STAT 824 sp 2023 Lec 11 slides

Bootstrap for regression

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

"Sieve" methods $n \to \infty$

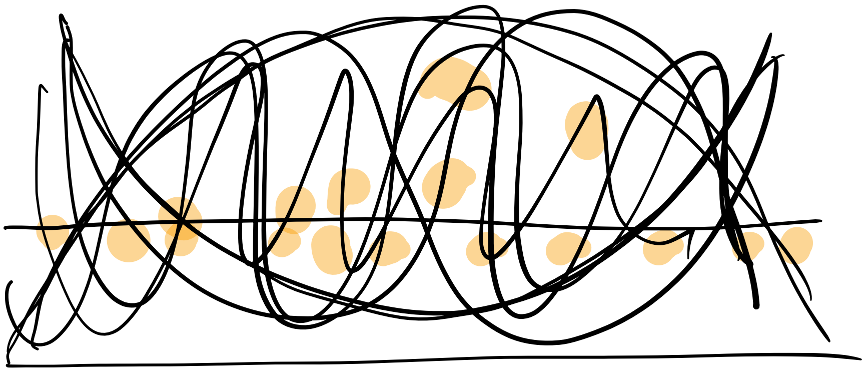
B-splines

$$\underset{n\times N}{B} = \begin{bmatrix} \varphi_1(x_1) \cdots & \varphi_N(x_1) \\ \vdots & \vdots \\ \varphi_1(x_n) \cdots & \varphi_N(x_n) \end{bmatrix}$$

$$\hat{\underset{\sim}{f}}{}_n^N = \begin{bmatrix} \hat f_n^N(x_1) \\ \vdots \\ \hat f_n^N(x_n) \end{bmatrix} = \frac{1}{n} B B^T Y = \frac{1}{n} B \underbrace{\left(\frac{1}{n} B^T B\right)^{-1}}_{I} B^T Y$$

$$\frac{1}{n} B^T B = I \qquad = B \underbrace{(B^T B)^{-1} B^T Y}_{\hat{\underset{\sim}{\alpha}}},$$

$$\hat{\underset{\sim}{\alpha}} = \underset{\underset{\sim}{\alpha}}{\operatorname{argmin}} \left\| Y - B \underset{\sim}{\alpha} \right\|_2^2 = (B^T B)^{-1} B^T \underset{\sim}{Y}$$



Build C.I. for θ_0.

① **pivotal:** $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \sim \underline{\underline{G_n}} \xleftarrow{\text{use bootstrap to estimate the dist.}}$

bootstrap version $\boxed{\sqrt{n}\left(\hat{\theta}_n^{**} - \hat{\theta}_n\right)} \sim \hat{G}_n \leftarrow$ approximate via Monte Carlo.

C.I.: $\left(\hat{\theta}_n - \overset{\text{unknown}}{\overset{\downarrow}{G_{n,\alpha/2}}} \frac{1}{\sqrt{n}},\ \hat{\theta}_n - \overset{\text{unknown}}{\overset{\downarrow}{G_{n,1-\alpha/2}}} \frac{1}{\sqrt{n}}\right)$

bootstrap C.I.: $\left(\hat{\theta}_n - \underset{\sqrt{n}\left(\hat{\theta}_n^{*(\lceil(1-\alpha/2)B\rceil)} - \hat{\theta}_n\right)}{\underset{\uparrow}{\hat{G}_{n,\alpha/2}}} \frac{1}{\sqrt{n}},\ \hat{\theta}_n - \underset{\sqrt{n}\left(\hat{\theta}^{*(\lceil(\alpha/2)B\rceil)} - \hat{\theta}_n\right)}{\underset{\uparrow}{\hat{G}_{n,1-\alpha/2}}} \frac{1}{\sqrt{n}}\right)$

② **percentile:** Draw $\hat{\theta}_n^{*(1)}, \ldots, \hat{\theta}_n^{*(B)}$

$$\boxed{\left(\hat{\theta}_n^{*(\lceil(\alpha/2)B\rceil)},\ \hat{\theta}_n^{*(\lceil(1-\alpha/2)B\rceil)}\right)}$$

$$\hat{p} \pm 1.96\sqrt{\frac{p(1-p)}{500}}$$

Table of Contents

1. Bootstrap in multiple linear regression
   - Constant error term variance
   - Heteroscedastic error term variances
   - Resample pairs
2. Pointwise confidence intervals in nonparametric regression
3. Confidence bands in nonparametric regression
   - Asymptotic distribution approach
   - Bootstrap confidence bands

Linear regression model

Let $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_n, Y_n)$ be data pairs such that

fixed

$$\underset{n\times 1}{\underset{\sim}{Y}} = \underset{n\times p}{\overset{\frown}{X}} \underset{p\times 1}{\underset{\sim}{\beta}_0} + \underset{n\times 1}{\underset{\sim}{\varepsilon}}$$

$$Y_i = \mathbf{x}_i^T \beta_0 + \varepsilon_i, \quad i = 1, \ldots, n,$$

$$\mathrm{Cov}(\underset{\sim}{Y}) = \mathrm{Cov}(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$$

with $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ deterministic, $\varepsilon_1, \ldots, \varepsilon_n$ iid with $\mathbb{E}\varepsilon_1 = 0$, $\mathbb{E}\varepsilon_1^2 = \sigma^2 < \infty$.

Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T$ and

$$\underset{\sim}{c}^T \underset{\sim}{\beta}_0 = [0\,0 \ldots \underset{\overset{\uparrow}{j}}{1}\;0\,0]\, \underset{\sim}{\beta}_0 = \beta_{0j}$$

$$\hat{\beta}_n = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

$$\underset{\sim}{c}^T \underset{\sim}{\beta}_0 = [-1\;1\;0 \ldots 0] = \beta_{02} - \beta_{01}$$

$$\hat{\sigma}_n^2 = (n-p)^{-1}\|\mathbf{Y} - \mathbf{X}\hat{\beta}_n\|_2^2$$

Exercise: Consider estimating contrasts $\mathbf{c}^T\beta_0$ for $\mathbf{c} \in \mathbb{R}^p$ with $\mathbf{c}^T\hat{\beta}_n$.

1. Come up with pivot quantities relevant for making inferences.
2. Give the form of a confidence interval for $\mathbf{c}^T\beta_0$.

$$\hat{\underset{\sim}{\beta}}_n = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \underset{\sim}{Y} .$$

$$\begin{aligned}
\mathrm{Cov}(\hat{\underset{\sim}{\beta}}_n) &= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \,\mathrm{Cov}(\underset{\sim}{Y})\, \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1} \\
&= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \;\sigma^2 \cdot I_n \,\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1} \\
&= \sigma^2(\mathbb{X}^T\mathbb{X})^{-1}
\end{aligned}$$

Suppose $\underset{\sim}{\varepsilon} \sim N(0, \sigma^2 I_n)$. Then

$$\begin{aligned}
\hat{\underset{\sim}{\beta}}_n &= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \underset{\sim}{Y} \\
&= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T(\mathbb{X}\underset{\sim}{\beta}_0 + \underset{\sim}{\varepsilon}) \\
&= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{X}\underset{\sim}{\beta}_0 + (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\underset{\sim}{\varepsilon} \\
&= \underset{\sim}{\beta}_0 + \underbrace{(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\underset{\sim}{\varepsilon}}_{\sim N(\underset{\sim}{0},\ \sigma^2(\mathbb{X}^T\mathbb{X})^{-1})}
\end{aligned}$$

$$\hat{\underset{\sim}{\beta}}_n \sim N(\underset{\sim}{\beta}_0, \sigma^2(\mathbb{X}^T\mathbb{X})^{-1}) ,$$

$$\underset{\sim}{c}^T\hat{\underset{\sim}{\beta}}_n \sim N(\underset{\sim}{c}^T\underset{\sim}{\beta}_0,\ \sigma^2 \underbrace{\underset{\sim}{c}^T(\mathbb{X}^T\mathbb{X})^{-1}\underset{\sim}{c}}_{\text{scalar}})$$

pivotal quantity for $\underset{\sim}{c}^T\underset{\sim}{\beta}_0$

$$\frac{(\underset{\sim}{c}^T\hat{\beta}_n - \underset{\sim}{c}^T\underset{\sim}{\beta}_0)}{\sigma\sqrt{\underset{\sim}{c}^T(\mathbb{X}^T\mathbb{X})^{-1}\underset{\sim}{c}}} \sim N(0,1) .$$

Can show :

$$\frac{\left(\underset{\sim}{c}^T \hat{\beta}_n - \underset{\sim}{c}^T \underset{\sim}{\beta}_0\right)}{\hat{\sigma}_n \sqrt{\underset{\sim}{c}^T (X^T X)^{-1} \underset{\sim}{c}}} \sim t_{n-p}$$

$$\hat{\sigma}_n^2 = \frac{1}{n-p} \| Y - X \underset{\sim}{\hat{\beta}}_n \|_2^2$$

$\Rightarrow$ $(1-\alpha)^* 100\%$ C.I. for $\underset{\sim}{c}^T \underset{\sim}{\beta}_0$ is

$$\underset{\sim}{c}^T \underset{\sim}{\hat{\beta}}_n \pm t_{n-p, \alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}} \sqrt{\underset{\sim}{c}^T \underbrace{\left(\frac{1}{n} X^T X\right)^{-1}}_{\hat{\Omega}} \underset{\sim}{c}}$$

