Lecture 10 | 30.04.2024

Regression models

beyond linearity

Linear regression models

Normal linear regression model

- generic regression model $Y = \boldsymbol{X}^\top\beta + \varepsilon$, for $\varepsilon \sim N(0, \sigma^2)$
- random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ 1 = 1, \dots, n\}$ from $F_{(Y,\boldsymbol{X})}$
- conditional distribution of $Y|\boldsymbol{X}$ is normal, i.e., $Y|\boldsymbol{X} \sim N(\boldsymbol{X}^\top\beta, \sigma^2)$
- parameter estimates (LSE/MLE) are BLUE and normally distributed
- easy and straightforward statistical inference

Linear regression model without normality

- generic regression model $Y = \boldsymbol{X}^\top\beta + \varepsilon$, for $\varepsilon \sim (0, \sigma^2(\boldsymbol{X}))$
- mean ($E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$) and variance ($Var(Y|\boldsymbol{X}) = \sigma^2(\boldsymbol{X})$) specification
- conditional distribution of $Y|\boldsymbol{X}$ is left unspecified (LSE only)
- parameter estimates (MLE) are BLUE and asymptotically normal
- relatively straightforward inference based on the central limit theorem

Linear regression models

Normal linear regression model

- generic regression model $Y = \boldsymbol{X}^\top \beta + \varepsilon$, for $\varepsilon \sim N(0, \sigma^2)$
- random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ 1 = 1, \dots, n\}$ from $F_{(Y,\boldsymbol{X})}$
- conditional distribution of $Y|\boldsymbol{X}$ is normal, i.e., $Y|\boldsymbol{X} \sim N(\boldsymbol{X}^\top\beta, \sigma^2)$
- parameter estimates (LSE/MLE) are BLUE and normally distributed
- easy and straightforward statistical inference

Linear regression model without normality

- generic regression model $Y = \boldsymbol{X}^\top \beta + \varepsilon$, for $\varepsilon \sim (0, \sigma^2(\boldsymbol{X}))$
- mean ($E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$) and variance ($Var(Y|\boldsymbol{X}) = \sigma^2(\boldsymbol{X})$) specification
- conditional distribution of $Y|\boldsymbol{X}$ is left unspecified (LSE only)
- parameter estimates (MLE) are BLUE and asymptotically normal
- relatively straightforward inference based on the central limit theorem

Recall, that **linearity** + **normality** = **"lightness of being"** but linear regression models without the assumptions of normality introduce just a minor complication...
Thus, the linearity property is way more crucial!

Beyond linearity

- In practice, however: **The truth is (almost) never linear!**
  (however, the linearity assumption is a good and easy approximation)
- **What to do, when the linearity assumption fails?**
  (the answer usually depends on the reason why the linearity fails)
- **Note, that there are a few levels of linearity in the model**
  (linearity of the predictor, linearity of the expectation, linearity of LS)

Beyond linearity

- In practice, however: **The truth is (almost) never linear!**
  (however, the linearity assumption is a good and easy approximation)
- **What to do, when the linearity assumption fails?**
  (the answer usually depends on the reason why the linearity fails)
- **Note, that there are a few levels of linearity in the model**
  (linearity of the predictor, linearity of the expectation, linearity of LS)

  - the data are too flexible (higher order approximations/splines)
  - the data are too irregular (piecewise approximation)
  - the data are too complex (additive models)
  - the data are too volatile (robust estimation approaches)
  - the data contradicts the linear model (GLM)
  - the data can not be approximated by a linear model (nonlinear models)

Beyond linearity

- In practice, however: **The truth is (almost) never linear!**
  (however, the linearity assumption is a good and easy approximation)
- **What to do, when the linearity assumption fails?**
  (the answer usually depends on the reason why the linearity fails)
- **Note, that there are a few levels of linearity in the model**
  (linearity of the predictor, linearity of the expectation, linearity of LS)

  - the data are too flexible (higher order approximations/splines)
  - the data are too irregular (piecewise approximation)
  - the data are too complex (additive models)
  - the data are too volatile (robust estimation approaches)
  - the data contradicts the linear model (GLM)
  - the data can not be approximated by a linear model (nonlinear models)
  - and many more reasons (and way more alternatives)

Generalized linear models (GLM)

So far, all regression models concerned the response variable $Y \in \mathbb{R}$ that was apriori assumed to be continuous and the conditional distribution of $Y|\boldsymbol{X}$ was assumed to be normal or, at least, close to normal...

