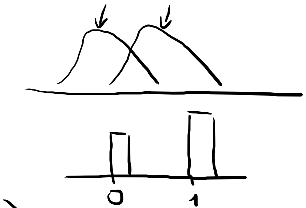
Y response , $X_1, \dots, X_p$ predictors 8/3/21



① LINEAR REGRESSION $\quad Y \mid X = \underline{x} \sim N(\underline{x}^t \underline{\beta}, \sigma^2)$

② LOGISTIC REGRESSION $\quad Y \mid X = \underline{x} \sim \text{Bernoulli}\left(\dfrac{e^{\underline{x}^t \underline{\beta}}}{1 + e^{\underline{x}^t \underline{\beta}}}\right)$

$$\eta_i = \underline{x}_i^t \underline{\beta} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \qquad \underline{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} \qquad \underline{x}_i = \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$$

LINEAR PREDICTOR

LM

① $E[Y \mid X = \underline{x}_i] = \eta_i$ deterministic linear

② $Var[Y \mid X = \underline{x}_i] = \sigma^2$ (homoskedasticity)

③ $Y \mid X = \underline{x}_i \sim N(\cdot, \cdot)$ Gaussian

stochastic

GLM

① $\mu_i = E[Y \mid X = \underline{x}_i]$
$\eta_i = g(\mu_i)$ with g a <u>link function</u> (one-to-one, differentiable)

② $Var(Y \mid X = \underline{x}_i) = \underbrace{V(\mu_i)}_{\text{variance function}} \cdot a_i(\phi)$ (heteroskedasticity)

$\phi \to$ dispersion parameter

③ $Y \mid X = \underline{x}_i \sim \underline{f}$ with f is a member of the Exponential Dispersion Family

EXAMPLE : Bernoulli $\quad y = \begin{cases} 1 & p \\ 0 & 1-p \end{cases}$

$Y \mid X = \underline{x}_i = Y_i$

CONDITIONAL DISTRIBUTION $\quad Y_i \sim \text{Bernoulli}(\mu_i) \quad , \quad i = 1, \dots, n$

CONDITIONAL MEAN $\quad \mu_i = E[Y_i] = P(Y_i = 1) \in [0,1]$

CONDITIONAL VARIANCE $\quad Var(Y_i) = \mu_i(1-\mu_i)$

- Not constant
- Depends on μ_i
- There is no dispersion parameter. We can say that $\phi = 1$ (known)

LINEAR PREDICTOR $\quad \eta_i = \underline{x}_i^t \beta$ which we want to link to μ_i

$\eta_i \in (-\infty, \infty)$

We need to find a link function :

$$\eta_i = g(\mu_i)$$

LINK FUNCTION g.

$\eta_i = g(\mu_i)$

Any cdf could do for $\mu_i = F(\eta_i)$ with $F : (-\infty, \infty) \to [0,1]$

Consider a cdf F of an absolutely continous random variable.

Then

$\eta_i = F^{-1}(\mu_i)$

a choice of g

F^{-1} is also called the quantile function

$F^{-1} : [0,1] \to (-\infty, \infty)$

Two popular options :

① $F = \Phi$ cdf of $N(0,1)$

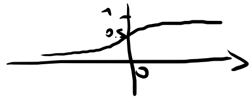
$\eta_i = \Phi^{-1}(\mu_i)$ PROBIT REGRESSION MODEL

② LOGISTIC DISTRIBUTION

density $\quad f(y;\mu,\sigma) = \dfrac{\exp((y-\mu)/\sigma)}{\sigma^2\left(1+\exp\left(\frac{y-\mu}{\sigma}\right)\right)^2} \qquad -\infty < y < +\infty$

cdf of standard logistic ($\mu=0, \sigma=1$) $\quad F(y;0,1) = \dfrac{e^y}{1+e^y} \qquad -\infty < y < +\infty$

