european
social fund in the
czech republic

EUROPEAN UNION

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

OP Education
for Competitiveness

INVESTMENTS
IN EDUCATION
DEVELOPMENT

**Streamlining the Applied Mathematics Studies
at Faculty of Science of Palacký University in Olomouc
CZ.1.07/2.2.00/15.0243**

International Conference
Olomoucian Days of Applied Mathematics

ODAM 2011

Department of Mathematical analysis
and Applications of Mathematics
Faculty of Science
Palacký University Olomouc

Strange Design Points in Linear Regression

Andrej Pázman

Department of Applied Mathematics and Statistics
Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava

ODAM Olomouc 26.-28.1. 2011

Introduction

- Model under consideration

$$y_x = \mathbf{f}^{\mathsf{T}}(x)\boldsymbol{\theta} + \varepsilon_x, \; x \in \mathcal{X}$$

  - $\boldsymbol{\theta} \in \mathbb{R}^m$ – vector of unknown parameters,
  - $y_{x_1}, \ldots, y_{x_N}$ – observations at N design points $x_1, \ldots, x_N \in \mathcal{X}$,
  - $\mathcal{X}$ – design space.
- Without loss of generality $\mathcal{X} = \langle a, b \rangle$.
- In a vector notation

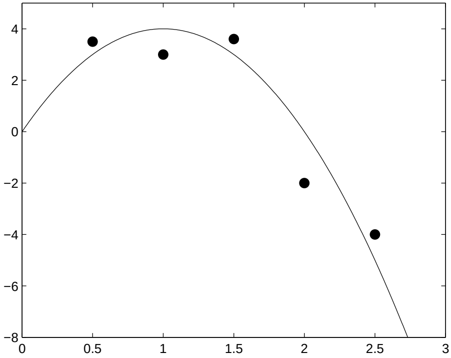
$$\begin{aligned}
\mathbf{y} &= \mathbf{F}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \text{ where} \\
\mathbf{y} &= (y_{x_1}, \ldots, y_{x_N})^{\mathsf{T}}, \\
\mathbf{F} &= \begin{pmatrix} \mathbf{f}^{\mathsf{T}}(x_1) \\ \vdots \\ \mathbf{f}^{\mathsf{T}}(x_N) \end{pmatrix} \\
\boldsymbol{\varepsilon} &= (\varepsilon_{x_1}, \ldots, \varepsilon_{x_N})^{\mathsf{T}}, \; \mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}, \; \mathbb{V}\mathrm{ar}[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{W}.
\end{aligned}$$

Introduction

- Particular case: Polynomial regression on an interval

$$y_x = \theta_1 + \theta_2 x + \cdots + \theta_m x^{m-1} + \varepsilon_x$$

  - $x \in \langle a, b \rangle$,
  - $\mathbf{f}(x) = (1, x, \ldots, x^{m-1})^{\mathrm{T}}$.



Introduction

- The estimator for $\boldsymbol{\theta}$ is given by (weighted) **least squares** (LS) method

$$\hat{\theta} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^m} (\mathbf{y} - \mathbf{F}\theta)^\top \mathbf{W}^{-1} (\mathbf{y} - \mathbf{F}\theta) = \mathbf{M}^{-1}\mathbf{F}^\top\mathbf{W}^{-1}\mathbf{y}.$$

  - $\mathbf{M} \equiv \mathbf{M}(x_1, \ldots, x_N) = \mathbf{F}^T\mathbf{W}^{-1}\mathbf{F}$ – the information matrix (for $\sigma = 1$).
- The Gauss-Markov theorem: $\hat{\theta}$ is BLUE.
- We have

$$\begin{aligned} \mathbb{V}\mathrm{ar}[\hat{\theta}] &= \sigma^2\mathbf{M}^{-1}, \text{ if } \det[\mathbf{M}] \neq 0, \\ \mathbb{V}\mathrm{ar}[\mathbf{h}^\top\hat{\theta}] &= \sigma^2\mathbf{h}^\top\mathbf{M}^-\mathbf{h}, \text{ if } \mathbf{h} \in \mathscr{M}(\mathbf{M}), \end{aligned}$$