In practical applications, however, the domain of Y can be also more restricted...

- $Y \in \mathbb{N} \cup \{0\}$ (counts)
- $Y \in \{1, \ldots, K\}$ for $K \in \mathbb{N}$ (categories/label)
- $Y \in \{0, 1\}$ (true/false)
- ...

Generalized linear models (GLM)

So far, all regression models concerned the response variable $Y \in \mathbb{R}$ that was apriori assumed to be continuous and the conditional distribution of $Y|\boldsymbol{X}$ was assumed to be normal or, at least, close to normal...

In practical applications, however, the domain of Y can be also more restricted...

- $Y \in \mathbb{N} \cup \{0\}$ (counts)
- $Y \in \{1, \dots, K\}$ for $K \in \mathbb{N}$ (categories/label)
- $Y \in \{0, 1\}$ (true/false)
- ...

Note, that despite the fact that the domain of Y is restricted (discrete or even finite), the mean parameter of Y (the conditional mean if $Y|\boldsymbol{X}$ respectively) is still assumed to be from some well defined compact subset, $\mathcal{M} \subset \mathbb{R}$...

This will be also used in the following models...

Linear models with a flavour of nonlinearity

- in a standard linear model (OLM)—the conditional mean is modelled as

$$E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top \beta, \quad \text{for } \beta \in \mathbb{R}^p$$

  while the variance structure $Var[Y|\boldsymbol{X}]$ is modeled separately and independently from the mean structure (e.g., $Var[Y|\boldsymbol{X}] = \sigma^2$)

- in a generalized linear model (GLM)—the conditional mean is modelled as

$$g(E[Y|\boldsymbol{X}]) = \boldsymbol{X}^\top \beta, \quad \text{for } \beta \in \mathbb{R}^p$$

  for some non-linear link function g, where $g^{-1} : \mathbb{R} \to \mathcal{M}$ (typically continuous, smooth, monotone, but nonlinear) and the variance structure depends on the mean (i.e., $Var[Y|\boldsymbol{X}] = v(E[Y|\boldsymbol{X}])\phi$)

Example 1: Logistic regression

Logistic regression

- the response variable $Y \in \mathbb{R}$ takes only two possible values, $Y \in \{0,1\}$
- the conditional distribution of $Y|\boldsymbol{X}$ is alternative, with the probability of success $p_x = P[Y=1|\boldsymbol{X}] \in \mathcal{M}$, where $\mathcal{M} = [0,1]$
- the conditional mean $\mu_x = E[Y|\boldsymbol{X}] = P[Y=1|\boldsymbol{X}]$ is modeled with the linear predictor $\boldsymbol{X}^\top\beta$ using the logit link function $g(x) = \log[x/(1-x)]$
- the model assumes the mean structure

$$logit(\mu_x) = \log\frac{E[Y|\boldsymbol{X}]}{1-E[Y|\boldsymbol{X}]} = \log\frac{P[Y=1|\boldsymbol{X}]}{1-P[Y=1|\boldsymbol{X}]} = \boldsymbol{X}^\top\beta$$

- note, that for the link function it holds that $g^{-1}: \mathbb{R} \to \mathcal{M}$
- the model postulates the variance structure in a form

$$Var[Y|\boldsymbol{X}] = v(\mu_x) = \mu_x(1-\mu_x)$$

  (which fully corresponds with the mean/variance structure of some random variable with the alternative distribution)
- the model is interpreted in terms of multiplicative comparisons and the parameters are interpreted in terms of the odds ratios (probabilities resp.)

Example 2: Poisson regression

Logistic regression

- the response variable $Y \in \mathbb{N} \cup \{0\}$ represents integer counts (including 0)
- the conditional distribution of $Y|\boldsymbol{X}$ is Poisson, with $\lambda_x = E[Y|\boldsymbol{X}]$
- the conditional mean $\lambda_x = E[Y|\boldsymbol{X}] \in \mathcal{M} \equiv \mathbb{R}_+$ is modeled with the linear predictor $\boldsymbol{X}^\top \beta$ using the log link function $g(x) = \log x$
- the model assumes the mean structure

$$log(\lambda_x) = \log E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top \beta$$

- note, that for the link function it holds that $g^{-1} : \mathbb{R} \to \mathcal{M}$
- the model assumes the variance structure which depends on the mean λ_x and some additional dispersion parameter $\phi > 0$

$$Var[Y|\boldsymbol{X}] = v(\lambda_x)\phi = \phi\lambda_x$$

  (which fully corresponds with the mean/variance structure of some random variable with the Poisson (overdispersed) distribution)
- the model is interpreted in terms of multiplicative comparisons and the parameters are interpreted in terms of the proportional changes of the conditional expectations

