Optimal experimental design that minimizes the width of simultaneous confidence bands

Satoshi Kuriki (Inst. Statist. Math., Tokyo)

Oberwolfach workshop Algebraic Statistics

Tue 20th April 2017

Joint with Henry Wynn (LSE, UK)

Contents of talk

1. New criterion for optimal experimental design
2. Polynomial regression and Möbius group action
3. Optimization over the cross-section space
4. Summary and open problems

1. New criterion for optimal experimental design

Very brief introduction to Optimal Design

- The data $(x_i, y_i)_{1 \le i \le N} \in \mathcal{X} \times \mathbb{R}$
  ($\mathcal{X} \subset (-\infty, \infty)$ is the domain of explanatory variable x.)
- Assume a regression model

$$y_i = b^\top f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2(x_i)) \text{ i.i.d.}$$

  $b \in \mathbb{R}^{n \times 1}$ (unknown), $f(x) \in \mathbb{R}^{n \times 1}$ (known), $\sigma^2(x) > 0$ (known)
- The LS: $\widehat{b} \sim N(b, \Sigma)$, where

$$\Sigma = M^{-1}, \quad M = \sum_{i=1}^{N} f(x_i) f(x_i)^\top \frac{1}{\sigma^2(x_i)} \text{ (information matrix)}$$

- "Optimal design" is to find a good $\{x_1, \ldots, x_N\} \in \mathcal{X}^N$ and hence good M, e.g., that maximizes

$$\det(\mathrm{Var}(\widehat{b}))^{-1} = \det(M) \quad \text{(D-optimal criterion)}$$

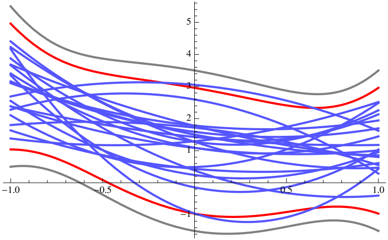
Simultaneous confidence bands

- $100(1-\alpha)\%$-simultaneous confidence band:

$$P\Big(b^\top f(x) \in \underbrace{\widehat{b}^\top f(x) \pm \|M^{-\frac{1}{2}} f(x)\| \cdot c_\alpha}_{\text{confidence band}} \ \text{for all } x \in \mathcal{X}\Big) = 1-\alpha$$

- Example of a confidence band:

$$b^\top f(x) = x^2 - x + 1, \ \ \sigma^2(x) \equiv 1, \ \ x_i = 1, 0, -1, \ \ \mathcal{X} = [-1, 1]$$



Blue	: Estimated curves (20 times)
Red	: 90% Confidence band (Naiman's volume-of-tube method)
Black	: Conservative 90% confidence band (Scheffè)

- The trajectory of the normalized regression basis vector

$$\Gamma = \left\{ \pm \frac{M^{-\frac{1}{2}} f(x)}{\|M^{-\frac{1}{2}} f(x)\|} \mid x \in \mathcal{X} \right\} \subset \mathbb{S}^{n-1} \text{ (unit sphere in } \mathbb{R}^n)$$

Let

$\mathrm{Vol}_1(\Gamma)$: 1-dim volume (length) of Γ

$\chi(\Gamma)$: the number of connected components of Γ

Proposition 1 (Naiman's (1986) volume-of-tube method)

$$P\Big(b^\top f(x) \in \widehat{b}^\top f(x) \pm \|M^{-\frac{1}{2}} f(x)\| \cdot c \ \ \textit{for all } x \in \mathcal{X} \Big)$$
$$\gtrapprox 1 - \frac{\mathrm{Vol}_1(\Gamma)}{2\pi} P(\chi_2^2 > c^2) - \frac{\chi(\Gamma)}{2} P(\chi_1^2 > c^2) \qquad (1)$$

("$\geq$" holds for all $c \geq 0$, "$\approx$" holds when c is large)

- By equating the RHS (1) to $1 - \alpha$, $c = c_\alpha$ is obtained.

Volume criterion — New optimal design criterion

- Naiman's formula (redisplay)

$$P\Big(b^\top f(x) \in \widehat{b}^\top f(x) \pm \underbrace{\|M^{-\frac{1}{2}} f(x)\| \cdot c_\alpha}_{\text{band width}} \ \text{ for all } x \in \mathcal{X}\Big)$$

$$\gtrapprox 1 - \frac{\mathrm{Vol}_1(\Gamma)}{2\pi} P\big(\chi_2^2 > c_\alpha^2\big) - \frac{\chi(\Gamma)}{2} P\big(\chi_1^2 > c_\alpha^2\big) \ (= 1 - \alpha)$$

- Propose a new criterion of optimal experimental design so that the width of confidence band is minimal.
  That is, the design minimizes $\|M^{-\frac{1}{2}} f(x)\|$ and $\mathrm{Vol}_1(\Gamma)$
- The design that minimizes $\max_{x\in\mathcal{X}} \|M^{-\frac{1}{2}} f(x)\|$ is known to be the D-optimal design (Kiefer-Wolfowitz's equivalence theorem).
- We propose the optimal design that minimizes $\mathrm{Vol}_1(\Gamma)$.