<u>If dist. of $\underset{\sim}{\varepsilon}$ is unknown :</u>

$$T_{n,S} = \frac{\sqrt{n}\left(\underset{\sim}{c}^T \hat{\beta}_n - \underset{\sim}{c}^T \underset{\sim}{\beta}_0\right)}{\hat{\sigma}_n \sqrt{\underset{\sim}{c}^T \left(\frac{1}{n} X^T X\right)^{-1} \underset{\sim}{c}}} \xrightarrow{D} N(0,1) \quad \text{under conditions}$$

$$T_{n,U} = \sqrt{n}\left(\underset{\sim}{c}^T \hat{\beta}_n - \underset{\sim}{c}^T \underset{\sim}{\beta}_0\right) \xrightarrow{D} N(0, \vartheta)$$

$$\mathbb{E}\; \underset{\sim}{c}^T \underset{\sim}{\hat{\beta}}_n = \underset{\sim}{c}^T\, \mathbb{E}\, \underset{\sim}{\hat{\beta}}_n = c^T \underset{\sim}{\beta}_0$$

$$\mathrm{Var}\left(\underset{\sim}{c}^T \underset{\sim}{\hat{\beta}}_n \right) = \underset{\sim}{c}^T \, \mathrm{Cov}\left(\underset{\sim}{\hat{\beta}}_n \right) \underset{\sim}{c} \;,$$

$$= \underset{\sim}{c}^T \left[\sigma^2 \left(\mathbb{X}^T \mathbb{X} \right)^{-1} \right] \underset{\sim}{c}$$

$$= \sigma^2 \; \underset{\sim}{c}^T \left(\mathbb{X}^T \mathbb{X} \right)^{-1} \underset{\sim}{c}$$

$$Y = X_1 \beta_1 + \cdots + X_p \beta_p + \varepsilon \qquad \mathbb{E}\,\varepsilon = 0 \quad \mathrm{Var}\,\varepsilon = \sigma^2.$$

$$\varepsilon \sim \mathrm{Dist}\,?$$

$$\underset{\sim}{\hat{\beta}}_n = \left(\mathbb{X}^T \mathbb{X} \right)^{-1} \mathbb{X}^T \underset{\sim}{Y} \quad \left(\begin{array}{l} \sim N\left(\underset{\sim}{\beta},\; \sigma^2 \left(\mathbb{X}^T \mathbb{X} \right)^{-1} \right) \\ \text{if} \quad \underset{\sim}{\varepsilon} \sim N\left(0, I_n \sigma^2\right) \end{array} \right) \uparrow$$

asymptotically pivotal

$$\boxed{\frac{\underset{\sim}{c}^T \underset{\sim}{\hat{\beta}}_n - \underset{\sim}{c}^T \underset{\sim}{\beta}}{\hat{\sigma}_n \sqrt{\underset{\sim}{c}^T \left(\mathbb{X}^T \mathbb{X} \right)^{-1} \underset{\sim}{c}}}} \xrightarrow{D} N(0,1) \quad \text{as} \quad n \to \infty.$$

$\Rightarrow$

A <u>large-n</u> $(1-\alpha)\cdot 100\%$ C.I. for $\underset{\sim}{c}^T \underset{\sim}{\beta}$ is

$$\underset{\sim}{c}^T \underset{\sim}{\hat{\beta}}_n \pm z_{\alpha/2} \; \hat{\sigma}_n \sqrt{\underset{\sim}{c}^T \left(\mathbb{X}^T \mathbb{X} \right)^{-1} \underset{\sim}{c}}$$

$$\frac{\underset{\sim}{c}^T \underset{\sim}{\hat\beta}_n - \underset{\sim}{c}^T \underset{\sim}{\beta}}{\hat\sigma_n \sqrt{\underset{\sim}{c}^T (X^T X)^{-1} \underset{\sim}{c}}} \sim G_{n,S}$$ ← unknown, but we estimate with bootstrap.

If we knew $G_{n,S}$, then $(1-\alpha)$ C.I. for $\underset{\sim}{c}^T \underset{\sim}{\beta}$ is

$$\left(\underset{\sim}{c}^T \underset{\sim}{\hat\beta} - G_{n,S,\alpha/2}\, \hat\sigma_n \sqrt{\underset{\sim}{c}^T (X^T X)^{-1} \underset{\sim}{c}} \,,\; \underset{\sim}{c}^T \underset{\sim}{\hat\beta} - G_{n,S,1-\alpha/2}\, \hat\sigma_n \sqrt{\underset{\sim}{c}^T (X^T X)^{-1} \underset{\sim}{c}} \right)$$

$$\frac{\underset{\sim}{c}^T \underset{\sim}{\hat\beta}_n^{*} - \underset{\sim}{c}^T \underset{\sim}{\hat\beta}_n}{\hat\sigma_n^{*} \sqrt{\underset{\sim}{c}^T (X^T X)^{-1} \underset{\sim}{c}}} \sim \hat G_{n,S}$$ ← Bootstrap estimate of $G_{n,S}$.

Define

$$\sigma^2_{c,n} = [\mathbf{c}^T(n^{-1}\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}]^{1/2} \cdot \sigma^2$$

$$\hat{\sigma}^2_{c,n} = [\mathbf{c}^T(n^{-1}\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}]^{1/2} \cdot \hat{\sigma}^2_n$$

Theorem (Asymptotically Normal pivots for least-squares coefficients)

For $\mathbf{c} \in \mathbb{R}^p$*, the pivot quantities*

$\sqrt{n}\left(\underset{\sim}{c}^T \hat{\beta}_n - \underset{\sim}{c}^T \beta_0\right)$

$$T_{n,U} = \sqrt{n} \cdot \mathbf{c}^T(\hat{\beta}_n - \beta_0) \quad \text{and} \quad \boxed{T_{n,S} = \sqrt{n} \cdot \mathbf{c}^T(\hat{\beta}_n - \beta_0)/\boxed{\hat{\sigma}_{c,n}}}$$

are asymptotically $\text{Normal}(0, \sigma^2_{c,\infty})$ *and* $\text{Normal}(0,1)$ *rvs, respectively, provided*

$$\max_{1 \le i \le n} h_{ii} \to 0 \quad \text{and} \quad \sigma^2_{c,n} \to \sigma^2_{c,\infty} \in (0,\infty) \quad \text{as } n \to \infty,$$

where h_{ii}, $i = 1, \dots, n$ *is the* i*th diagonal entry of* $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

Denote by $G_{n,U}$ and $G_{n,S}$ the dists of $T_{n,U}$ and $T_{n,S}$, respectively.

Bootstrap for regression

Design matrix X random ↙

Paired bootstrap

Resample with replacement from

$(X_1, Y_1), \dots, (X_n, Y_n)$

→ $(X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*)$.

OR

Wild bootstrap

(Design-matrix X is "fixed" or "deterministic")

Residual bootstrap

(i) Fit model — get $\hat{\beta}_n$.

(ii) Get residuals

$\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$

(iii) Sample with replacement from

$\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$

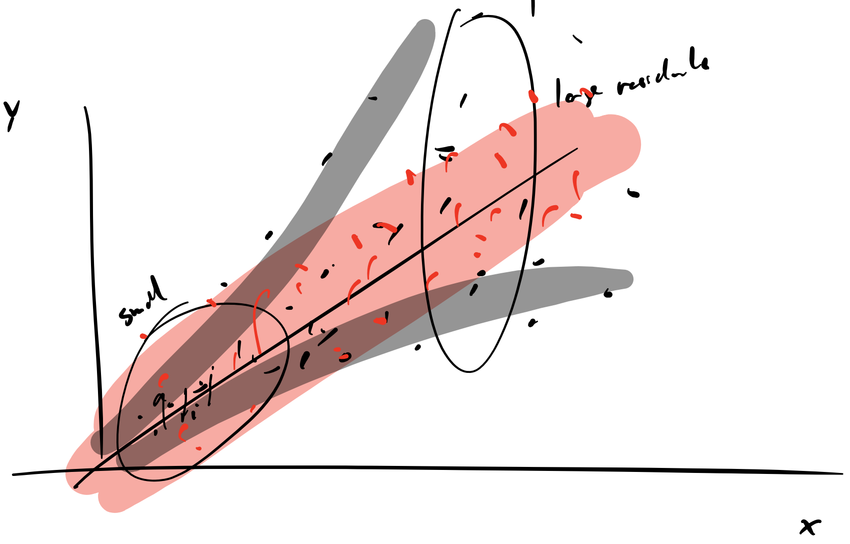
→ get $\varepsilon_1^*, \dots, \varepsilon_n^*$

(iv) Build bootstrap responses

$$Y_i^* = x_i \hat{\beta}_n + \varepsilon_i^*$$

gives

$(x_1, Y_1^*), \dots, (x_n, Y_n^*)$



Residual bootstrap for linear regression

1. Draw $\varepsilon_1^*,\dots,\varepsilon_n^*$ with repl. from $\hat{\varepsilon}_i = Y_i - \mathbf{x}_i^T\hat{\beta}_n$, $i = 1,\dots,n$
2. Set $\boxed{Y_i^* = \mathbf{x}_i^T\hat{\beta}_n + \varepsilon_i^*}$ for $i = 1,\dots,n$.
3. Compute $\hat{\beta}_n^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}^*$ and $(\hat{\sigma}_n^*)^2 = (n-p)^{-1}\|\mathbf{Y}^* - \mathbf{X}\hat{\beta}_n^*\|_2^2$.
4. Compute the bootstrap versions of the pivots $T_{n,U}$ and $T_{n,S}$ given by

$$T_{n,U}^* = \sqrt{n}\cdot\mathbf{c}^T(\hat{\beta}_n^* - \hat{\beta}_n) \quad \text{and} \quad \boxed{T_{n,S}^* = \sqrt{n}\cdot\mathbf{c}^T(\hat{\beta}_n^* - \hat{\beta}_n)/\hat{\sigma}_{c,n}^*,}$$

where $(\hat{\sigma}_{c,n}^*)^2 = [\mathbf{c}^T(n^{-1}\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}]^{1/2}\cdot(\hat{\sigma}_n^*)^2$.

Denote by $\hat{G}_{n,U}$ and $\hat{G}_{n,S}$ the dists of $T_{n,U}^*$ and $T_{n,S}^*$ conditional on the data.

