]^{-1}\}^2$
(more flat than $\phi_1(x,\theta)$ around x^*)



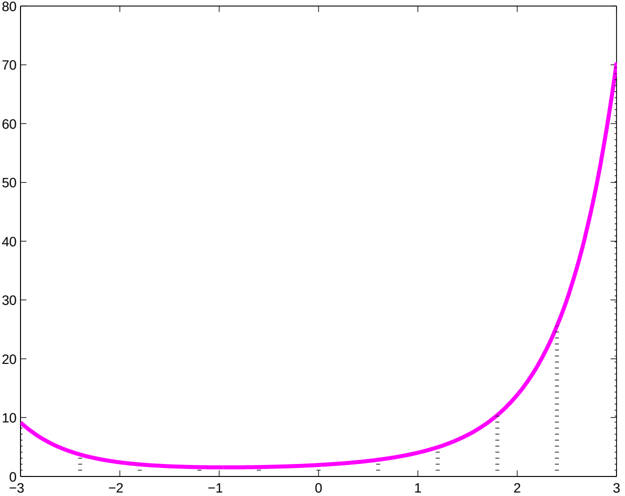
$\lambda \in [1,100]$

$\phi_2(x,\theta) = \{\pi_{10}^{-1}(x,\theta) - [\max_x \pi_{10}(x,\theta)]^{-1}\}^2$
(more flat than $\phi_1(x,\theta)$ around x^*)

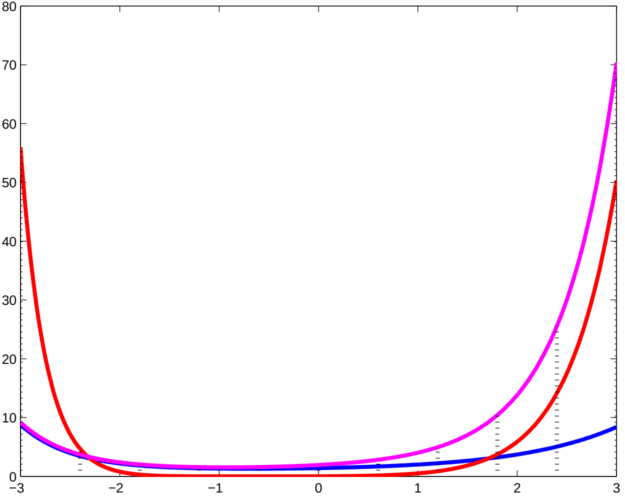


$\lambda \in [1,1000]$

yet another cost function: (more importance given to toxicity)
$\phi_3(x,\theta) = \pi_{10}^{-1}(x,\theta)[1 - \pi_{\cdot 1}(x,\theta)]^{-1}$ with
$\pi_{\cdot 1}(x,\theta) = \pi_{01}(x,\theta) + \pi_{11}(x,\theta)$ = marginal probability of toxicity ($\phi_3(x,\theta)$ is minimum at $x^{(4)} = -1.2$)

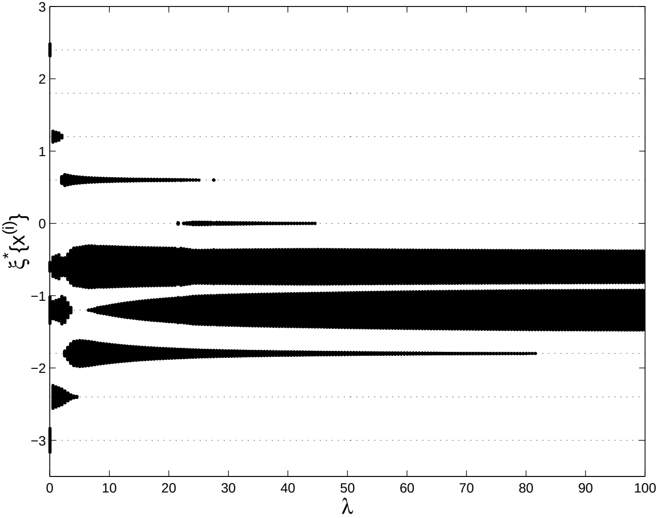


yet another cost function: (more importance given to toxicity)
$\phi_3(x,\theta) = \pi_{10}^{-1}(x,\theta)[1 - \pi_{.1}(x,\theta)]^{-1}$ with
$\pi_{.1}(x,\theta) = \pi_{01}(x,\theta) + \pi_{11}(x,\theta)$ = marginal probability of toxicity ($\phi_3(x,\theta)$ is minimum at $x^{(4)} = -1.2$)



INI
08/2011

$\phi_3(x,\theta) = \pi_{10}^{-1}(x,\theta)[1 - \pi_{.1}(x,\theta)]^{-1}$ with $\pi_{.1}(x,\theta) = \pi_{01}(x,\theta) + \pi_{11}(x,\theta)$ = marginal probability of toxicity ($\phi_3(x,\theta)$ is minimum at $x^{(4)} = -1.2$)



Motivation: why sequential design?

Regression model $\eta(x,\theta)$, nonlinear in $\theta \Rightarrow$ the optimal design for estimating θ depends on θ!
Information matrix $\mathbf{M}(\xi,\theta) = \int_{\mathcal{X}} \mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)\,\xi(dx)$ with

- ξ a probability measure on $\mathcal{X}$
- $\mathbf{f}_\theta(x) = \frac{1}{\sigma}\frac{\partial\eta(x,\theta)}{\partial\theta}$ (cont. diff. w.r.t. θ for any $x \in \mathcal{X}$)

$\xi_D^*(\theta)$ is D-optimal for θ: $\xi_D^*(\theta)$ maximizes $\log\det[\mathbf{M}(\xi,\theta)]$

Examples

- [Box & Lucas, 1959]:
  $\eta(x,\theta) = \frac{\theta_1}{\theta_1-\theta_2}\,[\exp(-\theta_2\, x) - \exp(-\theta_1\, x)]$
  $\rightarrow$ for $\theta = (0.7\,,\, 0.2)^\top$, $\xi_D^* = \frac{1}{2}\delta_{x^{(1)}} + \frac{1}{2}\delta_{x^{(2)}}$ with $x^{(1)} \simeq 1.25$ and $x^{(2)} \simeq 6.60$
- Michaelis-Menten : $\eta(x,\theta) = \frac{\theta_1\, x}{\theta_2 + x}$, $x \in (0,\overline{x}]$, $\theta_1,\theta_2 > 0$
  $\rightarrow \xi_D^* = \frac{1}{2}\delta_{x^{(1)}} + \frac{1}{2}\delta_{x^{(2)}}$ with $x^{(1)} = \frac{\theta_2\overline{x}}{2\theta_2+\overline{x}}$ and $x^{(2)} = \overline{x}$
- Exponential decrease: $\eta(x,\theta) = \theta_1\, \exp(-\theta_2\, x)$, $x \geq \underline{x}$, $\theta_1,\theta_2 > 0$
  $\rightarrow \xi_D^* = \frac{1}{2}\delta_{x^{(1)}} + \frac{1}{2}\delta_{x^{(2)}}$ with $x^{(1)} = \underline{x}$ and $x^{(2)} = \underline{x} + \frac{1}{\theta_2}$
- ...

