I. Linear Model

1.1 Regression analysis

2 - 8

Basic methods of regression analysis assume:

DATA: $(Y_i, Z_i^T)^T \overset{iid}{\sim} (Y, Z^T)^T, \quad i = 1, \dots, n$

$Z = (Z_1, \dots, Z_p)^T$

response covariates

$$P(Z \in \mathcal{Z}) = 1, \quad \mathcal{Z} \subset \mathbb{R}^p$$

sample space of covariates

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbb{Z} = \begin{pmatrix} Z_{11} & \dots & Z_{1p} \\ \vdots & & \\ Z_{n1} & \dots & Z_{np} \end{pmatrix} = \begin{pmatrix} Z_1^T \\ \vdots \\ Z_n^T \end{pmatrix}$$

$$= (Z^1, \dots, Z^p)$$

covariate matrix

THIS LECTURE: Y: continuous

Z_j: $j = 1, \dots, p$ - whatever

Probabilistic model for the data

Statistical modelling is based on specifying a probabilistic model for data

prob. model for data $\equiv$ joint distribution of $(Y, Z^T)^T$

$\equiv$ joint density $f_{Y,Z}(y,z)$, $y \in \mathbb{R}$, $z \in \mathcal{Z}$

(with respect to some σ-finite measure $\lambda_y \times \lambda_z$

$$f_{Y,Z}(y,z) = \underbrace{f_{Y|Z}(y|z)}_{\text{interest of this lecture}} \cdot f_Z(z)$$

f_Z : nuisance

(does not have to be specified for many things)

MAIN INTEREST (here) : $E(Y/Z)$

$=$ one of characteristics of $f_{Y|Z}$

Regressors

- $\mathbb{E}(Y|Z)$ = function of Z
- To model $\mathbb{E}(Y|Z)$, transformations of Z might be necessary (why? wait a minute)

$t: \mathcal{Z} \to \mathcal{X} \subseteq \mathbb{R}^k \qquad t = (t_0, \dots, t_{k-1})^T$

(Borel) measurable function

$$X = (X_0, \dots, X_{k-1})^T := (t_0(Z), \dots, t_{k-1}(Z))^T$$
$$= t(Z)$$

$$X_i = (X_{i0}, \dots, X_{ik-1})^T := t(Z_i), \quad i = 1, \dots, n$$

ASSUMPTION : $\mathbb{E}(Y|Z) = m(\underbrace{t(Z)}_{X})$

measurable function

X, X_i : regressor vectors

m : regression function

$$X^j := \begin{pmatrix} X_{1j} \\ \vdots \\ X_{nj} \end{pmatrix}, \quad j = 0, \dots, k-1$$

$= j^{th}$ regressor vector

THIS LECTURE: t assumed to be known (for all theory, not in practice)

DATA: Either $(Y_i, Z_i^T)^T \overset{iid}{\sim} (Y, Z^T)^T$

or $(Y_i, X_i^T)^T \overset{iid}{\sim} (Y, X^T)^T$

$$f_{Y,Z}(y,z) \xrightarrow{\text{transf. theorem}} f_{Y,X}(y,x) = f_{Y|X}(y|x)\, f(x)$$

$$\downarrow$$

$$E(Y|X)$$

ASSUMPTION (the whole lecture):

$$E(Y|Z=z) = E(Y|X=t(z))$$

for almost all $z \in \mathcal{Z}$

Hence to model $E(Y|Z)$ we can also model $E(Y|X)$ using (transformed) data

$(Y_i, X_i^T)^T, \ i=1,\dots,n$

Reminder of the lecture:

If not needed to mention t, data will directly be represented by $(Y_i, X_i^T)^T$.