Exercise: Simulate performance of residual bootstrap CIs for (i) $\mathbf{c}^T\beta_0$ and (ii) β_{0j}.

$$T_{n,s} = \frac{\sqrt{n}\left(\underset{\sim}{c}^T \hat{\underset{\sim}{\beta}}_n - \underset{\sim}{c}^T \underset{\sim}{\beta}_0\right)}{\hat{\sigma}_n \sqrt{\underset{\sim}{c}^T \left(\frac{1}{n} \underset{\sim}{X}^T \underset{\sim}{X}\right)^{-1} \underset{\sim}{c}}} \sim \overset{\text{unknown}}{G_{n,S}} \quad \left(\xrightarrow{D} N(0,1)\right)$$

$$P\left(G_{n,S,1-\alpha/2} < \frac{\sqrt{n}\left(\underset{\sim}{c}^T \hat{\underset{\sim}{\beta}}_n - \underset{\sim}{c}^T \underset{\sim}{\beta}_0\right)}{\hat{\sigma}_n \sqrt{\underset{\sim}{c}^T \left(\frac{1}{n} \underset{\sim}{X}^T \underset{\sim}{X}\right)^{-1} \underset{\sim}{c}}} < G_{n,S,\alpha/2} \right) = 1-\alpha$$

$$= P\left(\bigcirc < \underset{\sim}{c}^T \underset{\sim}{\beta} < \bigcirc \right) = 1-\alpha$$

Then $\left(\underset{\sim}{c}^T \hat{\underset{\sim}{\beta}}_n - G_{n,S,\alpha/2} \frac{\hat{\sigma}_n}{n} \sqrt{\underset{\sim}{c}^T \left(\frac{1}{n} \underset{\sim}{X}^T \underset{\sim}{X}\right)^{-1} \underset{\sim}{c}},\ \underset{\sim}{c}^T \hat{\underset{\sim}{\beta}}_n - G_{n,S,1-\alpha/2} \frac{\hat{\sigma}_n}{n} \sqrt{\underset{\sim}{c}^T \left(\frac{1}{n} \underset{\sim}{X}^T \underset{\sim}{X}\right)^{-1} \underset{\sim}{c}} \right)$

is a $(1-\alpha)100\%$ C.I. for $\underset{\sim}{c}^T \underset{\sim}{\beta}_0$.

Bootstrap version

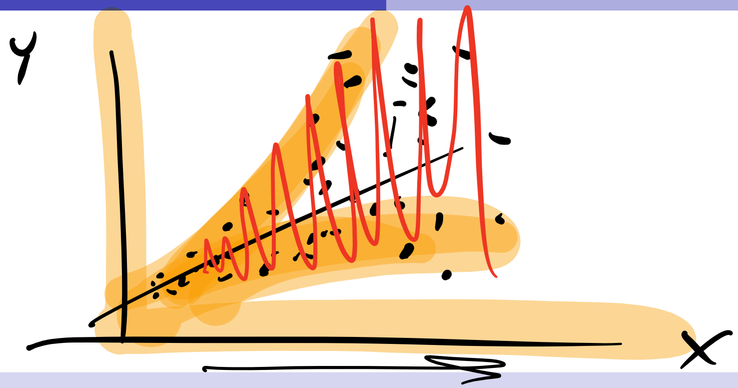
$$\left(\underset{\sim}{c}^T \hat{\underset{\sim}{\beta}}_n - \underset{\underset{T^{*(1)}_{n,s}, \ldots, T^{*(B)}_{n,s}}{\uparrow}}{\overset{T_{n,s}^{*(\lceil(1-\alpha/2)B\rceil)}}{\hat{G}_{n,S,\alpha/2}}} \frac{\hat{\sigma}_n}{n} \sqrt{\underset{\sim}{c}^T \left(\frac{1}{n} \underset{\sim}{X}^T \underset{\sim}{X}\right)^{-1} \underset{\sim}{c}},\ \underset{\sim}{c}^T \hat{\underset{\sim}{\beta}}_n - \overset{\overset{T_{n,s}^{*(\lceil(\alpha/2)B\rceil)}}{\downarrow}}{\hat{G}_{n,S,1-\alpha/2}} \frac{\hat{\sigma}_n}{n} \sqrt{\underset{\sim}{c}^T \left(\frac{1}{n} \underset{\sim}{X}^T \underset{\sim}{X}\right)^{-1} \underset{\sim}{c}} \right)$$

$$\Sigma = \left(\left(\tfrac{1}{2}\right)^{|i-j|} \right)_{1 \le i,j \le}$$

1. Bootstrap in multiple linear regression
    - Constant error term variance
    - Heteroscedastic error term variances
    - Resample pairs

2. Pointwise confidence intervals in nonparametric regression

3. Confidence bands in nonparametric regression
    - Asymptotic distribution approach
    - Bootstrap confidence bands



Linear regression model with heteroscedasticity

Let $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_n, Y_n)$ be data pairs such that

$$\text{Cov}(\underset{\sim}{\varepsilon}) = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}$$

$$Y_i = \mathbf{x}_i^T \beta_0 + \varepsilon_i, \quad i = 1, \ldots, n,$$

with $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ deterministic, $\mathbb{E}\varepsilon_i = 0$ and $\mathbb{E}\varepsilon_i^2 = \sigma_i^2 \in (0, \infty)$, $i = 1, \ldots, n$.

Exercise: Give an expression for $\text{Var}(\cancel{\sqrt{n}} \cdot \mathbf{c}^T \hat{\beta}_n)$ for $\mathbf{c} \in \mathbb{R}^p$.

$$\text{Var}(\underset{\sim}{c}^T \hat{\underset{\sim}{\beta}}_n) = \underset{\sim}{c}^T \text{Cov}(\hat{\underset{\sim}{\beta}}_n) \underset{\sim}{c}$$

$$= \underset{\sim}{c}^T (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix} \mathbb{X} (\mathbb{X}^T \mathbb{X})^{-1} \underset{\sim}{c}$$

$$\underset{\sim}{c}^T \hat{\underset{\sim}{\beta}}_n - \underset{\sim}{c}^T \underset{\sim}{\beta} \xrightarrow{D} ?$$

$$\hat{\underset{\sim}{\beta}}_n = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \underset{\sim}{Y}$$

$$\text{Cov}(\hat{\underset{\sim}{\beta}}_n) = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \overset{\overset{\text{Cov}(\underset{\sim}{\varepsilon})}{\shortparallel}}{\text{Cov}(\underset{\sim}{Y})}\, \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}$$

$$= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \begin{pmatrix}\sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2\end{pmatrix} \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}$$

$$\frac{\underset{\sim}{c}^T\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{c}^T\underset{\sim}{\beta}_n}{\sqrt{\text{Var}(\underset{\sim}{c}^T\hat{\underset{\sim}{\beta}}_n)}} \xrightarrow{?\,D} N(0,1).$$

$$=$$

Under some conditions this true.

$$\downarrow$$

$$\frac{\underset{\sim}{c}^T\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{c}^T\underset{\sim}{\beta}}{\sqrt{\underbrace{\underset{\sim}{c}^T(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \begin{pmatrix}\sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2\end{pmatrix} \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\underset{\sim}{c}}_{v_{n,\underset{\sim}{c}}}}} \xrightarrow{D} N(0,1).$$

$$\boxed{\frac{\sqrt{n}\,(\underset{\sim}{c}^T\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{c}^T\underset{\sim}{\beta})}{\sqrt{n\cdot \underbrace{\underset{\sim}{c}^T(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \begin{pmatrix}\hat{\varepsilon}_1^2 & & \\ & \ddots & \\ & & \hat{\varepsilon}_n^2\end{pmatrix} \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\underset{\sim}{c}}_{\hat{v}_{n,\underset{\sim}{c}}}}} \xrightarrow{D} N\underline{(0,1)}}$$

Define
$$\vartheta^2_{c,n} = n \cdot \mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \cdot \operatorname{diag}(\sigma_1^2, \dots, \sigma_n^2) \cdot \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}$$
$$\hat{\vartheta}^2_{c,n} = n \cdot \mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \cdot \operatorname{diag}(\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_n^2) \cdot \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}$$

Theorem (Asymptotically Normal pivots for LS coefs under hetsc.)

For $\mathbf{c} \in \mathbb{R}^p$, *the pivot quantities*

$$T^{\text{het}}_{n,U} = \sqrt{n} \cdot \mathbf{c}^T(\hat{\beta}_n - \beta_0) \quad \textit{and} \quad \boxed{T^{\text{het}}_{n,S} = \sqrt{n} \cdot \mathbf{c}^T(\hat{\beta}_n - \beta_0)/\hat{\vartheta}_{c,n}}$$

are asymptotically $\text{Normal}(0, \vartheta^2_{c,\infty})$ *and* $\boxed{\text{Normal}(0,1)}$ *rvs, respectively, provided*

$$\max_{1 \le i \le n} h^{\sigma}_{ii}/\sigma_i^2 \to 0 \quad \textit{and} \quad \vartheta^2_{c,n} \to \vartheta^2_{c,\infty} \in (0, \infty)$$

as $n \to \infty$, *where* h^{σ}_{ii} *is* i*th diag. entry of* $\mathbf{X}(\mathbf{X}^T \cdot \operatorname{diag}(\sigma_1^2, \dots, \sigma_2^2) \cdot \mathbf{X})^{-1}\mathbf{X}^T$.

(some additional moment assumps. necessary for $\hat{\vartheta}_{c,n}/\vartheta_{c,n} \to^p 1$, *[5])*

Denote by $G^{\text{het}}_{n,U}$ and $G^{\text{het}}_{n,S}$ the distributions of $T^{\text{het}}_{n,U}$ and $T^{\text{het}}_{n,S}$, respectively.