Examples

- [Box & Lucas, 1959]:
  $\eta(x,\theta) = \frac{\theta_1}{\theta_1-\theta_2}\,[\exp(-\theta_2\,x) - \exp(-\theta_1\,x)]$
  $\rightarrow$ for $\theta = (0.7\,,\,0.2)^\top$, $\xi_D^* = \frac{1}{2}\delta_{x^{(1)}} + \frac{1}{2}\delta_{x^{(2)}}$ with $x^{(1)} \simeq 1.25$ and $x^{(2)} \simeq 6.60$
- Michaelis-Menten : $\eta(x,\theta) = \frac{\theta_1\,x}{\theta_2+x}$, $x \in (0,\overline{x}]$, $\theta_1,\theta_2 > 0$
  $\rightarrow \xi_D^* = \frac{1}{2}\delta_{x^{(1)}} + \frac{1}{2}\delta_{x^{(2)}}$ with $x^{(1)} = \frac{\theta_2\overline{x}}{2\theta_2+\overline{x}}$ and $x^{(2)} = \overline{x}$
- Exponential decrease: $\eta(x,\theta) = \theta_1\,\exp(-\theta_2\,x)$, $x \geq \underline{x}$, $\theta_1,\theta_2 > 0$
  $\rightarrow \xi_D^* = \frac{1}{2}\delta_{x^{(1)}} + \frac{1}{2}\delta_{x^{(2)}}$ with $x^{(1)} = \underline{x}$ and $x^{(2)} = \underline{x} + \frac{1}{\theta_2}$
- ...

Examples

- [Box & Lucas, 1959]:
  $\eta(x,\theta) = \frac{\theta_1}{\theta_1-\theta_2}\,[\exp(-\theta_2\, x) - \exp(-\theta_1\, x)]$
  $\rightarrow$ for $\theta = (0.7\,,\,0.2)^\top$, $\xi_D^* = \frac{1}{2}\delta_{x^{(1)}} + \frac{1}{2}\delta_{x^{(2)}}$ with $x^{(1)} \simeq 1.25$ and $x^{(2)} \simeq 6.60$
- Michaelis-Menten : $\eta(x,\theta) = \frac{\theta_1\, x}{\theta_2 + x}$, $x \in (0,\overline{x}]$, $\theta_1,\theta_2 > 0$
  $\rightarrow \xi_D^* = \frac{1}{2}\delta_{x^{(1)}} + \frac{1}{2}\delta_{x^{(2)}}$ with $x^{(1)} = \frac{\theta_2\overline{x}}{2\theta_2+\overline{x}}$ and $x^{(2)} = \overline{x}$
- Exponential decrease: $\eta(x,\theta) = \theta_1\, \exp(-\theta_2\, x)$, $x \geq \underline{x}$, $\theta_1,\theta_2 > 0$
  $\rightarrow \xi_D^* = \frac{1}{2}\delta_{x^{(1)}} + \frac{1}{2}\delta_{x^{(2)}}$ with $x^{(1)} = \underline{x}$ and $x^{(2)} = \underline{x} + \frac{1}{\theta_2}$
- ...

full sequential D-optimal design: choose $x_1,\ldots,x_{n_0}$, estimate $\hat\theta^{n_0}$, set $k = n_0$ then

- design x_{k+1}
- observe Y_{k+1}
- re-estimate $\hat\theta^{k+1}$
- $k \leftarrow k+1 \ldots$

D-optimality:

$$x_{k+1} = \arg\max_{x\in\mathcal{X}} \det\left[\sum_{i=1}^k \mathbf{f}_{\hat\theta^k}(x_i)\mathbf{f}^\top_{\hat\theta^k}(x_i) + \mathbf{f}_{\hat\theta^k}(x)\mathbf{f}^\top_{\hat\theta^k}(x)\right]$$

or equivalently

$$\boxed{x_{k+1} = \arg\max_{x\in\mathcal{X}} \mathbf{f}^\top_{\hat\theta^k}(x)\mathbf{M}^{-1}(\xi_k,\hat\theta^k)\mathbf{f}_{\hat\theta^k}(x)}$$ with

$\xi_k = \frac{1}{k}\sum_{i=1}^k \delta_{x_i}$ the empirical measure defined by $x_1,\ldots,x_k$

$\rightarrow \mathbf{M}(\xi_k,\theta) = \frac{1}{k}\sum_{i=1}^k \mathbf{f}_\theta(x_i)\mathbf{f}^\top_\theta(x_i)$

we hope that $\hat\theta^n \xrightarrow{\text{a.s.}} \bar\theta$ and $\sqrt{n}(\hat\theta^n - \bar\theta) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0},\mathbf{M}^{-1}[\xi_D^*(\bar\theta),\bar\theta])$

full sequential D-optimal design: choose $x_1, \ldots, x_{n_0}$, estimate $\hat{\theta}^{n_0}$, set $k = n_0$ then

- design x_{k+1}
- observe Y_{k+1}
- re-estimate $\hat{\theta}^{k+1}$
- $k \leftarrow k+1 \ldots$

D-optimality:

$$x_{k+1} = \arg\max_{x \in \mathcal{X}} \det \left[\sum_{i=1}^{k} \mathbf{f}_{\hat{\theta}^k}(x_i) \mathbf{f}^\top_{\hat{\theta}^k}(x_i) + \mathbf{f}_{\hat{\theta}^k}(x) \mathbf{f}^\top_{\hat{\theta}^k}(x) \right]$$

or equivalently

$$\boxed{x_{k+1} = \arg\max_{x \in \mathcal{X}} \mathbf{f}^\top_{\hat{\theta}^k}(x) \mathbf{M}^{-1}(\xi_k, \hat{\theta}^k) \mathbf{f}_{\hat{\theta}^k}(x)}$$ with

$\xi_k = \frac{1}{k} \sum_{i=1}^{k} \delta_{x_i}$ the empirical measure defined by $x_1, \ldots, x_k$

