Analysis of Categorical Data
Chapter 7: Alternative Modeling of Binary Response Data

Shaobo Jin

Department of Mathematics

Intended Learning Outcome

Through this chapter, you should be able to

1. fit binary data models other than the logit model,
2. describe conditional maximum inference,
3. apply conditional maximum likelihood.

Canonical Link

If the response variable Y_i belongs to exponential dispersion family, its pmf/pdf is of the form

$$f\left(y_i;\theta_i,\phi_i\right)=\exp\left\{\frac{y_i\theta_i-b\left(\theta_i\right)}{\phi_i}+c\left(y_i,\phi_i\right)\right\},$$

where θ_i is the natural parameter. The link function of a GLM transforms the mean μ_i to the linear predictor $\eta_i=g\left(\mu_i\right)$.
The canonical link function transforms the mean μ_i to the natural parameter θ_i. Hence,

$$\begin{aligned}
\theta_i &= g\left(\mu_i\right)=\boldsymbol{x}_i^T\boldsymbol{\beta}, && \text{canonical link,}\\
\theta_i &\neq g\left(\mu_i\right)=\boldsymbol{x}_i^T\boldsymbol{\beta}, && \text{otherwise.}
\end{aligned}$$

Logit Link As Canonical Link

Let $Z_i \sim \text{Bin}\left(n_i, \pi_i\right)$ and $Y_i = Z_i / n_i$. The pmf of Y_i is

$$
\begin{aligned}
P\left(Y_i = y_i\right) &= \exp\left\{\frac{y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log\left(1-\pi_i\right)}{1/n_i} + \log\begin{pmatrix} n_i \\ n_i y_i \end{pmatrix}\right\} \\
&= \exp\left\{\frac{y_i\theta_i - \log\left[1+\exp\left(\theta_i\right)\right]}{1/n_i} + \log\begin{pmatrix} n_i \\ n_i y_i \end{pmatrix}\right\},
\end{aligned}
$$

where $\theta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ is the natural parameter.
The canonical link satisfies

$$
\theta_i \quad = \quad g\left(\pi_i\right) = \boldsymbol{x}_i^T \boldsymbol{\beta}.
$$

Hence, the logit link is the canonical link.

Behind Link Function: Distribution Assumption!

- Suppose that, in an ideal world, we could observe continuous y_i^* and we could use the linear model

$$y_i^* = \boldsymbol{x}_i^T \boldsymbol{\beta} - \varepsilon_i.$$

- However, in reality, we only observe y_i such that

$$y_i = \begin{cases} 0, & \text{if } y_i^* < 0, \\ 1, & \text{if } y_i^* \geq 0. \end{cases}$$

- In such a case, we often assume that $Y_i \sim \text{Bernoulli}(\pi_i)$.
- Note that

$$\pi_i = P(Y_i = 1) = P(Y_i^* \geq 0) = P\left(\varepsilon_i \leq \boldsymbol{x}_i^T \boldsymbol{\beta}\right) = F_{\varepsilon_i}\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right).$$

- The link function corresponds to the distribution assumption that we put on ε_i.

Link Functions For Binary Data

- logit link (logistic model) is inverse function of logistic cdf

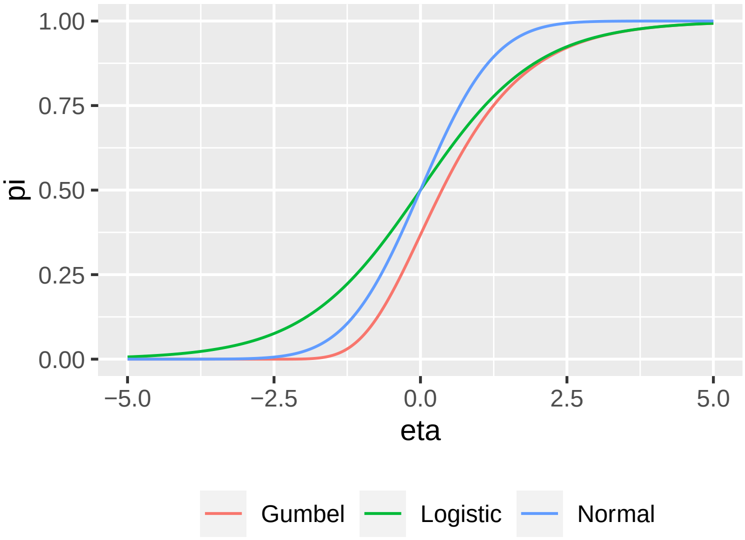
$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right).$$

