Optimal design for linear models with correlated observations

Holger Dette, Ruhr-Universität Bochum
Andrey Pepelyshev, RWTH Aachen
Anatoly Zhigljavsky, Cardiff University

Nice, July 2013

Outline

Linear regression model

- Common linear regression model

$$y(x) = \theta_1 f_1(x) + \ldots + \theta_m f_m(x) + \varepsilon(x) \ ,$$

  - $f_1, \ldots, f_m$ are linearly independent, continuous (regression) functions
  - $\theta_1, \ldots, \theta_m$ are unknown parameters
  - N observations

$$y_1 = y(x_1), \ldots, y_N = y(x_N)$$

at experimental conditions $x_1, \ldots, x_N \in \mathcal{X} \subset \mathbb{R}^d$

Correlation

- Correlation structure

$$\mathrm{E}[\varepsilon(x_i)] = 0, \quad \mathrm{E}[\varepsilon(x_i)\varepsilon(x_j)] = K(x_i, x_j)\,; \quad x_i, x_j \in \mathcal{X}$$

- Here K is a kernel representing the covariance structure, which satisifies
  - positive definite
  - $K(u, v) \neq 0$ for all $(u, v) \in \mathcal{X} \times \mathcal{X}$
  - continuous at all points $(u, v) \in \mathcal{X} \times \mathcal{X}$ except possibly at the diagonal points (u, u)
- **Design problem:** optimal allocation of $x_1, \ldots, x_N$ for most efficient estimation of $\theta_1, \ldots, \theta_m$

Estimation

- Least squares estimation (LSE)

$$\tilde{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

where

  - $\mathbf{X} = (f_i(x_j))_{j=1,\dots,N}^{i=1,\dots,m}$
  - $\mathbf{Y} = (y_1, \dots, y_N)^T$

- Covariance matrix of $\tilde{\theta}$

$$\mathrm{Var}(\tilde{\theta}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}\mathbf{X}\,(\mathbf{X}^T\mathbf{X})^{-1}$$

where

$$\mathbf{\Sigma} = (K(x_i, x_j))_{i,j=1,\dots,N}$$

Weighted versus unweighted least squares

- Weighted least squares estimation (BLUE)

$$\hat{\theta} = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{Y}$$

- Covariance matrix of $\hat{\theta}$

$$\mathrm{Var}(\hat{\theta}) = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \leq \mathrm{Var}(\tilde{\theta})$$

where

$$\mathbf{\Sigma} = (K(x_i, x_j))_{i,j=1,\ldots,N}$$

- **Note:** We focus on ordinary least squares estimation (LSE)

Weighted versus unweighted least squares

- Weighted least squares estimation (BLUE)

$$\hat{\theta} = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{Y}$$

- Covariance matrix of $\hat{\theta}$

$$\mathrm{Var}(\hat{\theta}) = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \leq \mathrm{Var}(\tilde{\theta})$$

where

$$\mathbf{\Sigma} = (K(x_i, x_j))_{i,j=1,\dots,N}$$

- **Note:** We focus on ordinary least squares estimation (LSE)

Weighted versus unweighted least squares

- **Note:** We focus on ordinary least squares estimation (LSE) because

  (1) BLUE is often sensitive with respect to misspecification of Σ (LSE is more robust)

  (2) The difference between BLUE and LSE is often surprisingly small [Rao (1967), Kruskal (1968)]

  (3) We will give a heuristic explanation of this phenomenon and will additionally derive conditions such that

$$LSE + \text{optimal design} \;=\; BLUE + \text{optimal design}$$

Motivation (one dimensional case)

- $a : \mathcal{X} \to [0, 1]$ distribution function on $\mathcal{X} \subset \mathbb{R}$
- Design points are quantiles of a, that is

$$x_i = a^{-1}((i-1)/(N-1)), \; i = 1, \dots, N,$$

- If ξ_N is the probability measure with masses $1/N$ at x_i, then

$$\mathrm{Var}(\tilde{\theta}) = D(\xi_N) = M^{-1}(\xi_N) B(\xi_N, \xi_N) M^{-1}(\xi_N)$$

where

- $M(\xi_N) = \int_{\mathcal{X}} f(u) f^T(u) \xi_N(\mathrm{d}u)$
- $B(\xi_N, \xi_N) = \int\int K(u, v) f(u) f^T(v) \xi_N(\mathrm{d}u) \xi_N(\mathrm{d}v)$

and $f = (f_1, \dots, f_m)^T$ is the vector of regression functions.