$\rightarrow \mathbf{M}(\xi_k, \theta) = \frac{1}{k} \sum_{i=1}^{k} \mathbf{f}_\theta(x_i) \mathbf{f}^\top_\theta(x_i)$

we hope that $\hat{\theta}^n \stackrel{\text{a.s.}}{\rightarrow} \bar{\theta}$ and $\sqrt{n}(\hat{\theta}^n - \bar{\theta}) \stackrel{\text{d}}{\rightarrow} \mathcal{N}(\mathbf{0}, \mathbf{M}^{-1}[\xi^*_D(\bar{\theta}), \bar{\theta}])$

LS estimation in nonlinear regression

$Y_i = \eta(x_i,\bar{\theta}) + \varepsilon_i$, $x_i \in \mathcal{X} \subset \mathbb{R}^d$, $\bar{\theta} \in \Theta \subset \mathbb{R}^p$, $\{\varepsilon_i\}$ i.i.d., variance σ^2

LS estimation: $\hat{\theta}^n = \arg\min_{\theta\in\Theta} S_n(\theta)$ with $S_n(\theta) = \sum_{i=1}^n [Y_i - \eta(x_i,\theta)]^2$

Conditions for strong consistency of $\hat{\theta}^n$ ($\hat{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}$) very much differ depending whether the x_k are constants or depend on ε_i, $i < k$

Sequential D-optimal design

For $n_0 \leq k < n$,

$$x_{k+1} = \arg\max_{x\in\mathcal{X}} \mathbf{f}^{\top}_{\hat{\theta}^k}(x)\mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k}(x)$$

How to ensure that $\hat{\theta}^n \xrightarrow{\text{a.s.}} \bar{\theta}$ and $\sqrt{n}(\hat{\theta}^n - \bar{\theta}) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0},\mathbf{M}^{-1}[\xi^*_D(\bar{\theta}),\bar{\theta}])$ when $n \to \infty$?

1. deterministic choice of x_k when $k \in \{k_1,k_2,\ldots\}$, with $k_i \sim i^{\alpha}$, $\alpha \in (1,2)$ [Lai 1994]
2. let n_0 tend to ∞ when $n \to \infty$
3. suppose that $\mathcal{X}$ is a finite set
   $\Rightarrow$ replication of observations

Sequential D-optimal design

For $n_0 \leq k < n$,

$$x_{k+1} = \arg\max_{x \in \mathcal{X}} \mathbf{f}^\top_{\hat{\theta}^k}(x) \mathbf{M}^{-1}(\xi_k, \hat{\theta}^k) \mathbf{f}_{\hat{\theta}^k}(x)$$

How to ensure that $\hat{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}$ and $\sqrt{n}(\hat{\theta}^n - \bar{\theta}) \stackrel{\text{d}}{\to} \mathcal{N}(\mathbf{0}, \mathbf{M}^{-1}[\xi^*_D(\bar{\theta}), \bar{\theta}])$ when $n \to \infty$?

1. deterministic choice of x_k when $k \in \{k_1, k_2, \ldots\}$, with $k_i \sim i^\alpha$, $\alpha \in (1,2)$ [Lai 1994]
2. let n_0 tend to ∞ when $n \to \infty$
3. suppose that $\mathcal{X}$ is a finite set
   $\Rightarrow$ replication of observations

Sequential *D*-optimal design

For $n_0 \leq k < n$,

$$x_{k+1} = \arg\max_{x \in \mathcal{X}} \mathbf{f}^{\top}_{\hat{\theta}^k}(x) \mathbf{M}^{-1}(\xi_k, \hat{\theta}^k) \mathbf{f}_{\hat{\theta}^k}(x)$$

How to ensure that $\hat{\theta}^n \xrightarrow{\text{a.s.}} \bar{\theta}$ and $\sqrt{n}(\hat{\theta}^n - \bar{\theta}) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{M}^{-1}[\xi^*_D(\bar{\theta}), \bar{\theta}])$ when $n \to \infty$?

1. deterministic choice of x_k when $k \in \{k_1, k_2, \ldots\}$, with $k_i \sim i^{\alpha}$, $\alpha \in (1,2)$ [Lai 1994]
2. let n_0 tend to ∞ when $n \to \infty$
3. suppose that $\mathcal{X}$ is a finite set
   $\Rightarrow$ replication of observations

Sequential D-optimal design

For $n_0 \leq k < n$,

$$x_{k+1} = \arg\max_{x \in \mathcal{X}} \mathbf{f}^\top_{\hat{\theta}^k}(x) \mathbf{M}^{-1}(\xi_k, \hat{\theta}^k) \mathbf{f}_{\hat{\theta}^k}(x)$$

How to ensure that $\hat{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}$ and $\sqrt{n}(\hat{\theta}^n - \bar{\theta}) \stackrel{\text{d}}{\to} \mathcal{N}(\mathbf{0}, \mathbf{M}^{-1}[\xi^*_D(\bar{\theta}), \bar{\theta}])$ when $n \to \infty$?

1. deterministic choice of x_k when $k \in \{k_1, k_2, \ldots\}$, with $k_i \sim i^\alpha$, $\alpha \in (1,2)$ [Lai 1994]
2. let n_0 tend to ∞ when $n \to \infty$
3. suppose that $\mathcal{X}$ is a finite set
   $\Rightarrow$ replication of observations

$\mathcal{X}$ is a finite set $\mathcal{X} = \{x^{(1)}, \ldots, x^{(K)}\}$

Convergence of $\hat{\theta}^n$ to $\bar{\theta}$: everything is fine if $D_n(\theta,\bar{\theta}) = \sum_{i=1}^{n}[\eta(x_i,\theta) - \eta(x_i,\bar{\theta})]^2$ grows to ∞ fast enough for all $\theta \neq \bar{\theta}$

Theorem 1: convergence [LP, S&P Letters, 2009]

If $D_n(\theta,\bar{\theta}) = \sum_{i=1}^{n}[\eta(x_i,\theta) - \eta(x_i,\bar{\theta})]^2$ satisfies

$$\text{for all } \delta > 0 \,, \left[\inf_{\|\theta-\bar{\theta}\| \geq \delta/\tau_n} D_n(\theta,\bar{\theta}) \right] / (\log\log n) \stackrel{\text{a.s.}}{\rightarrow} \infty$$

with $\mathcal{X}$ finite and $\{\tau_n\}$ a non-decreasing sequence of positive constants, then $\hat{\theta}^n$ satisfies $\tau_n \|\hat{\theta}^n - \bar{\theta}\| \stackrel{\text{a.s.}}{\rightarrow} 0$

$\mathcal{X}$ is a finite set $\mathcal{X} = \{x^{(1)}, \ldots, x^{(K)}\}$

Convergence of $\hat{\theta}^n$ to $\bar{\theta}$: everything is fine if $D_n(\theta,\bar{\theta}) = \sum_{i=1}^n [\eta(x_i,\theta) - \eta(x_i,\bar{\theta})]^2$ grows to ∞ fast enough for all $\theta \neq \bar{\theta}$

