北京航空航天大學
经济管理学院
BEIHANG UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT

Generalized Linear Models

Lecture 3: Binary response

Outline

Logistic regression

Suppose response variable Y_i takes values 0 or 1 with probability $P(Y_i = 1) = p_i$. (i.e., a Bernoulli distribution)

We relate p_i to the predictors:

$$\eta_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_q x_{i,q}$$

$$\eta_i = g(p_i)$$

- The link function g should be monotone.
- $0 \leq g^{-1}(\eta) \leq 1$ for any η.

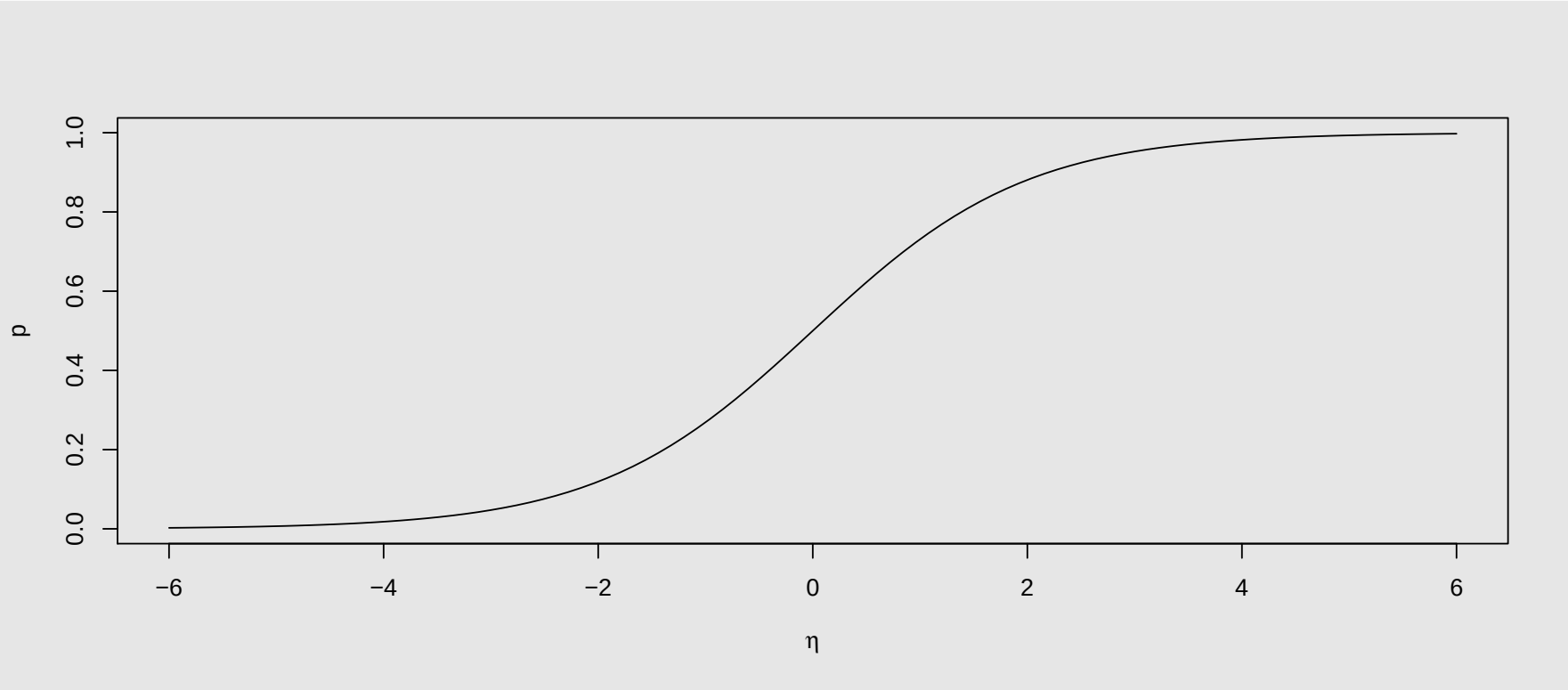
The choice of link function

logit link function

- $\eta = g(p) = \log(\frac{p}{1-p})$ is the *logit* function. It maps $(0, 1) \to \mathbb{R}$.
- The inverse logit is $g^{-1}(\eta) = \frac{e^{\eta}}{1+e^{\eta}}$ which maps $\mathbb{R} \to (0, 1)$.
- $p_i = \frac{e^{\eta_i}}{1+e^{\eta_i}} = P(Y_i = 1)$.

Inverse logit function

```
curve(ilogit(x), -6, 6, xlab=expression(eta), ylab='p')
```



- If $p_i \approx 0.5$, logistic and linear regression similar.

Log-likelihood

$$\eta_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_q x_{i,q}$$

$$\begin{aligned}
\log L(\beta) &= \sum_{i=1}^{n} \log(p_i) 1_{Y_i=1} + \log(1 - p_i) 1_{Y_i=0} \\
&= \sum_{i=1}^{n} y_i[\eta_i - \log(1 + e^{\eta_i})] + (1 - y_i) \log\left(1 - \frac{e^{\eta_i}}{1 + e^{\eta_i}}\right) \\
&= \sum_{i=1}^{n} y_i[\eta_i - \log(1 + e^{\eta_i})] + (1 - y_i) \log\left(\frac{1}{1 + e^{\eta_i}}\right) \\
&= \sum_{i=1}^{n} y_i[\eta_i - \log(1 + e^{\eta_i})] - (1 - y_i) \log(1 + e^{\eta_i}) \\
&= \sum_{i=1}^{n} [y_i \eta_i - \log(1 + e^{\eta_i})]
\end{aligned}$$