Motivation (one dimensional case)

- $a : \mathcal{X} \to [0,1]$ distribution function on $\mathcal{X} \subset \mathbb{R}$
- Design points are quantiles of a, that is

$$x_i = a^{-1}((i-1)/(N-1)), \ i = 1, \dots, N,$$

- If ξ_N is the probability measure with masses $1/N$ at x_i, then

$$\mathrm{Var}(\tilde{\theta}) = D(\xi_N) = M^{-1}(\xi_N) B(\xi_N, \xi_N) M^{-1}(\xi_N)$$

where

- $M(\xi_N) = \int_{\mathcal{X}} f(u) f^T(u) \xi_N(\mathrm{d}u)$
- $B(\xi_N, \xi_N) = \int\int K(u,v) f(u) f^T(v) \xi_N(\mathrm{d}u) \xi_N(\mathrm{d}v)$

and $f = (f_1, \dots, f_m)^T$ is the vector of regression functions.

Approximate (continuous) designs

- For a probability measure ξ on $\mathcal{X}$ (more precisely on its Borel field) the matrix

$$D(\xi) = M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)$$

is called the **information matrix** (for LSE) of the design ξ, where

  - $M(\xi) = \int_{\mathcal{X}} f(u)f^T(u)\xi(\mathrm{d}u)$
  - $B(\xi,\xi) = \int\int K(u,v)f(u)f^T(v)\xi(\mathrm{d}u)\xi(\mathrm{d}v)$

Admissible designs

- Define

$$\mathcal{X}_1 = \mathcal{X} \backslash \mathcal{X}_0 = \{x \in \mathcal{X} : f(x) \neq 0\}$$

- Assume that designs ξ_0 and ξ_1 are concentrated on $\mathcal{X}_0$ and $\mathcal{X}_1$ correspondingly.
- The design $\xi_\alpha = \alpha\xi_0 + (1-\alpha)\xi_1$ satisfies

$$D(\xi_\alpha) = M^{-1}(\xi_\alpha)B(\xi_\alpha, \xi_\alpha)M^{-1}(\xi_\alpha) = D(\xi_1)$$

(for all $0 \leq \alpha < 1$)

- For the theoretical part of this talk we assume $f(x) \neq 0$ for all $x \in \mathcal{X}$

Admissible designs

- Define
$$\mathcal{X}_1 = \mathcal{X} \backslash \mathcal{X}_0 = \{x \in \mathcal{X} : f(x) \neq 0\}$$
- Assume that designs ξ_0 and ξ_1 are concentrated on $\mathcal{X}_0$ and $\mathcal{X}_1$ correspondingly.
- The design $\xi_\alpha = \alpha\xi_0 + (1-\alpha)\xi_1$ satisfies
$$D(\xi_\alpha) = M^{-1}(\xi_\alpha)B(\xi_\alpha, \xi_\alpha)M^{-1}(\xi_\alpha) = D(\xi_1)$$
(for all $0 \le \alpha < 1$)
- For the theoretical part of this talk we assume $f(x) \neq 0$ for all $x \in \mathcal{X}$

Optimal design

- Let $\Phi(\cdot)$ be a monotone, convex real valued functional defined on the space of symmetric $m \times m$ matrices
- The design ξ is **Φ-optimal**, if it minimizes the function

$$\Phi(D(\xi)) = \Phi(M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi))$$

among all designs on the design space $\mathcal{X}$, where

  - $M(\xi) = \int_{\mathcal{X}} f(u)f^T(u)\xi(\mathrm{d}u)$
  - $B(\xi,\xi) = \int\int K(u,v)f(u)f^T(v)\xi(\mathrm{d}u)\xi(\mathrm{d}v)$