Theorem 1: convergence [LP, S&P Letters, 2009]

If $D_n(\theta,\bar{\theta}) = \sum_{i=1}^n [\eta(x_i,\theta) - \eta(x_i,\bar{\theta})]^2$ satisfies

$$\text{for all } \ \delta > 0\,, \ \left[\inf_{\|\theta-\bar{\theta}\| \geq \delta/\tau_n} D_n(\theta,\bar{\theta}) \right] /(\log\log n) \stackrel{\text{a.s.}}{\to} \infty$$

with $\mathcal{X}$ finite and $\{\tau_n\}$ a non-decreasing sequence of positive constants, then $\hat{\theta}^n$ satisfies $\tau_n \|\hat{\theta}^n - \bar{\theta}\| \stackrel{\text{a.s.}}{\to} 0$

Theorem 2: asymptotic normality [LP, S&P Letters, 2009]

If there exists a sequence of matrices $\mathbf{C}_n$ symmetric pos. def. such that $\mathbf{C}_n^{-1}\mathbf{M}^{1/2}(\xi_n,\bar{\theta}) \stackrel{\text{p}}{\rightarrow} \mathbf{I}$
with $c_n = \lambda_{\min}(\mathbf{C}_n)$ and $D_n(\theta,\bar{\theta})$ satisfying $n^{1/4}c_n \rightarrow \infty$ and

$$\text{for all } \delta > 0\,, \left[\inf_{\|\theta-\bar{\theta}\|\geq c_n^2\delta} D_n(\theta,\bar{\theta})\right] / (\log\log n) \stackrel{\text{a.s.}}{\rightarrow} \infty$$

then $\hat{\theta}^n$ satisfies $\sqrt{n}\,\mathbf{M}^{1/2}(\xi_n,\hat{\theta}^n)(\hat{\theta}^n - \bar{\theta}) \stackrel{\text{d}}{\rightarrow} \mathcal{N}(\mathbf{0},\mathbf{I})$

☞ Apply that to

$$x_{k+1} = \arg\max_{x\in\mathcal{X}} \mathbf{f}_{\hat{\theta}^k}^\top(x)\mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k}(x)$$ with $\mathcal{X}$ finite

➠ for any sequence $\{\hat{\theta}^k\}$, the sampling rate of a non-singular design is > 0

Theorem 2: asymptotic normality [LP, S&P Letters, 2009]

If there exists a sequence of matrices $\mathbf{C}_n$ symmetric pos. def. such that $\mathbf{C}_n^{-1}\mathbf{M}^{1/2}(\xi_n,\bar{\theta}) \xrightarrow{\mathrm{p}} \mathbf{I}$
with $c_n = \lambda_{\min}(\mathbf{C}_n)$ and $D_n(\theta,\bar{\theta})$ satisfying $n^{1/4}c_n \to \infty$ and

$$\text{for all } \ \delta > 0\,, \ \left[\inf_{\|\theta-\bar{\theta}\|\geq c_n^2\delta} D_n(\theta,\bar{\theta})\right]/(\log\log n) \xrightarrow{\mathrm{a.s.}} \infty$$

then $\hat{\theta}^n$ satisfies $\sqrt{n}\,\mathbf{M}^{1/2}(\xi_n,\hat{\theta}^n)(\hat{\theta}^n - \bar{\theta}) \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0},\mathbf{I})$

☞ Apply that to

$$\boxed{x_{k+1} = \arg\max_{x\in\mathcal{X}} \mathbf{f}_{\hat{\theta}^k}^\top(x)\mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k}(x)}$$ with $\mathcal{X}$ finite

➟ for any sequence $\{\hat{\theta}^k\}$, the sampling rate of a non-singular design is > 0

$$x^{(1)} \quad x^{(2)} \quad \ldots \quad x^{(j)} \quad \ldots \quad x^{(l)} \quad \ldots \quad x^{(K)}$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\liminf_{n\to\infty} n_i/n > \alpha > 0}$$

➠ $\hat{\theta}^n \xrightarrow{\text{a.s.}} \bar{\theta}$

➠ $\mathbf{M}(\xi_n,\hat{\theta}^n) \xrightarrow{\text{a.s.}} \mathbf{M}[\xi_D^*(\bar{\theta}),\bar{\theta}]$

➠ $[n\mathbf{M}(\xi_n,\hat{\theta}^n)]^{1/2}(\hat{\theta}^n - \bar{\theta}) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0},\mathbf{I})$

$$\underbrace{x^{(1)} \quad x^{(2)} \quad \dots \quad x^{(j)} \quad \dots \quad x^{(l)} \quad \dots \quad x^{(K)}}_{\liminf_{n\to\infty} n_i/n > \alpha > 0}$$

- ➠ $\hat{\theta}^n \stackrel{\text{a.s.}}{\rightarrow} \bar{\theta}$
- ➠ $\mathbf{M}(\xi_n, \hat{\theta}^n) \stackrel{\text{a.s.}}{\rightarrow} \mathbf{M}[\xi_D^*(\bar{\theta}), \bar{\theta}]$
- ➠ $[n\mathbf{M}(\xi_n, \hat{\theta}^n)]^{1/2}(\hat{\theta}^n - \bar{\theta}) \stackrel{\text{d}}{\rightarrow} \mathcal{N}(\mathbf{0}, \mathbf{I})$

$$\underbrace{x^{(1)} \quad x^{(2)} \quad \ldots \quad x^{(j)} \quad \ldots \quad x^{(l)} \quad \ldots \quad x^{(K)}}_{\liminf_{n\to\infty} n_i/n > \alpha > 0}$$

- $\hat{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}$
- $\mathbf{M}(\xi_n, \hat{\theta}^n) \stackrel{\text{a.s.}}{\to} \mathbf{M}[\xi_D^*(\bar{\theta}), \bar{\theta}]$
- $[n\mathbf{M}(\xi_n, \hat{\theta}^n)]^{1/2}(\hat{\theta}^n - \bar{\theta}) \stackrel{\text{d}}{\to} \mathcal{N}(\mathbf{0}, \mathbf{I})$

$x^{(1)} \quad x^{(2)} \quad \ldots \quad x^{(j)} \quad \ldots \quad x^{(l)} \quad \ldots \quad x^{(K)}$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\liminf_{n\to\infty} n_i/n > \alpha > 0}$$

- ➠ $\hat{\theta}^n \overset{\text{a.s.}}{\to} \bar{\theta}$
- ➠ $\mathbf{M}(\xi_n, \hat{\theta}^n) \overset{\text{a.s.}}{\to} \mathbf{M}[\xi_D^*(\bar{\theta}), \bar{\theta}]$
- ➠ $[n\mathbf{M}(\xi_n, \hat{\theta}^n)]^{1/2}(\hat{\theta}^n - \bar{\theta}) \overset{\text{d}}{\to} \mathcal{N}(\mathbf{0}, \mathbf{I})$

