Planification d'expériences numériques : quelques tendances et questions ouvertes

Luc Pronzato

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

Outline

Possible objectives

Computer experiment = simulations

$x \in \mathbb{R}^d \rightarrow$ observation $\mathscr{Y}(x)$ (real phenomenon, physical system)

Numerical simulation: $x \rightarrow Y(x) = f(x)$

Pairs $(X_i, f(X_i))$, $i = 1, 2, \ldots, n \rightarrow$ approximation $\eta_n(\cdot)$ of $f(\cdot)$ ($\rightarrow$ epistemic uncertainty – due to simulator, model $\eta_n(\cdot)$, finite data set…)

Possible objectives

Computer experiment = simulations

$x \in \mathbb{R}^d \rightarrow$ observation $\mathscr{Y}(x)$ (real phenomenon, physical system)
Numerical simulation: $x \rightarrow Y(x) = f(x)$
Pairs $(X_i, f(X_i))$, $i = 1, 2, \ldots, n \rightarrow$ approximation $\eta_n(\cdot)$ of $f(\cdot)$ ($\rightarrow$ epistemic uncertainty – due to simulator, model $\eta_n(\cdot)$, finite data set. . .)

- optimization: find $x^* = \arg\max_{x\in\mathscr{X}} f(x)$ (hopefully close to $\arg\max_{x\in\mathscr{X}} \mathscr{Y}(x)$)
- inversion: reconstruct $\{x \in \mathscr{X} : f(x) = T\}$
- estimation of failure prob. $\text{Prob}\{f(x) > C\}$ when $x \sim \phi(\cdot)$ (intrinsic uncertainty due to input variability)
- sensitivity analysis (functional variance analysis)
- approximation/interpolation

Luc Pronzato

Objective = approximation/interpolation

$f(x)$ some unknown function defined for $x \in \mathscr{X} \subset \mathbb{R}^d$
construct a 'good' approximation $\eta_n(\cdot)$ of $f(\cdot)$ over $\mathscr{X}$ from pairs $(X_i, f(X_i))$, $i = 1, 2, \ldots, n$ (n not necessarily fixed beforehand)

Since $f(\cdot)$ is unknown $\rightarrow$ put n points $\mathbf{X}_n = (X_1, \ldots, X_n)$ in $\mathscr{X}$ as dispersed as possible (can be justified precisely [Biedermann & Dette, 2001])

Luc Pronzato

Classical measures of dispersion (often $\mathscr{X} = [0,1]^d$)

① miniMax distance: minimize
$\Phi_{mM}(\mathbf{X}_n) = \max_{x \in \mathscr{X}} \min_i \|x - X_i\|$
$d = 1 \Leftrightarrow X_i = (2i-1)/(2n),\ i = 1, \ldots, n$
$\Rightarrow \Phi^*_{mM} = 1/(2n)$
$d > 1 \Leftrightarrow$ sphere-covering

Luc Pronzato

Workshop de clôture ANR OPUS 21/10/2011

miniMax and Maximin

Classical measures of dispersion (often $\mathscr{X} = [0,1]^d$)

① miniMax distance: minimize

$$\Phi_{mM}(\mathbf{X}_n) = \max_{x \in \mathscr{X}} \min_i \|x - X_i\|$$

$d = 1 \Leftrightarrow X_i = (2i-1)/(2n),\ i = 1,\ldots,n$

$\Rightarrow \Phi^*_{mM} = 1/(2n)$

$d > 1 \Leftrightarrow$ sphere-covering

② Maximin distance: maximize

$$\Phi_{Mm}(\mathbf{X}_n) = \min_{i \neq j} d_{ij} = \min_{i \neq j} \|X_i - X_j\|$$

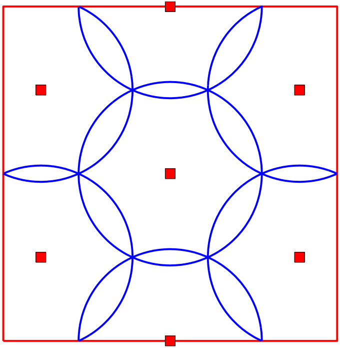
$d = 1 \Leftrightarrow X_i = (i-1)/(n-1),\ i = 1,\ldots,n$

$\Rightarrow \Phi^*_{Mm} = 1/(n-1)$

$d > 1 \Leftrightarrow$ sphere-packing

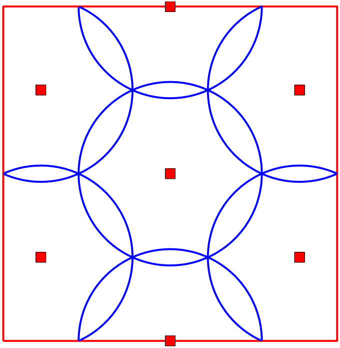
Luc Pronzato

① miniMax $d = 2, n = 7$ (radius=$\phi_{mM}(\mathbf{X})$)

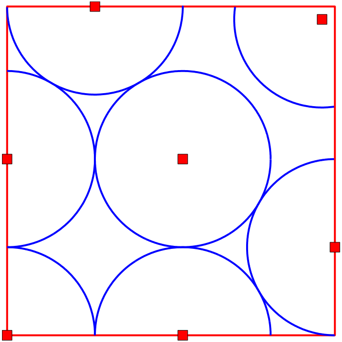


miniMax and Maximin (2)

① miniMax $d = 2, n = 7$
(radius=$\phi_{mM}(\mathbf{X})$)



② Maximin $d = 2, n = 7$
(radius=$\phi_{Mm}(\mathbf{X})/2$)



① miniMax criterion Φ_{mM}

➠ Φ_{mM} has nice properties in terms of approximation theory, but is difficult to compute

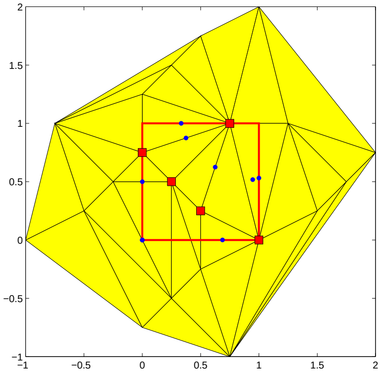
Possible evaluation via Delaunay triangulation (tessellation)

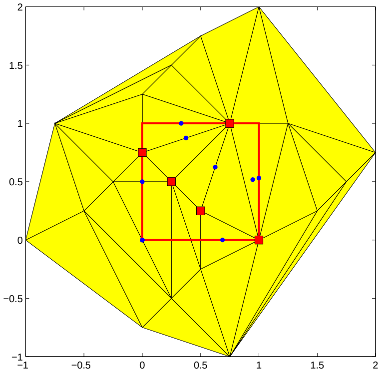
① miniMax criterion Φ_{mM}

➠ Φ_{mM} has nice properties in terms of approximation theory, but is difficult to compute
Possible evaluation via Delaunay triangulation (tessellation)

- $\mathbf{X}_n$ ($= n$ points in $\mathscr{X} = [0,1]^d$) $\rightarrow$ take all symmetric points w.r.t. $2d$ faces of $\mathscr{X}$
- Compute the Delauany triangulation (tessellation)
- candidates for $\arg\max_{x\in\mathscr{X}} \min_i \|x - X_i\|$ are centers of a circumscribed spheres

miniMax with Delaunay tessellation: 5-point Lh



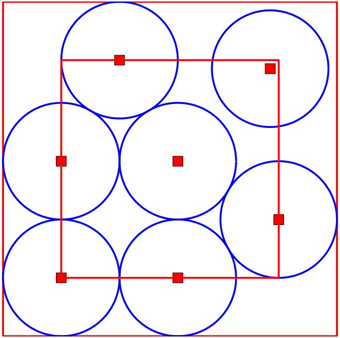


$\Phi_{Mm}(\mathbf{X})$ can be calculated... but remains computationally costly: up to $M^{\lceil d/2 \rceil}$ simplices (and circumscribed spheres) with $M = (2d+1)n \rightarrow$ computing time $\mathcal{O}(M \log M + M^{\lceil d/2 \rceil})$

② Maximin criterion Φ_{Mm}

Relation between Maximin-optimal design and sphere packing:

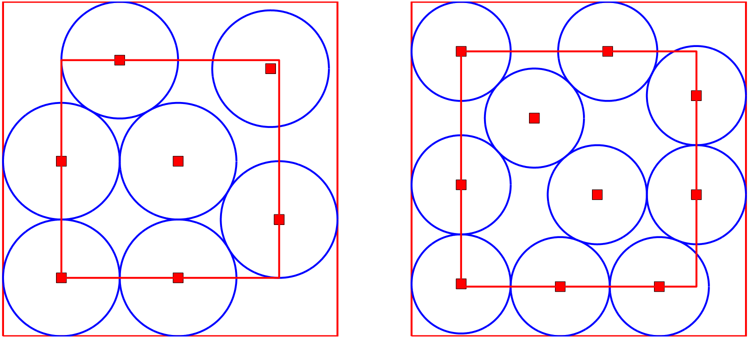
$d = 2, n = 7$



② Maximin criterion Φ_{Mm}

Relation between Maximin-optimal design and sphere packing:

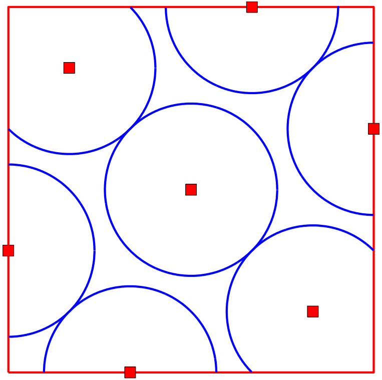
$d = 2, n = 7$ $\qquad$ $d = 2, n = 10$



http://www.packomania.com/

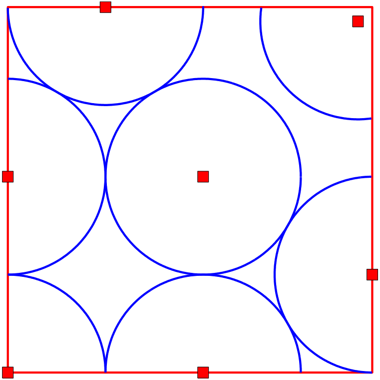
To ensure good projective properties along all axes: make each one-dimensional projection Maximin-optimal ($X_i = \frac{i-1}{n-1}$) ($\rightarrow$ finite set, with $(n!)^{d-1}$ elements)

Maximin Lh ($d = 2, n = 7$, radius=$\phi_{Mm}(\mathbf{X})/2$)



To ensure good projective properties along all axes: make each one-dimensional projection Maximin-optimal ($X_i = \frac{i-1}{n-1}$) ($\rightarrow$ finite set, with $(n!)^{d-1}$ elements)

Maximin, not Lh, ($d = 2, n = 7$, radius=$\phi_{Mm}(\mathbf{X})/2$)



2 Regularized Maximin

Define $d_{ij} = \|X_i - X_j\|$, so that $\Phi_{Mm}(\mathbf{X}) = \min_{i \neq j} d_{ij}$

$$\underline{\phi}_{[q]}(\mathbf{X}) = \left[\sum_{i<j} d_{ij}^{-q} \right]^{-1/q} \quad \text{and} \quad \overline{\phi}_{[q]}(\mathbf{X}) = \left[\sum_{i<j} \mu_{ij}\, d_{ij}^{-q} \right]^{-1/q}$$

with $\mu_{ij} > 0$ and $\sum_{i<j} \mu_{ij} = 1$

2 Regularized Maximin

Define $d_{ij} = \|X_i - X_j\|$, so that $\Phi_{Mm}(\mathbf{X}) = \min_{i \neq j} d_{ij}$

$$\underline{\phi}_{[q]}(\mathbf{X}) = \left[\sum_{i<j} d_{ij}^{-q}\right]^{-1/q} \text{ and } \overline{\phi}_{[q]}(\mathbf{X}) = \left[\sum_{i<j} \mu_{ij}\, d_{ij}^{-q}\right]^{-1/q}$$

with $\mu_{ij} > 0$ and $\sum_{i<j} \mu_{ij} = 1$

Then

$$\boxed{\underline{\phi}_{[q]}(\mathbf{X}) \leq \phi_{Mm}(\mathbf{X}) \leq \overline{\phi}_{[q]}(\mathbf{X}) \leq \underline{\mu}^{-1/q}\, \underline{\phi}_{[q]}(\mathbf{X})}\,,\ q > 0\,,$$

with $\underline{\mu} = \min_{i<j} \mu_{ij}$ (convergence monotonic in q from both sides as $q \to \infty$)

2 Regularized Maximin

Define $d_{ij} = \|X_i - X_j\|$, so that $\Phi_{Mm}(\mathbf{X}) = \min_{i\neq j} d_{ij}$

$$\underline{\phi}_{[q]}(\mathbf{X}) = \left[\sum_{i<j} d_{ij}^{-q}\right]^{-1/q} \text{ and } \overline{\phi}_{[q]}(\mathbf{X}) = \left[\sum_{i<j} \mu_{ij}\, d_{ij}^{-q}\right]^{-1/q}$$

with $\mu_{ij} > 0$ and $\sum_{i<j} \mu_{ij} = 1$

Then

$$\boxed{\underline{\phi}_{[q]}(\mathbf{X}) \leq \phi_{Mm}(\mathbf{X}) \leq \overline{\phi}_{[q]}(\mathbf{X}) \leq \underline{\mu}^{-1/q}\, \underline{\phi}_{[q]}(\mathbf{X})}\,,\ q > 0\,,$$

with $\underline{\mu} = \min_{i<j} \mu_{ij}$ (convergence monotonic in q from both sides as $q \to \infty$)

By continuity: $\overline{\phi}_{[0]}(\mathbf{X}) = \exp\left[\sum_{i<j} \mu_{ij} \log(d_{ij})\right]$

Regularized Φ_{Mm}

μ = uniform measure ($\mu_{ij} = \underline{\mu} = \binom{n}{2}^{-1}$ for all $i < j$) $\Rightarrow$

$$\frac{\phi_{Mm}(\underline{\mathbf{X}}^*_{[q]})}{\phi^*_{Mm}} \geq \binom{n}{2}^{-1/q},$$

with $\underline{\mathbf{X}}^*_{[q]}$ optimal for $\underline{\phi}_{[q]}$

(Maximin efficiency $> 1 - \epsilon$ for $q > \frac{2\log(n)}{\epsilon}$)

Regularized Φ_{Mm}

μ = uniform measure ($\mu_{ij} = \underline{\mu} = \binom{n}{2}^{-1}$ for all $i < j$) $\Rightarrow$

$$\frac{\phi_{Mm}(\underline{\mathbf{X}}^*_{[q]})}{\phi^*_{Mm}} \geq \binom{n}{2}^{-1/q},$$

with $\underline{\mathbf{X}}^*_{[q]}$ optimal for $\underline{\phi}_{[q]}$

(Maximin efficiency $> 1 - \epsilon$ for $q > \frac{2\log(n)}{\epsilon}$)

$q = 2 \rightarrow$ "Energy criterion" of [Audze & Eglais, 1977]

for $q \lesssim 5$ easier optimization (Lh designs) than for ϕ_{Mm} [Morris & Mitchell, 1995]

☞ use the smallest q such that the optimum designs coincide

Regularized Φ_{Mm}

μ = uniform measure ($\mu_{ij} = \underline{\mu} = \binom{n}{2}^{-1}$ for all $i < j$) $\Rightarrow$

$$\frac{\phi_{Mm}(\underline{\mathbf{X}}^*_{[q]})}{\phi^*_{Mm}} \geq \binom{n}{2}^{-1/q},$$

with $\underline{\mathbf{X}}^*_{[q]}$ optimal for $\underline{\phi}_{[q]}$

(Maximin efficiency $> 1 - \epsilon$ for $q > \frac{2\log(n)}{\epsilon}$)

$q = 2 \rightarrow$ "Energy criterion" of [Audze & Eglais, 1977]

for $q \lesssim 5$ easier optimization (Lh designs) than for ϕ_{Mm} [Morris & Mitchell, 1995]

☞ use the smallest q such that the optimum designs coincide

Continuous versions: ξ a probability measure on $\mathscr{X}$,

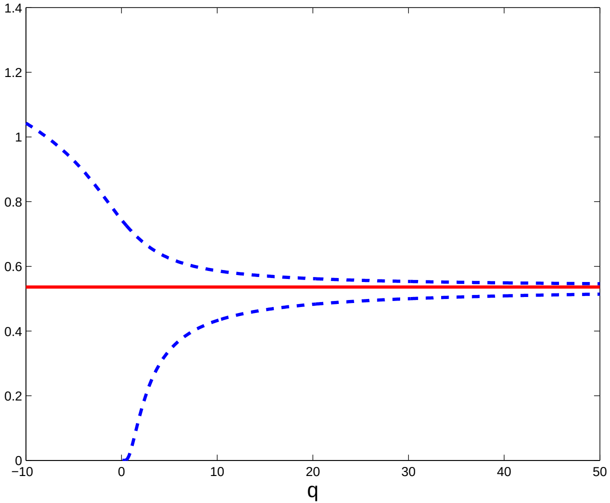
$\tilde{\phi}_{[q]}(\xi) = \left[\int_{\mathscr{X}} \int_{\mathscr{X}} \|x - y\|^{-q}\, \xi(dx)\, \xi(dy)\right]^{-1/q}$

$\tilde{\phi}_{[0]}(\xi) = \exp \int_{\mathscr{X}} \int_{\mathscr{X}} \log \|x - y\|\, \xi(dx)\, \xi(dy)$

Regularized Φ_{Mm} (2)

7-point maximin-distance design ($d = 2$):

Φ_{Mm} and bounds $\underline{\phi}_{[q]}$ and $\overline{\phi}_{[q]}$



A less severe regularization of Φ_{Mm} with NN

Write $\Phi_{Mm}(\mathbf{X}) = \min_i d_i^*$ with $d_i^* = \min_{j \neq i} \|X_i - X_j\|$
(= Nearest Neighbor distance to X_i)

Define

$$\underline{\phi}_{[NN,q]}(\mathbf{X}) = \left[\sum_{i=1}^{n} (d_i^*)^{-q}\right]^{-1/q} , \overline{\phi}_{[NN,q]}(\mathbf{X}) = \left[\sum_{i=1}^{n} \frac{(d_i^*)^{-q}}{n}\right]^{-1/q}$$

