Lecture 12 | 14.05.2024

Regression models

beyond the expectation

Linear regression models

- **Random sample** $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ 1 = 1, \dots, N\}$ drawn from the joint distribution of some generic random vector $(Y, \boldsymbol{X}^\top)^\top \sim F_{(Y,\boldsymbol{X})}$
- **The conditional distribution of $Y|\boldsymbol{X}$ is assumed to be normal** (or at least close to normal) such that $Y|\boldsymbol{X} \sim N(\boldsymbol{X}^\top\beta, \sigma^2)$
- **The model specifies the conditional expectation** $E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$ or, alternatively, in terms of the empirical data $E[Y_i|\boldsymbol{X}_i] = \boldsymbol{X}_i^\top\beta$
- **The unknown parameters $\beta \in \mathbb{R}^p$ and $\sigma^2 > 0$** are obtained either by minimizing the least squares criterion or by maximizing the likelihood
- **The estimates $\widehat{\beta} = (\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$ and $\widehat{\sigma^2} = SSe/(n-p)$** are unbiased and consistent estimates of their theoretical (unknown) counterparts

Least squares vs. maximum likelihood

- In an ordinary linear regression model (without the normality assumption) the likelihood can not be obtained and the estimates for $\beta \in \mathbb{R}^p$ are obtained by minimizing least squares

$$\widehat{\beta} = \operatorname*{Arg\,min}_{\beta \in \mathbb{R}^p} \sum_{i=1}^{N} (Y_i - \boldsymbol{X}_i^\top \beta)^2$$

- In a normal linear regression model (under the normality assumption) the full likelihood for β and σ^2 can be formulated and the estimates are obtained by maximizing the likelihood function

$$\widehat{\beta} = \operatorname*{Arg\,max}_{\beta \in \mathbb{R}^p} (2\pi\sigma^2)^{-N/2} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (Y_i - \boldsymbol{X}_i^\top \beta)^2 \right\}$$

Least squares vs. maximum likelihood

- In an ordinary linear regression model (without the normality assumption) the likelihood can not be obtained and the estimates for $\beta \in \mathbb{R}^p$ are obtained by minimizing least squares

$$\widehat{\beta} = \operatorname*{Arg\,min}_{\beta \in \mathbb{R}^p} \sum_{i=1}^{N} (Y_i - \boldsymbol{X}_i^\top \beta)^2$$

- In a normal linear regression model (under the normality assumption) the full likelihood for β and σ^2 can be formulated and the estimates are obtained by maximizing the likelihood function

$$\widehat{\beta} = \operatorname*{Arg\,max}_{\beta \in \mathbb{R}^p} (2\pi\sigma^2)^{-N/2} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (Y_i - \boldsymbol{X}_i^\top \beta)^2 \right\}$$

- Note, that under the normal linear model the least squares estimation and the maximum likelihood estimation are two equivalent formulation of the same problem—the minimization of the sum of squared residuals

Simple expectation and beyond

- For some real random variable $X \sim F$ (and the density f with respect to the Lebesgue or count measure) and some measurable funtion $h: \mathbb{R} \to \mathbb{R}$ we can obtain the expectation (if the integral exists) as

$$Eh(X) = \int_{\mathbb{R}} h(x) \mathrm{d}F(x) = \int_{\setminus} h(x) f(x) \mathrm{d}x$$

- For the random sample $X_1, \dots, X_N$ drawn from the same distribution as the distribution of X we can construct the empirical distribution function F_N and the empirical counterpart for $Eh(X)$ (i.e., the empirical estimate)

$$\widehat{Eh(X)} = \int_{\mathbb{R}} h(x) \mathrm{d}F_N(x) = \sum_{i=1}^{N} h(X_i)$$

- The quantity (parameter) $\mu_h = Eh(X)$ is sometimes called the theoretical functional of the distribution F while the quantity $\widehat{\mu_h} = \widehat{Eh(X)}$ is called the (empirical) functional of the empirical distribution F_N
- Note, that different fuctions can be used in place of h in both expressions (e.g., $h(x) = x$ gives the mean EX and the corresponding sample mean (average) $\overline{X}_N$)

Different choices of the function h

- The mean (theoretical functional) and the average (empirical functional) can be also obtained via different formulation for $h(x) = (x-a)^2$ as

$$E[X] = \operatorname*{Arg\,min}_{a \in \mathbb{R}} E(X-a)^2 = \operatorname*{Arg\,min}_{a \in \mathbb{R}} \int_{\mathbb{R}} (x-a)^2 \mathrm{d}F(x)$$

and, correspondingly also

$$\overline{X}_N = \operatorname*{Arg\,min}_{a \in \mathbb{R}} \int_{\mathbb{R}} (x-a)^2 \mathrm{d}F_N(x) = \operatorname*{Arg\,min}_{a \in \mathbb{R}} \sum_{i=1}^{N} (X_i - a)^2$$