  - $\mathbf{M}^-$ – an arbitrary g-inverse of $\mathbf{M}$,
  - $\mathscr{M}(\mathbf{M})$ – the column space of $\mathbf{M}$,
  - $\hat{\theta}$ solves the normal equation

$$\mathbf{M}\theta = \mathbf{F}^\top\mathbf{W}^{-1}\mathbf{y}.$$

Introduction

- Evidently, the position of the design points $x_1, \dots, x_N$ influences the variances, as well as the form of the estimated regression function

$$\eta(x, \hat{\boldsymbol{\theta}}) = \mathbf{f}^\top(x)\hat{\boldsymbol{\theta}}.$$

- **It seems, however, that nothing surprising can be found in this model. We shall try to prove the opposite.**

Design points, which cannot be approximated by neighborhood design points

Consider the following setting:

- The model = a quadratic polynomial without intercept:

$$\eta(x, \boldsymbol{\theta}) = \theta_1 x + \theta_2 x^2 \tag{1}$$

- $\mathcal{X} = \langle 0, 10 \rangle$.
- The aim is to estimate the value of $\eta(x, \boldsymbol{\theta})$ at $\bar{x} = 1$.
- We have to perform 10 independent observations.

First design: all observations at one point

- "Natural" design: $x_1 = \ldots = x_{10} = \bar{x}$.
- Then we have

$$\begin{aligned}
\mathbf{M}(\bar{x}) &= 10 \begin{pmatrix} \bar{x} \\ \bar{x}^2 \end{pmatrix} \begin{pmatrix} \bar{x} & \bar{x}^2 \end{pmatrix} = 10 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \\
\mathbf{M}^{-}(\bar{x}) &= \frac{1}{40} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \\
\mathbb{V}\mathrm{ar}[\eta(\bar{x}, \hat{\boldsymbol{\theta}})] &= \sigma^2 \begin{pmatrix} 1 & 1 \end{pmatrix} \frac{1}{40} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{\sigma^2}{10} = \mathbb{V}\mathrm{ar}\left[\frac{1}{10} \sum_{i=1}^{10} y_i\right].
\end{aligned}$$

Design points, which cannot be approximated by neighborhood design points

Observations at two points

Second design: observations at two points

- 5 observations at $x_1 = \bar{x} + t$ and 5 at $x_2 = \bar{x} + ct$, $t > 0$ small, $c \in \langle -1, 1 \rangle$.
- Then we have

$$
\begin{aligned}
\mathbf{M}(x_1, x_2) &= 5 \begin{pmatrix} x_1^2 & x_1^3 \\ x_1^3 & x_1^4 \end{pmatrix} + 5 \begin{pmatrix} x_2^2 & x_2^3 \\ x_2^3 & x_2^4 \end{pmatrix}, \\
\mathbf{M}^{-1}(x_1, x_2) &= \frac{1}{5x_1^2 x_2^2 (x_1 - x_2)^2} \begin{pmatrix} x_1^4 + x_2^4 & -x_1^3 - x_2^3 \\ -x_1^3 - x_2^3 & x_1^2 + x_2^2 \end{pmatrix}, \\
\mathbb{V}\mathrm{ar}[\eta(\bar{x}, \hat{\boldsymbol{\theta}})] &= \sigma^2 \begin{pmatrix} 1 & 1 \end{pmatrix} \mathbf{M}^{-1}(x_1, x_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
&= \frac{\sigma^2}{5(1-c)^2} \left[\frac{1}{(1+ct)^2} + \frac{c^2}{(1+t)^2} \right],
\end{aligned}
$$

- In the limit we have

$$\mathbb{V}\mathrm{ar}[\eta(\bar{x}, \hat{\boldsymbol{\theta}})] \to \frac{\sigma^2(1+c^2)}{5(1-c)^2}$$

  for $t \to 0$.
- Consequently, $\mathbb{V}\mathrm{ar}[\eta(\bar{x}, \hat{\boldsymbol{\theta}}]$ **is arbitrary large if we take** c **sufficiently close to** 1**, regardless how small is** t**, i.e., how close are the points** x_1 **and** x_2 **to** $\bar{x} = 1$.
- For any $c \neq -1$ the limit variance is larger than when performing all observations at one point $\bar{x}$.