4) Sequential penalized D-optimal design

$$x_{k+1} = \arg\max_{x\in\mathcal{X}} \left\{ \mathbf{f}_{\hat{\theta}^k}^\top \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\, \phi(x,\hat{\theta}^k) \right\}$$

Again, sequential construction $\Rightarrow$ independency is lost

☞ Use the assumption that $\mathcal{X}$ is finite: $\mathcal{X} = \{x^{(1)},\ldots,x^{(K)}\}$

- If λ_k = constant λ:
  ➠ $\hat{\theta}^n \overset{\text{a.s.}}{\to} \bar{\theta}$ and $\mathbf{M}(\xi_n,\bar{\theta}) \to \mathbf{M}^*(\bar{\theta})$
  (optimal for criterion $\log\det\mathbf{M}(\xi,\bar{\theta}) - \lambda\,\Phi(\xi,\bar{\theta})$) and
  $\sqrt{n}\mathbf{M}^{1/2}(\xi_n,\hat{\theta}^n)(\hat{\theta}^n - \bar{\theta}) \overset{\text{d}}{\to} \mathcal{N}(\mathbf{0},\mathbf{I})$
- Also true if
  λ_k = bounded measurable function of $x_1,Y_1,\ldots,x_k,Y_k$
  (e.g., $\lambda_k = \lambda^*(\hat{\theta}^k)$ = optimal Lagrange coefficient for minimization of $\log\det\mathbf{M}(\xi,\hat{\theta}^k)$ under the constraint $\Phi(\xi,\hat{\theta}^k) \le C$)

4) Sequential penalized D-optimal design

$$x_{k+1} = \arg\max_{x \in \mathcal{X}} \left\{ \mathbf{f}_{\hat{\theta}^k}^\top \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\, \phi(x,\hat{\theta}^k) \right\}$$

Again, sequential construction ⇒ independency is lost

☞ Use the assumption that $\mathcal{X}$ is finite: $\mathcal{X} = \{x^{(1)},\ldots,x^{(K)}\}$

- If λ_k = constant λ :
  ➠ $\hat{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}$ and $\mathbf{M}(\xi_n,\bar{\theta}) \to \mathbf{M}^*(\bar{\theta})$
  (optimal for criterion $\log\det \mathbf{M}(\xi,\bar{\theta}) - \lambda\, \Phi(\xi,\bar{\theta})$) and
  $\sqrt{n}\mathbf{M}^{1/2}(\xi_n,\hat{\theta}^n)(\hat{\theta}^n - \bar{\theta}) \stackrel{\text{d}}{\to} \mathcal{N}(\mathbf{0},\mathbf{I})$
- Also true if
  λ_k = bounded measurable function of $x_1,Y_1,\ldots,x_k,Y_k$
  (e.g., $\lambda_k = \lambda^*(\hat{\theta}^k)$ = optimal Lagrange coefficient for minimization of $\log\det \mathbf{M}(\xi,\hat{\theta}^k)$ under the constraint $\Phi(\xi,\hat{\theta}^k) \leq C$)

4) Sequential penalized D-optimal design

$$x_{k+1} = \arg\max_{x \in \mathcal{X}} \left\{ \mathbf{f}_{\hat{\theta}^k}^\top \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\, \phi(x,\hat{\theta}^k) \right\}$$

Again, sequential construction ⇒ independency is lost

☞ Use the assumption that $\mathcal{X}$ is finite: $\mathcal{X} = \{x^{(1)},\ldots,x^{(K)}\}$

- If λ_k = constant λ :
  ➠ $\hat{\theta}^n \overset{\text{a.s.}}{\rightarrow} \bar{\theta}$ and $\mathbf{M}(\xi_n,\bar{\theta}) \rightarrow \mathbf{M}^*(\bar{\theta})$
  (optimal for criterion $\log\det \mathbf{M}(\xi,\bar{\theta}) - \lambda\, \Phi(\xi,\bar{\theta})$) and
  $\sqrt{n}\mathbf{M}^{1/2}(\xi_n,\hat{\theta}^n)(\hat{\theta}^n - \bar{\theta}) \overset{\text{d}}{\rightarrow} \mathcal{N}(\mathbf{0},\mathbf{I})$
- Also true if
  λ_k = bounded measurable function of $x_1,Y_1,\ldots,x_k,Y_k$
  (e.g., $\lambda_k = \lambda^*(\hat{\theta}^k)$ = optimal Lagrange coefficient for minimization of $\log\det \mathbf{M}(\xi,\hat{\theta}^k)$ under the constraint $\Phi(\xi,\hat{\theta}^k) \leq C$)

- If $\lambda_k \nearrow \infty$, $(\lambda_k \log\log k)/k \to 0$, then $\boxed{\hat{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}}$

  moreover, convergence to minimum-cost design:

  $$\boxed{\Phi(\xi_n,\bar{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \phi(x_i,\bar{\theta}) \stackrel{\text{a.s.}}{\to} \phi_{\bar{\theta}}^* = \min_{x\in\mathcal{X}} \phi(x,\bar{\theta})}$$

  ... and $\boxed{\xi_N(x^{(i^*)}) \stackrel{\text{a.s.}}{\to} 1}$ if $\phi(x,\bar{\theta})$ has a unique minimum at $x^{(i^*)} \in \mathcal{X} = \{x^{(1)},\ldots,x^{(K)}\}$

$\Rightarrow$ We can thus optimize $\sum_{i=1}^{n} \phi(x_i,\bar{\theta})$ without knowing $\bar{\theta}$: self-tuning optimization

$$\boxed{x_{k+1} = \arg\max_{x\in\mathcal{X}} \left\{ \mathbf{f}_{\hat{\theta}^k}^\top \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\, \phi(x,\hat{\theta}^k) \right\}}$$

Already suggested for linear regression ($\eta(x,\theta)$ linear in θ) [Åström & Wittenmark, 1989], condition on λ_k in [LP, AS 2000]

Here, LS in nonlinear regression, or ML, but with $\mathcal{X}$ finite

Beware: $x_{k+1} = \arg\min_{x\in\mathcal{X}} \phi(x,\hat{\theta}^k)$ may not work!