- Probit link (probit model) is the inverse function of the standard normal cdf

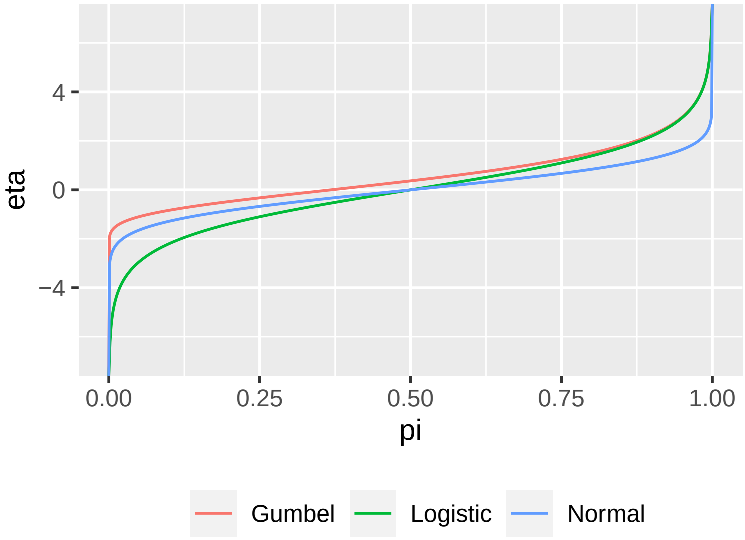
$$g(\pi_i) = \Phi^{-1}(\pi_i).$$

- Identity link (linear probability model) is the inverse function of the uniform distribution cdf $g(\pi_i) = \pi_i$.
- Log-log link is the inverse function of the Gumbel distribution cdf $g(\pi_i) = -\log[-\log(\pi_i)]$.
    - Complementary log-log link: $g(\pi_i) = \log[-\log(1-\pi_i)]$.

Different Distribution Functions



Different Link Functions



Symmetric Link Functions

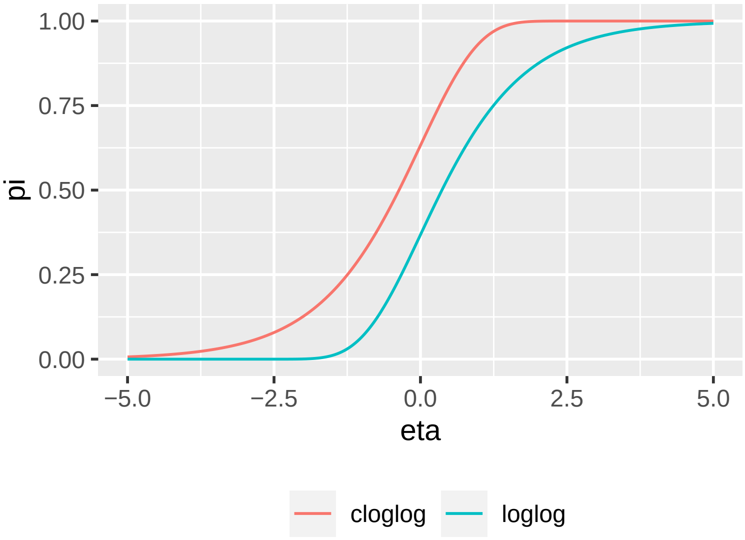
A link function is symmetric about 0.5 if

$$g(\pi) = -g(1-\pi).$$

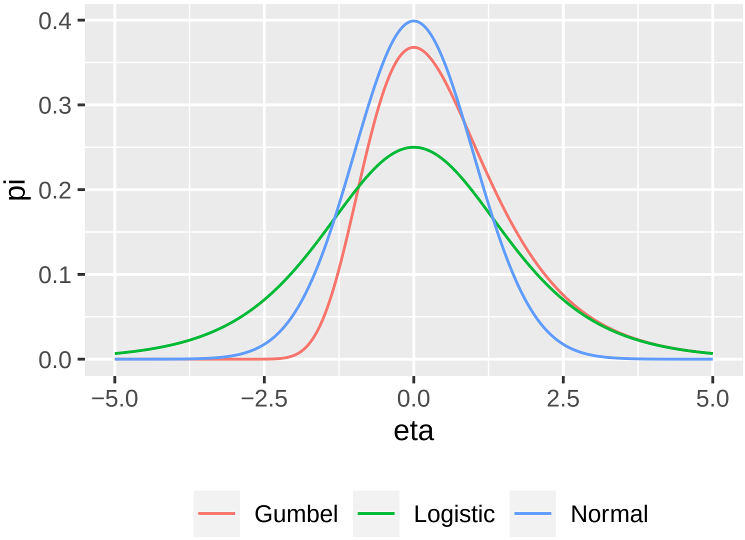
It means that the response curve when π approaches 0 has a similar appearance to the response curve when π approaches 1.
The log-log link and the clog-log link are not symmetric.