A less severe regularization of Φ_{Mm} with NN

Write $\Phi_{Mm}(\mathbf{X}) = \min_i d_i^*$ with $d_i^* = \min_{j \neq i} \|X_i - X_j\|$
(= Nearest Neighbor distance to X_i)

Define

$$\underline{\phi}_{[NN,q]}(\mathbf{X}) = \left[\sum_{i=1}^{n} (d_i^*)^{-q}\right]^{-1/q}, \overline{\phi}_{[NN,q]}(\mathbf{X}) = \left[\sum_{i=1}^{n} \frac{(d_i^*)^{-q}}{n}\right]^{-1/q}$$

Then $\boxed{\underline{\phi}_{[NN,q]}(\mathbf{X}) \leq \phi_{Mm}(\mathbf{X}) \leq n^{1/q}\, \underline{\phi}_{[NN,q]}(\mathbf{X})}$, $q > 0$,
(convergence monotonic in q from both sides as $q \to \infty$)

A less severe regularization of Φ_{Mm} with NN

Write $\Phi_{Mm}(\mathbf{X}) = \min_i d_i^*$ with $d_i^* = \min_{j\neq i} \|X_i - X_j\|$
(= Nearest Neighbor distance to X_i)

Define

$$\underline{\phi}_{[NN,q]}(\mathbf{X}) = \left[\sum_{i=1}^n (d_i^*)^{-q}\right]^{-1/q} , \overline{\phi}_{[NN,q]}(\mathbf{X}) = \left[\sum_{i=1}^n \frac{(d_i^*)^{-q}}{n}\right]^{-1/q}$$

Then $\boxed{\underline{\phi}_{[NN,q]}(\mathbf{X}) \leq \phi_{Mm}(\mathbf{X}) \leq n^{1/q}\, \underline{\phi}_{[NN,q]}(\mathbf{X})}$, $q > 0$,
(convergence monotonic in q from both sides as $q \to \infty$)

$$\frac{\phi_{Mm}(\underline{\mathbf{X}}^*_{[NN,q]})}{\phi^*_{Mm}} \geq n^{-1/q} ,$$

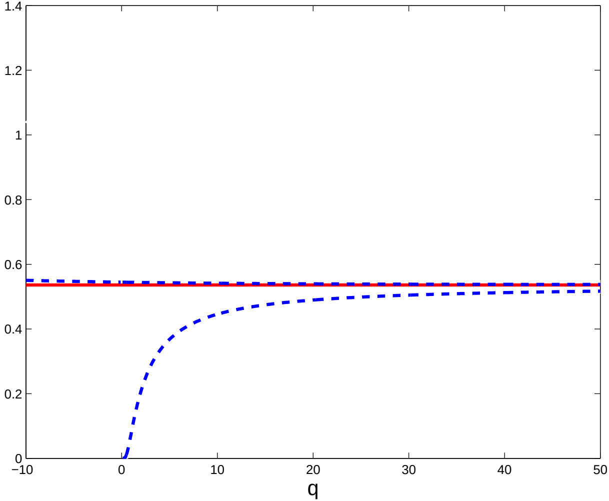
with $\underline{\mathbf{X}}^*_{[NN,q]}$ optimal for $\underline{\phi}_{[NN,q]}$

(Maximin efficiency $> 1 - \epsilon$ for $q > \frac{\log(n)}{\epsilon}$ $\to$ gain of factor 2)

A less severe regularization of Φ_{Mm} with NN (2)

7-point maximin-distance design ($d = 2$):

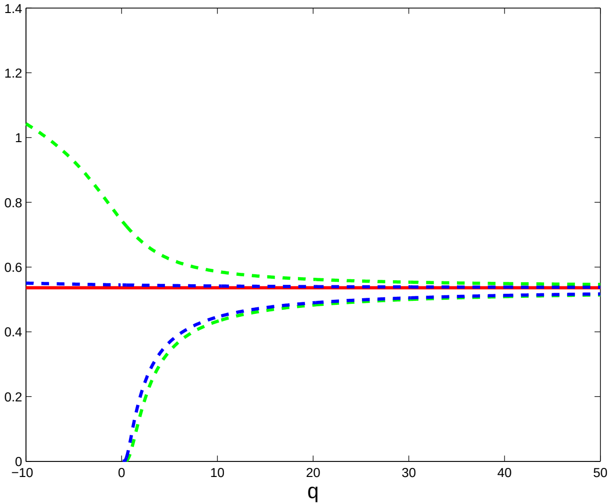
Φ_{Mm} and regularization via $\underline{\phi}_{[NN,q]}$



A less severe regularization of Φ_{Mm} with NN (2)

7-point maximin-distance design ($d = 2$):

Φ_{Mm} and regularization via $\underline{\phi}_{[NN,q]}$ and $\underline{\phi}_{[q]}$



[Beardwood, Halton, Hammersley 1959]: X_i i.i.d., p.d.f. φ, TSP graph $\mathcal{G}_{TSP}(\mathbf{X})$

$$\frac{\sum_{e_i \in \mathcal{G}_{TSP}(\mathbf{X})} |e_i|}{n^{(d-1)/d}} \to C(d) \int \varphi^{(d-1)/d}(x)\, dx \text{ a.s. }, \; n \to \infty$$

then [Steele, 1981] for other Euclidean functionals on $\mathbf{X}$, [Redmond & Yukich, 1994] using quasi-additivity

Relation with MST and other graphs & entropy

[Beardwood, Halton, Hammersley 1959]: X_i i.i.d., p.d.f. φ, TSP graph $\mathcal{G}_{TSP}(\mathbf{X})$

$$\frac{\sum_{e_i \in \mathcal{G}_{TSP}(\mathbf{X})} |e_i|}{n^{(d-1)/d}} \to C(d) \int \varphi^{(d-1)/d}(x)\, dx \text{ a.s. }, \; n \to \infty$$

then [Steele, 1981] for other Euclidean functionals on $\mathbf{X}$, [Redmond & Yukich, 1994] using quasi-additivity

[Redmond & Yukich , 1996], [Yukich, 1998], [Penrose & Yukich 2003...2011], [Wade, 2011]:

$$\frac{\sum_{e_i \in \mathcal{G}(\mathbf{X})} |e_i|^\beta}{n^{(d-\beta)/d}} \to C(\beta, d) \int \varphi^{(d-\beta)/d}(x)\, dx \,, \; n \to \infty$$

with $\mathcal{G}(\mathbf{X})$ Minimum Spanning Tree (MST), NN, TSP, Voronoi, Delaunay, Sphere of Influence, Gabriel... (different types of convergence (L_p), different conditions on φ and β...)

Graphs & entropy (2)

$$\frac{\sum_{e_i \in \mathcal{G}(\mathbf{X})} |e_i|^\beta}{n^{(d-\beta)/d}} \to C(\beta, d) \int \varphi^{(d-\beta)/d}(x)\, dx\,, \;\; n \to \infty$$

- $\Rightarrow$ estimation of Rényi entropy
  $H_\alpha{}^*(\varphi) = \frac{1}{1-\alpha} \log \int \varphi^\alpha(t)\, dt$ with $\alpha = (d-\beta)/d$
  $1 \leq \beta < d \Rightarrow 0 < \alpha \leq 1 - 1/d$, RHS max for $\varphi =$ ct. (uniform)
  $\to$ maximize the LHS

Graphs & entropy (2)

$$\frac{\sum_{e_i \in \mathcal{G}(\mathbf{X})} |e_i|^\beta}{n^{(d-\beta)/d}} \to C(\beta,d) \int \varphi^{(d-\beta)/d}(x)\, dx\,,\ \ n \to \infty$$

- $\Rightarrow$ estimation of Rényi entropy
  $H_\alpha{}^*(\varphi) = \frac{1}{1-\alpha} \log \int \varphi^\alpha(t)\, dt$ with $\alpha = (d-\beta)/d$
  $1 \leq \beta < d \Rightarrow 0 < \alpha \leq 1 - 1/d$, RHS max for $\varphi =$ ct. (uniform)
  $\longrightarrow$ maximize the LHS
- NN graph: $|e_i| = d_i^* \to$ maximize $\underline{\phi}_{[NN,q]}$ with $q = -\beta$

Graphs & entropy (2)

$$\frac{\sum_{e_i \in \mathcal{G}(\mathbf{X})} |e_i|^\beta}{n^{(d-\beta)/d}} \to C(\beta,d) \int \varphi^{(d-\beta)/d}(x)\, dx\,, \ \ n \to \infty$$

- $\Rightarrow$ estimation of Rényi entropy $H_\alpha{}^*(\varphi) = \frac{1}{1-\alpha} \log \int \varphi^\alpha(t)\, dt$ with $\alpha = (d-\beta)/d$
  $1 \leq \beta < d \Rightarrow 0 < \alpha \leq 1 - 1/d$, RHS max for $\varphi =$ ct. (uniform)
  $\longrightarrow$ maximize the LHS
- NN graph: $|e_i| = d_i^*$ $\to$ maximize $\underline{\phi}_{[NN,q]}$ with $q = -\beta$
- $(1/n) \sum_{e_i \in \mathcal{G}_{MST}(\mathbf{X})} |e_i|$ used in [Franco, Ph.D., 2008] to classify designs (also considers $\mathrm{Var}_{\mathcal{G}_{MST}(\mathbf{X})}\{|e_i|\}$)