- A further definition:

$$B(\xi,\nu) = \int_{\mathcal{X}}\int_{\mathcal{X}} K(u,v)f(u)f^T(v)\xi(\mathrm{d}u)\nu(\mathrm{d}v),$$

Optimal design

- Let $\Phi(\cdot)$ be a monotone, convex real valued functional defined on the space of symmetric $m \times m$ matrices
- The design ξ is **Φ-optimal**, if it minimizes the function

$$\Phi(D(\xi)) = \Phi(M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi))$$

among all designs on the design space $\mathcal{X}$, where
  - $M(\xi) = \int_{\mathcal{X}} f(u)f^T(u)\xi(\mathrm{d}u)$
  - $B(\xi,\xi) = \int\int K(u,v)f(u)f^T(v)\xi(\mathrm{d}u)\xi(\mathrm{d}v)$

- A further definition:

$$B(\xi,\nu) = \int_{\mathcal{X}}\int_{\mathcal{X}} K(u,v)f(u)f^T(v)\xi(\mathrm{d}u)\nu(\mathrm{d}v),$$

A necessary condition

Theorem

If the matrix of derivatives

$$C = \frac{\partial \Phi(D)}{\partial D} = \left(\frac{\partial \Phi(D)}{\partial D_{ij}} \right)_{i,j=1,\dots,m}$$

exists and ξ^* *minimizes* $\Phi(D(\xi))$, *then the inequality*

$$f^T(x) D(\xi^*)\, C(\xi^*) M^{-1}(\xi^*) f(x) \;\le\; tr\big(C(\xi^*) M^{-1}(\xi^*) B(\xi^*, \xi_x) M^{-1}(\xi^*)\big) \qquad (1)$$

holds for all $x \in \mathcal{X}$, *where*

$$B(\xi^*, \xi_x) = \int_{\mathcal{X}} K(u, x) f(u) \xi^*(\mathrm{d}u) f^T(x).$$

Moreover, there is equality in (1) *for* ξ^**-almost all* x

Two examples:

- The necessary condition is of the form

$$d(x,\xi^*) \;\le\; b(x,\xi^*) \quad \text{for all} \quad x \in \mathcal{X}$$

- D-optimality; $\Phi(D(\xi)) = -\log\det(D(\xi))$

$$f^T(x)M^{-1}(\xi^*)f(x) \;\le\; f^T(x)B^{-1}(\xi^*,\xi^*)\int K(u,x)f(u)\xi^*(\mathrm{d}u)$$

- c-optimality (for a given $c \in \mathbb{R}^m$); $\Phi(D(\xi)) = c^T D(\xi)c$

$$f^T(x)M^{-1}(\xi^*)cc^T M^{-1}(\xi^*) \times\Big(\int K(x,u)f(u)\xi^*(du) - B(\xi^*,\xi^*)M^{-1}(\xi^*)f(x)\Big) \;\ge 0$$

Quadratic regression on the interval $[-1, 1]$

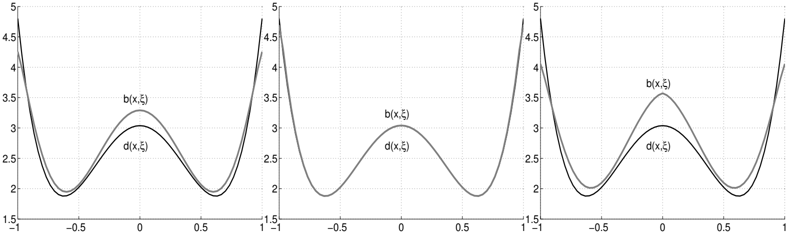


Figure: *The functions $b(x, \xi)$ and $d(x, \xi)$ in the necessary condition*

$$d(x, \xi^*) \ \leq \ b(x, \xi^*)$$

for the covariance kernels $K(u, v) = e^{-|u-v|}$, $K(u, v) = -\log(u - v)^2$ and $K(u, v) = \max(0, 1 - |u - v|)$. ξ^* is arcsine design, i.e.