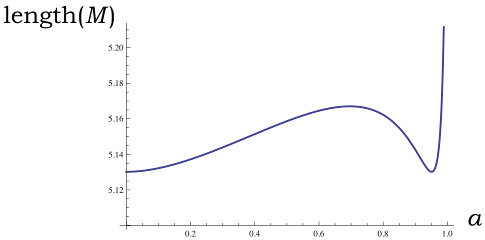
Problem is not convex!

- From now on, consider a polynomial regression (cf. Dette, et al., 1999)

$$f(x) = (1, x, \ldots, x^{n-1})^\top$$

- For example, $\mathrm{length}(M) := \mathrm{Vol}_1(M)$ for

$$M = (1-\alpha)\frac{1}{8}\begin{pmatrix}3 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 3\end{pmatrix} + \alpha\frac{1}{72}\begin{pmatrix}15+4\sqrt{3} & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 7-4\sqrt{3}\end{pmatrix}$$



- The problem is not convex.

2. Polynomial regression and Möbius group action

Volume criterion — New optimal design criterion (contd)

- The length of the curve Γ is

$$\mathrm{length}(M) := \mathrm{Vol}_1(\Gamma) = \int_{\mathcal{X}} \left\| \frac{d}{dx}\left(\frac{M^{-\frac{1}{2}} f(x)}{\|M^{-\frac{1}{2}} f(x)\|} \right) \right\| dx$$

$$= \int_{\mathcal{X}} \frac{\sqrt{f^\top M^{-1} f \cdot g^\top M^{-1} g - (f M^{-1} g)^2}}{f^\top M^{-1} f} \, dx$$

where $f(x) = (1, \ldots, x^{n-1})^\top$, $g(x) = \frac{d}{dx} f(x)$

- Problem: Minimize

$$\mathrm{length}(M)$$

subject to $M \in \mathcal{M}$, where

$$\mathcal{M} = \left\{ M = \int_{\mathcal{X}} f(x) f(x)^\top dP(x) \mid P : \text{positive measure} \right\}$$

(moment cone)

- Note: Instead of $\{x_i\}_{1 \le i \le N}$, we use the design measure P.
- $\mathrm{length}(M)$ is homogeneous in M, $\mathcal{M}$ can be a cone.

Volume optimal design (polynomial regression)

- The moment cone $\mathcal{M}$ is equivalent to the set of positive definite Hankel matrices (Karlin and Studden, 1966)

$$M \in \mathcal{M} = \left\{ \begin{pmatrix} m_0 & m_1 & & m_{n-1} \\ m_1 & & \iddots & \\ & \iddots & & m_{2n-3} \\ m_{n-1} & & m_{2n-3} & m_{2n-2} \end{pmatrix}_{n\times n} \succ 0 \right\}$$

- Objective function $\mathrm{length}(M)$ is an elliptic integral

Möbius group action

- Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. The map $\varphi : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$:
$$x \mapsto \varphi(x) = \varphi(x; a, b, c, d) = \frac{ax+b}{cx+d} \quad (ad - bc \neq 0)$$
the (real) Möbius transform
- $f(x) = (1, x, \ldots, x^{n-1})^\top$
Define a matrix $A = A(a, b, c, d)$ by
$$f(\varphi(x)) = Af(x)\frac{1}{(cx+d)^{n-1}}$$
e.g., when $n = 3$,
$$\underbrace{\begin{pmatrix} 1 \\ \frac{ax+b}{cx+d} \\ (\frac{ax+b}{cx+d})^2 \end{pmatrix}}_{f(\varphi(x))} = \underbrace{\begin{pmatrix} d^2 & 2cd & c^2 \\ bd & bc+ad & ac \\ b^2 & 2ab & a^2 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}}_{f(x)} \frac{1}{(cx+d)^2}$$
- $\mathcal{A} = \{A \mid ad - bc \neq 0\}$ forms a group (a representation of $GL(2)$)

Möbius group action (contd)

Lemma 1

If M is Hankel, then $AMA^\top$ $(A \in \mathcal{A})$ is Hankel. That is, the group $\mathcal{A}$ acts onto $\mathcal{M}$.

