IV. Parameterizations of Covariates

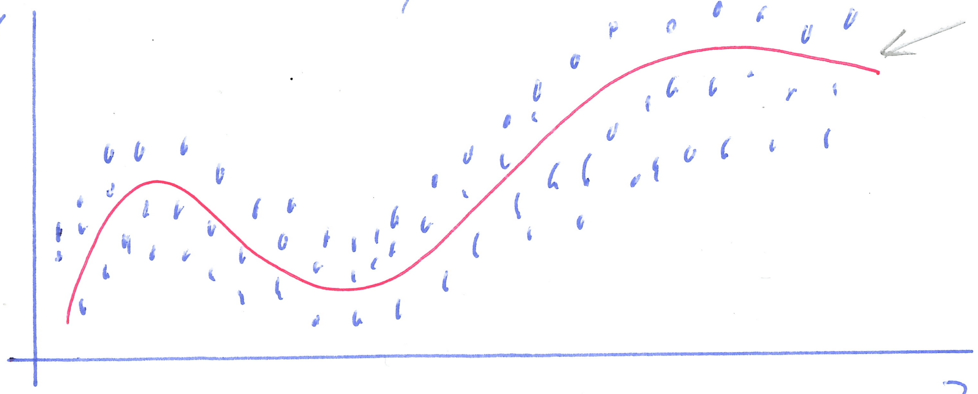
0

4.1 Linearization of the dependence of the response on the covariates

1

2

Data: $(Y_i, Z_i^T)^T$, $i = 1, \dots, n$



$Z = (Z_1, \dots, Z_p)^T \in \mathcal{Z} \subset \mathbb{R}^p$

AIM: Find/propose $t(z)$ such that we can write (model/approximate)

$$E(Y_i \mid Z_i = z) = (t(z))^T \beta \quad \text{for some } \beta \in \mathbb{R}^k$$

$$= E(Y_i \mid X_i = \underbrace{t(z)}_{x}) = x^T \beta \quad (=: m(z) \text{ regression fun.})$$

That is, find/propose $t(z)$ such that transformed data $(Y_i, X_i^T)^T$, $X_i = t(Z_i)$ satisfy a linear model:

$$E(Y \mid X) = X\beta \qquad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} t^T(Z_1) \\ \vdots \\ t^T(Z_n) \end{pmatrix}$$

$$\mathrm{var}(Y \mid X) = \sigma^2 I_n$$

summary 3

4.2

2

4.2 Parameterization of a single covariate

4

NOW: $p=1$, $Z_i \in \mathcal{Z} \subset \mathbb{R}$

AIM: propose $\mathbb{t}: \mathcal{Z} \to \mathbb{R}^k$ such that we can write $\mathbb{E}(Y / Z = z) = \underbrace{\beta_0 t_0(z) + \dots + \beta_{k-1} t_{k-1}(z)}_{=: m(z)}$

generic response / covariate

$=: m(z)$ (regression funct.)

MOST CASES: intercept explicitly included in the model

$$\Rightarrow m(z) = \beta_0 + \beta_1 t_1(z) + \dots + \beta_{k-1} t_{k-1}(z)$$

$$\mathbb{s} = (s_1, \quad \dots \quad , s_{k-1})^T$$

$$\mathbb{s}: \mathcal{Z} \to \mathbb{R}^{k-1}$$

– non-intercept parts of transformation $\mathbb{t}$

→ Parameterization of a covariate

4.2.1 Parameterization

Def 4.1 <u>Parameterization of a covariate</u>

Let $Z_1, \dots, Z_n$ be values of a given covariate $Z \in \mathcal{Z} \subseteq \mathbb{R}$. By a <u>parameterization</u> of this covariate we ~~at~~ mean

(i) the function $\mathbf{s} : \mathcal{Z} \to \mathbb{R}^{k-1}$,
$\mathbf{s}(z) = \big(s_1(z), \dots, s_{k-1}(z)\big)^T$, $z \in \mathcal{Z}$, where all $s_1, \dots, s_{k-1}$ are non-constant functions on $\mathcal{Z}$, and