Graphs & entropy (2)

$$\frac{\sum_{e_i \in \mathcal{G}(\mathbf{X})} |e_i|^\beta}{n^{(d-\beta)/d}} \to C(\beta, d) \int \varphi^{(d-\beta)/d}(x)\, dx\,, \;\; n \to \infty$$

- $\Rightarrow$ estimation of Rényi entropy
  $H_\alpha{}^*(\varphi) = \frac{1}{1-\alpha} \log \int \varphi^\alpha(t)\, dt$ with $\alpha = (d-\beta)/d$
  $1 \leq \beta < d \Rightarrow 0 < \alpha \leq 1 - 1/d$, RHS max for $\varphi =$ ct. (uniform)
  $\longrightarrow$ maximize the LHS
- NN graph: $|e_i| = d_i^*$ $\to$ maximize $\underline{\phi}_{[NN,q]}$ with $q = -\beta$
- $(1/n) \sum_{e_i \in \mathcal{G}_{MST}(\mathbf{X})} |e_i|$ used in [Franco, Ph.D., 2008] to classify designs (also considers $\mathrm{Var}_{\mathcal{G}_{MST}(\mathbf{X})}\{|e_i|\}$)
- Shannon entropy with kernel estimator in [Jourdan & Franco, 2009, 2010]

Luc Pronzato

Workshop de clôture ANR OPUS 21/10/2011

Maximin optimal design is model-free,

$\longrightarrow$ **model-specific design?**

Luc Pronzato

3 Beyond space filling

Maximin optimal design is model-free,
$\longrightarrow$ **model-specific design?**

Model for $f(\cdot)$

$f(x) = \mathbf{r}^\top(x)\beta + Z(x,\omega)$ with
$\mathbf{r}(x)$ a known (vector of) functions(s) of x
$Z(x,\omega)$ = realization of a (second-order stationary) **Gaussian process** (random field)
$\mathbb{E}\{Z(x,\omega)\} = 0$, $\mathbb{E}\{Z(x,\omega)Z(u,\omega)\} = \sigma^2 C((x-u);\theta)$

3 Beyond space filling

Maximin optimal design is model-free,
$\longrightarrow$ **model-specific design?**

Model for $f(\cdot)$

$f(x) = \mathbf{r}^\top(x)\beta + Z(x,\omega)$ with
$\mathbf{r}(x)$ a known (vector of) functions(s) of x
$Z(x,\omega)$ = realization of a (second-order stationary) **Gaussian process** (random field)
$\mathbb{E}\{Z(x,\omega)\} = 0$, $\mathbb{E}\{Z(x,\omega)Z(u,\omega)\} = \sigma^2 C((x-u);\theta)$

Computer experiments

[Sacks, Welch, Mitchell, Wynn, 1989]: Take $C(\delta;\theta)$ continuous at $\delta = 0$, $C(0;\theta) = 1 \rightarrow$ two repetitions for the same x give the same $f(x)$

Kriging

Objective

Interpolate (or extrapolate): construct a prediction $\eta_n(x)$ for one particular realization of $Z(\cdot)$

$\neq$ situation: prediction for future realizations

($\longrightarrow$ simply estimate β!)

Kriging

Objective

Interpolate (or extrapolate): construct a prediction $\eta_n(x)$ for one particular realization of $Z(\cdot)$

$\neq$ situation: prediction for future realizations

($\rightarrow$ simply estimate β!)

Ordinary kriging:

$f(x) = \beta + Z(x, \omega) \longrightarrow \eta_n(x) = \eta_n[f](x)$

BLUP at x: $\boxed{\eta_n(x) = \mathbf{v}_n^\top(x)\mathbf{y}_n}$ where

- $\mathbf{y}_n = (f(X_1), \ldots, f(X_n))^\top$
- $\mathbf{v}_n(x)$ minimizes $\mathbb{E}\{(\mathbf{v}_n^\top \mathbf{y}_n - [\beta + Z(x, \omega)])^2\}$
- under the constraint $\mathbb{E}\{\mathbf{v}_n^\top \mathbf{y}_n\} = \beta \sum_{i=1}^n \{\mathbf{v}_n\}_i = \mathbb{E}\{f(x)\} = \beta$, i.e., $\sum_{i=1}^n \{\mathbf{v}_n\}_i = 1$

Kriging (2)

Prediction: $\eta_n(x) = \hat{\beta}^n + \mathbf{c}_n^\top(x)\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n \mathbf{1})$

MSPE: proportional to

$$\rho_n(x) = \left(1 - [\mathbf{c}_n^\top(x)\ 1] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(x) \\ 1 \end{bmatrix}\right)$$

[with $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j);\theta)$, $\{\mathbf{c}_n(x)\}_i = C((X_i - x);\theta)$, $\hat{\beta}^n = (\mathbf{1}^\top \mathbf{C}_n^{-1}\mathbf{y}_n)/(\mathbf{1}^\top \mathbf{C}_n^{-1}\mathbf{1})$ (WLS) and $\mathbf{1} = (1, \ldots, 1)^\top$]

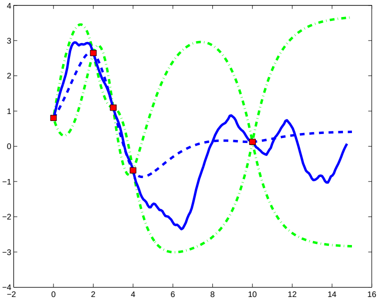
Kriging (2)

Prediction: $\eta_n(x) = \hat{\beta}^n + \mathbf{c}_n^\top(x)\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})$

MSPE: proportional to

$$\rho_n(x) = \left(1 - [\mathbf{c}_n^\top(x)\ 1] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(x) \\ 1 \end{bmatrix}\right)$$

[with $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j);\theta)$, $\{\mathbf{c}_n(x)\}_i = C((X_i - x);\theta)$, $\hat{\beta}^n = (\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{y}_n)/(\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{1})$ (WLS) and $\mathbf{1} = (1,\ldots,1)^\top$]



Typical criteria: based on MSPE $\rho_n(x)$

E.g., minimize Max. MSPE: $\max_{x\in\mathscr{X}} \rho_n(x)$
with $\rho_k(X_i) = 0,\ i = 1, \ldots, k$

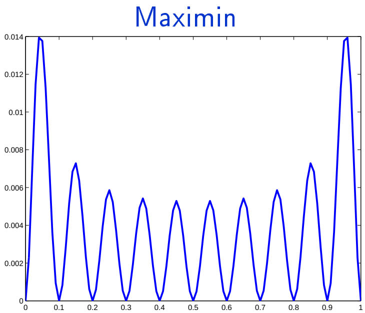
Typical criteria: based on MSPE $\rho_n(x)$

E.g., minimize Max. MSPE: $\max_{x \in \mathscr{X}} \rho_n(x)$
with $\rho_k(X_i) = 0$, $i = 1, \ldots, k$

The designs constructed are typically space-filling

Boundary effect

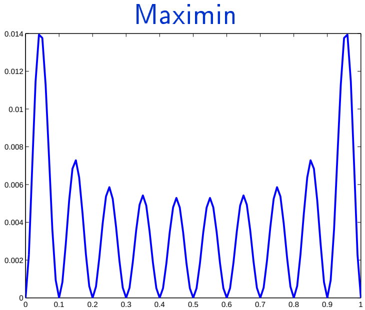
$d = 1$ $n = 11$ observations in $[0, 1]$, $C(t) = \exp(-50\, t^2)$
plot of $\rho_n(x)$

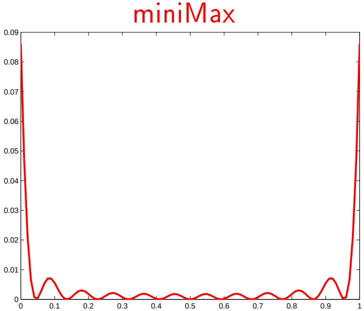


Boundary effect

$d = 1$ $n = 11$ observations in $[0, 1]$, $C(t) = \exp(-50\, t^2)$

plot of $\rho_n(x)$





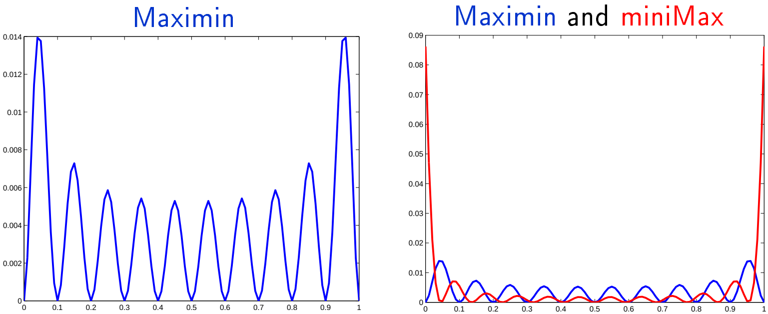
Uniform distribution of points

$\Rightarrow$ less points available for prediction near the boundary

$\Rightarrow$ larger uncertainty near the boundary

Boundary effect

$d = 1$ $n = 11$ observations in $[0,1]$, $C(t) = \exp(-50\,t^2)$
plot of $\rho_n(x)$