Log-likelihood

$$\eta_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_q x_{i,q}$$

$$\begin{aligned}
\log L(\beta) &= \sum_{i=1}^{n} \log(p_i) 1_{Y_i=1} + \log(1-p_i) 1_{Y_i=0} \\
&= \sum_{i=1}^{n} y_i[\eta_i - \log(1+e^{\eta_i})] + (1-y_i) \log\left(1 - \frac{e^{\eta_i}}{1+e^{\eta_i}}\right) \\
&= \sum_{i=1}^{n} y_i[\eta_i - \log(1+e^{\eta_i})] + (1-y_i) \log\left(\frac{1}{1+e^{\eta_i}}\right) \\
&= \sum_{i=1}^{n} y_i[\eta_i - \log(1+e^{\eta_i})] - (1-y_i) \log(1+e^{\eta_i}) \\
&= \sum_{i=1}^{n} [y_i \eta_i - \log(1+e^{\eta_i})]
\end{aligned}$$

- Uses MLE to obtain $\hat{\beta}$
- See Chapter 8 for estimation details.

Logistic regression in R

```
fit <- glm(y ~ x1 + x2, family=binomial,
           data=df)
```

- Bernoulli is equivalent to Binomial with only two levels.
- First (alphabetical) level is set to 0, other to 1. Use `relevel` if you want to change it.
- `glm` uses MLE with a logit link function (when `family=binomial`).

Interpreting a logistic regression

Odds = $p/(1-p)$.

- Odds are unbounded.
- log(odds) = log($p/(1-p)$) = $\beta_0 + \beta_1 x_1 + \beta_2 x_2$
- A unit increase in x_1 with x_2 held fixed increases log-odds of success by β_1.
- A unit increase in x_1 with x_2 held fixed increases odds of success by a factor of e^{β_1}.

Outline

Inference

- The likelihood ratio statisic: $2 \log \frac{L_L}{L_S}$, where L_L and L_S are likelihood of a larger model with l parameters and a smaller model with s parameters.

Inference

- The likelihood ratio statisic: $2 \log \frac{L_L}{L_S}$, where L_L and L_S are likelihood of a larger model with l parameters and a smaller model with s parameters.
- Now suppose we choose a saturated larger model, which has as many parameters as cases and has fitted values $\tilde{p}_i = y_i$. Then the statistic becomes:

$$
\begin{aligned}
D &= 2 \log L_L - 2 \log L_S \\
&= -2 \log L_S \\
&= -2 \sum_{i=1}^{n} \tilde{p}_i \text{logit}(\hat{p}_i) + \log(1 - \hat{p}_i)
\end{aligned}
$$

Deviance

D is called the Deviance.

- Difference between deviances equivalent to a likelihood ratio test.
- $D_S - D_L \sim \chi^2_{l-s}$, where l and s are the number of parameters, assuming
  1. smaller model is correct
  2. models are nested
  3. distributional assumptions true
- Null deviance is for model with only an intercept.

Deviance test

In R:

```
fit <- glm(y ~ x1 + x2, family='binomial',
           data=df)
anova(fit, test="Chisq")
drop1(fit, test="Chisq")
anova(fit1, fit2, test="Chisq")
```

- NOT equivalent to t-tests on coefficients
- Deviance tests preferred

Confidence intervals for coefficients

- Constructed using normal approximations for the parameter estimates.
- A $100(1-\alpha)\%$ confidence interval for β_i would be :

$$\hat{\beta}_i \pm z^{\alpha/2} se(\hat{\beta}_i).$$

- Implemented in R using `confint`.

Outline

OLS Residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{y}_i$$

OLS Residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{y}_i$$

Pearson residuals: Standardized

$$r_i = e_i / \hat{\sigma}$$

OLS Residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{y}_i$$

Pearson residuals: Standardized

$$r_i = e_i / \hat{\sigma}$$

- Mean 0, variance 1.

OLS Residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{y}_i$$

Pearson residuals: Standardized

$$r_i = e_i / \hat{\sigma}$$

- Mean 0, variance 1.

Deviance residuals: Signed root contribution to $-2 \log L$.

$$-2 \log L = c + \frac{1}{\hat{\sigma}^2} \sum e_i^2 = c + \sum d_i^2$$

$$d_i = e_i / \hat{\sigma}$$

Logistic regression residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{p}_i$$

Logistic regression residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{p}_i$$

Pearson residuals: Standardized

$$r_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)}}$$

Logistic regression residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{p}_i$$

Pearson residuals: Standardized

$$r_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)}}$$

- Mean 0, variance 1.

Logistic regression residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{p}_i$$

Pearson residuals: Standardized

$$r_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)}}$$

- Mean 0, variance 1.

Deviance residuals: Signed root contribution to $-2 \log L$.

$$-2 \log L = -2 \sum [\log(\hat{p}_i) y_i + \log(1 - \hat{p}_i)(1 - y_i)]$$

$$d_i = \text{sign}(y_i - \hat{p}_i) \sqrt{-2 \left[y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i) \right]}$$

Logistic regression residuals

In R:

```
fit <- glm(y ~ x1 + x2, family='binomial',
           data=df)
e <- residuals(fit, type='response')
r <- residuals(fit, type='pearson')
d <- residuals(fit, type='deviance')
```

- `deviance` is the default
- Residual plots can be hard to interpret
- Don’t expect residuals to be normally distributed

Outline

Model selection

Akaike's Information Criterion

$$\text{AIC} = -2 \log L + 2q$$

or

$$\text{AIC} = \text{deviance} + 2q$$

- Select model with smallest AIC
- Beware that model selection is not magical