Consequences:

i) **mathematically**, we have a very clear discontinuity of the variance when considering it as a function of the design,

ii) **we have a conflict between singular and regular regression models.**

- We observe just at one point $\bar{x} \Rightarrow$ singular regression model, since just one important parameter, namely $\eta(\bar{x}, \boldsymbol{\theta})$.
- We observe at two points, we have a regular model with two parameters to be estimated. If the points are very close to $\bar{x}$, the model is "bad conditioned" $\Rightarrow$ some functions of the parameters are estimated with high variances, in our case $\eta(\bar{x}, \boldsymbol{\theta})$.

Design points, which cannot be approximated by neighborhood design points

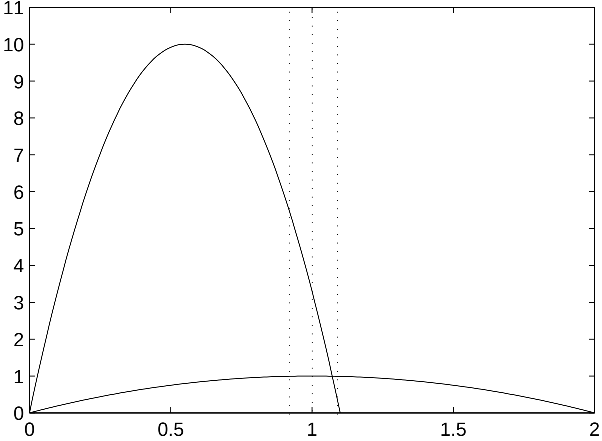
Consequences

iii) We have a statistical paradox on an elementary level.

- If x_1, x_2 are very close to $\bar{x}$, in practice one can not be sure whether we observed at $\bar{x}$ or at x_1, and x_2.
- the estimator $\eta(\bar{x}, \hat{\boldsymbol{\theta}})$ should not be so sensitive to the choice of design points as given by the theory.
- However, this reflection contains implicitly the a priori assumption that the values $\eta(\bar{x}, \boldsymbol{\theta})$, $\eta(x_1, \boldsymbol{\theta})$, $\eta(x_2, \boldsymbol{\theta})$ do not differ very much.
- Hence, quadratic model (1) is not adequate for such an assumption (it does not exclude a very narrow parabola) $\Rightarrow$ gives very different values of $\eta(\bar{x}, \boldsymbol{\theta})$, $\eta(x_1, \boldsymbol{\theta})$, $\eta(x_2, \boldsymbol{\theta})$ even if $\bar{x}$, x_1, x_2 are very close to each other.

Design points, which cannot be approximated by neighborhood design points

Consequences



Design points, which cannot be approximated by neighborhood design points

Consequences

So...

... the "paradoxical" result obtained from the theory is statistically correct, but the model does not correspond to our "intuitive" assumptions.

To go out we either must

- use a **Bayesian modelling** rejecting "a priori" the possibility of a narrow parabola, or we must
- use another regression model, say we must suppose that for x in a neighborhood of the point $\bar{x}$ a one-parameter model

$$\eta(x, \boldsymbol{\theta}) = \theta_1$$

is acceptable.

Design points, which cannot be approximated by neighborhood design points

Nonlinear parametric function

The situation is even worse when we want to estimate **a nonlinear function of θ_1 and θ_2**:

- Suppose we want to estimate the position x_0 of the extreme point:

$$\left.\frac{\mathrm{d}(\theta_1 x + \theta_2 x^2)}{\mathrm{d}x}\right|_{x=x_0} = 0 \;\Rightarrow\; x_0 = -\frac{\theta_1}{2\theta_2} \;\Rightarrow\; \hat{x}_0 = -\frac{\hat{\theta}_1}{2\hat{\theta}_2}.$$

- AP&LP (Statistics & Probability Letters (2006)): If we observe m times at the "true" point x_0 and $N - m$ times at a point x_1 which is close to x_0, it may happen that the limit distribution of $-\frac{\hat{\theta}_1}{2\hat{\theta}_2}$ is asymtotically not normal for $N \to \infty$, or still normal but with a "strange" variance. All depends on the behavior of m/N.

Design points which have to be replaced by distant design points

- Again consider the quadratic model

$$y_x = \eta(x, \boldsymbol{\theta}) + \varepsilon_x = \theta_1 x + \theta_2 x^2 + \varepsilon_x,$$

with $N = 10$ independent observations at hand, $\bar{x} = 1$, and we estimate the value of $\eta(\bar{x}, \boldsymbol{\theta})$.

