Part II: Generalized Linear Models

Chapter II.2

Theory of Generalized Linear Models

Topics

To be discussed...

- fundamentals of GLMs: definition, components, representative types
- exponential dispersion family of distributions
- link functions
- estimation of the parameter vector
- evaluation of the model fit & model selection
- residuals

Linear models (LMs)

Recall Definition I.4.2/I.4.3 (Normal LM)

Let $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ with Var $\epsilon_1 = \sigma^2 > 0$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \Theta \subseteq \mathbb{R}^p$ be a parameter space and $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a (known) matrix.

Then, we call

- $Y = \mathbf{X}\beta + \varepsilon$ **normal linear model**,
- $Y = (Y_1, \ldots, Y_n)'$ vector of observations,
- $\mathbf{X}$ design matrix or model matrix,
- β parameter vector
- ε error term.

Normal LMs are fundamental for modeling the relationship between a response variable and one or more explanatory variables, when

- the relationship is of a form supported by LMs,
- the response variable is normal distributed.

Non-linear models

II.2.1 Remark (linear vs. non-linear models)

Linear Models: A model is linear with respect to its parameters!

Example: $Y = \beta_0 + \beta_1 X + \beta_2 X^2$ is a LM (quadratic and polynomial regression; see I.5.6-I.5.9)

Non-linear Models: Some non-linear models can be inverted into a linear form by suitable mathematical transformation.

- $Y = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2} \quad \Rightarrow \quad \log(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$
- $I \times J$ contingency table $\boldsymbol{\pi} = (\pi_{ij})$, with π_{ij} the probability of cell (i,j):
  The row and column classification variables X and Y are stochastically independent if $\pi_{ij} = \pi_{i+}\pi_{+j}$ for $i = 1, \ldots, I$, $j = 1, \ldots, J$.
  $$\pi_{ij} = \pi_{i+}\pi_{+j} \;\Rightarrow\; \log(\pi_{ij}) = \log(\pi_{i+}) + \log(\pi_{+j}) \quad ☞ \quad \log(\pi_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$$
  log-linear model

Not every non-linear model can be transformed to a LM!

Example: $Y = \beta_0 + \beta_1 e^{\beta_2 X_1} + \beta_3 e^{\beta_4 X_2}$

Fundamentdals of GLMs

Normal **LM:** $\mu = \mathrm{E}(\mathbf{Y}) = \mathbf{X}\beta$

$\downarrow$ *random component* $\searrow$ *systematic component*

GLM (Nelder & Wedderburn, 1972 ; McCullagh & Nelder, 1989)

- *relaxes the assumption of normal distribution* for the response Y,
- *links* the systematic component to a function g of μ:

$$\boxed{\eta = g(\mu) = g[\mathrm{E}(\mathbf{Y})] = \mathbf{X}\beta}$$

link function $\swarrow$ $\downarrow$ *random component* $\searrow$ *linear predictor*

It indicates how the expected/ predicted value of the response relates to the linear combination of predictor variables.

- Normal LM
  - identity link: $g(\mu) = \mu$
  - random component: $Y_1, \ldots, Y_n \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, \sigma^2)$ with constant $\sigma^2 > 0$

Important GLMs

Random Component	Link	Systematic Component	Model	
Normal	Identity	Continuous	Regression	Part I
Normal	Identity	Categorical	Analysis of variance	
Normal	Identity	Mixed	Analysis of covariance	
Binomial	Logit	Mixed	Logistic regression	Part II
Multinomial	Generalized logit	Mixed	Logistic regression for multinomial response	
Poisson	Log	Categorical	Log linear	

Systematic component

For the GLM $\eta = g(\mu) = g[E(Y)] = \mathbf{X}\beta$:

- The data vector $\mathbf{y} = (y_1, \ldots, y_n)'$ and the vector $g(\mu)$ are points of the n-dimensional Euclidean space ($\mathfrak{R}^n$).
- For a particular model matrix $\mathbf{X}_{n \times p}$, the values of $\mathbf{X}\beta$ for all possible $\beta = (\beta_1, \ldots, \beta_p)'$ values generate a vector space that is a linear subspace of $\mathfrak{R}^n$.
  Column space generated by $\mathbf{X}$:

$$C(\mathbf{X}) = \{\eta : \text{ there is a } \beta \text{ such that } \eta = \mathbf{X}\beta\ \}$$

☞ Assumptions and properties of the systematic component of a LM apply.