1.2 Linear model : Basics 9

Def 1.1 Linear model with iid data 10

The data $(Y_i, X_i^T)^T \overset{iid}{\sim} (Y, X^T)^T$, $i = 1, \dots, n$ satisfy a linear model if

$$E(Y|X) = X^T\beta \quad , \quad var(Y|X) = \sigma^2, \quad (a.s.)$$

$\beta = (\beta_0, \dots, \beta_{k-1})^T \in \mathbb{R}^k$, $0 < \sigma^2 < \infty$ unknown par.

↑ regression coef. ↑ residual variance

$$\sigma = \sqrt{\sigma^2} \equiv \text{residual std. deviation}$$

- LM : only E and var of $f_{Y|X}$ is specified

$$f_{X,Y} = f_{Y|X} \cdot f_X$$

  - other aspects of $f_{Y|X}$ not of interest
  - f_X not of interest

- regression function $m(x) = x^T\beta$, $x \in \mathcal{X}$

linear in β

$x = t(z)$ not neces. linear!

LM with intercept : $t_0(z) = 1$

$$X = (1, X_1, \dots, X_{k-1})^T$$

intercept term

$$m(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_{k-1} x_{k-1}$$

Interpretation of regres. coefs.

$\mathbf{x} = (x_0, \ldots, x_j, \ldots, x_{k-1})^T \in \mathcal{X}$

$x^{j(+1)} := (x_0, \ldots, \underline{x_j + 1}, \ldots, x_{k-1})^T \in \mathcal{X}$

clearly: $\mathbb{E}(Y \mid \mathbb{X} = \underline{x^{j(+1)}}) - \mathbb{E}(Y \mid \mathbb{X} = x) = \underline{\beta_j}$

$x^{x_j(+\delta)} := (x_0, \ldots, x_j + \delta, \ldots, x_{k-1})^T \in \mathcal{X}$

$$\mathbb{E}(Y \mid \mathbb{X} = \underline{x^{j(+\delta)}}) - \mathbb{E}(Y \mid \mathbb{X} = x) = \underline{\delta \cdot \beta_j}$$

(i) change in $\mathbb{E}(Y \mid \ldots)$ does not depend on x_j

(ii) change in $\mathbb{E}(Y \mid \ldots)$ does not depend on the remaining regressors

$\beta_j \equiv$ effect of the jth regressor

! interpretation depends on units of x_j !

LM with intercept:

it makes sense only if $(1, 0, \ldots, 0)^T \in \mathcal{X}$:

$$\mathbb{E}(Y \mid \mathbb{X} = (1, 0, \ldots, 0)^T) = \beta_0$$

Linear model with general data

DATA (as assumed by now): $(Y_i, X_i^T)^T \overset{iid}{\sim} (Y, X^T)^T$

NOTATION: $\mathbb{X} = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix}$ row i = vector of regressors for observation i

model matrix

$= (\mathbb{X}^0, \mathbb{X}^1, \dots, \mathbb{X}^{k-1})$

column j = j^{th} regressor

$\mathbb{Y} = (Y_1, \dots, Y_n)^T$, $\mathbb{X}$ are (assumed to be) random

→ their (joint) density

$$f_{\mathbb{Y},\mathbb{X}}(y, x) = f_{\mathbb{Y}|\mathbb{X}}(y|x)\, f_{\mathbb{X}}(x)$$

$$\overset{\text{if iid}}{=} \prod_{i=1}^{n} f_{Y|X}(y_i|x_i) \cdot \prod_{i=1}^{n} f_X(x_i)$$

LM implies some properties of $f_{Y|X}$

Lemma 1.1 Conditional mean and covariance matrix of the response vector

Let $(Y_i, X_i^T)^T \overset{iid}{\sim} (Y, X^T)^T$, $i = 1, \dots, n$ satisfy a linear model, i.e. $E(Y|X) = X^T\beta$, $var(Y|X) = \sigma^2$ for some $\beta \in \mathbb{R}^k$, $0 < \sigma^2 < \infty$.

Then $\mathbb{E}(\mathbb{Y}|\mathbb{X}) = \mathbb{X}\beta$, $var(\mathbb{Y}|\mathbb{X}) = \sigma^2 I_n$.