Is the strategy to perform all N observations at $\bar{x}$ optimal?

...

No!

- A better strategy: 87% of observations at $x_1 = 4.14$ and 13% of observations at $x_2 = 10$ (although our aim is to estimate the polynomial at $\bar{x} = 1$).
  This follows from the following Elfving construction:

Design points which have to be replaced by distant design points

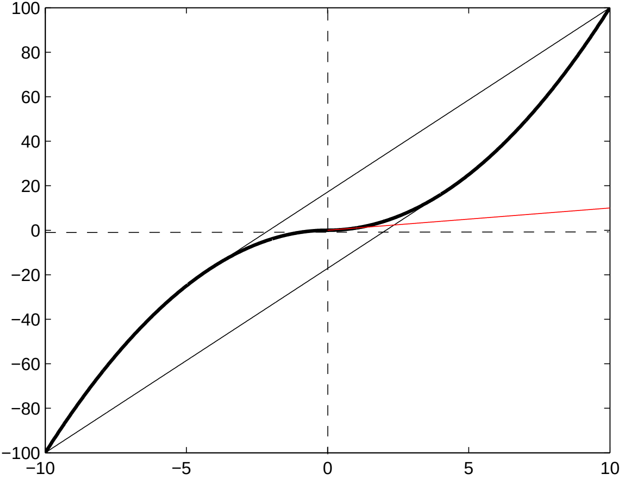
Elfving Theorem

Let ...

- $y = \mathbf{f}^{\mathsf{T}}(x)\boldsymbol{\theta} + \varepsilon_x$,
- $\mathbb{V}\mathrm{ar}[\varepsilon_x] = \sigma^2$,
- observations be independent,
- $\mathcal{S} = \mathrm{co}(\{\mathbf{f}(x) : x \in \mathcal{X}\} \cup \{-\mathbf{f}(x) : x \in \mathcal{X}\})$
  $\mathrm{co}(T)$ – minimal convex set containing T = convex hull of T.

Design points which have to be replaced by distant design points

Elfving Theorem



Design points which have to be replaced by distant design points

Elfving Theorem

Theorem

Suppose that we want to estimate $\mathbf{h}^\top \boldsymbol{\theta}$ *optimally. Take* $\beta \geq 0$ *such that the point* $\beta \mathbf{h}$ *is on the boundary of the set* $\mathcal{S}$. *Take points* $x_1^*, \ldots, x_k^* \in \mathcal{X}$ *such that*

1. *the points* $\mathbf{f}(x_1^*), \ldots, \mathbf{f}(x_k^*)$ *are on the boundary of* $\mathcal{S}$,
2. *the vector* $\beta \mathbf{h}$ *can be expressed as* $\beta \mathbf{h} = \sum_{i=1}^k \gamma_i \mathbf{f}(x_i^*)$, *with* $\gamma_i \in \langle -1, 1 \rangle$ *and* $\sum_{i=1}^k |\gamma_i| = 1$.

Then

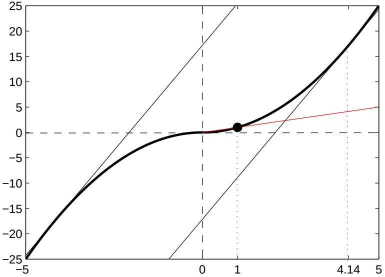
1. *the design giving* $|\gamma_i| \times 100\,\%$ *of (uncorrelated) observations at* x_i^* *is the design minimizing* $\mathbb{V}\mathrm{ar}[\mathbf{h}^\top \hat{\boldsymbol{\theta}}]$, *and*
2. *the corresponding minimal variance is equal to* $\frac{\sigma^2}{N\beta^2}$.