II.2.2 Definition (Parameters' Identifiability; s. also I.4.16)

For a GLM with linear predictor $\mathbf{X}\beta$, with $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\beta \in \mathbb{R}^p$, β is identifiable if

$$\beta_1 \neq \beta_2 \ \Rightarrow\ \mathbf{X}\beta_1 \neq \mathbf{X}\beta_2\ ,$$

for any $\beta_1,\ \beta_2 \in \mathbb{R}^p$.

Distribution of the random component of a GLM

II.2.3 Definition (Exponential Dispersion Family)

Each component of the random response vector $\mathbf{Y} \in \mathbb{R}^n$ of a GLM has a distribution in the *exponential dispersion family* (EDF), with probability density function (pdf) or probability mass function (pmf) of the form

$$f(y;\ \vartheta, \phi) = \exp\left\{\frac{y\vartheta - b(\vartheta)}{a(\phi)} + c(y, \phi)\right\}, \tag{II.1}$$

for some specific functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$, giving rise to different distributions in the family. We call

- ϑ **natural** parameter [determines $\mu = E(Y)$]
- ϕ **dispersion** parameter [controls $\mathrm{Var}(Y)$]

II.2.4 Proposition (Mean and Variance of Y)

The expected value and variance of a random variable with EDF distribution are given by expressions of terms in (II.1). In particular it holds:

- $\mu = \mathsf{E}(Y) = b'(\vartheta)$
- $\mathsf{Var}(Y) = b''(\vartheta)a(\phi)$

Proof

$\mathsf{E}(Y)$ and $\mathsf{Var}(Y)$ are derived from well known results of the score function (see II.1.14 and II.1.10):

$$\mathsf{E}S(\vartheta) = \mathsf{E}\left(\frac{\partial \ell(\vartheta)}{\partial \vartheta}\right) = 0 \tag{II.2}$$

$$\mathsf{E}\left(\frac{\partial \ell(\vartheta)}{\partial \vartheta}\right)^2 = -\mathsf{E}\left(\frac{\partial^2 \ell(\vartheta)}{\partial \vartheta^2}\right) . \tag{II.3}$$

We have from (II.1) that

$$\ell(\vartheta|y) = \frac{y\vartheta - b(\vartheta)}{a(\phi)} + c(y, \phi)$$

and hence

Proof (continues)

$$\frac{\partial \ell(\vartheta)}{\partial \vartheta} = \frac{y - b'(\vartheta)}{a(\phi)} \quad \text{and} \quad \frac{\partial^2 \ell(\vartheta)}{\partial \vartheta^2} = \frac{-b''(\vartheta)}{a(\phi)} \; .$$

From (II.2) we have

$$E\left(\frac{\partial \ell(\vartheta)}{\partial \vartheta}\right) = 0 \;\Rightarrow\; \frac{\mu - b'(\vartheta)}{a(\phi)} = 0 \;\Rightarrow\; \mu = EY = b'(\vartheta) \; .$$

Furthermore it holds

$$-E\left(\frac{\partial^2 \ell(\vartheta)}{\partial \vartheta^2}\right) = -E\left(\frac{-b''(\vartheta)}{a(\phi)}\right) = \frac{b''(\vartheta)}{a(\phi)} \quad \text{and} \quad E\left(\frac{\partial \ell(\vartheta)}{\partial \vartheta}\right)^2 = E\left(\frac{Y - b'(\vartheta)}{a(\phi)}\right)^2 \; .$$

Thus from (II.3) we have

$$\frac{b''(\vartheta)}{a(\phi)} = \frac{E\left(Y - b'(\vartheta)\right)^2}{\left(a(\phi)\right)^2}$$

and since $\mu = b'(\vartheta)$, we finally have $b''(\vartheta) = \frac{\mathrm{Var}(Y)}{a(\phi)}$, which completes the proof.