Different choices of the function h

- The mean (theoretical functional) and the average (empirical functional) can be also obtained via different formulation for $h(x) = (x - a)^2$ as

$$E[X] = \operatorname*{Arg\,min}_{a \in \mathbb{R}} E(X - a)^2 = \operatorname*{Arg\,min}_{a \in \mathbb{R}} \int_{\mathbb{R}} (x - a)^2 \mathrm{d}F(x)$$

  and, correspondingly also

$$\overline{X}_N = \operatorname*{Arg\,min}_{a \in \mathbb{R}} \int_{\mathbb{R}} (x - a)^2 \mathrm{d}F_N(x) = \operatorname*{Arg\,min}_{a \in \mathbb{R}} \sum_{i=1}^{N} (X_i - a)^2$$

- Note, that in both cases we actually formulate the least squares problem (theoretical and empirical) and the solution is the theoretical mean and the empirical average (i.e., the estimate for the mean)

Different choices of the function h

- The mean (theoretical functional) and the average (empirical functional) can be also obtained via different formulation for $h(x) = (x - a)^2$ as

$$E[X] = \operatorname*{Arg\,min}_{a \in \mathbb{R}} E(X - a)^2 = \operatorname*{Arg\,min}_{a \in \mathbb{R}} \int_{\mathbb{R}} (x - a)^2 \mathrm{d}F(x)$$

and, correspondingly also

$$\overline{X}_N = \operatorname*{Arg\,min}_{a \in \mathbb{R}} \int_{\mathbb{R}} (x - a)^2 \mathrm{d}F_N(x) = \operatorname*{Arg\,min}_{a \in \mathbb{R}} \sum_{i=1}^{N} (X_i - a)^2$$

- Note, that in both cases we actually formulate the least squares problem (theoretical and empirical) and the solution is the theoretical mean and the empirical average (i.e., the estimate for the mean)
- **This principle can be generalized even further—for the regression concepts and different forms of the function** h

Regression in terms of the functional representation

- For the (population) regression model $Y = \boldsymbol{X}\beta + \varepsilon$, where $\varepsilon \sim (0, \sigma^2)$ and $E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$ the true (unknown) vector of parameters $\beta \in \mathbb{R}^p$ satisfies

$$\beta = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} E[Y - \boldsymbol{X}^\top\boldsymbol{b}]^2 = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} \int_{\mathbb{R}^{1+p}} (y - \boldsymbol{x}^\top\boldsymbol{b})^2 \mathrm{d}F_{(Y,\boldsymbol{X})}(y, \boldsymbol{x})$$

where $\beta = \mathcal{F}(F_{(Y,\boldsymbol{X})})$ can be seen as a (theoretical) functional (TF) of the joint distribution $F_{(Y,\boldsymbol{X})}$

- For the random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ 1 = 1, \dots, N\}$ drawn from the same distribution $F_{(Y,\boldsymbol{X})}$ we can formulate the empirical counterpart as

$$\widehat{\beta}_N = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} \int_{\mathbb{R}^{1+p}} (y - \boldsymbol{x}^\top\boldsymbol{b})^2 \mathrm{d}F_N^{(Y,\boldsymbol{X})}(y, \boldsymbol{x}) = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} \sum_{i=1}^{N} (Y_i - \boldsymbol{X}_i^\top\boldsymbol{b})^2$$

where $\widehat{\beta}_N = \mathcal{F}(F_N^{(Y,\boldsymbol{X})})$ can be now seen as the corresponding (empirical) functional (EF) of the empirical distribution function $F_N^{(Y,\boldsymbol{X})}$

Regression models beyond the expectation I

Median regression

$$\text{TF: } \beta = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} E|Y - \boldsymbol{X}^\top \boldsymbol{b}| \qquad \text{and} \qquad \text{EF: } \widehat{\beta}_N = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} \sum_{i=1}^N |Y_i - \boldsymbol{X}_i^\top \boldsymbol{b}|$$

for the choice $h(x) = |x - a|$ where $\boldsymbol{X}^\top\beta$ is the conditional median of Y given $\boldsymbol{X}$ and $\widehat{\beta}_N$ is the corresponding (empirical) estimate

Quantile regression

$$\text{TF: } \beta = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} h_\tau(Y - \boldsymbol{X}^\top \boldsymbol{b}) \qquad \text{and} \qquad \text{EF: } \widehat{\beta}_N = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} \sum_{i=1}^N h_\tau(Y_i - \boldsymbol{X}_i^\top \boldsymbol{b})$$

for the so called quantile check function $h_\tau(x) = \tau(x - \mathbb{I}_{\{x<0\}})$ and some quantile level $\tau \in (0,1)$ where $\boldsymbol{X}^\top\beta$ is the conditional τ-level quantile of Y given $\boldsymbol{X}$ and $\widehat{\beta}_N$ is the corresponding (empirical) estimate