$$\frac{d\xi^*}{dx} = \frac{1}{\pi\sqrt{1 - x^2}}$$

Comments: the lack of convexity

- **Note:** The conditions are "only" necessary. This means:
  - The arcsine design is **not** D-optimal for quadratic regression with a covariance kernel

$$K(u,v) = e^{-|u-v|} \text{ or } K(u,v) = \max(0, 1-|u-v|)$$

  - For the logarithmic kernel

$$K(u,v) = -\log(u-v)^2$$

  we observe equality in the necessary condition for all x.

  $\rightarrow$ The arcsine design **might** be D-optimal for quadratic regression with logarithmic kernel

Comments: the lack of convexity

- Optimality results are only available for the location model

$$y(x) = \theta + \varepsilon(x)$$

(in this case the criterion is fact convex).

- In the following discussion we propose a method for deriving optimality results for more general models:
    - regression models with more than one regression function and an associated covariance kernel
    - universally optimal designs

Comments: the lack of convexity

- Optimality results are only available for the location model

$$y(x) = \theta + \varepsilon(x)$$

  (in this case the criterion is fact convex).

- In the following discussion we propose a method for deriving optimality results for more general models:
  - regression models with more than one regression function and an associated covariance kernel
  - universally optimal designs

Universally optimal designs

- A design ξ^* is **universally** optimal if and only if

$$D(\xi^*) \;\leq\; D(\xi)$$

in the sense of the Loewner ordering for any design $\xi \in \Xi$, that is

$$c^T D(\xi^*) c \;\leq\; c^T D(\xi) c$$

for all $c \in \mathbb{R}^m$.

- A design ξ^* is universally optimal if and only if it is c-optimal for all $c \in \mathbb{R}^m$.

Universally optimal designs

- A design ξ^* is **universally** optimal if and only if

$$D(\xi^*) \;\leq\; D(\xi)$$

  in the sense of the Loewner ordering for any design $\xi \in \Xi$, that is

$$c^T D(\xi^*) c \;\leq\; c^T D(\xi) c$$

  for all $c \in \mathbb{R}^m$.

- A design ξ^* is universally optimal if and only if it is c-optimal for all $c \in \mathbb{R}^m$.

A crucial representation

- For any design ξ we have the representation

$$\int K(x,u)f(u)\xi(du) = \Lambda f(x) + g_\xi(x)\,, \quad x \in \mathcal{X}\,,$$

where $\Lambda = B(\xi,\xi)M^{-1}(\xi)$ and the function g_ξ satisfies.

$$\int g_\xi(x) f^T(x)\xi(\mathrm{d}x) = 0$$

- **Note:**
  - The function g_ξ depends on the design ξ and the kernel K
  - If $g_\xi \equiv 0$ and Λ is diagonal, then the regression functions $f = (f_1, \ldots, f_m)^T$ are eigenfunctions of the integral operator associated with the kernel K and the design ξ

A crucial representation

- For any design ξ we have the representation

$$\int K(x,u)f(u)\xi(du) = \Lambda f(x) + g_\xi(x)\,, \quad x \in \mathcal{X}\,,$$

where $\Lambda = B(\xi,\xi)M^{-1}(\xi)$ and the function g_ξ satisfies.

$$\int g_\xi(x)f^T(x)\xi(\mathrm{d}x) = 0$$

- **Note:**
  - The function g_ξ depends on the design ξ and the kernel K
  - If $g_\xi \equiv 0$ and Λ is diagonal, then the regression functions $f = (f_1,\ldots,f_m)^T$ are eigenfunctions of the integral operator associated with the kernel K and the design ξ

$g_\xi \equiv 0$ is sufficient for universal optimality

Theorem

Consider the linear regression model with a covariance kernel K, a design $\xi \in \Xi$ and the corresponding the vector-function $g_\xi(\cdot)$ defined by

$$g_\xi(x) = \int K(x,u) f(u) \xi(du) - \Lambda f(x), \quad x \in \mathcal{X},$$

If $g_\xi(x) = 0$ for all $x \in \mathcal{X}$, then the design ξ is universally optimal.