Uniform distribution of points

$\Rightarrow$ less points available for prediction near the boundary

$\Rightarrow$ larger uncertainty near the boundary

$\longrightarrow$ Put more points near the boundaries of $\mathscr{X}$

Similar to polynomial regression (e.g. D-optimality):

- put design points at the roots of some orthogonal polynomials
  (e.g., roots of $t(t-1)P'_{p-1}(2t-1)$ for D-optimal design on $[0,1]$ with P_n= n-th Legendre polynomial on $[-1,1]$)

$\rightarrow$ Put more points near the boundaries of $\mathscr{X}$

Similar to polynomial regression (e.g. D-optimality):

- put design points at the roots of some orthogonal polynomials
  (e.g., roots of $t(t-1)P'_{p-1}(2t-1)$ for D-optimal design on $[0,1]$ with P_n= n-th Legendre polynomial on $[-1,1]$)
- Erdös-Turan theorem: roots r of orthogonal polynomials on $[0,1]$ are asymptotically distributed with the arcsine density $\varphi(r) = \frac{1}{\pi\sqrt{r(1-r)}}$

→ Put more points near the boundaries of $\mathscr{X}$

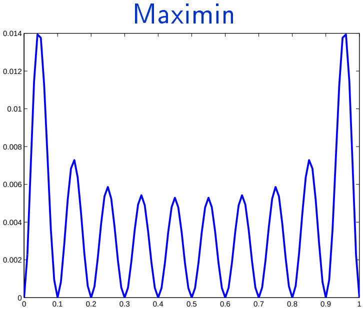
Similar to polynomial regression (e.g. D-optimality):

- put design points at the roots of some orthogonal polynomials
  (e.g., roots of $t(t-1)P'_{p-1}(2t-1)$ for D-optimal design on $[0,1]$ with P_n= n-th Legendre polynomial on $[-1,1]$)
- Erdös-Turan theorem: roots r of orthogonal polynomials on $[0,1]$ are asymptotically distributed with the arcsine density $\varphi(r) = \frac{1}{\pi\sqrt{r(1-r)}}$
- [Dette & Pepelyshev, 2010]:
  - use a space filling-design (e.g., Maximin Lh),
  - for all $j = 1, \ldots, d$, transform the j-th coordinates $\{X_i\}_j$ by $T : x \mapsto z = T(x) = \frac{1+\cos(\pi x)}{2}$
    ($x \sim$ uniformly $\rightarrow z \sim$ arcsine),
  - use the transformed design points Z_i

Correcting the boundary effect (2)

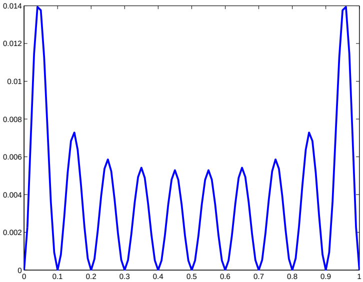
…The arcsine transformation can be too severe for small n

Back to the example, plot of $\rho_n(x)$:

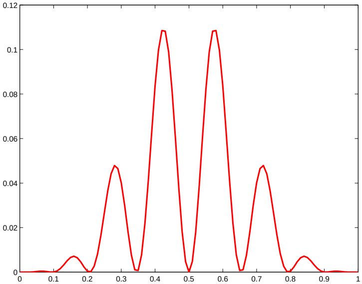


...The arcsine transformation can be too severe for small n
Back to the example, plot of $\rho_n(x)$:

Maximin

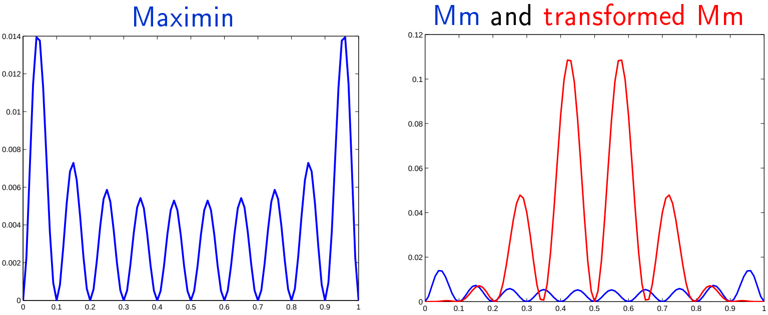


transformed Maximin (arcsine)



Correcting the boundary effect (2)

... The arcsine transformation can be too severe for small n
Back to the example, plot of $\rho_n(x)$:



Arcsine distribution: maximizes
$\tilde{\Phi}_{[0]}(\xi) = \exp\left[\int_0^1 \int_0^1 \log\|x-y\|\,\xi(dx)\,\xi(dy)\right]$
(continuous version of $\overline{\phi}_{[0]}(\mathbf{X}) = \exp\left[\sum_{i<j} \mu_{ij}\, \log(d_{ij})\right]$)

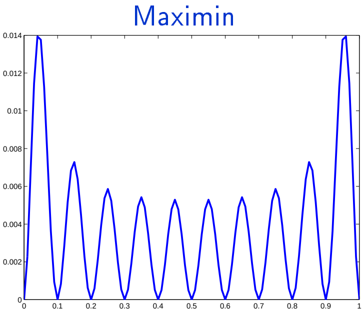
Arcsine distribution: maximizes
$\tilde{\Phi}_{[0]}(\xi) = \exp\left[\int_0^1 \int_0^1 \log\|x-y\|\,\xi(dx)\,\xi(dy)\right]$
(continuous version of $\overline{\phi}_{[0]}(\mathbf{X}) = \exp\left[\sum_{i<j} \mu_{ij}\, \log(d_{ij})\right]$)

Maximization of

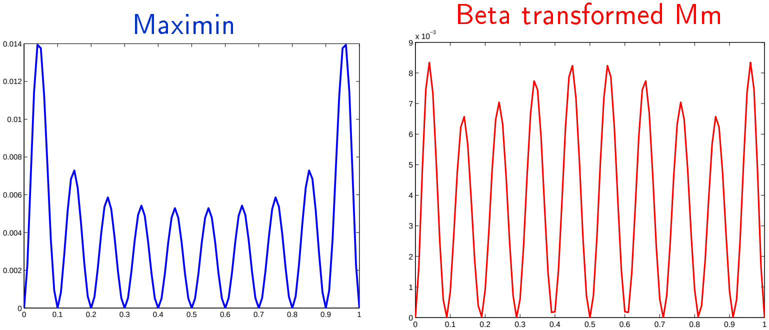
$$\tilde{\Phi}_{[q]}(\xi) = \left[\int_0^1 \int_0^1 \|x-y\|^{-q}\,\xi(dx)\,\xi(dy)\right]^{-1/q}, \quad 0<q<1$$

is obtained for ξ having the density $\varphi(x) = \frac{x^{(q-1)/2}(1-x)^{(q-1)/2}}{B(\frac{q+1}{2},\frac{q+1}{2})}$
(Beta distribution) [Dette, Pepelyshev, Zhigljavsky, 2009]
(tends to arcsine for $q \to 0$ and to uniform for $q \to 1$)

. . . For a suitable Beta transformation ($q = 0.84$)

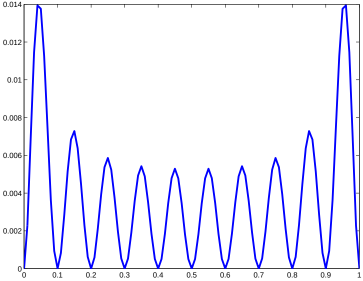


... For a suitable Beta transformation ($q = 0.84$)

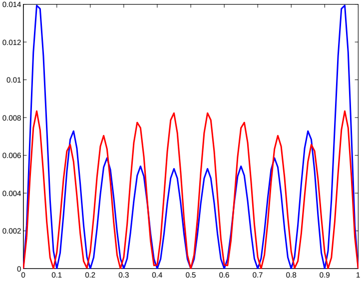


...For a suitable Beta transformation ($q = 0.84$)

Maximin



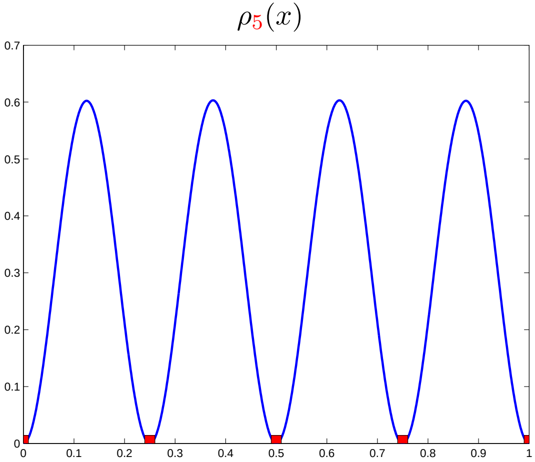
Mm and Beta transformed Mm



Choice of a suitable q?
(depends on the correlation function $C(\cdot, \theta)$)