(ii) an $n \times (k-1)$ matrix $\mathbb{S}$, where

$$\mathbb{S} = \begin{pmatrix} \mathbf{s}^T(Z_1) \\ \vdots \\ \mathbf{s}^T(Z_n) \end{pmatrix} = \begin{pmatrix} s_1(Z_1) & \dots & s_{k-1}(Z_1) \\ \vdots & & \\ s_1(Z_n) & \dots & s_{k-1}(Z_n) \end{pmatrix}.$$

$= \left(X^1, \dots, X^{k-1}\right)$ reparameterizing matrix of a covariate

regressors of the model

$\Rightarrow$ <u>MODEL MATRIX</u> is $\mathbb{X} = \left(\mathbb{1}, \mathbb{S}\right) = \begin{pmatrix} 1 & \mathbf{s}^T(Z_1) \\ \vdots & \\ 1 & \mathbf{s}^T(Z_n) \end{pmatrix}$

<u>Mostly</u>: $\mathbb{S}$ chosen such that

$\mathbb{1} \notin \mathcal{M}(\mathbb{S})$ (a.s.)

$\text{rank}(\mathbb{S}) = k-1$ (a.s.)

$\Rightarrow \text{rank}(\mathbb{X}) = k$ (a.s.)

SUMMARY:

with chosen $\mathcal{S}$, the regression function which models $\mathbb{E}(Y|Z=z)$ is

$$m(z) = \beta_0 + \beta_1 s_1(z) + \cdots + \beta_{k-1} s_{k-1}(z),$$

the model matrix is

$$\mathbb{X} = \begin{pmatrix} 1 & s_1(z_1) & \dots & s_{k-1}(z_1) \\ \vdots & \vdots & & \vdots \\ 1 & s_1(z_n) & & s_{k-1}(z_n) \end{pmatrix}$$

To evaluate appropriateness of chosen $\mathcal{S}$ or to compare different choices of $\mathcal{S}$, we will develop methods to test

(a) $H_0: \beta_j = 0$ (for chosen $j \in \{1, \dots, k-1\}$)

$\rightarrow$ t-test based on

$$T_j = \frac{\hat{\beta}_j}{\sqrt{MS_e v_{jj}}}, \text{ where } \mathbb{V} = (\mathbb{X}^T\mathbb{X})^{-1}$$

(b) $H_0: \beta_{[sub]} = 0$

$\rightarrow$ (Wald) F-test based on

$$F = \frac{1}{m} \hat{\beta}_{[sub]}^T \left(MS_e \mathbb{V}_{[sub,sub]}\right)^{-1} \hat{\beta}_{[sub]},$$

$$m = \# \beta_{sub}$$

1c) X^0, X^1: two model matrices based on two transformations such that

~~M_0~~ $\mathcal{M}(X^0) \subset \mathcal{M}(X^1)$

M_0: $Y/Z \sim (X^0\beta^0, \sigma^2 I_n)$

M_1: $\sim (X^1\beta^1, \sigma^2 I_n)$

model M_0 is simpler model for $\mathbb{E}(Y/Z)$

H_0: $\mathbb{E}(Y/Z) \in \mathcal{M}(X^0)$

F-test on submodel based on

$$F = \frac{\frac{SS_e^0 - SS_e^1}{r_1 - r_0}}{MS_e^0}$$

$r_0 = \text{rank}(X^0)$

$r_1 = \text{rank}(X^1)$

SS_e^0, SS_e^1: residual sums of squares for the two models

$MS_e^0 = \frac{SS_e^0}{n - r_0}$

4.2.2 Covariate types

6

Covariates of different type require different choices of reparameterizations (function g)

→ two main categories of covariates

(1) NUMERIC

(i) continuous

(ii) discrete

(2) CATEGORICAL

7

(i) nominal

(ii) ordinal

Cars2004nh examples

8-12