Proof (first idea)

- Check c-optimality for any $c \in \mathbb{R}^m$
- Necessary condition:

$$f^T(x)M^{-1}(\xi)cc^TM^{-1}(\xi)\underbrace{\Big(\int K(x,u)f(u)\xi(du) - B(\xi,\xi)M^{-1}(\xi)f(x)\Big)}_{g_\xi(x)\equiv 0} \geq 0$$

- ξ is a candidate for universal optimality!
- However, the criterion is **not** convex!

Proof (idea)

- **Idea of a rigorous proof:** simultaneous optimal estimation and optimization of the design in the model

$$y(x) = \theta^T f(x) + \varepsilon(x)$$

where the full trajectory $\{y(x) | x \in \mathcal{X}\}$ can be observed.

- Arbitrary (linear) estimate: if $\mu = (\mu_1, \ldots, \mu_m)^T$ is a **vector of signed measures**

$$\hat{\theta}(\mu) = \int y(x)\mu(\mathrm{d}x)$$

- Unbiasedness means here

$$\int \mu(\mathrm{d}x) f^T(x) = \int f(x)\mu^T(\mathrm{d}x) = I_m \,,$$

- E.g. $\mu_\xi(dx) = M^{-1}(\xi) f(x)\xi(dx)$ gives LSE for the design ξ

Proof (idea)

- **Idea of a rigorous proof:** simultaneous optimal estimation and optimization of the design in the model

$$y(x) = \theta^T f(x) + \varepsilon(x)$$

where the full trajectory $\{y(x) | x \in \mathcal{X}\}$ can be observed.

- Arbitrary (linear) estimate: if $\mu = (\mu_1, \ldots, \mu_m)^T$ is a **vector of signed measures**

$$\hat{\theta}(\mu) = \int y(x)\mu(\mathrm{d}x)$$

- Unbiasedness means here

$$\int \mu(\mathrm{d}x) f^T(x) = \int f(x)\mu^T(\mathrm{d}x) = I_m \,,$$

- E.g. $\mu_\xi(dx) = M^{-1}(\xi) f(x) \xi(dx)$ gives LSE for the design ξ

Proof (idea)

- **Note:** The variance of $c^T\hat{\theta}(\mu)$ is given by

$$\begin{aligned}\mathrm{Var}(c^T\hat{\theta}(\mu)) &= c^T\int\int \mathrm{E}[\varepsilon(x)\varepsilon(u)]\mu(dx)\mu^T(du)c \\ &= c^T\int\int K(x,u)\mu(\mathrm{d}x)\mu^T(\mathrm{d}u)c \;=:\; \Phi_c(\mu)\end{aligned}$$

- **This function is convex with respect to μ!**

Proof (idea)

- Standard equivalence theory (convex optimization) is applicable!

- A vector of signed measures μ^* minimizes

$$\Phi_c(\mu) = c^T \int \int K(x,u)\mu(\mathrm{d}x)\mu^T(\mathrm{d}u)c$$

if and only if the inequality

$$c^T \int \int K(x,u)\mu^*(\mathrm{d}x)\nu^T(\mathrm{d}u)c \geq \Phi_c(\mu^*)$$

holds for all vector valued signed measures ν corresponding to unbiased estimates.