So for large n: $(1-\alpha)$ C.I. for $\underset{\sim}{c}^T\underset{\sim}{\beta}$ is

$$\boxed{\underset{\sim}{c}^T\hat{\underset{\sim}{\beta}}_n \pm z_{\alpha/2}\sqrt{\underset{\sim}{c}^T(X^TX)^{-1}X^T\begin{pmatrix}\hat\varepsilon_1^2 & & \\ & \ddots & \\ & & \hat\varepsilon_n^2\end{pmatrix}X(X^TX)^{-1}\underset{\sim}{c}}}$$

For each $n \geq 1$,

$$\frac{\sqrt{n}\left(\underset{\sim}{c}^T\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{c}^T\underset{\sim}{\beta}\right)}{\sqrt{\underbrace{n\cdot\underset{\sim}{c}^T(X^TX)^{-1}X^T\begin{pmatrix}\hat\varepsilon_1^2 & & \\ & \ddots & \\ & & \hat\varepsilon_n^2\end{pmatrix}X(X^TX)^{-1}\underset{\sim}{c}}_{\hat\vartheta_{c,n}}}} \sim G^{het}_{n,s}$$

Bootstrap version:

$$\frac{\sqrt{n}\left(\underset{\sim}{c}^T\hat{\underset{\sim}{\beta}}_n^{*W} - \underset{\sim}{c}^T\hat{\underset{\sim}{\beta}}\right)}{\sqrt{n\cdot\underset{\sim}{c}^T(X^TX)^{-1}X^T\begin{pmatrix}\hat\varepsilon_1^{*W2} & & \\ & \ddots & \\ & & \hat\varepsilon_n^{*W2}\end{pmatrix}X(X^TX)^{-1}\underset{\sim}{c}}} \sim \hat G^{het}_{n,s}$$

$(1-\alpha)$ C.I. for $\underset{\sim}{c}^T\underset{\sim}{\beta}$ is

$$\left(\underset{\sim}{c}^T\hat{\underset{\sim}{\beta}} - \hat G_{n,s,\alpha/2}\,\hat\vartheta_{c,n}\,,\ \underset{\sim}{c}^T\hat{\underset{\sim}{\beta}} - \hat G_{n,s,1-\alpha/2}\,\hat\vartheta_{c,n}\right)$$

get ordinary residuals

Wild bootstrap for linear regression

1. Generate indep. bootstrap residuals $\varepsilon_1^{*W}, \dots, \varepsilon_n^{*W}$ satisfying $\mathbb{E}_*[\varepsilon_i^{*W}] = 0$, $\mathbb{E}_*[(\varepsilon_i^{*W})^2] = \hat{\varepsilon}_i^2$, and $\mathbb{E}_*[(\varepsilon_i^{*W})^3] = \hat{\varepsilon}_i^3$, where $\boxed{\hat{\varepsilon}_i = Y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_n}$, $i = 1, \dots, n$.
2. Set $Y_i^{*W} = \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_n + \varepsilon_i^{*W}$, $i = 1, \dots, n$.
3. Compute $\hat{\boldsymbol{\beta}}_n^{*W} = \boxed{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T}\mathbf{Y}^{*W}$, $\hat{\varepsilon}_i^{*W} = Y_i^{*W} - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_n^{*W}$, $i = 1, \dots, n$.
4. Compute the bootstrap versions of the pivots $T_{n,U}$ and $T_{n,S}$ given by

$$T_{n,U}^{\text{het}\,*} = \sqrt{n}\cdot\mathbf{c}^T(\hat{\boldsymbol{\beta}}_n^{*W} - \hat{\boldsymbol{\beta}}_n) \quad \text{and} \quad T_{n,S}^{\text{het}\,*} = \sqrt{n}\cdot\mathbf{c}^T(\hat{\boldsymbol{\beta}}_n^{*W} - \hat{\boldsymbol{\beta}}_n)/\hat{\vartheta}_{c,n}^{*W},$$

where $(\hat{\vartheta}_{c,n}^{*W})^2 = n\cdot\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\cdot\text{diag}((\hat{\varepsilon}_1^{*W})^2, \dots, (\hat{\varepsilon}_n^{*W})^2)\cdot\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}$.

Denote by $\hat{G}_{n,U}^{\text{het}}$ and $\hat{G}_{n,S}^{\text{het}}$ the dists of $T_{n,U}^{\text{het}\,*}$ and $T_{n,S}^{\text{het}\,*}$ conditional on the data.

Two ways to obtain wild bootstrap residuals [1], [2]

1. Mammen (1993): For $i = 1, \ldots, n$, get $V_{i,1}, V_{i,2} \overset{\text{ind}}{\sim} \text{Normal}(0,1)$. Then set

$$U_i = (\delta_1 + V_{i,1}/\sqrt{2})(\delta_2 + V_{i,2}/\sqrt{2}) - \delta_1\delta_2,$$

where $\delta_1 = (3/4 + \sqrt{17}/12)^{1/2}$, $\delta_2 = (3/4 - \sqrt{17}/12)^{1/2}$. Then let

$$\varepsilon_i^{*W} = \hat{\varepsilon}_i \cdot U_i.$$

2. Das et al. (2019): For $i = 1, \ldots, n$, generate $U_i \sim \text{Beta}(\underline{1/2}, \underline{3/2})$. Then set

$$\varepsilon_i^{*W} = \hat{\varepsilon}_i \cdot 4\boxed{(U_i - 1/4)}.$$

$$\frac{\alpha}{\alpha+\beta} = \frac{1/2}{1/2 + 3/2} = \frac{1}{4}$$

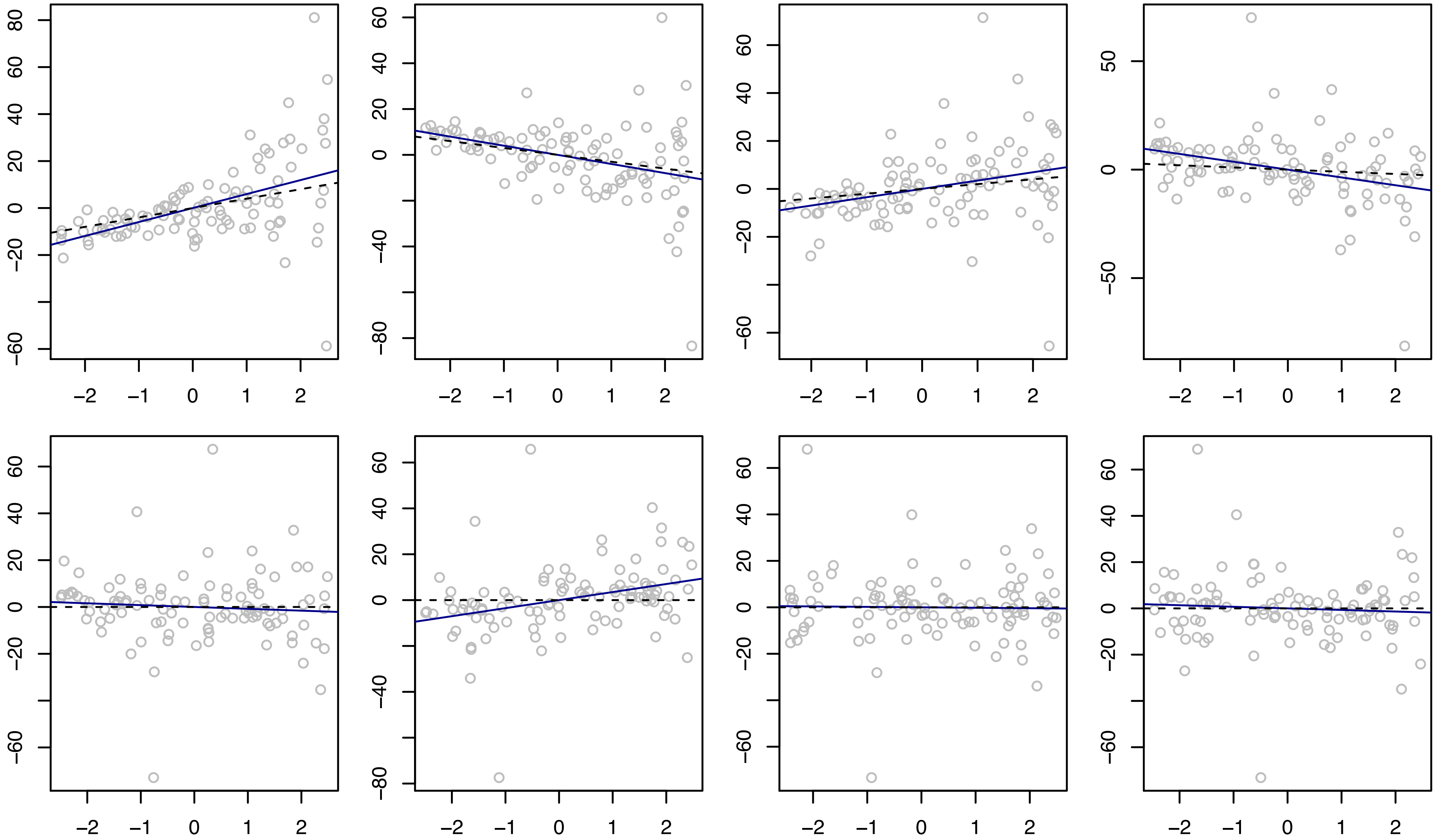
Exercise: Simulate performance of residual bootstrap CIs for (i) $\mathbf{c}^T\beta_0$ and (ii) β_{0j}.

Discuss: Comparison of methods in simulation.

```
r <- .7
R <- r^abs( outer(1:8,1:8,"-"))
P <- 2*sin( R * pi / 6)
X <- cbind(1,(pnorm( matrix(rnorm(n*8),ncol = 8) %*% chol(P)) - .5) * 5)
beta <- c(-1,c(4:1)*(-1)^(4:1),0,0,0,0)
sigma <- 1/4 + abs(X[,2] + 2.5)^2
error <- rnorm(n,0,sigma)
Y <- as.numeric(X %*% beta) + error
```

Coverage of 95% confidence intervals for β_{03} at sample sizes $n = 10, 20, \dots, 100$.

	n									
method	10	20	30	40	50	60	70	80	90	100
$T^*_{n,U}$	0.34	0.81	0.88	0.91	0.93	0.95	0.93	0.95	0.95	0.95
resid $T^*_{n,S}$	0.98	0.99	0.98	0.97	0.97	0.97	0.97	0.98	0.97	0.96
$T_{n,S}$ as $N(0,1)$	0.66	0.92	0.94	0.96	0.95	0.97	0.96	0.97	0.96	0.95
$T^{\text{het}\,*}_{n,U}$	0.31	0.79	0.87	0.90	0.93	0.95	0.93	0.95	0.93	0.93
$T^{\text{het}\,*}_{n,S}$	0.90	0.91	0.92	0.93	0.93	0.95	0.94	0.94	0.94	0.93
$T^{\text{het}}_{n,S}$ as $N(0,1)$	0.30	0.79	0.86	0.90	0.93	0.94	0.93	0.93	0.93	0.93


80
60
40
20
0
−20
−60
−2
−1
0
1
2
60
40
20
0
−40
−80
−2
−1
0
1
2
60
40
20
0
−20
−60
−2
−1
0
1
2
50
0
−50
−2
−1
0
1
2
60
40
20
0
−20
−60
−2
−1
0
1
2
60
40
20
0
−40
−80
−2
−1
0
1
2
60
40
20
0
−20
−60
−2
−1
0
1
2
60
40
20
0
−20
−60
−2
−1
0
1
2

$$\underset{n\times p}{X} \xrightarrow{\text{sampling rows with repl.}} \underset{n\times p}{X^*} \qquad (X^{*T}X^*)^{-1}$$

If p is large but still $p < n$.