- We can define a equivalence relation:

$$M_1 \sim M_2 \quad \Leftrightarrow \quad M_2 = AM_1A^\top,\ \exists A \in \mathcal{A}$$

(M_1 and M_2 are on the same orbit)

Theorem 1

The group $\mathcal{A}$ remains the length $\mathrm{length}(M)$ invariant, i.e.,

$$\mathrm{length}(M_1) = \mathrm{length}(M_2) \quad \textit{if } M_1 \sim M_2$$

- We can reduce the dimension of the optimization problem.

3. Optimization over the cross-section space

Orbital decomposition

- From now on, we restricted our attention to the case $n = 3$.

Theorem 2

The moment cone $\mathcal{M}$ (= set of positive definite Hankel matrices) has an orbital decomposition

$$\mathcal{M} = \bigsqcup_{v \in (0, \frac{1}{3}]} \{M \mid M \sim M_v\}$$

where

$$M_v = \begin{pmatrix} 1 & 0 & v \\ 0 & v & 0 \\ v & 0 & 1 \end{pmatrix} \in \mathcal{M}$$

Orbital decomposition (contd)

Proof of Theorem 2.

We can show that for any $M \in \mathcal{M}$,

$$M \overset{(i)}{\sim} \begin{pmatrix} u_1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & w_1 \end{pmatrix} \overset{(ii)}{\sim} \begin{pmatrix} u_2 & 0 & v_2 \\ 0 & v_2 & 0 \\ v_2 & 0 & w_2 \end{pmatrix} \overset{(iii)}{\sim} \begin{pmatrix} 1 & 0 & v_3 \\ 0 & v_3 & 0 \\ v_3 & 0 & 1 \end{pmatrix} = M_{v_3}$$

(i) and (iii) are easy.

For (ii), we need to find $A = A(a, b, c, d)$ such that

$$A \begin{pmatrix} u_1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & w_1 \end{pmatrix} A^\top = \begin{pmatrix} u_2 & 0 & v_2 \\ 0 & v_2 & 0 \\ v_2 & 0 & w_2 \end{pmatrix}$$

by solving algebraic equations with checking resultants so that $ad - bc \neq 0$. (algebraic statistics here?) □

Optimization on the cross-section space

- The volume (length) of Γ at M_v is

$$\mathrm{length}(M_v) = \int_{-\infty}^{\infty} s(x;v)dx$$

where

$$s(x;v) = \frac{\sqrt{\frac{1-v^2}{v}}\sqrt{1+6vx^2+x^4}}{1+(\frac{1}{v}-3v)x^2+x^4}$$

(still elliptic integral...)

Theorem 3

Over $v \in (0, 1/3]$, the minimum of $\mathrm{length}(M_v)$ is attained if and only if $v = 1/3$, i.e.,

$$M = M_{1/3} = \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 1 \end{pmatrix}$$

The minimum is $2\pi\sqrt{\frac{2}{3}}$.

Proof of Theorem 3.

Using $1/\sqrt{1+z} \geq 1 - z/2$, construct a lower bound:

$$s(x;v) \geq \underline{s}(x;v) = \frac{\sqrt{\frac{1-v^2}{v}}(1+6vx^2+x^4)}{(1+(\frac{1}{v}-3v)x^2+x^4)(1+x^2)}\Big(1+\frac{(1-3v)x^2}{(1+x^2)^2}\Big)$$

(equality iff $v = 1/3$)

$\underline{s}(x;v)$ is a rational function and the integral can be evaluated by residues.

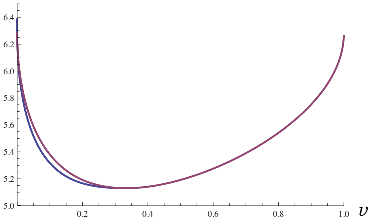
Fortunately,

$$\min_{v\in(0,1/3]} \int_{-\infty}^{\infty} \underline{s}(x;v)dx \quad \text{attains at } v = 1/3$$

and

$$\int_{-\infty}^{\infty} s(x;v)dx = \int_{-\infty}^{\infty} \underline{s}(x;v)dx \quad \text{at } v = 1/3$$

Optimization on the cross-section space (contd)



Objective function and its (integrable) lower bound

$$\text{length}(M_v) = \int_{-\infty}^{\infty} s(x;v)dx \;\geq\; \int_{-\infty}^{\infty} \underline{s}(x;v)dx$$

4. Summary and open problems

Optimal design in the polynomial regression ($n = 3$)

Theorem 4 (polynomial regression)

M is volume-mimimum optimal design iff

$$M = A M_{1/3} A^{\top}, \quad \exists A \in \mathcal{A}, \quad M_{1/3} = \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 1 \end{pmatrix}$$

Concretely

$$M = k \begin{pmatrix} 1 & r & \frac{q^2}{3} + r^2 \\ r & \frac{q^2}{3} + r^2 & r(q^2 + r^2) \\ \frac{q^2}{3} + r^2 & r(q^2 + r^2) & (q^2 + r^2)^2 \end{pmatrix}, \quad q \neq 0, \quad k > 0$$

Proof.