Sequential construction

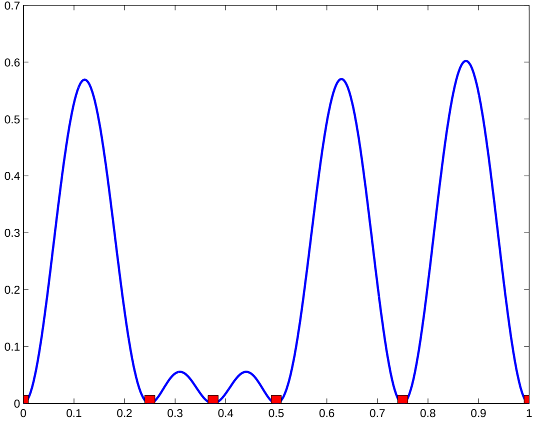
$$x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)$$ (initialized with $\mathbf{X}_n$, n small)



Sequential construction

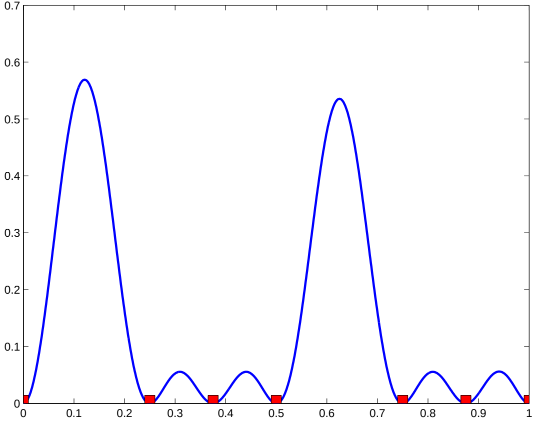
$$x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)$$ (initialized with $\mathbf{X}_n$, n small)

$\rho_6(x)$



$$\boxed{x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)}$$ (initialized with $\mathbf{X}_n$, n small)

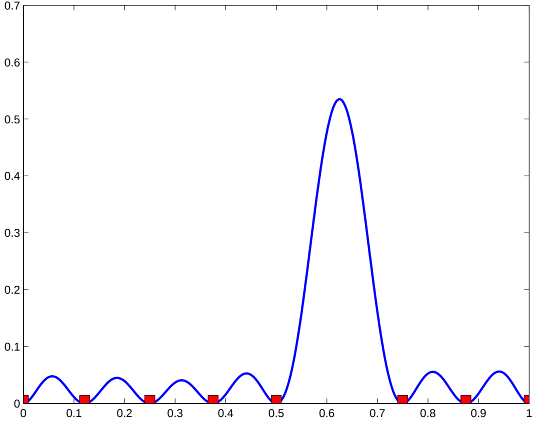
$\rho_7(x)$



Sequential construction

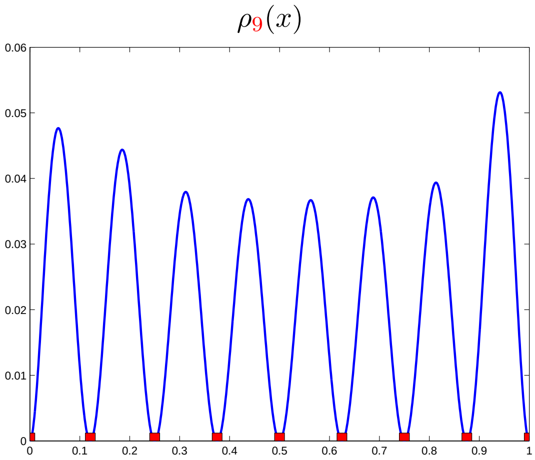
$$\boxed{x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)}$$ (initialized with $\mathbf{X}_n$, n small)

$\rho_8(x)$



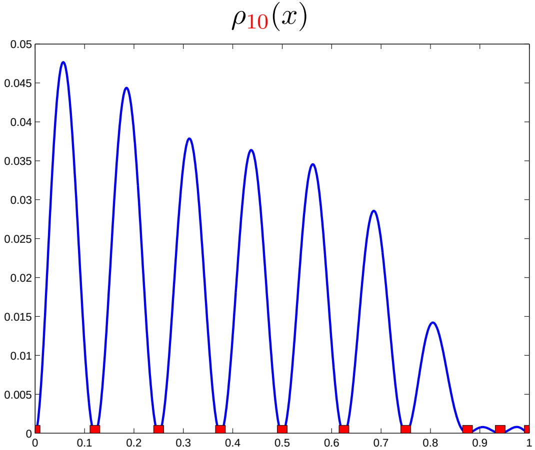
Sequential construction

$$\boxed{x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)}$$ (initialized with $\mathbf{X}_n$, n small)



Sequential construction

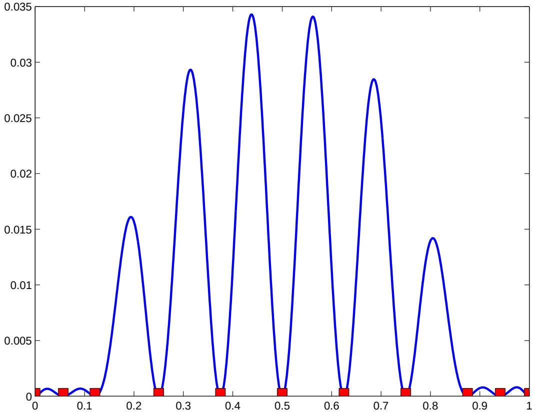
$$x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)$$ (initialized with $\mathbf{X}_n$, n small)



Sequential construction

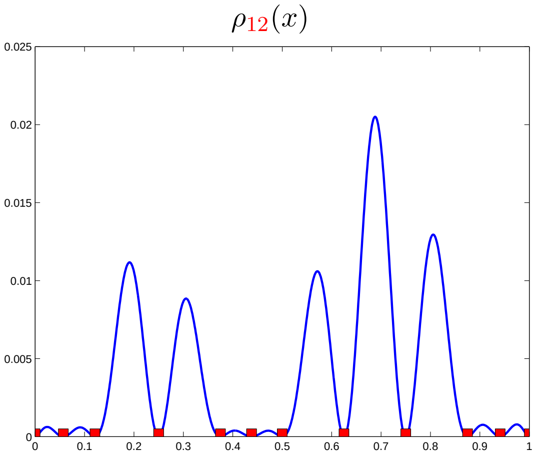
$$x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)$$ (initialized with $\mathbf{X}_n$, n small)

$\rho_{11}(x)$



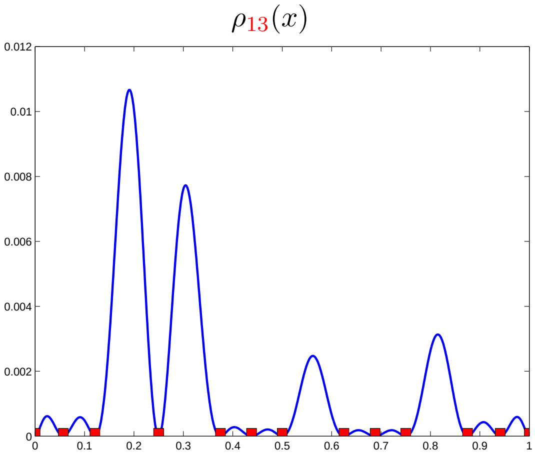
Sequential construction

$$x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)$$ (initialized with $\mathbf{X}_n$, n small)



Sequential construction

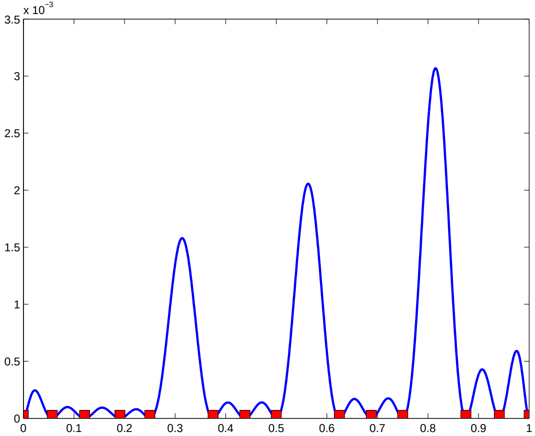
$$\boxed{x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)}$$ (initialized with $\mathbf{X}_n$, n small)



Sequential construction

$$x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)$$ (initialized with $\mathbf{X}_n$, n small)

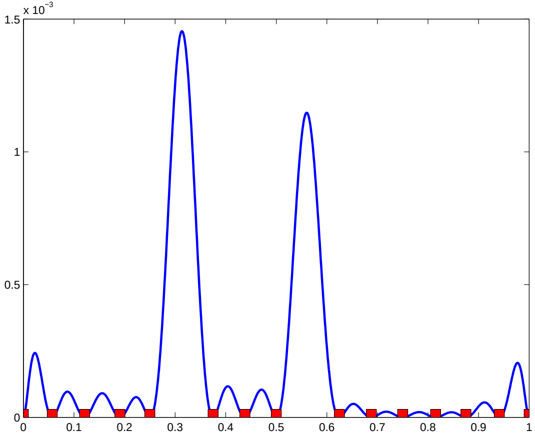
$\rho_{14}(x)$



Sequential construction

$$\boxed{x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)}$$ (initialized with $\mathbf{X}_n$, n small)