Proof: trivial... $\boxed{\text{M}}$

iid assumption not really needed, many results can be derived just from assuming WEAKER

$$\mathbb{E}(\mathbb{Y}|\mathbb{X}) = \mathbb{X}\beta, \quad var(\mathbb{Y}|\mathbb{X}) = \sigma^2 I_n.$$

Def. 1.2 Linear model with general data [13]

The data $(\mathbb{Y}, \mathbb{X})$ satisfy a linear model if $\mathbb{E}(\mathbb{Y}|\mathbb{X}) = \mathbb{X}\beta$, $var(\mathbb{Y}|\mathbb{X}) = \sigma^2 I_n$, where $\beta = (\beta_0, \dots, \beta_{k-1})^T \in \mathbb{R}^k$ and $0 < \sigma^2 < \infty$ are unknown parameters.

NOTATION

- $(Y_i, X_i^T)^T \overset{iid}{\sim} (Y, X^T)^T, \quad i = 1, \dots, n$

  $Y|X \sim (X^T\beta, \sigma^2) \quad \equiv$ LM with iid data

- $\underline{\mathbb{Y}|\mathbb{X} \sim (\mathbb{X}\beta, \sigma^2 I_n)} \quad \equiv$ LM with general data

  $\downarrow$

  here, in $f_{\mathbb{Y},\mathbb{X}}(y, x) = \underbrace{f_{\mathbb{Y}|\mathbb{X}}(y|x)}_{\text{not neces. } \prod_{i=1}^n f_{Y|X}(y_i|x_i)} \cdot \underbrace{f_{\mathbb{X}}(x)}_{\text{not neces. } \prod_{i=1}^n f_X(x_i)}$

Rank of the model

Regressor vectors $X_1, \dots, X_n$ ($\equiv$ model matrix $\mathbb{X}$)

$\sim$ jointly $f_{\mathbb{X}}(x_1, \dots, x_n) = f_{\mathbb{X}}(x)$

ASSUMPTIONS

- $n > k$
- $P(\operatorname{rank}(\mathbb{X}) = r) = 1$ for some $r \leq k$

column rank

Def 1.3 Full-rank model

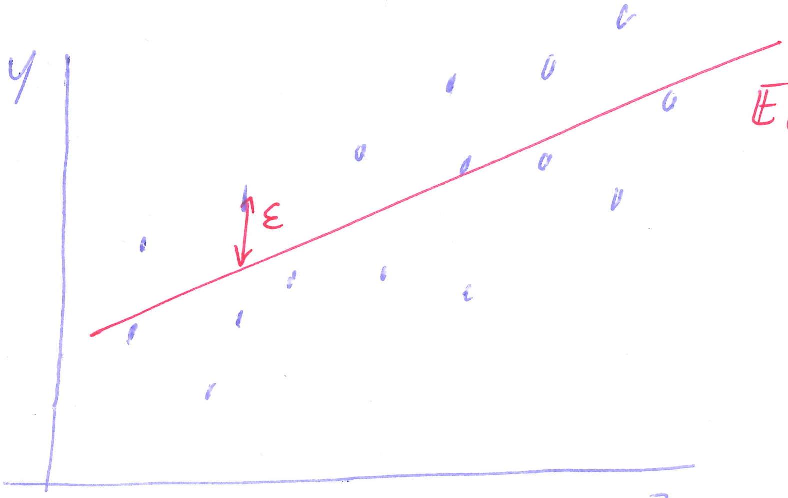
A full-rank linear model is such a linear model where $r = k$.

$\equiv$ columns of $\mathbb{X}$ are (a.s.) linearly independent vectors

---

In most practical situations, full-rank models are sufficient. In the rest of lecture, we will mostly (always) assume that $r = k$.