Linear regression model with heteroscedasticity and a random design

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be data pairs such that

$$Y_i = X_i^T \beta_0 + \varepsilon_i, \quad i = 1, \dots, n,$$

with $X_1, \dots, X_n \in \mathbb{R}^p$ rvs, $\mathbb{E}[\varepsilon_i | X_i] = 0$ and $\mathbb{E}[\varepsilon_i^2 | X_i] = \sigma_i^2 \in (0, \infty)$, $i = 1, \dots, n$.

A random design is often more realistic (but does it really matter?).

Mammen (1993) showed that the wild bootstrap works in the above setting.

The resampling pairs bootstrap

1. Draw $(X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*)$ with replacement from $(X_1, Y_1), \dots, (X_n, Y_n)$.
2. Then let $\hat{\beta}_n^* = (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{Y}^*$.

Note that we must compute the inverse $(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}$ for every bootstrap resample!

This is taken from Mammen (1993). “Bootstrap” is the resampling pairs bootstrap.

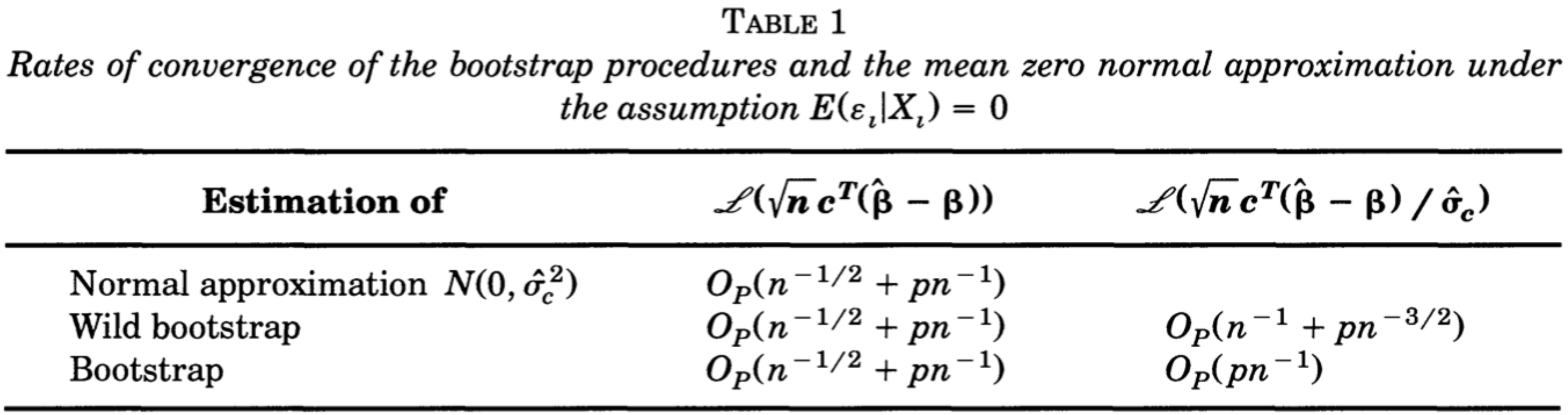


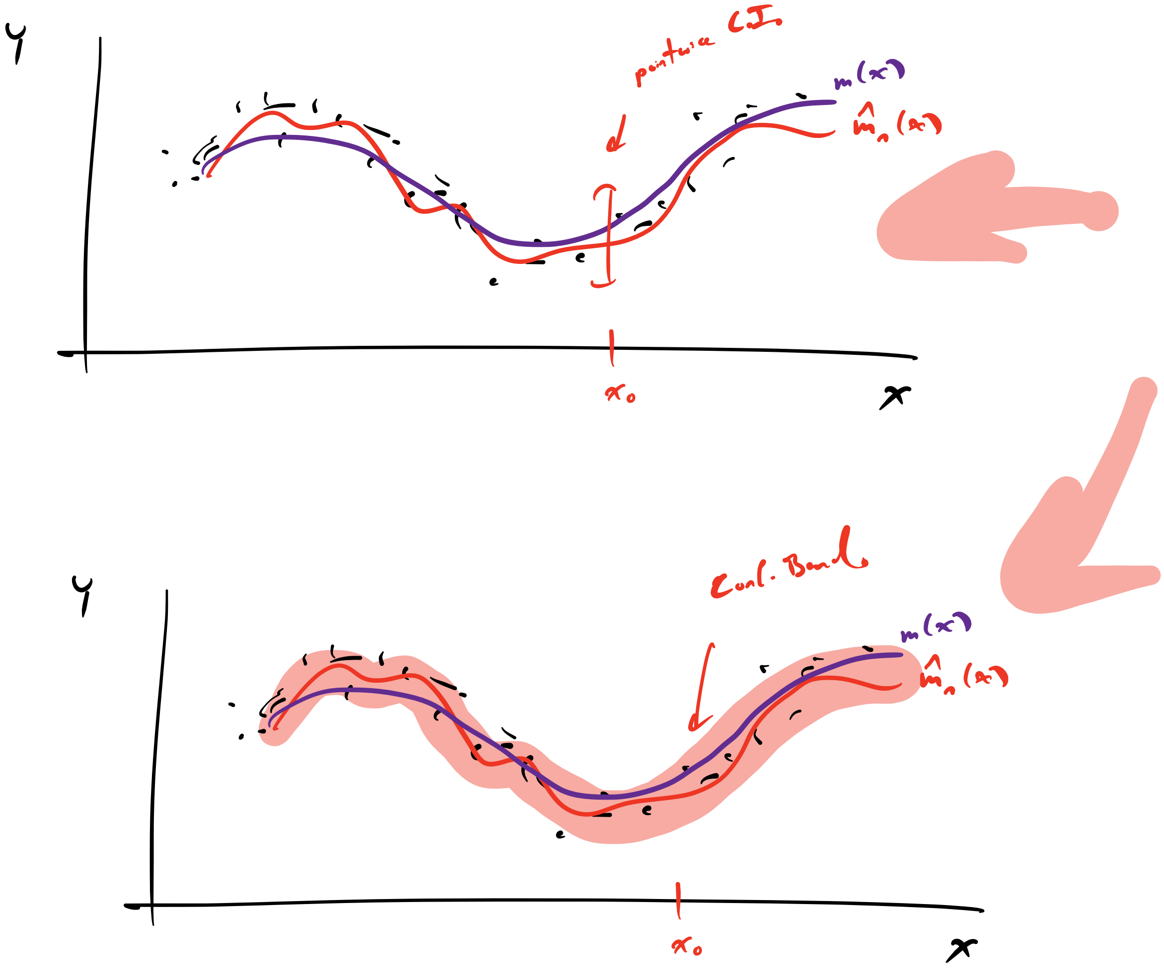
TABLE 1

Rates of convergence of the bootstrap procedures and the mean zero normal approximation under the assumption $E(\varepsilon_i|X_i) = 0$

Estimation of	$\mathcal{L}(\sqrt{n}\,\mathbf{c}^T(\hat{\beta} - \beta))$	$\mathcal{L}(\sqrt{n}\,\mathbf{c}^T(\hat{\beta} - \beta)\,/\,\hat{\sigma}_c)$
Normal approximation $N(0, \hat{\sigma}_c^2)$	$O_P(n^{-1/2} + pn^{-1})$	
Wild bootstrap	$O_P(n^{-1/2} + pn^{-1})$	$O_P(n^{-1} + pn^{-3/2})$
Bootstrap	$O_P(n^{-1/2} + pn^{-1})$	$O_P(pn^{-1})$

In this paper the affect of the dimension p is tracked along with that of n.

The studentized resampling pairs bootstrap is more adversely affected by high dimension than the wild bootstrap.


Y
pointwise C.I.
m(x)
$\hat{m}_n(x)$
x_0
x
Y
Conf. Band.
m(x)
$\hat{m}_n(x)$
x_0
x

Nonparametric regression model

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be data pairs such that

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

with $X_1, \dots, X_n \in [0,1]$ deterministic, $\mathbb{E}\varepsilon_i = 0$, $\mathbb{E}\varepsilon_i^2 = \sigma_i^2 \in (0, \infty)$, $i = 1, \dots, n$.

Consider *linear estimators*, i.e. estimators of the form

$$\hat{m}_n(x) = \sum_{i=1}^n W_{ni}(\underset{\uparrow}{x}) Y_i, \qquad \text{for } x \in [0,1].$$

Exercise: Discuss estimators of $\operatorname{Var} \hat{m}_n(x)$ in the cases

1. $\sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2 \in (0, \infty)$.
2. $\sigma_1^2, \dots, \sigma_n^2$ are heteroscedastic.

replace σ_i^2 with $\hat{\varepsilon}_i^2$.

$$\operatorname{Var} \hat{m}_n(x) = \operatorname{Var}\left(\sum_{i=1}^n W_{ni}(x) Y_i \right) = \sum_{i=1}^n W_{ni}^2(x) \operatorname{Var} Y_i = \sum_{i=1}^n W_{ni}(x) \sigma_i^2$$

Suppose we wish to build a confidence interval for $m(x)$ at some $x \in [0,1]$.