- If $\lambda_k \nearrow \infty$, $(\lambda_k \log\log k)/k \to 0$, then $\boxed{\hat{\theta}^n \overset{\text{a.s.}}{\to} \bar{\theta}}$
  moreover, convergence to minimum-cost design:

  $$\boxed{\Phi(\xi_n,\bar{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \phi(x_i,\bar{\theta}) \overset{\text{a.s.}}{\to} \phi^*_{\bar{\theta}} = \min_{x\in\mathcal{X}} \phi(x,\bar{\theta})}$$

  ...and $\boxed{\xi_N(x^{(i^*)}) \overset{\text{a.s.}}{\to} 1}$ if $\phi(x,\bar{\theta})$ has a unique minimum at $x^{(i^*)} \in \mathcal{X} = \{x^{(1)},\ldots,x^{(K)}\}$

$\Rightarrow$ We can thus optimize $\sum_{i=1}^{n} \phi(x_i,\bar{\theta})$ without knowing $\bar{\theta}$: **self-tuning optimization**

$$\boxed{x_{k+1} = \arg\max_{x\in\mathcal{X}} \left\{ \mathbf{f}^\top_{\hat{\theta}^k} \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\, \phi(x,\hat{\theta}^k) \right\}}$$

Already suggested for linear regression ($\eta(x,\theta)$ linear in θ) [Åström & Wittenmark, 1989], condition on λ_k in [LP, AS 2000]

Here, LS in nonlinear regression, or ML, but with $\mathcal{X}$ finite

Beware: $x_{k+1} = \arg\min_{x\in\mathcal{X}} \phi(x,\hat{\theta}^k)$ may not work!

- If $\lambda_k \nearrow \infty$, $(\lambda_k \log\log k)/k \to 0$, then $\boxed{\hat{\theta}^n \stackrel{\text{a.s.}}{\to} \bar{\theta}}$
  moreover, convergence to minimum-cost design:
  $$\boxed{\Phi(\xi_n,\bar{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \phi(x_i,\bar{\theta}) \stackrel{\text{a.s.}}{\to} \phi^*_{\bar{\theta}} = \min_{x\in\mathcal{X}} \phi(x,\bar{\theta})}$$
  ... and $\boxed{\xi_N(x^{(i^*)}) \stackrel{\text{a.s.}}{\to} 1}$ if $\phi(x,\bar{\theta})$ has a unique minimum at $x^{(i^*)} \in \mathcal{X} = \{x^{(1)},\ldots,x^{(K)}\}$

$\Rightarrow$ We can thus optimize $\sum_{i=1}^{n} \phi(x_i,\bar{\theta})$ without knowing $\bar{\theta}$: self-tuning optimization

$$\boxed{x_{k+1} = \arg\max_{x\in\mathcal{X}} \left\{ \mathbf{f}_{\hat{\theta}^k}^\top \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\, \phi(x,\hat{\theta}^k) \right\}}$$

Already suggested for linear regression ($\eta(x,\theta)$ linear in θ) [Åström & Wittenmark, 1989], condition on λ_k in [LP, AS 2000]

Here, LS in nonlinear regression, or ML, but with $\mathcal{X}$ finite

Beware: $x_{k+1} = \arg\min_{x\in\mathcal{X}} \phi(x,\hat{\theta}^k)$ may not work!

Example 2 (self-tuning regulation)

- Observe $Y_i = \frac{\bar{\theta}_1 x}{\bar{\theta}_2 + x} + \varepsilon_i$, $\{\varepsilon_i\}$ i.i.d. $\mathcal{N}(0,0.1)$
- Find x^* such that $\Psi(x,\bar{\theta}) = T$

$$\Psi(x,\theta) = \theta_1[1 - \exp(-\theta_2 x/3)] \neq \eta(x,\bar{\theta})$$

$(\bar{\theta} = (1,1)^\top \Rightarrow \Psi(x^*,\bar{\theta}) = 1/2$ for $x^* = 3\log(2) \simeq 2.08)$

- $\rightarrow$ minimise $\phi(x,\theta) = [\Psi(x,\theta) - T]^2$ — note that we do not observe $\Psi(x,\bar{\theta})$!
- $x_1 = 1$, $x_2 = 10$, then $x_{k+1} = \arg\max_{x \in \mathcal{X}} \left\{ \mathbf{f}_{\hat{\theta}^k}^\top \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k \, \phi(x,\hat{\theta}^k) \right\}$ for three sequences $\{\lambda_k\}$
  - (a) $\lambda_k = \log^2 k$
  - (b) $\lambda_k = k/(1 + \log^2 k)$
  - (c) $\lambda_k = k^{1.1}$

Example 2 (self-tuning regulation)

- Observe $Y_i = \frac{\bar{\theta}_1 x}{\bar{\theta}_2 + x} + \varepsilon_i$, $\{\varepsilon_i\}$ i.i.d. $\mathcal{N}(0,0.1)$
- Find x^* such that $\Psi(x,\bar{\theta}) = T$

$$\Psi(x,\theta) = \theta_1[1 - \exp(-\theta_2 x/3)] \neq \eta(x,\bar{\theta})$$

$(\bar{\theta} = (1,1)^\top \Rightarrow \Psi(x^*,\bar{\theta}) = 1/2$ for $x^* = 3\log(2) \simeq 2.08)$

- $\to$ minimise $\phi(x,\theta) = [\Psi(x,\theta) - T]^2$ — note that we do not observe $\Psi(x,\bar{\theta})$!
- $x_1 = 1$, $x_2 = 10$, then $x_{k+1} = \arg\max_{x \in \mathcal{X}} \left\{ \mathbf{f}_{\hat{\theta}^k}^\top \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\, \phi(x,\hat{\theta}^k) \right\}$ for three sequences $\{\lambda_k\}$
  - (a) $\lambda_k = \log^2 k$
  - (b) $\lambda_k = k/(1 + \log^2 k)$
  - (c) $\lambda_k = k^{1.1}$

5) Examples

Example 2 (self-tuning regulation)

- Observe $Y_i = \frac{\bar{\theta}_1 x}{\bar{\theta}_2 + x} + \varepsilon_i$, $\{\varepsilon_i\}$ i.i.d. $\mathcal{N}(0,0.1)$
- Find x^* such that $\Psi(x,\bar{\theta}) = T$

$$\Psi(x,\theta) = \theta_1[1 - \exp(-\theta_2 x/3)] \neq \eta(x,\bar{\theta})$$

$(\bar{\theta} = (1,1)^\top \Rightarrow \Psi(x^*,\bar{\theta}) = 1/2$ for $x^* = 3\log(2) \simeq 2.08)$

- $\rightarrow$ minimise $\phi(x,\theta) = [\Psi(x,\theta) - T]^2$ — note that we do not observe $\Psi(x,\bar{\theta})$!
- $x_1 = 1$, $x_2 = 10$, then $x_{k+1} = \arg\max_{x \in \mathcal{X}} \left\{ \mathbf{f}_{\hat{\theta}^k}^\top \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\, \phi(x,\hat{\theta}^k) \right\}$ for three sequences $\{\lambda_k\}$
  - (a) $\lambda_k = \log^2 k$
  - (b) $\lambda_k = k/(1 + \log^2 k)$
  - (c) $\lambda_k = k^{1.1}$