Proof (idea)

- We use
$$\mu^*(\mathrm{d}x) = M^{-1}(\xi) f(x) \xi(\mathrm{d}x), \tag{2}$$
which yields an unbiased estimator
- Note that ($g_\xi \equiv 0$, by assumption of the Theorem)
$$\int K(x,u) f(x) \xi^*(dx) = \Lambda f(u) \tag{3}$$
- Left hand side of equivalence theorem
$$
\begin{aligned}
& c^T \int\int K(x,u) \mu^*(\mathrm{d}x) \nu^T(\mathrm{d}u) c \\
\stackrel{(2)}{=} \; & c^T M^{-1}(\xi) \int\int K(x,u) f(x) \xi(\mathrm{d}x) \nu^T(\mathrm{d}u) c \\
\stackrel{(3)}{=} \; & c^T M^{-1}(\xi) \int \Lambda f(u) \nu^T(\mathrm{d}u) c \stackrel{unbiased}{=} c^T M^{-1}(\xi) \Lambda c
\end{aligned}
$$

Proof (idea)

- We use

$$\mu^*(\mathrm{d}x) = M^{-1}(\xi) f(x) \xi(\mathrm{d}x), \qquad (2)$$

  which yields an unbiased estimator
- Note that ($g_\xi \equiv 0$, by assumption of the Theorem)

$$\int K(x,u) f(x) \xi^*(dx) = \Lambda f(u) \qquad (3)$$

- Left hand side of equivalence theorem

$$
\begin{aligned}
& c^T \int \int K(x,u) \mu^*(\mathrm{d}x) \nu^T(\mathrm{d}u) c \\
\stackrel{(2)}{=} \; & c^T M^{-1}(\xi) \int \int K(x,u) f(x) \xi(\mathrm{d}x) \nu^T(\mathrm{d}u) c \\
\stackrel{(3)}{=} \; & c^T M^{-1}(\xi) \int \Lambda f(u) \nu^T(\mathrm{d}u) c \stackrel{unbiased}{=} c^T M^{-1}(\xi) \Lambda c
\end{aligned}
$$

Proof (idea)

- We use
$$\mu^*(\mathrm{d}x) = M^{-1}(\xi^*)f(x)\xi^*(\mathrm{d}x), \tag{4}$$
- Right hand side of equivalence theorem (with similar arguments)
$$\begin{aligned} \Phi_c(\mu^*) &= c^T M^{-1}(\xi)\Lambda c \\ &= c^T M^{-1}(\xi)B(\xi,\xi)M^{-1}(\xi)c = D(\xi) \end{aligned}$$
- μ^* minimizes Φ_c in the class of all vector valued signed measures corresponding to unbiased estimates!

Proof (idea)

- Now return to the minimization of $D(\eta)$ in the class of all designs $\eta \in \Xi$.
- For any $\eta \in \Xi$ consider the corresponding vector-valued signed measure $\mu_\eta(\mathrm{d}x) = M^{-1}(\eta)f(x)\eta(dx)$, then

$$\begin{aligned} c^T D(\eta)c &= c^T M^{-1}(\eta)B(\eta,\eta)M^{-1}(\eta)c = \Phi_c(\mu_\eta) \\ &\geq \min_\mu \Phi_c(\mu) = \Phi_c(\mu^*) = c^T D(\xi)c. \end{aligned}$$

- Since the design ξ does not depend on the particular vector c, it follows that ξ is universally optimal.

$g_\xi \equiv 0$ is "necessary" for universal optimality

Theorem

Consider the linear regression model with a covariance kernel K, a design $\xi \in \Xi$ and the corresponding function $g_\xi(\cdot)$ defined by

$$g_\xi(x) = \int K(x,u)f(u)\xi(du) - \Lambda f(x)\,, \;\; x \in \mathcal{X}\,,$$

If the design ξ is universally optimal, then the function $g_\xi(\cdot)$ can be represented in the form

$$g_\xi(x) = \gamma(x)f(x),$$

where $\gamma(x)$ is a non-negative function such that $\gamma(x) = 0$ for all x in the support of the design ξ.