Consider our discussions from Lecture 4: We have

$$\frac{\hat{m}_n(x) - m_n(x)}{\sqrt{\operatorname{Var} \hat{m}_n(x)}} = \underbrace{\frac{\hat{m}_n(x) - \mathbb{E}\hat{m}_n(x)}{\sqrt{\operatorname{Var} \hat{m}_n(x)}}}_{\to^D N(0,1)} + \underbrace{\cancel{\frac{\mathbb{E}\hat{m}_n(x) - m_n(x)}{\sqrt{\operatorname{Var} \hat{m}_n(x)}}}}_{\to^p \text{ 0 if } \hat{m}_n \text{ undersmoothed}}.$$

Strategy: Undersmooth and pretend $\mathbb{E}\hat{m}_n(x)$ is equal to $m(x)$.

Studentized pivots under constant and heteroscedastic variances

$$T_{n,x} = \frac{\hat{m}_n(x) - \mathbb{E}\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^n W_{ni}^2(x)}} \quad \text{and} \quad \boxed{T_{n,x}^{\text{het}} = \frac{\hat{m}_n(x) - \mathbb{E}\hat{m}_n(x)}{\sqrt{\sum_{i=1}^n W_{ni}^2(x)\hat{\varepsilon}_i^2}}}$$

const. variance

non constant var.

$\Downarrow$

$$T_{n,x} \overset{\text{approx}}{\sim} N(0,1)$$

$$\lim_{n\to\infty} P\left(-z_{\alpha/2} < \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}} < z_{\alpha/2}\right) = 1-\alpha$$

$$\lim_{n\to\infty} P\left(\hat{m}_n(x) - z_{\alpha/2}\,\hat{\sigma}\sqrt{\sum_{i=1}^{n} W_{ni}^2(x)} < \overset{\approx m(x)}{\mathbb{E}\,\hat{m}_n(x)} < \hat{m}_n(x) + z_{\alpha/2}\,\hat{\sigma}_n\sqrt{\sum W_{ni}^2(x)}\right)$$

$\rightarrow$ C.I. for $m(x)$ is

$$\boxed{\hat{m}_n(x) \pm z_{\alpha/2}\,\hat{\sigma}_n \underbrace{\sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}}_{s.\underset{\sim}{x}}} .$$

$$\hat{\sigma}_n^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \left(Y_{(i+1)} - Y_{(i)}\right)^2$$

Non const. variance : ..

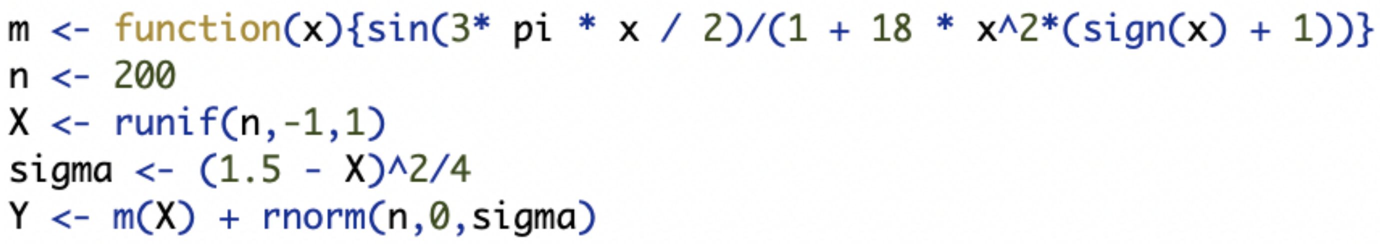
$$\hat{m}_n(x) \pm z_{\alpha/2} \underbrace{\sqrt{\sum_{i=1}^{n} W_{ni}^2(x)\,\hat{\varepsilon}_i^2}}_{s.het.x} .$$

$$\text{NW:} \quad \hat{m}_n(x) \quad \frac{\sum_{i=1}^{n} \overset{W_{ni}(x)}{K\left(\frac{X_i - x}{h}\right)} Y_i}{\sum_{j=1}^{n} K\left(\frac{X_j - x}{h}\right)} = \sum_{i=1}^{n} W_{ni}(x)\, Y_i$$

$$\underset{N\times 1}{\begin{bmatrix} \hat{m}_n(x_1) \\ \vdots \\ \hat{m}_n(x_N) \end{bmatrix}} = \underbrace{\underset{N\times n}{\begin{bmatrix} W_{n1}(x_1) & \cdots & W_{nn}(x_1) \\ & & \\ W_{n1}(x_N) & \cdots & W_{nn}(x_N) \end{bmatrix}}}_{S_{\underset{\sim}{x}}} \underset{n\times 1}{\begin{bmatrix} Y_1 \\ \\ Y_n \end{bmatrix}}$$

$$S = \begin{bmatrix} W_{n1}(X_1) & \cdots & W_{nn}(X_1) \\ & & \\ W_{n1}(X_n) & \cdots & W_{nn}(X_n) \end{bmatrix}$$

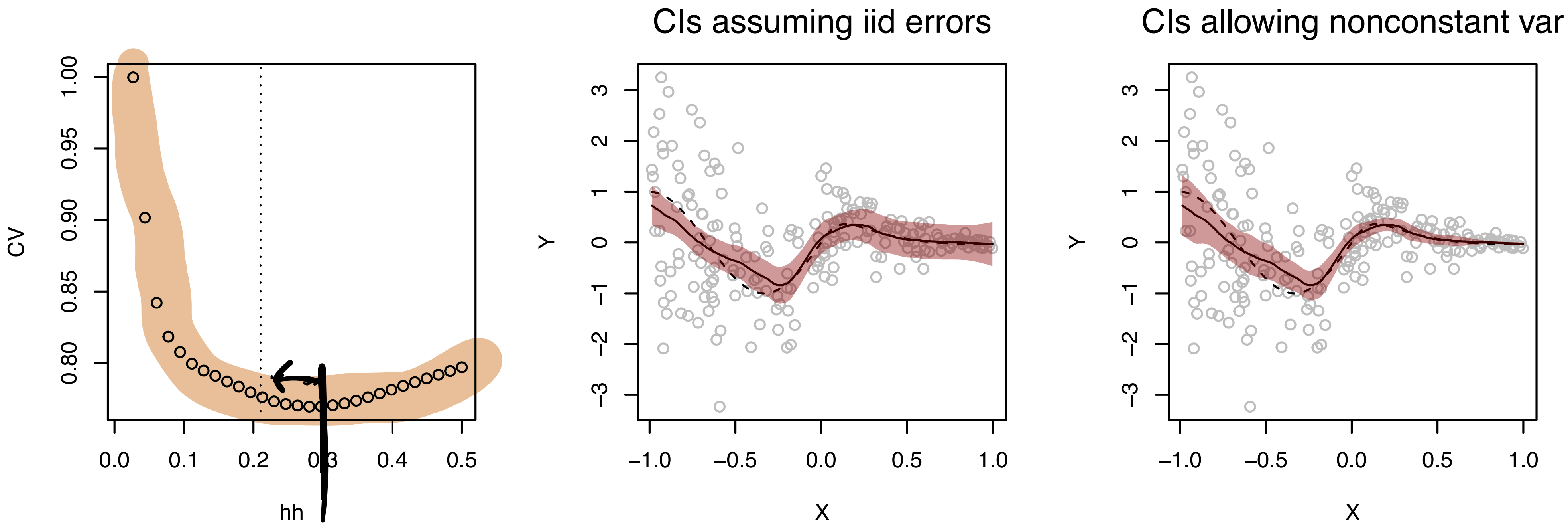
$$\begin{bmatrix} W_{n1}^2(X_1)\,\hat{\varepsilon}_1^2 & & W_{nn}^2(X_1)\,\hat{\varepsilon}_1^2 \\ & & \\ W_{n1}^2(X_n)\,\hat{\varepsilon}_n^2 & \cdots & W_{nn}^2(X_n)\,\hat{\varepsilon}_n^2 \end{bmatrix}$$



```
m <- function(x){sin(3* pi * x / 2)/(1 + 18 * x^2*(sign(x) + 1))}
n <- 200
X <- runif(n,-1,1)
sigma <- (1.5 - X)^2/4
Y <- m(X) + rnorm(n,0,sigma)
```

The asymptotic Normality of the pivots suggests the pointwise CIs

$$\left[\hat{m}_n(x) \pm z_{\alpha/2}\hat{\sigma}_n\sqrt{\textstyle\sum_{i=1}^n W_{ni}^2(x)}\right] \quad \text{and} \quad \left[\hat{m}_n(x) \pm z_{\alpha/2}\sqrt{\textstyle\sum_{i=1}^n W_{ni}^2(x)\hat{\varepsilon}_i^2}\right].$$



1. Bootstrap in multiple linear regression
    - Constant error term variance
    - Heteroscedastic error term variances
    - Resample pairs

2. Pointwise confidence intervals in nonparametric regression

3. Confidence bands in nonparametric regression
    - Asymptotic distribution approach
    - Bootstrap confidence bands

Consider the constant variances case $\sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2 \in (0, \infty)$.

Assume $\hat{m}_n$ is undersmoothed and pretend $\mathbb{E}\hat{m}_n(x)$ equals $m(x)$ (i.e. ignore bias).

A pivot for building confidence bands

Define

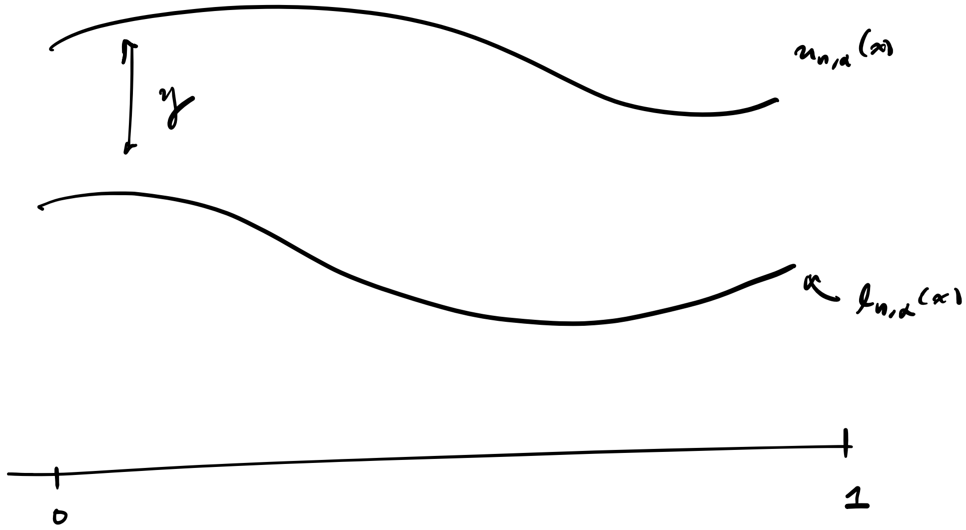
$$T_n = \sup_{x\in[0,1]} \left| \frac{\hat{m}_n(x) - \mathbb{E}\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^n W_{ni}^2(x)}} \right|$$

and denote by G_n the distribution of T_n.

