Categorical Data Analysis Note
Edgard Maboudou

Chapter 3. Generalized Linear Models

1- Linear Models (LM)

- Better approach to analyze data to build models.
- Classical general linear models: N observations of a response variable $Y_1, Y_2, \ldots, Y_N$, k explanatory variables $X_1, X_2, \ldots, X_k$, unknown parameters $\beta_0, \beta_1, \ldots, \beta_k$
- 3 parts to the classical general linear model.
  1. Random component (Y's): identify the response variable (Y) and specify or assume a probability distribution for it, i.e distribution of Y_i's. Y_i is independent *normal* with constant variance σ^2. In that case, we are interested in
  $$E[Y] = \mu.$$
  2. Systematic component: specify what the explanatory or predictor variables are. These explanatory variables enter in a linear manner.
  $$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$
  3. Link: Specify the relationship between the mean or expected value of the random component (i.e. $E(Y)$) and the systematic components. Use the identity link $E(Y) = g(\mu) = \mu$.
  $$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$
- This model includes
  1. regression: both response and explanatory variables are continuous.
  2. Analysis of variance (ANOVA): response is continuous, explanatory variables are categorical.
  3. Analysis of covariance (ANCOVA): response is continuous, some explanatory variables are continuous and some are categorical.

2- Generalized Linear Models (GLM)

If Y_i's is nonnormal,

- Generalize the general linear model in 2 ways

  1. Y_i's are independent and have some distribution other than the normal distribution – any distribution within a class of distributions known as "exponential family of distributions".
  2. The relationship between the response (Y) and explanatory variables need not be simple ("identity"). For example, instead of

     $$Y = \alpha + \beta x,$$

     we can allow for transformations of Y

     $$g(Y) = \alpha + \beta x.$$

  3. Random component (Y's): identify the response variable (Y) and specify or assume a probability distribution for it, i.e distribution of Y_i's.
  4. Use a link other than the identity.
  5. We 'link' the two sides (random component and systematic component) using a link function $E(Y) = g(\mu)$

     $$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

- We will study 2 GLM.

  1. Y_i's are Bernoulli random variables (binary).

     $$P(Y_i = 1) = \pi \quad \text{and} \quad P(Y_i = 0) = 1 - \pi.$$

     This is the logistic regression. Instead of modeling $E(Y_i) = \pi$, we model the odds of success ($Y_i = 1$):

     $$\log\left(\frac{\pi}{1-\pi}\right),$$

     e.g. $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x.$
  2. Y_i's are independent Poisson random variables with mean μ. This is the loglinear model. Instead of modeling $E(Y_i) = \mu$, we model $\log \mu$, e.g. $\log \mu = \beta_0 + \beta_1 x$

a- Logistic Regression

- Many variables have only 2 possible outcomes, Bernoulli random variables.
- When we have n independent trials and take the sum of the Bernoulli trials, we have a Binomial distribution
- We are typically interested in π.
- We will consider models for π, which can vary according to some of the values of an explanatory variable(s) (i.e., $x_1, \ldots, x_k$).
- To emphasis that π changes with x's, we write $\pi(x)$
- Assume the model logit $(\pi(x)) = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x$.
- The random component is Bernoulli.
- Systematic component is a linear predictor such as

$$\alpha + \beta x$$

  which can be any real number and yield a π within (0, 1).
- The logit transformation is the link function.
- The model can be equivalently written as

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}.$$

- What does a plot of $\pi(x)$ vs x look like?

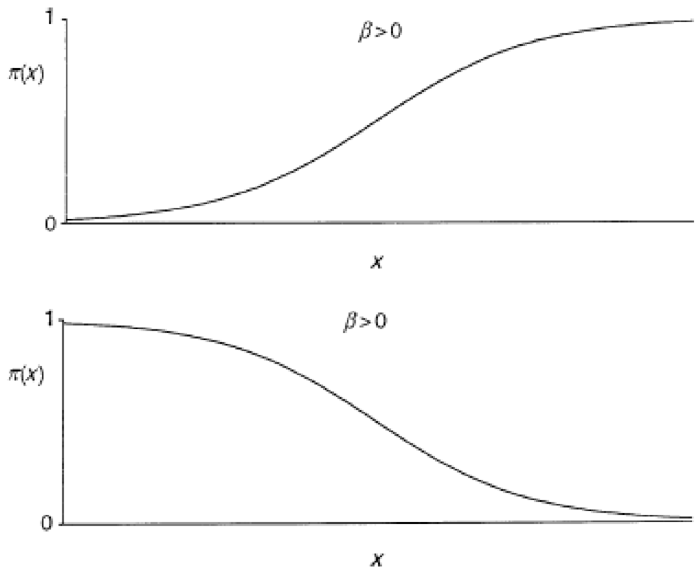

Figure 3.2. Logistic regression functions.

