Generalized linear models

Jean-Michel Marin

University of Montpellier
Faculty of Sciences

HAX912X - 2023/2024

Why generalized linear models

In many applications, the response does not vary in all $\mathbb{R}$ but in $\mathbb{R}^+$, in $\mathbb{N}$, in $\{0, 1\}$...

The Gaussian model is not suited to this situation

Why generalized linear models

$y = (y_1, \dots, y_n)$ the vector of responses
X the matrix of explanatory variables

The distribution of y_i, $(\mathbb{P}_{\theta_i})_{\theta_i \in \mathbb{R}}$ must be specified
$\mathscr{P}(\theta_i)$, $\mathscr{E}(\theta_i)$, $\mathscr{B}(\theta_i)$, $\mathscr{N}(\theta_i, 1)$, ...

The link between θ_i and X must also be specified

Why generalized linear models

We assume that $\theta_i = \gamma(x_i\beta)$
$\gamma(\cdot)$ is called the link function

A GLM is fully specified by

- a probability family
- a link function

Gaussian linear model
$\mathbb{P}_\theta = \mathcal{N}(\theta, \sigma^2)$
$\gamma(x_i\beta) = x_i\beta$

Why generalized linear models

Examples

- Gaussian linear model
- Logistic regression model
- Poisson regression model

Scalar exponential families

Let $\nu(dx)$ be a reference measure on $\mathbb{R}$,

$$b(\theta) = \log\left(\int \exp(\theta y)\nu(dy)\right)$$

and

$$D_\nu = \{\theta | b(\theta) < \infty\} \subseteq \mathbb{R}$$

Scalar exponential families

Definition

A family of probability distribution $\mathbb{P}_\theta$ is said to belong to the scalar exponential family if

- for each element of the family there exist a $\theta \in D_\nu$ such that the probability distribution can be written in the form

$$\mathbb{P}_\theta(dx) = \exp(\theta x - b(\theta))\nu(dx)$$

- to any value of θ corresponds one and only one element of the family

θ is called the natural parameter of the exponential family
The exponential family is said to be regular if D_ν is open

Scalar exponential families

If θ is an interior point of D_ν then

$$b'(\theta) = \mathbb{E}_\theta(y)$$

$$b''(\theta) = \mathbb{V}_\theta(y)$$

Scalar exponential families

The function $b(\theta)$ is strictly convex

The strictly convex nature of $b(\theta)$ means that $b'(\theta)$ is bijective

We can also consider $\mu = \mathbb{E}_\theta(y)$ as a parameter

Scalar exponential families

Examples

- Poisson distribution with parameter $\lambda > 0$
- Binomial distribution with parameters (m, p) where m is fixed and $p \in]0, 1[$
- Gaussian distribution with parameters (μ, σ^2) where σ^2 is known and $\mu \in \mathbb{R}$

Scalar exponential families

Maximum likelihood estimation of θ

Let $y_1, \dots, y_n$ be an n-sample from $\mathbb{P}_{\theta^*}$

If $\mathbb{P}_\theta$ belongs to the scalar exponential family with θ as the natural parameter, then $\hat{\theta}_n$ the MLE of θ^* is such that

$$\frac{1}{n}\sum_{i=1}^{n} y_i = b'(\hat{\theta}_n)$$

Exponential families with a nuisance parameter

$$D_{\nu,\phi} = \left\{ \theta \,\middle|\, \int \exp\left[\frac{x\theta - b(\theta)}{\phi} + c(x,\phi) \right] \nu(dx) < \infty \right\}$$

Definition

A family of probability distribution $\mathbb{P}_{(\theta,\phi)}$ is said to belong to the exponential family with nuisance parameter ϕ if

- for each element of the family there exist a $\theta \in D_{\nu,\phi}$ and a $\phi \in \mathbb{R}^+$ such that the probability distribution can be written in the form

$$\mathbb{P}_{\theta,\phi}(dx) = \exp\left\{ \frac{x\theta - b(\theta)}{\phi} + c(x,\phi) \right\} \nu(dx)$$

- to any pair of $\theta \in D_{\nu,\phi}$ and $\phi \in \mathbb{R}^+$ corresponds one and only one element of the family

Exponential families with a nuisance parameter

We have

$$b'(\theta) = \mathbb{E}_\theta(y)$$

$$b''(\theta) = \frac{\mathbb{V}_\theta(y)}{\phi}$$

Exponential families with a nuisance parameter

Examples

- Gaussian distribution with parameters (μ, σ^2) where $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$
- Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$

Exponential families with a nuisance parameter

Maximum likelihood estimation of θ

Let $y_1, \dots, y_n$ be an n-sample from $f(y; \theta^*, \phi^*)\nu(dx)$

For any ϕ^*, $\hat{\theta}_n$ the MLE of θ^* is such that

$$\frac{1}{n}\sum_{i=1}^{n} y_i = b'(\hat{\theta}_n)$$

Definition of generalized linear models

Consider the n-sample $(x_i, y_i)_{i=1,\dots,n}$ from (x, y) where x is the vector of explanatory variables and y the corresponding response

Definition

Choosing a generalized linear model corresponds to choosing a conditional probability distribution for $y|x$. For the class of generalized linear model this conditional distribution is such that

- the distribution of $y|x$ belongs to an exponential family with a nuisance parameter
- $$\gamma(\mathbb{E}(y|x)) = x\beta$$

  $\gamma(\cdot)$ is called the link function

Definition of generalized linear models

1) Choosing the exponential family
determined in most cases by the values taken by y; if several choices are possible, the plots of the residuals can be used to decide which family is the most appropriate

2) Choice of link function:
we can use the canonical link: $\gamma(\cdot) = b'(\cdot)$
in this case we have $\theta = x\beta$
that is a natural and advantageous choice, many formulas are simplified

Classical examples

Logistic regression
$\mathbb{P}_\theta = \mathscr{B}(\theta)$
$\gamma(u) = \log(u/(1-u))$
$\mathbb{E}(y|x) = \exp(x\beta)/(1+\exp(x\beta))$

Poisson regression $\mathbb{P}_\theta = \mathscr{P}(\theta)$
$\gamma(u) = \log(u)$
$\mathbb{E}(y|x) = \exp(x\beta)$