Regression models beyond the expectation II

Expectile regression

$$\text{TF: } \beta = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} h_\tau(Y - \boldsymbol{X}^\top \boldsymbol{b}) \qquad \text{and} \qquad \text{EF: } \widehat{\beta}_N = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} \sum_{i=1}^{N} h_\tau(Y_i - \boldsymbol{X}_i^\top \boldsymbol{b})$$

for the choice $h_\tau(x) = |\tau - \mathbb{I}_{\{x<0\}}|x^2$ and some expectile level $\tau \in (0,1)$ where $\boldsymbol{X}^\top\beta$ is the conditional expectile of Y given $\boldsymbol{X}$ and $\widehat{\beta}_N$ is the corresponding (empirical) estimate

Robust regression

$$\text{TF: } \beta = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} \rho(Y - \boldsymbol{X}^\top \boldsymbol{b}) \qquad \text{and} \qquad \text{EF: } \widehat{\beta}_N = \operatorname*{Arg\,min}_{\boldsymbol{b}\in\mathbb{R}^p} \sum_{i=1}^{N} \rho(Y_i - \boldsymbol{X}_i^\top \boldsymbol{b})$$

for some convex (and robust) loss function $\rho(x)$ where the interpretation of $\boldsymbol{X}^\top\beta$ and its empirical counterpart $\widehat{\beta}_N$ now depends on the choice of the loss function ρ

Basic properties of the regression variants

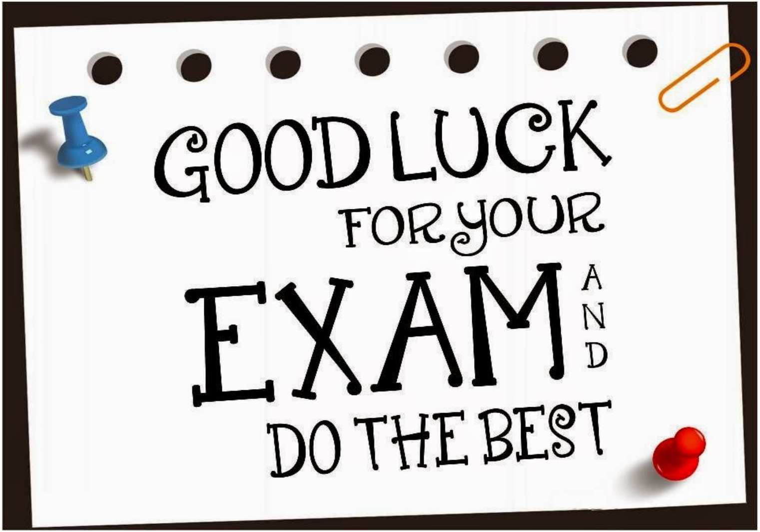
- **Median regression**
  - More robust than the standard least squares regression
  - For symmetric error distributions the median corresponds with the mean
  - Easy and straightforward intepretation of the estimated parameters
- **Quantile regression**
  - Generalization of the median regression (which is obtained for $\tau = 0.5$)
  - Provideds a complex insight about the conditional distribution of $Y|\boldsymbol{X}$
  - Relatively easy interpretation but not that much popular in practice
- **Expectle regression**
  - Generalization of the least squares (which are obtained for $\tau = 0.5$)
  - Expectiles form elastic and elucitable risk measures (unlike quantiles)
  - Relatively difficult interpretation of β but very popular in risk theory
- **Robust regression**
  - Generalization of the regression for outlyiers and heavy-talied distributions
  - Least squares for $\rho(x) = x^2$, median regression for $\rho(x) = |x|$, maximum likelihood for $\rho(x) = -log(x)$
  - Other choices are common in practice as well (e.g., Huber function, Tukey function, Andrew's function)

Exam terms

- **Exam 1**
  Thursday, 16.05.2024 | Lecture room Praktikum KPMS | Start at 12:20
- **Exam 2**
  Tuesday, 28.05.2024 | Lecture room K11 | Start at 09:00
- **Exam 3**
  Tuesday, 30.05.2024 | Lecture room K11 | Start at 09:00

- **Exam 4**
  ↪ will be schedulled later (in the week 10.06 – 14.06, 2024)
- **Exam 5**
  ↪ will be schedulled later (in september 2024)

Registration for any exam via the SIS system only!


GOOD LUCK
FOR YOUR
EXAM
AND
DO THE BEST