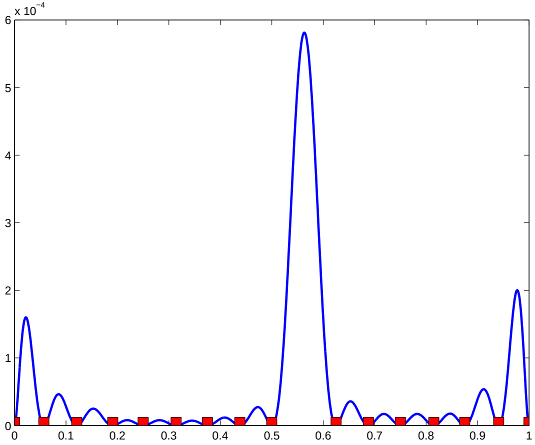
$\rho_{15}(x)$



Sequential construction

$$x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)$$ (initialized with $\mathbf{X}_n$, n small)

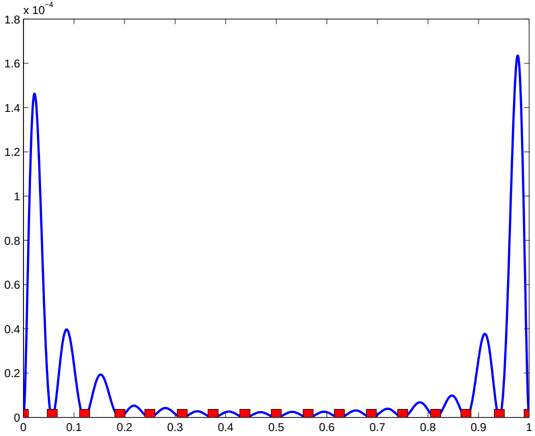
$\rho_{16}(x)$



Sequential construction

$$\boxed{x_{k+1} = \arg\max_{x \in \mathscr{X}} \rho_k(x)}$$ (initialized with $\mathbf{X}_n$, n small)

$\rho_{17}(x)$



Some remarks

Luc Pronzato

- Irregularities in $\rho_k(x)$ are difficult to avoid $\rightarrow$ related to a property of low discrepancy sequences:
  $\mathbf{X}^*_{mM}$ miniMax optimal ($x_i = (2i-1)/(2n)$) minimizes $\mathcal{D}(\mathbf{X}) = \max_{x \in \mathbf{X}} |F_n(x) - U(x)|$ ($U(\cdot)$=cdf of the uniform distribution, $F_n(\cdot)$ empirical cdf for $\mathbf{X}$) and $\mathcal{D}(\mathbf{X}^*_{mM}) = 1/(2n)$
  but $\mathcal{D}(x_1, x_2, \ldots, x_n) \geq 0.06 \log(n)/n$ for any *sequence* of n points

Some remarks (2)

- [Vazquez & Bect, 2011]
  $\rho_n(x) = \sup_{f\in\mathcal{H}_1} |f(x) - \eta_n[f](x)|^2$, with $\mathcal{H}_1$ the unit ball of the RKHS $\mathcal{H}$ of functions generated by $k(u,v) = C((u-v);\theta)$

Some remarks (2)

- [Vazquez & Bect, 2011] $\rho_n(x) = \sup_{f\in\mathcal{H}_1} |f(x) - \eta_n[f](x)|^2$, with $\mathcal{H}_1$ the unit ball of the RKHS $\mathcal{H}$ of functions generated by $k(u,v) = C((u-v);\theta)$
- [Vazquez & Bect, 2011] If $t \mapsto C(t;\theta)$ has a Fourier transform $\tilde{C}(u) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} C(t;\theta)\, e^{\imath(t,u)}\, dt$ satisfying $c_1(1+\|u\|_2^2)^{-s} \leq \tilde{C}(u) \leq c_2(1+\|u\|_2^2)^{-s}\,,\ u\in\mathbb{R}^d$ with $s > d/2$, $0 < c_1 \leq c_2$, then,

$$x_{k+1} = \arg\max_{x\in\mathscr{X}} \rho_k(x) \Rightarrow \sup_{x\in\mathscr{X}} \rho_n(x) = \mathcal{O}(n^{1-2s/d})$$

— same rate as for non-sequential construction —

Some remarks (3)

- Comparison with non-parametric estimation with i.i.d. additive errors [Tsybakov, 2004]:
  - $\sup_{f\in\Sigma(L,\alpha)} \mathbb{E}\|f-\eta_n[f]\|_2^2 = \mathcal{O}(n^{-2\alpha/(2\alpha+d)})$

    ($\Sigma(L,\alpha)$=Hölder smooth, $\mathbb{E} \leftrightarrow$ presence of random errors)

    = best achievable rate for passive (non-sequential) or active (sequential) constructions [Castro, Willett, Nowak, 2005]

Some remarks (3)

- Comparison with non-parametric estimation with i.i.d. additive errors [Tsybakov, 2004]:
  - $\sup_{f\in\Sigma(L,\alpha)} \mathbb{E}\|f-\eta_n[f]\|_2^2 = \mathcal{O}(n^{-2\alpha/(2\alpha+d)})$

    ($\Sigma(L,\alpha)$=Hölder smooth, $\mathbb{E} \leftrightarrow$ presence of random errors)

    = best achievable rate for passive (non-sequential) or active (sequential) constructions [Castro, Willett, Nowak, 2005]
  - *[different for piecewise smooth functions: [Castro, Willett, Nowak, 2005]*

    $\mathcal{O}(\max\{n^{-2\alpha/(2\alpha+d)}, n^{-1/d}\})$ *for passive construction*

    $\mathcal{O}(\max\{n^{-2\alpha/(2\alpha+d)}, n^{-1/(d-1)}\})$ *for active construction*

    — not important for α small, some effect when f is regular enough]

Some remarks (3)

- Comparison with non-parametric estimation with i.i.d. additive errors [Tsybakov, 2004]:
  - $\sup_{f\in\Sigma(L,\alpha)} \mathbb{E}\|f-\eta_n[f]\|_2^2 = \mathcal{O}(n^{-2\alpha/(2\alpha+d)})$
    ($\Sigma(L,\alpha)$=Hölder smooth, $\mathbb{E} \leftrightarrow$ presence of random errors)
    = best achievable rate for passive (non-sequential) or active (sequential) constructions [Castro, Willett, Nowak, 2005]
  - *[different for piecewise smooth functions: [Castro, Willett, Nowak, 2005]*
    $\mathcal{O}(\max\{n^{-2\alpha/(2\alpha+d)}, n^{-1/d}\})$ *for passive construction*
    $\mathcal{O}(\max\{n^{-2\alpha/(2\alpha+d)}, n^{-1/(d-1)}\})$ *for active construction*
    — not important for α small, some effect when f is regular enough]
- But sequential construction is useful when objective = optimization, inversion, estimation of failure prob.

- Comparison with non-parametric estimation with i.i.d. additive errors [Tsybakov, 2004]:
  - $\sup_{f\in\Sigma(L,\alpha)} \mathbb{E}\|f-\eta_n[f]\|_2^2 = \mathcal{O}(n^{-2\alpha/(2\alpha+d)})$
    ($\Sigma(L,\alpha)$=Hölder smooth, $\mathbb{E} \leftrightarrow$ presence of random errors)
    = best achievable rate for passive (non-sequential) or active (sequential) constructions [Castro, Willett, Nowak, 2005]
  - *[different for piecewise smooth functions: [Castro, Willett, Nowak, 2005]*
    $\mathcal{O}(\max\{n^{-2\alpha/(2\alpha+d)}, n^{-1/d}\})$ *for passive construction*
    $\mathcal{O}(\max\{n^{-2\alpha/(2\alpha+d)}, n^{-1/(d-1)}\})$ *for active construction*
    — not important for α small, some effect when f is regular enough]
- But sequential construction is useful when objective = optimization, inversion, estimation of failure prob.
- Importance of asymptotic statements? (how large n should be?)

