THE BOOT STRAP METHOD AND MLE

AHC-LAB CC-GROUP
M1 RUI HIRAOKA

The Bootstrap and Maximum likelihood methods

- 8.2.1 A Smoothing Example
  - Non-Parametric bootstrap
  - Parametric bootstrap

- 8.2.3 MLE vs The bootstrap

THE BOOTSTRAP METHOD

B. Efron (1979)	Resampling
Simple	computational
Maximum - likelihood	Interval-estimation

8.2 The Bootstrap and Maximum likelihood methods

- 8.2.1 A Smoothing Example
  - Non-Parametric bootstrap
  - Parametric bootstrap
- 8.2.3 MLE VS The bootstrap

EXAMPLE (2-DEMENTION)

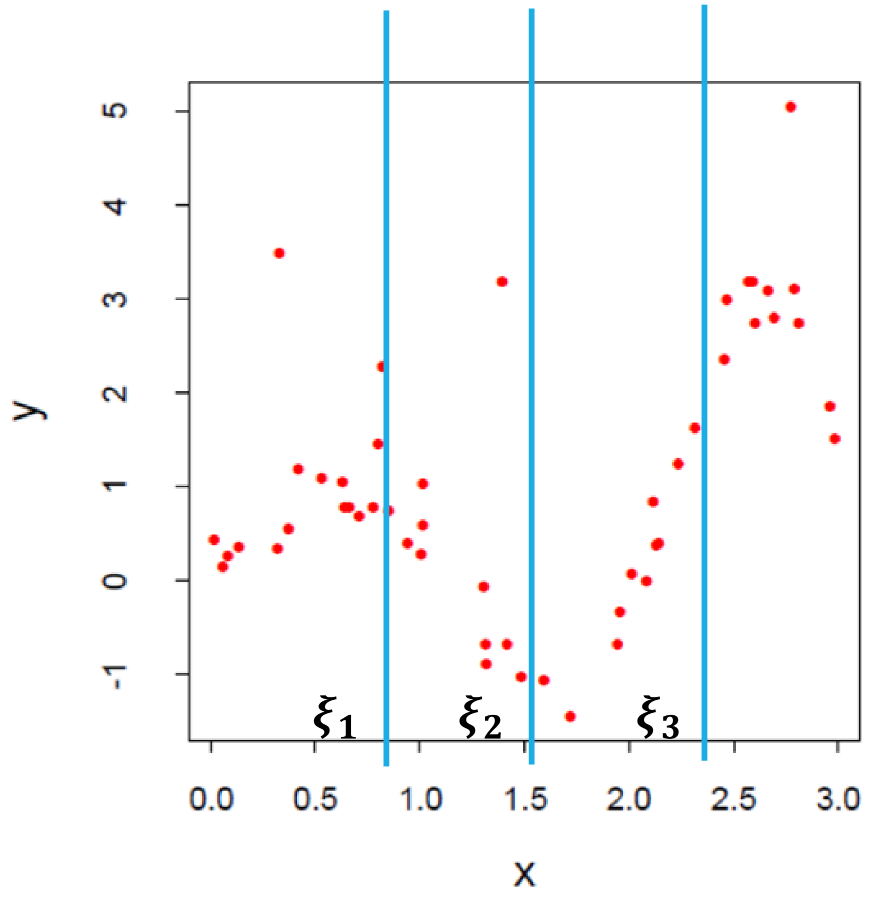
2-dementional training DATA **Z** (red bots)

$\boldsymbol{Z_i = (x_i, y_i)} \quad \boldsymbol{i = 1 \sim 50}$

$\boldsymbol{x_i}$: 1-D input

Apply a Cubic spline with 3 knots (blue lines)