Exercise: Propose a confidence band for $m(x)$, $x \in [0, 1]$ of the form

$$\mathcal{B}_{n,\alpha} = \{(x, y) : \underset{\uparrow}{\underline{l_{n,\alpha}(x)}} \underline{\leq y \leq} \underset{\uparrow}{\underline{u_{n,\alpha}(x)}}, \underline{x \in [0, 1]}\}$$

assuming the distribution G_n is known.



$$P\left(m(x) \in [l_{n,\alpha}(x), u_{n,\alpha}(x)] \quad \forall \quad x \in [0,1] \right) = 1-\alpha$$

Start with

$$P\left(\sup_{x\in[0,1]} \left| \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}} \right| < c_{n,\alpha} \right) = 1-\alpha.$$

$\iff$

$$P\left(\left| \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}} \right| < c_{n,\alpha} \quad \forall \; x\in[0,1] \right) = 1-\alpha$$

$$P\left(-G_{n,\alpha} < \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}} < G_{n,\alpha} \quad \forall\ x \in [0,1]\right) = 1-\alpha$$

$$P\left(\hat{m}_n(x) - G_{n,\alpha}\,\sigma\sqrt{\sum_{i=1}^{n} w_{ni}^2(x)} < \mathbb{E}\,\hat{m}_n(x) < \hat{m}_n(x) + G_{n,\alpha}\,\sigma\sqrt{\sum_{i=1}^{n} w_{ni}^2(x)} \quad \text{for all } x \in [0,1]\right) = 1-\alpha.$$

$$\Longleftrightarrow$$

$$P\left(\underbrace{\mathbb{E}\,\hat{m}_n(x)}_{\uparrow\ \text{pretend } \mathbb{E}\,\hat{m}_n(x) = m(x).} \in \left[\hat{m}_n(x) \pm G_{n,\alpha}\,\sigma\sqrt{\sum_{i=1}^{n} w_{ni}^2(x)}\right] \ \forall\ x \in [0,1]\right) = 1-\alpha$$

But how do we know $\boxed{G_n}$?

Asymptotic result of Sun and Loader (1994) [3], [4].

For $Z_1, \dots, Z_n \overset{\text{ind}}{\sim} \text{Normal}(0,1)$, we have

$$P\left(\sup_{x\in[0,1]}\left|\sum_{i=1}^n M_{ni}(x)Z_i\right| > c\right) \approx 2(1-\Phi(c)) + \frac{\kappa_0}{\pi}e^{-c^2/2}$$

for large c and large enough n, where

$$\kappa_0 = \int_0^1 \sqrt{\sum_{i=1}^n \left[\frac{\partial}{\partial x}M_{ni}(x)\right]^2}dx.$$

Exercise:

1. Find $M_{ni}(x)$, $i = 1, \dots, n$, such that we may write
$$T_n = \sup_{x\in[0,1]}\left|\sum_{i=1}^n M_{ni}(x)(\varepsilon_i/\hat{\sigma}_n)\right|$$
2. Propose a conf. band for $m(x), x \in [0,1]$ based on the above result.

$$T_n = \sup_{x\in[0,1]} \left| \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}} \right|$$

$$= \sup_{x\in[0,1]} \left| \frac{\sum_{i=1}^{n} W_{ni}(x)\ \varepsilon_i}{\hat{\sigma}_n \sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}} \right|$$

$$= \sup_{x\in[0,1]} \left| \sum_{i=1}^{n} M_{ni}(x) \overbrace{(\varepsilon_i / \hat{\sigma}_n)}^{\approx z_i} \right| ,$$

$$M_{ni}(x) = \frac{W_{ni}(x)}{\sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}} , \quad i = 1, \dots, n.$$

$$\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x) = \sum_{i=1}^{n} W_{ni}(x)\, Y_i - \mathbb{E} \sum_{i=1}^{n} W_{ni}(x)\, Y_i$$

$$= \sum_{i=1}^{n} W_{ni}(x) \left[Y_i - \mathbb{E}\, Y_i \right]$$

$$= \sum_{i=1}^{n} W_{ni}(x) \left[Y_i - m(x_i) \right]$$

$$= \sum_{i=1}^{n} W_{ni}(x)\ \varepsilon_i$$

$$) \left(\sup_{x\in[0,1]} \left| \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}} \right| > \overset{\text{like ``c''}}{c_{n,\alpha}} \right) = \begin{array}{c} \alpha \\ \shortparallel \\ 2(1-\Phi(c)) + \frac{K_0}{\pi} e^{-c^2/2} \end{array}$$

Asymp. Conf. band is

$$\left[\hat{m}_n(x) \pm c_\alpha \hat{\sigma}_n \sqrt{\sum_{i=1}^{n} w_{ni}^2(x)}\right] \quad \forall x \in [0,1]$$

where c_α solves $\boxed{2(1-\Phi(c)) + \frac{K_0}{\pi} e^{-c^2/2}} = \alpha$.

Need to get K_0 ...

Approximation to κ_0

To compute κ_0 in practice is hard. We can get an approximation to it as

$$\kappa_0 \approx \sum_{j=1}^{N-1} \sqrt{\sum_{i=1}^{n} [M_{ni}(x_{j+1}) - M_{ni}(x_j)]^2}$$

given a grid of values $x_1, \dots, x_N \in [0,1]$, for some large N.

Exercise:

1. Justify the above approximation to κ_0.
2. Construct the asymptotic $(1-\alpha)100\%$ confidence band

$$\left\{(x,y) : y \in \left[\ \hat{m}_n(x) \pm c_\alpha \hat{\sigma}_n \sqrt{\textstyle\sum_{i=1}^n W_{ni}^2(x)}\ \right],\quad x \in [0,1]\right\}$$

on a simulated data set, where c_α is from the Sun and Loader method.

Consider the case in which $\sigma_1^2, \dots, \sigma_n^2 \in (0, \infty)$ are heteroscedastic.

Assume $\hat{m}_n$ is undersmoothed and pretend $\mathbb{E}\hat{m}_n(x)$ equals $m(x)$ (i.e. ignore bias).

A pivot for building confidence bands under heteroscedasticity

Define

$$T_n^{\text{het}} = \sup_{x \in [0,1]} \left| \frac{\hat{m}_n(x) - \mathbb{E}\hat{m}_n(x)}{\sqrt{\sum_{i=1}^n W_{ni}^2(x)\hat{\varepsilon}_i^2}} \right|$$

and denote by G_n^{het} the distribution of T_n^{het}.

Exercise: Propose a confidence band for $m(x)$, $x \in [0, 1]$ of the form

$$\mathcal{B}_{n,\alpha}^{\text{het}} = \{(x, y) : l_{n,\alpha}^{\text{het}}(x) \leq y \leq u_{n,\alpha}^{\text{het}}(x), x \in [0, 1]\}$$

assuming the distribution G_n^{het} is known.

$$T_n^{\text{boot}} = \sup_{x \in [0,1]} \left| \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\sqrt{\sum_{i=1}^{n} W_{ni}^2(x)\, \hat{\varepsilon}_i^2}} \right| = \sup_{x} \left| \sum_{i=1}^{n} \tilde{M}_{ni}(x)\, z_i \right|$$

$$= \sup_{x \in [0,1]} \left| \frac{\sum_{i=1}^{n} W_{ni}(x)\, \sigma_i \overset{\approx z_i}{\left(\varepsilon_i / \sigma_i\right)}}{\sqrt{\sum_{i=1}^{n} W_{ni}^2(x)\, \hat{\varepsilon}_i^2}} \right|$$

$$= \sup_{x \in [0,1]} \left| \sum_{i=1}^{n} \tilde{M}_{ni}(x)\, z_i \right|$$

unknown

$$\tilde{M}_{ni}(x) = \frac{W_{ni}(x)\, \sigma_i}{\sqrt{\sum_{i=1}^{n} W_{ni}^2(x)\, \hat{\varepsilon}_i^2}} \, .$$

$$\sigma_i^2 = \hat{\varepsilon}_i^2 \, , \quad \sigma_i = |\hat{\varepsilon}_i|$$

Use instead

$$\boxed{\hat{M}_{ni}(x) = \frac{W_{ni}(x)\, |\hat{\varepsilon}_i|}{\sqrt{\sum_{i=1}^{n} W_{ni}^2(x)\, \hat{\varepsilon}_i^2}} \, .}$$

Asymp. Conf. band is

$$\left[\hat{m}_n(x) \pm c_\alpha \sqrt{\sum_{i=1}^{n} w_{ni}^2(x)\, \hat{\varepsilon}_i^2}\right] \quad \forall\, x \in [0,1]$$

where c_α solves $2(1-\Phi(c)) + \frac{K_0}{\pi} e^{-c^2/2} = \alpha$.

Need to get K_0 ... from $\hat{M}_{ni}(x)$ ...

Exercise:

1. To build a conf. band à la Sun and Loader, we need $\bar{M}_{ni}(x)$ such that

$$T_n^{\text{het}} = \sup_{x \in [0,1]} \left| \sum_{i=1}^n \bar{M}_{ni}(x) Z_i \right|$$

for some rvs $Z_1, \dots, Z_n \overset{\text{ind}}{\sim} \text{Normal}(0,1)$. Find $\bar{M}_{ni}(x)$, $i = 1, \dots, n$.