- For the clog-log link, π approaches 0 slowly, but approaches 1 quickly.
- For the log-log link, π approaches 1 slowly, but approaches 0 quickly.

Different Link Functions



Different Density Functions



Example: Different Link Functions

Number of beetles killed after exposure to gaseous carbon disulfide

	Response	
Dose	Killed	Not Killed
1.6907	6	53
1.7242	13	47
1.7552	18	44
1.7842	28	28
1.8113	52	11
1.8369	53	6
1.8610	61	1
1.8839	60	0

Likelihood

The model

$$\operatorname{logit} P\left(Y_i = 1\right) = \alpha + \sum_{j=1}^{p} \beta_j x_{ij}$$

is equivalent to

$$P\left(Y_i = y_i\right) = \frac{\exp\left[y_i\left(\alpha + \sum_{j=1}^{p} \beta_j x_{ij}\right)\right]}{1 + \exp\left(\alpha + \sum_{j=1}^{p} \beta_j x_{ij}\right)}.$$

For N independent observations, the likelihood is

$$P\left(Y_1 = y_1, \cdots, Y_n = y_n\right) = \frac{\exp\left[\left(\sum_{i=1}^{N} y_i\right)\alpha + \sum_{j=1}^{p}\left(\sum_{i=1}^{N} y_i x_{ij}\right)\beta_j\right]}{\prod_{i=1}^{N}\left[1 + \exp\left(\alpha + \sum_{j=1}^{p} \beta_j x_{ij}\right)\right]}.$$

Sufficient Statistics

By the factorization theorem of sufficient statistics,

$$P\left(Y_1 = y_1, \cdots, Y_n = y_n\right) = \frac{\exp\left[\left(\sum_{i=1}^N y_i\right)\alpha + \sum_{j=1}^p \left(\sum_{i=1}^N y_i x_{ij}\right)\beta_j\right]}{\prod_{i=1}^N \left[1 + \exp\left(\alpha + \sum_{j=1}^p \beta_j x_{ij}\right)\right]}$$

implies that $\left(\sum_{i=1}^N y_i, \sum_{i=1}^N y_i x_{i1}, \cdots, \sum_{i=1}^N y_i x_{ip}\right)$ is a sufficient statistic for $(\alpha, \beta_1, \cdots, \beta_p)$. In fact,

$$\frac{P\left(Y_1 = y_1, \cdots, Y_n = y_n\right)}{P\left(Y_1 = y_1', \cdots, Y_n = y_n'\right)} = \frac{\exp\left[\left(\sum_{i=1}^N y_i\right)\alpha + \sum_{j=1}^p \left(\sum_{i=1}^N y_i x_{ij}\right)\beta_j\right]}{\exp\left[\left(\sum_{i=1}^N y_i'\right)\alpha + \sum_{j=1}^p \left(\sum_{i=1}^N y_i' x_{ij}\right)\beta_j\right]}$$

does not depend on the parameters if and only if $\sum_{i=1}^N y_i = \sum_{i=1}^N y_i'$ and $\sum_{i=1}^N y_i x_{ij} = \sum_{i=1}^N y_i' x_{ij}$ for all j. Hence, they are also minimal sufficient.

Nuisance Parameter

Suppose that β_1 is the focus parameter and all others are nuisance parameters. Let

$$S = \left\{ (y_1^*, \cdots, y_N^*) : \sum_{i=1}^{N} y_i^* = t_0, \sum_{i=1}^{N} y_i^* x_{ij} = t_j, j = 2, .., p \right\}.$$

Then,

$$\begin{aligned} & P\left(Y_1 = y_1, \cdots, Y_n = y_n \mid \sum_{i=1}^{N} y_i = t_0, \sum_{i=1}^{N} y_i x_{ij} = t_j, j = 2, .., p\right) \\ = \; & \frac{P\left(Y_1 = y_1, \cdots, Y_n = y_n, \sum_{i=1}^{N} y_i = t_0, \sum_{i=1}^{N} y_i x_{ij} = t_j, j = 2, .., p\right)}{P\left(\sum_{i=1}^{N} y_i = t_0, \sum_{i=1}^{N} y_i x_{ij} = t_j, j = 2, .., p\right)} \\ = \; & \frac{\exp\left[\left(\sum_{i=1}^{N} y_i x_{i1}\right) \beta_1\right]}{\sum_S \exp\left[\left(\sum_{i=1}^{N} y_i^* x_{i1}\right) \beta_1\right]}, \end{aligned}$$

which depends only on β_1.