Error terms



$\mathbb{E}(Y|Z=z) = \beta_0 + \beta_1 z = X^T\beta$

$z \quad , \quad X = (1, z)$

$\varepsilon_i := Y_i - X_i^T\beta \; , \; i = 1, \dots, n \qquad$ = error terms

$\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T = \mathbb{Y} - \mathbb{X}\beta \qquad$ = error term vector

---

Lemma 1.2 Moments of the error terms

Let $\mathbb{Y}|\mathbb{X} \sim (\mathbb{X}\beta, \sigma^2 I_n)$. Then

$$\mathbb{E}(\boldsymbol{\varepsilon}|\mathbb{X}) = 0_n \quad , \quad \mathbb{E}(\boldsymbol{\varepsilon}) = 0_n$$

$$var(\boldsymbol{\varepsilon}|\mathbb{X}) = \sigma^2 I_n \quad ; \quad var(\boldsymbol{\varepsilon}) = \sigma^2 I_n .$$

Proof: • $\mathbb{E}(\boldsymbol{\varepsilon}|\mathbb{X}) = \mathbb{E}(\mathbb{Y} - \mathbb{X}\beta|\mathbb{X}) = \mathbb{E}(\mathbb{Y}|\mathbb{X}) - \mathbb{E}(\mathbb{X}\beta|\mathbb{X})$

$$= \mathbb{X}\beta - \mathbb{X}\beta = 0_n$$

• $var(\boldsymbol{\varepsilon}|\mathbb{X}) = var(\mathbb{Y} - \mathbb{X}\beta|\mathbb{X}) = var(\mathbb{Y}|\mathbb{X}) = \sigma^2 I_n$

• $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbb{E}(\mathbb{E}(\boldsymbol{\varepsilon}|\mathbb{X})) = \mathbb{E}(0_n) = 0_n$

• $var(\boldsymbol{\varepsilon}) = \mathbb{E}(var(\boldsymbol{\varepsilon}|\mathbb{X})) + var(\mathbb{E}(\boldsymbol{\varepsilon}|\mathbb{X})) =$

$$= \mathbb{E}(\sigma^2 I_n) + var(0_n) = \sigma^2 I_n \qquad \blacksquare$$

NOTE: $(Y_i, X_i^T)^T \overset{iid}{\sim} (Y, X^T)^T$

$$\Rightarrow \varepsilon_i \overset{iid}{\sim} \varepsilon \; , \quad \varepsilon \sim (0, \sigma^2) .$$

Distributional assumptions

16

To derive some results, we will have to assume more than just

$$E(\mathbb{Y}|\mathbb{X}) = \mathbb{X}\beta \ ; \ \mathrm{var}(\mathbb{Y}|\mathbb{X}) = \sigma^2 I_n .$$

Sometimes, we will assume

$$(Y_i, X_i^T) \overset{iid}{\sim} (Y, X^T)^T , \quad Y|X \sim N(X^T\beta, \sigma^2)$$

Which implies $\mathbb{Y}|\mathbb{X} \sim N(\mathbb{X}\beta, \sigma^2 I_n)$

Fixed or random covariates

17

Sometimes, covariates are <u>not</u> random, they are pre-specified by the analyst (designed experiments, ...)

If interest in $Y|X$ or $\mathbb{Y}|\mathbb{X}$, it is not really necessary to distinguish whether X is random or not.

$$X \sim f_X(x) \equiv \text{nuisance for us}$$

Limitations of a linear model

18

George Box

~~[illegible]~~ If linear model applied to some dataset, it is assumed (≡ it is believed) that

(1) $E(Y|Z=z) = E(Y|\cancel{X}=t(z)) =$ 19

$=$ linear function of $x=t(z)$ 20

21

(2) $var(Y|X=x) = \sigma^2 = const$ 22

$var(Y|X) = \sigma^2 I_n$ ≡ diagonal matrix

≡ uncorrelated obs.

(3) $Y|X \sim N(\quad)$ 23

Sometimes: interest in modelling also other features of $f_{Y|X}$ than just E and var.