II.2.5 Example (common univariate distributions in the EDF)

Members of the exponential dispersion family (among others):
Gaussian (or normal), gamma, inverse Gaussian, Poisson and binomial distributions

$$f(y;\ \theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

	θ	$b(\theta)$	$a(\phi)$	$c(y,\phi)$
$Y \sim \mathcal{N}(\mu,\sigma^2)$	μ	$\frac{\mu^2}{2} = \frac{\theta^2}{2}$	σ^2	$-\frac{\log(2\pi\sigma^2))}{2} - \frac{y^2}{2\sigma^2}$
$Y \sim \mathcal{P}(\mu)$	$\log(\mu)$	e^θ	1	$-\log(y!)$
$Y \sim \mathcal{B}(m,\pi)/m$	$\log(\frac{\pi}{1-\pi})$	$\log[1+e^\theta]$	$1/m$	$\log\binom{m}{my}$

Mean and Variance of Y:

	$E(Y) = b'(\theta)$	$Var(Y) = a(\phi)b''(\theta)$
$Y \sim \mathcal{N}(\mu_i,\sigma^2)$	$\mu = \theta$	σ^2
$Y \sim \mathcal{P}(\mu_i)$	$\mu = e^\theta$	$e^\theta = \mu$
$Y \sim \mathcal{B}(m,\pi)/m$	$\mu = \pi = \frac{e^\theta}{1+e^\theta}$	$\frac{1}{m}\frac{e^\theta}{[1+e^\theta]^2} = \frac{\pi(1-\pi)}{m}$

II.2.6 Remark (natural exponential family of distributions)

When the dispersion parameter ϕ is known, the EDF becomes an exponential family model with canonical parameter ϑ, called the *natural exponential family* (NEF). Random variables in this family have a probability density or mass function of the form

$$f(y;\vartheta) = h(y)\exp\left[y\vartheta - b(\vartheta)\right] .$$

☞ If ϕ is unknown, the EDF can or cannot be a two-parameter exponential family.

II.2.7 Definition (k-parameter exponential family)

In general, $\{P_\vartheta, \vartheta \in \Theta\}$, $\Theta \subset \mathbb{R}^k$ is a k-parameter exponential family if the probability density or mass function of $\mathbf{Y} \sim P_\vartheta$, with $\mathbf{Y} \in \mathbb{R}^n$ is

$$f(\mathbf{y};\ \boldsymbol{\vartheta}) = h(\mathbf{y})\exp\left[\sum_{j=1}^{k} \eta_j(\boldsymbol{\vartheta})T_j(\mathbf{y}) - B(\boldsymbol{\vartheta})\right] ,$$

where

$\eta_1,\ldots,\eta_k$ and B are real-valued functions mapping $\Theta \to \mathbb{R}$,

$T_1,\ldots,T_k$ and h are real-valued functions mapping $\mathbb{R}^n \to \mathbb{R}$.

II.2.8 Remark

For one-parameter families the (known) dispersion parameter ϕ is fixed. The function $a(\phi)$ is commonly of the form $a(\phi) = \phi/w$, where w (weight) may vary from observation to observation.

- For example, $\phi = 1$ for Poisson $\mathcal{P}(\mu)$ and binomial $\mathcal{B}(m, \pi)$, for fixed m.
- Furthermore, for the Poisson $w = 1$ while for the binomial $w = m$ when as response is considered the success proportion, i.e. $Y \in \{0, \frac{1}{m}, \dots, 1\}$.
  - If $\mathbf{Y} = (Y_1, \dots, Y_n)'$, where Y_i are stochastically independent success proportions with $Y_i \sim \mathcal{B}(m_i, \pi_i)/m_i$, then $w_i = m_i$ (observation dependent).