SIGMOID FUNCTION



In the context of glm:

$$\mu = F(\eta) = \frac{e^\eta}{1+e^\eta} \implies \mu(1+e^\eta) = e^\eta$$

$$\mu = e^\eta(1-\mu)$$

$$\implies e^\eta = \frac{\mu}{1-\mu}$$

$$\implies \eta = \log\left(\frac{\mu}{1-\mu}\right) = g(\mu)$$

LOGIT LINK FUNCTION $\rightarrow$ LOGISTIC REGRESSION

③ There is a third common link function, which can be justified as follows:

Let $z_i \sim \text{Poisson}(\lambda_i)$ eg z_i number of bacteria on a plate (DILUTION BIOASSAY)

Assume that the instrument for counting olny observes

$$y_i = \begin{cases} 0 & \text{if } z_i = 0 \text{ (absent)} \\ 1 & \text{if } z_i > 0 \text{ (present)} \end{cases}$$

Dilute bacteria at different stages / time points, so there is a "covariate"

$x = 0, 1, 2, \dots$

We look for a link function $\eta_i = g(\mu_i)$ where $\mu_i = E[Y \mid X = x_i]$.

We do this indirectly by linking η_i to λ_i and λ_i to μ_i.

A natural link function between η_i and λ_i is the log:

$$\underbrace{\eta_i}_{(-\infty,\infty)} = \log(\underbrace{\lambda_i}_{(0,\infty)})$$

$$\mu_i = E[Y_i] = P(Y_i = 1) = 1 - P(Y_i = 0) = 1 - P(Z_i = 0) \overset{\text{Z Poisson}}{=} 1 - e^{-\lambda_i}$$

$$\Rightarrow \mu_i = 1 - e^{-\lambda_i} \Rightarrow e^{-\lambda_i} = 1 - \mu_i = -\lambda_i = \log(1-\mu_i) \Rightarrow \lambda_i = -\log(1-\mu_i)$$

$$\Rightarrow \eta_i = \underbrace{\log(-\log(1-\mu_i))}_{g(\mu_i)}$$

COMPLEMENTARY LOG-LOG LINK

EXPONENTIAL DISPERSION FAMILY

Def $y_i \sim ED(\mu_i, \phi, w_i)$

if

EXPECTATION $E[y_i] = \mu_i$

canonical parameter

DENSITY $$f_{y_i}(y_i; \theta_i, \phi) = \exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i; \phi) \right\}$$

dispersion parameter

for some functions a_i, b, c and parameters θ_i.

In most cases, we will assume $a_i(\phi) = \frac{\phi}{w_i}$ with known weights $w_i > 0$.

① NORMAL DISTRIBUTION is a member of the EDF

Density

$$f(y_i; \mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_i-\mu_i)^2}{\sigma^2}\right\}$$

$$= \exp\left\{-\frac{1}{2}\frac{(y_i-\mu_i)^2}{\sigma^2} + \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)\right\}$$

$$= \exp\left\{-\frac{1}{2}\frac{y_i^2-2y_i\mu_i+\mu_i^2}{\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\}$$

$$= \exp\left\{\frac{y_i\mu_i - \mu_i^2/2}{\sigma^2} \underbrace{- \frac{y_i^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)}\right\}$$

Map this formulation to the general form of an EDF:

$$\theta_i = \mu_i, \quad b(\theta_i) = \frac{\theta_i^2}{2}, \quad \phi = \sigma^2, \quad a_i(\phi) = \phi,$$

$$c(y_i; \phi) = -\frac{y_i^2}{2\phi} - \frac{1}{2}\log(2\pi\phi)$$

② BINOMIAL DISTRIBUTION is a member of EDF

$$f_{Y_i}(y_i; \pi_i, n_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{n_i - y_i}$$

$$= \exp\left\{ \log\binom{n_i}{y_i} + y_i \log(\pi_i) + (n_i - y_i)\log(1-\pi_i) \right\}$$

$$= \exp\left\{ y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log\binom{n_i}{y_i} + n_i \log(1-\pi_i) \right\}$$

$$= \exp\left\{ y_i \underbrace{\log\left(\frac{\pi_i}{1-\pi_i}\right)}_{\theta_i} \underbrace{+\, n_i \log(1-\pi_i)}_{-b(\theta_i)} + \underbrace{\log\binom{n_i}{y_i}}_{c(y_i;\, \phi)} \right\}$$