X= ξ1, ξ2, ξ3



VARIABLES OF CUBIC SPLINES

Cubic B-spline with 3 knots

For each node

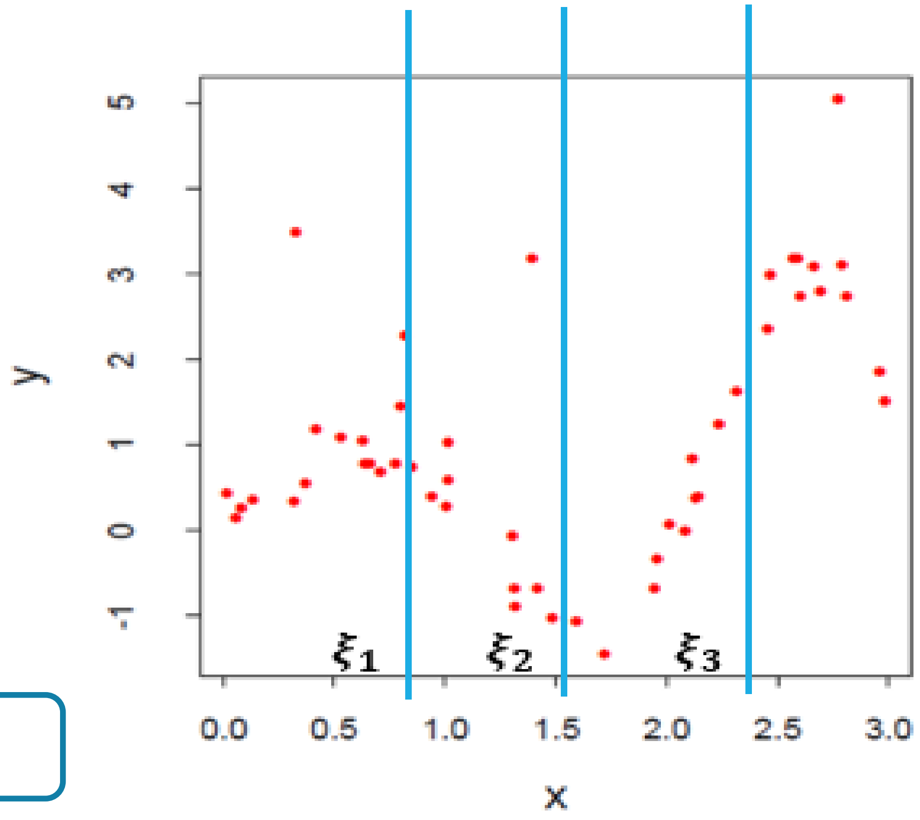
1 set of 3 Constrains

(continuity condition)

For each area

1 set of 4 variables

$(a_0 + \alpha_1 x + a_2 x^2 + a_3 x^3)$



Freedom

4*4 variables - 3*3 constrains=16-9=7

A CUBIC SPLINES MODEL

$$\boxed{\mu(x) = \sum_{j=1}^{7} \beta_j h_j(x). \qquad (8.1)} \longleftarrow$$

Model $\boldsymbol{y} = \boldsymbol{\mu}(\boldsymbol{x})$

$h_j(x)$ is given by $(x^0, x^1, x^2, x^3, \xi_1, \xi_2, \xi_3)$

Need to determine coefficients

$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^T$

INITIAL ESTIMATION (LSE)

$$\mu(x) = \sum_{j=1}^{7} \beta_j h_j(x). \qquad (8.1)$$

By LSE (Least square error)

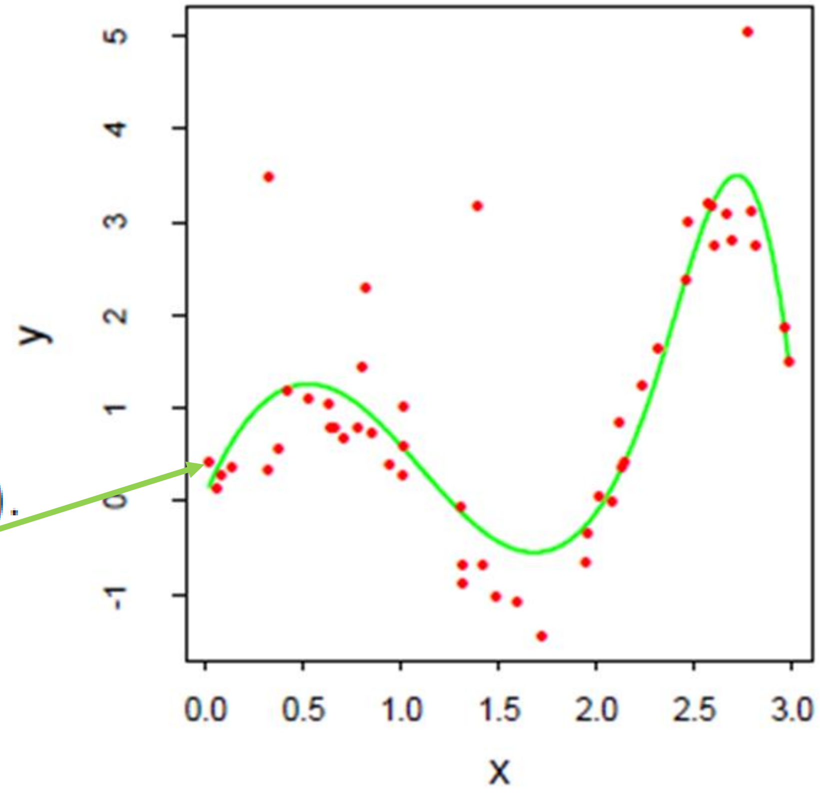
$$\hat{\beta} = \operatorname{arg}min \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{7} \beta_j h_j(x_i)\right)^2$$