Design points which have to be replaced by distant design points

Elfving Theorem

- In our case the vector $\mathbf{f}(\bar{x})$ is in the interior of $\mathcal{S}$. So we have to take $\beta > 1$ and the variance of the optimal design gives

$$\mathbb{V}\mathrm{ar}[\eta(\bar{x}, \hat{\boldsymbol{\theta}})] < \mathbb{V}\mathrm{ar}\left(\frac{1}{10}\sum_{i=1}^{10} y_i\right)$$



Design points giving zero information about $\boldsymbol{\theta}$

- General model
$$y_x = \mathbf{f}^{\mathsf{T}}(x)\boldsymbol{\theta} + \varepsilon_x,$$
$\mathbb{E}[\varepsilon_x] = 0.$
- If $\mathbb{V}\mathrm{ar}[\varepsilon_x] = \sigma^2 \mathbf{I}$ and observations are uncorrelated, then the information matrix
$$\mathbf{M}(x_1, \ldots, x_N) = \sum_{i=1}^{N} \mathbf{f}(x_i)\mathbf{f}^{\mathsf{T}}(x_i).$$
- $\mathbf{f}(x_i)\mathbf{f}^{\mathsf{T}}(x_i)$ – information at one design point x_i, does not depend on the used design.

In the uncorrelated case a design point x_i gives zero information about $\boldsymbol{\theta}$ if and only if $\mathbf{f}(x_i) = \mathbf{0}$.

Design points giving zero information about $\boldsymbol{\theta}$

The situation may be quite different, and somehow surprising, when the observations are correlated:

- $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$, $\mathbb{V}\mathrm{ar}[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{W}$, $\det[\mathbf{W}] \neq 0$.
- $$\mathbf{M}(x_1, \ldots, x_N) = \mathbf{F}^\top \mathbf{W}^{-1} \mathbf{F} = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{f}(x_i) \{\mathbf{W}^{-1}\}_{ij} \mathbf{f}^\top(x_j).$$
- Even when $\mathbf{W}$ is known, it is not quite transparent, which design points give zero information about $\boldsymbol{\theta}$.

Intuitively, one can perhaps argue also here that $\mathbf{f}(x) = \mathbf{0}$ implies that in the model $y_x = \mathbf{f}^\top(x)\boldsymbol{\theta} + \varepsilon_x$, $x \in \mathcal{X}$, y_x is not influenced by the value of $\boldsymbol{\theta}$, hence should give no information about $\boldsymbol{\theta}$.
This intuitive approach is false.

Design points giving zero information about θ

Example:

- $\theta \in \mathbb{R}$, take $\{x, z\}$ a two point design such that $f(x) = 0$, $f(z) = 1$, and suppose that $W_{xx} = W_{zz} = 1$, but $W_{xz} \neq 0$.
- We have

$$\mathbf{M}(x,z) = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & W_{xz} \\ W_{xz} & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{1-(W_{xz})^2} > 1 = \mathbf{M}(\{z\}).$$

- So although $f(x) = 0$, by deleting the point x from the design we can lose much information. The contribution of the point x to $\mathbf{M}(x,z)$ is very large even if $f(x) = 0$, when the observations y_x and y_z are highly correlated.

Design points giving zero information about $\boldsymbol{\theta}$

Proposition

(AP, Kybernetika (2010)) Suppose that the point x_k is deleted from the design $D = \{x1, \dots, x_N\}$. Then the resulting information matrix is

$$\mathbf{M}(D - \{x_k\}) = \mathbf{M}(D) - \frac{\mathbf{a}(x_k)\mathbf{a}^\mathsf{T}(x_k)}{\{\mathbf{W}^{-1}\}_{k,k}},$$

where

$$\mathbf{a}(x_k) = \sum_{i=1}^{N} \{\mathbf{W}^{-1}\}_{k,i}\mathbf{f}(x_i).$$

Consequently, the design point x_k gives zero information about $\boldsymbol{\theta}$ if and only if $\mathbf{a}(x_k) = \mathbf{0}$.

Design points giving zero information about θ

Proof.