2. Construct the asymptotic $(1-\alpha)100\%$ confidence band

$$\left\{ (x,y) : y \in \left[\hat{m}_n(x) \pm c_\alpha \sqrt{\sum_{i=1}^n W_{ni}^2(x) \hat{\varepsilon}_i^2} \;\right], \; x \in [0,1] \right\}$$

on a simulated data set, where c_α is from the S&L method with κ_0 based on

$$\hat{M}_{ni}(x) = \frac{W_{ni}(x)|\hat{\varepsilon}_i|}{\sqrt{\sum_{i=1}^n W_{ni}^2(x)\hat{\varepsilon}_i^2}}.$$

Consider the constant variances case $\sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2 \in (0, \infty)$.

Residual bootstrap for nonparametric regression

1. Draw $\varepsilon_1^*, \dots, \varepsilon_n^*$ w/repl from $\hat{\varepsilon}_i = Y_i - \hat{m}_n(X_i)$, $i = 1, \dots, n$.
2. Set $Y_i^* = \hat{m}_n(X_i) + \varepsilon_i^*$, $i = 1, \dots, n$.
3. Compute $\hat{m}_n^*(x) = \sum_{i=1}^n W_{ni}(x) Y_i^*$ and $\hat{\sigma}_n^*$ based on $(X_1, Y_1^*), \dots, (X_n, Y_n^*)$.
4. Compute bootstrap version of $T_{n,x}$ given by

$$T_{n,x}^* = \frac{\boxed{\hat{m}_n^*(x) - \mathbb{E}_* \hat{m}_n^*(x)}}{\hat{\sigma}_n^* \sqrt{\sum_{i=1}^n W_{ni}^2(x)}}.$$

$$T_{n,x} = \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^n W_{ni}^2(x)}}$$

Exercise: Show that $\hat{m}_n^*(x) - \mathbb{E}_* \hat{m}_n^*(x) = \boxed{\sum_{i=1}^n W_{ni}(x)\varepsilon_i^*,}$ provided $\sum_{i=1}^n \hat{\varepsilon}_i = 0$.

$$T_{n,x}^* = \frac{\sum_{i=1}^n W_{ni}(x)\,\varepsilon_i^*}{\hat{\sigma}_n^* \sqrt{\sum_{i=1}^n W_{ni}^2(x)}}$$

Consider the case in which $\sigma_1^2, \dots, \sigma_n^2 \in (0, \infty)$ are heteroscedastic.

Wild bootstrap for nonparametric regression

1. Generate indep. bootstrap residuals $\varepsilon_1^{*W}, \dots, \varepsilon_n^{*W}$ satisfying $\mathbb{E}_*[\varepsilon_i^{*W}] = 0$, $\mathbb{E}_*[(\varepsilon_i^{*W})^2] = \hat{\varepsilon}_i^2$, and $\mathbb{E}_*[(\varepsilon_i^{*W})^3] = \hat{\varepsilon}_i^3$, where $\hat{\varepsilon}_i = Y_i - \hat{m}_n(X_i)$, $i = 1, \dots, n$.
2. Set $Y_i^{*W} = \hat{m}_n(X_i) + \varepsilon_i^{*W}$, $i = 1, \dots, n$.
3. Compute $\hat{m}_n^{*W}(x) = \sum_{i=1}^n W_{ni}(x) Y_i^{*W}$.
4. Compute bootstrap version of $T_{n,x}^{\text{het}}$ given by

$$T_{n,x}^{\text{het}\,*} = \frac{\hat{m}_n^{*W}(x) - \mathbb{E}_* \hat{m}_n^{*W}(x)}{\sqrt{\sum_{i=1}^n W_{ni}^2(x)(\hat{\varepsilon}_i^{*W})^2}}$$

$$T_{n,x}^{het} = \frac{\sum_{i=1}^{n} W_{ni}(x)\, \varepsilon_i^{*W}}{\sqrt{\sum_{i=1}^{n} W_{ni}^2(x) (\hat{\varepsilon}_i^{*W})^2}} .$$

Consider constructing bootstrap versions of the pivots T_n and T_n^{het}.

Residual/wild bootstrap versions of T_n and T_n^{het}

Given a grid $x_1, \dots, x_N$ of values in $[0, 1]$, define

$$T_n^* = \max_{1 \leq j \leq N} \left| \frac{\sum_{i=1}^n W_{ni}(x_j)\varepsilon_i^*}{\hat{\sigma}_n^* \sqrt{\sum_{i=1}^n W_{ni}^2(x_j)}} \right| \text{ and } T_n^{\text{het}*} = \max_{1 \leq j \leq N} \left| \frac{\sum_{i=1}^n W_{ni}(x_j)\varepsilon_i^{*W}}{\sqrt{\sum_{i=1}^n W_{ni}^2(x_j)(\hat{\varepsilon}_i^{*W})^2}} \right|$$

and denote by $\hat{G}_n$ and $\hat{G}_n^{\text{het}}$ the dists of T_n^* and $T_n^{\text{het}*}$ conditional on the data.

Exercise:

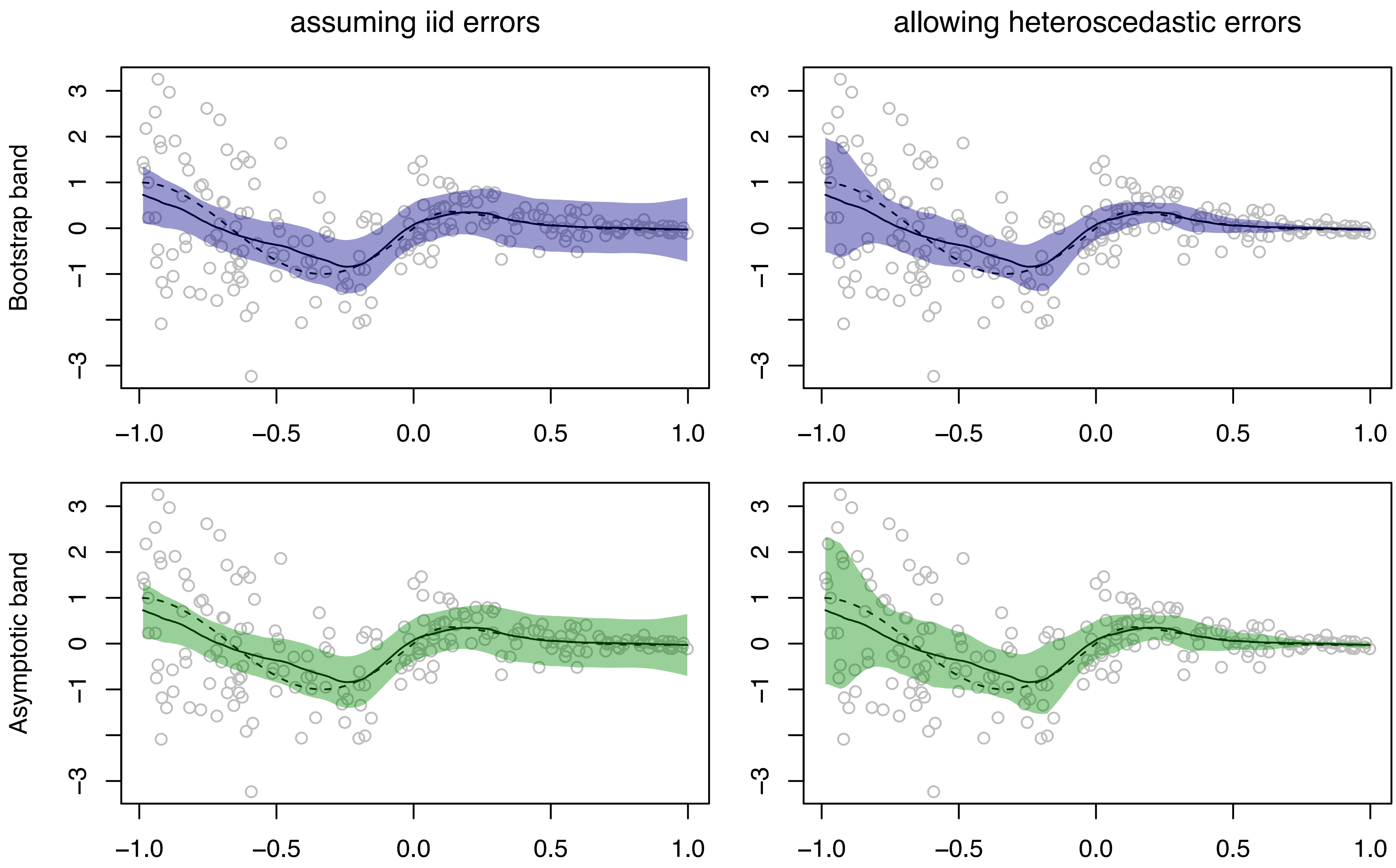
1. Give the form of $(1 - \alpha)100\%$ bootstrap confidence bands under constant variances and heteroscedasticity.
2. Demonstrate on simulated data.

Boot conf band under const. variance

$$\left[\hat{m}_n(x) \pm T_n^{*}(\lceil(1-\alpha)B\rceil)\, \hat{\sigma}_n \sqrt{\sum_{i=1}^{n} w_{ni}^2(x)}\right] \forall x \in [0,1]$$

$$\uparrow$$

$$\left[\hat{m}_n(x) \pm T_n^{*\text{het}}(\lceil(1-\alpha)B\rceil) \sqrt{\sum_{i=1}^{n} w_{ni}^2(x)\, \hat{\varepsilon}_i^2}\right] \forall x \in [0,1]$$


assuming iid errors
allowing heteroscedastic errors
Bootstrap band
Asymptotic band
3
2
1
0
−1
−3
−1.0
−0.5
0.0
0.5
1.0

Debraj Das, Karl Gregory, and Soumendra Lahiri.
Perturbation bootstrap in adaptive lasso.
The Annals of Statistics, 47(4):2080–2116, 2019.

Enno Mammen.
Bootstrap and wild bootstrap for high dimensional linear models.
The Annals of Statistics, pages 255–285, 1993.

Jiayang Sun, Clive R Loader, et al.
Simultaneous confidence bands for linear regression and smoothing.
The Annals of Statistics, 22(3):1328–1345, 1994.

Larry Wasserman.
All of nonparametric statistics.
Springer Science & Business Media, 2006.

Halbert White.
A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity.
Econometrica: Journal of the Econometric Society, pages 817–838, 1980.