$$\hat{\beta} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}. \qquad (8.2)$$

Let $\mathbf{H}$ be the $N \times 7$ matrix with ijth element $h_j(x_i)$.

(——) in the left Fig

$\hat{\mu}(x) = \sum_{j=1}^{7} \hat{\beta}_j h_j(x)$



INITIAL ESTIMATION 2 (LSE)

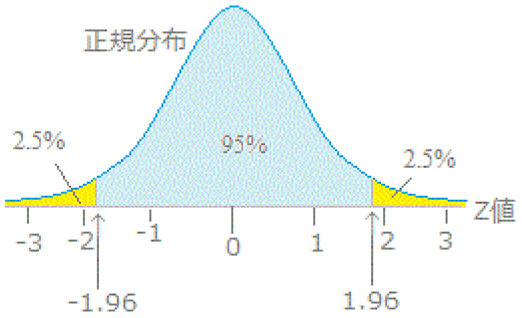


Fig Gaussian [1]

Covariance

$$\widehat{\mathrm{Var}}(\hat{\beta}) = (\mathbf{H}^T\mathbf{H})^{-1}\hat{\sigma}^2, \quad (8.3)$$

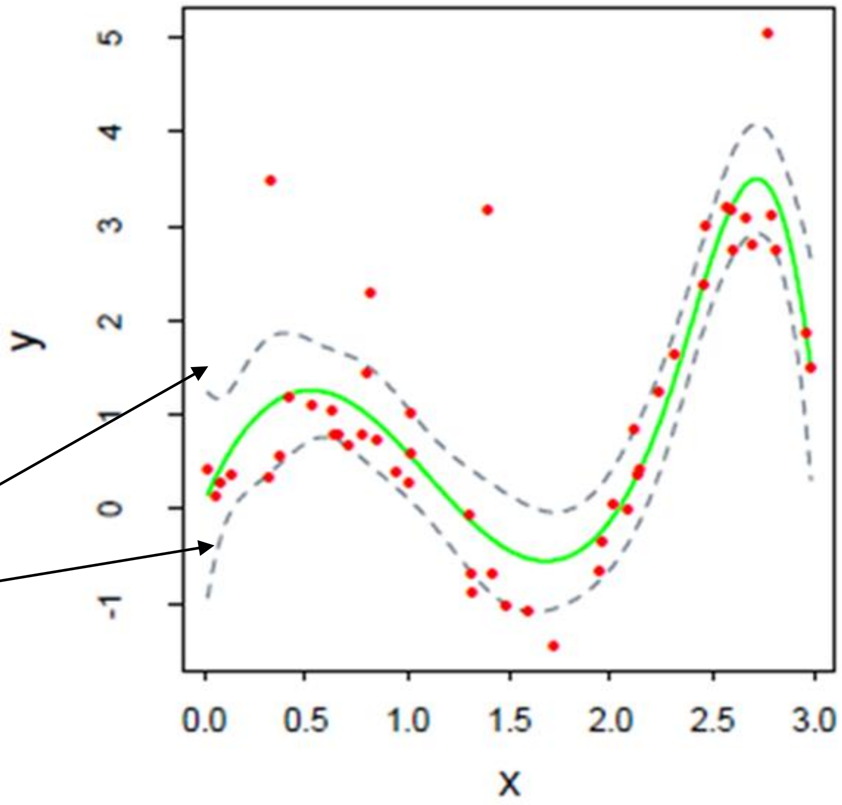
$$\hat{\sigma}^2 = \textstyle\sum_{i=1}^N (y_i - \hat{\mu}(x_i))^2/N.$$