Take $\mathbf{s} = (0, \ldots, 0, c, 0, \ldots, 0)^\mathsf{T}$ having $c > 0$ at the kth coordinate. We have (cf. Harville)

$$(\mathbf{W} + \mathbf{s}\mathbf{s}^\mathsf{T})^{-1} = \mathbf{W}^{-1} - \frac{\mathbf{W}^{-1}\mathbf{s}\mathbf{s}^\mathsf{T}\mathbf{W}^{-1}}{1 + \mathbf{s}^\mathsf{T}\mathbf{W}^{-1}\mathbf{s}}.$$

Hence

$$\mathbf{M}(D - \{x_k\}) = \lim_{c\to\infty} \mathbf{F}^\mathsf{T}(\mathbf{W} + \mathbf{s}\mathbf{s}^\mathsf{T})^{-1}\mathbf{F} = \mathbf{M}(D) - \lim_{c\to\infty} \frac{c^2\mathbf{a}(x_k)\mathbf{a}^\mathsf{T}(x_k)}{1 + c^2\{\mathbf{W}^{-1}\}_{k,k}}.$$

$\square$

Design points giving zero information about $\boldsymbol{\theta}$

- To express that the information at a design point x_k is small we need measures of information which are one dimensional (scalars) – we shall consider information functionals, which are concave, monotone, real-valued functions defined on the set of positive definite matrices (see Pukelsheim (1993)).
- The gradient of Φ is $\nabla\Phi[\mathbf{M}]$, with

$$\{\nabla\Phi[\mathbf{M}]\}_{ij} = \frac{\partial\Phi[\mathbf{M}]}{\partial\{M\}_{ij}}.$$

	function	**gradient**
D-optimality	$\ln\det(\mathbf{M})$	$\mathbf{M}^{-1}$
A-optimality	$-\mathrm{tr}(\mathbf{M}^{-1})$	$\mathbf{M}^{-2}$

Design points giving zero information about $\boldsymbol{\theta}$

- For a fixed design $D = \{x_1, \ldots, x_N\}$ we denote

$$\|\mathbf{a}(x_k)\|_\Phi^2 = \mathbf{a}^\mathsf{T}(x_k)\nabla\Phi[\mathbf{M}(D)]\mathbf{a}(x_k),$$

  which is a (pseudo)norm, since concavity of Φ implies that the gradient $\nabla\Phi[\mathbf{M}]$ is a positive (semi)definite matrix of $\mathbf{M}$ (see WM&AP, Biometrika (2003)).

- Example: $\Phi[\mathbf{M}] = \ln\det(\mathbf{M})$

$$\|\mathbf{a}(x_k)\|_\Phi^2 = \mathbf{a}^\mathsf{T}(x_k)\mathbf{M}^{-1}(D)\mathbf{a}(x_k).$$

Design points giving zero information about θ

Proposition

We have

$$\Phi[\mathbf{M}(D - \{x_k\})] = \Phi[\mathbf{M}(D)] - \frac{\|\mathbf{a}(x_k)\|_\Phi^2}{\{\mathbf{W}^{-1}\}_{x_k,x_k}} + o(\|\mathbf{a}(x_k)\|_\Phi^3),$$

with $\lim_{t\to 0} o(t)/t = 0$. Consequently, the amount of information obtained from the design point x_k is small iff the expression $\frac{\|\mathbf{a}(x_k)\|_\Phi^2}{\{\mathbf{W}^{-1}\}_{x_k,x_k}}$ is small.

Design points giving zero information about $\boldsymbol{\theta}$

Proof.

From the Taylor formula applied to the matrix function $\mathbf{M} \mapsto \Phi[\mathbf{M}]$ we obtain

$$
\begin{aligned}
\Phi[\mathbf{M}(D - \{x_k\})] &= \Phi\left[\mathbf{M}(D) - \frac{\mathbf{a}(x_k)\mathbf{a}^{\mathsf{T}}(x_k)}{\{\mathbf{W}^{-1}\}_{k,k}}\right] \\
&= \Phi[\mathbf{M}(D)] - \mathrm{tr}\left\{\nabla\Phi[\mathbf{M}(D)]\left[\frac{\mathbf{a}(x_k)\mathbf{a}^{\mathsf{T}}(x_k)}{\{\mathbf{W}^{-1}\}_{k,k}}\right]\right\} \\
&\quad + \text{term of order } \|\mathbf{a}(x_k)\|_{\Phi}^{4} \\
&= \Phi[\mathbf{M}(D)] - \frac{\|\mathbf{a}(x_k)\|_{\Phi}^{2}}{\{\mathbf{W}^{-1}\}_{k,k}}.
\end{aligned}
$$

