Bayesian Analysis

in

the Generalized Linear Model

Group member:

马茹

陈文琳

Our goals

provide enough guidance so we can combine generalized linear models with the ideas of Bayesian analysis.

Overviews

1	Linear regression model
2	Introduction of GLM
3	Example

Linear regression model

Classical linear regression model

$$Y = X\beta + \varepsilon$$

Where we obtain the following variables

Y —— **endogenous variable**

X —— **exogenous variable**

β —— **the regression coefficient**

ε —— **the random error**

Linear regression model

A possible assumption of ε

$$\varepsilon \overset{iid}{\sim} N(0, \sigma^2)$$

We denote that $\mu = X\beta$

$$\Rightarrow \mu = E(Y \mid X)$$

Linear regression model

Several properties of linear regression model

- $E(Y) = \mu = X\beta$
- X, Y **are continuous variables;**
- $Y \sim N(\mu, \sigma^2)$

Linear regression model

shortage of linear regression model

- The assumption that Y have normal distribution is impractical
- Exogenous variable can only effect on endogenous variable through addition
- Endogenous variable must be continuous variable

Linear regression model

discontinuous characters

- Counted variables

  0-1variables $z \sim P(\lambda)$

- variables do no follow the normal distribution

  variables follow Gamma distribution ,binomial distribution

And so on

Introduction of GLM

Difference between linear model and GLM

1.Endogenous variables can follow any distribution in exponential family of distributions ;

2.We induct link function $\theta(\mu) = X\beta$ to measure the mean of endogenous

Introduction of GLM

Several properties of linear regression model

- x, y can be continuous or discrete variables
- $$f(y,\mu,\phi) = \exp\left[\frac{y(\mu) - b(\mu)}{a(\phi)} + c(y,\phi)\right]$$
- $$E(y) = \mu, \theta(\mu) = X\beta$$

Introduction of GLM

A generalized linear model is specified in three stages:

1. The linear predictor $\theta = \mathrm{X}\beta$
2. The link function g(·) that relates the linear predictor to the mean of the outcome variable: $\mu = \theta^{-1}(\mathrm{X}\beta)$
3. The random component specifying the distribution of the outcome variable y with mean E(y|X) = μ. The distribution can also depend on a dispersion parameter, φ.

Introduction of GLM

The exponential family

The probability density function of the exponential family

$$f(y|\theta,\Phi) = \exp\{\frac{y\theta - b(\theta)}{\Phi} + c(y,\Phi)\}$$

θ is a canonical parameter and a function of the mean of the outcome variable μ

$b(\theta)$ is a function of θ and has positive second order derivative

Φ is a dispersion parameter that plays a role in defining the variance of y

$c(y, \Phi)$ is a function of y and Φ

Introduction of GLM

Link function , expectation and variance of y

We can use normal distribution $f(y|\mu,\sigma^2)$ to find connections with the exponential family.

$$\mu = \theta, \qquad b(\theta) = \frac{\theta^2}{2}, \qquad \Phi = \sigma^2$$

More generally, we can set θ as linear predictor and get link function:

$$\theta(\mu) = X\beta$$

Analogously, we can get the expectation and variance of y:

$$E(y) = b'(\theta), \qquad Var(y) = b''(\theta)\Phi$$

Introduction of GLM

Likelihoods

The mean of the distribution of y, given X, is determined by Xβ: E(y | X) = θ^{-1}(Xβ).We use the same notation as in linear regression whenever possible, so that X is the n × p matrix of explanatory variables and $\theta = X\beta$ is the vector of n linear predictor values.

If we denote the linear predictor for the ith case by $X_i\beta$ and the variance or dispersion parameter(if present) by Φ, then the data distribution takes the form

$$p(y|X,\beta,\Phi) = \prod_{i=1}^{n} p(y_i|X_i\beta,\Phi)$$

Introduction of GLM

Standard GLM likelihoods

Normal distribution

Normal distribution has an identity link function $\theta(\mu) = \mu$.

With a assumption $\varepsilon \sim N(0, \sigma^2)$,we can learn that linear regression model $y = X\beta + \varepsilon$ is a special case of the generalized linear model, for

$$y|\beta, X, \sigma^2 \sim \mathrm{N}(X\beta, \sigma^2).$$

Introduction of GLM

Standard GLM likelihoods

Poisson distribution

Counted data are often modeled using a Poisson model. The Poisson generalized linear model, often called the Poisson regression model, assumes that y is Poisson with mean μ (and therefore variance μ).

The link function is typically chosen to be the logarithm, so that log μ = Xβ. The distribution for data y = $(y_1, \dots\dots, y_n)$ is thus

$$p(y|\beta) = \prod_{i=1}^{n} \frac{1}{y_i!} e^{-\exp(\theta_i)} (\exp(\theta_i))^{y_i}$$

where θ_i = Xiβ is the linear predictor for the i-th case.

Introduction of GLM

Standard GLM likelihoods

Binomial distribution

Suppose that $y_i \sim \mathrm{Bin}(n_i, \mu_i)$ with n_i known. It is common to specify the model in terms of the mean of the proportions $y_i\,/n_i$, rather than the mean of y_i. Choosing the logit transformation of the probability of success, $\theta(\mu_i) = \log(\mu_i/(1-\mu_i))$, as the link function leads to the logistic regression model.

The distribution for data y is

$$p(y|\beta) = \prod_{i=1}^{n} \binom{n_i}{y_i} \left(\frac{e^{\theta_i}}{1+e^{\theta_i}}\right)^{y_i} \left(\frac{1}{1+e^{\theta_i}}\right)^{n_i - y_i}$$

Introduction of GLM

The combination of GLM and Bayesian Analysis

GLM:
nonlinear model;
Discrete data
…

Bayesian Analysis:
Small sample size
Too many parameters
…

example

reserve for outstanding losses

To fail to draw or carry down various kinds of liability reserves or fail to draw reserve for outstanding losses according to the provisions of this law. In general, we take the unbiased estimator of expected value of outstanding losses as the reserve for outstanding losses.

example

Triangular flow

事故年	进展年(j)					
(i)	1	2	3	4	…	n
1	C_{11}	C_{12}	C_{13}	C_{14}	…	$C_{1,n}$
2	C_{21}	C_{22}	C_{23}	C_{24}	…	
3	C_{31}	C_{32}	C_{33}	…		
…	…	…	…			
n	C_{n1}					

example

We define the following variables

x_i the aggregate amount of outstanding losses in the ith year

y_j the probability of outstanding losses in the jth year

c_{ij} the expected amount of outstanding losses in the jth year when one insured in the ith year

example

Over-dispersed Poisson model

$$z \sim p(\lambda)$$

$$x = \varphi z, (\varphi \geq 1)$$

The prior distribution

$$x_j \sim Gamma(\alpha_i, \beta_i)$$

example

Then we obtain that

$$f(x,y \mid c_{ij}, i=1,\cdots,n, j=1,\cdots,n-i+1,\varphi) \propto \prod_{i=1}^{n}\prod_{j=1}^{n-i+1} f(c_{ij} \mid x,y,\varphi)\prod_{i=1}^{n} f\left(x_i\right) f\left(y_i\right)$$

Therefore, we can measure the expected amount of outstanding losses in the jth year when one insured in the ith year

$$\hat{c}_{ij} = E(c_{ij}) = \exp(\hat{\eta}_{ij})$$

We can measure the reserve for outstanding losses

$$R = \sum c_{ij}$$

Thank you!