Remarks:

- **Note:** If $g_\xi \equiv 0$ then LSE with the optimal design can **not** be improved by any BLUE!

$$LSE + \text{optimal design} \;=\; BLUE + \text{optimal design}$$

- Mercer's theorem provides numerous models for which universally optimal designs can be identified explicitly [see e.g. Kanwal (1997)]

Remarks:

- Integral operator on $L_2(\xi)$

$$T_K(f)(\cdot) = \int_{\mathcal{X}} K(\cdot, u) f(u) \xi(\mathrm{d}u)$$

Under certain assumptions on the kernel T_K defines a symmetric, compact self-adjoint operator.

- Mercer's theorem: there exist a countable number of eigenfunctions

$$\varphi_1, \varphi_2, \cdots$$

with positive eigenvalues

$$\lambda_1, \lambda_2, \ldots$$

of the operator K

Optimal designs for kernels corresponding to integral operators

Theorem

- *Assume that the covariance kernel $K(x,u)$ defines an integral operator T_K with corresponding eigenfunctions $\varphi_1, \varphi_2, \ldots$*
- *For any non-singular matrix $L \in \mathbb{R}^{m \times m}$ consider the linear regression model*

$$\theta^T f(x) = \theta^T L(\varphi_{i_1}(x), \ldots, \varphi_{i_m}(x))^T$$

  with covariance kernel $K(x,u)$.
- **Then the design ξ is universally optimal!**

Example: series estimation/nonparametric regression

- Consider the regression functions

$$f_j(x) = \begin{cases} 1 & \text{if} \quad j = 1 \\ \sqrt{2}\cos(2\pi(j-1)x) & \text{if} \quad j \geq 2 \end{cases} \tag{5}$$

  on the design space $\mathcal{X} = [0, 1]$.

- **Note:** Linear models with regression functions (5) are widely applied in series estimation in nonparametric regression [see e.g. Efromovich (1999), Tsybakov (2009)].

- If $K(x, y) = \rho(x - y)$ (stationarity) where ρ is periodic with period 1
  $\rightarrow$ **the uniform design is universally optimal!**

Example: polynomial regression

- Consider the regression functions

$$f_j(x) = x^{j-1}\ ,\ j = 1, \ldots, m+1 \tag{6}$$

on the design space $\mathcal{X} = [-1, 1]$.

- If $K(x, y) = -\log|x - y|$ (stationarity)

- $\rightarrow$ **the arcsine design is universally optimal!**

$$\frac{d\xi^*}{dx} = \frac{1}{\pi\sqrt{1-x^2}}$$

Example: spherical descriptors

- For $n = 0, 1, \dots ;\ m = -n, -n+2, \dots, n-2, n$ define

$$Y_n^m(\varphi, \phi) = \sqrt{\frac{2n+1}{4\pi} \frac{n-|m|}{n+|m|}} P_n^{|m|}(\cos\varphi) \exp(im\psi)$$

where $\varphi \in [0, \pi]$, $\psi \in [0, 2\pi]$,

$$P_n^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{d^m x} P_n(x)$$

and P_n is the nth Legendre polynomial.

- **The uniform distribution on $[0, \pi] \times [0, 2\pi]$ is universally optimal for the kernels**

$$K(u, v) = \exp(-||u-v||^2)\ , \quad K(u, v) = (1 + \langle u, v \rangle)^d \ (d \in \mathbb{N})$$

Future research: g-Optimal Designs

- **Recall:** the condition

$$g_\xi(x) = \int_{\mathcal{X}} K(x,u) f(u) \xi(du) - B(\xi,\xi) M^{-1}(\xi) f(x) \equiv 0$$

is "necessary and sufficient" for universal optimality

- A g-**optimal design** minimizes

$$||g_\xi||_2^2 = \int_{\mathcal{X}} |g_\xi(x)|^2 d\xi(x)$$

- **Note:** This criterion seeks for designs "close" to universal optimality
- A multiplicative algorithm is available, which yields g-optimal designs.
- We expect that these designs have "good" with respect to many optimality criteria