$$\underline{\theta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right)}$$

$$\downarrow$$

$$e^{\theta_i} = \frac{\pi_i}{1-\pi_i}$$

$$\Rightarrow (1-\pi_i) e^{\theta_i} = \pi_i$$

$$\Rightarrow e^{\theta_i} = \pi_i (1 + e^{\theta_i})$$

$$\Rightarrow \pi_i = \frac{e^{\theta_i}}{1+e^{\theta_i}} \Rightarrow 1 - \pi_i = 1 - \frac{e^{\theta_i}}{1+e^{\theta_i}} = \frac{1}{1+e^{\theta_i}}$$

$$b(\theta_i) = -n_i \log(1-\pi_i)$$

$$= \underline{n_i \log(1+e^{\theta_i})}$$

$$\phi = 1$$

$$a_i(\phi) = \phi$$

$$c(y_i; \phi) = \log\binom{n_i}{y_i}$$

LOG-LIKELIHOOD INFERENCE

Suppose we have a dataset $(\underline{x}_i, y_i)$, $i = 1, \dots, n$.

$Y \mid X = \underline{x}$ belongs to EDF with parameters θ_i, ϕ

Then log-likelihood is given by

$$\ell(\underline{\theta}; y_i, \underline{x}_i) = \sum_{i=1}^{n} \log f_{Y_i}(y_i; \theta_i, \phi)$$

$$= \sum_{i=1}^{n} \left[\frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i; \phi) \right]$$

Note:

- θ_i is a function of μ_i and $\overset{x_i^t \beta}{\overset{\shortparallel}{\eta_i}} = g(\mu_i)$
- So $\ell(\underline{\theta})$ is also a function of $\underline{\beta}$
- Estimates of β are found by taking derivatives of $\ell(\underline{\theta})$ wrt $\underline{\beta}$

CHOICE OF LINK FUNCTION

Consider $\dfrac{\partial \ell(\theta)}{\partial \theta_i} = \dfrac{y_i - b'(\theta_i)}{a_i(\phi)}$ with $b'(\theta_i) = \dfrac{d\, b(\theta_i)}{d\theta_i}$

For "nicely" behaved distributions:

$$E_{y_i}\left[\frac{\partial \ell(\theta)}{\partial \theta_i}\right] = 0$$

$$\shortparallel$$

$$E_{y_i}\left[\frac{y_i - b'(\theta_i)}{a_i(\phi)}\right] = \frac{E[y_i] - b'(\theta_i)}{a_i(\phi)}$$

$\Rightarrow E[y_i] = b'(\theta_i)$

$\shortparallel$

μ_i

EX :

① Gaussian: $b(\theta_i) = \dfrac{\theta_i^2}{2} \Rightarrow b'(\theta_i) = \theta_i = \mu_i$

② Binomial $b(\theta_i) = n_i \log(1 + e^{\theta_i}) \Rightarrow b'(\theta_i) = \dfrac{n_i e^{\theta_i}}{1 + e^{\theta_i}} = n_i \pi_i = \mu_i$

CANONICAL LINK FUNCTION

$$\eta_i \xrightarrow{g} \mu_i \xrightarrow{b'} \theta_i$$

The link function is canonical if $\eta_i = \theta_i$

$\Rightarrow g = (b')^{-1} \qquad \eta_i = g(\mu_i)$

VARIANCE FUNCTION

Bartlet identity

$$\underbrace{\mathrm{Var}_{y_i}\left(\frac{\partial \ell}{\partial \theta_i}\right)}_{\frac{\mathrm{var}(y_i)}{a_i(\phi)^2}} = E\left[\left(\frac{\partial \ell}{\partial \theta_i}\right)^2\right] = -\underbrace{E\left[\frac{\partial^2 \ell}{\partial \theta_i^2}\right]}_{\frac{b''(\theta_i)}{a_i(\phi)}}$$

$$\Rightarrow \mathrm{Var}(y_i) = a_i(\phi)\underbrace{b''(\theta_i)}_{\text{function of } \mu_i}$$

$V(\mu_i)$ variance function

→ Heteroskedasticity is naturally embedded in generalised linear models.