Standard error

$$\widehat{\mathrm{se}}[\hat{\mu}(x)] = [h(x)^T(\mathbf{H}^T\mathbf{H})^{-1}h(x)]^{\frac{1}{2}}\hat{\sigma}. \quad (8.4)$$

95% pointwise band (--- in left Fig)

$$\hat{\mu}(x) \pm 1.96 \cdot \widehat{\mathrm{se}}[\hat{\mu}(x)]$$



HOW TO APPLY THE BOOTSTRAP METHOD?

ALGORITHM (NON-PARAMETRIC)

<Given a dataset Z= $(z_1, z_2, z_3 \dots z_n)$ >

Make a model with parameter β (8.1)

Estimate β with raw data Z ← (LSE) : Done

For b in range B (iteration)

Z_b^* =a set of N random Resamples from Z

Non-Parametric

$\beta_b^* \ = \ \beta(Z_b^*)$ ← (LSE)

Result $\boldsymbol{\beta_1^*, \beta_2^*, \beta_3^* \dots \beta_B^*}$

SIMPLE EXAMPLE : HOW TO EXTRACT DATA

<Given a dataset Z= $\left(\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\4\end{pmatrix} \dots\dots\dots\dots\dots\dots\dots\dots \begin{pmatrix}n-1\\(n-1)^2\end{pmatrix}, \begin{pmatrix}n\\n^2\end{pmatrix}\right)$

Make a model with parameter β

Allow to overlap

Estimate β with raw data Z

$$Z^* = \underbrace{\left(\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\4\end{pmatrix}, \begin{pmatrix}2\\4\end{pmatrix} \dots \begin{pmatrix}n-1\\(n-1)^2\end{pmatrix}\right)}_{n}$$