Example 2 (self-tuning regulation)

- Observe $Y_i = \frac{\bar{\theta}_1 x}{\bar{\theta}_2 + x} + \varepsilon_i$, $\{\varepsilon_i\}$ i.i.d. $\mathcal{N}(0,0.1)$
- Find x^* such that $\Psi(x,\bar{\theta}) = T$

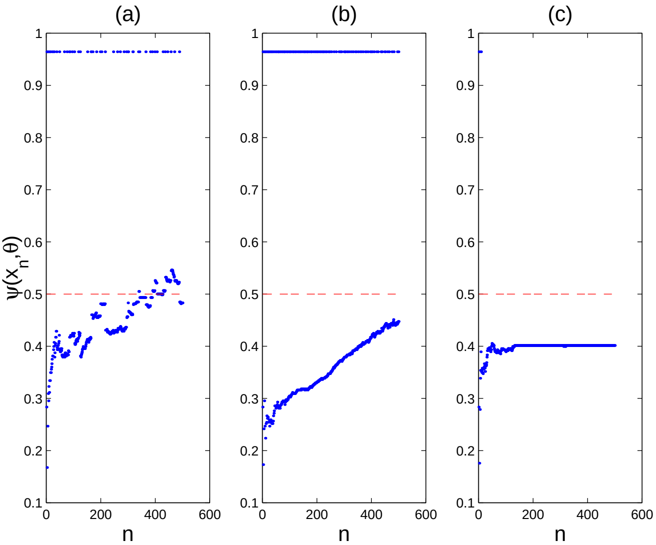
$$\Psi(x,\theta) = \theta_1[1 - \exp(-\theta_2\, x/3)] \neq \eta(x,\bar{\theta})$$

$(\bar{\theta} = (1,1)^\top \Rightarrow \Psi(x^*,\bar{\theta}) = 1/2$ for $x^* = 3\log(2) \simeq 2.08)$

- $\to$ minimise $\phi(x,\theta) = [\Psi(x,\theta) - T]^2$ — note that we do not observe $\Psi(x,\bar{\theta})$!
- $x_1 = 1$, $x_2 = 10$, then $x_{k+1} = \arg\max_{x\in\mathcal{X}} \left\{ \mathbf{f}_{\hat{\theta}^k}^\top \mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\, \phi(x,\hat{\theta}^k) \right\}$ for three sequences $\{\lambda_k\}$
  - (a) $\lambda_k = \log^2 k$
  - (b) $\lambda_k = k/(1+\log^2 k)$
  - (c) $\lambda_k = k^{1.1}$

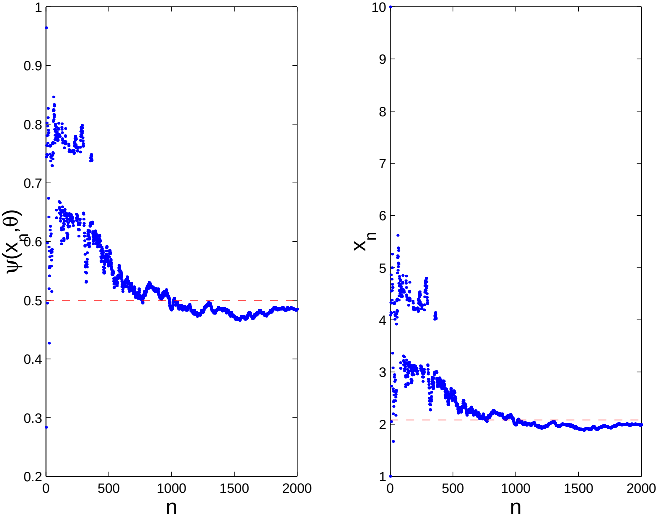
$\rightarrow \Psi(x_k, \bar{\theta}),\ k = 1, \ldots, 500$

(a) $\lambda_k = \log^2 k$ (b) $\lambda_k = k/(1 + \log^2 k)$ (c) $\lambda_k = k^{1.1}$



Replace $\phi(x,\theta) = [\psi(x,\theta) - T]^2$ by $\phi(x,\theta) = [\psi(x,\theta) - T]^4$, take $\lambda_k = 10^3 \log^2 k$

$\rightarrow$ the support points of ξ_k tend to concentrate around x^*



Example 1 [Dragalin & Fedorov 2006] (continued)

Clinical trials: 11 doses, Y for efficacy, Z for toxicity, 36 patients

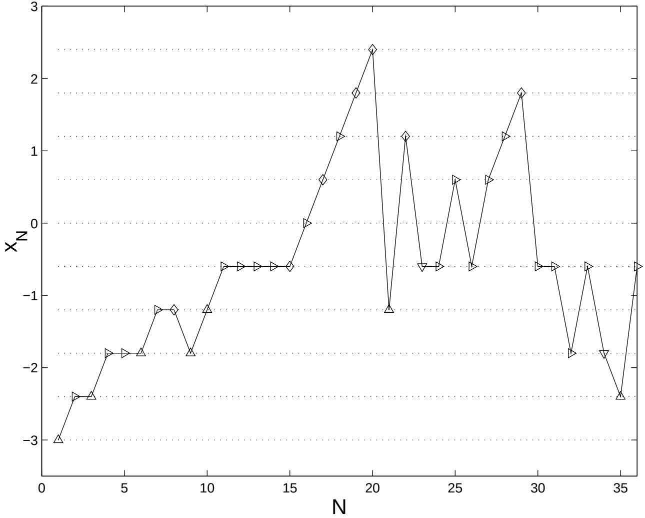
up and down method [Ivanova 2003]

$$x_{N+1} = \begin{cases} \max\{x^{(i_N-1)}, x^{(1)}\} & \searrow & \text{si } Z_N = 1\,, \\ x^{(i_N)} & \longrightarrow & \text{si } Y_N = 1 \text{ and } Z_N = 0\,, \\ \min\{x^{(i_N+1)}, x^{(11)}\} & \nearrow & \text{si } Y_N = 0 \text{ and } Z_N = 0\,, \end{cases}$$