$A \in \mathcal{A}$ has a decomposition

$$A(a,b,c,d) = k \underbrace{A(q,r,0,1)}_{\text{affine}} \; \underbrace{A(\pm s, \mp t, t, s)}_{O(2) \text{ (isotropy group)}}, \quad \begin{cases} s = \cos\theta \\ t = \sin\theta \end{cases}$$

Optimal design in the polynomial regression ($n = 3$) (contd)

Remark 1

It is known that

$$M_{1/3} = \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 1 \end{pmatrix}$$

is the D-optimal information matrix.

$$\Big\{\textit{Whole designs}\Big\} \quad \textit{(4-dim)}$$
$$\supsetneq \Big\{\textit{Minimum-volume optimal designs}\Big\} \quad \textit{(2-dim)}$$
$$\supsetneq \Big\{\textit{D-optimal design}\Big\} \quad \textit{(0-dim)}$$

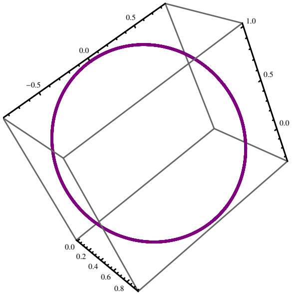
The D-optimal design is a universal optimal design that optimizes both D-criterion and volume criterion (hence, minimizes the width of simultaneous confidence bands).

Remark 2

The minimum-volume design $M \in \mathcal{M}$ is attained iff the curve

$$\Gamma_+ = \left\{ M^{-\frac{1}{2}} f(x) / \| M^{-\frac{1}{2}} f(x) \| \mid x \in \mathcal{X} \right\}$$

forms a circle.



Future topic: Multivariate extension

- $x = (x_i)^\top_{1 \le i \le p}$

$$f(x) = (1, (x_i)_{1 \le i \le p}, (x_i x_j)_{1 \le i \le j \le p}, \ldots, (x_{i_1} \cdots x_{i_d})_{1 \le i_1 \le \cdots \le i_d})^\top \in \mathbb{R}[x]^{\binom{p+d}{d}}$$

- Möbius transform $\overline{\mathbb{R}}^p \to \overline{\mathbb{R}}^p$

$$\varphi(x; A, b, c, d) = \frac{Ax + b}{c^\top x + d}, \quad A \in \mathbb{R}^{p \times p}, \quad b,\, c \in \mathbb{R}^{p \times 1}, \quad d \in \mathbb{R}$$

- "Volume preserving property" holds:

$$\mathrm{Vol}(M) = \mathrm{Vol}(AMA^\top), \quad A \in \mathcal{A},\; M \in \mathcal{M}$$

- Moment cone:

$$\mathcal{M} = \left\{ \int_{\mathcal{X} \subset \mathbb{R}^{\binom{p+d}{d}}} f(x) f(x)^\top dP(x) \mid P : \text{positive measure} \right\} = ?$$

Summary

- We proposed a new optimal design criterion — volume criterion.
- For the polynomial regression problems, the Möbius group acts on the moment cone $\mathcal{M}$, and keeps our problem invariant.
- When $n = 3$, by the optimization over cross-section space, we found the Möbius group orbit passing through the D-optimal design are minimum-volume optimal.
  (We conjecture that this is true for arbitrary n.)

References

- Kuriki, S. and Wynn, P. H., arXiv:1704.03995

Simultaneous confidence bands:

- Naiman, D. Q. (1986). Conservative confidence bands in curvilinear regression. *Ann. Statist.*, **14**, 896–906.
- Wynn, H. P. and Bloomfield, P. (1971). Simultaneous confidence bands in regression analysis (with discussions). *J. Roy. Statist. Soc., Ser. B*, **33**, 202–221.

Volume-of-tube method (≈ Expected Euler characteristic heuristic):

- Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*, Springer.
- Takemura, A. and Kuriki, S. (2002). On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Ann. Appl. Probab.*, **12** (2), 768–796.

Polynomial optimal design:

- Dette, H., Haines, L. M. and Imhof, L. (1999). Optimal designs for rational models and weighted polynomial regression, *Ann. Statist.*, **27** (4), 1272–1293.