Resampling example

For b in range B

$Z_b^* \;$ = a set of N random Resample from Z

$\beta_b^* = \beta(Z_b^*)$

Result $\boldsymbol{\beta_1^*, \beta_2^*, \beta_3^* \dots \beta_B^*}$

10 DRAWING EXAMPLES

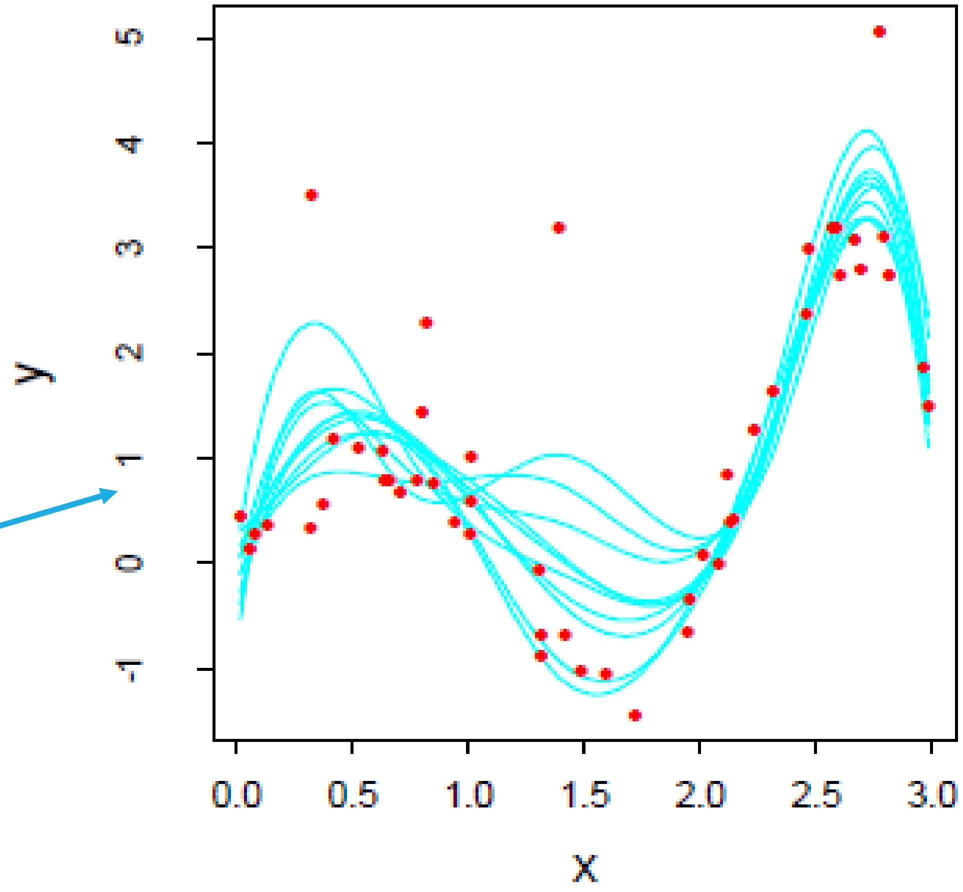
B=10 in previous algorithm

(——) lines in the left Fig

$$\mu(x) = \sum_{j=1}^{7} \beta^*_{j,b} h_j(x)$$

$$b = 1 \sim 10$$