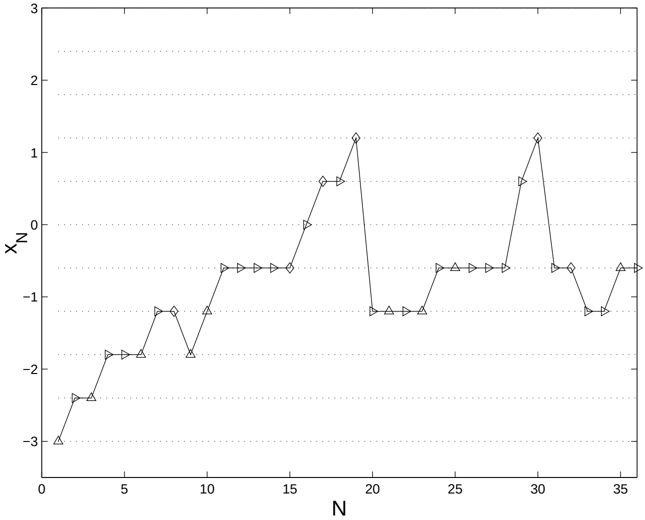
for the first 10 patients, then sequential penalized D-optimal design — switching at first observed toxicity (with maximum increase of one dose at each step)[Dragalin & Fedorov 2006])

$\triangle$ for $(Y = 0, Z = 0)$, $\triangleright$ for $(Y = 1, Z = 0)$, $\diamond$ for $(Y = 1, Z = 1)$ and $\triangledown$ for $(Y = 0, Z = 1)$

Penalty function $\phi_1(x,\theta) = \pi_{10}^{-1}(x,\theta)$



Penalty function $\phi_3(x,\theta) = \pi_{10}^{-1}(x,\theta)[1 - \pi_{\cdot 1}(x,\theta)]^{-1}$



(36 patients, 1000 repetitions)

design	$\Phi_1(\xi,\theta)$	$J(\xi,\theta)$	$\widehat{x^*}_{\{t<4\}}$	$\widehat{x^*}_{\{t=4\}}$	$\widehat{x^*}_{\{t=5\}}$	$\widehat{x^*}_{\{t=6\}}$	$\widehat{x^*}_{\{t>6\}}$	$\#x^{(11)}$
S1	1.87	28.02	2%	38.6%	36.9%	8.6%	13.9%	0
$\xi_{u\&d}(\bar{\theta})$	1.47	29.4						
S2	3.16	17.23	0	19.8%	70.5%	7.8%	1.9%	5%
$\xi^*_D(\bar{\theta})$	4.45	14.99						
S3	2.38	18.78	0	22.3%	68.2%	7%	2.5%	2.3%
$\xi^*_{\lambda=2}(\bar{\theta})$	1.97	17.00						

$$\phi_1(\xi,\theta) = \pi_{10}^{-1}(x,\theta),\ J(\xi,\theta) = \det{}^{-1/6}[\mathbf{M}(\xi,\theta)],\ \text{OSD} = x^{(5)}$$

S1 = up & down

S2 = up & down + sequential *D*-optimal

S3 = up & down + sequential penalized *D*-optimal

(240 patients, 150 repetitions)

	$\Phi_1(\xi,\theta)$	$J(\xi,\theta)$	$\widehat{x^*}_{\{t<4\}}$	$\widehat{x^*}_{\{t=4\}}$	$\widehat{x^*}_{\{t=5\}}$	$\widehat{x^*}_{\{t=6\}}$	$\widehat{x^*}_{\{t>6\}}$	$\#x^{(11)}$
S1	1.54	29.04	0	14%	77.3%	7.3%	1.3%	0
$\xi_{u\&d}(\bar{\theta})$	1.47	29.4						
S4	1.52	27.87	0	8.7%	90%	0.7%	0.7%	0.1%

S4 = up & down + sequential penalized D-optimal with logarithmic increase $\lambda_N \nearrow$

- switching when σ(optimal estimated dose) $< \Delta_x$

  Δ_x = interval between 2 consecutive doses
- cautious increase of doses when σ(optimal estimated dose) $> \Delta_x/2$

S4 better than S1 (= up & down)

6) Conclusions

- Convergence and asymptotic normality of estimators under sequential *D*-optimal design

$$x_{k+1} = \arg\max_{x\in\mathcal{X}} \mathbf{f}^\top_{\hat\theta^k}(x)\mathbf{M}^{-1}(\xi_k,\hat\theta^k)\mathbf{f}_{\hat\theta^k}(x)$$

and sequential penalized *D*-optimal design

$$x_{k+1} = \arg\max_{x\in\mathcal{X}} \left\{\mathbf{f}^\top_{\hat\theta^k}\mathbf{M}^{-1}(\xi_k,\hat\theta^k)\mathbf{f}_{\hat\theta^k} - \lambda_k\,\phi(x,\hat\theta^k)\right\}$$

when λ_k bounded

...assuming that $\mathcal{X}$ is finite (only a technical assumption?)

- $\lambda_n \to \infty$ not too fast $\to$ self-tuning regulation/optimization

6) Conclusions

- Convergence and asymptotic normality of estimators under sequential *D*-optimal design

  $$x_{k+1} = \arg\max_{x \in \mathcal{X}} \mathbf{f}^{\top}_{\hat{\theta}^k}(x)\mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k}(x)$$

  and sequential penalized *D*-optimal design

  $$x_{k+1} = \arg\max_{x \in \mathcal{X}} \left\{\mathbf{f}^{\top}_{\hat{\theta}^k}\mathbf{M}^{-1}(\xi_k,\hat{\theta}^k)\mathbf{f}_{\hat{\theta}^k} - \lambda_k\,\phi(x,\hat{\theta}^k)\right\}$$

  when λ_k bounded

  ...assuming that $\mathcal{X}$ is finite (only a technical assumption?)
- $\lambda_n \to \infty$ not too fast $\to$ self-tuning regulation/optimization

- Asymptotic normality when $\lambda_n \to \infty$ slowly enough?
  $\lambda_{\min}[\mathbf{M}(\xi_n,\hat{\theta}^n)] \sim A/\lambda_n$?
  Find a sequence of matrices $\mathbf{C}_n$ symmetric pos. def. such that $\mathbf{C}_n^{-1}\mathbf{M}^{1/2}(\xi_n,\bar{\theta}) \stackrel{\mathrm{p}}{\to} \mathbf{I}$:
  - construct $\mathbf{C}_n$ from
    $x_{k+1} = \arg\max_{x\in\mathcal{X}} \left\{\mathbf{f}_{\bar{\theta}}^\top \mathbf{M}^{-1}(\xi_k,\bar{\theta})\mathbf{f}_{\bar{\theta}} - \lambda_k\, \phi(x,\bar{\theta})\right\}$?
  - use $\mathbf{C}_n = \mathbf{M}^{1/2}(\xi^*(\bar{\theta},\lambda_n),\bar{\theta})$ with $\xi^*(\theta,\lambda)$ maximizing $\log\det\mathbf{M}(\xi,\theta) - \lambda\Phi(\xi,\theta)$?

Thank you for your attention!