Hidden difficulty: θ in $C(\cdot;\theta)$ is unknown

$\rightarrow$ We must estimate θ from the same data as those use to construct $\eta_n(x)$

Hidden difficulty: θ in $C(\cdot;\theta)$ is unknown

$\longrightarrow$ We must estimate θ from the same data as those use to construct $\eta_n(x)$

$\longrightarrow$ Maximum likelihood estimator $\hat{\theta}^n$
($Z(x,\omega)$ is Gaussian)

$\longrightarrow$ corrective term [Harville & Jeske, 1992; Abt 1999]:

$$\hat{\rho}_n(x;\theta) = \rho_n(x;\theta) + \mathrm{trace}\{\mathbf{M}_\theta^{-1} \frac{\partial \mathbf{v}_n(x;\theta)}{\partial \theta} \mathbf{C}_n \frac{\partial \mathbf{v}_n(x;\theta)}{\partial \theta^\top}\}$$

where:

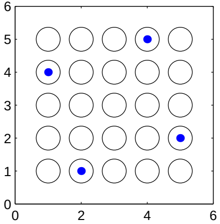
$\mathbf{v}_n(x;\theta)$ such that $\eta_n(x) = \mathbf{v}_n^\top(x;\theta)\mathbf{y}_n$

$\mathbf{M}_\theta$ = Fisher information matrix for θ

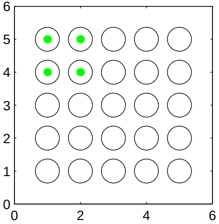
Example: [Zimmerman, *Envirometrics* 2006]

$\mathbb{E}\{Z(x,\omega)Z(u,\omega)\} = \sigma^2\theta^{\|x-u\|}$, $\theta = 0.3$ (σ^2 known),
$\mathscr{X}$ = regular grid 5×5

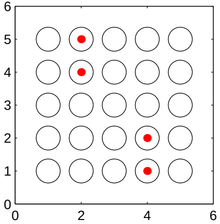
prediction for θ known



estimation of θ



prediction with θ estimated



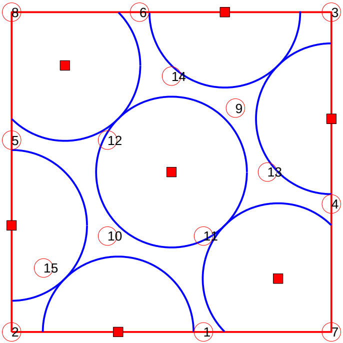
prediction for θ known:
$\mathbf{X}_4$ minimizes $\max_{x\in\mathscr{X}} \rho_4(x)$
estimation of θ:
$\mathbf{X}_4$ maximizes $\det \mathbf{M}_\theta$
prediction with θ estimated:
$\mathbf{X}_4$ minimizes $\max_{x\in\mathscr{X}} \hat{\rho}_4(x;\theta)$

Sequential construction based on $\max_{x\in\mathscr{X}} \hat{\rho}_k(x;\theta)$

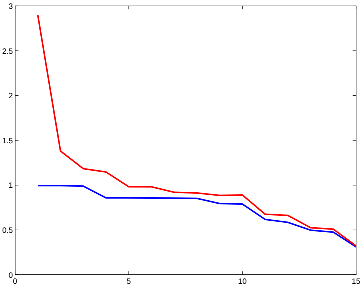
$d=2$, $C(t;\theta)=\exp(-\theta\|t\|_2)$, $\theta=0.7$, $\mathbf{X}_7=$Mm Lh, then

$x_{7+j} = \arg\max_{x\in\mathscr{X}} \hat{\rho}_{7+j-1}(x;\theta)$, $j=1,2\ldots$

x_{7+j}



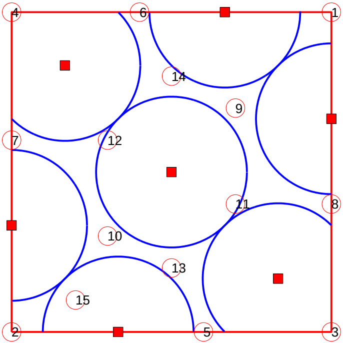
$\max_{x\in\mathscr{X}} \rho_{7+j}(x;\theta)$ and $\max_{x\in\mathscr{X}} \hat{\rho}_{7+j}(x;\theta)$



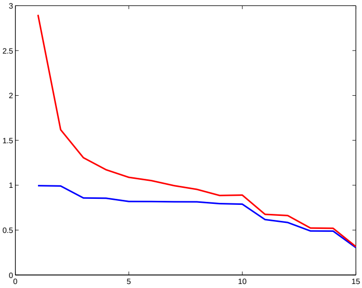
$d = 2$, $C(t;\theta) = \exp(-\theta\|t\|_2)$, $\theta = 0.7$, $\mathbf{X}_7$ =Mm Lh, then

$x_{7+j} = \arg\max_{x\in\mathscr{X}} \rho_{7+j-1}(x;\theta)$, $j = 1, 2\ldots$

x_{7+j}



$\max_{x\in\mathscr{X}} \rho_{7+j}(x;\theta)$ and $\max_{x\in\mathscr{X}} \hat{\rho}_{7+j}(x;\theta)$



An alternative criterion

Choosing $\mathbf{X}_n$ that minimizes $\max_{x\in\mathscr{X}} \hat{\rho}_n(x;\theta)$ is difficult

An alternative criterion

Choosing $\mathbf{X}_n$ that minimizes $\max_{x\in\mathscr{X}} \hat{\rho}_n(x;\theta)$ is difficult

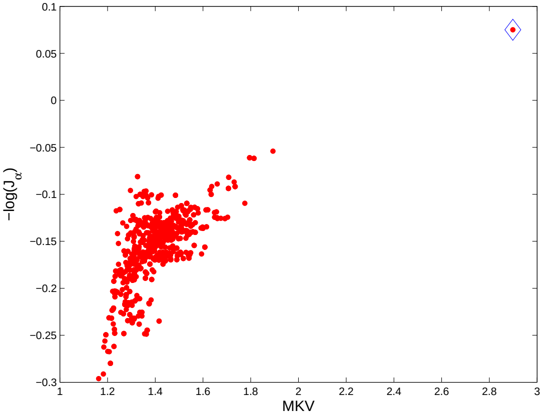
$\rightarrow$ Use a criterion that makes compromise between space filling and clustering, e.g., choose $\mathbf{X}_n$ that maximizes $\gamma \log\det(\mathbf{M}_\beta) + (1-\gamma)\log\det(\mathbf{M}_\theta)$ [Müller et al., 2010, 2011], with

- $\mathbf{M}_\beta$= FIM for trend parameters (maximization $\rightarrow$ space filling)
- $\mathbf{M}_\theta$= FIM for correlation parameters (maximization $\rightarrow$ clustering)

An alternative criterion (2)

Example: $n = 7$, $d = 2$, $C(t; \theta) = \exp(-\theta \|t\|_2)$, $\theta = 0.7$, 1000 Lh (999 random + $\diamondsuit$ for Mm optimal)

MKV$= \max_{x \in \mathscr{X}} \hat{\rho}_n(x; \theta)$, $J_\alpha = \det^\alpha(\mathbf{M}_\beta) + \det^{1-\alpha}(\mathbf{M}_\theta)$ ($\alpha = 0.8$)



4 Conclusions

- Only considered design for approximation/interpolation

4 Conclusions

- Only considered design for approximation/interpolation
- Nothing on algorithms (heuristics — genetic, taboo search, SA —, MCMC — spatial point process —, local search)

4 Conclusions

- Only considered design for approximation/interpolation
- Nothing on algorithms (heuristics — genetic, taboo search, SA —, MCMC — spatial point process —, local search)
- Nothing on Bayesian methods:
  - Karhunen-Loève representation + Bayesian optimal design [Fedorov & Müller, 2007]
  - Maximum Entropy Sampling [Shewry & Wynn, 1987; Wynn & Youssef, 2011] (with substitution of INLA for MCMC to compute posterior means?)

Conclusions (2)

- Basic guidelines:
  - Precision of the construction ($\leftrightarrow$ final objective) for a given horizon n?
    ($\mathbb{E}$ or worst-case) — e.g., $\max_{x \in \mathscr{X}} \rho_n(x)$
  - $\longrightarrow$ design criterion

Conclusions (2)

- Basic guidelines:
  - Precision of the construction ($\leftrightarrow$ final objective) for a given horizon n?
    ($\mathbb{E}$ or worst-case) — e.g., $\max_{x \in \mathscr{X}} \rho_n(x)$
  - $\rightarrow$ design criterion
  - ... taking all sources of uncertainty into account (estimation of parameters, simulations with variable precision...) — e.g., $\max_{x \in \mathscr{X}} \hat{\rho}_n(x)$

Conclusions (2)

- Basic guidelines:
  - Precision of the construction ($\leftrightarrow$ final objective) for a given horizon n?
    ($\mathbb{E}$ or worst-case) — e.g., $\max_{x \in \mathscr{X}} \rho_n(x)$
  - $\rightarrow$ design criterion
  - ...taking all sources of uncertainty into account (estimation of parameters, simulations with variable precision...) — e.g., $\max_{x \in \mathscr{X}} \hat{\rho}_n(x)$
  - Some contraints may exist on real physical experiments (e.g., dynamical constraints $\rightarrow$ mobile sensors)

Conclusions (2)

- Basic guidelines:
  - Precision of the construction ($\leftrightarrow$ final objective) for a given horizon n?
    ($\mathbb{E}$ or worst-case) — e.g., $\max_{x \in \mathscr{X}} \rho_n(x)$
  - $\rightarrow$ design criterion
  - . . . taking all sources of uncertainty into account (estimation of parameters, simulations with variable precision. . .) — e.g., $\max_{x \in \mathscr{X}} \hat{\rho}_n(x)$
  - Some contraints may exist on real physical experiments (e.g., dynamical constraints $\rightarrow$ mobile sensors)
- Only simplest rules and constructions will be used!

- Basic guidelines:
  - Precision of the construction ($\leftrightarrow$ final objective) for a given horizon n?
    ($\mathbb{E}$ or worst-case) — e.g., $\max_{x \in \mathscr{X}} \rho_n(x)$
  - $\to$ design criterion
  - ...taking all sources of uncertainty into account (estimation of parameters, simulations with variable precision...) — e.g., $\max_{x \in \mathscr{X}} \hat{\rho}_n(x)$
  - Some contraints may exist on real physical experiments (e.g., dynamical constraints $\to$ mobile sensors)
- Only simplest rules and constructions will be used!

Thank you for your attention!