10 lines



200 DRAWING EXAMPLES

B=200 in previous algorithm

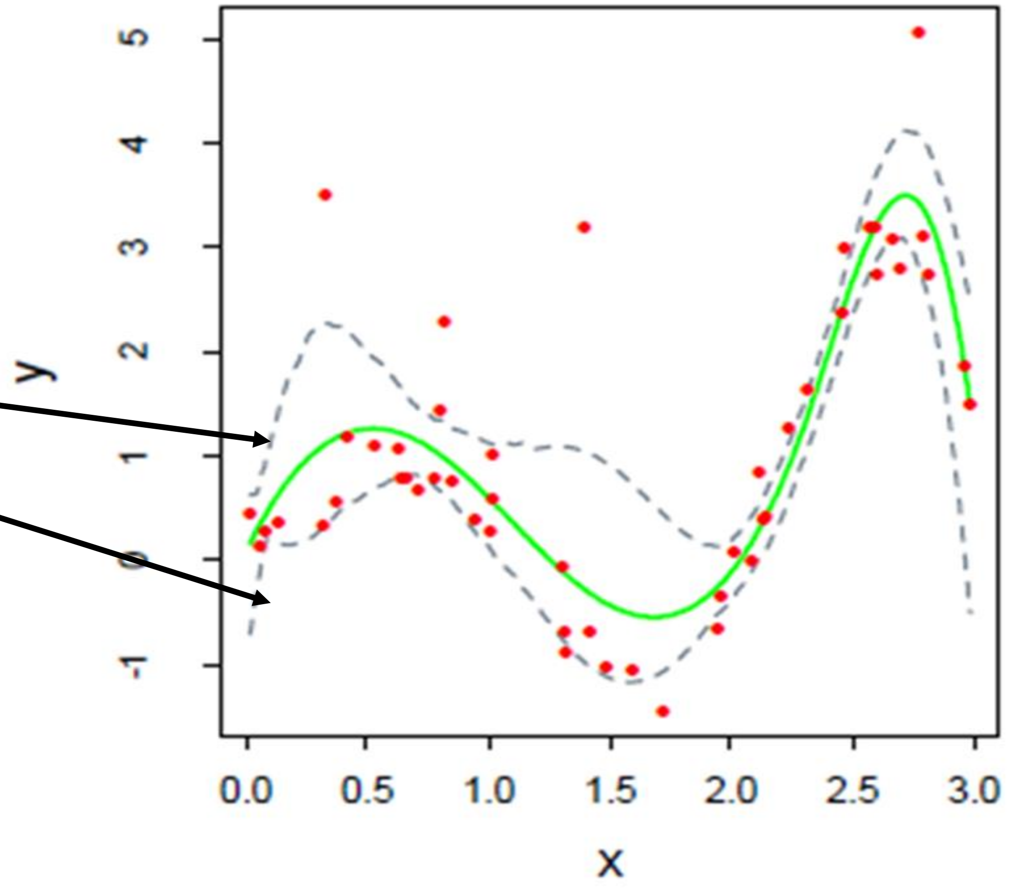
(-----) lines in the left Fig

<u>95%* pointwise band</u>

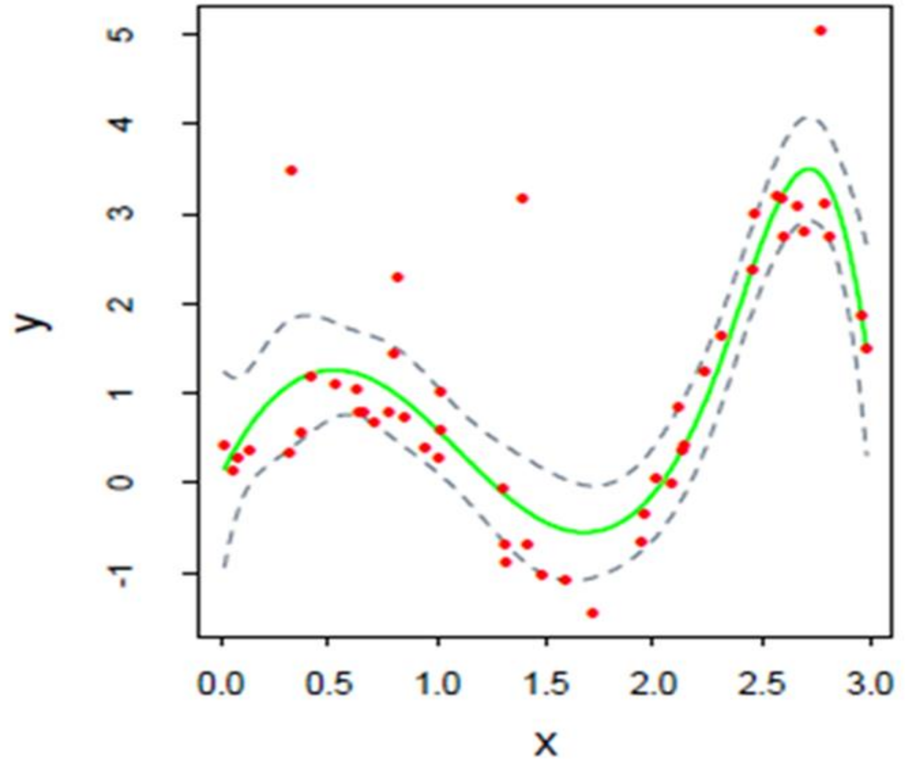
$$\boldsymbol{\mu}(\boldsymbol{x}) = \sum_{j=1}^{7} \boldsymbol{\beta}^{*}_{j,b} \boldsymbol{h}_j(\boldsymbol{x}) \;\; \boldsymbol{b} = \mathbf{1} \sim \mathbf{200}$$