(a) e.g. $\alpha = 1$ and $\beta = .5 > 0$ there is a positive relationship between x and $\pi(x)$. (b) e.g. $\alpha = 1$ and $\beta = -.5 < 0$ there is a negative relationship between x and $\pi(x)$.

- The shape of the function is similar to an S. It is symmetric shape about $\pi(x) = .5$. $0 < \pi(x) < 1$.

Parameter Estimation

- Parameter estimates can be found from maximum likelihood estimation.
- The MLE of α and β are the values which maximize the log-likelihood.
- These estimates can only be found using numerically methods, so parameter estimates are found by many software packages.

b- Alternative Binary Links

- Many other link functions can be used to model binary data.
- These link functions use the "cumulative distribution function" or CDF.

i) Probit - based on the CDF of the standard normal

name comes from "probability unit" (Hubert, 1992)

- Random component: $Y \sim$ Bernoulli.
- Systematic component: $\alpha + \beta x$.
- Link function: probit transformation:

  $$\pi(x) = \phi(\alpha + \beta x)$$

  where $\phi(\ )$ is the CDF of a standard normal.
- $\phi^{-1}(\pi(x)) = \alpha + \beta x$
- $\phi^{-1}(\ )$ is often called the probit transformation and denoted by probit.
- So $\text{probit}(\pi(x)) = \alpha + \beta x$

ii) Complementary Log-Log - based on $1-$CDF of the Gumbel distribution

CDF of a Gumbel distribution is

$$F(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$$

for $-\infty < \mu < \infty$ and $\sigma^2 > 0$. So

$$1 - F(x) = 1 - \exp\left(-\exp\left(\alpha + \beta x\right)\right) \begin{cases} \alpha = \mu/\sigma \\ \sigma = -1/\beta \end{cases} .$$

- Random component: $Y \sim$ Bernoulli.
- Systematic component: $\alpha + \beta x$.
- Link function: complementary log-log transformation:

  $$\pi(x) = 1 - \exp\left(-\exp\left(\alpha + \beta x\right)\right)$$

  $$\log(-\log(1 - \pi(x))) = \alpha + \beta x$$

3- Statistical Inference and Model Checking

- One of the best things about GLMs is that they provide a unified approach to test model parameters, check goodness of fit, examine residuals, estimate parameters.

Wald and Likelihood Ratio Tests

- A hypothesis test commonly of interest is $H_0 : \beta = \beta_0$ vs. $H_a : \beta \neq \beta_0$.

a- Wald Test

$$Z = \frac{\widehat{\beta} - \beta_0}{SE(\widehat{\beta})}$$

$Z \sim N(0, 1)$ if H_0 is true.

b- Likelihood Ratio Test with $0 < \Lambda < 1$

$$\Lambda = \frac{\text{max likelihood when parameters satisfies } H_0}{\text{max likelihood when parameters satisfies } H_0 \text{ and } H_a}$$

- The actual test statistic used for a LRT is $-2\log(\Lambda)$, which is often denoted in Categorical Data Analysis (CDA) as G^2.
- Because for large n, it has an approximate χ^2 distribution.
- Often, in computer output, $-2\log(\Lambda)$ is not given directly.
- Instead, what is often given is the "null deviance" and the "residual deviance".
- These are $-2\log(\Lambda)$ statistics themselves, but for testing a different set of hypothesis.
- The $G^2 = -2\log(\Lambda)$ for a test $H_0 : \beta = 0$ vs. $H_a : \beta \neq 0$ is null deviance - residual deviance.

Remark: The word "deviance" is used because the statistics gives a measurement of how much the observed data "deviates" from the model's fit.

4- Models Residuals

- Pearson residuals can be calculated in a similar manner as described in ch.2,

$$e_{ij} = \frac{n_{ij} - \widehat{\mu}_{ij}}{\sqrt{\widehat{\mu}_{ij}}}.$$

- For a Poisson regression model, the Pearson residual is

$$e_i = \frac{Y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i}},$$

where Y_i is the i^{th} observed value and $\widehat{\mu}_i$ its predicted value.

- The standardized residual is

$$stdres = \frac{Y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i(1 - h_i)}},$$

where the leverage h_i is the i^{th} diagonal value of the hat matrix.

- The standardized residual is also called "adjusted Pearson residual", "adjusted residual", or "studentized residual".

- The standardized residual has a distribution that is closer to a standard normal than the Pearson residual.