Future research: g-Optimal Designs

- **Recall:** the condition

$$g_\xi(x) = \int_{\mathcal{X}} K(x,u) f(u) \xi(du) - B(\xi,\xi) M^{-1}(\xi) f(x) \equiv 0$$

  is "necessary and sufficient" for universal optimality

- A g-**optimal design** minimizes

$$||g_\xi||_2^2 = \int_{\mathcal{X}} |g_\xi(x)|^2 d\xi(x)$$

- **Note:** This criterion seeks for designs "close" to universal optimality
- A multiplicative algorithm is available, which yields g-optimal designs.
- We expect that these designs have "good" with respect to many optimality criteria

Future research: g-Optimal Designs

- **Recall:** the condition

$$g_\xi(x) = \int_{\mathcal{X}} K(x,u)f(u)\xi(du) - B(\xi,\xi)M^{-1}(\xi)f(x) \equiv 0$$

is "necessary and sufficient" for universal optimality

- A g-**optimal design** minimizes

$$||g_\xi||_2^2 = \int_{\mathcal{X}} |g_\xi(x)|^2 d\xi(x)$$

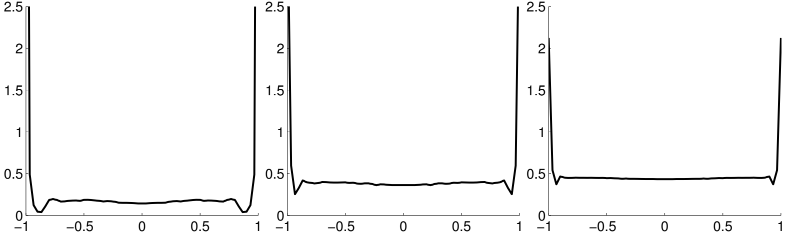
- **Note:** This criterion seeks for designs "close" to universal optimality
- A multiplicative algorithm is available, which yields g-optimal designs.
- We expect that these designs have "good" with respect to many optimality criteria

g-optimal designs for quadratic regression

- Quadratic regression model with correlation function

$$K(x, y) = \exp(-\lambda|x - y])$$

- $\mathcal{X} = [-1, 1]$
- $g-$optimal designs for $\lambda = 1$ (left), $\lambda = 4$ (middle) and $\lambda = 8$ (right).



g-optimal designs for quadratic regression

- Quadratic regression model with correlation function

$$K(x, y) = \exp(-\lambda|x - y|)$$

- $\mathcal{X} = [-1, 1]$
- D-, A-efficiency of the g-optimal and uniform design.

	$\lambda = 1$		$\lambda = 4$		$\lambda = 8$	
ξ	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$
ξ_g^*	0.996	0.993	0.998	0.996	0.999	0.998
ξ_u	0.821	0.832	0.851	0.822	0.910	0.881

Some selected references

H. Dette, A. Pepelyshev, A. Zhigljavksy (2013). Optimal design for linear models with correlated observations. Annals of Statistics, Vol. 41(1), 143-176.

H. Dette, A. Pepelyshev, A. Zhigljavksy (2013). "Nearly" universally optimal designs for models with correlated, Computational Statistic and Data Analysis. to appear.

S. Efromovich (1999). Nonparametric Curve Estimation, Springer Series in Statistics, Springer, NY.

R.P. Kanwal (1997). Linear Integral Equations. Boston, Birkhauser.

W. Kruskal (1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. Annals of Mathematical Statistics, 70-75.

C.R. Rao (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Proc. Fifth Berkeley Sympos., Univ. California Press, Berkeley, Calif., 355-372.

A.B. Tsybakov (2009). Introduction to Nonparametric Estimation. Springer Series in Statistics, Springer, NY.

A. Zhigljavksy, H. Dette, A. Pepelyshev (2010). A new approach to optimal design for linear models with correlated observations. Journal of the American Statistical Association, Vol. 105(491), 1093-1103.