*top 5 and bottom 5 are eliminated

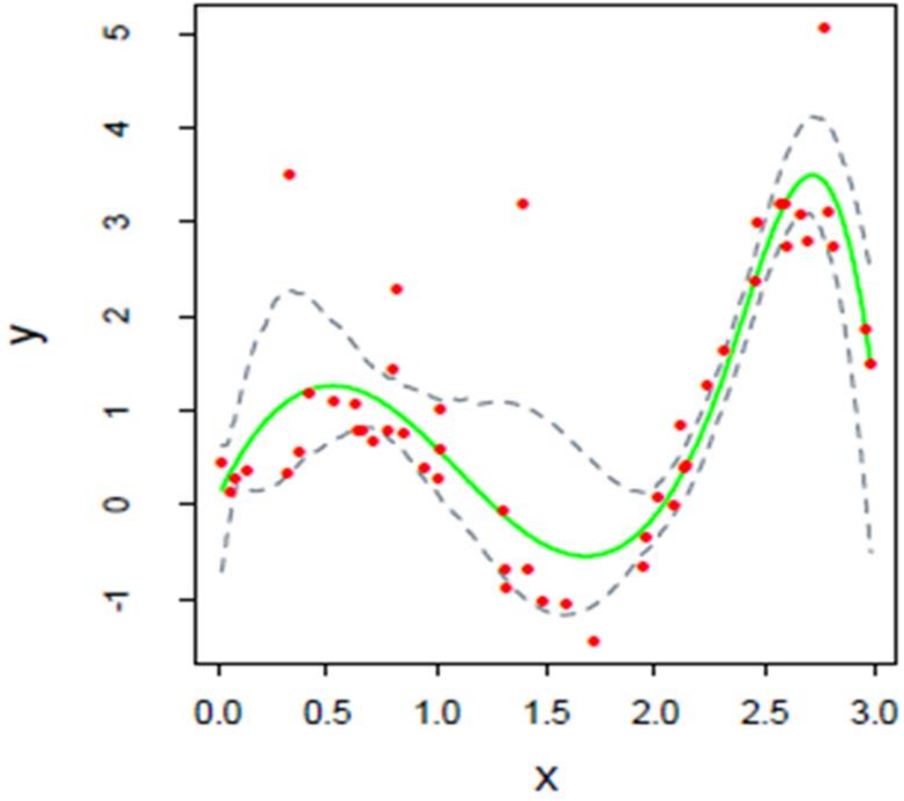
<u>(200-10)/200 = 95%</u>



COMPARISON : 95% POINTWISE BANDS



LSE



B=200 The Bootstrap

ALGORITHM (PARAMETRIC)

<Given a dataset Z= $(z_1, z_2, z_3 \ldots z_n)$ >

Make a model with parameter β (8.1)

Estimate $\hat{\beta}$ with raw data Z ← (LSE or MLE) $\boldsymbol{\hat{\mu}(x) = \sum_{j=1}^{7} \hat{\beta}_j \; h_j(x)}$

For b in range B

$$\boldsymbol{Z_b^*} = \left(\begin{pmatrix} \boldsymbol{x_1^*} \\ \boldsymbol{\hat{\mu}(x_1^*) + \varepsilon_1} \end{pmatrix}, \begin{pmatrix} \boldsymbol{x_2^*} \\ \boldsymbol{\hat{\mu}(x_2^*) + \varepsilon_2} \end{pmatrix} \cdots \begin{pmatrix} \boldsymbol{x_n^*} \\ \boldsymbol{\hat{\mu}(x_n^*) + \varepsilon_n} \end{pmatrix} \right)$$

$\epsilon : N(0, \sigma)$Gaussian noise

Parametric

$$\beta_b^* = \beta(Z_b^*)$$

Result $\boldsymbol{\beta_1^*, \beta_2^*, \beta_3^* \ldots \beta_B^*}$

8.2 The Bootstrap and Maximum likelihood methods

- 8.2.1 A Smoothing Example
  - Parametric bootstrap
  - Non-parametric bootstrap

- 8.2.3 MLE and The bootstrap

MLE AND BOOTSTRAP

Parametric bootstrap ≈ MLE

(Not always parametric bootstrap = LSE)

For this case

MLE = LSE

$\therefore$ ≈ parametric bootstrap

RELATION

Parametric bootstrap

When number of bootstrap sample is large enough, $B \rightarrow \infty$

$$\boldsymbol{\beta}^* \approx \mathrm{N}(\hat{\boldsymbol{\beta}}, \widehat{\boldsymbol{Var}(\hat{\boldsymbol{\beta}}}))$$

Maximum Likelihood estimation

When maximizing likelihood function (detail later)

$$\boldsymbol{\beta} \sim \mathrm{N}(\hat{\boldsymbol{\beta}}, \widehat{\boldsymbol{Var}(\hat{\boldsymbol{\beta}}}))$$

SUMMARY

The Bootstrap method (parametric and non-parametric)

Computer implementation of MLE (when B is large enough)

Pro

Allows us to compute MLE standard error or other quantities without difficult formulas

Effective method for **Interval estimation (区間推定)**

QUESTIONS?

REFERENCE

[B.Efron Bootstrap Methods - CMU Statistics [English]]

~http://www.stat.cmu.edu/~fienberg/Statistics36-756/Efron1979.pdf

[下平英寿 ２１世紀の統計科学３　数理・計算の統計科学 [Japanese]]

~http://ebsa.ism.ac.jp/ebooks/sites/default/files/ebook/1881/pdf/vol3_ch8.pdf