Conditional ML Estimator

The conditional ML estimator of β_1 maximizes the conditional likelihood

$$P\left(Y_1 = y_1, \cdots, Y_n = y_n \mid \sum_{i=1}^{N} y_i = t_0, \sum_{i=1}^{N} y_i x_{ij} = t_j, j = 2, .., p\right).$$

When the sample size n is not large enough and the number of nuisance parameters is large, ML for logistic regression may not perform so well. The conditional maximum likelihood tends to perform better.

Conditional Inference for 2×2 Tables

Consider the 2×2 table with independent binomial sampling

	Y		
X	1	0	Total
1	t	$n_1 - t$	n_1
0	s	$n_2 - s$	n_2

Suppose that

$$\operatorname{logit} P\left(Y_i = 1\right) = \alpha + \beta x_i,$$

where $x_1 = 1$ and $x_2 = 0$. To eliminate α, we conditional on its sufficient statistic $\sum_{i=1}^{N} y_i = s + t$. Then,

$$P\left(t \mid t + s, n_1, n_2\right) = \frac{\binom{n_1}{t}\binom{n_2}{s}\exp\left\{\beta t\right\}}{\sum_u \binom{n_1}{u}\binom{n_2}{s+t-u}\exp\left\{\beta u\right\}}.$$

If $\beta = 0$, we obtain the Fisher's exact test for 2×2 tables.

Conditional Inference for $2 \times 2 \times K$ Tables

In a $2 \times 2 \times K$ table, consider the logistic model

$$\text{logit}\pi_{ik} = \alpha + \beta x_i + \beta_k^Z,$$

where $x_1 = 1$ and $x_2 = 0$. Our focus parameter is often β. The sufficient statistics for $\left\{\beta_k^Z\right\}$ are $\{n_{+jk}\}$.

- When we treat n_{i+k} as fixed at each XZ combination in binomial sampling, small sample inference about β conditions on the row and column totals in each stratum.
- Conditional on the strata margins, an exact test uses $T = \sum_k n_{11k}$. The Cochran-Mantel-Haenszel test statistic

$$\text{CMH} = \frac{\left[\sum_k (n_{11k} - \mu_{11k})\right]^2}{\sum_k \text{var}(n_{11k})}$$

  is based on $\sum_k n_{11k}$.

Sufficient Statistic For Sparse Tables

The idea of conditional ML is welcome especially when the contingency tables are sparse. Consider a $2 \times 2 \times K$ table and the model

$$\operatorname{logit} P\left(Y_{ik} = 1\right) = \alpha_k + \beta x_i, \quad i = 1, 2,\ k = 1, ..., K,$$

where x_i is 0 or 1, and k means partial table k.

- In the extreme case where the row sums in each partial table are $(1, 1)$, the joint likelihood is

$$\begin{aligned} & \prod_{k=1}^{K} P\left(Y_{1k} = y_{1k}, Y_{2k} = y_{2k}\right) \\ = & \prod_{k=1}^{K} \left\{ \frac{\exp\left[y_{1k}\left(\alpha_k + \beta\right)\right]}{1 + \exp\left(\alpha_k + \beta x_i\right)} \times \frac{\exp\left[y_{2k}\alpha_k\right]}{1 + \exp\left(\alpha_k\right)} \right\} \\ = & \frac{\exp\left(\sum_k y_{1k}\beta\right) \exp\left[\sum_k \left(y_{1k} + y_{2k}\right)\alpha_k\right]}{\prod_{k=1}^{K}\left[1 + \exp\left(\alpha_k + \beta\right)\right] \prod_{k=1}^{K}\left[1 + \exp\left(\alpha_k\right)\right]}. \end{aligned}$$

- The sufficient statistics for $\{\alpha_k\}$ are $\{y_{1k} + y_{2k}\}$.

Conditional ML For Sparse Tables

Suppose that partial tables are independent of each other. Then

$$\begin{aligned} & P\left(Y_{11}=y_{11}, Y_{21}=y_{21} \cdots, Y_{2k}=y_{2K} \mid y_{1k}+y_{2k}=t_k, k=1,..,K\right) \\ = & \prod_{k=1}^{K} P\left(Y_{1k}=y_{1k}, Y_{2k}=y_{2k} \mid y_{1k}+y_{2k}=t_k\right) \\ = & \frac{\exp\left[\sum_{k=1}^{K} I\left(t_k=1, y_{1k}=1\right)\beta\right]}{\left[1+\exp\left(\beta\right)\right]^{\sum_{k=1}^{K} I\left(t_k=1\right)}}, \end{aligned}$$

which depends only on β. Its maximizer is the conditional MLE of β.

Even though $K \to \infty$, the number of parameters in the conditional likelihood is still 1. In contrast, the number of parameters in the likelihood is $K+1$.