Example 3: Special cases

- **Classical linear regression model**
  - continuous response $Y \in \mathbb{R}$
  - the mean parameter $\mu_x = E[Y|\boldsymbol{X}] \in \mathcal{M} \equiv \mathbb{R}$
  - identity link function $g(x) = x$
  - constant variance function $v(x) = 1$ and $\phi = \sigma^2$
- **Multinomial regression model**
- **Exponential data model**
- **Gamma model**
- **...**

Nonlinear regression models

- In linear models and generalized linear models as well, the conditional mean is modeled (using a proper link function) as a linear combination of the response variables and the subset of unknown parameters...
- If the class of available models is not reach enough (and we still prefer a parametric model structure) $\Longrightarrow$ nonlinear (parametric) regression models can serve a a good alternative/compromise...
- The principal idea of the nonlinear models is to use a general parametic (but nonlinear) function $f : \mathbb{R}^{p \times q} \to \mathcal{M} \subseteq \mathbb{R}$, such that

$$E[Y|\boldsymbol{X}] = f(\boldsymbol{X}, \beta),$$

  where $\boldsymbol{X} \in \mathbb{R}^p$ and $\boldsymbol{\beta} \in \mathbb{R}^q$
- Note, that nonlinear element (the nonlinear function f) is now introduced on the other side of the classical regression model formula and the dimensions of $\boldsymbol{X} \in \mathbb{R}^p$ and $\boldsymbol{\beta} \in \mathbb{R}^q$ may now differ
- The primary interest is on the mean structure modeling

Nonlinear regression: Some examples

There are, of course, plenty of different models with various analytical structure and different regularity properties (smoothness, continuity, isotonic properties, etc).
Typical nonlinear models are, for instance, various population growth models...

- **Exponential growth model**

$$f(x, \beta, \alpha) = \alpha \exp\{X\beta\}$$

  $\rightarrow$ for some parameters $a > 0$ and $\beta > 0$;

- **Logistic growth model**

$$f(X, \beta, \alpha, K) = \frac{K}{1 + be^{-X\beta}};$$

  $\rightarrow$ for some parameters $\alpha, \beta, K > 0$;

- **Gomertz growth model**

$$f(X, \beta, \alpha, K) = K \cdot exp\{-\beta e^{-\alpha t}\};$$

  $\rightarrow$ for some parameters $\alpha, \beta, K > 0$;

Solutions for nonlinear regression models

- Note, that all three nonlinear models above can not be solved by using classical method of the least squares...
  (no explicit solution can be obtained)
- Thus, different computation strategies must be used to obtain the model solution—the estimates for the unkown parameters $\alpha, \beta, K > 0$
- Such computational methods may involve:
  - reparametrization into a linear model and applying least squares
  - model approximation and least squares
  - various iterative solutions
- Note, that as far as the unknown regression function is unspecified, the corresponding minimization problem may not even be convex!

Generalized nonlinear models

- **Advanced, but still possible**....

$$g(E[Y|\boldsymbol{X}]) = f(\boldsymbol{X}, \beta)$$

where two additional sources of nonlinearity are introduced at the same time—the nonlinear link function g and the nonlinear predictor function f

- **Some challenges**
  - mostly, the interpretation of $\beta \in \mathbb{R}$ is not straightforward
  - due to nonlinearity, various computational issues and solution instability
  - difficult statistical inference typically perfomed by simulations

Summary

- **Linear regression models**
  The term “linear” primarily refers to the linearity of the predictor $\boldsymbol{X}^\top\beta$ which is a linear combination of the unknown parameters $\beta \in \mathbb{R}^p$ and some information from the data
- **Nonlinear regression models**
  The term “nonlinear” primarily refers to the fact, that the regression model uses a predictor $f(\boldsymbol{X}, \beta)$ which is a nonlinear function of the unknown parameters $\beta \in \mathbb{R}^p$
- **Generalized linear models**
  The term “generalized” refer to a class of regression models where the conditional mean of the response Y is modeled with the linear predictor $\boldsymbol{X}^\top\beta$ using some nonlinear link function g
- **General linear models**
  The term “general” refers to a sub-class of linear regression models (the first item) where the conditional variance structure of the response variable Y is modeled by a general variance matrix $\sigma^2\mathbb{W}$