$\square$

Design points sensitive to outliers

Model:

$$\begin{aligned} y_i &= \mathbf{f}^\mathsf{T}(x_i)\boldsymbol{\theta} + \varepsilon_i, \; i = 1, \ldots, N, \\ & \mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}, \mathbb{V}\mathrm{ar}[\boldsymbol{\varepsilon}] = \sigma^2\mathbf{W}. \end{aligned}$$

Proposition

Suppose that the information matrix $\mathbf{M}$ is nonsingular. Then for any $i = 1, \ldots, N$ we have

$$\mathbf{f}^\mathsf{T}(x_i)\mathbf{M}^{-1}\mathbf{f}(x_i) \leq \{\mathbf{W}\}_{ii}.$$

In the extreme case that

$$\mathbf{f}^\mathsf{T}(x_i)\mathbf{M}^{-1}\mathbf{f}(x_i) = \{\mathbf{W}\}_{ii},$$

the graph of the estimated regression function $x \in \langle a, b\rangle \mapsto \eta(x, \hat{\boldsymbol{\theta}})$ contains the point $[x_i, y_{x_i}]$.

Design points sensitive to outliers

Proof.

Using the expression for the estimate $\hat{\boldsymbol{\theta}}$ we have

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\mathbf{y}-\mathbf{F}\hat{\boldsymbol{\theta}}] &= \mathbb{V}\mathrm{ar}[(\mathbf{I}-\mathbf{F}\mathbf{M}^{-1}\mathbf{F}^{\mathsf{T}}\mathbf{W}^{-1})\mathbf{y}] \\ &= \sigma^2(\mathbf{W}-\mathbf{F}\mathbf{M}^{-1}\mathbf{F}^{\mathsf{T}}).\end{aligned}$$

Consequently, $\mathbf{f}^{\mathsf{T}}(x_i)\mathbf{M}^{-1}\mathbf{f}(x_i) = \{\mathbf{F}\mathbf{M}^{-1}\mathbf{F}^{\mathsf{T}}\} \leq \{\mathbf{W}\}_{ii}$. If $\{\mathbf{F}\mathbf{M}^{-1}\mathbf{F}^{\mathsf{T}}\} = \{\mathbf{W}\}_{ii}$, then $\mathbb{E}[(y_i - \mathbf{F}_{i\cdot}\hat{\boldsymbol{\theta}})^2] = \mathbb{V}\mathrm{ar}[y_i - \mathbf{F}_{i\cdot}\hat{\boldsymbol{\theta}}] = 0$, hence

$$y_i = \mathbf{F}_{i\cdot}\hat{\boldsymbol{\theta}} = \eta(x_i, \hat{\boldsymbol{\theta}}),$$

with probability 1. □

Design points sensitive to outliers

Corollary

If $\mathbf{f}^{\mathsf{T}}(x_i)\mathbf{M}^{-1}\mathbf{f}(x_i)$ is close to $\{\mathbf{W}\}_{ii}$, then, even before performing the experiment, we know that the whole estimated regression function is strongly influenced by y_i, even if y_i is an outlier.

Remark

In the case that $\mathbf{W} = \mathbf{I}$ (case of uncorrelated observations with constant variances), this emphasizes the importance of the use the G-optimality criterion of optimality

$$\max_{x \in \langle a,b \rangle} \mathbf{f}^{\mathsf{T}}(x)\mathbf{M}^{-1}(x_1, \ldots, x_N)\mathbf{f}(x).$$

The minimization of this expression with respect to $x_1, \ldots, x_N$ gives a design which is good not only for the precision of the response function, but also for its robustness with respect to outliers.

Conclusion

In such a simple model as is a polynomial regression on a real line we described three kinds of “strange” design points. So “one can be never too careful” even with linear models.

References

- Harville D.A.: Matrix Algebra from a Statistician's Perspective. Springer, New York, 1997.
- Müller W.G., Pázman A.: Measures for designs in experiments with correlated errors. *Biometrika* **90** (2003), 423–445.
- Pázman A.: Information contained in design points of experiments with correlated observations. *Kybernetika* **46** (2010), 769–781.
- Pázman A., Pronzato L.: On the irregular behaviour of LS estimators for asymptotically singular designs. *Statistics and Probability Letters* **76** (2006), 1089–1096.
- Pukelsheim F.: Optimal Design of Experiments. Wiley, New York, 1993.
- Zvára K.: Regresní analýza. Academie Praha, 1989.